WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/175514

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Identifying overlooked mechanisms for organic phosphorus and nitrogen acquisition in environmental bacteria

by

Andrew Robert Joseph Murphy

Submitted for the degree of Doctor of Philosophy by Published Research

THE UNIVERSITY OF WARWICK

School of Life Sciences

July 2022

Table of Contents

List of Figures

List of Tables

WARWICK
THE UNIVERSITY OF WARWICK

7/14/2022

Subject: Confirmation letter supporting the following statement made by Andrew Murphy on behalf of all his co-authors

To whom it may concern:

This is confirmation of the permission given by all co-authors concerning veracity of statements made by Andrew of his involvement in each paper submitted for the PhD. I have spoken with all authors and on their behalf, I confirm the statement made by Andrew are all accurate and a true representation of his inputs.

Yours faithfully

Professor E M H Wellington

e-mail: e.m.h.wellington@warwick.ac.uk
http://www2.warwick.ac.uk/fac/sci/lifesci/people/ewellington/

Professor E M H Wellington
School of Life Sciences
The University of Warwick
Coventry CV4 7AL United Kingdom
Tel: 00442476 523184
Fax: 00442476 523701
Email: e.m.h.wellington@warwick.ac.uk

Acknowledgements

I have worked in scientific research at the University of Warwick for 11 years and so I have many mentors, colleagues, friends and collaborators to thank and acknowledge. First of all, I would like to thank my primary supervisor Professor Elizabeth Wellington, with whom I worked for over 8-9 years on a variety of projects, for all her help, support, encouragement, and entertainment – you always manage to keep things interesting and exciting. I would also like to thank Professor Yin Chen, and Professor Dave Scanlan, for offering me continued employment, and their help, support, expertise and contributions over the past few years. Finally, I would like to thank Dr Ian Lidbury, who I worked with while he was a post-doc in the Wellington group, for all his efforts, without which I would not be submitting this thesis.

I'd also like to thank all the people I've met along the way, for all the discussions, activities, and events, including, in no particular order, Will, Helen, Phill, Hayley, Emma, Vicky, Luci, Tanya, Muriel, Greg, Chiara, Gemma, Siân, Rob, Séverine, Valentin, David, Archer, Lidija, James, Richard, Mussa, Linda, Eleonora, Holly, Rachel, Rebekah, Alberto, Rebecca, Nello, Rich, Branko, Michi, Rhiannon, Izzy, Macarius, Paula, Jingfan, Madhav, Liv, Jack, Alevtina, Zuzi, Gerald, Angus, and all past and present members of the Wellington, Chen, and Scanlan groups.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

As this is a submission for the degree of Doctor of Philosophy by published research any data generated and analysed within said published research that is not specifically mentioned as carried out by me should be assumed to be the work of one or more co-authors. Work specifically mentioned in this thesis not performed by me is outlined below:

- Binding affinities of AepX

Expression, purification, and microscale thermophoresis analysis of AepX, performed by William Cadman, Dr Nathan Adams, Dr Andrew Hitchcock and Dr Ian Lidbury at the University of Sheffield

- Construction of *Pseudomonas putida* BIRD-1 Δ*phoBR* (multiple sections)

This mutant was made by Dr Ian Lidbury

Parts of this thesis have been published by the author:

Lidbury, I.D.E.A., Murphy, A.R.J., Scanlan, D.J., Bending, G.D., Jones, A.M., Moore, J.D. et al. (2016) Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. *Environ Microbiol* **18**: 3535-3549.

Lidbury, I.D.E.A., Murphy, A.R.J., Fraser, T.D., Bending, G.D., Jones, A.M.E., Moore, J.D. et al. (2017b) Identification of extracellular glycerophosphodiesterases in *Pseudomonas* and their role in soil organic phosphorus remineralisation. *Sci Rep* **7**: 2179.

Lidbury, I.D.E.A., Borsetto, C., Murphy, A.R.J., Bottrill, A., Jones, A.M.E., Bending, G.D. et al. (2020) Niche-adaptation in plant-associated Bacteroidetes favours specialisation in organic phosphorus mineralisation. *ISME J*.

Jones, R.A., Shropshire, H., Zhao, C., Murphy, A., Lidbury, I., Wei, T. et al. (2021) Phosphorus stress induces the synthesis of novel glycolipids in *Pseudomonas aeruginosa* that confer protection against a last-resort antibiotic. *ISME J*.

Murphy, A.R.J., Scanlan, D.J., Chen, Y., Adams, N.B.P., Cadman, W.A., Bottrill, A. et al. (2021) Transporter characterisation reveals aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle. *Nature Comms* **12**: 4554.

Smith, A.F., Silvano, E., Päuker, O., Guillonneau, R., Quareshy, M., Murphy, A. et al. (2021) A novel class of sulfur-containing aminolipids widespread in marine roseobacters. *ISME J*. **15** (8) 2440-2453

Lidbury, I.D.E.A., Scanlan, D.J., Murphy, A.R.J., Christie-Oleza, J.A., Aguilo-Ferretjans, M.M., Hitchcock, A., and Daniell, T.J. (2022) A widely distributed phosphate-insensitive phosphatase presents a route for rapid organophosphorus remineralization in the biosphere. *Proceedings of the National Academy of Sciences* **119**: e2118122119.

Murphy, A.R.J., Scanlan, D.J., Chen, Y., Bending, G.D., Hammond, J.P., Wellington, E.M.H., and Lidbury, I.D.E.A. (2022) 2-Aminoethylphosphonate utilization in *Pseudomonas putida* BIRD-1 is controlled by multiple master regulators. *Environ Microbiol* **24**: 1902-1917.

Summary

All living organisms require phosphorus (P) and nitrogen (N), which exist in many forms, from inorganic phosphate (Pi) or ammonium, to multiple and diverse organic P (Po) and N molecules. As essential elements, bacteria possess multiple mechanisms that allow them to acquire P and N from their surrounding environment. Many environmental niches, including regions of the global ocean and the rhizosphere of crop plants, exhibit N and P limitation, and thus bacteria must possess mechanisms to overcome this nutrient limitation. Over the past seven years, I have contributed to, and led, research uncovering hitherto overlooked mechanisms for N and P acquisition and characterising others. This research has resulted in several published papers, including two first author papers within this field as outputs. In particular, I have contributed to research into the mechanisms displayed by rhizobacteria associated with oilseed rape (*Brassica napus*), including *Pseudomonas* and *Flavobacteria* spp., using analysis of proteomics data. The biological insights provided from this work were used to develop further foci of study, including the activity, functional role and distribution of a newly identified phosphatase, and the extracellular activity of Po-mineralising enzymes and its consequences for Pi availability. I have also led research into the identification, characterisation, and environmental distribution of 2-aminoethylphosphonate (2AEP) transporters, showing a novel 2AEP transporter to be the most transcribed phosphonate transporter in the global ocean. I followed this research with further investigation into the molecular mechanisms underpinning the regulation of 2AEP transport and degradation in *Pseudomonas putida* BIRD-1. By utilising my developed skill sets in both wet lab and bioinformatics techniques, I also contributed to papers investigating lipid remodelling, a mechanism by which bacteria reduce the quantity of phospholipids in their lipid membranes, in both environmental and pathogenic bacteria. Additionally, I have contributed to, and written, other papers within the field of environmental microbiology, which go beyond the scope of this thesis. As such, the works presented in this thesis represent a coherent body of work demonstrating my contribution to the research field.

Covering Document

Introduction

Phosphorus (P) is an essential element across all domains of life. It is one of the most abundant elements in cells, and forms an essential part of nucleic acids, phospholipids, and proteins. The preferred form of phosphorus for most organisms is inorganic orthophosphate (Pi) [1]. Across a number of environments Pi is often present or available at low concentrations, and can be the limiting nutrient in both marine [2, 3] and terrestrial environments [4, 5]. Phosphorus also exists in a variety of organic forms (Po), which form cellular structures including phospholipids, nucleic acids, and phospho-glycans and -proteins. Po possesses a number of bond structures including mono-esters (C-O-P), di-esters (C-O-P-O-C), and phosphonates (C-P) (Fig. 1). To acquire phosphorus from these organic sources, bacteria have evolved both specific and general mechanisms for the liberation and acquisition of Pi [1]. Additionally, bacteria have evolved adaptations that allow them to release Pi from insoluble Pi-metal complexes [6, 7], and many possess the ability to remodel their lipid bilayers, replacing phospholipids with other head groups [8, 9], and thus liberating Pi for other cellular processes (Fig. 2).

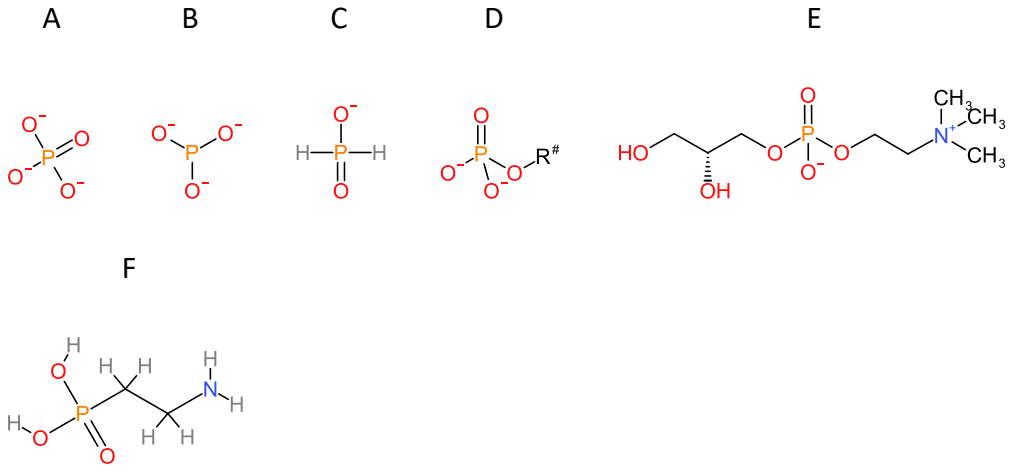


Figure 1. Structures of inorganic and organic phosphorus molecules. (A) phosphate, (B) phosphite, (C) hypophosphite, (D) phosphate monoester, (E) phosphate diester (glycerophosphorylcholine), (F) phosphonate (2-aminoethylphosphonate).

Pi is also the preferred P source of plants, including crops used for agriculture. While P can be relatively abundant in soils bioavailable Pi, as opposed to stable minerals or metal phosphates, is typically present at low concentrations [10]. Chemical P fertilisers are thus used to promote plant growth, an essential component of current agricultural systems [11, 12]. These fertilizers are derived from non-renewable rock phosphate, leading to concerns over the sustainability of existing agricultural practices [13, 14]. Additionally, only a small proportion of applied P fertiliser (up to 20%) is taken up by plants in the first year [15]. This leads to "P-loading" of soils, and subsequent run-off can lead to phosphate pollution of nearby waterways and estuaries [16]. This pollution leads to eutrophication, causing negative environmental impacts including oxygen depletion, reductions in diversity of ecosystems [17], and harmful algal blooms [18], with substantial economic impact. Reducing the quantity of phosphate fertiliser applied, by increasing the efficiency of its use, and unlocking otherwise unavailable P in soils would therefore reduce the severity of these problems.

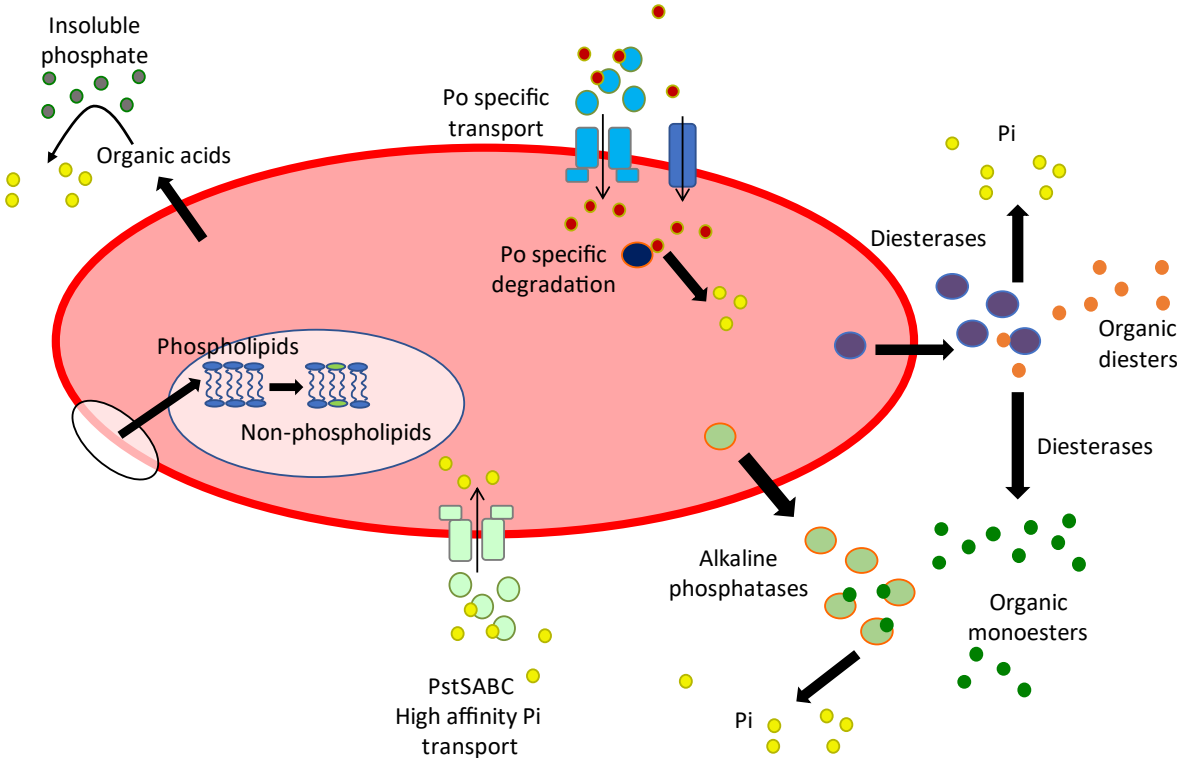


Figure 2. Bacterial mechanisms for phosphate acquisition. Mechanisms include the expression of high affinity phosphate transporters (PstSABC), mono- and di-esterases (e.g. alkaline phosphatases and glycerophosphodiesterases, respectively),

specific transport and degradation mechanisms for organic phosphate species (e.g. phosphonate systems), production of organic acids (e.g. gluconic acid) to liberate Pi from insoluble minerals, and lipid remodelling to replace phospholipids with non-phospholipids to reduce the phosphorus requirements of the cell.

Plant growth promoting rhizobacteria (PGPR) are defined as bacteria that increase plant yields. This occurs through a variety of mechanisms, including disease suppression, degradation of organic pollutants, and direct phytostimulation via hormone production [19, 20], but also nutrient provision, including phosphate mineralization [6, 21]. Oilseed Rape (*Brassica napus* - OSR) is an economically important crop in the United Kingdom [22]. Two abundant groups in the rhizosphere/root microbiome of OSR are *Pseudomonas* and *Bacteroidetes*, particularly *Flavobacteria* [23, 24], both of which are known to contain PGPR [25, 26], although *Pseudomonas* has been studied in much greater detail. As part of the Soil and Rhizosphere Interactions for Sustainable Agri-ecosystems (SARISA) project, Phosphorus cycling in the soil-microbe-plant continuum of agri-ecosystems, I worked on several objectives:

1) Characterise rhizosphere metagenome and root/rhizosphere metaproteome changes in response to soil P availability.
2) Assess the microbial taxonomic groups of oilseed rape rhizospheres during crop development and their impacts on rhizosphere P cycling.

To investigate the first objective, we acquired fundamental knowledge of the proteins *Pseudomonas* and *Flavobacteria* species synthesise for their abilities to acquire P when Pi is limiting. This involved characterising their (exo-) proteomic response to P limitation under laboratory conditions. Using this information, we developed metaexoproteomic techniques (which identify the total exoprotein of a community) to investigate the proteins that rhizosphere bacteria express *in situ* in the plant rhizosphere. Exoproteomics identifies secreted, periplasmic, and membrane proteins, and these protein groups constitute the primary mechanisms by which bacteria interact with their environment [27]. As such, exoproteomics can identify proteins related to niche specialisation, and provide insight into the ecological strategy bacteria undertake. Metaexoproteomics, while more challenging

[28], allows for the identification of bacterial mechanisms of nutrient mineralisation that are active in the rhizosphere, and reveals the key bacterial taxa that provide these functions. This is important as, while PGPR are thought to mineralise P in the rhizosphere [6, 21], recommended agricultural soil Pi concentrations (Olsen-P 15-25 mg L^{-1} – 168-263 µM) [29] in the UK are actually above the threshold concentration for the activation of the Pho regulon (4 µM) [30], the regulatory mechanism by which bacteria coordinate their response to Pi limitation [1]. It should be considered, however, that soil is a heterogeneous environment, with complex spatial distributions of both minerals and microbes, so localised Pi limitation can occur in the midst of general Pi abundance. As such, bacterial Pi liberation mechanisms should only occur if ambient Pi concentrations within the rhizosphere drop substantially below levels in the surrounding bulk soil – as commonly occurs within 2.5-3 mm of the root hair [31, 32].

Proteomic investigation of the Pi limitation response of *Pseudomonas* and *Flavobacteria*

As stated, *Pseudomonas* and *Flavobacteria* are important components of the OSR rhizosphere microbiome [23, 24]. As such we investigated three *Pseudomonas* rhizosphere isolates, *Pseudomonas putida* BIRD-1 (BIRD-1) (isolated from garden soil) [33], *Pseudomonas fluorescens* SBW25 (SBW25) (isolated from sugar beet) [34], and *Pseudomonas stutzeri* DSM4166 (DSM4166) (isolated from Sorghum) [35]. Exoproteomics was used to determine the Pi depletion response of all three strains, and, additionally, the Pho regulon of *P. putida* BIRD-1 was determined by mutagenesis of *PhoBR*, the two-component master regulator that coordinates the response to Pi depletion, together with total protein proteomics. This work resulted in the publication of the paper 'Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria' in Environmental Microbiology [36] in 2016. At the time of publication Environmental Microbiology had an impact factor of 5.556, and this paper has received 51 citations, demonstrating the value of the work to the field. I contributed to this publication by performing the proteomics experiments, and initial analysis of the resulting

datasets. This work revealed a number of conserved mechanisms related to phosphate acquisition, but also that the *Pseudomonas* strains differed not only in their complement of Pi liberation/acquisition mechanisms, but also in their expression of shared mechanisms (Lidbury *et al*. 2016 Figure 3, page 43, Figure 4, page 43, and Table 1, page 44) [36]. For example, both DSM4166 and BIRD-1 possess pyroloquinoline quinone (PQQ)-dependent alcohol dehydrogenases related to organic acid production, but only DSM4166 expressed them in response to Pi limitation. Both DSM4166 and SBW25 synthesise PhoD, and it is the second most abundant exoprotein in DSM4166, whereas in SBW25 it is not abundant at all (ranked 355th in its exoproteome). Additionally, proteomic data from the BIRD-1 Δ*phoBR* mutant revealed imperfections in bioinformatic identification of *pho* regulon members (Lidbury *et al.* 2016 Table 2, page 47) [36] – some proteins predicted to be pho regulon members [33] were found not to be over-produced in response to Pi limitation, including a putative acid phosphatase (PPUBIRD1_2395). In contrast, some proteins that were not predicted to be pho regulon members were identified as such due to their expression profile, including an as yet uncharacterised porin (PPUBIRD1_1354). This study thus revealed insights into the mechanisms this genus uses to acquire P, and provided hints as to what mechanisms may be important in rhizosphere soil that may be relevant to increasing Pi availability in the rhizosphere.

A similar approach was taken to the analysis of the Pi limitation response of *Flavobacterium*. In addition to the type strain *Flavobacterium johnsoniae* DSM2064 (DSM2064), the rhizosphere isolates *Flavobacterium* sp. F52 (F52) [37] and *Flavobacterium* sp. L0A5 (Cytryn, unpublished data) were donated by the Cytryn laboratory, Agricultural Research Organisation, Israel. I also isolated several *Bacteroidetes* strains from the rhizosphere of OSR from two field sites, one at the Warwick Crop Centre, Wellesbourne campus, University of Warwick, and one at Sonning Farm, University of Reading. In addition to proteomic analysis of the response to Pi limitation, transcriptomic analysis was also performed. This work resulted in the publication of the paper 'Niche-adaptation in plant-associated *Bacteroidetes* favours specialisation in organic phosphorus mineralisation' in the

ISME Journal [38] in 2020. At the time of publication, the ISME Journal had an impact factor of 10.302, and the paper has since received 8 citations and attracted almost 5000 online reads. In addition to the proteomic experiments and analysis, I also suggested and led the protein fractionation work using size exclusion chromatography as well as affinity purification to identify novel phosphatases in DSM2064. This was performed with the aid of Dr Mussa Quareshy who helped me with the equipment used in this experiment. Following fractionation, I performed colorimetric enzyme assays using p-Nitrophenyl Phosphate (PnPP) as the artificial substrate, identifying phosphatase-active protein fractions. Proteomics was performed on these fractions of reduced complexity in order to look for proteins whose abundance was positively correlated with activity (Lidbury *et al*. 2021, Figure S6, page 155) [38]. As such, I demonstrated increased independence as a researcher in my approach to this work. Importantly, this work identified a unique phosphatase PafA [38] in *Flavobacterium* and laid the foundation for future research, which resulted in Dr. Lidbury receiving a BBSRC-funded Discovery fellowship, and the publication of a recent paper I co-authored in PNAS [39].

These proteomic and transcriptomic datasets revealed diverse characterised and novel members of the *Flavobacterium pho* regulon. *Flavobacterium* possess several phosphatases, including divergent members of the PhoA and PhoX families, and the aforementioned Pi-insensitive phosphatase, PafA. Additionally, a few SusCD proteins (for starch utilisation system), hereafter termed PusCD (phosphorus utilisation system), were upregulated in response to Pi limitation (Lidbury *et al.* 2021 Figure 3, page 62) [38]. These proteins are outer membrane transporters, which, in characterised systems, transport polysaccharides allowing for intracellular degradation [40-42]. These PusCD systems, some of which exist in operons with the identified phosphatases, as well as putative phosphonate degrading enzymes, suggest a role in the breakdown of organic phosphate molecules, although further research is required to elucidate their roles. Interestingly, bioinformatic analysis revealed an expansion of the repertoire of these *pusCD* systems in the genomes of rhizosphere associated *Bacteroidetes* strains relative to aquatic or pathogenic strains (Lidbury *et al*. 2021 Figure 5, page 64, and Figure 6, page 65) [38].

In order to transfer these insights on global responses to Pi limitation in these important rhizosphere phyla into an understanding of the processes and mechanisms occurring within the OSR rhizosphere, I further developed the meta-exoproteomic extraction method used by Johnson-Rollings *et al*. [43] to extract protein from the rhizosphere of OSR grown under field conditions at Sonning Farm, University of Reading. A revised version is currently in-review in the journal mSystems after provisional acceptance following minor corrections, and so does not fall within the works to be assessed for this PhD by published research, however a version is available at BioRXiv [44]. This work demonstrates the validity of the approach taken in terms of determining the proteomic response to P limitation under controlled conditions prior to applying the knowledge gained to a field setting.

Application of proteomic insights drove further research avenues

The data generated from the proteomics experiments resulted in the development of several new avenues of inquiry. One of the most abundant exoproteins expressed by DSM4166 in response to P limitation was a putative glycerophosphodiester phosphodiesterase (GlpQ) [36, 45]. Additionally, SBW25 possesses a distant homolog of this putative GlpQ which, while not abundant, was also responsive to P limitation [36, 45]. These enzymes cleave the phosphate diester bond found in lipid head groups, including glycerophosphorylcholine (GPC), resulting in the release of *sn*-glycerol-3-phosphate (G3P) or phosphocholine (PC). The fact that the putative GlpQ appeared to be an exoenzyme therefore suggested that this degradation of GPC, and subsequent release of Pi from the breakdown products, could occur extracellularly, and thus supported growth of non-GPC degrading organisms, potentially including plants (Lidbury *et al*. 2017 Figure 3, page 74) [45]. The investigation into these putative GlpQ enzymes resulted in the publication 'Identification of extracellular glycerophosphodiesterases in *Pseudomonas* and their role in soil organic phosphorus remineralisation' in Scientific Reports in 2017 [45]. At the time of publication, Scientific Reports had an impact factor of 4.61, and the paper has since received 20 citations. I designed and made the plasmid constructs for the knockout mutants used in this work, and

additionally performed and analysed the growth curves and competition experiments that collectively demonstrate the extracellular release of Pi from GPC in a *glpQ* dependent manner.

Proteomics data were also used to identify the alkaline phosphatases of *Flavobacterium johnsoniae* DSM 2064, discussed earlier. I also contributed to work on the environmental abundance of PafA, a novel phosphomonoesterase which is both functional and expressed under high Pi conditions. This enzyme is required for growth on phosphorylated carbon sources and is constitutively expressed by *F. johnsoniae*. I compared the environmental abundance of *pafA* with that of other alkaline phosphatases (specifically *phoA*, *phoX*, and *phoD*), in addition to quantifying the abundance by phylogenetic origin, using the TARA oceans dataset (Lidbury *et al*. 2022 Figure 5, page 86) [46]. *pafA* was most abundant in the polar oceans in both the metagenome and metatranscriptome, with the majority of sequences falling within the class *Flavobacteriia*. This work was published in the Proceedings of the National Academy of Sciences of the United States of America (PNAS) with the title "A widely distributed phosphate-insensitive phosphatase presents a route for rapid organophosphorus remineralization in the biosphere" in 2022 [39], with 2 citations and over 1600 downloads at the time of writing. At the time of publication, PNAS had an impact factor of 11.2.

An additional avenue of inquiry, which forms the main part of my submitted works, came from the observation that the 2-aminoethylphosphonate (2AEP)-pyruvate transaminase (PhnW) and the phosphonoacetaldehyde hydrolase (PhnX) were both upregulated in response to Pi limitation in BIRD-1, but that no characterised 2AEP transporter was present on the genome of BIRD-1 [36], despite our work identifying several candidates. These observations resulted in me beginning a large project investigating 2AEP utilisation in soil and marine bacteria, which I developed into two studies, (alongside academic support in the form of meetings on a monthly or less frequent basis). I lead the research in the areas of aminoethylphosphonate transport and regulation characterisation, and from 2019 onwards no PDRA was employed on the project.

Characterization of aminoethylphosphonate transporters in *Pseudomonas*

Phosphonates are organic phosphorus molecules that possess a direct carbon (C) – P bond, unlike the more common phosphate ester bond (C – O (oxygen) – P) [47]. This bond is resistant to lysis, and cannot be degraded by common mechanisms to liberate Po, such as alkaline phosphatases [48, 49]. The P in phosphonates exists in a different redox state (3+) compared to Pi and other Po (5+), and must be oxidised as part of phosphonate catabolism in order to liberate Pi [47]. 2AEP is considered to be the most abundant naturally occurring phosphonate [50], although abundant sources of methyl-, hydroxymethyl-, and hydroxyethyl-phosphonate (MP, HMP, and HEP, respectively) are also known [51, 52]. A number of phosphonate degradation mechanisms have been characterised, including the C-P lyase, which can liberate Pi from a variety of phosphonates in a PhoBR (i.e. P limitation) dependent manner [50]. Additionally, there are two 2AEP-specific systems that begin with PhnW, the aforementioned PhnWX [53, 54], and the PhnW - phosphonoacetaldehyde dehydrogenase (PhnY) – phosphonoacetate hydrolase (PhnA) system [55-57]. Some members of the phosphonate dioxygenase (PhnY* - unrelated to PhnY described above) – phosphohydrolase (PhnZ) system degrade 2AEP [58, 59], while others degrade MP and HMP [60, 61]. Finally, a (R)-1-hydroxy-2-aminoethylphosphonate ammonia lyase (PbfA) can be found in some, but not all, PhnWX and PhnWAY operons [62].

Despite the diversity of degradation mechanisms, only two phosphonate transporters, both ATP-binding cassette (ABC) systems, which consist of a periplasmic substrate-binding protein, a transmembrane permease, and an ATP-binding protein, have been characterised. The first, termed PhnCDE is part of the C-P lyase operon and capable of transporting multiple species of phosphonate [63, 64], receiving by far the bulk of research interest. The second, termed PhnSTUV was found in an operon together with PhnWX, and is specific for 2AEP [53]. No homologs of either transporter are present in the genome of BIRD-1, however an ABC transporter, hereafter aminoethylphosphonateXVW (AepXVW), was found to be synthesised in response to Pi limitation; identified in our initial proteomics dataset [36]. AepXVW has a homolog whose genes are localised in the *phnWAY*

operon in *Sinorhizobium meliloti* [57]. Additionally, two alternative substrate-binding proteins of distinct ABC transporters were identified based on bearing some similarity to either PhnD (hereafter $PhnC_2D_2E_2$ – similarity of substrate binding protein 25%, $5e^{-13}$), or sharing the pfam domain 13343 present in AepX (hereafter AepSTU).

I initially designed constructs to knockout these transporters, both individually and in parallel, and assayed their ability to grow on 2AEP (Table 1). Δ*aepXVW* demonstrated a significantly reduced growth rate on 2AEP as the sole P source, while neither other mutant differed from the wild-type. However, the triple mutant Δ*aepXVW*:Δ*phnC2D2E2*:Δ*aepSTU*, showed a further reduced growth rate compared to both Δ*aepXVW* and Δ*aepXVW*:Δ*phnC2D2E2*, which did not differ, when grown on 2AEP as the sole P source, indicating that AepSTU, but not PhnC2D2E2, can transport 2AEP. Complementation of Δ*aepXVW* using *pBBR1MCS*-km restored full growth to the mutant. Interestingly, none of these mutants showed any defect when grown on 2AEP as the sole nitrogen (N) source, suggesting the existence of another 2AEP transporter in BIRD-1.

Table 1. Putative 2AEP transporter knockout mutant constructs in P. putida BIRD-1

Mutant	Locus tags	Description
Δ*phnC2D2E2*	PPUBIRD1_0873-PPUBIRD1-0876	Closest blastP match to the characterised PhnCDE, no impact on growth on 2AEP
Δ*aepXVW*	PPUBIRD1_4925-PPUBIRD1_4927	Closest blastP match to the *Sinorhizobium meliloti* ABC transporter found as part of a *phnWAY* operon
Δ*aepSTU*	PPUBIRD1_3891-PPUBIRD1_3895	Substrate binding protein shares a common pfam domain with AepX

Δ*aepP*	PPUBIRD1_3129	Upregulated in membrane enrichment proteomics when BIRD-1 is grown on 2AEP as the sole N source
Δ*aepXVW*:Δ*phnC2D2E2*	See above	Double mutant
Δ*aepXVW*:Δ*phnC2D2E2*:Δ*aepSTU*	See above	Triple mutant
Δ*aepXVW*:Δ*aepP*	See above	Double mutant
Δ*aepXVW*:Δ*aepP*:Δ*aepSTU*	See above	Triple mutant

In order to identify the transporter responsible for growth on 2AEP as a N source, I grew the Δ*aepXVW* mutant on 2AEP as the sole N source and performed proteomics on the cultures. The substrate-binding protein (locus tags; PPUBIRD1_4628 and PPUBIRD1_4654) of two candidate ABC transporters, both annotated as branched chain amino acid transporters (COG ID COG0683), were identified as upregulated. However knockout mutants, including a double mutant, showed no phenotypic difference to the wild-type. As such, I decided to perform membrane enrichment proteomics to preferentially identify membrane proteins enriched when grown on 2AEP as the sole N source. This successfully identified PPUBIRD1_3129, hereafter AepP for aminoethylphosphonate permease, as well as PPUBIRD1_3128, as being upregulated only when 2AEP was the sole N source, not in the absence of 2AEP, nor when 2AEP was the sole P source – though AepP was detected in this condition. Knockout mutants of these two genes were constructed and, while no phenotype was identified for PPUBIRD_3128, Δ*aepP* was unable to grow on 2AEP as the sole N source, with the wild-type phenotype restored by complementation. Additionally, Δ*aepXVW*:Δ*aepP* showed a further growth defect compared to Δ*aepXVW* when grown on 2AEP as the sole P source, and Δ*aepXVW*:Δ*aepP*:Δ*aepSTU* was unable to grow under this condition. As such, both AepP and AepSTU contribute to the growth on 2AEP as the sole P source that occurs in the absence of AepXVW (Murphy *et al*. 2021 Figure 1, page 92) [65].

Taken together, this demonstrates BIRD-1 possesses three transporters capable of acting on 2AEP, with AepXVW being the main transporter responsible for

the utilisation of 2AEP as the sole P source, and AepP being the sole transporter responsible for the utilisation of 2AEP as the sole N source. Bacteria possessing multiple differentially synthesised transporters for the same molecule is not unprecedented – indeed the transport of glycerol-3-phosphate (G3P) by *Escherichia coli* provides an interesting parallel. *E. coli* can use G3P as a source of either P or C, and uses the UgpABCE ABC transporter to acquire G3P as the sole P source [66], and the G3P/Pi antiporter GlpT (related to AepP – 26% identity, $4e^{-22}$, pfam 07690) to transport G3P as the sole C and P source [67, 68]. These transporters are inversely stimulated by internal cellular Pi concentrations, with UgpABCE showing increased activity under low internal Pi concentrations, and GlpT showing increased activity under high internal Pi concentrations [69]. The actions of these transporters allows internal Pi concentrations to be maintained within a range – although this concentration does increase in response to UgpABCE activity [69]. Mutants of *phoU,* the product of which transfers the (unknown) signal of low Pi concentrations from PstSCAB, the high affinity Pi transport system, to PhoR, show constitutive de-repression of the *pho* regulon. These mutants typically show poor growth until compensatory mutations in *pstSCAB* (or *phoBR*) occur [70]. Taken together with the observation that internal phosphate concentrations in response to G3P metabolism have an upper limit [69], this suggests that excessively high internal phosphate concentrations are deleterious to the cell. As such, during growth on organophosphorus, with high higher Pi:C or Pi:N stoichiometry, as C or N sources, a mechanism for the efflux of excess Pi may be essential.

Characterisation of 2AEP transport in *Stappia stellulata* DSM 5886

Marine *Alphaproteobacteria* are capable of using 2AEP in a Pi-independent manner, using PhnWAY [71]. Therefore, the related model marine bacterium *Stappia stellulata* DSM 5886 was chosen to identify the transporter responsible for this phenotype, due to possession of genes likely encoding AepXVW and PhnWAY. In addition, *S. stellulata* encodes the C-P lyase and two homologs of PhnCDE – one similar to the characterised *E. coli* PhnD, and one similar to that found in SAR11. Due to the absence of established genetic methods in *S. stellulata*, proteomic analysis was used to identify which transporters were upregulated in response to

growth on 2AEP. Interestingly, and in contrast to BIRD-1, AepXVW was upregulated in response to growth on 2AEP as either the sole P or sole N source, but neither PhnCDE homologs were upregulated (Murphy *et al*. 2021 Figure 3, page 96) [65]. PhnWAY was also upregulated in response to growth on 2AEP, in contrast to C-P lyase components, which were not upregulated. This suggested that, in *S. stellulata*, AepXVW is the sole transporter responsible for 2AEP utilisation. The heterologous expression of *aepXVW*^{Stappia} in BIRD-1 Δ*aepXVW*:Δ*aepP*:Δ*aepSTU* was able to restore growth on 2AEP as the sole P source, but not the sole N source, demonstrating it functions as a 2AEP transporter (Murphy *et al*. 2021 Figure 2, page 95) [65]. Both BIRD-1 and *S. stellulata* exported Pi when grown on 2AEP as the sole N (or sole N and P) source, due to the greater cellular requirement for N over P (Murphy *et al*. 2021 Figure S1, page 161, and Figure S6, page 165) [65]. It is possible that they use different mechanisms, as AepP is closely related to GlpT (Murphy *et al.* 2021 Figure S4, page 163) [65], a glycerol-3-phosphate:Pi antiporter, and so could function as a 2AEP:Pi antiporter, thus explaining this phenotype and the failure of *aepXVW*^{Stappia} to complement its function. However, we did not attempt to experimentally verify the role AepP plays in Pi export. *S. stellulata* could use several phosphonates as sole P sources, indicating its C-P lyase (and PhnCDE transporters) are functional (Murphy *et al*. 2021 Figure S6, page 165) [65].

Abundance and expression of phosphonate transporters in the global ocean

To determine how widespread phosphonate utilisation is in the ocean, I analysed the Tara Oceans dataset to determine the abundance and subsequent transcription of genes encoding these newly identified 2AEP transporters, by normalising the substrate binding proteins read count to the median values of 10 single copy marker genes. In addition, I performed the same analysis on known phosphonate transporters as well as phosphonate degradation enzymes. This allowed me to compare the relative importance of the distinct phosphonate utilisation pathways. *phnS*, *aepS*, and *aepP* were of relatively low abundance throughout the global ocean (Murphy *et al*. 2021 Figure S12, page 173, and Figure 5, page 98) [65]. Meanwhile, *phnD* was abundant only in regions typified by P limitation, such as the Mediterranean Sea, and the North Atlantic Gyre. *aepX*, in

contrast, was abundant throughout the global ocean, regardless of region or sampling depth (Murphy *et al*. 2021 Figure 5, page 98) [65]. *phnD* abundance and transcription showed a negative relationship with Pi concentration, whilst *aepX* abundance showed a weak positive relationship with Pi concentration (Murphy *et al*. 2021 Figure 5, page 98) [65]. Next, I compared these gene abundances with R*, a measure of relative P:N limitation, calculated as the sum of standing stock nitrate plus nitrite concentrations - 16 × Pi concentration [72]. *phnD* gene and transcript abundance had a clear positive relationship with R*. On the other hand, *aepX* showed a negative relationship with R*, though its transcription showed no significant relationships (Murphy *et al*. 2021 Figure 5, page 98) [65]. Taken together, this indicates that expression of *aepX* is ubiquitous and independent of nutrient limitation, in contrast to *phnD*, which is expressed highly only in Pi deplete regions. As such, 2AEP mineralization, based on marker gene abundance and transcription, appears to be a ubiquitous process, and given that Pi efflux occurs under P replete conditions this suggests that Pi is liberated from 2AEP by degrading organisms and recycled into the surrounding water column.

To corroborate this finding, I performed equivalent analyses for the phosphonate degradation gene markers *phnJ* (part of the C-P lyase), *phnA*, *phnX*, *phnZ*, and *tmpB*, which functions on (R)-1-hydroxy-2-(trimethylammonio)-ethylphosphonate [(R)-OH-TMAEP], which is produced from (*N*) trimethyl-2-aminoethylphosphonate (TMAEP) oxidation [73]. Unsurprisingly, and corroborating previous research [74], *phnJ* showed similar patterns to *phnD*, as the transporter also forms part of the C-P lyase system (Murphy *et al*. 2021 Figure S9, page 168) [65]. Of the catabolic enzyme systems associated with *aepX*, the most abundant and highly transcribed was *phnA*, which was ubiquitous, similar to *aepX*. *phnX* was less abundant but also showed no relationship with Pi concentration (Murphy *et al*. 2021 Figure S9, page 168) [65]. A phylogenetic tree of *phnZ* homologs showed those capable of 2AEP degradation were almost absent from the dataset, whereas *phnZ* homologs related to MP degradation were abundant in Pi deplete regions, and showed a significant negative relationship between abundance and expression, and Pi concentration, suggesting that they function to liberate Pi (Murphy *et al*.

2021 Figure S11, page 171) [65]. *tmpB* homologs (which are phylogenetically related to *phnZ*, and identified with the same profile hmm), however, showed no such relationship, and were most abundant in the polar oceans (Murphy *et al*. 2021 Figure S11, page 171) [65], suggesting they function in the acquisition of C or N rather than P, and that sources of TMAEP are most abundant in these regions.

Phylogeny of 2AEP transporters within the global ocean

I investigated the phylogeny of AepX and AepP ORFs, in both genomes of sequenced bacterial strains, and in the TARA oceans dataset. The hmms were used to identify homologs in refseq in addition to the TARA metagenome/ metatranscriptome environmental sequences which were clustered at 0.8 similarity using CD-HIT and representative sequences included in the alignment, alongside homologs found in metagenome-assembled-genomes (MAGs) and single cell assembled genomes from oceanic sources. Trees were constructed using IQTree. AepX was found in distantly related taxa, including throughout the Proteobacteria, but also Firmicutes and Cyanobacteria (Murphy *et al*. 2021 Figure 4, page 97) [65]. The AepX contained several sub-clades, with *S. stellulata* and *P. putida* BIRD-1 sequences well separated. 2AEP degradation operons including *phnWAY*, *phnWX*, and *pbfA*, alongside as yet unidentified genes, were found in the immediate genomic neighbourhood of *aepX* in the majority of refseq sequences – the separation of *aepXVW* and *phnWX*, found in BIRD-1, appears rare. Interestingly, two well separated *aepX* homologs were found to replace *phnCDE* in the C-P lyase operon of *Vibrio nereis* DSM 19584 and *Laribacter hongkongensis* sp. HLHK9. Where possible, the genomic neighbourhoods of TARA sequences were also investigated, which showed a similar pattern, with 2AEP degradation operons frequently found (Murphy *et al*. 2021 Figure 4, page 97) [65]. Given that most sequences within the TARA database are short, however, this was only possible to do for a small number of sequences.

The majority of *aepX* abundance within the TARA metatranscriptomes, both meso- and epi-pelagic, occurred within sequences related to two clades of oligotrophic Alphaproteobacteria – though one of those clades also contained deltaproteobacterial sequences that appear to have been acquired through

horizontal gene transfer, and so the phylogeny of the organisms containing them is thus uncertain (Murphy *et al*. 2021 Figure 4, page 97) [65]. Copiotrophic *Roseobacter* clades contained much of the rest of the *aepX* sequences, particularly within the epipelagic waters. Interestingly, a significant fraction of epipelagic *aepX* was found within sequences related to Cyanobacteria. Analysis of the distribution of these sequences revealed them to be exclusively found within the Mediterranean Sea, a P-deplete region, suggesting that they function to acquire 2AEP as a P source within that environment, and the global pattern of *aepX* expression serves to obscure that relationship. Overall, though, the presence of *aepX* in oligotrophic organisms confirms the global pattern of expression discussed earlier.

A similar approach was taken for AepP. AepP was widely distributed through diverse groups, including *Gammaproteobacteria*, *Acidobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, and the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) group (Murphy *et al*. 2021 Figure S8, page 167) [65]. While AepP sequences were more divergent than AepX sequences, 2AEP degradation operons were again found within their genomic neighbourhoods in both refseq, and one environmental sequence, thus indicating that these divergent homologs were also involved in 2AEP uptake. Most of the abundance within both epipelagic and mesopelagic waters was found within two clades containing PVC sequences, with much of the rest found within a sister group to the *Gammaproteobacteria-Actinobacteria* clade.

2AEP transporter abundance and phylogeny within a soil metagenome

In addition to analysing the TARA oceans dataset I also performed equivalent analyses on a metagenome sequenced from an agricultural field site used for growing oilseed rape. This metagenome was sequenced for use as a template for rhizosphere metaexoproteomics within this field setting [44]. As this dataset is not on the same scale as the TARA oceans dataset, it was not included in the paper, but nonetheless represents an insight into phosphonate transport and degradation within soil systems. No difference in abundance between bulk and rhizosphere soil was found, so the sample types were combined for the purposes of abundance analyses. As in the ocean, *aepX* was the most abundant transporter (Fig.

3A), however *aepP* was also abundant. *phnD* was the least abundant transporter, though interestingly abundances of the two subtypes – two hmms were required to capture *phnD* sequences as described above - were similar (data not shown). *phnA* was also the most abundant degradation mechanism, as in the TARA dataset, with both *phnJ* and *phnX* less abundant (Fig. 3B).

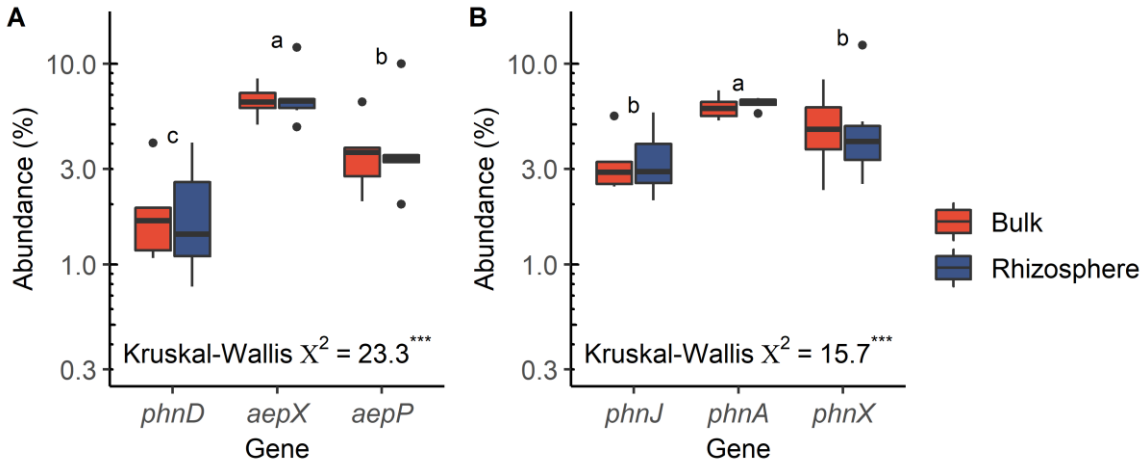


*Figure 3. The abundance of phosphonate transporters (**A**) and degradation enzymes (**B**) within a soil metagenome sequenced from bulk and rhizosphere agricultural soil growing oilseed rape (Brassica napus) plants near the University of Reading. Kruskal-Wallis X^2 comparisons between combined bulk and rhizosphere samples are plotted. Significant differences between gene abundance (by post-hoc Dunn's test with Holm's correction) are represented by individual letters.*

Environmental sequences for *aepX* (Fig. 4) and *aepP* (Fig. 5) were also included in phylogenetic trees to determine their identity. Abundances were normalised to the median abundance of 10 single copy marker genes, as in the analysis of the TARA metagenome, and sequences were clustered with CD-HIT at an identity of 80%, with representative sequences included in the alignments and subsequent trees. Within the environmental soil sequences, much of the abundance of both *aepX* and *aepP* fell within the *Pseudomonas* cluster, indicating that, within the site tested, *Pseudomonas* possesses the greatest potential for 2AEP transport. This is in contrast to oceanic *aepX* which is most abundant in oligo- and copio-trophic *Alphaproteobacteria* (Murphy *et al*. 2021 Figure 4, page 97) [65].

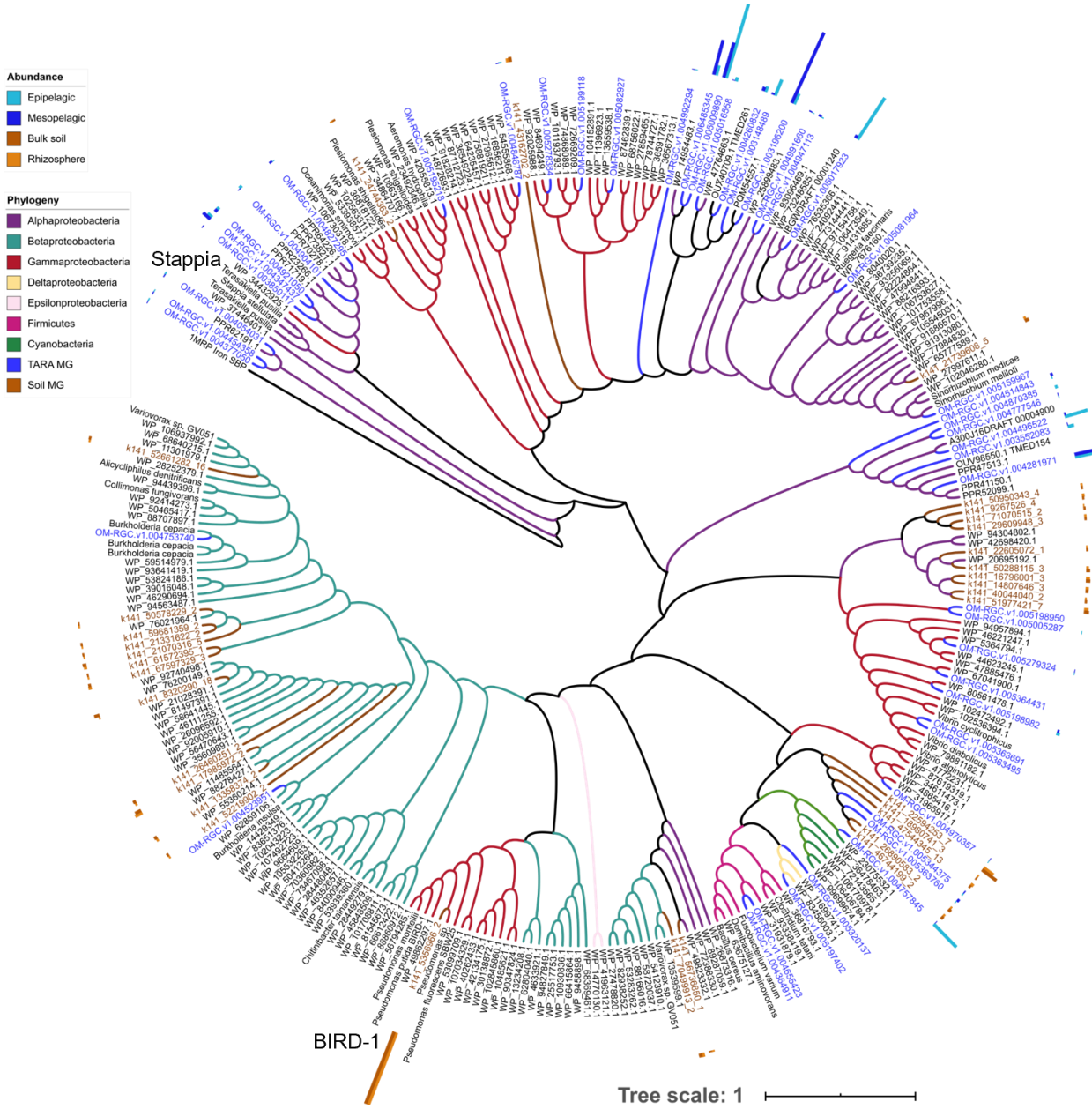


Figure 4. Abundance of aepX by phylogeny in the agricultural soil and the Tara oceans metagenomes. Phylogenetic assignment of reference sequences are shown by colour. Environmental soil sequences are coloured brown, Tara sequences are coloured blue. Relative abundance is shown by the bars. Dark brown represents bulk soil, light brown represents rhizosphere. Dark blue represents mesopelagic water, light blue represents epipelagic water.

Figure 5. Abundance of aepP by phylogeny in the agricultural soil and the Tara oceans metagenomes. Phylogenetic assignment of reference sequences are shown by colour. Environmental soil sequences are coloured brown. Relative abundance is shown by the bars. Dark brown represents bulk soil, light brown represents rhizosphere. Dark blue represents mesopelagic water, light blue represents epipelagic water.

Binding affinities of AepX

Determination for *S. stellulata* AepX binding affinity for various phosphonate species was performed by collaborators at the University of Sheffield. The main aim of this work was to confirm the hypothesis that AepX is a substrate-specific phosphonate binding protein, unlike PhnD. Briefly, binding affinities for a range of phosphonate molecules, including 2AEP, ethylphosphonate, aminomethylphosphonate, and MP were determined using microscale thermophoresis on purified protein labelled with a fluorescent dye. Microscale thermophoresis works by determining the proportion of a fluorescently labelled protein that is bound to an unlabelled ligand by comparison to the normalized fluorescence in both its bound and unbound state, thus allowing calculation of the affinity of the interaction [75]. AepX was found to be highly specific for 2AEP, with a K_d of 23±4 nM, whereas binding affinities to other phosphonates were at least four orders of magnitude higher. This contrasts with published values for *E. coli* PhnD, which, while it has a K_d of 50-100 nM for 2AEP, is much better at binding other phosphonates. This agrees with the role of PhnD in providing multiple substrates for the non-specific C-P lyase. Notably, however, given the differences between *E. coli* PhnD and the more abundant SAR11-related *phnD* sequences found within the TARA oceans dataset, it cannot be assumed that these are also most sensitive to 2AEP. Indeed it is likely that they bind MP and/or ethylphosphonate better, given their role in the degradation of these compounds in P deplete oceanic regions. Indeed, personal communication with collaborators in Sheffield and York indicate that key mutations in the binding sites of marine and soil PhnD homologs do indeed decrease affinity for 2AEP, whilst increasing affinity for alkylphosphonates. Taken together, this provides further evidence for the specificity of oceanic AepX proteins for 2AEP, although a role in binding related compounds such as 1-OH-2AEP cannot be ruled out, especially given the presence of *pbfA* in 2AEP transport and degradation operons.

Collectively, these observations led to a first author paper "Transporter characterisation reveals aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle" published in Nature Communications [65]. At the time of publication Nature Communications had an impact factor of 14.919, and the paper has thus far received 3 citations. Except for the microscale thermophoresis, and the construction of some of the mutants, all experimental work and data analyses was performed by me.

The regulation of 2AEP utilisation in the rhizobacteria *Pseudomonas putida* BIRD-1

Master regulators are required for efficient 2AEP utilisation

Regulation of 2AEP utilisation has been thought to be either under the control of PhoBR, the master regulator that coordinates responses to Pi scarcity in the environment, or otherwise substrate inducible. However for BIRD-1, the use of different 2AEP transporters in response to different nutrient stresses suggested alternative master regulators may be involved in the regulation of 2AEP transport and degradation. *Pseudomonas* has previously been demonstrated to use 2AEP as a source of both N and C [76]. As such, I investigated the role of three master regulators: PhoBR, NtrBC, which coordinates a response to nitrogen limitation, and CbrAB, which coordinates a response to carbon limitation (Murphy *et al*. 2022 Figure 1, page 104) [77]. This involved the construction of numerous mutants containing various combinations of master regulator disruption. While PhoBR was essential for the utilisation of 2AEP as the sole P source, neither Δ*ntrBC* and Δ*cbrAB* mutants eliminated growth on 2AEP as the sole N or C source, respectively, but instead demonstrated slower growth than the wild type (Murphy *et al*. 2022 Figure 2, page 106) [77]. Given previous research had demonstrated co-regulation of some genes by CbrAB and NtrBC, allowing compensation in the absence of one of the regulators, I also constructed a Δ*ntrBC*:Δ*cbrAB* double mutant, however the double mutant did not differ from the phenotype of the corresponding single mutant when grown on 2AEP as the sole N or C source (Murphy *et al*. 2022 Figure 2, page 106) [77].

Aminoethylphosphonate operons in BIRD-1 are substrate-inducible

After demonstrating the role of master regulators in 2AEP utilisation as their corresponding nutrient source, I sought to investigate additional regulatory mechanisms governing 2AEP utilisation in BIRD-1. Our first proteomics dataset [36], showed very low levels of PhnWX synthesis during P limitation, however this production was significantly increased in the presence of 2AEP (Murphy *et al*. 2021 Figure S3, page 162) [65]. As such I cloned the aminoethylphosphonate promoter regions (250 bp upstream of the start codon) of BIRD-1 into the pBIO1878 plasmid [78], forming the *lacZ* fusion plasmids pBIO-*phnWX*-pr, pBIO-*aepXVW*-pr, and pBIO-*aepP*-pr. These plasmids enable the detection of β-galactosidase activity, and thus serve as a proxy for transcriptional activation. These plasmids were then transformed into the wild type, as well as the master regulator single mutants. As not all mutants grow on 2AEP, or not at a similar rate, I decided to grow cultures under replete conditions prior to washing them in fresh media and exposing them to various nutrient conditions to determine whether β-galactosidase activity was induced. In the wild-type, absence of nutrient alone was not sufficient to induce β-galactosidase activity from either transporter promoter, or from the *phnWX* promoter, for any nutrient limitation (Murphy *et al*. 2022 Figure 4, page 109) [77]. In contrast, β-galactosidase activity was induced from the *phnWX* promoter in response to nutrient limitation plus 2AEP, for P, N, and C limitation, while activity was induced from the *aepXVW* promoter in response to P limitation plus 2AEP, and from the *aepP* promoter in response to N, and C limitation plus 2AEP (Murphy *et al*. 2022 Figure 4, page 109) [77]. As such, mechanisms for substrate induction at the aminoethylphosphonate operons must therefore exist in BIRD-1.

Interactions between master regulators and substrate-induction

Given the clear role played by the master regulators in growth on 2AEP, it is clear that substrate induction alone cannot compensate for their absence. Nonetheless, I investigated the relationship between the master regulators and

substrate induction using the promoter plasmids transformed into the master regulator mutants. Neither of the *aepXVW* or *phnWX* promoters induced expression (under Pi limitation plus 2AEP) in the Δ*phoBR* background (Murphy *et al*. 2022 Figure 4, page 109) [77], suggesting the substrate inducing protein coordinates with PhoB to drive expression from these promoters under conditions of Pi limitation. In contrast, substrate induction still occurred from both the *aepP* and *phnWX* promoter in the Δ*ntrBC* background in response to N limitation plus 2AEP, albeit to a significantly lesser extent than in the wild type (Murphy *et al*. 2022 Figure 4, page 109) [77], consistent with the observed phenotype that the Δ*ntrBC* mutant shows reduced growth on 2AEP as the sole N source compared to the wild-type. Similarly, substrate induction still occurred from both the *aepP* and *phnWX* promoter in the Δ*cbrAB* background in response to C limitation plus 2AEP, albeit to a significantly lesser extent than in the wild-type (Murphy *et al*. 2022 Figure 4, page 109) [77], again consistent with the observed phenotype that the Δ*cbrAB* mutant shows reduced growth on 2AEP as the sole C source compared to the wild-type.

AepR^{WX} is responsible for substrate induction at the *phnWX* promoter

The aminoethylphosphonate operons of BIRD-1 and other organisms, including copiotrophic marine *Alphaproteobacteria*, possess LysR homologs which we named AepR (Murphy *et al*. 2021 Figure 2, page 95) [65]. It has been suggested that these function as substrate-inducible regulators for 2AEP [57] in a manner similar to PhnR [79] and PalR [80] function for phosphonoacetate and phosphonoalanine, respectively. However, these three proteins, though all members of the LysR family, are not closely related (Murphy *et al*. 2021 Figure 2, page 95) [65]. To determine if they are indeed substrate-inducible regulators, I constructed a knockout mutant of the *aepR* gene present at the *phnWX* operon (hereafter *aepR^{WX}*). This mutant was still capable of growth on 2AEP as the sole P, N, or C source, though a substantial increase in the lag time was observed in all cases (Murphy *et al*. 2022 Figure 6, page 112) [77]. Next, I transformed Δ*aepR^{WX}* with the *lacZ* fusion plasmid pBIO-*phnWX*-pr. β-galactosidase activity was not detectable above baseline levels in the presence of 2AEP as the sole P or N source (Murphy *et al*. 2022 Figure 6, page 112) [77]. While above-background activity in

the presence of 2AEP as the sole C source was detected, it was significantly lower than the wild-type (Murphy *et al*. 2022 Figure 6, page 112) [77]. As such, I concluded that AepR^{WX} does indeed function as a substrate-inducible regulator, and so the homologous proteins AepR^{XVW} (encoded next to *aepXVW*) and AepR^{P} (encoded next to *aepP*) likely function similarly, although they are apparently incapable of substituting for AepR^{WX}. Given the absence of substrate induction in the Δ*aepR*^{WX} mutant, it is likely that AepR^{WX} functions as an activator, which interacts with the *phnWX* promoter directly in the presence of the substrate, although the details of its interactions with the master regulators remain undetermined (Murphy *et al*. 2022 Figure 7, page 113) [77]. Notably, however, the *phnWX* promoter is clearly active in the presence of 2AEP in response to both N and C limitation, but not in response to P limitation, in the absence of the corresponding master regulators. As PhoB interacts with σ^{70} whereas NtrC and CbrB interact with σ^{54}, this suggests that AepR^{WX} can interact with σ^{54}, but not σ^{70}, independently of the master regulator, albeit not as efficiently. Similarly, AepR^{WX} may help recruit the master regulators to the *phnWX* promoter, although it is not essential for this, explaining the difference in substrate induction and growth between the wild-type and Δ*aepR*^{WX} mutant. Further experimental evidence is required to determine these hypotheses.

2AEP is not the sole molecule responsible for substrate induction at BIRD-1's aminoethylphosphonate operons

To determine if 2AEP was responsible for substrate induction (as opposed to a breakdown product, as in the *lac* operon of *Escherichia coli*), I transformed Δ*phnWX* with the *lacZ* fusion plasmids pBIO-*phnWX*-pr, pBIO-*aepXVW*-pr, and pBIO-*aepP*-pr. This strain cannot grow on 2AEP as either the sole P or N [65], or C source (Murphy *et al*. 2022 Figure 2, page 106) [77], yet as it contains intact transporters it should remain capable of 2AEP import. As such, Δ*phnWX* allowed me to test the hypothesis that 2AEP alone is responsible for substrate induction. β-galactosidase activity was significantly lower than in the wild-type from all aminoethylphosphonate promoters under their respective nutrient conditions, however induction nonetheless still occurred (Murphy *et al*. 2022 Figure 5, page

111) [77]. As such, while 2AEP can interact with the substrate inducers of the three aminoethylphosphonate operons, under the experimental conditions tested, it is not solely responsible for substrate-induction. While the molecule that accounts for the remaining activity is unidentified, I hypothesise that phosphonoacetaldehyde is the most likely candidate, as it is common to several pathways of aminoethylphosphonate degradation, including 2AEP degradation via PhnWAY and PhnWX, and 1-hydroxy-2AEP degradation via PbfA [62]. *aepR* homologs are found together with both *phnWAY* and *phnWX* [65], and *pbfA* is also found in both *phnWAY* and *phnWX* operons [62], so a molecule common to all pathways makes a promising candidate (Fig. 6). Additionally, the final breakdown products of aminoethylphosphonate breakdown, alanine, acetaldehyde/acetate, and inorganic phosphate, are not specific to phosphonate catabolism, and therefore make unlikely candidates for substrate-induction.

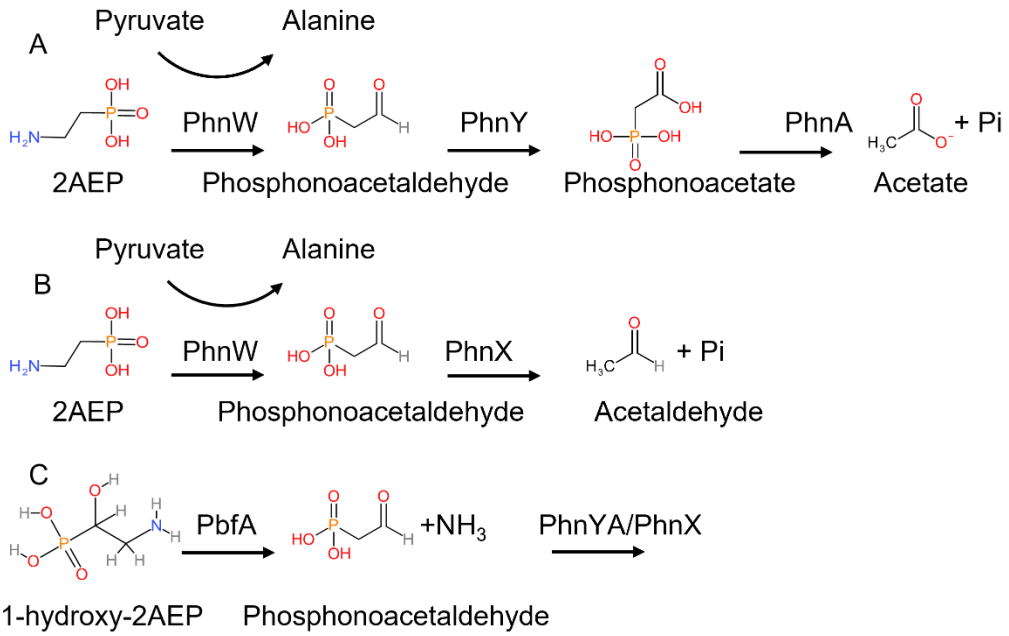


Figure 6. Pathways of 2AEP degradation by PhnWAY *(**A**), and* PhnWX *(**B**), and 1-hydroxy-2AEP degradation by* PbfA *(**C**). Phosphonoacetaldehyde is common to all pathways. N.B. phosphonoacetaldehyde produced via pathway C is then degraded via either pathway A or B.*

In summary, this second body of research demonstrated that BIRD-1 uses a dual-mechanism of regulation to efficiently utilise 2AEP as a nutrient source in

response to nutrient stress. This includes coordination between the master regulators as well as specific substrate-induction required to best use 2AEP. Together, these observations formed the basis for a first-author paper, entitled "2-aminoethylphosphonate utilisation in *Pseudomonas putida* BIRD-1 is controlled by multiple master regulators" published in Environmental Microbiology. This journal has an impact factor of 5.491 at the time of publication. I performed all of the experimental work associated with this paper, with the exception of the construction of the Δ*phoBR* mutant (performed by Dr Ian Lidbury), and performed all analysis. All authors provided feedback on initial manuscripts and contributed to experimental design.

Bacterial lipid remodelling

Concurrent with my work on 2AEP transportation and regulation, I worked in the laboratory of Professor Yin Chen, on projects related to lipid remodelling in environmental and pathogenic bacteria. Bacteria remodel their membrane lipids in response to Pi limitation, partially replacing phospholipids headgroups with non-phosphate containing headgroups, thus reducing the P needs of the cell. However, bacteria can possess constitutively synthesised non-P containing lipids, though the mechanisms of their synthesis, and their biological role, remain poorly understood. The copiotrophic marine alphaproteobacterium *Ruegeria pomeroyi* DSS-3 synthesises several non-P lipids, including glutamine and ornithine lipids. Additionally, a new sulfur containing aminolipid (SAL) was recently identified, published as "A novel class of sulfur-containing aminolipids widespread in marine roseobacters" in the ISME Journal [81]. I contributed to this paper by constructing a knockout mutant of the biosynthetic gene responsible for SAL production in both *R. pomeroyi* DSS-3 and the related bacterium *Phaeobacter inhibens* DSM 17395, as well as constructing complementation plasmids. Using lipidomics, SAL was shown to be present within the majority of Roseobacters, with the exception of *Dinoroseobacter shibae*, and was also absent in the closely related *S. stellulata* (Smith *et al*. 2021 Figure 4, page 125) [81]. I quantified the abundance of the biosynthetic gene, *salA*, in the TARA Oceans metagenome and metatranscriptomes, using the analyses I had developed during my work on phosphonate transporter

abundance (Smith *et al*. 2021 Figure 7, page 129) [65, 81]. This revealed *salA* to be both present and transcribed throughout the ocean, but especially in epipelagic waters. Interestingly, the *salA* mutant of *P. inhibens* showed a defect in biofilm formation relative to the wild-type. Particle-associated biofilm formation is an important part of the lifestyle strategy of marine Roseobacters, and so this may explain the constitutive synthesis of this lipid, however this does not preclude other roles for this lipid which remain to be elucidated.

Additionally, I contributed to a paper on non-P lipids in the opportunistic pathogen *Pseudomonas aeruginosa*. *P. aeruginosa* replaces phospholipids with the glycolipids monoglucosyldiacylglycerol and glucuronic acid-diacylglycerol in response to Pi limitation, a mechanism that is conserved across all *P. aeruginosa* isolates, environmental or clinical. The production of these glycolipids is of clinical relevance, as in the lungs of cystic fibrosis patients, *P. aeruginosa* experiences significant Pi limitation. In this work, the glycosyltransferases Agt1 and Agt2 were confirmed to be responsible for the biosynthesis of these glycolipids by mutagenesis, but also by heterologous expression in *E. coli*. I contributed to this paper by performing antibiotic sensitivity assays to polymyxin B, an antibiotic of last resort against *P. aeruginosa*, on these heterologously expressing strains. This demonstrated that biosynthesis of these glycolipids conferred increased resistance to polymyxin B (Jones *et al*. 2021 Figure 5, page 142) [82], consistent with data showing increased sensitivity in Δ*agt1* and Δ*agt2* mutants in *P. aeruginosa*. This work was published as "Phosphorus stress induces the synthesis of novel glycolipids in *Pseudomonas aeruginosa* that confer protection against a last-resort antibiotic" in the ISME Journal [82].

Conclusions

Building mechanistic insights into microbial elemental cycling is key to understanding the controls on marine and terrestrial productivity, including the agricultural systems societies depend upon. Pi can be a limiting nutrient in some areas of the global ocean, as well as in rhizosphere soil. Existing models of P cycling include the liberation of bioavailable Pi from organophosphorus molecules in response to Pi limitation. As the work presented here shows, however, Pi can be

released from organic P molecules that are used for other purposes, as in the case of 2AEP as a N or C source, or phosphorylated C utilisation by *Bacteroidetes* using PafA, both processes that appear important in the global ocean, and in agricultural systems. As such, the cycling of organophosphorus molecules occurs at least partially in a Pi independent manner, and models should be updated to reflect the flux between organic and inorganic forms of phosphorus that occurs in response to other factors than Pi limitations. Ideally this contribution should be quantified, although this is not a trivial matter.

There are numerous potential future directions of work relating to new scientific questions emerging from the 2AEP transport and regulation work. While we demonstrated AepX binds with high-affinity to 2AEP, we were unable to assay its binding to 1-OH-2AEP. Genes capable of 1-OH-2AEP degradation (*pbfA*) are found within many (though not all) *phnWX* and *phnWAY* operons adjacent to the *aepXVW* operon – including that of *Stappia stellulata*. As such, it seems likely that 1-OH-2AEP is a substrate of at least a fraction of AepX proteins. While BIRD-1 lacks *pbfA* it is possible it possesses other, uncharacterised mechanisms for 1-OH-2AEP degradation. An FAD dependent oxidoreductase, PPUBIRD1_3128, found in an operon with AepP, was upregulated during growth on 2AEP as the sole N source, but also during growth on 2AEP as the sole P source. It could therefore be an alternative mechanism for the utilisation of 1-OH-2AEP. Given that the phosphonate substrate binding proteins PhnD, PhnS, and AepX are unrelated, they represent the convergent evolution of phosphonate binding within the ABC transporter class. As such, mechanisms of binding by AepX and PhnS warrant further investigation.

Another avenue of investigation involves the mechanism of Pi export during growth on 2AEP as the sole N/C source in organisms that possess or don't possess AepP. As a member of the organophosphate:Pi antiporter family, closely related to the characterised glycerol-3-phosphate:phosphate antiporter, GlpT, it is likely that AepP functions in a similar fashion, although this has yet to be demonstrated experimentally. The mechanism employed by *S. stellulata*, which uses the AepXVW transporter when grown on 2AEP as either the sole P or N source, to export Pi is

unknown, however. A putative sodium:phosphate transporter is found within the aminophosphonate operon of *S. stellulata*, which provides a candidate, but this protein was not detected in any proteomics dataset. As a membrane component, however, it may require membrane enrichment to detect, similar to AepP, so it cannot be ruled out. Finally, the regulation of aminophosphonate regulation in marine bacteria remains to be determined. *aepR* homologs are found in proximity to the aminoethylphosphonate operons of copiotrophic *Roseobacter* species, however the most abundant sequences in the Tara Oceans dataset belonged to oligotrophic *Alphaproteobacteria* spp. While sequences with a full genetic neighbourhood are generally lacking for these organisms, it is notable that *aepR* sequences are absent from the few that exist. As such, the mechanism or mechanisms of regulation in these organisms cannot be assumed, and, given the ubiquitous transcription of *aepX*, it cannot be ruled out that it is transcribed constitutively.

Aims and objectives

The aims and objectives of the work discussed here can be summarized as follows:

1) Contribute to high quality scientific research projects
2) Develop as a self-directed researcher
3) Contribute to research outcomes with real world impact and applications

I believe the research I have presented here fulfils all of these aims and objectives.

1) I have detailed a number of papers published in high quality journals within the theme of environmental phosphate limitation adaptation within this covering document, including two first name author papers to which I was the principal contributor. Additionally I have contributed to other published papers within the domain of environmental microbiology, including one other first name author paper, full details of which can be found in Appendix I.
2) During the time I have worked in academic science I have developed as an independent scientist. I made contributions to a number of the

papers discussed here whilst working with Dr Ian Lidbury in his role as a post-doctoral Research Associate, but for the time period beginning in January 2019 until the present I have been engaged in more self-directed research, complemented with meetings with academics at a no more than monthly average frequency. Additionally, my other first name author paper was based on a short project I organized, and all analysis was performed by me.

3) Some of the research I have contributed to has the potential to lead to important real world applications. In particular, the PafA phosphatase has potential in terms of increasing the bioavailability of organic phosphate to plants in agricultural settings. This is due to both its regulation and activity being unaffected by ambient inorganic phosphate concentrations. As such, either a bioinoculant of *pafA* possessing *Bacteroidetes*, or direct application of PafA enzyme could release Pi from Po for agricultural plants to utilise.

Future perspectives

It is clear that much remains to be understood about the metabolism of organophosphates by microorganisms, and the impact of this metabolism on the biogeochemical cycling of essential elements. Insights into these processes are essential to predicting changes in biogeochemical cycles in response to climate change. Additionally, characterising novel phosphate acquisition mechanisms may allow for the development of tools to increase agricultural productivity without increasing demand for non-renewable resources such as rock phosphate. For example, the Pi-independent phosphatase PafA could prove able to liberate Pi for plant use even under high ambient Pi concentrations, in contrast to other phosphatases. Mechanistic understanding of microbial processes releasing nutrients in the rhizosphere is key to improving plant productivity through biological means, such as bioinoculants, or enzyme application.

References

1. Santos-Beneit, F., *The Pho regulon: a huge regulatory network in bacteria.* Front. Microbiol., 2015. **6**: p. 402.
2. Moore, C.M., et al., *Processes and patterns of oceanic nutrient limitation.* Nat. Geosci, 2013. **6**(9): p. 701-710.
3. Duhamel, S., et al., *Phosphorus as an integral component of global marine biogeochemistry.* Nat. Geosci, 2021. **14**(6): p. 359-368.
4. Goll, D.S., et al., *Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling.* Biogeosciences, 2012. **9**(9): p. 3547-3569.
5. Terrer, C., et al., *Nitrogen and phosphorus constrain the CO2 fertilization of global plant biomass.* Nat. Clim. Change, 2019. **9**(9): p. 684-689.
6. Rodríguez, H. and R. Fraga, *Phosphate solubilizing bacteria and their role in plant growth promotion.* Biotechnol. Adv., 1999. **17**(4): p. 319-339.
7. Otieno, N., et al., *Plant growth promotion induced by phosphate solubilizing endophytic Pseudomonas isolates.* Front. Microbiol., 2015. **6**: p. 745.
8. Sebastián, M., et al., *Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency.* ISME J, 2016. **10**(4): p. 968-978.
9. Carini, P., et al., *SAR11 lipid renovation in response to phosphate starvation.* Proc Natl Acad Sci USA, 2015. **112**(25): p. 7767.
10. Shen, J., et al., *Phosphorus dynamics: from soil to plant.* Plant Physiol, 2011. **156**(3): p. 997-1005.
11. Ha, S. and L.-S. Tran, *Understanding plant responses to phosphorus starvation for improvement of plant tolerance to phosphorus deficiency by biotechnological approaches.* Crit. Rev. Biotechnol, 2014. **34**(1): p. 16-30.
12. Vance, C.P., C. Uhde-Stone, and D.L. Allan, *Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource.* New Phytol., 2003. **157**(3): p. 423-447.
13. Cordell, D., J.-O. Drangert, and S. White, *The story of phosphorus: Global food security and food for thought.* Glob. Environ. Change, 2009. **19**(2): p. 292-305.
14. Daneshgar, S., et al., *The potential phosphorus crisis: resource conservation and possible escape technologies: a review.* Resources, 2018. **7**(2).
15. Russell, J.E., *Soil conditions and plants growth*. 2002: Daya Books.

16. White, P.J. and J.P. Hammond, *The sources of phosphorus in the waters of Great Britain.* J. Environ. Qual., 2009. **38**(1): p. 13-26.
17. Smith, V.H. and D.W. Schindler, *Eutrophication science: where do we go from here?* Trends Ecol. Evol., 2009. **24**(4): p. 201-207.
18. Wurtsbaugh, W.A., H.W. Paerl, and W.K. Dodds, *Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum.* WIREs Water, 2019. **6**(5): p. e1373.
19. Backer, R., et al., *Plant growth-promoting rhizobacteria: context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture.* Front. Plant Sci., 2018. **9**: p. 1473.
20. Gouda, S., et al., *Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture.* Microbiol. Res., 2018. **206**: p. 131-140.
21. Granada, C.E., et al., *Is phosphate solubilization the forgotten child of plant growth-promoting rhizobacteria?* Front. Microbiol., 2018. **9**: p. 2054.
22. Neik, T.X., M.J. Barbetti, and J. Batley, *Current status and challenges in identifying disease resistance genes in Brassica napus.* Front. Plant Sci., 2017. **8**.
23. Gkarmiri, K., et al., *Identifying the active microbiome associated with roots and rhizosphere soil of Oilseed Rape.* Appl. Environ. Microbiol., 2017. **83**(22): p. e01938-17.
24. Rathore, R., et al., *Crop establishment practices are a driver of the plant microbiota in winter oilseed rape (Brassica napus).* Front. Microbiol., 2017. **8**: p. 1489.
25. Miller, S.H., et al., *Biochemical and genomic comparison of inorganic phosphate solubilization in Pseudomonas species.* Environ. Microbiol. Rep., 2010. **2**(3): p. 403-411.
26. Kolton, M., et al., *The Flavobacterium genus in the plant holobiont: ecological, physiological, and applicative insights*, in *Microbial Models: From Environmental to Industrial Sustainability*, S. Castro-Sowinski, Editor. 2016, Springer Singapore: Singapore. p. 189-207.
27. Armengaud, J., et al., *Exoproteomics: exploring the world around biological systems.* Expert Rev. Proteomics, 2012. **9**(5): p. 561-575.
28. Abiraami, T.V., S. Singh, and L. Nain, *Soil metaproteomics as a tool for monitoring functional microbial communities: promises and challenges.* Rev. Environ. Sci. Biotechnol., 2020. **19**(1): p. 73-102.

29. Withers, P.J.A., et al., *Reducing soil phosphorus fertility brings potential long-term environmental gains: A UK analysis.* Environ. Res. Lett., 2017. **12**(6): p. 063001.
30. Wanner, B.L., *Phosphorus assimilation and control of the phosphate regulon.* Escherichia coli and Salmonella: Cellular and Molecular Biology, 1996. **1**: p. 1357-1381.
31. Kreuzeder, A., et al., *In situ observation of localized, sub-mm scale changes of phosphorus biogeochemistry in the rhizosphere.* Plant Soil, 2018. **424**(1): p. 573-589.
32. Santner, J., et al., *High-resolution chemical imaging of labile phosphorus in the rhizosphere of Brassica napus L. cultivars.* Environ. Exp. Bot., 2012. **77**: p. 219-226.
33. Roca, A., et al., *Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium Pseudomonas putida BIRD-1.* Environ. Microbiol., 2013. **15**(3): p. 780-794.
34. Naseby, D.C., et al., *Biocontrol of Pythium in the pea rhizosphere by antifungal metabolite producing and non-producing Pseudomonas strains.* J. Appl. Microbiol., 2001. **90**(3): p. 421-429.
35. Yu, H., et al., *Complete genome sequence of the nitrogen-fixing and rhizosphere-associated bacterium Pseudomonas stutzeri strain DSM4166.* J. Bacteriol., 2011. **193**(13): p. 3422-3423.
36. Lidbury, I.D., et al., *Comparative genomic, proteomic and exoproteomic analyses of three Pseudomonas strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria.* Environ. Microbiol., 2016. **18**(10): p. 3535-3549.
37. Kolton, M., et al., *Draft genome sequence of Flavobacterium sp. strain F52, isolated from the rhizosphere of Bell Pepper (Capsicum annuum L. cv. Maccabi).* J. Bacteriol., 2012. **194**(19): p. 5462-5463.
38. Lidbury, I.D.E.A., et al., *Niche-adaptation in plant-associated Bacteroidetes favours specialisation in organic phosphorus mineralisation.* ISME J, 2020.
39. Lidbury, I.D.E.A., et al., *A widely distributed phosphate-insensitive phosphatase presents a route for rapid organophosphorus remineralization in the biosphere.* Proc. Natl. Acad. Sci. U.S.A., 2022. **119**(5): p. e2118122119.
40. Glenwright, A.J., et al., *Structural basis for nutrient acquisition by dominant members of the human gut microbiota.* Nature, 2017. **541**(7637): p. 407-411.

41. Kappelmann, L., et al., *Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans.* ISME J, 2019. **13**(1): p. 76-91.
42. Cuskin, F., et al., *Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism.* Nature, 2015. **517**(7533): p. 165-169.
43. Johnson-Rollings, A.S., et al., *Exploring the functional soil-microbe interface and exoenzymes through soil metaexoproteomics.* ISME J, 2014. **8**(10): p. 2148-2150.
44. Lidbury, I.D.E.A., et al., *Meta-exoproteomics identifies active plant-microbe interactions operating in the rhizosphere.* bioRxiv, 2021: p. 2021.09.01.458574.
45. Lidbury, I.D.E.A., et al., *Identification of extracellular glycerophosphodiesterases in Pseudomonas and their role in soil organic phosphorus remineralisation.* Sci. Rep., 2017. **7**(1): p. 2179.
46. Villar, E., et al., *The Ocean Gene Atlas: exploring the biogeography of plankton genes online.* Nucleic Acids Res., 2018. **46**(W1): p. W289-W295.
47. Villarreal-Chiu, J.F., J.P. Quinn, and J.W. McGrath, *The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment.* Front. Microbiol., 2012. **3**: p. 19.
48. Wanner, B.L., *Molecular genetics of carbon-phosphorus bond cleavage in bacteria.* Biodegradation, 1994. **5**(3-4): p. 175-84.
49. Kononova, S.V. and M.A. Nesmeyanova, *Phosphonates and their degradation by microorganisms.* Biochemistry (Mosc), 2002. **67**(2): p. 184-95.
50. White, A.K. and W.W. Metcalf, *Microbial metabolism of reduced phosphorus compounds.* Annu. Rev. Microbiol., 2007. **61**: p. 379-400.
51. Metcalf, W.W., et al., *Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean.* Science, 2012. **337**(6098): p. 1104-1107.
52. Born, D.A., et al., *Structural basis for methylphosphonate biosynthesis.* Science, 2017. **358**(6368): p. 1336.
53. Jiang, W., et al., *Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of Salmonella typhimurium LT2.* J. Bacteriol., 1995. **177**(22): p. 6411.
54. Kim, A.D., et al., *The 2-aminoethylphosphonate-specific transaminase of the 2-aminoethylphosphonate degradation pathway.* J. Bacteriol., 2002. **184**(15): p. 4134.

55. Cooley, N.A., et al., *Phosphonoacetate biosynthesis: In vitro detection of a novel $NADP^+$-dependent phosphonoacetaldehyde-oxidizing activity in cell-extracts of a marine Roseobacter.* Microbiology, 2011. **80**(3): p. 335-340.
56. Agarwal, V., et al., *Structure and Function of Phosphonoacetaldehyde Dehydrogenase: The Missing Link in Phosphonoacetate Formation.* Chem. & Biol., 2014. **21**(1): p. 125-135.
57. Borisova, S.A., et al., *Genetic and biochemical characterization of a pathway for the degradation of 2-aminoethylphosphonate in Sinorhizobium meliloti 1021.* J. Biol. Chem., 2011. **286**(25): p. 22283-90.
58. McSorley, F.R., et al., *PhnY and PhnZ comprise a new oxidative pathway for enzymatic cleavage of a carbon-phosphorus bond.* J. Am. Chem. Soc., 2012. **134**(20): p. 8364-7.
59. van Staalduinen, L.M., et al., *Crystal structure of PhnZ in complex with substrate reveals a di-iron oxygenase mechanism for catabolism of organophosphonates.* Proc. Natl. Acad. Sci. U.S.A., 2014. **111**(14): p. 5171-6.
60. Sosa, O.A., J.R. Casey, and D.M. Karl, *Methylphosphonate oxidation in Prochlorococcus strain MIT9301 supports phosphate acquisition, formate excretion, and carbon assimilation into purines.* Appl. Environ. Microbiol., 2019. **85**(13): p. e00289-19.
61. Gama, S.R., et al., *An oxidative pathway for microbial utilization of methylphosphonic acid as a phosphate source.* ACS Chem. Biol., 2019. **14**(4): p. 735-741.
62. Zangelmi, E., Stanković, T., Malatesta, M., Acquotti, D., Pallitsch, K., and Peracchi, A., *Discovery of a new, recurrent enzyme in bacterial phosphonate degradation: (R)-1-hydroxy-2-aminoethylphosphonate ammonia-lyase.* Biochemistry, 2021.
63. Alicea, I., et al., *Structure of the Escherichia coli phosphonate binding protein PhnD and rationally optimized phosphonate biosensors.* J. Mol. Biol., 2011. **414**(3): p. 356-69.
64. Rizk, S.S., M.J. Cuneo, and H.W. Hellinga, *Identification of cognate ligands for the Escherichia coli phnD protein product and engineering of a reagentless fluorescent biosensor for phosphonates.* Protein Sci, 2006. **15**(7): p. 1745-51.
65. Murphy, A.R.J., et al., *Transporter characterisation reveals aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle.* Nat. Commun., 2021. **12**(1): p. 4554.

66. Brzoska, P., et al., *The pho regulon-dependent Ugp uptake system for glycerol-3-phosphate in Escherichia coli is trans inhibited by Pi.* J Bact, 1994. **176**(1): p. 15-20.
67. Hayashi, S., J.P. Koch, and E.C. Lin, *Active transport of L-alpha-glycerophosphate in Escherichia coli.* J. Biol. Chem., 1964. **239**: p. 3098-105.
68. Elvin, C.M., C.M. Hardy, and H. Rosenberg, *Pi exchange mediated by the GlpT-dependent sn-glycerol-3-phosphate transport system in Escherichia coli.* J. Bacteriol., 1985. **161**(3): p. 1054-8.
69. Xavier, K.B., et al., *Kinetic analysis by in vivo 31P nuclear magnetic resonance of internal Pi during the uptake of sn-glycerol-3-phosphate by the pho regulon-dependent Ugp system and the glp regulon-dependent GlpT system.* J. Bacteriol., 1995. **177**(3): p. 699-704.
70. Gardner, S. and W. McCleary, *Control of the phoBR Regulon in Escherichia coli.* EcoSal Plus, 2019.
71. Chin, J.P., J.P. Quinn, and J.W. McGrath, *Phosphate insensitive aminophosphonate mineralisation within oceanic nutrient cycles.* ISME J, 2018. **12**(4): p. 973-980.
72. Smith, A.F., et al., *Elucidation of glutamine lipid biosynthesis in marine bacteria reveals its importance under phosphorus deplete growth in Rhodobacteraceae.* ISME J, 2019. **13**(1): p. 39-49.
73. Rajakovich, L.J., et al., *A new microbial pathway for organophosphonate degradation catalyzed by two previously misannotated non-heme-iron oxygenases.* Biochem, 2019. **58**(12): p. 1627-1647.
74. Sosa, O.A., et al., *Phosphate-limited ocean regions select for bacterial populations enriched in the carbon-phosphorus lyase pathway for phosphonate degradation.* Environ. Microbiol., 2019. **21**(7): p. 2402-2414.
75. Jerabek-Willemsen, M., et al., *Molecular interaction studies using microscale thermophoresis.* Assay Drug Dev. Technol., 2011. **9**(4): p. 342-53.
76. Ternan, N.G. and J.P. Quinn, *Phosphate starvation-independent 2-aminoethylphosphonic acid biodegradation in a newly isolated strain of Pseudomonas putida, NG2.* Syst. Appl. Microbiol., 1998. **21**(3): p. 346-52.
77. Murphy, A.R.J., et al., *2-Aminoethylphosphonate utilization in Pseudomonas putida BIRD-1 is controlled by multiple master regulators.* Environ. Microbiol., 2022. **24**(4): p. 1902-1917.
78. Todd, J.D., et al., *DddW, a third DMSP lyase in a model Roseobacter marine bacterium, Ruegeria pomeroyi DSS-3.* ISME J, 2012. **6**(1): p. 223-226.

79. Kulakova, A.N., et al., *Structural and functional analysis of the phosphonoacetate hydrolase (phnA) gene region in Pseudomonas fluorescens 23F.* J Bacteriol., 2001. **183**(11): p. 3268.
80. Kulakova, A.N., et al., *Expression of the phosphonoalanine-degradative gene cluster from Variovorax sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate.* FEMS Microbiol. Lett., 2009. **292**(1): p. 100-106.
81. Smith, A.F., et al., *A novel class of sulfur-containing aminolipids widespread in marine roseobacters.* ISME J, 2021.
82. Jones, R.A., et al., *Phosphorus stress induces the synthesis of novel glycolipids in Pseudomonas aeruginosa that confer protection against a last-resort antibiotic.* ISME J, 2021.

Appendix I

environmental microbiology

Environmental Microbiology (2016) **18**(10), 3535–3549 doi:10.1111/1462-2920.13390

Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria

Ian D. E. A. Lidbury,[1]* Andrew R. J. Murphy,[1] David J. Scanlan,[1] Gary D. Bending,[1] Alexandra M. E. Jones,[1] Jonathan D. Moore,[2] Andrew Goodall,[3] John P. Hammond[3,4] and Elizabeth M. H. Wellington[1]

[1]*School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, West Midlands CV4 7AL, UK.*

[2]*The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK.*

[3]*School of Agriculture, Policy, and Development, University of Reading, Earley Gate, Whiteknights, Reading RG6 6AR, UK.*

[4]*Southern Cross Plant Science, Southern Cross University, Lismore NSW 2480, Australia.*

Summary

Bacteria that inhabit the rhizosphere of agricultural crops can have a beneficial effect on crop growth. One such mechanism is the microbial-driven solubilization and remineralization of complex forms of phosphorus (P). It is known that bacteria secrete various phosphatases in response to low P conditions. However, our understanding of their global proteomic response to P stress is limited. Here, exoproteomic analysis of *Pseudomonas putida* BIRD-1 (BIRD-1), *Pseudomonas fluorescens* SBW25 and *Pseudomonas stutzeri* DSM4166 was performed in unison with whole-cell proteomic analysis of BIRD-1 grown under phosphate (Pi) replete and Pi deplete conditions. Comparative exoproteomics revealed marked heterogeneity in the exoproteomes of each *Pseudomonas* strain in response to Pi depletion. In addition to well-characterized members of the PHO regulon such as alkaline phosphatases, several proteins, previously not associated with the response to Pi depletion, were also identified. These included putative nucleases, phosphotriesterases, putative phosphonate transporters and outer membrane proteins. Moreover, in BIRD-1, mutagenesis of the master regulator, *phoBR*, led us to confirm the addition of several novel PHO-dependent proteins. Our data expands knowledge of the *Pseudomonas* PHO regulon, including species that are frequently used as bioinoculants, opening up the potential for more efficient and complete use of soil complexed P.

Introduction

Phosphorus (P) is an essential macroelement for all living biota. In soil, microorganisms and plants compete for P. It is therefore essential that a sufficient amount of P is available for agricultural crops to sustain their yields. Plants acquire P as inorganic orthophosphate (Pi) from the soil solution (Vance *et al.*, 2003; White and Hammond, 2008). The concentration of Pi in the soil solution is controlled by chemical and biological processes which fix and release Pi through complex interactions between the soil, soil microorganisms and plant roots (Richardson *et al.*, 2009; Shen *et al.*, 2011). To overcome these limitations in agricultural systems, inorganic fertilizers, derived from non-renewable Pi rocks, are supplied to crops and pastures (Vance *et al.*, 2003; López-Arredondo *et al.*, 2014). Over 85% of mined P is used in food production (Heffer *et al.*, 2006) and consumption of this non-renewable resource could lead to a peak P scenario (akin to peak oil; Raven, 2008; Cordell *et al.*, 2009). It is, therefore, likely that there will be increasing pressures on Pi fertilizer availability and, consequently, cost in the future. These pressures will be exacerbated by increasing demand on food production systems as the human population increases and by fluctuation in oil prices (Cordell *et al.*, 2009). Inappropriate use of inorganic Pi fertilizers can also perturb the nutrient balance of natural ecosystems and reduce biodiversity (White and Hammond, 2008; 2009). Hence, it is desirable to increase the efficiency by which plants can access the many forms of unavailable P that reside within soil, thus, reducing the

*For correspondence. E-mail i.lidbury@warwick.ac.uk; Tel. +44 (0) 24 765 75874; Fax +44 (0)24 7652 2052.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

requirements for Pi fertilizer application (Richardson *et al*., 2009; Stutter *et al*., 2012).

Plant-growth promoting rhizobacteria (PGPR) are bacteria that can enhance crop yields through a variety of mechanisms, including P mobilization, and their identification has led to the notion that bacteria are an integral part of the plant-root interface (rhizosphere) (Lugtenberg and Kamilova, 2009; Miller *et al*., 2010). One advantage of utilizing P-liberating bacteria in agricultural systems is that they can have synergistic beneficial effects, for example, pathogen suppression (Vassilev *et al*., 2006). The majority of research into P mobilization by PGPR has focused on the solubilization of inorganic P through acidification of the surrounding soil via the release of organic acids, namely gluconic acid (Rodríguez and Fraga, 1999; Miller *et al*., 2010; Oteino *et al*., 2015). The genus *Pseudomonas* represents one soil bacterial group that is frequently associated with plant-growth promotion, including P solubilization (Miller *et al*., 2010). Various *Pseudomonas* strains can also degrade organic P compounds, such as phytate, phosphonates and phosphites (Ternan and Quinn, 1998; White and Metcalf, 2004; 2007). Three *Pseudomonas* strains*, Pseudomonas putida* BIRD-1, *Pseudomonas fluorescens* SBW25 and *Pseudomonas stutzeri* DSM4166 (hereafter, BIRD-1, SBW25 and DSM4166 respectively) are three examples of PGPR (Naseby *et al*., 2001; Hass and Keel, 2003; Preston, 2004; Yu *et al*., 2011; Roca *et al*., 2013). SBW25 inhabits the rhizosphere of Pea plants and is antagonistic towards the pathogen *Pythium ultimum* (Naseby *et al*., 2001), whereas DSM4166, an 'unusual' nitrogen-fixing bacterium, was isolated from a cultivar of *Sorghum nutans* (Yu *et al*., 2011). BIRD-1, a P-solubilising bacterium, has been previously utilized as a bioinoculant, since it can significantly improve the germination rates, growth and yields of various agricultural crops (Roca *et al*., 2013). BIRD-1 can remineralize Pi from the plant P-storage compound phytate, a major source of organic P in some soils (up to 50%) (Stutter *et al*., 2012). Furthermore, when BIRD-1 was used as an inoculant, phosphatase activity was greater in the rhizosphere compared with bulk soil (Roca *et al*., 2013). Although *Pseudomonas* have been implicated in plant-growth promotion, partially attributed to their effect on P mobilization, the precise mechanisms behind this process remain largely unknown.

The majority of Bacteria studied can undergo a physiological response to Pi depletion controlled by a two-component regulatory system (PhoBR) encoded by *phoBR*. PhoBR regulates a large set of genes (the PHO regulon) in response to low P concentrations (Baek and Lee, 2006; Monds *et al*., 2006; Su *et al*., 2007). The majority of studies have focused on specific functions/mechanisms that are regulated by PhoBR, for example, the Pi specific transport (Pst) system, motility and swarming, and expression of alkaline phosphatases (APases), acid phosphatases or secondary metabolites (Rittmann *et al*., 2005; Monds *et al*., 2006; Sola-Landa *et al*., 2008; Furtwängler *et al*., 2010; Zavaleta-Pastor *et al*., 2010). Fewer studies have experimentally confirmed the global PHO regulon using recently developed 'omics' techniques. Microarrays were employed to identify genes regulated at the transcriptional level in *Pseudomonas aeruginosa* (Bains *et al*., 2012), *E. coli* (Baek and Lee, 2006) and *Synechococcus* sp. WH8102 (Tetu *et al*., 2009; Ostrowski *et al*., 2010) while traditional 2D-gel electrophoresis was performed to make a qualitative assessment of the *Bacillus subtilis* proteome in response to low Pi (Antelmann *et al*., 2000). These studies identified a number of genes/proteins involved in Pi scavenging underlining the importance of performing 'omics' to study Pi acquisition. In *Pseudomonas*, it has been shown that phosphate binding proteins (PBPs), APases, and virulence factors are associated with phosphate-stress and regulated by PhoBR (Monds *et al*., 2006; Bains *et al*., 2012; Putker *et al*., 2013: Santos-Beneit, 2015).

Exoproteomics captures the extracellular protein fraction that results from active secretion, cell lysis or leakage during cell division (Armengaud *et al*., 2012; Ebner *et al*., 2016). Removing the cellular fraction prior to protein extraction can help to identify exoproteins that are involved in the interaction of microorganisms with their environment (Christie-Oleza *et al*., 2012; 2015). The majority of exoproteins detected are associated with nutrient acquisition, motility, cell attachment, defence, communication as well as antagonism (Christie-Oleza *et al*., 2012). Therefore, exoproteomics is the ideal method of choice to study the bacterial mechanisms for scavenging extracellular P. Due to the complexity and technical challenges associated with metaproteomics (Muth *et al*., 2015), this study aimed to identify Pi-responsive proteins that can be used as markers for future studies investigating Pi mobilization within the rhizosphere. Therefore, exoproteomic analyses were performed on three plant-associated *Pseudomonas* strains grown under Pi-deplete conditions (50 μM). We hypothesized that *Pseudomonas* strains harbour a number of common Pi-scavenging enzymes that would be expressed during Pi depletion. In reality, clear evidence for intra-genus-level heterogeneity in their exoproteomes was observed and a number of novel PHO-regulon members specifically linked with the PHO regulon in *Pseudomonas putida* BIRD-1 were also determined.

Results

Effects of P-limitation on the growth of Pseudomonas *strains*

We investigated the effect of Pi stress on three strains, DSM4166, SBW25) and BIRD-1 by comparing growth under Pi-replete (1.4 mM) or Pi-deplete (50 μM) conditions ($n = 3$). Both the growth rates and growth yields of all three *Pseudomonas* strains showed a significant decrease

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

(*t*-test score, $P < 0.01$) under Pi-deplete growth conditions (Fig. 1A–C). As expected, Pi-deplete cultures of all three strains demonstrated a significant increase in the level of APase activity (Fig. 1D) due to the induction of the PHO regulon (Monds *et al.*, 2006; Putker *et al.*, 2013).

General characteristics of the exoproteomes of the three Pseudomonas *species*

Based on qualitative 1D SDS-PAGE analysis, there was visible evidence for heterogeneity in the profiles of *Pseudomonas* exoproteomes in response to Pi-depletion (Fig. 2). Samples were subsequently processed for peptide identification using LC-MS/MS. Proteins were considered present based on a minimum of at least two unique peptides. Exoproteins were identified by the presence of a signal peptide sequence (IMG/JGI). Proteins detected in the exoproteome that did not possess a signal peptide sequence were further analysed using SecretomeP and LipaseP to determine if they are secreted in a non-classical manner (Christie-Oleza and Armengaud, 2010; Christie-Oleza *et al.*, 2015). In Pi-deplete cultures, 'predicted' exoproteins comprised 89.9%, 68.3% or 97.6% of the top-60 most abundant proteins detected in the exoproteomes of BIRD-1, SBW25 and DSM4166 respectively (Supporting Information Fig. S1), while the remaining fraction was comprised of cytoplasmic proteins. In BIRD-1, SBW25 and DSM4166 29, 52 and 54 proteins were significantly enriched (*t*-test, P value ≤ 0.05, fold-change (log2) ≥ 1.5) in response to Pi-depletion respectively (Supporting Information Tables S1–S3). In all three *Pseudomonas* exoproteomes a suite of previously characterized PHO-dependent proteins (Santos-Beneit, 2015) were enriched in their exoproteomes during growth under Pi depletion. These included the high affinity periplasmic substrate binding protein (SBP) subunits of the high affinity Pi transporter (PstS) and phosphonate transporter (PhnD), APases (PhoX, PhoD), 5'-nucleotidase (UshA) and glycerolphosphodiesterase (GlpQ) (Figs. 3 and 4 and Supporting Information Fig. S2). However, the genomic content (Table 1) and thus exoproteomic response to Pi depletion varied between the three *Pseudomonas* strains (Fig. 3), revealing significant inter-genus level heterogeneity. We should point out though that the exoproteome of DSM4166 was harvested after a longer period (DSM4166, 25 h; BIRD-1 & SBW25, 7–11 h respectively) of Pi stress (Fig. 1) and as a result a higher percentage of its exoproteome was associated with Pi-scavenging compared with either BIRD-1 or SBW25 (Fig. 4).

Pseudomonads show heterogeneity in their resource allocation towards organic P scavenging

Bacteria possess a number of different APases (PhoA, PhoD, PhoX) with different phosphomonoesterase and phosphodiesterase activities (Brickman and Beckwith, 1975; Scott and Wu, 2005). Again, there is genomic and thus exoproteomic variation between the three *Pseudomonas* strains with respect to the catabolism of organic P (Fig. 3 and Table 1). For example, BIRD-1 possesses PhoX but lacks PhoD whereas DSM4166 and SBW25 possess both exoenzymes. Furthermore, PhoD was the second most abundant protein in DSM4166 while it was ranked 355th in the exoproteome of SBW25. In addition, DSM4166 also harbours a distinctive PhoX homolog that is duly expressed under Pi depletion (Fig. 4). DSM4166 also expressed homologs of GlpQ (Larson *et al.*, 1983) and UshA (Zalkin and Nygaard, 1996; Rittmann *et al.*, 2005; Pinchuk *et al.*, 2008) while BIRD-1 and SBW25 do not possess either of these exoenzymes. BIRD-1 did possess a gene encoding a predicted exoprotein containing the same domains as UshA (Pfam00149 – metallophos; Pfam02872 – 5_nucleotid_C), but did not secrete this protein in response to low Pi.

Pseudomonads harbour a number of phosphate binding proteins (PBPs) that are enriched in their exoproteome in response to Pi-depletion

In total, the three *Pseudomonas* strains increased the secretion of four different Pi binding proteins (PBP) containing the Pfam01249 domain, in response to Pi-depletion (Figs. 3 and 4). All PBPs found in the genomes of the three *Pseudomonas* strains contain the key residues associated with Pi binding (Supporting Information Fig. S3) (Berna *et al.*, 2008; Liebschner *et al.*, 2009). DSM4166 has a single operon encoding Pst (*pstSCAB*), whereas BIRD-1 and SBW25 have two operons encoding two separate Pst systems (Fig. 5A). All PstS homologs were secreted in large quantities during Pi depletion (Fig. 4). The two operons were named *pst1* and *pst2*. PstS1 is closely related to PstS found in *E. coli* (Wanner, 1990) and *Synechocystis* sp. PCC6803 (herein, *Synechocystis*) (Pitt *et al.*, 2010), whereas PstS2 is phylogenetically distinct and closely related to *Vibrio* PstS (Pratt *et al.*, 2010) (Fig. 5B). Interestingly, *pstSCAB1* is not present in all Pseudomonads whereas *pstSCAB2* is (Table 1).

In SBW25, another PBP (encoded by PFLU2427), hereafter referred to as Psp, was heavily secreted in Pi-deplete cultures (Fig. 4) alongside a large hypothetical exoprotein (encoded by PFLU2428) that contains a domain related to an adhesion virulence factor (Inatsuka *et al.*, 2005). All *Pseudomonas* screened encode Psp3. However, only DSM4166 appeared to secrete this exoprotein (PSTAA_2217) in response to Pi stress. Psp3 contains an outer membrane protein A (OmpA) domain implying that this protein may be located in the outer membrane.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

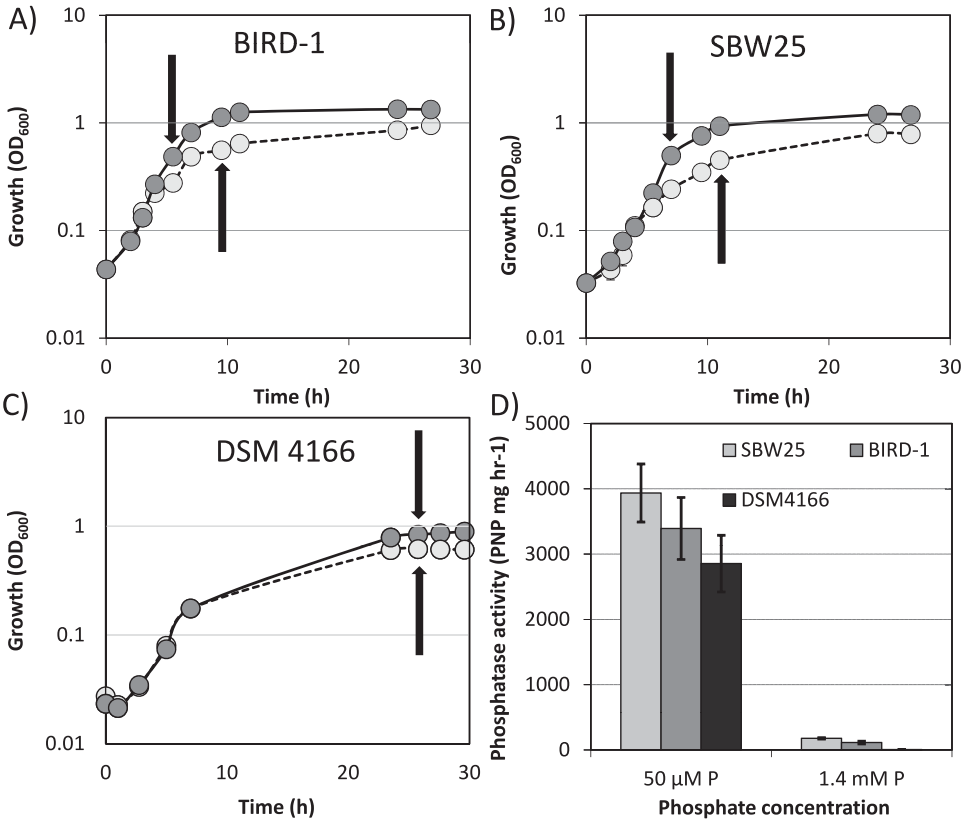


Fig. 1. A–C. Growth of the three *Pseudomonas* strains under either Pi replete (1.4 mM) or Pi deplete (50 μM) growth conditions. Arrows denote sampling points for exoproteomics. D. During the growth experiments, alkaline phosphatase activity was quantified as a proxy for determining the activation of the PHO regulon. The values shown represent a given time point when the maximal alkaline phosphatase activity detected for each strain was obtained. Results presented are the mean of triplicate cultures. Error bars denote standard deviation.

Acquisition of phosphonates in phosphate-depleted Pseudomonas *cells*

To date, PhnD, which is located in an operon with genes encoding the promiscuous C-P lyase (PhnF-M), is the only characterized phosphonate transporter (Baker *et al*., 1998). A number of putative SBPs responsible for phosphonate transport were identified in the *Pseudomonas* strains genomes (Fig. 6A) and detected in their exoproteomes, with almost all showing a positive response to Pi-depletion (Supporting Information Fig. S4). Phylogenetic analysis categorized these homologs into five groups, hereafter referred to as PhnD, PhnD2, PhnD3, PhnD4 and PhnD5. However, none of the three *Pseudomonas* strains possesses all five homologs in their genomes (Fig. 6B and Table 1). PhnD and PhnD2 both contain the Pfam012974 domain (phosphonate-binding domain) and these two homologs are mutually exclusive with one another among the genomes of *Pseudomonas* strains. Only SBW25 possesses and secreted PhnD, a known PHO-regulon member (Baker *et al*., 1998), during Pi-depletion (Fig. 6). BIRD-1 and DSM4166 both possess PhnD2, which showed a modest increase in abundance in both strains under Pi depletion (Supporting Information Fig. S4). The

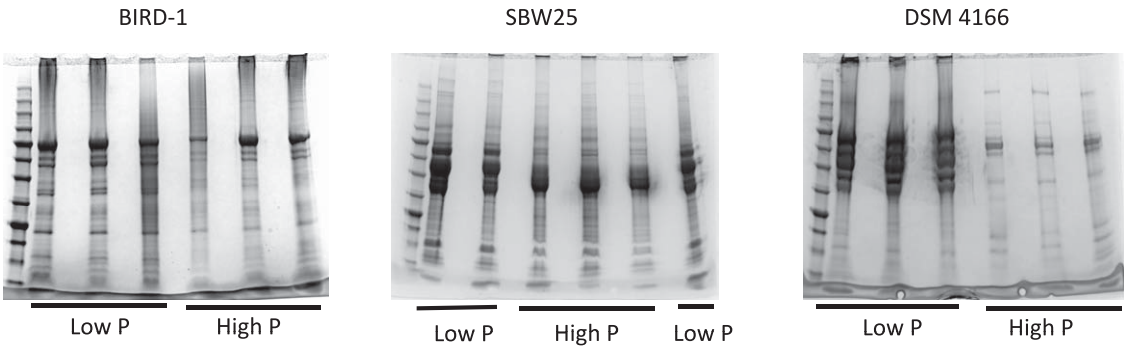


Fig. 2. A qualitative assessment, using 1D-SDS PAGE, of the exoproteomes of all three *Pseudomonas* strains examined prior to HPLC 2D-MS/MS. Each gel lane represents 20 ml culture supernatant. For both Pi deplete and Pi replete growth conditions, three biological replicates were performed.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

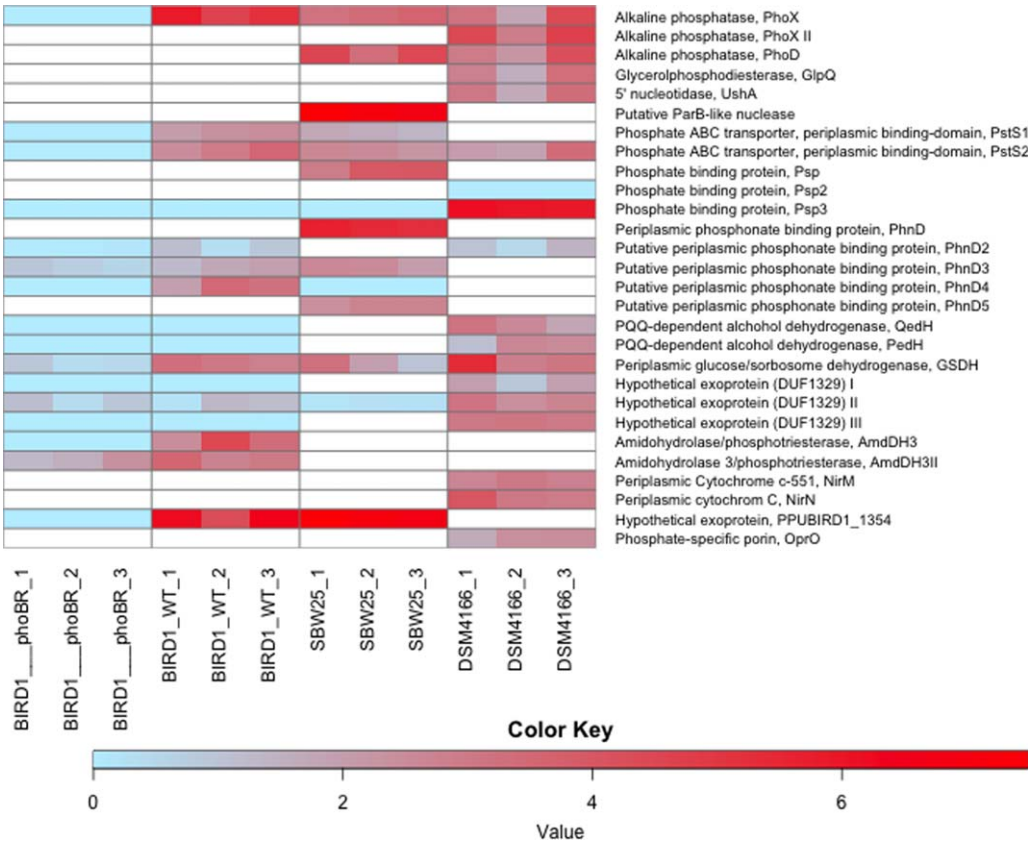


Fig. 3. Protein expression analyses in response to Pi depletion of the three *Pseudomonas* exoproteomes in addition to the exoproteome of the *phoBR* mutant of *P. putida* BIRD-1. White spaces represent the absence of genes encoding the corresponding proteins from their genomes. Each individual biological replicate is displayed. The colour key represents Log2 transformations of protein fold change.

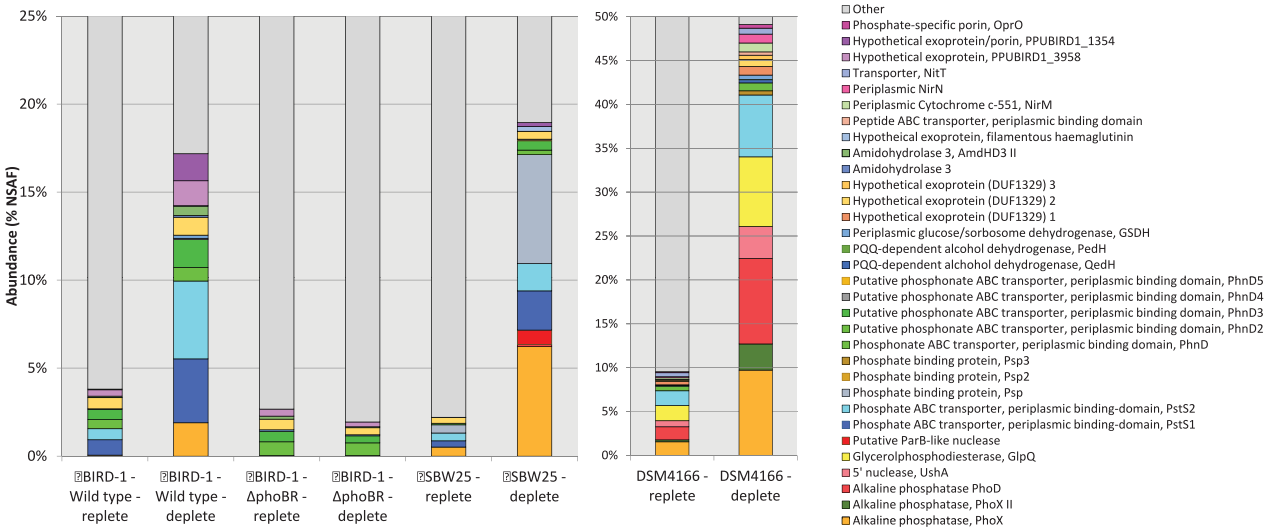


Fig. 4. The relative abundance of Pi scavenging proteins detected in the exoproteomes of the three *Pseudomonas* strains and the *phoBR* mutant grown in both Pi replete and Pi deplete growth conditions. The normalized spectral abundance factor (NSAF) was calculated using Scaffold 4. Values displayed are the mean of triplicate cultures.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Table 1. Comparative genomic analysis of selected proteins involved in the recycling of P among *Pseudomonas* isolates.

BLASTP was performed using the IMG/JGI database on selected strains whose genome was marked as 'finished'. Multiple diamonds indicate two or more homologs.
Abbreviations are the same as in Figure 2 as well as: LapA/B, low molecular weight phosphatase; PhoT5, Flavobacterial putative Pi-binding protein form V; PhnF-M, C-P lyase; PhnXW, phosphonatase; PhnWAY, alternative 2-aminoethylphosphonatase; PhnCDE, phosphonate transporter I, $PhnC_2D_2E_2$, putative phosphonate transporter II; PalA, phosphonopyruvate hydrolase; PhoA, alkaline phosphatase; PlcP, intracellular phospholipid phosphodiesterase; DagK, diacylglycerol kinase; OlsA, Lyso-ornithine lipid:acyl-ACP O-acyltransferase; OlsB, Ornithine:acyl-ACP N-acyltransferase; OlsF, bifunctional ornithine acyltransferase; SqdBCD, sulfolipid biosynthesis; Cfa, cyclo-propane fatty acid synthase; BtaAB, diacylglycerol trimethylhomoserine biosynthesis.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

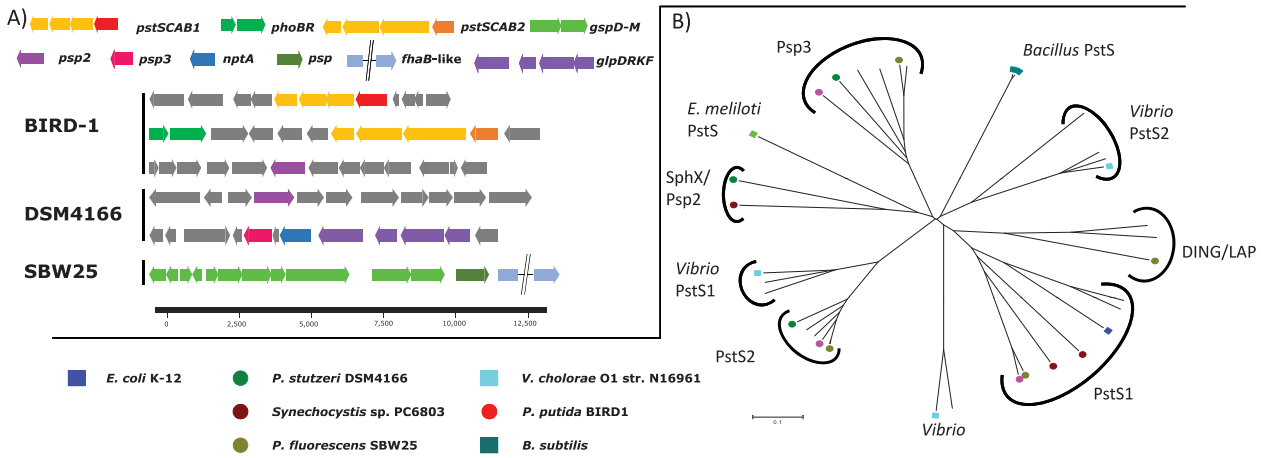


Fig. 5. Genomic analyses of the Pi binding proteins found in the three *Pseudomonas* strains (A) The genetic neighbourhood profiles of the different Pi binding proteins located in the three *Pseudomonas* strains (B) The diversity of proteins that contain the Pfam12849 domain using a number of genome-sequenced soil bacteria with the inclusion of characterized Pi binding proteins. Abbreviations; *pstSCAB*1/2, Pi-specific ABC transporter; *psp*, DING-family Pi binding protein; *psp2/3*, uncharacterized Pi binding protein; *nptA*, NA^{+}/Pi co-transporter; *glp*D, glycerol 3-phosphate dehydrogenase; *glpR*, transcriptional regulator; glycerol kinase; *glpK*, glycerol uptake facilitator; *fhaB*-like, putative filamentous haemagglutinin; *phoBR*, two component regulator; *gsp*, type II secretion system.

abundance of PhnD3 also increased in Pi-deplete SBW25 and BIRD-1 exoproteomes. PhnD4 is closely related to a SBP (SM_b21540) located upstream of genes (*phnWAY*) encoding an alternative pathway for the degradation of 2-AEP in *Sinorhizobium meliloti* (Fig. 6A) (Borisova *et al.*, 2011) and was only detected in Pi-deplete BIRD-1 cultures, albeit at a low abundance.

Organic acid production

The Pi solubilization potential of the three Pseudomonads showed marked differences at both the genomic and exoproteomic level. Only DSM4166 induced secretion of two extracellular proteins annotated as pyroloquinoline quinone (PQQ)-dependent alcohol dehydrogenases (QedH, PSTAA_2299; PedH, PSTAA_2293) in response to Pi stress (Fig. 3). Although BIRD-1 does possess homologs of both of these proteins neither were detected in its exoproteome. All three strains expressed and secreted an exoprotein under Pi-depletion that contained the Pfam07995 domain (glucose/sorbosone dehydrogenase) and this enzyme may represent a novel mechanism for extracellular organic acid production.

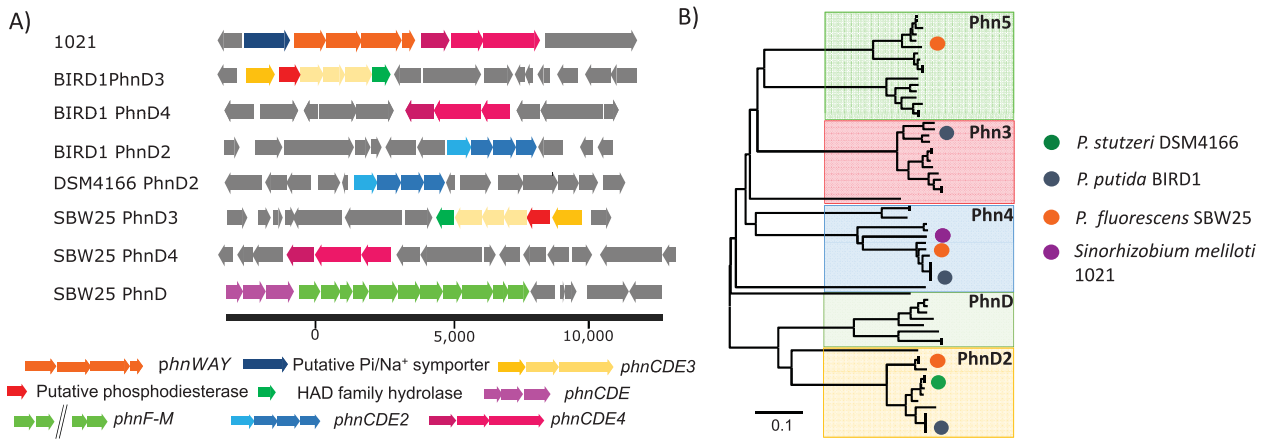


Fig. 6. Genomic and proteomic analyses of the phosphonate binding proteins found in the three *Pseudomonas* strains (A) The genetic neighbourhood profiles of the different phosphonate binding proteins located in the three *Pseudomonas* strains (BIRD-1, SBW25, DSM4166) as well as *Sinorhizobium meliloti* 1021 (1021). (B) Neighbour-joining phylogenetic analysis of the different phosphonate binding proteins detected in the *Pseudomonas* strains outlined in Table 1 with the addition of various *Burkholderia* and *Flavobacteria* strains. Bootstrap values (500 runs) have been omitted for clarity. IMG accession numbers have been included as a reference. Abbreviations: *phnWAY*, alternative 2-aminoethylphosphonate degradation pathway; *phnDCE1/2/3/4*, phosphonate ABC transporter; *phnF-M*, C-P lyase.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Identification of novel Pi-responsive exoproteins in Pseudomonas *strains*

In response to Pi-depletion BIRD-1 and SBW25 expressed a hypothetical outer membrane protein (PPUBIRD1_1354, PFLU4536) containing the Pfam16930 domain which is linked to porin structure (Figs. 3 and 4). In BIRD-1 this was one of the most abundant proteins in the exoproteome of P-stressed cells (Fig. 3). P-stressed SBW25 cells secreted an uncharacterized lipoprotein containing a ParB-like nuclease domain (Pfam08857), which may have a similar function to UshA (Rittmann *et al*., 2005) (Fig. 3). BIRD-1 also secreted a putative extracellular phosphotriesterase/amidohydrolase (AmdDH3 II, PPUBIRD1_5046) that has the potential to cleave the C-O-P bond of certain pesticides (Sun *et al*., 2004). In DSM4166, three hypothetical exoproteins containing the domain of unknown function (DUF) 1329 increased in abundance in response to low Pi. However, these proteins did not increase in abundance in the exoproteomes of either of the other two *Pseudomonas* isolates. Finally, a number of putative extracellular proteases were enriched in the exoproteome of SBW25 in response to Pi depletion (Supporting Information Table S2).

Analysis of the Pi-responsive whole-cell proteome in BIRD-1

To gain a deeper understanding of the Pi-responsive proteome, the whole-cell proteome of BIRD-1, which included both cytosolic and membrane protein fractions, was analysed. The majority of abundant proteins detected in the proteome of BIRD-1 were housekeeping and central metabolism proteins (e.g. GyrB, DnaK, GroEL, RpoD, RpoB, FusA, IleS, GuaA, Tuf, SdhA, SdhB, AtpA, AtpB, RpsA, RpsC, SucC, SucA) whose abundance was not significantly affected by differing Pi regime (Supporting Information Table S4). A total of 267 proteins were significantly enriched [*t*-test, P value ≤ 0.05, fold-change (log2) ≥ 1.5] during Pi depletion and there was concordance between the two datasets. These enriched proteins included the transmembrane and ATP-binding domains of the Pst system (PstC, PstA, PstB), a 2-aminoethylphosphonate (2-AEP)–specific phosphonatase (PhnX, PhnW) (Jiang *et al*., 1995; Baker *et al*., 1998; White and Metcalf, 2007), proteins involved in lipid remodelling (PlcP, DagK, OlsA, OlsB, Cfa, TauD) (Liu and Hulett, 1998; Antelmann *et al*., 2000; Zavaleta-Pastor *et al*., 2010; Carini *et al*., 2015; Sebastian *et al*., 2016), a putative intracellular phosphatase (UxpA) and the twin-arginine translocation (TAT) pathway (Putker *et al*., 2013) (Table 2). Proteins for both starch (MalQ, GlgE, GlgX, GlpA and GlpB) and polyhydroxyalkanoic acid (PhaA, PhaG, PhaC) biosynthesis (carbon storage) were also enriched during Pi stress (Supporting Information Table S4). The abundance of a cytoplasmic glucose-6-phosphate dehydrogenase (Zwf), as well as two distinct membrane-bound (PPUBIRD1_4115, PPYBiRD1_2225) glucose dehydrogenases (Gcd, GcdII respectively), all of which are known to play a role in Pi solubilization through gluconic acid production (Miller *et al*., 2010; Roca *et al*., 2013), was also greater in Pi-deplete cells. Finally, another HAD-family phosphatase (encoded by PPUBIRD1_3492), similar to PhnX was only detected in the proteome of Pi-deplete cells (Table 2).

Identification of PHO-regulated Pi-responsive proteins in BIRD-1

To assess how the Pi-responsive proteome and exoproteome in BIRD-1 is regulated we disrupted the genes encoding the master regulator of the PHO regulon, *phoBR*. Compared with the wild type, the *phoBR* mutant showed a substantial reduction in final growth yield when grown under Pi deplete, but not Pi replete conditions (Supporting Information Fig. S5). As with the wild type strain, we performed exoproteome and whole-cell proteome analysis for the *phoBR* strain. In the *phoBR* mutant, three categories of Pi responsive proteins were determined: (1) proteins that were absent (below detection level) in the mutant and observed in the wild type (2) proteins that no longer increased in abundance under Pi stress and (3) proteins whose abundance in both Pi-replete and Pi-deplete growth conditions was similar to the wild type. In the whole-cell proteome, PstSCAB1&2, PhoX, UxpA, PhoBR, PhoU, TatADG, PhnXW, PlcP, OlsAB, TuaD were all absent during Pi-deplete growth of the mutant (Table 2). With respect to Pi solubilization, GcdII was also absent, whereas Gcd was not enriched in Pi-deplete cells, unlike the wild type. However, the abundance of Zwf was unaffected by mutation of *phoBR*. The putative HAD-like phosphatase (PPUBIRD-1_3492), as well as the hypothetical outer membrane protein PPBUBIRD1_1354, were also absent in the mutant, suggesting that these two proteins are novel members of the pho regulon (Table 2). Interestingly, AmdHd3II was still enriched during Pi-stress suggesting that it is not regulated by *phoBR*. The abundance of several other proteins was also not affected by mutation of *phoBR* (Table 2 and Supporting Information Table S5) indicating that a PHO-independent response to Pi stress occurs in BIRD-1. For example, all the carbon storage proteins and proteins linked with biofilm formation (encoded by PPUBIRD1_2591-2608) still showed a PHO-independent response to Pi stress.

In the exoproteome of the BIRD-1 *phoBR* mutant, PstS1, PhoX, the hypothetical exoprotein/porin (PPUBIRD1_1354), PhnD4 were all absent from either growth condition. Meanwhile, the abundance of PstS2 PhnD2, PhnD3, GSDH, AmdHd3, AmdHdII and the hypothetical exoprotein (PPUBIRD1_3958) were all reduced in Pi

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Table 2. P-responsive proteins in BIRD-1 that are under the control of the PHO regulator, PhoBR.

Identified proteins		Accession number	Locus tag	Fold change – WT*	Fold change – mutant*
Proteins silenced in the *phoBR* mutant					
Hypothetical protein, conserved		ADR57845	PPUBIRD1_0136	4.12	ND
Taurine dioxygenase	*tauD*	ADR57960	PPUBIRD1_0256	2.31	ND
Phosphate-specific methyl-accepting chemotaxis protein	*ctpL*	ADR58302	PPUBIRD1_0612	4.95	ND
Dehydratase		ADR58320	PPUBIRD1_0630	2.32	ND
HlyD family type I secretion membrane fusion protein		ADR58534	PPUBIRD1_0849	3.42	ND
ABC transporter related protein		ADR58535	PPUBIRD1_0850	4.34	ND
lyso-ornithine lipid acyltransferase	*olsA*	ADR58658	PPUBIRD1_0974	2.18	ND
ornithine-acyl[acyl carrier protein] N-acyltransferase	*olsB*	ADR58659	PPUBIRD1_0975	4.14	ND
L-serine dehydratase		ADR58719	PPUBIRD1_1037	2.05	ND
Arginine/ornithine antiporter	*arcD*	ADR58734	PPUBIRD1_1052	3.04	ND
General secretion pathway protein K	*tatA*	ADR58773	PPUBIRD1_1091	3.34	ND
Alkaline phosphatase	*phoX*	ADR58775	PPUBIRD1_1093	5.14	ND
2',3'-cyclic-nucleotide 2'-phosphodiesterase	*uxpA*	ADR58776	PPUBIRD1_1094	3.24	ND
General secretion pathway protein G	*tatG*	ADR58781	PPUBIRD1_1099	4.60	ND
Hypothetical protein, conserved		ADR59032	PPUBIRD1_1354	5.10	ND
Metallophosphoesterase	*plcP*	ADR59065	PPUBIRD1_1390	4.59	ND
Gluconate 2-dehydrogenase acceptor subunit		ADR59804	PPUBIRD1_2165	3.15	ND
2Fe-2S iron-sulfur cluster binding domain		ADR59806	PPUBIRD1_2167	4.47	ND
Probable quinate dehydrogenase	*gcdII*	ADR59861	PPUBIRD1_2225	3.26	ND
UDP-glucose 4-epimerase		ADR60229	PPUBIRD1_2604	2.14	ND
Response regulator		ADR60354	PPUBIRD1_2733	3.33	ND
UDP-glucose 6-dehydrogenase	*tuaD*	ADR60422	PPUBIRD1_2809	5.63	ND
Putative cyclopropane fatty acid synthase A	*cfa2*	ADR60552	PPUBIRD1_2940	2.93	ND
Hypothetical protein, conserved		ADR60627	PPUBIRD1_3016	4.96	ND
Phosphate ABC transporter, ATP-binding domain	*pstB1*	ADR60628	PPUBIRD1_3017	6.70	ND
Phosphate ABC transporter, transmembrane domain	*pstA1*	ADR60629	PPUBIRD1_3018	4.52	ND
Phosphate ABC transporter transmembrane domain	*pstC1*	ADR60630	PPUBIRD1_3019	3.91	ND
Phosphate ABC transporter, periplasmic binding domain	*pstS1*	ADR60631	PPUBIRD1_3020	5.08	ND
2-aminoethylphosphonate–pyruvate transaminase	*phnW*	ADR61037	PPUBIRD1_3442	3.87	ND
Phosphonoacetaldehyde hydrolase	*phnX*	ADR61038	PPUBIRD1_3443	4.76	ND
Hypothetical protein, conserved (HAD-like domain)		ADR61085	PPUBIRD1_3492	3.081	ND
Cation/acetate symporter actP	*actP*	ADR61460	PPUBIRD1_3874	2.74	ND
GntR family transcriptional regulator		ADR61476	PPUBIRD1_3890	3.57	ND
ABC transporter ATP-binding protein		ADR61481	PPUBIRD1_3895	4.72	ND
Metallopeptidase, zinc binding protein		ADR61927	PPUBIRD1_4353	2.26	ND
Fe3+ ABC transporter, periplasmic binding domain		ADR62476	PPUBIRD1_4925	2.04	ND
Fe3+ ABC transporter, ATP-binding domain		ADR62478	PPUBIRD1_4927	2.85	ND
Arylesterase, putative		ADR62594	PPUBIRD1_5047	3.47	ND
Winged helix family regulator	*phoB*	ADR62659	PPUBIRD1_5112	4.38	ND
Phosphate regulon sensor protein	*phoR*	ADR62660	PPUBIRD1_5113	4.64	ND
Phosphate ABC transporter, ATP-binding domain	*pstB2*	ADR62665	PPUBIRD1_5118	5.16	ND
Phosphate ABC transporter, transmembrane domain	*pstA2*	ADR62666	PPUBIRD1_5119	4.93	ND
Phosphate ABC transporter, transmembrane domain	*pstC2*	ADR62667	PPUBIRD1_5120	6.46	ND
Phosphate ABC transporter, periplasmic binding domain	*pstS2*	ADR62668	PPUBIRD1_5121	3.55	ND
Proteins that were down-regulated compared with the WT					
Taurine ABC transport, periplasmic binding domain	*tauA*	ADR57963	PPUBIRD1_0259	3.10	−1.14
Quinoprotein glucose dehydrogenase A	*gcd*	ADR61697	PPUBIRD1_4115	2.03	0.86
Methyl-accepting chemotaxis sensory transducer		ADR61928	PPUBIRD1_4354	1.99	−0.38
Phosphate transport regulator	*phoU*	ADR62664	PPUBIRD1_5117	2.07	−0.32

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Table 2. *cont.*

Identified proteins		Accession number	Locus tag	Fold change – WT*	Fold change – mutant*
Proteins whose expression was not affected by PhoBR					
Cyclopropane-fatty-acyl-phospholipid synthase	*cfa1*	ADR57745	PPUBIRD1_0035	2.71	4.55
Phosphonate ABC transporter, periplasmic binding domain	*phnD2*	ADR58557	PPUBIRD1_0873	0.62	1.16
ABC-type Fe3+ transport system periplasmic binding domain	*phnD3*	ADR61477	PPUBIRD1_3891	2.69	2.31
Exopolyphosphatase	*ppX*	ADR62560	PPUBIRD1_5012	0.44	0.66
Polyphosphate kinase	*ppK*	ADR62561	PPUBIRD1_5013	1.34	0.89
Amidohydrolase 3	*amdhd3II*	ADR62593	PPUBIRD1_5046	4.45	3.97
PhoH family protein	*phoH*	ADR61846	PPUBIRD1_4270	1.38	2.78

ND, not detected; WT, wild type.
A number of P-responsive proteins of interest that are not regulated by PhoBR are also listed. The abundance of these proteins within the Proteome of BIRD-1 is shown in Table S4. The accession number shown refers to the Uniprot database. * represents Log2 transformation of fold change values that are the mean of triplicate cultures. All proteins displayed in the Table with a Log2 value ≥ 1.5, were also statistically significantly enriched under Pi depletion (*t*-test, *P* value ≤ 0.05).

deplete cultures compared with the wild type, but were still detected (Figs. 3 and 4).

Discussion

Characterizing the exoproteomes, and thus the functional entities associated with environmental interactions (Armengaud *et al.*, 2012), of various *Pseudomonas* strains allowed us to deepen our understanding of Pi-regulated protein expression in this genus (Figs. 3 and 4). While, genomic comparisons, based on known proteins in the literature, allowed us to access the heterogeneity in their P-mobilizing and P-scavenging 'potential', proteomic analyses revealed both differences in their global regulatory networks and also helped to identify novel Pi-responsive proteins, which may be of further biotechnological interest. For example, a novel Pi-responsive extracellular nuclease in SBW25 was discovered that was not identified through our comparative genomic analysis. Importantly, the genes (PPUBIRD1_5077, PPUBIRD1_0727, PPUBIRD1_2395, PPUBIRD1_0951, PPUBIRD1_0932) identified in BIRD-1, based solely on *in silico* annotation (Roca *et al.*, 2013), were not members of the PHO regulon, highlighting the need for auxiliary studies to confirm genomic annotation. Furthermore, the strong secretion of exoproteins, such as PstS and PhoX, may serve as markers for characterizing complex communities in soil/rhizosphere to enable identification of the key microbial taxa involved in P recycling.

Soil organic P exists in many forms and frequently accounts for 30%–65% of total P in soils (Harrison, 1987) and its mineralization to Pi can have a great impact on total P bioavailability (Turner *et al.*, 2002; Shen *et al.*, 2011). From the genomic comparison of each strain, it appears that DSM4166 has the greatest ability to degrade organic P compounds as it contains two distinct PhoX homologs as well as PhoD, UshA and GlpQ (Larson *et al.*, 1983; Antelmann *et al.*, 2000; Rittmann *et al.*, 2005; Monds *et al.*, 2006; Pinchuk *et al.*, 2008; Putker *et al.*, 2013). We also identified an UshA-like homolog and a GlpQ-like homolog (PFLU4789) in the genomes of BIRD-1 and SBW25, respectively, but in contrast to DSM4166, neither or these homologs were secreted in response to Pi depletion. SBW25 also heavily secreted Psp, a phosphate-binding protein (Scott and Wu, 2005), which is closely related to the low molecular weight phosphatases, LapA and LapB, in *P. aeruginosa* (Tan and Worobec, 1993; Ball *et al.*, 2002). Psp may, therefore, be the exoenzyme responsible for the unnaccounted APase activity detected in *P. fluorescens* Pf0-1 (Monds *et al.*, 2006). Interestingly, it was SBW25 that elicited the strongest APase activity towards *p*NPP (phosphomonoesterase activity). To date, little is known about the natural substrate range of these promiscuous enzymes in soil and it is likely that differences occur between the different PhoX homologs. In support of this hypothesis, SBW25 PhoX is phylogenetically distinct (Supporting Information Fig. S6) from either BIRD-1 or DSM4166 homologs.

Although P solubilization through organic acid production has been well studied in *Pseudomonas* (Rodríguez and Fraga, 1999; Miller *et al.*, 2010) several putatively new P solubilizing proteins likely involved in this process were still identified at the genomic level, e.g. QedH, PedH, and an alternative PHO-regulated Gcd, GcdII (Fig. 2; Table 2). However, a similar discordance between genomic prediction and proteomic abundance was observed for Pi-solubilizing enzymes. For example, QedH and PedH were detected in Pi-depleted DSM4166 cultures, but not in BIRD-1. GSDH, another previously uncharacterized protein with respect to the PHO regulon, and not identified in our genomic assessment, was secreted in all three strains

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

in response to Pi-depletion and likely has a role in Pi-solubilization though organic acid release. The lower abundance of Pi-solubilizing proteins may have resulted from an absence of glucose in the growth medium the precursor substrate for these Pi-solubilizing enzymes.

Although the three *Pseudomonas* strains can grow on phytate as a source of P (Lim *et al*., 2007; Roca *et al*., 2013), there was no evidence that either the known phytase of DSM4166 and SBW25, or the predicted phytase, responsible for the growth of BIRD-1 on this substrate (Roca *et al*., 2013) are regulated by PhoBR. Considering that phytate is usually ubiquitous in soils (Stutter *et al*., 2012), Pseudomonads have likely adapted to express their respective phytases solely in response to the presence of this compound and may explain why these Pi-mobilizing enzymes are unexpectedly not part of the *Pseudomonas* PHO regulon.

The existence of two distinct Pst systems in both BIRD-1 and SBW25 is similar to that of *Synechocystis*, *Vibrio cholerae* and the Archaeon, *Halobacterium salinarium* R1 (Furtwängler *et al*., 2010; Pitt *et al*., 2010; Mudrak and Tamayo, 2012). As Pst2 is present in all *Pseudomonas* strains, while Pst1 appears in only a few, we hypotheize that Pst2 is essential for efficient uptake of Pi in *Pseudomonas*. However, Pst1 must clearly have a role in Pi uptake as we detected PstS1 as well as PstS2 in both BIRD-1 and SBW25 exoproteomes. Furthermore, although disruption of PstS1 in SBW25 did not affect growth on low Pi in isolation, it did confer a fitness reduction in the presence of the wild type (Zhang *et al*., 2007). In *Synechocystis, Vibrio* and *H. salinarium*, Pst1 and Pst2 either have different kinetic parameters for the uptake of Pi (Furtwängler *et al*., 2010; Pitt *et al*., 2010) or are expressed during different growth phases (planktonic v biofilm) (Pratt *et al*., 2010; Mudrak and Tamayo, 2012). The data presented in this study favours the hypothesis that they have different kinetic parameters as both were expressed in BIRD-1 and SBW25 during planktonic growth.

Only BIRD-1 and SBW25 have the genetic potential to catabolize phosphonates (Table 1), and in BIRD-1, phosphonatase (PhnWX) was PHO-regulated. Based on our data, we cannot rule out the possibility that DSM4166 can also grow on phosphonates as a source of P for two reasons: (i) Although phosphonate degradation has been well documented in recent years (Jiang *et al*., 1995; White and Metcalf, 2007; Villarreal-Chiu *et al*., 2012; McGrath *et al*., 2013), bacteria capable of growing on phosphonates that do not possess any of the characterized genes/proteins have been isolated, demonstrating alternative pathways for phosphonate degradation must exist in nature (Fox and Mendz, 2006); (ii) DSM4166 possesses and expressed a number of putative phosphonate transporters (Table 1 and Fig. 6). Characterizing these putative transporters will surely enhance our knowledge regarding phosphonate degradation in *Pseudomonas* strains and may provide new molecular markers for investigating the *in situ* cycling of these compounds (Mauchline *et al*., 2006; Christie-Oleza and Armengaud, 2010; Lidbury *et al*., 2014).

Bacteria that can remodel their lipid membranes in order to reduce their ratio of P-containing:non P-containing lipids (Zavaleta-Pastor *et al*., 2010; Carini *et al*., 2015; Sebastian *et al*., 2016) are desirable to use as PGPR as their requirement for P is reduced. We found genes encoding for the key proteins required for lipid remodelling (PlcP, DagK, OlsA and OlsB) in the genomes of all *Pseudomonas* strains scrutinised (Gao *et al*., 2004; Zavaleta-Pastor *et al*., 2010). Furthermore, in BIRD-1 these proteins were expressed in a PHO-dependent manner. Interestingly, in BIRD-1 UDP-glucose 6-phosphate dehydrogenase (TuaD) was also PHO-regulated. In *B. subtilis*, TuaD is encoded by the *tua* operon that is involved in the production of teichuronic acid lipids during Pi-depletion (Liu and Hulett, 1998; Antelmann *et al*., 2000). The rest of the genes required for teichuronic acid were absent, therefore, making the role of TuaD somewhat unclear in BIRD-1. As all of these proteins were silenced in the *phoBR* mutant, it is likely that remodelling the lipid membrane accounts for the gross difference observed in growth between this strain and the wild type when grown under Pi-deplete conditions (Fig. 7).

The observed heterogeneity in the Pi responsive portion of the proteome of the three *Pseudomonas* strains in this study highlights how the utilization of different PGPR can have potentially different effects in the rhizosphere. For example, based on our genomic and proteomics data, differences in the ability of the strains to degrade phospholipids, nucleic acids and organopesticides (Singh and Walker, 2006; Bigley and Raushel, 2013), as well as a likely difference in their broad organic P substrate range may have marked effects on their ability to mobilize P for a plant host under certain environmental conditions. For example, the addition of a strain comparable to DSM4166, expressing UshA and GlpQ, may increase the acquisition of P in plants when using manure as the nutrient source, which is rich in nucleic acids and phospholipids (Turner and Leytem, 2004; Shen *et al*., 2011). Furthermore, DSM4166 is a known nitrogen fixer and in certain soils (nitrogen-limited) employing this strain over that of either BIRD-1 or SBW25 may provide more efficient plant-growth promotion. Likewise, in soils contaminated with phosphorus-containing organopesticides, it may be more suitable to deploy a strain similar to SBW25, which contains the promiscuous C-P lyase, capable of degrading these compounds, or BIRD-1 which possesses a Pi-responsive putative phosophotriesterase.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Conclusions

Observing the global exoproteomic response of just three *Pseudomonas* species revealed new insights into the P scavenging capabilities of this genus and has provided a number of markers (Muth *et al*., 2015) that can be utilized to investigate P-mobilization directly in the rhizosphere. Given the enormous task of identifying proteins *in situ* from complex communities, the data presented in this paper will serve as a platform to investigate the key enzymes and microbial taxa involved in P-mobilization at the level of functional entities (proteins) *in situ*. Meta-exoproteomics has already identified differences between genomic and proteomic assessments of soil chintase-degrading communites (Johnson-Rollings *et al*., 2014) and should also shed light on the 'black box' concerning P-mobilization in the rhizosphere.

Experimental procedures

Growth and maintenance of bacterial strains

All three *Pseudomonas* strains were maintained on Luria Bertani (LB) agar (1.5% w/v) medium at 30°C. To investigate the effect of Pi-depletion on the three strains, each was grown ($n = 3$) in an adapted Minimal A medium comprising: Na-Succinate 5.4 g l^{-1}, NaCl 200 mg l^{-1}, NH_4Cl 450 mg l^{-1}, $CaCl_2$ 200 mg l^{-1}, KCL mg l^{-1} $MgCl_2$ 450 mg l^{-1}, $FeCl_2$ 10 mg l^{-1}, $MnCl_2$ 10 mg l^{-1}, 10 mM 4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid (HEPES) pH 7.2, with KH_2PO_4 added to a final concentration of either 50 μM or 1.4 mM. Each strain was pre-cultured in minimal A medium containing 400 μM Pi to ensure cells had adequate Pi while minimizing the potential for carry over of residual Pi into triplicate experimental cultures.

Quantification of alkaline phosphatase activity

A 0.5 ml culture ($n = 3$) was incubated with 20 μl *para*-nitrophenyl phosphate (*p*NPP) (final conc. 4mM) and incubated at room temperature for 1 h or when colour development started to occur. The reaction was stopped using 25 μl NaOH (2 mM) and incubated for 10 min. Cell debris and precipitants were removed via centrifugation (2 min, 8,000 × *g*) prior to spectrophotometry (optical density 405 nm). A standard curve for *para*-nitrophenol was generated using a range of known concentrations (0, 4, 8, 25, 50, 75, 100 mg ml^{-1}).

Preparation of exoproteomes, trypsin in-gel proteolysis, nano-LC-MS/MS analysis and peptide identification through MS/MS database searching

Exoproteomes were analysed using modified methods of Christie-Oleza and Armengaud (2010). The recorded MS/MS spectra were searched against the protein sequence database (*P. putida* BIRD-1, NC_017530.1; *P. fluorescens* SBW25, NC_012660.1; *P. stutzeri* DSM4166, NC_017532.1). Full details of the protocol and parameters used for peptide identification can be found in the supplementary materials and methods.

Quantification of detected proteins

The Normalized Spectral Abundance Factor (NSAF) values were calculated using SCAFFOLD v4.0 according to software defaults. For the exoproteomes, no further normalization was performed. For whole-cell proteomics, 25 μg of protein was loaded onto SDS-PAGE gels prior to identification. No further normalization was performed. However, we examined the abundance of several housekeeping proteins and central metabolic enzymes and did not observe substantial change in their abundance. For determining the proportion of proteins within the exoproteome, replicate cultures ($n = 3$) were averaged. The proteomics data has been deposited in the Proteomics Identification (PRIDE) database (Martens *et al*., 2005) with the following accession numbers: PXD004065, PXD004064, PXD003830, PXD003829, PXD003828, PXD003827, PXD003826.

Bioinformatics analysis of detected proteins and comparative genomics

The majority of analyses were performed using the Integrated Microbial Genomes Database at the Joint Genome Institute (IMG/JGI) server (http://img.jgi.doe.gov/). Please refer to supplementary information for a detailed summary of the bioinformatics approaches used. IMG/JGI was also used for comparative genomic analyses. BLASTP (expected value, e-30, minimum identity = 20%) searches were performed using the proteins detected in the exoproteomes/proteome. In some cases other proteins identified from the literature known to be involved in Pi scavenging/recycling were used as queries for BLASTP analysis. For the Pi-binding proteins (PBP), a function search using the IMG/JGI database was performed using the Pfam domain, 12849 as the query.

Genetic manipulation of P. putida *BIRD-1*

To construct a *phoBR* mutant of *P. putida* BIRD-1, the method outlined by Lidbury *et al*. (2014) was adapted. Please refer to the supplementary information for a detailed procedure.

Acknowledgements

A special thank you goes to Prof. Juan Luis Ramos and Dr. Andrew Spiers for their kind donations of *P. putida* BIRD-1 and *P. fluorescens* SBW25, respectively. This study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under the project code BB/L026074/1. We also acknowledge the WPH Proteomics RTP, specifically Cleidiane Zampronio, for mass spectrometry and proteomic analysis.

References

Antelmann, H., Scharf, C., and Hecker, M. (2000) Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis. *J Bacteriol* **182:** 4478–4490.

Armengaud, J., Christie-Oleza, J.A., Clair, G., Malard, V., and Duport, C. (2012) Exoproteomics: exploring the world around biological systems. *Expert Rev Proteomics* **9:** 561–575.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Bains, M., Fernández, L., and Hancock, R.E.W. (2012) Phosphate starvation promotes swarming motility and cytotoxicity of *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **78:** 6762–6768.

Baker, A.S., Ciocci, M.J., Metcalf, W.W., Kim, J., Babbitt, P.C., Wanner, B.L., *et al*. (1998) Insights into the mechanism of catalysis by the P−C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry* **37:** 9305–9315.

Ball, G., Durand, É., Lazdunski, A., and Filloux, A. (2002) A novel type II secretion system in *Pseudomonas aeruginosa*. *Mol Microbiol* **43:** 475–485.

Berna, A., Bernier, F., Chabrière, E., Perera, T., and Scott, K. (2008) DING proteins; novel members of a prokaryotic phosphate-binding protein superfamily which extends into the eukaryotic kingdom. *Int J Biochem Cell B* **40:** 170–175.

Bigley, A.N., and Raushel, F.M. (2013) Catalytic mechanisms for phosphotriesterases. *Biochim Biophys Acta* **1834:** 443–453.

Borisova, S.A., Christman, H.D., Metcalf, M.E.M., Zulkepli, N.A., Zhang, J.K., van der Donk, W.A., and Metcalf, W.W. (2011) Genetic and biochemical characterization of a pathway for the degradation of 2-aminoethylphosphonate in *Sinorhizobium meliloti* 1021. *J Biol Chem* **286:** 22283–22290.

Brickman, E., and Beckwith, J. (1975) Analysis of the regulation of *Escherichia coli* alkaline phosphatase synthesis using deletions and φ80 transducing phages. *J Mol Biol* **96:** 307–316.

Carini, P., Van Mooy, B.A.S., Thrash, J.C., White, A., Zhao, Y., Campbell, E.O., *et al*. (2015) SAR11 lipid renovation in response to phosphate starvation. *Proc Natl Acad Sci USA* **112:** 7767–7772.

Christie-Oleza, J.A., and Armengaud, J. (2010) In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. *Mar Drugs* **8:** 2223–2239.

Christie-Oleza, J.A., Piña-Villalonga, J.M., Bosch, R., Nogales, B., and Armengaud, J. (2012) Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol Cell Proteomics* **11:** M111.013110.

Christie-Oleza, J.A., Armengaud, J., Guerin, P., and Scanlan, D.J. (2015) Functional distinctness in the exoproteomes of marine *Synechococcus*. *Environ Microbiol* **17:** 3781–3794.

Cordell, D., Drangert, J.-O., and White, S. (2009) The story of phosphorus: global food security and food for thought. *Global Environ Change* **19:** 292–305.

Ebner, P., Rinker, J., and Götz, F. (2016) Excretion of cytoplasmic proteins in *Staphylococcus* is most likely not due to cell lysis. *Curr Genet* **62:** 19–23

Fox, E.M., and Mendz, G.L. (2006) Phosphonate degradation in microorganisms. *Enzyme Microb Technol* **40:** 145–150.

Furtwängler, K., Tarasov, V., Wende, A., Schwarz, C., and Oesterhelt, D. (2010) Regulation of phosphate uptake via Pst transporters in *Halobacterium salinarum* R1. *Mol Microbiol* **76:** 378–392.

Gao, J.-L., Weissenmayer, B., Taylor, A.M., Thomas-Oates, J., López-Lara, I.M., and Geiger, O. (2004) Identification of a gene required for the formation of lyso-ornithine lipid, an intermediate in the biosynthesis of ornithine-containing lipids. *Mol Microbiol* **53:** 1757–1770.

Harrison, A.F. (1987) *Soil Organic Phosphorus – A Review of the Literature*. Wallingford, Oxon, UK: CAB International, p. 257

Hass, D., and Keel, C. (2003) Regulation of antibiotic production in root-colonising *Pseudomonas* spp. and relevance for biological control of plant disease. *Annu Rev Phytopathol* **41:** 117–153.

Heffer, P. (2006) *Phosphorus fertilisation: issues and outlook : proceedings [of a] paper presented to the International Fertiliser Society at a conference in Cambridge, on 14th December 2006*. York: International Fertiliser Society.

Hwan Baek, J., and Yup Lee, S. (2006) Novel gene members in the Pho regulon of Escherichia coli. *FEMS Microbiology Letters* **264**: 104–109.

Inatsuka, C.S., Julio, S.M., and Cotter, P.A. (2005) *Bordetella* filamentous hemagglutinin plays a critical role in immunomodulation, suggesting a mechanism for host specificity. *Proc Natl Acad Sci USA* **102:** 18578–18583.

Jiang, W., Metcalf, W.W., Lee, K.S., and Wanner, B.L. (1995) Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of *Salmonella typhimurium* LT2. *J Bacteriol* **177:** 6411–6421.

Johnson-Rollings, A.S., Wright, H., Masciandaro, G., Macci, C., Doni, S., Calvo-Bado, L.A., *et al*. (2014) Exploring the functional soil-microbe interface and exoenzymes through soil metaexoproteomics. *ISME J* **8:** 2148–2150.

Larson, T.J., Ehrmann, M., and Boos, W. (1983) Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the glp regulon. *J Biol Chem* **258:** 5428–5432.

Lidbury, I., Murrell, J.C., and Chen, Y. (2014) Trimethylamine *N*-oxide metabolism by abundant marine heterotrophic bacteria. *Proc Natl Acad Sci USA* **111:** 2710–2715.

Liebschner, D., Elias, M., Moniot, S., Fournier, B., Scott, K., Jelsch, C., *et al*. (2009) Elucidation of the phosphate binding mode of DING proteins revealed by subangstrom X-ray crystallography. *J Am Chem Soc* **131:** 7879–7886.

Lim, B.L., Yeung, P., Cheng, C., and Hill, J.E. (2007) Distribution and diversity of phytate-mineralizing bacteria. *ISME J* **1:** 321–330.

Liu, W., and Hulett, F.M. (1998) Comparison of PhoP binding to the *tuaA* promoter with PhoP binding to other Pho-regulon promoters establishes a *Bacillus subtilis* Pho core binding site. *Microbiology* **144:** 1443–1450.

Lugtenberg, B., and Kamilova, F. (2009) Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* **63:** 541–556.

López-Arredondo, D.L., Leyva-González, M.A., González-Morales, S.I., López-Bucio, J., and Herrera-Estrella, L. (2014) Phosphate nutrition: improving low-phosphate tolerance in crops. *Annu Rev Plant Biol* **65:** 95–123.

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., *et al*. (2005) PRIDE: the proteomics identifications database. *Proteomics* **5:** 3537–3545.

Mauchline, T.H., Fowler, J.E., East, A.K., Sartor, A.L., Zaheer, R., Hosie, A.H.F., *et al*. (2006) Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc Natl Acad Sci USA* **103:** 17933–17938.

McGrath, J.W., Chin, J.P., and Quinn, J.P. (2013) Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nat Rev Microbiol* **11:** 412–419.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Miller, S.H., Browne, P., Prigent-Combaret, C., Combes-Meynet, E., Morrissey, J.P., and O'Gara, F. (2010) Biochemical and genomic comparison of inorganic phosphate solubilization in *Pseudomonas* species. *Environ Microbiol Rep* **2:** 403–411.

Monds, R.D., Newell, P.D., Schwartzman, J.A., and O'Toole, G.A. (2006) Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. *Appl Environ Microbiol* **72:** 1910–1924.

Mudrak, B., and Tamayo, R. (2012) The *Vibrio cholerae* Pst2 phosphate transport system is upregulated in biofilms and contributes to biofilm-induced hyperinfectivity. *Infect Immun* **80:** 1794–1802.

Muth, T., Kolmeder, C.A., Salojärvi, J., Keskitalo, M., Varjosalo, M., Froukje J., *et al*. (2015) Navigating through metaproteomics data – a logbook of database searching. *Proteomics* **15:** 3439–3453.

Naseby, D.C., Way, J.A., Bainton, N.J., and Lynch, J.M. (2001) Biocontrol of *Phythium* in the pea rhizosphere by antifungal metabolite producing and non-producing *Pseudomonas* strains. *J Appl Microbiol* **90:** 421–429.

Ostrowski, M., Mazard, S., Tetu, S.G., Phillippy, K., Johnson, A., Palenik, B., *et al*. (2010) PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. *ISME J* **4:** 908–921.

Oteino, N., Lally, R.D., Kiwanuka, S., Lloyd, A., Ryan, D., Germaine, K.J., and Dowling, D.N. (2015) Plant growth promotion induced by phosphate solubilizing endophytic *Pseudomonas* isolates. *Front Microbiol* **6:** 745.

Pinchuk, G.E., Ammons, C., Culley, D.E., Li, S.-M.W., McLean, J.S., Romine, M.F., *et al*. (2008) Utilization of DNA as a sole source of phosphorus, carbon, and energy by *Shewanella* spp.: ecological and physiological implications for dissimilatory metal reduction. *Appl Environ Microbiol* **74:** 1198–1208.

Pitt, F.D., Mazard, S., Humphreys, L., and Scanlan, D.J. (2010) Functional characterization of *Synechocystis* sp. strain PCC 6803 *pst1* and *pst2* gene clusters reveals a novel strategy for phosphate uptake in a freshwater cyanobacterium. *J Bacteriol* **192:** 3512–3523.

Pratt, J.T., Ismail, A.M., and Camilli, A. (2010) PhoB regulates both environmental and virulence gene expression in *Vibrio cholerae*. *Mol Microbiol* **77:** 1595–1605.

Preston, G.M. (2004) Plant perceptions of plant growth-promoting *Pseudomonas*. *Philos Trans R Soc Lond B Biol Sci* **359:** 907–918.

Putker, F., Tommassen-van Boxtel, R., Stork, M., Rodríguez-Herva, J.J., Koster, M., and Tommassen, J. (2013) The type II secretion system (Xcp) of *Pseudomonas putida* is active and involved in the secretion of phosphatases. *Environ Microbiol* **15:** 2658–2671.

Raven, J.A. (2008) Phosphorus and the future. In: White PJ, Hammond JP, editors, The Ecophysiology of Plant-Phosphorus Interactions. Dordrecht: Springer. pp. 271–283.

Richardson, A.E., Barea, J.-M., McNeill, A.M., and Prigent-Combaret, C. (2009) Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. *Plant Soil* **321:** 305–339.

Rittmann, D., Sorger-Herrmann, U., and Wendisch, V.F. (2005) Phosphate starvation-inducible gene *ushA* encodes a 5′ nucleotidase required for growth of *Corynebacterium glutamicum* on media with nucleotides as the phosphorus source. *Appl Environ Microbiol* **71:** 4339–4344.

Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M.A., Molina-Henares, M.A., *et al*. (2013) Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. *Environ Microbiol* **15:** 780–794.

Rodrıguez, H., and Fraga, R. (1999) Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnol Adv* **17:** 319–339.

Santos-Beneit, F. (2015) The Pho regulon: a huge regulatory network in bacteria. *Front Microbiol* **6:** 402.

Scott, K., and Wu, L. (2005) Functional properties of a recombinant bacterial DING protein: comparison with a homologous human protein. *Biochim Biophys Acta Mol Cell Res* **1744:** 234–244.

Sebastian, M., Smith, A.F., Gonzalez, J.M., Fredricks, H.F., Van Mooy, B., Koblizek, M., *et al*. (2016) Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. *ISME J.* **10:** 968–978.

Shen, J., Yuan, L., Zhang, J., Li, H., Bai, Z., Chen, X., *et al*. (2011) Phosphorus dynamics: from soil to plant. *Plant Physiol* **156:** 997–1005.

Singh, B.K., and Walker, A. (2006) Microbial degradation of organophosphorus compounds. *FEMS Microbiol Rev* **30:** 428–471.

Sola-Landa, A., Rodríguez-García, A., Apel, A.K., and Martín, J.F. (2008) Target genes and structure of the direct repeats in the DNA-binding sequences of the response regulator PhoP in *Streptomyces coelicolor*. *Nucleic Acids Res* **36:** 1358–1368.

Stutter, M.I., Shand, C.A., George, T.S., Blackwell, M.S.A., Bol, R., MacKay, R.L., *et al*. (2012) Recovering phosphorus from soil: a root solution? *Environ Sci Technol* **46:**1977–1978.

Su, Z., Olman, V., and Xu, Y. (2007) Computational prediction of Pho regulons in cyanobacteria. *BMC Genomics* **8:** 156.

Sun, L., Dong, Y., Zhou, Y., Yang, M., Zhang, C., Rao, Z., and Zhang, X. (2004) Crystallization and preliminary X-ray studies of methyl parathion hydrolase from *Pseudomonas* sp. WBC-3. *Acta Crystallogr C Struct Chem D* **60:** 954–956.

Tan, A.S.P., and Worobec, E.A. (1993) Isolation and characterization of two immunochemically distinct alkaline phosphatases from *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* **106:** 281–286.

Ternan, N.G., and Quinn, J.P. (1998) Phosphate starvation-independent 2-aminoethylphosphonic acid biodegradation in a newly isolated strain of *Pseudomonas putida*, NG2. *Syst Appl Microbiol* **21:** 346–352.

Tetu, S.G., Brahamsha, B., Johnson, D.A., Tai, V., Phillippy, K., Palenik, B., and Paulsen, I.T. (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* **3:** 835–849.

Turner, B.L., and Leytem, A.B. (2004) Phosphorus compounds in sequential extracts of animal manures: chemical speciation and a novel factionation procedure. *Environ Sci Technol* **38:** 6101–6108.

Turner, B.L., Richardson, A.E., and Mullaney, E.J. (2002) Inositol phosphates in the environment. *Philos Trans R Soc Lond B Biol Sci* **357:** 449–469.

Vance, C.P., Uhde-Stone, C., and Allan, D.L. (2003) Phosphorus acquisition and use: critical adaptations by plants for

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

securing a nonrenewable resource. *New Phytol* **157:** 423–447.
Vassilev, N., Vassileva, M., and Nikolaeva, I. (2006) Simultaneous P-solubilizing and biocontrol activity of microorganisms: potentials and future trends. *Appl Microbiol Biotechnol* **71:** 137–144.
Villarreal-Chiu, J.F., Quinn, J.P., and McGrath, J.W. (2012) The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Front Microbiol* **3:** 19.
Wanner, B.L. (1990) Phosphorus assimilation and its control of gene expression in *Escherichia coli*. In *The Molecular Basis of Bacterial Metabolism*. Hauska, G., and Thauer, R. (eds). Berlin, Heidelberg: Springer, pp. 152–163.
White, A.K., and Metcalf, W.W. (2004) The *htx* and *ptx* operons of *Pseudomonas stutzeri* WM88 are new members of the Pho regulon. *J Bacteriol* **186:** 5876–5882.
White, A.K., and Metcalf, W.W. (2007) Microbial metabolism of reduced phosphorus compounds. *Annu Rev Microbiol* **61:** 379–400.
White, P.J., and Hammond, J.P. (2008) Phosphorus nutrition of terrestrial plants. In *The Ecophysiology of Plant-Phosphorus Interactions*. White, P.J., and Hammond, J.P. (eds). Dordrecht, Netherlands: Springer, pp. 51–81.
White, P.J., and Hammond, J.P. (2009) The sources of phosphorus in the waters of Great Britain. *J Environ Qual* **38:** 13–26.
Yu, H., Yuan, M., Lu, W., Yang, J., Dai, S., Li, Q., *et al.* (2011) Complete genome sequence of the nitrogen-fixing and rhizosphere-associated bacterium *Pseudomonas stutzeri* strain DSM4166. *J Bacteriol* **193:** 3422–3423.
Zalkin, H., and Nygaard, P. (1996) Biosynthesis of purine nucleotides. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., *et al.* (eds). Washington, DC: ASM Press, pp. 561–579.
Zavaleta-Pastor, M., Sohlenkamp, C., Gao, J.-L., Guan, Z., Zaheer, R., Finan, T.M., *et al.* (2010) *Sinorhizobium meliloti* phospholipase C required for lipid remodeling during phosphorus limitation. *Proc Natl Acad Sci USA* **107:** 302–307.
Zhang, X.-X., Scott, K., Meffin, R., and Rainey, P.B. (2007) Genetic characterization of *psp* encoding the DING protein in *Pseudomonas fluorescens* SBW25. *BMC Microbiol* **7:** 114.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Proportion of the extracellular and intracellular proteins detected in the four protein fractions extracted during this study. Only the top-60 most abundant proteins were included in the analyses. The values displayed are taken from Pi-deplete and Pi-replete cultures. Results presented are the mean of triplicate cultures.
Fig. S2. The abundance of proteins in both the high Pi and low Pi exoproteomes of the three *Pseudomonas* strains. **(A)** *Pseudomonas fluorescens* SBW25, **(B)** *Pseudomonas putida* BIRD-1, **(C)** *Pseudomonas stutzeri* DSM4166. Results are the mean of triplicate cultures and error bars denote standard deviation.
Fig. S3. Conservation of the key residues (highlighted in red) involved in phosphate binding among the periplasmic binding proteins containing the domain, Pfam12849- PBP. Locus tags are used as the identifier. Abbreviations: VP, *Vibrio parahaemolyticus*; VC/VCA, *V. cholerae*; VAA, *V. anguillarum*; *V. harveyi* MYO, *Synechocystis* sp. PCC6803; PFLU, *P. fluorescens*; PA, *P. aeruginosa*; PPUBIRD1, *P. putida*; PSTAA, *P. stutzeri*; Psyr, *P. syringae*; EcDH1, *E. coli*; *P. Antarctica*; Smc; *Ensifer meliloti*.
Fig. S4. Semi-quantitative abundance analysis of the putative phosphonate substrate binding proteins detected in the exoproteomes of the three *Pseudomonas* strains. and *the phoBR* mutant. Results presented are the mean of triplicate cultures. Error bars denote standard deviation.
Fig. S5 Growth of the *phoBR* mutant strain of *P. putida* BIRD-1.
A. A comparison of the *phoBR* mutant grown under Pi-replete (Black circles) and Pi-deplete (Grey circles) conditions. Concentrations of Pi were the same as those used for the wild type. Black arrows indicated the times of sampling for proteomics and exoproteomics. The striped arrow indicates the addition of Pi (50 μM) to help generate enough biomass for sampling.
B. Growth yields of either the wild type or *phoBR* mutant sampled after 48 hours grown on Pi-replete or Pi-deplete growth media. Results presented are the mean of triplicate cultures. Error bars denote standard deviation.
Fig. S6. Evolutionary relationships of PhoX-like homologs. The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch length = 3.66263884 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method [2] and are in the units of the number of amino acid differences per site. The analysis involved 27 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 842 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [3].
Table S1. A rank-abundance profile of the identified proteins in the exoproteome of *Pseudomonas putida* BIRD-1.
Table S2. A rank-abundance profile of the identified proteins in the exoproteome of *Pseudomonas fluorescens* SBW25.
Table S3. A rank-abundance profile of the identified proteins in the exoproteome of *Pseudomonas stutzeri* DSM4166.
Table S4. A rank-abundance profile of the identified proteins in the proteome of *P. putida* BIRD-1.
Table S5. A rank-abundance profile of the identified proteins in the cellular proteome of the *P. putida* BIRD-1 *phoBR* mutant.
Table S6. A rank-abundance profile of the identified proteins in the exoproteome of the *P. putida* BIRD-1 *phoBR* mutant.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

The ISME Journal (2021) 15:1040–1055
https://doi.org/10.1038/s41396-020-00829-2

ARTICLE

Niche-adaptation in plant-associated *Bacteroidetes* favours specialisation in organic phosphorus mineralisation

Ian D. E. A. Lidbury[1,2] · Chiara Borsetto[1] · Andrew R. J. Murphy[1] · Andrew Bottrill[1] · Alexandra M. E. Jones[1] · Gary D. Bending[1] · John P. Hammond[3] · Yin Chen[1] · Elizabeth M. H. Wellington[1] · David J. Scanlan[1]

Received: 24 July 2020 / Accepted: 30 October 2020 / Published online: 30 November 2020
© The Author(s) 2020. This article is published with open access

Abstract

Bacteroidetes are abundant pathogen-suppressing members of the plant microbiome that contribute prominently to rhizosphere phosphorus mobilisation, a frequent growth-limiting nutrient in this niche. However, the genetic traits underpinning their success in this niche remain largely unknown, particularly regarding their phosphorus acquisition strategies. By combining cultivation, multi-layered omics and biochemical analyses we first discovered that all plant-associated *Bacteroidetes* express constitutive phosphatase activity, linked to the ubiquitous possession of a unique phosphatase, PafA. For the first time, we also reveal a subset of *Bacteroidetes* outer membrane SusCD-like complexes, typically associated with carbon acquisition, and several TonB-dependent transporters, are induced during Pi-depletion. Furthermore, in response to phosphate depletion, the plant-associated *Flavobacterium* used in this study expressed many previously characterised and novel proteins targeting organic phosphorus. Collectively, these enzymes exhibited superior phosphatase activity compared to plant-associated *Pseudomonas* spp. Importantly, several of the novel low-Pi-inducible phosphatases and transporters, belong to the *Bacteroidetes* auxiliary genome and are an adaptive genomic signature of plant-associated strains. In conclusion, niche adaptation to the plant microbiome thus appears to have resulted in the acquisition of unique phosphorus scavenging loci in *Bacteroidetes*, enhancing their phosphorus acquisition capabilities. These traits may enable their success in the rhizosphere and also present exciting avenues to develop sustainable agriculture.

Introduction

Flavobacteriaceae belong to the phylum *Bacteroidetes* and are dominant members of plant/soil and ocean microbiomes [1–5]. In the plant microbiome (rhizosphere, endosphere and phyllosphere), the abundance of *Flavobacteriaceae* and other members of the *Bacteroidetes* phylum is generally orders of magnitude greater than that of the surrounding bulk soil [2, 3, 6, 7]. *Flavobacteriaceae* typically represent 5–65% of the microbial community associated with various agriculturally important crops [8] and exhibit functionally important effects on plant health [8–11]. In Barley, 25% of the isolates obtained from the rhizosphere were related to *Flavobacteriaceae* and these contributed almost 50% of the potential microbial P-turnover [12]. In contrast to other important plant-associated microbiota [13–16], the genetic traits responsible for the success of *Flavobacteriaceae* in the plant microbiome, including their phosphorus acquisition strategies, have not been resolved.

In ocean and animal microbiomes, *Bacteroidetes* are key regulators of carbon cycling, and thus microbiome functioning, due to their enhanced ability to degrade complex algal and plant-derived polysaccharides [4, 5, 17]. Terrestrial *Flavobacteriaceae* have also been predicted to degrade plant-derived glycans due to the high number of glycan-specific hydrolytic enzymes encoded in their genomes

These authors contributed equally: Ian D. E. A. Lidbury, Chiara Borsetto.

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-00829-2) contains supplementary material, which is available to authorized users.

✉ Ian D. E. A. Lidbury
I.lidbury@sheffield.ac.uk

[1] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, UK

[2] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

[3] School of Agriculture, Policy, and Development, University of Reading, Earley Gate, Whiteknights, Reading, UK

[8, 18]. Polysaccharide degradation in *Bacteroidetes* utilises specialised gene clusters, termed polysaccharide utilisation loci (PULs) [19]. A prominent feature of PULs includes a two-component outer membrane (OM) transport system-termed SusCD (archetypal form, **S**tarch **u**tilisation **s**ystem). SusC is a transmembrane TonB-dependent transporter (TBDT) and SusD is the ligand-binding lipoprotein cap [20]. The specificity of the SusCD-like-complexes allows *Bacteroidetes* to grow on a variety of plant- [21, 22] algal- [23] and fungal-derived polysaccharides [24, 25]. The latter category includes the sole example of an entire PUL characterised in *Flavobacterium*.

For all biota on Earth, P is an essential macronutrient and its availability directly affects the global carbon cycle as well as governing global crop production [26–30]. Modern agriculture massively relies on the use of unsustainable chemical fertilisers, which has now created an emerging global phosphorus crisis [26, 31]. In soils, P exists in many immobilised organic and inorganic complexes whilst only a small fraction (<1%) exists in solution as orthophosphate (Pi) [32–34]. The rate plants take up labile Pi into their roots far exceeds the rate at which diffusion can replenish soil Pi surrounding the roots [32, 33]. This creates localised regions of Pi-depletion at the soil–plant interface (a region commonly referred to as the rhizosphere) [35]. Organic P (P_o) often accounts for the majority of total P in soils and soil microorganisms play a key role in solubilising this fraction before it becomes available to plants [33, 34, 36]. However, rhizosphere microbes also contribute to the immobilisation of labile Pi, thus competing with plants for P [32, 37]. Therefore, developing a greater understanding of soil microbial P dynamics at the soil–root interface is critical for improving sustainable agriculture [33, 38, 39] and improving models regarding plant's efficiency to sequester anthropogenic CO_2 emissions [29, 30].

Bacteria have evolved numerous strategies to overcome environmental Pi scarcity, which are almost exclusively regulated by a two-component system, usually referred to as PhoBR (PhoB, DNA binding response regulator; PhoR, transmembrane histidine kinase) [13]. Several P-scavenging strategies are common to almost all bacteria, the most common being possession of a high affinity Pi transport system, PstSABC, as well as inducing periplasmic and OM-bound phosphatases [13, 15, 40]. Several studies have shown that Bacteria predominantly secrete non-specific alkaline phosphatases (APases) comprising three distinct forms (PhoA, PhoD and PhoX) whilst a few cultured strains secrete acid phosphatases (AcPases, class I, II and III) in response to Pi limitation [13–16, 41]. However, recent evidence from genomic and metagenomic datasets suggest only half possess these characterised P-mobilising enzymes [42].

Here, we isolated *Bacteroidetes* strains from the rhizosphere of the economically important crop *Brassica napus* L. (Oilseed Rape, OSR), grown under field conditions, to determine the Pi-mobilisation potential of this enigmatic phylum. After initially observing atypical constitutive phosphatase activity in all our *Bacteroidetes* isolates we performed a comprehensive multi-layered omics approach to identify the genetic loci responsible for this unique trait. In doing so we also discovered *Flavobacterium* possess an extremely high potential for transforming rhizosphere-associated P_o through unique uptake and degradation systems.

Materials and methods

Isolation and growth medium used for cultivating *Bacteroidetes*

Flavobacterium johnsoniae DSM2064 (UW101) was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). *Flavobacterium* sp. F52, a bell pepper rhizobacterium [43] and *Flavobacterium* sp. L0A5, a desert plant rhizobacterium (Cytryn, unpublished data) were both kindly donated from the Cytryn Lab, Agricultural Research Organisation, Israel. Strains were routinely maintained on casitone yeast extract medium (CYE) [44]. *Pseudomonas* strains, previously described in [15], were maintained in Luria-Bertani medium. To isolate all other strains a modified R2A medium developed by [45] was used. Cyclohexamide (100 mg L^{-1}) and Kanamycin (40 mg L^{-1}) were added to reduce fungal and non-Flavobacterial growth, respectively. To investigate the effect of Pi-depletion on the *Flavobacterium* strains, each was grown ($n = 3$) in a Minimal A medium adapted from [15]. Carbon source, glucose 3.6 g L^{-1} and 20 mM Bis/Tris buffer pH 7.2 was used instead of HEPES and KH_2PO_4 added to a final concentration of either 50 µM or 1.4 mM. A final concentration of NH_4Cl 1 mM, $FeCl_2$ 10 mg L^{-1} and KH_2PO_4 1.4 mM was used for the N-limited medium. A final concentration of NH_4Cl 8.4 mM and KH_2PO_4 1.4 mM with no Fe added was used for the Fe-limited medium. The control contained an excess of all nutrients. Cells were harvested at mid exponential phase (OD_{600nm} 0.4–0.8) after 16–20 h.

Isolation of *Bacteroidetes*

For isolation, two field trials growing *Brassica napus* L. (OSR) were utilised. The first was located at Warwick Crop Centre (April 2017), University of Warwick, Wellesbourne and the other was located at Sonning Farm, University of Reading (Sept. 2017). Samples were collected during the flowering stage at Warwick Crop Centre and during the four-leaf stage at Sonning Farm. Plants were removed from soil and shaken to remove all loosely adhering soil. Roots

containing tightly adhering soil were cut and washed (shaken vigorously) in 10 mL phosphate-buffered saline solution to remove rhizosphere soil. A serial dilution of the resulting soil solution was set up and plated onto either modified R2A or CYE. Individual colonies were picked and sub-cultured prior to identification by 16S rRNA gene analysis.

Genome sequencing of the *Flavobacterium* isolates

The OSR isolates were grown on CYE and cells were collected from plates by scraping off bacterial colonies. Cells were then sent to the Microbes NG unit (https://microbesng.uk/) at the University of Birmingham and data are deposited in the European Nucleotide Archive database. For assistance, amino acid sequences for the complete CDS profile of all newly isolated strains are attached to the Tables file (S16–S22).

Quantification of phosphatase activity

The protocol was adapted from Lidbury et al. [15]. Briefly, 0.5 ml cell cultures ($n = 3$) for both Pi-replete and Pi-deplete growth conditions were directly incubated (30°C at 160 rpm) with 4 mM *para*-nitrophenyl phosphate (*p*NPP) using a 100 mM *p*NPP stock solution in 20 mM Tris HCl, pH 7.2. The pH of the reaction mixture was equivalent to the culture medium and was not altered by addition of *p*NPP. The reaction was stopped using 2 mM NaOH once a colour change was detected. A colour change typically occurred within 5–20 min unless activity was very low (*Pseudomonas* spp.), in which case reactions were stopped after 1 h. All raw absorbance values were lower than A405nm 1. For each strain and growth condition, A405nm measurements were corrected by subtracting A_{405nm} measurements for reactions immediately stopped with NaOH. Normalisation against the culture optical density (OD_{600}) was performed and the rate was calculated and expressed h^{-1}. A standard curve for *para*-nitrophenol was generated using a range of known concentrations (0.25, 0.5, 1, 2, 4 mM). A qualitative plate assay using the substrate XP (5-Bromo-4-chloro-3-indolyl phosphate) was also established to confirm phosphatase activity.

Multi-omics analysis of nutrient-limited *F. johnsoniae*

To capture the transcriptome, RNA was extracted using the RNeasy Mini Kit (QIAGEN) following the manufacturer's instructions. RNA-seq library preparation, sequencing and standard bioinformatics analysis were performed by Novogene according to the company pipeline. A full description of the method is provided in Supplementary materials. To capture the cellular proteome (CP) and exoproteome (XP) we followed the method described by Lidbury et al. [15]. A full description of the method is provided in Supplementary materials.

Protein fractionation of the soluble *F. johnsoniae* CP

Cells (500 ml culture) were grown under Pi-deplete conditions and harvested as already described above. Detailed methodology for all three fractionation steps are described in Supplementary materials. An initial fractionation step was performed using size-exclusion chromatography via gel filtration. Fractions were assayed for phosphatase activity as described above. 1D-SDS PAGE analysis revealed the active fractions (peaks 1 and 2) contained a complex mixture of proteins, therefore a second fractionation step was performed on each peak. Both phosphonate-affinity chromatography (peak 1) and anion-exchange (Na gradient) chromatography (peak 2) was performed prior to activity assays and peptide identification.

Comparative genomics analyses

The online platform **I**ntegrated **M**icrobial **G**enomes & Microbiomes server at the **J**oint **G**enome **I**nstitute (IMG/JGI) was used to perform the majority of comparative genomics analyses described in this study. Genomes were stored in Genome sets and BLAST searches (Min. similarity 30%, *E*-value e^{-50}) were set up using the 'jobs function'.

To determine genome completeness of the OSR isolates, orthologs of core metabolic genes [46] were identified in DSM2064 and F52 using IMG/JGI. Local BLASTP was then performed on the OSR isolates and LAO-5 using DSM2064 orthologs as the query. To generate and annotate ORFs, a local version of the Prokka pipeline [47] was performed. Comparative genomic analyses of the *Flavobacterium* strains used in this study were performed using a local version of BLAST+. In most cases ORFs found in DSM2064 were used as the query, unless this strain lacked a given ORF. A relatively high stringency was set (e^{-90}). Any hits with low scores were manually checked to determine whether they were true orthologs or paralogs.

Structural homology analysis of the various uncharacterised Pi-acquisition proteins was performed using the online servers for SWISS-model homologue analysis (https://swissmodel.expasy.org/) and **P**rotein **H**omology/analogy **R**ecognition **E**ngine Version 2.0 (PHYRE2, http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). All searches were performed using the default parameters.

Phylogenetic analyses were performed using IQ-Tree [48] using the parameters -m TEST -bb 1000 -alrt 1000. Thus, the most suitable model was chosen by the software.

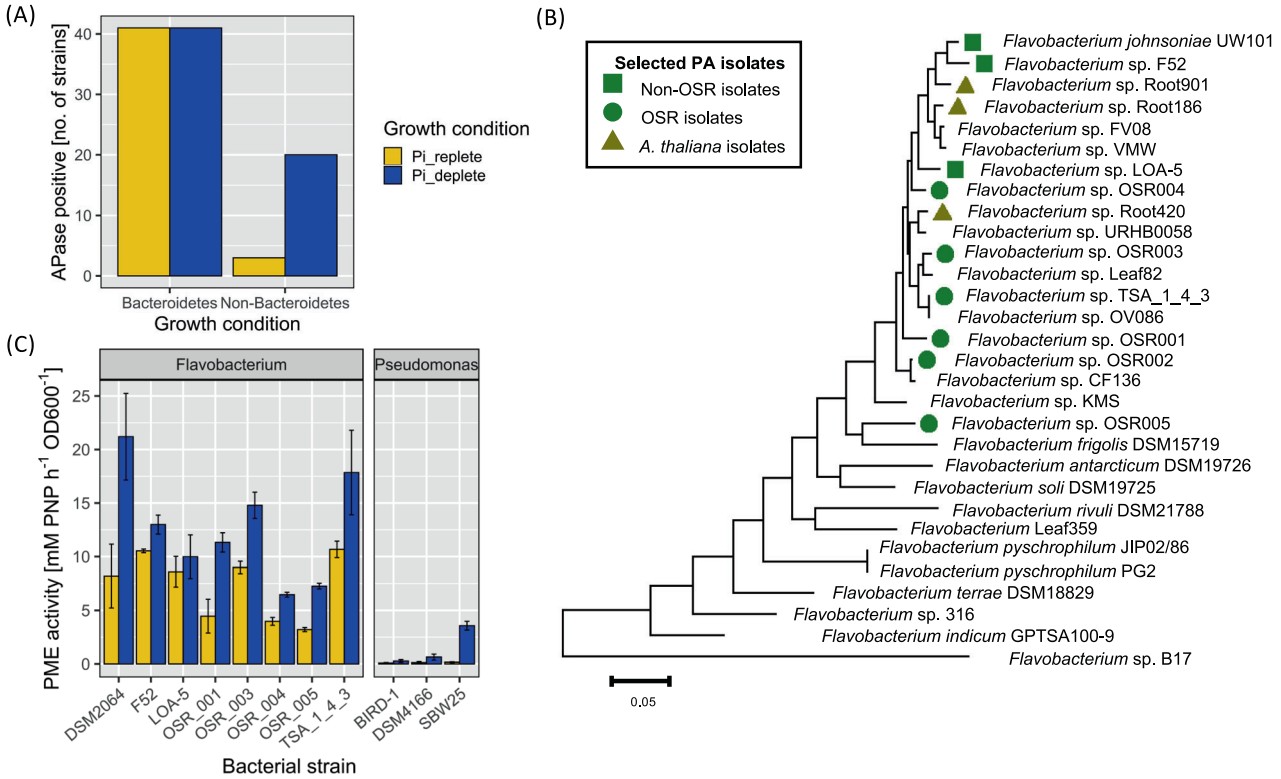


Fig. 1 Phosphomonoesterase activity observed in plant-associated *Bacteroidetes*. **A** Phosphatase (monoesterase) activity of *Bacteroidetes* ($n = 42$) and non-*Bacteroidetes* strains ($n = 20$) isolated from the rhizosphere of field-grown Oilseed Rape (OSR) grown on glucose as the carbon source. **B** Phylogenetic analysis of *Flavobacterium* isolates using a multi-locus approach with ten single-copy core genes (as previously described by [18]). Evolutionary distances were inferred using the maximum-likelihood method in MEGA 7. **C** Phosphomonoesterase activity of whole-genome-sequenced OSR strains, non-OSR *Flavobacterium* strains and three *Pseudomonas* rhizosphere strains grown on glucose as the carbon source. Results are the mean of triplicate assays and error bars denote ± standard deviation.

Evolutionary distances were inferred using maximum-likelihood analysis. Relationships were visualised using the online platform the **I**nteractive **T**ree **o**f **L**ife viewer (https://itol.embl.de/).

All statistical analyses and data visualisation were performed using the vegan, ggplot2, ggfortify, tidyr, plyr, serration packages in Rstudio (1.2.5033).

Results

Root-associated *Bacteroidetes* display 'unusual' constitutive and strong inducible phosphatase activity

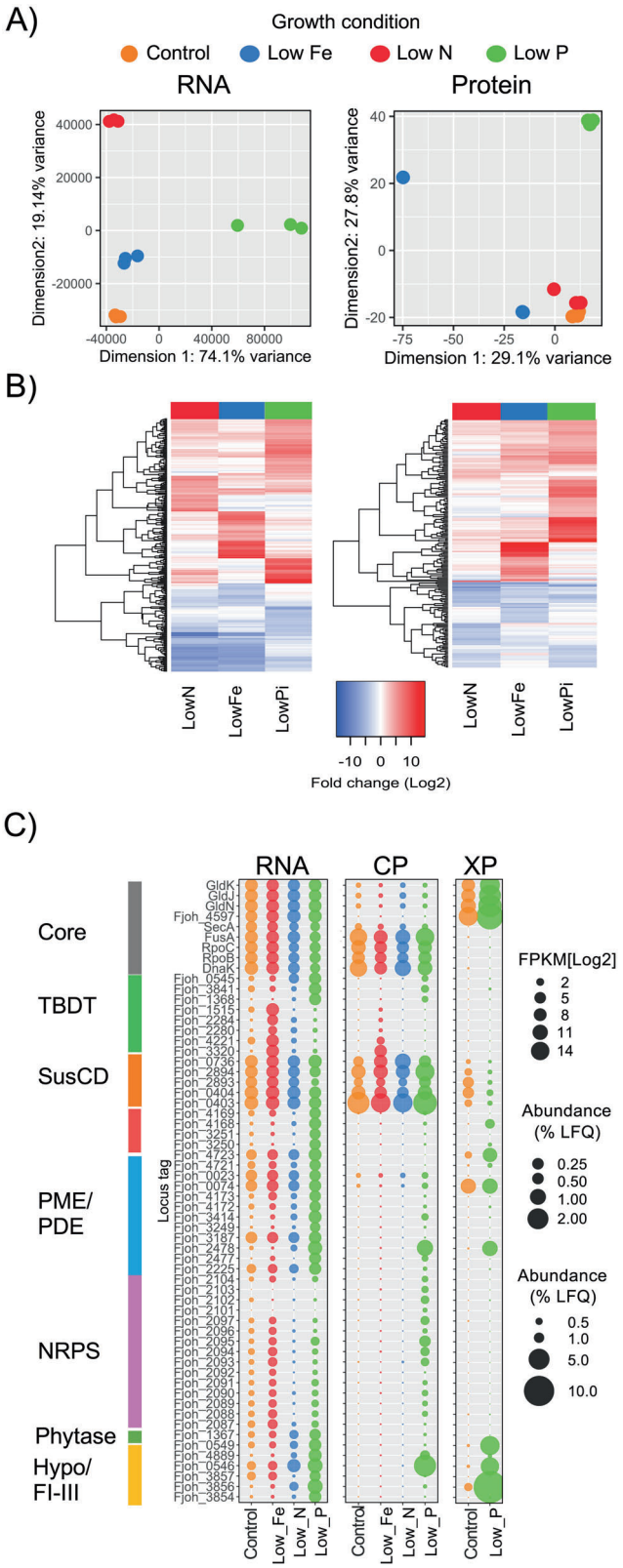
Using a selective medium [43], we isolated numerous *Bacteroidetes* strains from field-grown OSR rhizosphere soil (Two locations; Sonning Farm, Reading, UK and Warwick Crop Centre, Wellesbourne, UK). Most isolated *Bacteroidetes* were related to *Flavobacteriaceae* and to a lesser extent *Sphingobacteriaceae* (Fig. S1). In addition, several kanamycin-resistant non-*Bacteroidetes* strains were still isolated. A random subset of 42 *Bacteroidetes* isolates and 20 non-*Bacteroidetes* (predominantly *Gammaproteobacteria*) were screened for phosphatase (phosphomonoesterase [PME]) activity (Fig. 1A). Strikingly, all *Bacteroidetes* isolates tested exhibited constitutive phosphatase activity i.e. irrespective of Pi availability, in contrast to just 3/20 non-*Bacteroidetes* isolates. All OSR isolates exhibited inducible phosphatase activity when grown under Pi-deplete conditions.

Five OSR rhizosphere strains, representing the breadth of terrestrial *Flavobacterium* diversity (Fig. 1B), were selected for further analysis of their in vitro PME activity. For comparison, the model soil bacterium *Flavobacterium johnsoniae* DSM2064 (hereafter DSM2064) [42], two previously-isolated rhizosphere strains *Flavobacterium* sp. F52 (F52) and *Flavobacterium* sp. L0A5 (LOA5) [41], and three plant-growth promoting *Pseudomonas* spp. strains [15, 49] were also assayed. After overnight growth on glucose as a carbon source, due to the conversion of excess sugar to organic acid, the pH of the medium for all strains dropped to pH 6.1 ± 0.4. Whilst the strength of both the constitutive and inducible phosphatase activity varied, all *Flavobacterium* strains exhibited constitutive activity (Fig. 1C). The *Pseudomonas* isolates only possessed inducible phosphatase activity, consistent with their possession

of well-known Pi-sensitive phosphatases [15]. Furthermore, *Flavobacterium* inducible phosphatase activity was far superior to the *Pseudomonas* strains (Fig. 1C).

Phosphate depletion elicits a major regulatory response in *Flavobacterium johnsoniae*

To determine the phosphatases responsible for generating both the inducible and constitutive phosphatase activity, the model bacterium DSM2064 was challenged with Pi-limiting growth conditions and subjected to both transcriptomic and proteomic analyses. The latter comprised analysis of both CP and XP. Iron (Fe) and nitrogen (N)-limiting growth conditions were also established to act as additional controls in addition to the high Pi/Fe/N controls (Fig. 2). DSM2064 exhibited a clear transcriptomic response to all three growth-limiting conditions, but the proteomic response was most pronounced during Pi-depletion (Fig. 2A). Specific genes and gene clusters were differentially expressed under the three distinct nutrient-limiting growth conditions (Fig. 2B, C, Tables S1 and S2). Fe-regulated and N-regulated genes and their proteins are presented in Tables S1 and S2. Pi-depletion resulted in the greatest number of upregulated transcripts ($_{log2}$FC > 4, Q value $P < 0.05$; Pi = 95 [avg. FC = 6.54], N = 57 [avg. FC = 4.95], Fe = 47 [avg. FC = 7.69]) and corresponding proteins. This included an unusually high number of genes predicted to encode PMEs and phosphodiesterases (PDEs) (Table 1). Whilst several P-responsive loci have no known function, their tight repression under N and Fe limitation suggests these unique proteins play a primary role in scavenging P. These uncharacterised low-Pi-inducible loci included four hypothetical lipoproteins (Fjoh_0546, Fjoh_0549, Fjoh_3856, Fjoh_4889), two of which were located downstream of low-Pi-inducible TBDTs. Significantly, three of these hypothetical lipoproteins represented a major fraction of the Pi-deplete XP (Table 1 and Fig. 2C). Whilst some OM-transport systems related to TBDTs and SusCD-like complexes were either constitutively expressed or induced in response to Fe-depletion (archetypal function), we identified three TBDTs and two SusCD-like complexes (hereafter termed **P**hosphate **u**tilisation **s**ystem (Pus)), that were induced under Pi-depletion only (Table 1 and Fig. 2C). Interestingly, we also observed the expression of a non-ribosomal peptide cluster in response to Pi-depletion whose expression appeared to be post-transcriptionally regulated (Fig. 2C).



F. johnsoniae DSM2064 expresses a high number of phosphatases

Further in silico analyses of the putative PMEs and PDEs revealed DSM2064 has a very high potential to mineralise P_o (Table 1). Some were low-Pi-inducible whilst others were constitutively expressed, suggesting a constant reliance on P-acquisition from Po. We previously demonstrated that PME activity in *Pseudomonas* spp. was produced by

◀ **Fig. 2 Transcriptomic and proteomic analyses of *F. johnsoniae* (DSM2064) in response to growth under different nutrient stress conditions. A** Multidimensional scaling (euclidean distance) analysis of RNA and protein content in DSM2064 sampled after 18 h growth. One Fe replicate contained much less protein, hence its separation in the multidimensional scaling. **B** Selected genes and corresponding proteins showing significant (T test = $P < 0.05$, adjusted FDR = 0.05) differential expression/abundance from the control growth condition. **C** The relative abundance of selected genes in the transcriptome (% abundance based on FPKM [Log_2]) and corresponding proteins (% abundance based on raw label-free quantitative values) detected in either the total cellular proteome (CP) or exoproteome (XP). Genes/proteins were categorised into various functional guilds i.e. Core/housekeeping (grey), TonB-dependent transporter (TBDT, green), SusCD-like transport system (red), phosphatase-like enzymes (sky blue), a non-ribosomal secondary metabolite cluster (NRPS, purple), phytase (dark green), hypothetical lipoproteins, FI/II/III (orange). Results in (B) and (C) are the mean of triplicate cultures.

non-cytoplasmic PMEs, indicating *p*NPP does not cross the cytoplasmic membrane [15, 50]. Most PMEs and PDEs were predicted to possess N-terminal signal peptide sequences (periplasmic) and a subset of these also possessed predicted lipobox sequences (OM-anchored). Four putative PMEs likely contributed to the observable PME activity in DSM2064. Sequence and structural homology searches revealed two predicted periplasmic PMEs (Fjoh_3187, Fjoh_3249) were related to PhoA [51] (Fig. S2) and one predicted OM-anchored PME (Fjoh_2478) was distantly related to PhoX [16] (Fig. S3). Fjoh_3249 (PhoA1) Fjoh_3187 (PhoA2) had varied expression profiles and neither were abundant in the CP or XP (Table 1 and Fig. 2C). Fjoh_2478 was heavily transcribed and represented a major component of the XP during Pi-depletion (Table 1 and Fig. 2C). 9/10 key residues found in the PhoX active site were conserved in Fjoh_2478 (Fig. S3). However, phylogenetic analysis revealed *Flavobacterium* PhoX is distinct from the previously defined PhoXI or PhoXII clades [16] (Fig. S4). Indeed, the predicted active site architecture for DSM2064 PhoX and surface electrostatic potential are quite different to previously characterised PhoX (Fig. S3), suggesting a potential difference in enzyme kinetics, pH range and substrate preference [52]. Similar to *Pseudomonas* spp. OM-anchored PMEs [45], we hypothesised Fjoh_2478 contributed most of the inducible PME activity in DSM2064. The fourth putative PME (periplasmic), Fjoh_0023, was expressed in all growth conditions, irrespective of Pi availability (Fig. 2C and Table 1). This predicted periplasmic phosphatase has high sequence (44% identity, $e = 5^{e-159}$) and structural homology (Fig. S5) to the Pi-irrepressible phosphatase PafA, identified in the human opportunistic pathogen *Elizabethkingia* (formerly *Chryseobacterium*) *meningosepticum* [53].

PhoA, PafA and PhoX all mineralise *p*NPP in vitro [16, 40, 50, 53]. Therefore, to determine if these four putative PMEs contributed to PME activity in DSM2064, the soluble CP, extracted from cells grown under Pi-depletion, was fractionated prior to screening for PME activity. Two major peaks of PME activity were detected across the fractionated CP and both of these were subjected to a further fractionation step (See Supplementary methods and Fig. S6). For peak 1, Fjoh_0023 was the major protein whose abundance was positively correlated with PME activity ($P < 0.01$, Fig. S6, Table S4). For peak 2, PhoA2 (Fjoh_3187) was the second most abundant protein detected in the active fraction, whilst another inducible periplasmic APase Fjoh_2225 (Table 1) was also detected, albeit at a lower abundance (Fig. S6 and Table S5).

F. johnsoniae expresses distinct polysaccharide utilisation loci-type gene clusters in response to Pi-depletion

Bacteroidetes are best-known for their possession of numerous PULs, typified by the presence of *susCD-like* complexes, enabling them to efficiently acquire and degrade complex polymeric carbon molecules [8, 19]. We identified four gene clusters, heavily induced during Pi-depletion, that are predicted to share similar mechanisms with traditional PULs for OM-binding, OM-transport, and periplasmic degradation of organic substrates (Fig. 3). The first two, labelled PUSI and PUSII, possessed the two low-Pi-inducible PusCD complexes, PusCD1 (Fjoh_4168/9) and PusCD2 (Fjoh_3250/1) (Fig. 3). Unlike traditional PULs that contain hydrolytic enzymes found in the CAZy database [22, 54], PUSI and PUSII were co-localised with two of the differentially expressed periplasmic PDEs (PUSI, Fjoh_4172; Fjoh_4173), or the PME, PhoA1 (PUSII, Fjoh_3249). PhoA1 is predicted to possess an additional domain (pf13653), unique for PhoA, potentially extending its substrate range to target specific phosphodiesters (Table 1). In agreement, PhoA1 is predicted to have a more accessible active site, a prerequisite for catalysis of phosphodiesters [55] (Fig. S2).

Two more PUS gene clusters (here labelled PUSIII and PUSIV) were also identified. PUSIII contained a hypothetical lipoprotein, hereafter termed F(form)I, encoded by *fjoh_3856*, that was the most differentially transcribed gene and the most abundant exoprotein (18.7%) suggesting an important role in combatting Pi-depletion (Table 1 and Fig. 2C). The remaining ORFs in PUSIII encoded a TBDT/SusC-like porin, various OM hypothetical proteins, an OM protease and a putative cytoplasmic PDE, all of which were induced in response to Pi-depletion (Fig. 3). In PUSIV, the ORF encoding another related hypothetical lipoprotein, hereafter termed F(form)II, is co-localised in a Pi-sensitive gene cluster containing *phoBR* and ORFs encoding other heavily expressed proteins (Fig. 3). FI, FII and a third related orphan lipoprotein, termed FIII, all

Table 1 Expression (transcriptomic and proteomic) of P-acquisition and P-stress response loci in *F. johnsoniae* DSM0264.

						Transcriptome		Whole cell proteome			Exoproteome		
Predicted function	Locus_tag	Gene symbol/description	Genomic context	Predicted localisation	Pfam accession	FC	FDR (*p*-adj)	DE	−Log *p* value	*q* value	DE	−Log *p* value	*q* value
Phosphomonoesterase	Fjoh_0023	PafA	Orphan	Periplasm	pf01663	0.26	5.52E−01	0.82	3.61	0.00130526	-2.72	2.28	0.00
	Fjoh_1367	Phytase	Phytase	OM	pf02333	5.76	2.01E−36	6.83	0.51	9.31E−03	ND	ND	ND
	Fjoh_2478	PhoX_like	PhoX	OM	pf05787	12.50	2.74E−142	11.34	0.28	0.00E+00	8.23	4.22	0.00E+00
	Fjoh_2225	Putative APase	Orphan	Periplasm	pf01676	4.25	7.07E−34	6.20	0.39	0.00E+00	6.63	4.04	0.00E+00
	Fjoh_3187	PhoA2	Orphan	Periplasm	pf00245	1.59	0.28508	3.89	3.66	1.36E-03	0.03	0.01	9.73E-1
	Fjoh_3249	PhoA1 (putative diesterase activity)	PUSII	Periplasm	pf00245, pf13653	7.40	4.39E−61	4.87	0.11	0.00E+00	ND	ND	ND
	Fjoh_3414	Hypothetical Lipoprotein - putative phosphatase	Orphan	OM/IM	IPR011044, IPR011048, IPR015943	10.32	1.41E−90	9.09	0.08	0.00E+00	5.00	2.53	2.71E-03
Phosphodiesterase	Fjoh_0074	Endo/exonuclease/phosphatase	Orphan	OM	NA; IPR011050	1.15	4.17E−02	1.43	2.86	6.64E−03	0.57	0.56	0.23
	Fjoh_3851	Polyphosphate kinase_like/ PlcP D	FI cluster/ PUSIII	Cytoplasm	pf02503, pf13089, pf13090	7.80	6.37E−75	ND	ND	ND	ND	ND	ND
	Fjoh_4172	Phosphoinositide phospholipase C, Ca^{2+}-dependent, phosphodiesterases	PUSI	Periplasm	pf00149	6.44	2.87E−56	6.89	0.94	2.11E−03	6.03	1.03	6.15E-02
	Fjoh_4173	3′,5′-cyclic AMP phosphodiesterase CpdA	PUSI	Periplasm	pf16670	7.17	4.47E−69	4.48	0.37	3.79E−03	2.18	2.03	5.44E-03
	Fjoh_4721	Extracellular nuclease, EndA1	Orphan	OM	pf00041, pf00932, pf04231	4.66	3.19E−27	0.71	3.50	2.17E−02	3.70	3.92	0.00E+00
	Fjoh_4723	Extracellular nuclease, EndA2	Orphan	OM	pf00041, pf00932, pf04231	4.13	1.02E−32	ND	ND	ND	2.88	4.34	0.00E+00
Transporter/Binding	Fjoh_0542	OprP	PhoBR operon	OM	pf07396	-1.3	0.24533	ND	ND	ND	ND	ND	ND
	Fjoh_0545	Tdbt	PhoBR operon	OM	pf07715, pf13715, pf14905	8.02	8.21E−84	6.52	0.49	7.91E-03	5.38	5.68	0.00E+00
	Fjoh_0546	Hypothetical Lipoprotein - FII	PhoBR operon	OM	NA; IPR011050	9.91	3.67E−114	11.19	0.16	6.13E−03	6.47	4.57	0.00E+00
	Fjoh_0549	Hypothetical Lipoprotein	PhoBR operon	OM	pf07705	7.59	1.67E−71	5.06	1.72	3.57E−03	7.05	3.64	0.00E+00
	Fjoh_1368	Tbdt	Phytase	OM	pf07715; pf13715	9.22	1.84E−30	8.94	0.65	0.00E+00	3.72	2.41	5.29E-03
	Fjoh_3250	PusD2	PUSII	OM	pf07980, pf14322	8.94	1.49E−79	5.47	3.11	0.00E+00	8.85	4.43	0.00E+00
	Fjoh_3251	PusC2	PUSII	OM	pf00593, pf07715, pf13715, pf16344	9.03	7.60E−87	4.53	0.24	0.00E+00	ND	ND	ND
	Fjoh_3841	Tbdt	Orphan	OM	pf13620	7.69	2.64E−81	8.44	0.11	0.00E+00	8.34	4.60	0.00E+00
	Fjoh_3856	Hypothetical Lipoprotein - FI	FI cluster/ PUSIII	OM	NA, lipoprotein	14.47	5.36E−31	6.33	0.01	1.48E−3	5.39	3.07	0.00E+00
	Fjoh_3857	Tbdt	FI cluster/ PUSIII	OM	pf07715, pf13715	12.50	7.00E−100	5.18	4.08	1.41E-03	3.73	2.48	3.38E-03
	Fjoh_4168	PusD1	PUSI	OM	pf07980, pf14322	10.17	7.54E−94	6.23	1.84	0.00E+00	8.55	4.99	0.00E+00

Table 1 (continued)

Predicted function	Locus_tag	Gene symbol/description	Genomic context	Predicted localisation	Pfam accession	Transcriptome		Whole cell proteome			Exoproteome		
						FC	FDR (*p*-adj)	DE	−Log *p* value	*q* value	DE	−Log *p* value	*q* value
	Fjoh_4169	PusC1	PUSI	OM	pf00593, pf07715, pf13715	10.66	8.26E−101	3.17	0.01	1.33E−01	ND	ND	ND
	Fjoh_4889	Hypothetical Lipoprotein - FIII	Orphan	OM	NA	11.56	4.44E−118	9.13	1.42	0.00E+00	9.56	4.13	0.00E+00
Lipid remodelling	Fjoh_0541	Glycosyl transferase	PhoBR operon	Cytoplasm	pf04101	5.42	2.94E−49	7.04	5.08	0.00E+00	ND	ND	ND
	Fjoh_4577	PlcP	Orphan	Cytoplasm	pf12850	8.59	2.12E−77	ND	ND	ND	ND	ND	ND
Accessory	Fjoh_0543	PhoR	PhoBR operon	Cytoplasm	pf00512, pf02518	0.76	0.056401	ND	ND	ND	ND	ND	ND
	Fjoh_0544	PhoB	PhoBR operon	Cytoplasm	pf00072, pf00486	3.06	9.42E−20	3.23	4.49	0.00E+00	ND	ND	ND
	Fjoh_2477	MauG_like	PhoX	Cytoplasm	pf03150	9.58	3.54E−95	7.33	0.64	8.36E−03	ND	ND	ND
	Fjoh_3252	FecR_like regulator	PUSII	Cytoplasm	pf04773, pf16344	9.88	7.70E−61	ND	ND	ND	ND	ND	ND
	Fjoh_3853	C-terminal peptidase S41	FI cluster/ PUSIII	Periplasm	pf03572	9.08	9.92E−74	ND	ND	ND	4.44	2.79	0.00E+00
	Fjoh_3854	T9SS-secretion domain, lipoprotein	FI cluster/ PUSIII	OM	NA,	9.64	1.90E−66	3.30	0.65	6.04E−03	5.75	4.29	0.00E+00
	Fjoh_3855	T9SS-secretion domain,	FI cluster/ PUSIII	OM	NA	10.04	7.54E−94	ND	ND	ND	8.61	3.90	0.00E+00
	Fjoh_3858	FecR_like regulator	FI cluster/ PUSIII	Cytoplasm	pf04773, pf16344	9.79	9.56E−90	1.53	1.75	7.02E−02	ND	ND	ND
	Fjoh_3859	Sigma factor - regulator	FI cluster/ PUSIII	Cytoplasm	pf04542, pf08281	7.94	2.29E−70	ND	ND	ND	ND	ND	ND
	Fjoh_4171	ECF Sigma factor	PUSI	Cytoplasm	pf04542, pf08281	8.88	1.19E−70	0.05	3.82	9.50E−01	ND	ND	ND

Predicted function is based on in silico sequence analysis. Where applicable, supporting pfam or INTERPRO domains are provided. Genomic context is given in terms of the genetic neighbourhood each gene was identified. Predicted localisation is based on the presence of a leader sequence and/or lipobox. 'Orphan' denotes genes not associated with any other predicted P-acquisition genes.

OM outer membrane, *IM* inner membrane, *PUS* phosphate utilisation system, *DE* differential expression, *ND* not detected, *FC* fold change.

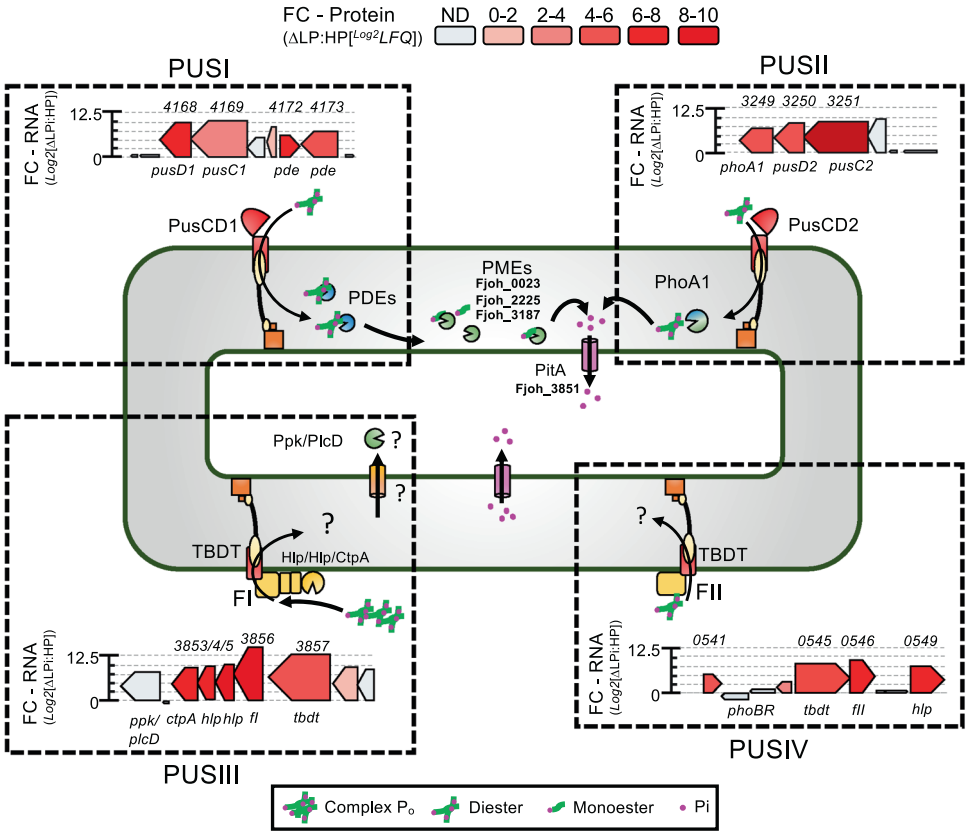


Fig. 3 Expression of putative Pi-sensitive polysaccharide utilisation loci in *F. johnsoniae*. Intergration of transcriptomic (*y* axis on genetic neighbourhood maps) and proteomic data (colour gradient on genetic neighbourhood maps) with in silico function prediction/cellular location used to generate a proposed model for the newly identified Phosphate utilisation systems in *Flavobacterium johnsoniae* DSM2064. Locus tags for selected ORFs are given, omitting 'fjoh_'. Results shown are the mean of triplicate cultures in each growth condition. Abbreviations: FC fold change, LFQ label-free quantification, PusCD phosphate utilisation system transporter complex – SusC forms the transmembrane pore and SusD is the cell-surface ligand-binding protein, FI/FII, hypothetical lipoproteins with homology to polygalacturanase, Ppk/PlcD polyphosphate kinase/phosphatidylinositol phospholipase-like, PME phosphomonoesterase, PDE phosphodiesterase, PitA high velocity inner membrane Pi symporter; *ctpA* – S41 peptidase.

appear to be related to each other (Fig. S7) and share modest sequence homology (31–36% identity, 50–54% coverage) with a low-Pi-inducible lipoprotein of *Caulobacter crescentus* that has an accessory role in producing phosphatase activity [56]. Furthermore, structural homology modelling revealed these proteins may have a role in binding or hydrolysing plant glycans or other complex plant-cell wall molecules, similar to that of the *Aspergillus niger* polygalacturanase [57] (Fig. S7). Combining all these data, it is clear that DSM2064 invests a large amount of cellular resources into expressing several PUS gene clusters that we predict target P_o molecules.

The unique *Flavobacterium* PHO regulon is largely conserved across plant-associated isolates

To better determine how important these unique Pi-acquisition loci are for root-associated *Flavobacterium*, we performed whole-genome sequencing and exoproteomics on the OSR rhizosphere strains previously tested for their phosphatase activity (Fig. 1B, C). General genomic characteristics for the newly isolated *Flavobacterium* strains are presented in Table S6. A comprehensive analysis of all the XP datasets are presented in Tables S7–S13. PafA was detected in the Pi-replete and Pi-deplete XPs of all seven (2 non-OSR and 5 OSR) *Flavobacterium* strains (Blue diamonds, Fig. 4A). Conservation in their response to Pi-depletion was also observed (Fig. 4A). This included the abundant expression of proteins associated with the PUSI/II/III/IV gene clusters, the distinct PhoX and other PME and PDEs (Fig. 4A). However, all *Flavobacterium* strains displayed microheterogeneity in their response to Pi-depletion. Most notably, this included the expression of phylogenetically distinct PusCD complexes, located within distinct PUS gene clusters (described in detail below) as well as the expression of strain-specific gene clusters.

This fine-scale heterogeneity in their response to Pi-depletion was also apparent in their genomic make up (Fig. 4B). In comparison to ORFs corresponding to core

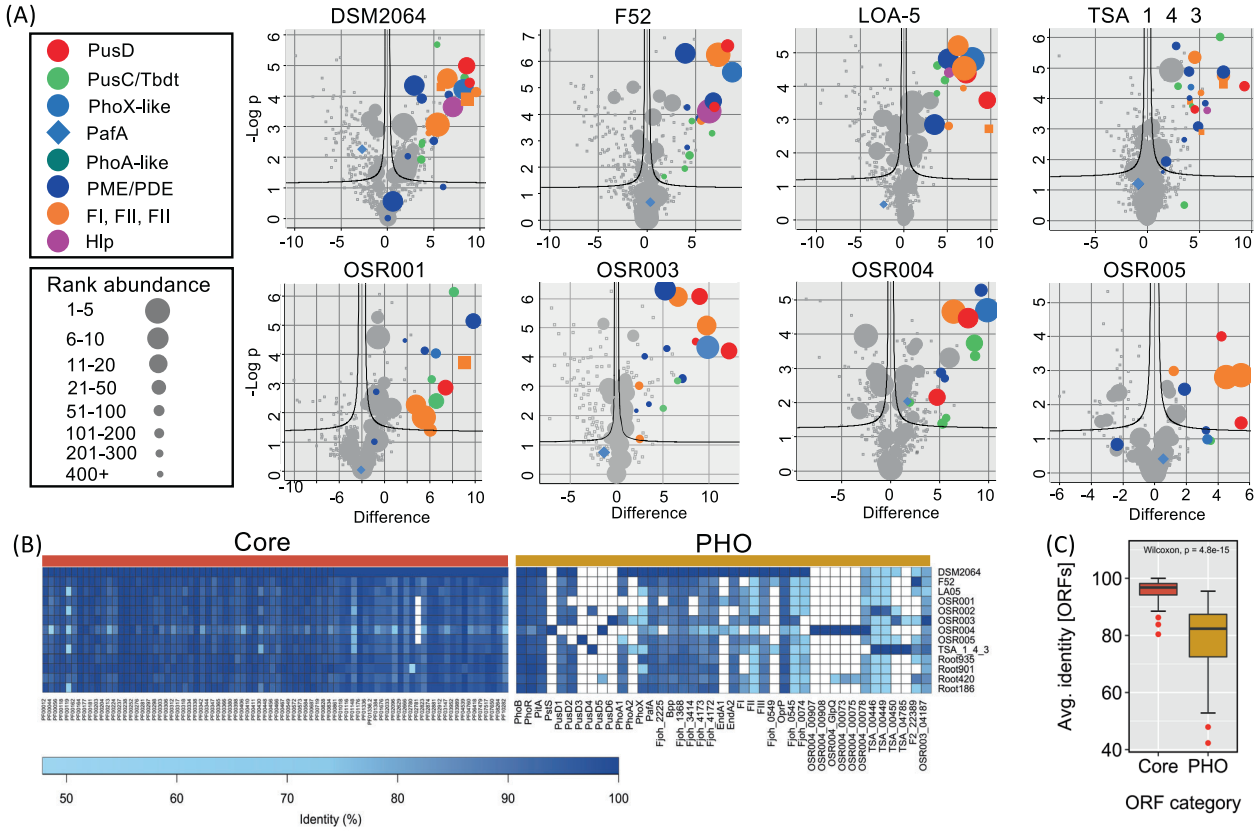


Fig. 4 Identification of P-acquisition proteins and their corresponding loci in several plant-associated *Flavobacterium* strains. **A** Volcano plots of exoproteome data illustrating both the difference in the $\log_2$ label-free quantitative values (LFQ) intensity between Pi-replete and Pi-deplete treatments and the log probability of the observed difference. Dashed lines represent the threshold for significant *P* values adjusted using FDR = 0.03 (*Q* values). The rank abundance of each protein is denoted by dot size. The top 30 proteins detected in the Pi-deplete exoproteomes are highlighted. In addition, the rank abundance of predicted P-acquisition proteins is also given. Data points represent the means of triplicate cultures sampled from each growth condition. Abbreviations: PusD phosphate utilisation system-like outer membrane binding domain, PhoX alkaline phosphatase PhoX-like, PusC/Tbdt including TonB-dependent transporter proteins, APases putative phosphatases, FI FII, FIII, hypothetical exoproteins that share weak homology with pectate lyases/polygalacturonases, PafA Pi-irrepressible phosphatase, Hlp hypothetical lipoprotein located in phoBR operon. **B** Occurrence and similarity (% identity) of ORFs related to core/housekeeping processes (maroon) and P-acquisition (gold) amongst the *Flavobacterium* strains. **C** Box–whisker plots illustrating the mean identity of core and P-acquisition (PHO) ORFs. Red dots denote outliers.

metabolic and housekeeping genes (~80 single-copy representatives), ORFs corresponding to low-Pi-inducible genes had fewer occurrences and greater sequence divergence across the strains (Fig. 4B). The average sequence identity of PHO regulon ORFs (77.9%) was significantly lower (Wilcox test, $P < 0.001$) than the average sequence identity of the core ORFs (95.9%) (Fig. 4C).

Highly expressed PHO regulon genes occur more frequently in plant-associated *Bacteroidetes* genomes

We further investigated the occurrence and diversity of these P-acquisition genes across the *Flavobacteriaceae* family. One hundred two genomes (predominantly from the genus *Flavobacterium*) previously deposited in the IMG/JGI database, representing strains isolated from different environmental niches (using 'Ecosystem Type'; 'Study name'; 'Habitat') were selected and analysed (Table S14). This included plant-associated, soil (terrestrial), marine (aquatic, sediment), freshwater (aquatic, sediment), fish (marine host-associated) and various 'other' microhabitats (wastewater treatment, etc.). The number of newly identified P-acquisition genes occurring in the plant-associated genomes was significantly greater (Wilcox test, $P = 0.00079$) than the number occurring in non-plant-associated genomes (Fig. 5A). This equated to greater similarity in the P-scavenging repertoire between plant-associated strains compared to non-plant-associated strains (Fig. 5B). Almost all plant-associated strains were found within a single subclade (subclade 1, Fig. 5C) demonstrating phylogeny contributed towards the occurrence of the newly identified Pi-acquisition genes within their genomes (Fig. 5C). However, niche adaptation was clearly a strong underlying

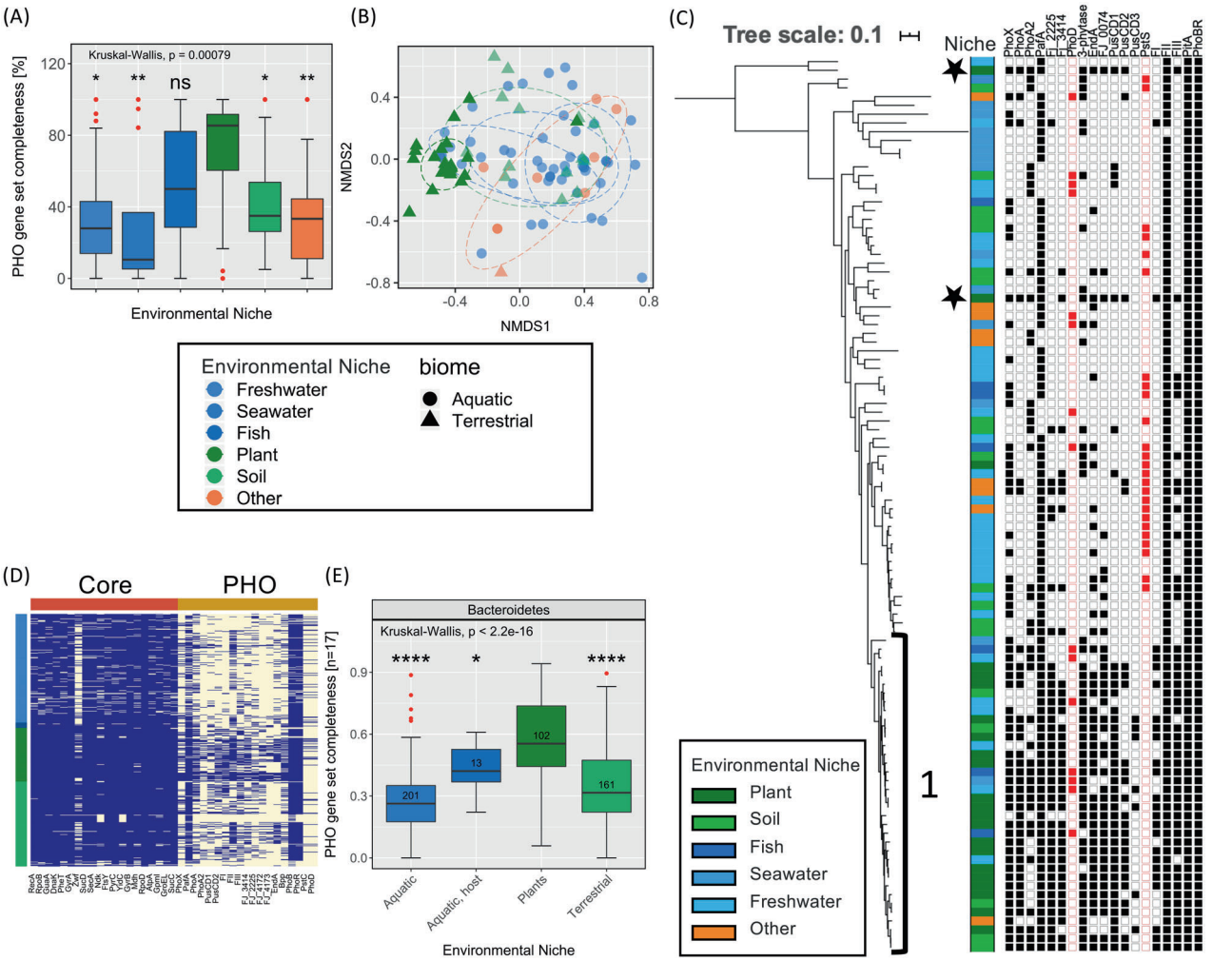


Fig. 5 Occurrence of P-acquisition (PHO) genes in the genomes of plant-associated and non-plant-associated *Bacteroidetes* strains. **A** Occurrence of all PHO genes among the 102 *Flavobacterium* strains. Red dots denote outliers in the data and asterisks denote significance levels (Kruskal Wallis; *<0.05, **<0.01, ***<0.001) for differences of occurrence in each group relative to the Plant-associated group. **B** Non-metric multidimensional scaling (Bray-Curtis distances) of P-acquisition gene content among the 102 *Flavobacterium* strains. Ellipses represent euclidean distances from the centre of groups (level = 0.7). **C** The multi-loci maximum-likelihood consensus tree was inferred from a simultaneous comparison of 10 housekeeping and core genes present in 102 *Flavobacterium* isolates. Tree topology and branch lengths were calculated by maximum likelihood using the VT + F + R5 model of evolution for amino acid sequences based on 1022 sites in lQ-TREE software [48]. The presence (filled symbol) or absence (hollow symbol) of selected P-acquisition genes is displayed. PstS and PhoD are highlighted in red. Subclade 1 is denoted by the numeric. Stars represent phylogenetically distinct plant-associated strains. **D** Occurrence of core and PHO genes among 468 *Bacteroidetes* strains isolated from the five environmental niches. **E** Occurrence of all PHO genes, excluding PitA and PhoBR, among the 468 *Bacteroidetes* strains. Red dots denote outliers in the data and asterisks denote significance levels (Kruskal Wallis; *<0.05, **<0.01, ***<0.001) for differences of occurrence in each group relative to the Plant-associated group.

driver as two plant-associated strains, *Flavobacterium* sp. B17 and *Flavobacterium chilense* DSM24724, that were distantly related to subclade 1, still possessed a high number of P-acquisition genes (black stars, Fig. 5C). Interestingly, ORFs corresponding to the constitutive APase PafA, were found in 90/102 strains suggesting that constitutive phosphatase expression is not confined to plant-associated strains only. The occurrence of the well-known APase *phoD* (15/102) was mostly associated with aquatic strains (13/15). *pstS* (26/102) was absent from subclade 1 and only found in one plant-associated strain (OSR_004) and this was coincident with a lack of *pusCD1*, and to a lesser extent *pusCD2* (Table S14).

Given the niche-associated diversity of PHO regulon genes within the *Flavobacterium* genus, we expanded our comparative genomics analysis across the entire *Bacteroidetes* phylum, still focusing on isolates retrieved from either plant, soil (terrestrial), freshwater or seawater (aquatic), or host-associated strains of aquatic organisms, such as algae or fish (aquatic, host) ($n = 468$, Table S15). Anaerobic

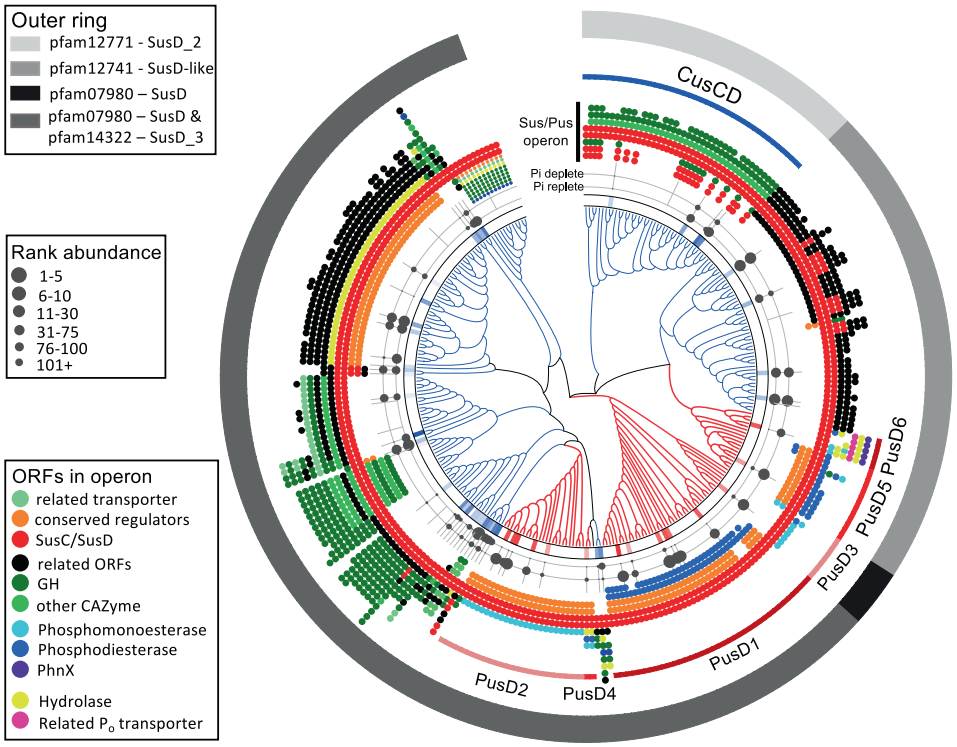


Fig. 6 Phylogenetic, expression and genomic relationship analyses of SusD-like homologues in *Flavobacterium*. The maximum-likelihood consensus tree was inferred from a simultaneous comparison of 310 protein sequences related to the SusD lipoprotein. Tree topology and branch lengths were calculated by maximum likelihood using the VT + F + R5 model of evolution for amino acid sequences based on 1022 sites in lQ-TREE software [48]. Clades containing low-Pi responsive PusD proteins are coloured red whilst clades containing Pi-insensitive SusD-like proteins are coloured blue. The first ring expresses the difference in abundance (Pi-deplete: Pi-replete) for all detected SusD-like domains. The second (Pi-replete) and third (Pi-deplete) rings represent the rank abundance against all proteins for detected SusD-like proteins in the various exoproteomes. The genetic neighbourhoods for each SusC/D-like complex are illustrated by the coloured dot plots. Numericals 1–6 denotes the six PusD forms identified in this study. CusCD denotes the experimentally characterised chitin utilisation system [25]. The outer ring defines the pfam domains associated with each SusD form (predicted by IMG/JGI).

isolates inhabiting animal and human gut microbiomes were not scrutinised. The greater heterogeneity in the occurrence of PHO genes versus the core genes was maintained at the phylum level (Fig. 5D), which again resulted in a significantly greater ($P < 2.2^{e-16}$) number of PHO genes occurring in plant-associated genomes compared to all non-plant-associated genomes (Fig. 5E). Finally, whilst the average occurrence of PHO regulon genes across the phylum was 35.2 ± 25%, the occurrence of genes encoding the constitutive APase PafA was much greater (78.7%). This suggests, the ubiquitous PafA- dependent constitutive phosphatase expression observed in plant-associated strains here, is a common trait in this phylum.

Plant-associated *Flavobacterium* have evolved a distinct set of SusCD-like complexes to facilitate Pi-scavenging

In total, we detected the expression of six distinct PusCD complexes across the eight *Flavobacterium* strains (Fig. 4A). All strains also expressed several Pi-insensitive SusCD-like complexes, in agreement with their predicted specialisation in plant glycan degradation [18] (Table S3). To compare the phylogeny of these Pi-sensitive and Pi-insensitive SusCD-like complexes, the SusD-like domains from all complexes detected in the *Flavobacterium* XPs were aligned (Fig. 6). In addition, the corresponding PusD/SusD-like homologues from the 102 *Flavobacterium* genomes were also included. The various PusD and SusD-like forms made up three distinct subclades evidenced by their different pfam domains (Fig. 6, outer ring). PusD was polyphyletic among various SusD-like subclades indicating multiple evolutionary events occurred during the generation of the various PusCD complexes. However, the predicted gene clusters containing the six PusCD forms were clearly distinct from the Pi-insensitive SusCD-like gene clusters. SusCD-like complexes were co-localised predominantly with carbohydrate processing enzymes, whilst all six PusCD complexes, as already described for PUSI and PUSII in DSM2064, were co-localised with PMEs, PDEs,

as well as phosphonate-degrading enzymes and inner membrane P_o transporters (Fig. 6). All PusCD-containing PUS gene clusters differed, suggesting a distinct role in P_o-acquisition for each. PUSI, containing PusCD1, and the two predicted periplasmic PDEs, was the most common PUS system found in *Flavobacterium* revealing these loci likely play an important role in scavenging ubiquitous phosphodiesters (most likely lipid headgroups and nucleic acids) found in the soil/rhizosphere.

Discussion

Competition for P in the rhizosphere is incredibly intense [30, 34]. Hence, understanding how *Bacteroidetes* competes for this scarce nutrient provides valuable insights into the underlying molecular mechanisms responsible for this phylum being a dominant member of the plant microbiome [1–3, 8]. This study provides strong evidence that plant-associated *Bacteroidetes* possess an incredibly high potential for transforming P_o to bioavailable Pi. This includes the apparent identification of novel molecular mechanisms for P uptake linked to TonB-dependent transport [58] that differ from previously known bacterial P transport mechanisms [13].

The constitutive PME expression displayed by all *Bacteroidetes* OSR isolates irrespective of Pi concentrations (Fig. 1) is a unique and potentially important trait in the context of improving plant phosphorus uptake at the root–soil interface [33, 39, 59]. Whilst inducible PME activity is high when *Pseudomonas* spp. are grown on succinate (growth medium pH 7–8) [15, 50], a lower pH range (pH 5–7) in the growth medium used here drastically inhibited their PME activity. Thus, *Flavobacterium* spp. may outperform traditional alkaline phosphatase-harbouring rhizobacteria, including *Pseudomonas* spp. [13–16], with respect to P-mobilisation, in an agricultural setting where rhizosphere pH is typically pH 5.5–7.5 [32]. This hypothesis warrants further investigation through protein purification and *in planta* experiments.

Plant-associated *Flavobacterium* expressed a far greater number of PMEs and PDEs than other well-known rhizobacteria [13–16] and not all were silenced by high concentrations of Pi (Table 1). Importantly, the majority of these P-acquisition loci appear to be niche-adapted traits that would allow plant-associated *Bacteroidetes* to overcome competition for P in the rhizosphere, where labile Pi is depleted and replaced by immobilised P_o [32, 37]. The *Flavobacterium* PhoX is a prime example of this niche adaptation to the plant microbiome. This phosphatase was one of the major proteins secreted by all plant-associated *Flavobacterium* strains in response to Pi-depletion (Fig. 4) but is only a common genetic feature in the genomes of plant-associated *Bacteroidetes* strains. PhoX is a lipoprotein [60], hence we did not detect this in our soluble fractions (Fig. S6). However, given its predicted localisation in the OM and structural similarity to other characterised PhoX homologues that are responsible for PME activity in other rhizobacteria [16, 47, 60, 61], we predict this enzyme is responsible for most of the elevated inducible PME activity observed in our *Flavobacterium* strains. *Flavobacterium* PhoX is both phylogenetically and structurally distinct from previously characterised PhoX and can be further subdivided into three subclades (Fig. S4). Each subclade is predicted to possess a different active site architecture and surface potential compared to the characterised PhoX [61] (Fig. S3), which may explain why *Flavobacterium* inducible activity is more resilient to lower pH conditions compared to *Pseudomonas* [15, 50].

In contrast, the widespread occurrence of PafA in *Bacteroidetes* isolated from all environmental niches reveals that the unusual constitutive phosphatase expression is almost certainly a common phenotypic trait in this phylum. PafA, whose expression was constant in all eight strains, is a Pi-irrepressible phosphatase that is a unique member of the alkaline phosphatase superfamily whose pH optimum is neutral to slightly acidic and also possesses diesterase activity [53, 55]. Using DSM2064 as the model, PafA was the major protein detected in one of the phosphatase-active protein fractions (Fig. S6), strongly suggesting this enzyme contributes towards the unique constitutive and inducible PME activity in our *Flavobacterium* isolates.

Why *Bacteroidetes* have such a huge diversity of TBDTs and SusCD-like complexes is a recurrent question [5, 22, 58]. Here, we present evidence that suggests novel OM transporters related to those traditionally associated with siderophore and oligosaccharide transport [20–25, 55, 62] may be involved in transporting P_o species. All plant-associated *Bacteroidetes* lack common organic phosphorus transporters and the high affinity PstSABC transporter for Pi [13, 15], but instead possess novel low-Pi-inducible PusCD complexes and orphan TBDTs, indicating *Bacteroidetes* favour an alternative mechanism for competitively acquiring P. Given, the pore size of transmembrane TBDTs and SusC-like domains and general role as polysaccharide transporters [19, 58], their co-localisation with ORFs encoding co-expressed predicted periplasmic PME and PDEs (Fig. 6) strongly suggests these transporters have evolved to target P_o. The co-localisation of a low-Pi-inducible TBDTs with 3-phytase (Fjoh_1368), a well-known P_o-degrading enzyme [63] supports this hypothesis. Further support is provided by the abundant expression of the uncharacterised FI and FII, located in PUSIII and PUSIV, that both share sequence homology with an uncharacterised PHO regulon lipoprotein in *C. crescentus* [56] as well as structural homology with plant-cell wall

degradation enzymes [57]. Thus, when facing Pi-depletion, *Flavobacterium* may target the large reservoir of P_o associated with plant root cell walls [64] and the consequence of this unique mechanism warrants further investigation.

Marine and Gut *Bacteroidetes* predominantly acquire carbohydrate substrates through a 'selfish uptake' mechanism using SusCD-like harbouring PULs [24, 65]. This selfish mechanism for degrading more recalcitrant organic carbon does not liberate low molecular weight carbon for scavengers in the way the 'sharing' mechanisms (extracellular enzyme secretion) can [65]. We have recently demonstrated 'sharing' of more recalcitrant P_o substrates occurs in *Pseudomonas* spp. through the production of periplasmic and extracellular phosphatases [50]. Based on the data presented here, *Flavobacterium* may be capable of both a sharing and selfish mechanism of P_o-utilisation that is partially dependent on external Pi concentrations, which may have divergent outcomes for rhizosphere P cycling.

Conclusions

P_o often represents a major fraction of the total P content in both agricultural and non-agricultural soils [59, 66]. This is set to rise with the increased use of organic P fertilisers primarily from manure and other waste materials [31, 59]. Our finding that abundant resident microorganisms of many cropping systems potentially possess novel mechanisms for liberating Pi from P_o may influence sustainable approaches to crop production and combat the global phosphorus crisis [24, 29, 34, 59]. Whether these novel and unique molecular mechanisms discovered in *Bacteroidetes* result in an enhanced or broader capability to degrade rhizosphere-associated P_o is still an open question, but one which has important consequences for sustainable agriculture. Not only for enhancing phosphorus acquisition in crops, but also developing 'designer rhizospheres' through selection, using plant genotyping and environmental conditioning, to enhance plant functionality, such as plant immunity or bioremediation [67]. This study paves the way for further exciting research on the P physiology of *Flavobacterium* strains and their potential exploitation as microbial tools in sustainable agriculture.

Data availability

The proteomics datasets supporting the conclusions of this article are available in the ProteomeXchange Consortium via the PRoteomics IDEntifications (PRIDE) partner repository with the dataset identifier PXD014380 and 10.6019/PXD014380. The transcriptomics dataset supporting the conclusions of this article are available in the NCBI short read archive (SRA) repository with the dataset identifier Bioproject accession PRJNA635152.

Acknowledgements We thank the Warwick Proteomics Research Facility, namely Dr. Cleidiane Zampronio and Dr. Juan Hernandez Fernaud for their assistance in generating and processing the mass-spectrometry data. We also thank Dr. Eddie Cytryn for kindly sending us various strains and plasmids used in this study. This study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under the project codes BB/L026074/1 and BB/T009152/1 linked to The Soil and Rhizosphere Interactions for Sustainable Agri-ecosystems (SARISA) programme and a Discovery Fellowship (IL), respectively.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature. 2012;488:91–5.
2. Bulgarelli D, Garrido-Oter R, Münch Philipp C, Weiman A, Dröge J, Pan Y, et al. Structure and function of the bacterial root microbiota in wild and domesticated barley. Cell Host Microbe. 2015;17:392–403.
3. Niu B, Paulson JN, Zheng X, Kolter R. Simplified and representative bacterial community of maize roots. PNAS. 2017;114:e2450–9.
4. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608–11.
5. Thomas F, Hehemann J-H, Rebuffet E, Czjzek M, Michel G. Environmental and gut *Bacteroidetes*: the food connection. Front Microbiol. 2011;2:93.
6. Yin C, Hulbert SH, Schroeder KL, Mavrodi O, Mavrodi D, Dhingra A, et al. Role of bacterial communities in the natural suppression of *Rhizoctonia solani* bare patch disease of wheat (*Triticum aestivum* L.). Appl Environ Microbiol. 2013;79:7428–38.
7. Gardner T, Acosta-Martinez V, Senwo Z, Dowd SE. Soil rhizosphere microbial communities and enzyme activities under organic farming in Alabama. Diversity. 2011;3:308–28.
8. Kolton M, Erlacher A, Berg G, Cytryn E. The *Flavobacterium* genus in the plant holobiont: ecological, physiological, and applicative insights. In: Castro-Sowinski S, editor. Microbial models:

from environmental to industrial sustainability. Singapore: Springer Singapore; 2016. p. 189–207.
9. Maimaiti J, Zhang Y, Yang J, Cen Y-P, Layzell DB, Peoples M, et al. Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. Environ Microbiol. 2007;9:435–44.
10. Kwak M-J, Kong HG, Choi K, Kwon S-K, Song JY, Lee J, et al. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat Biotechnol. 2018;36:1117.
11. Nishioka T, Marian M, Kobayashi I, Kobayashi Y, Yamamoto K, Tamaki H, et al. Microbial basis of *Fusarium wilt* suppression by Allium cultivation. Sci Rep. 2019;9:1715.
12. Johansen JE, Binnerup SJ. Contribution of Cytophaga-like bacteria to the potential turnover of carbon, nitrogen, and phosphorus by bacteria in the rhizosphere of barley (*Hordeum vulgare* L.). Microb Ecol. 2002;43:298–306.
13. Santos-Beneit F. The Pho regulon: a huge regulatory network in bacteria. Front Microbiol. 2015;6:402.
14. Antelmann H, Scharf C, Hecker M. Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis. J Bacteriol. 2000;182:4478–90.
15. Lidbury IDEA, Murphy ARJ, Scanlan DJ, Bending GD, Jones AME, Moore JD. et al. Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. Environ Microbiol. 2016;18:3535–49.
16. Zaheer R, Morton R, Proudfoot M, Yakunin A, Finan TM. Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. Environ Microbiol. 2009;11:1572–87.
17. Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E, et al. Whole genome analysis of the marine *Bacteroidetes* '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. Environ Microbiol. 2006;8:2201–13.
18. Kolton M, Sela N, Elad Y, Cytryn E. Comparative genomic analysis indicates that niche adaptation of terrestrial flavobacteria is strongly linked to plant glycan metabolism. PLoS ONE. 2013;8: e76704.
19. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. J Biol Chem. 2009;284:24673–7.
20. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature. 2017; 541:407–11.
21. Foley MH, Cockburn DW, Koropatkin NM. The Sus operon: a model system for starch uptake by the human gut *Bacteroidetes*. Cell Mol Life Sci. 2016;73:2603–1.
22. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2019;13:76–91.
23. Zhu Y, Thomas F, Larocque R, Li N, Duffieux D, Cladière L, et al. Genetic analyses unravel the crucial role of a horizontally acquired alginate lyase for brown algal biomass degradation by *Zobellia galactanivorans*. Environ Microbiol. 2017;19:2164–81.
24. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. Nature 2015;517:165–9.
25. Larsbrink J, Zhu Y, Kharade SS, Kwiatkowski KJ, Eijsink VGH, Koropatkin NM, et al. A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. Biotechnol Biofuels. 2016;9:2.
26. Cordell D, Drangert J-O, White S. The story of phosphorus: global food security and food for thought. Glob Environ Change. 2009;19:292–305.
27. Tyrrell T. The relative influences of nitrogen and phosphorus on oceanic primary production. Nature 1999;400:525–31.
28. Filippelli GM. The global phosphorus cycle: past, present, and future. Elements. 2008;4:89–95.
29. Yang N, Zavišić A, Pena R, Polle A. Phenology, photosynthesis, and phosphorus in European beech (*Fagus sylvatica* L.) in two forest soils with contrasting P contents. J Plant Nutr Soil Sci. 2016;179:151–8.
30. Goll DS, Brovkin V, Parida BR, Reick CH, Kattge J, Reich PB, et al. Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. Biogeosciences. 2012;9:3547–69.
31. Cordell D, White S. Tracking phosphorus security: indicators of phosphorus vulnerability in the global food system. Food Security. 2015;7:337–50.
32. White PJ, Hammond JP. Phosphorus nutrition of terrestrial plants. In: White PJ, Hammond JP, editors. The ecophysiology of plant-phosphorus interactions. Dordrecht: Springer Netherlands; 2008. p. 51–81.
33. Sharma SB, Sayyed RZ, Trivedi MH, Gobi TA. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. SpringerPlus. 2013;587.
34. Turner BL, Papházy MJ, Haygarth PM, McKelvie ID. Inositol phosphates in the environment. Philos Trans R Soc Lond B Biol Sci. 2002;357:449–69.
35. Schachtman DP, Reid RJ, Ayling SM. Phosphorus uptake by plants: from soil to cell. Plant Physiol. 1998;116:447–53.
36. Zou X, Binkley D, Doxtader KG. A new method for estimating gross phosphorus mineralization and immobilization rates in soils. Plant Soil. 1992;147:243–50.
37. Zoysa AKN, Loganathan P, Hedley MJ. A technique for studying rhizosphere processes in tree crops: soil phosphorus depletion around camellia (*Camellia japonica* L.) roots. Plant Soil. 1997; 190:253–65.
38. Shen J, Yuan L, Zhang J, Li H, Bai Z, Chen X, et al. Phosphorus dynamics: from soil to plant. Plant Physiol. 2011;156:997–1005.
39. Stutter MI, Shand CA, George TS, Blackwell MSA, Bol R, MacKay RL, et al. Recovering phosphorus from soil: a root solution? Environ Sci Technol. 2012;46:1977–8.
40. Wanner BL. Phosphorus assimilation and its control of gene expression in *Escherichia coli*. In: Hauska G, Thauer R, editors. The molecular basis of bacterial metabolism. Colloquium der Gesellschaft für Biologische Chemie 5.–7.1990. Mosbach/Baden: Springer Berlin Heidelberg; 1990. p. 152–63.
41. Gandhi NU, Chandra SB A comparative analysis of three classes of bacterial non-specific acid phosphatases and archaeal phosphoesterases: evolutionary perspective. Acta Inform Med. 2012;20:167–73.
42. IDEA Lidbury, Fraser T, Murphy ARJ, Scanlan DJ, Bending GD, AME Jones, et al. The 'known' genetic potential for microbial communities to degrade organic phosphorus is reduced in low-pH soils. Microbiol Open. 2017;6:e00474.
43. Kolton M, Green SJ, Harel YM, Sela N, Elad Y, Cytryn E. Draft genome sequence of *Flavobacterium* sp. strain F52, isolated from the rhizosphere of Bell Pepper (*Capsicum annuum* L. cv. Maccabi). J Bacteriol. 2012;194:5462–3.
44. McBride MJ, Kempf MJ. Development of techniques for the genetic manipulation of the gliding bacterium *Cytophaga johnsonae*. J Bacteriol. 1996;178:583–90.
45. Nishioka T, Elsharkawy MM, Suga H, Kageyama K, Hyakumachi M, Shimizu M. Development of culture medium for the isolation of *Flavobacterium* and *Chryseobacterium* from rhizosphere soil. Microbes Environ. 2016;31:104–10.
46. Gil R, Silva FJ, Peretó J, Moya A. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev. 2004;68:518–37.
47. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9.

48. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
49. Roca A, Pizarro-Tobías P, Udaondo Z, Fernández M, Matilla MA, Molina-Henares MA, et al. Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. Environ Microbiol. 2013;15:780–94.
50. IDEA Lidbury, Murphy ARJ, Fraser TD, et al. Identification of extracellular glycerophosphodiesterases in *Pseudomonas* and their role in soil organic phosphorus remineralisation. Sci Rep. 2017;7:2179. JD, et al.
51. Brickman E, Beckwith J. Analysis of the regulation of *Escherichia coli* alkaline phosphatase synthesis using deletions and φ80 transducing phages. J Mol Biol. 1975;96:307–15.
52. Skouri-Panet F, Benzerara K, Cosmidis J, Férard C, Caumes G, De Luca G, et al. In vitro and in silico evidence of phosphatase diversity in the biomineralizing bacterium *Ramlibacter tataouinensis*. Front Microbiol. 2018;8:2592.
53. Berlutti F, Passariello C, Selan L, Thaller MC, Rossolini GM. The *Chryseobacterium meningosepticum* PafA enzyme: prototype of a new enzyme family of prokaryotic phosphate-irrepressible alkaline phosphatases? Microbiology. 2001;147:2831–9.
54. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–5.
55. Sunden F, AlSadhan I, Lyubimov AY, Ressl S, Wiersma-Koch H, Borland J, et al. Mechanistic and evolutionary insights from comparative enzymology of phosphomonoesterases and phosphodiesterases across the alkaline phosphatase superfamily. J Am Chem Soc. 2016;138:14273–87.
56. Le Blastier S, Hamels A, Cabeen M, Schille L, Tilquin F, Dieu M, et al. Phosphate starvation triggers production and secretion of an extracellular lipoprotein in *Caulobacter crescentus*. PLoS ONE. 2010;5:e14198.
57. Maldonado MC, Strasser, de Saad AM. Production of pectinesterase and polygalacturonase by *Aspergillus niger* in submerged and solid state systems. J Ind Microbiol Biotechnol. 1998;20:34–8.
58. Bolam DN, van den Berg B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. Curr Opin Struct Biol. 2018;51:35–43.
59. Haygarth PM, Harrison AF, Turner BL. On the history and future of soil organic phosphorus research: a critique across three generations. Eur J Soil Sci. 2018;69:1.
60. Putker F, Tommassen-van Boxtel R, Stork m, Rodríguez-Herva JJ, Koster M, Tommassen J. The type II secretion system (Xcp) of *Pseudomonas putida* is active and involved in the secretion of phosphatases. Environ Microbiol. 2013;15:2658–71.
61. Yong SC, Roversi P, Lillington J, Rodriguez F, Krehenbrink M, Zeldin OB, et al. A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. Science. 2014;345:1170–3.
62. Joglekar P, Sonnenburg ED, Higginbottom SK, Earle KA, Morland C, Shapiro-Ward S, et al. Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroides thetaiotaomicron*. mSphere. 2018;3:e00185–18.
63. Lim BL, Yeung P, Cheng C, Hill JE. Distribution and diversity of phytate-mineralizing bacteria. ISME J. 2007;1:321–30.
64. Zhu XF, Wang ZW, Wan JX, Sun Y, Wu YR, Li GX, et al. Pectin enhances rice (*Oryza sativa*) root phosphorus remobilization. J Exp Bot. 2014;66:1017–24.
65. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J. 2017;11:1640–50.
66. Turner BL, Haygarth PM. Phosphatase activity in temperate pasture soils: Potential regulation of labile organic phosphorus turnover by phosphodiesterase activity. Sci Total Environ. 2005;344:27–36.
67. Sergaki C, Lagunas B, Lidbury I, Gifford ML, Schäfer P. Challenges and approaches in microbiome research: From fundamental to applied. Front Plant Sci. 2018;9:1205.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-24646-z OPEN

Transporter characterisation reveals aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle

Andrew R. J. Murphy[1], David J. Scanlan[1], Yin Chen[1], Nathan B. P. Adams[2,3], William A. Cadman[2], Andrew Bottrill[1], Gary Bending[1], John P. Hammond[4], Andrew Hitchcock[2], Elizabeth M. H. Wellington[1] & Ian D. E. A. Lidbury[5]✉

The planktonic synthesis of reduced organophosphorus molecules, such as alkylphosphonates and aminophosphonates, represents one half of a vast global oceanic phosphorus redox cycle. Whilst alkylphosphonates tend to accumulate in recalcitrant dissolved organic matter, aminophosphonates do not. Here, we identify three bacterial 2-aminoethylphosphonate (2AEP) transporters, named AepXVW, AepP and AepSTU, whose synthesis is independent of phosphate concentrations (phosphate-insensitive). AepXVW is found in diverse marine heterotrophs and is ubiquitously distributed in mesopelagic and epipelagic waters. Unlike the archetypal phosphonate binding protein, PhnD, AepX has high affinity and high specificity for 2AEP (*Stappia stellulata* AepX K_d 23 ± 4 nM; methylphosphonate K_d 3.4 ± 0.3 mM). In the global ocean, *aepX* is heavily transcribed (~100-fold>*phnD*) independently of phosphate and nitrogen concentrations. Collectively, our data identifies a mechanism responsible for a major oxidation process in the marine phosphorus redox cycle and suggests 2AEP may be an important source of regenerated phosphate and ammonium, which are required for oceanic primary production.

[1] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, UK. [2] Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK. [3] Nanotemper Technologies GmbH, Flößergasse 4, Munich, Germany. [4] School of Agriculture, Policy, and Development, University of Reading, Earley Gate, Whiteknights, Reading, UK. [5] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK. ✉email: I.lidbury@sheffield.ac.uk

Phosphonates are reduced organic phosphorus (P) molecules with a carbon (C)–P bond, as opposed to the more common C-oxygen (O)–P ester bonds found in many other organic P molecules[1]. Phosphonates are synthesised as both primary and secondary metabolites in various bacterial, archaeal and eukaryotic organisms[1–7] where they are incorporated into lipids (phosphonolipids) and glycans (phosphonoglycans)[4,8]. A significant proportion can also be released from the cell to facilitate favourable biotic interactions[9]. Thus, they are ubiquitous in terrestrial and aquatic ecosystems[10–14]. Phosphonates also represent a major fraction of the marine organic P pool[11,15–17] and recent studies have now identified several cosmopolitan marine microorganisms capable of synthesising significant quantities of these compounds[6,7,9,18,19]. Collectively, this synthesis drives a vast global oceanic P redox cycle with reduced P input in the surface ocean estimated to be an order of magnitude greater than (non-anthropogenic) riverine P input[9]. Whilst much attention has focused on the degradation of alkylphosphonates, such as (hydroxy-)methylphosphonate and 2-hydroxyethylphosphonate, as a source of P[16,17,20,21], the potential for aminophosphonates such as 2-aminoethylphosphonate (2AEP) to serve as sources of C and/or nitrogen (N) in the presence of inorganic phosphate (Pi), i.e. in a Pi-insensitive manner, has been neglected. However, emerging evidence suggests that Pi-insensitive 2AEP catabolism occurs in nature[22,23]. Notably, the fact that 2AEP is either absent from, or a minor component of, otherwise phosphonate rich high molecular weight dissolved organic matter (HMW DOM)[16,17], despite its supposed ubiquitous production[9,19,24,25], suggests preferential catabolism of this molecule in comparison to alkylphosphonates.

Unlike the majority of C–O–P monoester bonds, the C–P bond requires specific enzymes to break it, such as the C-P lyase[26,27]. Several 2AEP-specific phosphonate degradation systems have been characterised (Fig. 1a), including the 2AEP transaminase (PhnW) – phosphonoacetaldehyde hydrolase (phosphonatase-PhnX) system[28,29] and the PhnW – phosphonoacetaldehyde dehydrogenase (PhnY)-phosphonoacetate hydrolase (PhnA) system[30–32]. The phosphonate dioxygenase (PhnY*) - phosphohydrolase (PhnZ) system has also been shown to degrade 2AEP[33,34], though at least some homologs of this system are specific to (hydroxy-)methylphosphonate and cannot degrade 2AEP[35,36]. In addition, a gene encoding a recently characterised (*R*)-1-hydroxy-2-aminoethylphosphonate ammonia lyase (PbfA) is often found in *phnWX* and *phnWAY* operons[37], expanding the known repertoire of aminophosphonate degrading capabilites[37]. The C-P lyase, which is a non-specific promiscuous phosphonate degrading enzyme complex, is only induced in response to Pi-starvation, being regulated by the two-component master regulator of the Pi-stress response regulon, PhoBR[24]. In marine surface waters, genes encoding the C-P lyase are enriched in bacterial genomes found in regions typified by low Pi concentrations[20] where they are also heavily expressed[38]. Recent data has shown Pi-insensitive regulation of 2AEP degradation, facilitated by the 2AEP-specific systems, occurs in a few strains of bacteria related to marine *Alphaproteobacteria*[22] and a terrestrial gammaproteobacterium[23]. In both cases, a major consequence of Pi-insensitive 2AEP degradation was the remineralisation and release of labile Pi[22], due to the greater cellular demand for N over P and the 1:1 N:P stoichiometry of 2AEP.

When analytical methods are not sensitive enough to accurately quantify the concentration and turnover of specific environmental metabolites, screening for the expression of their respective uptake systems becomes an important tool in understanding their in situ cycling[38–41]. To date, only two 2-AEP transport systems have been identified. Both are ATP-binding cassette (ABC) transporters that consist of a periplasmic substrate-binding protein, an ATP-binding domain protein, and a transmembrane permease. The first is located within the C-P lyase operon (*phnCDEFGHIJKLMNOP*)[42,43] with genes encoding the substrate-binding protein, ATP-binding domain and transmembrane domains designated *phnD*, *phnC*, and *phnE*, respectively. The second is another ABC-transporter PhnSTUV shown by Jiang et al. to complement a C-P lyase knockout mutant of *E. coli* together with phnWX[28,29]. PhnD has a restricted distribution in seawater with an abundance that is highly correlated with regions of Pi-limitation[20]. However, no 2AEP transporter has been identified in the majority of bacteria possessing PhnWX and PhnWAY systems, which is surprising given the fact that 2AEP is a charged molecule and ubiquitous in marine and terrestrial ecosystems[1,20].

Here, we sought to identify transporters required for 2AEP catabolism in environmental bacteria lacking *PhnCDE* or *PhnSTU*. Through combining laboratory-based molecular and genetic analyses with environmental meta-omics, we identified three transporters that have a role in 2AEP uptake and revealed Pi-insensitive 2AEP catabolism is widespread in the global ocean, likely representing a major step in the marine P redox cycle.

Results

***Pseudomonas putida* BIRD-1 possesses a Pi-sensitive 2-aminoethylphosphonate ABC transporter, AepXVW**. We recently identified several candidate 2-AEP transporters (ABC-type) in *Pseudomonas* rhizobacteria that contain the PhnWX phosphonatase but lack both the archetypal PhnCDE transporter and PhnSTUV[44] (Fig. 1a). In *Pseudomonas putida* BIRD-1 (hereafter BIRD-1), a periplasmic substrate-binding protein associated with one of these putative transporters was induced under Pi-deplete growth conditions in a PhoBR-dependent manner[44]. We hereafter refer to this substrate-binding protein (PPUBIRD1_4925) as 2-**a**mino**e**thyl**p**hosphonate **X** (AepX). AepX belongs to the same family (pfam13343) as PhnS, iron and sulphate substrate-binding proteins, but is clearly distinct (Coverage = 40%, Identity = 25.09%, 1.1e^{-05}) (Fig. 1b).

BIRD-1 was capable of growth on 1.5 mM 2AEP as either a sole P, sole N or sole N and sole P source, the latter two resulting in mineralisation of excess Pi that was subsequently exported from the cell (Fig. 1c, Supplementary Fig. 1). Mutagenesis of *phnWX* confirmed this phosphonatase was essential for 2AEP catabolism under all growth conditions in this bacterium (Supplementary Fig. 2a, b). Next, we investigated if AepX and its corresponding ABC transporter components, the ATP binding domain protein (AepV), and the permease domain protein (AepW) were essential for its growth on 2AEP. Surprisingly, deletion of *aepXVW*^{BIRD} had no effect on growth as a sole N source (Fig. 1c). However, the mutant (Δ*aepXVW*^{BIRD}) had significantly ($p < 0.0001$) attenuated growth on 2AEP as a sole P source (Fig. 1c). The growth defect observed during growth on 2AEP as a sole P source was largely restored by complementing the mutant with a plasmid-encoded native homolog (Fig. 1c). These data suggest that whilst the AepXVW transporter is not essential, it is involved in 2AEP uptake as a sole P source in this bacterium but is not involved in growth as an N source. Therefore, another 2AEP transport system must also exist in this bacterium.

Identification of two Pi-insensitive 2-aminoethylphosphonate transporters in *P. putida* BIRD-1. Next, by subjecting the BIRD-1 Δ*aepXVW*^{BIRD} mutant to comparative proteomics, we identified a major facilitator-type transporter, (PPUBIRD1_3129), hereafter referred to as AepP for 2-**a**mino**e**thyl**p**hosphonate **p**ermease, whose expression was significantly increased during

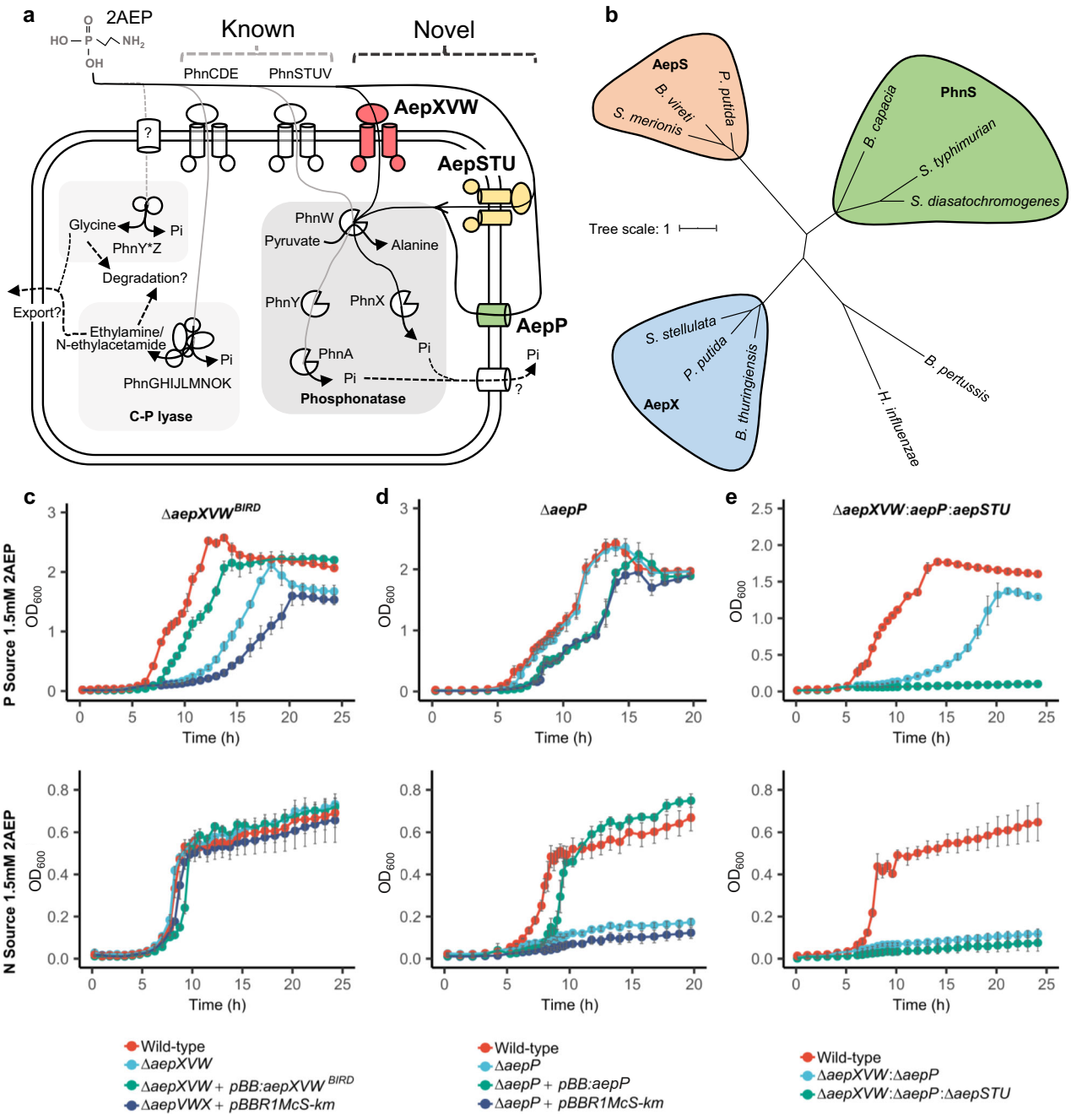


Fig. 1 2AEP transport and catabolism in *P. putida* BIRD-1. a Schematic representation of candidate (highlighted in bold and coloured) routes for 2AEP transport, together with existing characterised and putative 2AEP transport routes. Each catabolic system for degradation is highlighted and includes (i) the phosphonatase system comprising a 2AEP-pyruvate transaminase (PhnW) and either a phosphonoacetaldehyde hydrolase (PhnX)[28,29] or a NAD$^+$-dependent phosphonoacetaldehyde dehydrogenase (PhnY)[32] and a phosphonoacetate hydrolase (PhnA)[30,31], and (ii) the PhnY*Z system comprising phosphohydrolase (PhnZ)[33,34] and a 2-oxoglutarate dioxygenase (PhnY*)[33]. In addition, the promiscuous multi-subunit enzyme C-P lyase (PhnGHIJKLMN) can also act on 2AEP[26,27], as well as akylphosphonates. Pathways found in BIRD-1 are represented by black lines; pathways absent from BIRD-1 are shaded grey. Characterised pathways are shown with solid lines; uncharacterised pathways are shown with dashed lines. Transporters found in BIRD-1 are red if Pi-sensitive, green if Pi-insensitive, and yellow if constitutive. Unknown mechanisms are denoted by a '?'. **b** Phylogenetic tree of AepX, PhnS, and AepS, using the characterised Fe^{3+} substrate-binding protein FbpA from *Haemophilus influenzae* and *Bordetella pertussis* as an outgroup. *P. putida* = *Pseudomonas putida* BIRD-1, *S. stellulata* = *Stappia stellulata*, *B. cepacia* = *Burkholderia cepacia*, *B. vireti* = *Bacillus*, *S. merionis* = *Streptococcus merionis*, *S. typhimurium* = *Salmonella typhimurium*, *S. diastatochromogenes* = *Streptomyces diastatochromogenes*, *B. thuringiensis* = *Bacillus thuringiensis*. **c** Growth ($n = 4$) of *P. putida* BIRD-1 wild type, Δ*aepXVW*, and the complemented mutant. **d**, Growth ($n = 4$) of *P. putida* BIRD-1 wild type, Δ*aepP* and the complemented mutant, and (**e**), Growth ($n = 4$) of *P. putida* BIRD-1 wild type, the double mutant, Δ*aepXVW:*Δ*aepP*, and the 2AEP null transporter mutant, Δ*aepXVW*:Δ*aepP:*Δ*aepSTU*. All strains used 2AEP as a sole P (top panel) or N (bottom panel) source. Error bars denote standard deviation of the mean.

growth on 2AEP as a sole N source (Pi-insensitive) (Supplementary Fig. 3, Supplementary Table 3a). AepP is related to the glycerol-3-phosphate (G3P): Pi antiporter, GlpT[45,46] (identity = 27.75%, $9e^{-37}$), a member of the organophosphate: Pi antiporter family of major facilitator transporters. AepP shares conserved residues essential for binding Pi and the Pi moiety of G3P[47–50] with GlpT, whereas residues that impact the binding affinity to the glycerol moiety of G3P but not Pi are not conserved[49] (Supplementary Fig. 4). Mutation of *aepP* in either the wild type parental strain (Δ*aepP*, Fig. 1d) or the *aepXVW* mutant (Δ*aepXVW* Δ*aepP*, Fig. 1e) led to an inability to grow on 2AEP as sole N source. Subsequent complementation of Δ*aepP* with its native homolog (Δ*aepP* + *pBB:aepP*) restored growth on 2AEP as a sole N source (Fig. 1d). Interestingly, delayed but significant growth on 2AEP as a P source still occurred in the Δ*aepXVW* Δ*aepP* double mutant, revealing the presence of a third 2AEP transporter in this bacterium.

To identify the unknown 2AEP transporter, we reanalysed our proteomics data (Supplementary Fig. 3). Another substrate binding protein (PPUBIRD1_3891) containing the same pfam domain (pfam13343) as AepX (Fig. 1b), hereafter named AepS, was constitutively synthesised in all growth conditions. In order to uncover the role of AepS in the utilisation of 2AEP as a sole P source, a triple mutant Δ*aepXVW* Δ*aepP* Δ*aepSTU* was generated in BIRD-1 (Fig. 1e). This triple knockout mutant was unable to grow on 2AEP as a sole P source (Fig. 1e), suggesting AepSTU is a functional 2AEP transporter. However, generation of a single Δ*aepSTU* knockout mutant did not affect Pi-sensitive growth compared to the wild type (Supplementary Fig. 2c), suggesting AepXVW is the major transporter involved in Pi-sensitive 2AEP uptake and AepSTU only has an auxiliary role in 2AEP uptake. Despite production of AepS during Pi-insensitive growth (Supplementary Fig. 3), this transporter was unable to facilitate growth on 2AEP as a sole N source in the absence of AepP. Together, these data reveal the presence of three differentially regulated 2AEP transporters in BIRD-1. AepXVW is the primary Pi-sensitive 2AEP transporter with AepSTU having an auxiliary role, whereas AepP is essential for Pi-insensitive growth on 2AEP.

AepXVW is found in several marine bacteria capable of Pi-insensitive mineralisation and is a functional 2-aminoethylphosphonate transporter. Using PPUBIRD1_4925 (AepX) as the query, we scrutinised the genomes of several isolates related to marine *Rhodobacteraceae* (*Stappia* spp., *Terasakiella* spp., *Falsirhodobacter* spp.) capable of Pi-insensitive phosphonate catabolism[22]. ORFs encoding orthologs of AepX were identified in the genomes of *Stappia stellulata* DSM 5886 and *Terasakiella pusilla* DSM 6293, in addition to several other marine Roseobacter strains: *Aliiroseovarius crassostreae* (DSM 16950), *Aliiroseovarius sediminilitoris* (DSM 29439), *Shimia marina* (DSM 26895), and *Thalassobius aestuarii* (DSM 15283) (Fig. 2a). We also found an orthologous ORF in the model rhizosphere alphaproteobacterium *Sinorhizobium meliloti* strain 1021 that is capable of 2AEP catabolism via PhnWAY[31] (Fig. 2a). In all cases, ORFs encoding AepXVW were located adjacent to ORFs encoding PhnWAY or PhnWX, strongly suggesting a role in 2AEP transport. *S. stellulata* DSM 5886, *A. crassostreae* DSM 16950, *A. sediminilitoris* DSM 29439, *S. marina* DSM 26895, and *T. aestuarii* DSM 15283 were all capable of growth on 2AEP as either the sole N or P source (Supplementary Table 1, Supplementary Fig. 5). Indeed, both *Aliiroseovarius* strains lack other characterised 2AEP transport and degradation systems (Supplementary Table 2). As previously reported, Pi was exported from cells and accumulated in the medium during growth on 2AEP as a sole N source (Supplementary Fig. 6a). To confirm that *S. stellulata* AepXVW can take up 2AEP, we complemented the BIRD-1 null 2AEP transporter mutant (Δ*aepXVW:*Δ*aepP:*Δ*aepSTU*) with this transporter fused with the $aepXVW^{BIRD}$ promoter (Fig. 2b). This duly restored growth of the triple mutant confirming *S. stellulata* AepXVW is also a functional 2AEP transporter.

In many 2AEP gene clusters, we identified LysR-type regulators, which we refer to as AepR. A homolog of AepR has been shown to be essential for complementation of an *Escherichia coli* Δ*phnHIJKLMNOP* mutant with a Pseudomonad PhnWX[25], implying substrate inducible regulation. Additionally, PhnA activity has been shown to be induced by 2AEP in a marine *Falsirhodobacter* isolate even under nutrient replete conditions[22], though unfortunately no sequenced *Falsirhodobacter* strain possesses an aminophosphonate operon so it is not clear if AepR is responsible for this regulation. BIRD-1 and other Pseudomonads, whose 2AEP operons are often fragmented throughout the genome, possess up to three distinct genes encoding LysR-type regulators (Fig. 2a). In contrast, most *Alphaproteobacteria* only possess a single gene, located upstream of the *aepXVW-phnWAY* operon. These newly identified forms are phylogenetically distinct from either the archetypal PhnR[51] or PalR[52] found in *P. fluorescens* sp. 23F and *Variovorax* sp. PAL2, respectively (Fig. 2c). The three BIRD-1 LysR-like forms were clearly distinct from each other with the AepP-associated form being closely related to the single LysR-type regulator found in *Alphaproteobacteria*.

AepXVW is highly synthesised in the marine bacterium *Stappia stellulata* during Pi-sensitive and Pi-insensitive growth on 2-aminoethylphosphonate. In BIRD-1, AepXVW was only involved in Pi-sensitive growth whilst AepP was synthesised during Pi-insensitive and Pi-sensitive metabolism (Fig. 1). *S. stellulata* lacks AepP but is still capable of Pi-insensitive growth and Pi export when grown on 2AEP as a sole N, or sole N and sole P source (Supplementary Fig. 5a). In addition to *aepXVW* and genes encoding the 2AEP degradation system (*phnWAY*), *S. stellulata* also possesses genes (*phnCDEFGHIJKLMNO*) encoding the P-regulated C-P lyase operon and we also confirmed this strain grew on several other alkylphosphonates as sole P source (Supplementary Fig. 5b). Therefore, to determine which transport and degradation systems were upregulated during growth on 2AEP as either a sole N or P source, we performed comparative proteomics. Unlike BIRD-1, AepX was abundantly synthesised during growth on 2AEP as either a sole N or P source, as was PhnWAY and PbfA, the (*R*)-1-hydroxy-2AEP ammonia lyase, whilst the C-P lyase encoded proteins (including PhnD) were not (Fig. 3). Importantly, whilst we detected several general N stress–response proteins induced under Pi-insensitive growth, we did not identify any other potential 2AEP transporters (Supplementary Table 3b). Therefore, unlike in BIRD-1, AepXVW likely represents the major route for 2AEP uptake in this bacterium. These data are consistent with the hypothesis that *S. stellulata* 2AEP catabolism is regulated by a LysR-type regulator solely through substrate-induction, which will be investigated in a future study.

Finally, we determined the substrate specificity of recombinant *S. stellulata* AepX towards 2AEP and other (alkyl)phosphonates using microscale thermophoresis[53,54]. Unlike the relatively promiscuous phosphonate binding protein PhnD[43], AepX was highly specific for 2AEP with a K_d in the nanomolar range (Table 1, Supplementary Fig. 7), consistent with the observation that *aepXVW* is typically co-localised with either *phnWX* or *phnWAY* that encode 2AEP-specific degradation systems (Fig. 2a).

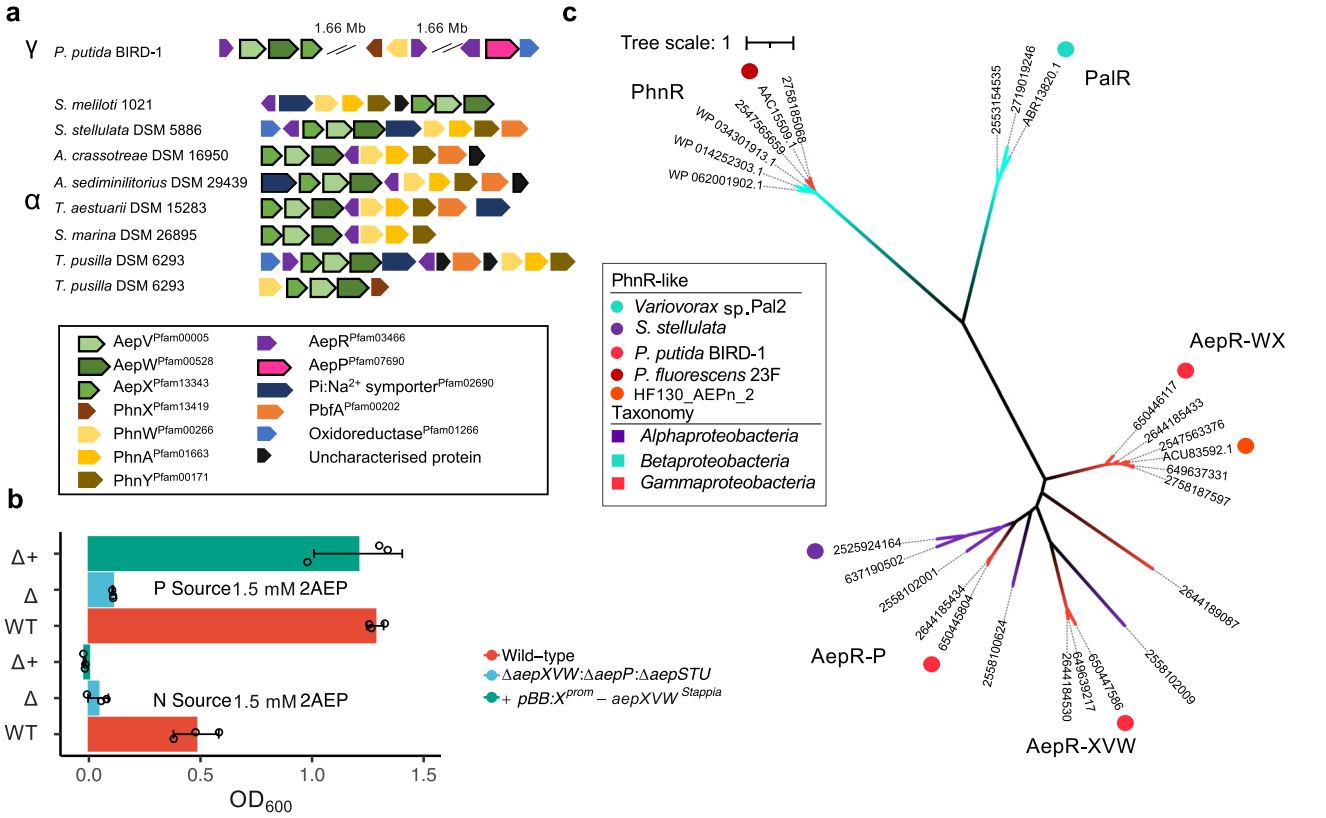


Fig. 2 Distribution and functional characterisation of AepXVW in marine bacteria. a Genetic neighbourhoods of *aepXVW* within marine *Alpha-* and terrestrial *Gamma-proteobacteria*. Strains shown are *Pseudomonas putida* BIRD-1, *Sinorhizobium meliloti* 1021, *Stappia stellulata* DSM 5886, *Aliiroseovarius crassostreae* DSM 16950, *Aliiroseovarius sediminilitoris* DSM 29439, *Thalassobius aestuarii* DSM 15283, and *Shimia marina* DSM 26895. ORFs separated on the genome are indicated by breaks with the corresponding gap given in megabases (Mb). **b** Growth of the *P. putida* BIRD-1 triple 2AEP transporter mutant (Δ*aepXVW*:Δ*aepP*:Δ*aepSTU*) complemented with *aepXVW*^{Stappia} concatenated with the promoter region from *aepXVW*^{BIRD} on 2AEP as either a sole N (60 h) or P (48 h) source. Data represent the mean of triplicates cultures. Error bars denote standard deviation. **c** Phylogeny of phosphonate-associated LysR-type regulators. Labels denote IMG/JGI gene IDs or Genbank accession numbers. Tree topology and branch lengths were calculated by maximum likelihood using the LG + G4 model of evolution for amino acid sequences based on 744 sites in IQ-TREE software[84]. Tree Scale represents the number of substitutions per site. A consensus tree was generated using 1000 bootstraps.

***aepX* and *aepP* are found in distantly related and cosmopolitan bacterial taxa**. Using the Integrated Microbial Genomes/Microbiomes from the Joint Genome Institute (IMG/M/JGI) database, we identified ORFs encoding AepX and AepP (but not AepS) in genomes retrieved from both taxonomically divergent isolates as well as single amplified genomes and metagenome-assembled genomes, which revealed an unexpected diversity for these substrate-binding proteins (Fig. 4). For AepX, this included cosmopolitan marine *Alphaproteobacteria* in addition to *Rhodobacteraceae*, marine *Deltaproteobacteria*, and marine *Vibrio* spp. AepX was also found in terrestrial *Betaproteobacteria*, *Firmicutes*, and other gram-positive bacteria (Fig. 4). AepX was partitioned into several subclades, with AepX^{Stappia} and AepX^{BIRD} well separated (Fig. 4). Many taxonomically divergent AepX encoding ORFs were co-localised with ORFs encoding the various 2AEP degradation systems, including the (*R*)-1-hydroxy-2-aminoethylphosphonate specific PbfA, the C-P lyase, or putative uncharacterised ORFs encoding potentially novel phosphonate catabolic enzymes, supporting a role in 2AEP transport (Fig. 4).

AepP was also found in a wide range of phylogenetically divergent taxa, such as *Acidobacteria* (*Granuliella mallensis*) and *Bacteroidetes* (*Kriegella aquimaris*), *Actinobacteria* (*Streptomyces albulus* CCRC 11814) and *Verrucomicrobia* (*Haloferula* sp. BvORR071 and *Verrucomicrobia sp.* SGGC AC-337 J20) (Supplementary Fig. 6). However unlike AepX, AepP was not found in cosmopolitan marine bacteria. Again, for all of these strains ORFs encoding AepP were co-localised with ORFs encoding PhnWAY or PhnWX, or putative catabolic enzymes (Supplementary Fig. 8). Notably, AepP was found in fewer marine isolates compared to AepX.

***aepX* gene and transcript abundance is far greater than the archetypal *phnD/phnS* in the global ocean**. Using the TARA oceans OM-RGCv2 + G metagenome (MG) and OM-RGCv2 + T metatranscriptome (MT) datasets[55], we calculated the abundance of *aepX*, *aepS* and *aepP* and compared this with *phnS* and the gene (*phnD*) encoding the archetypal phosphonate transporter, whose gene abundance in seawater was recently calculated[20]. The distribution of markers (*phnJ, phnA, phnX*) for the various phosphonate degradation mechanisms were also analysed (Supplementary Fig. 9). We analysed data from both the epipelagic and mesopelagic zones where phosphonate mineralisation is believed to occur[15]. Across all oceanic sampling sites in both the epipelagic and mesopelagic, *aepX* gene and transcript abundance was significantly greater (MG; post-hoc Dunn's test $z = 10.4$, $p < 0.001$ and $z = 4.8$, $p < 0.001$, respectively) than *phnD* (Fig. 5a, b). On average, in the mesopelagic almost 10% of bacterial cells possess *aepX* whilst only ~0.3% and 0.5% possess *phnD* and *aepP*, respectively (Fig. 5a). The transcript abundance of *aepX* was 40-fold and 140-fold greater than *phnD* in the epipelagic and mesopelagic, respectively (Fig. 5b). The majority of *aepX* sequences were related to the cosmopolitan *Alphaproteobacteria*, although some deltaproteobacterial

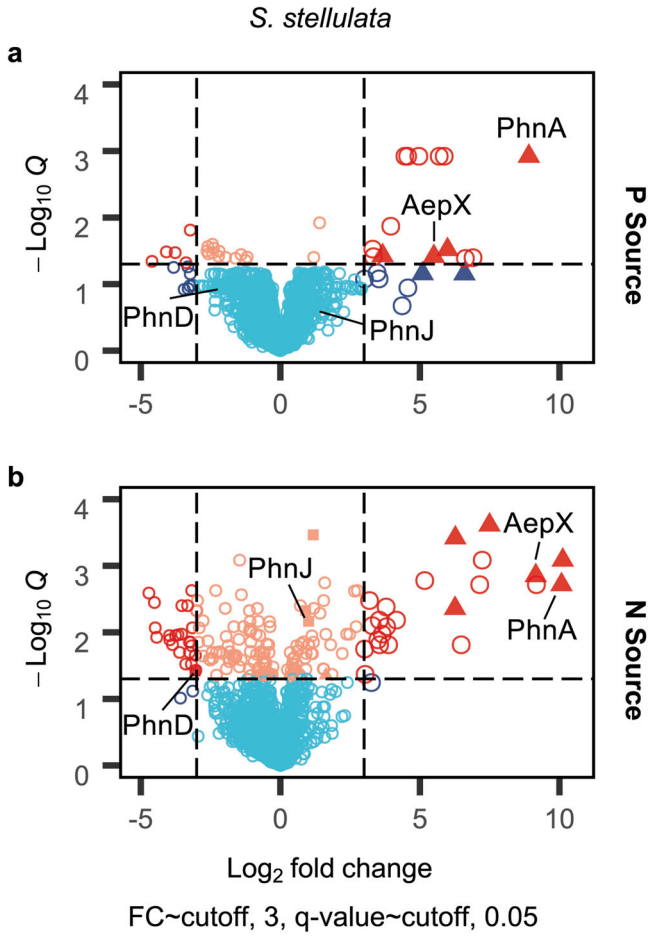


Fig. 3 Proteomic analysis of 2AEP-grown *S. stellulata* cells. Whole-cell protein profiles ($n = 3$) for *S. stellulata* grown using either Pi or 2AEP as sole P source (**a**) or NH_4 or 2AEP as the sole N source (**b**). Fold change (FC) represents the difference in Log_2 Label Free Quantification (LFQ) values between each treatment and the statistical value on the *Y* axis is generated from *Q* values (FDR corrected P values). Members of the *aepXVW-phnWAY* operon are shown as triangles, members of the C-P lyase operon are shown as squares, all other proteins are shown as open circles. Data plotted represents the mean of triplicate cultures. Vertical dashed lines represent an Log_2 LFQ difference > −3 or <3. The horizontal dashed line illustrates a cut off for a significant *Q* value ($p < 0.05$). Sky blue represents proteins showing no significant difference between treatments. Red indicates proteins significantly changing in abundance <−3-fold or >3-fold, respectively. Peach indicates significant changes less than 3-fold in either direction.

Table 1 Microscale thermophoresis determined dissociation constants (K_d) of the *S. stellulata* AepX for selected phosphonate ligands.

Ligand	K_d (μM)	
	AepX Ss	PhnD Ec
2-Aminoethylphosphonate	0.023 ± 0.004	>0.050–0.1
Methylphosphonate	3404.15 ± 280.99	1.3–18.4
Ethylphosphonate	145.96 ± 15.18	0.2–1.4
Aminomethylphosphonate	4490.91 ± 808.24	16.6

Results are compared against those obtained for *E. coli* PhnD in previous studies[42,65].

sequences fall within this clade (Fig. 4). We confirmed that these abundant environmental sequences were also co-localised with characterised and putative phosphonate degradation genes (Fig. 4). In broad agreement with *aepX*, the cumulative transcription of the two markers *phnA* and *phnX* is significantly greater than the C-P lyase marker *phnJ* (Kruskal–wallis $X^2 = 206.6$, $p < 0.001$), strengthening the observation that 2AEP mineralisation is a major oceanic process (Supplementary Fig. 9). The *phnZ* marker is split into several subclades, with only the original PhnY*Z specific for 2AEP (Supplementary Fig. 10). This 2AEP-specific form was found at very few sites (MG = 9; MT = 2) and in low abundance in both the MG and MT (Supplementary Fig. 10). Homologs related to the two PhnZ clades associated with either methylphosphonate[35,36] or (*N*)-trimethyl-2-aminoethylphosphonate[56] degradation were found at comparable gene and transcript abundances to *phnJ* and significantly lower than *phnA* (Supplementary Fig. 11).

The gene abundances of *aepP*, *aepS* and *phnS* were all significantly lower than both *aepX* and *phnD* (post-hoc Dunn's test $z = 13.1$ and 9.0, $p < 0.001$, $z = 14.5$ and 12.6, $p < 0.001$, and $z = 9.9$ and 8.9, $p < 0.001$ respectively) in the epipelagic, whilst only *aepS and phnS* were significantly lower in the mesopelagic (post-hoc Dunn's test $z = 10.1$ and 10.8, $p < 0.001$, and $z = 4.6$ and 5.5, $p < 0.001$ respectively) (Fig. 5a and S12).

Unlike *phnD* and *aepP*, *aepX* abundance was comparable across all oceanic regions within both MG and MT at each depth suggesting 2AEP mineralisation is a ubiquitous process in seawater (Fig. 5c, d). For all sites at each depth, the relative abundance of *aepX* transcripts was always significantly greater (Wilcoxon rank sum $W = 3771$, $p < 0.001$, estimated log_2 difference = 2.24 (95% CI 1.97–2.52)) than its own gene abundance. For *phnD*, we observed significantly greater transcript abundance compared to its own gene abundance only in the Mediterranean Sea, a region typified by Pi-depletion (Wilcoxon rank sum $W = 0$, $p < 0.001$, estimated log_2 difference = 3.11 (95% CI 1.90–4.42)). Finally, in agreement with previous work[20], *phnD* gene abundance was inversely correlated ($R^2 = 0.340$, $p < 0.001$) with standing stock concentrations of Pi (Fig. 5e), as was *phnD* transcript abundance (Fig. 5f). In contrast, *aepX* and *aepP* gene abundance were positively correlated ($R^2 = 0.098$, $p < 0.001$ and $R^2 = 0.291$, $p < 0.001$, respectively) with Pi concentration (Fig. 5e), whilst no significant relationship between Pi concentration and *aepX*/*aepP* transcript abundance was found (Fig. 5f), suggesting their expression is independent of Pi in seawater globally.

To better understand the parameters controlling 2AEP catabolism in the ocean, we compared both gene and transcript abundance in relation to R*, a measure of N vs P limitation calculated as $[NO_2] + [NO_3] - 16[PO_4]$ (adapted from Smith et al.[57]). As expected, *phnD* gene and transcript abundance were positively correlated with R* ($R^2 = 0.168$, $p < 0.001$ and $R^2 = 0.197$, $p < 0.001$, respectively) (Fig. 5g, h), i.e. regions typified by Pi depletion, though Pi concentration alone was a better predictor of *phnD* gene abundance (Fig. 5e). However, *aepX* and *aepP* gene abundance was (weakly) inversely correlated with R* ($R^2 = 0.029$, $p < 0.05$ and $R^2 = 0.108$, $p < 0.001$, respectively) (Fig. 5g) and no significant relationship was found between R* and *aepX*/*aepP* transcript abundance (Fig. 5h). These data are consistent with the proteomic response of *S. stellulata* under laboratory conditions and suggests AepX is induced in the presence of 2AEP (substrate-inducible) and not in response to nutrient limitation.

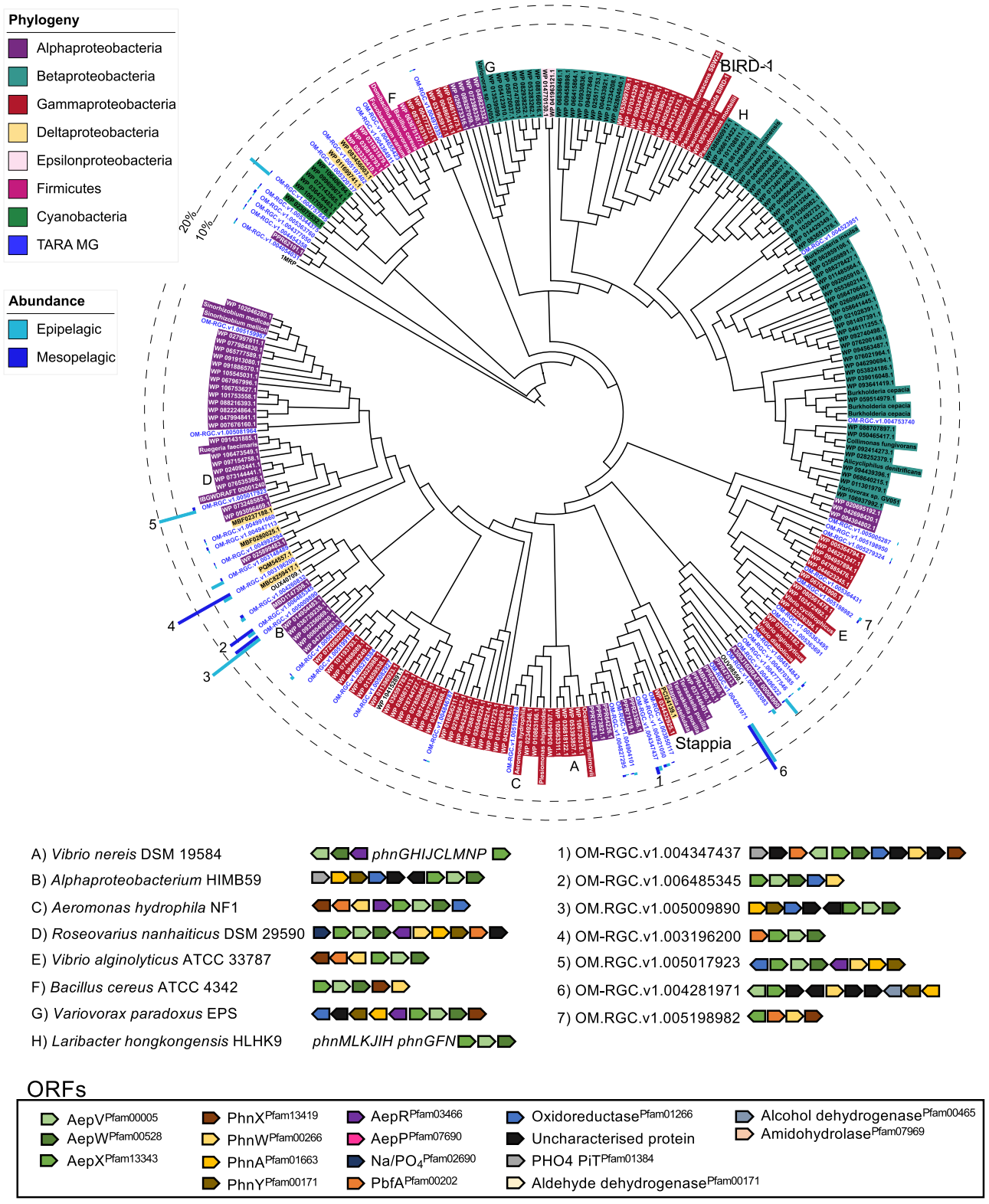


Fig. 4 Phylogenetic and genomic analyses of AepX in marine and terrestrial bacteria. Genetic neighbourhoods for selected AepX homologs are presented adjacent to trees. Numbers indicate environmental operational taxonomic units (OTUs) and letters indicate isolates, metagenome-assembled genomes (MAGs) or single-cell amplified genomes (SAGs). Tree topology and branch lengths were calculated by maximum likelihood using the LG + F + I + G4 model of evolution for amino acid sequences based on 744 sites in IQ-TREE software[84]. A consensus tree was generated using 1000 bootstraps. Branches representing isolates or MAGs/SAGs are colour coded based on their phylogenetic affiliation (see legends). Branches and identifiers for representative environmental OTU sequences (clustered at 0.8) retrieved from the TARA Oceans database are highlighted blue. The outer ring denotes the relative abundance of environmental AepX OTUs using the same colour scheme; 10% (dashed line) and 20% (filled line) thresholds are shown for scale. *S. stellulata* DSM 5886 and *P. putida* BIRD-1 AepX are labelled.

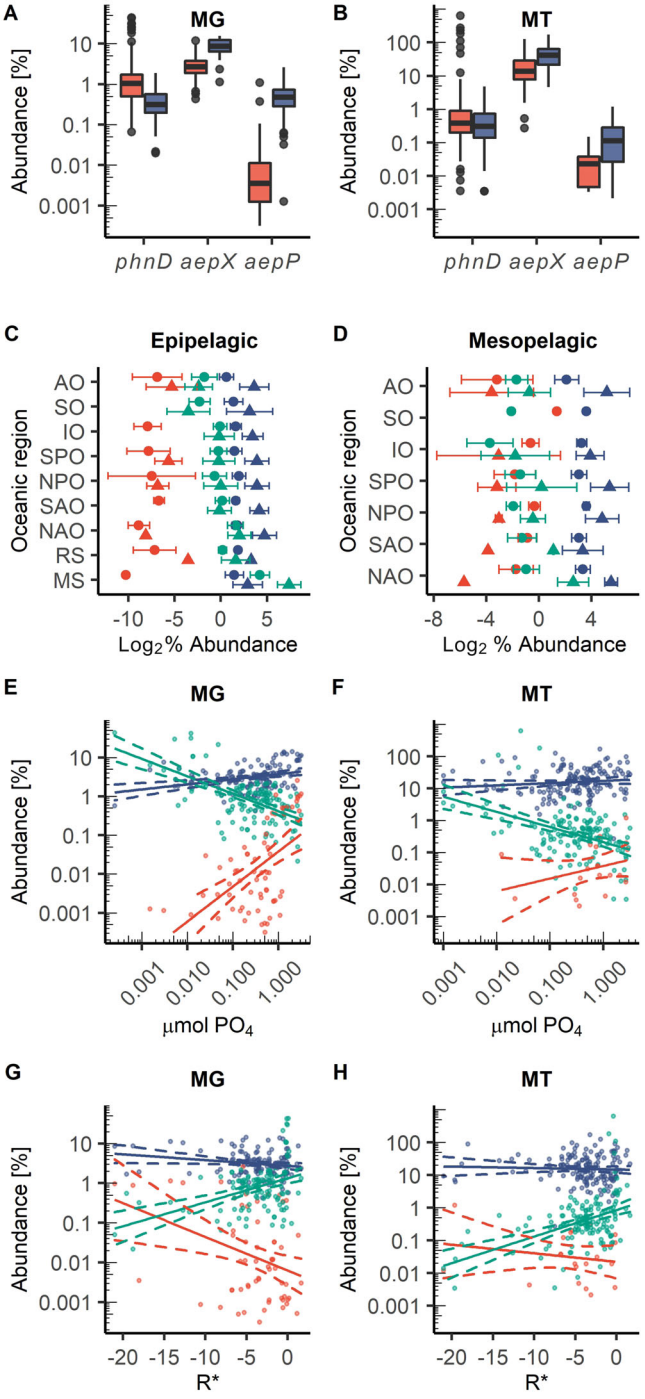


Fig. 5 Distribution and expression of phosphonate transporter genes in the global ocean. Abundance (% abundance [gene or transcript] relative to the median abundance [gene or transcript] of 10 single copy core genes) of *phnD*, *aepX*, *aepP* in marine epipelagic (red) and mesopelagic (blue) waters, split by metagenome (MG) (**a**) (epipelagic: *phnD* $n = 137$, *aepX* $n = 137$, *aepP* $n = 60$, mesopelagic: *phnD* $n = 43$, *aepX* $n = 43$, *aepP* $n = 42$, where n equals the number of biologically independent sampling sites where the genes were located), and metatranscriptome (MT) (**b**) (epipelagic: *phnD* $n = 148$, *aepX* $n = 154$, *aepP* $n = 11$, mesopelagic: *phnD* $n = 32$, *aepX* $n = 33$, *aepP* $n = 18$, where n equals the number of biologically independent sampling sites where the transcripts were located). Abundance of *phnD*, *aepX*, *aepP* in MG (circles) and MT (triangles) in epipelagic (**c**) and mesopelagic (**d**) waters, split by oceanic region. *aepP* (red), *phnD* (green), *aepX* (blue). AO Arctic Ocean, SO Southern Ocean, IO Indian Ocean, SPO South Pacific Ocean, NPO North Pacific Ocean, SAO South Atlantic Ocean, NAO North Atlantic Ocean, RS Red Sea, MS Mediterranean Sea. Circles are mean values of Log_2 abundance, error bars represent standard deviation of the mean. Epipelagic (**c**) *phnD* AO MG $n = 29$, MT $n = 26$, SO MG $n = 3$, MT $n = 8$, IO MG $n = 21$, MT $n = 18$, SPO MG $n = 25$, MT $n = 35$, NPO MG $n = 11$, MT $n = 20$, SAO MG $n = 14$, MT $n = 17$, NAO MG $n = 16$, MT $n = 17$, RS MG $n = 6$, MT $n = 3$, MS MG $n = 12$, MT $n = 7$, *aepX* AO MG $n = 29$, MT $n = 28$, SO MG $n = 3$, MT $n = 8$, IO MG $n = 21$, MT $n = 19$, SPO MG $n = 25$, MT $n = 35$, NPO MG $n = 11$, MT $n = 20$, SAO MG $n = 14$, MT $n = 17$, NAO MG $n = 16$, MT $n = 17$, RS MG $n = 6$, MT $n = 3$, MS MG $n = 12$, MT $n = 7$, *aepP* AO MG $n = 11$, MT $n = 3$, SO MG $n = 0$, MT $n = 0$, IO MG $n = 10$, MT $n = 0$, SPO MG $n = 13$, MT $n = 4$, NPO MG $n = 5$, MT $n = 2$, SAO MG $n = 4$, MT $n = 0$, NAO MG $n = 11$, MT $n = 1$, RS MG $n = 4$, MT $n = 1$, MS MG $n = 2$, MT $n = 0$. Mesopelagic (**d**) *phnD* AO MG $n = 9$, MT $n = 7$, SO MG $n = 1$, MT $n = 0$, IO MG $n = 6$, MT $n = 4$, SPO MG $n = 9$, MT $n = 9$, NPO MG $n = 5$, MT $n = 5$, SAO MG $n = 5$, MT $n = 2$, NAO MG $n = 8$, MT $n = 5$, *aepX* AO MG $n = 9$, MT $n = 7$, SO MG $n = 1$, MT $n = 0$, IO MG $n = 6$, MT $n = 4$, SPO MG $n = 9$, MT $n = 9$, NPO MG $n = 5$, MT $n = 5$, SAO MG $n = 5$, MT $n = 2$, NAO MG $n = 8$, MT $n = 6$, MT $n = 7$, *aepP* AO MG $n = 9$, MT $n = 6$, SO MG $n = 1$, MT $n = 0$, IO MG $n = 6$, MT $n = 2$, SPO MG $n = 9$, MT $n = 6$, NPO MG $n = 5$, MT $n = 2$, SAO MG $n = 5$, MT $n = 1$, NAO MG $n = 7$, MT $n = 1$. The relationship between the standing stock Pi concentration and transporter abundance, analysed by linear regression of Log_{10} Pi concentration and Log_{10} gene/transcript abundance, in the MG (**e**), (*aepX* $R^2 = 0.098$, $p = 1.477e\text{-}4$, *phnD* $R^2 = 0.340$, $p = 1.349e\text{-}13$, *aepP* $R^2 = 0.291$, $p = 1.466e\text{-}6$) and MT (**f**), (*aepX* $R^2 = 0.007$, $p = 0.1544$, *phnD* $R^2 = 0.203$, $p = 4.066e\text{-}9$, *aepP* $R^2 = 0.058$, $p = 0.1581$). *aepP* (red), *phnD* (green), *aepX* (blue). The relationship between R*, a measure of N vs P limitation defined as the sum of standing stock nitrate plus nitrite concentration minus 16x standing stock Pi concentration, and transporter abundance, analysed by linear regression of R* and Log_{10} gene/transcript abundance, in the MG (**g**) (*aepX* $R^2 = 0.029$, $p = 0.033$, *phnD* $R^2 = 0.168$, $p = 1.611e\text{-}6$, *aepP* $R^2 = 0.108$, $p = 5.31e\text{-}3$) and MT (**h**) (*aepX* $R^2 = -0.005$, $p = 0.606$, *phnD* $R^2 = 0.197$, $p = 1.693e\text{-}8$, *aepP* $R^2 = -0.014$, $p = 0.402$). *aepP* (red), *phnD* (green), *aepX* (blue). 95% confidence intervals are shown by dashed lines. In (**a**, **b**) data are represented as boxplots, where the middle line is the median and the upper and lower hinges correspond to the first and third quartiles. The upper whisker extends from the upper hinge to the largest value that is no more than 1.5 × IQR (inter-quartile range) from the upper hinge, and the lower whisker extends from the lower hinge to the smallest value that is no further than 1.5 × IQR from the lower hinge. Data beyond the ends of the whiskers are outlying points that are plotted individually. Note: *aepP* transcripts were not detected at numerous sites which is represented by an omission of data points.

Discussion

Both phosphonate biosynthesis[1,3,5,58] and catabolic[1,23,25,31,59–64] genes are ubiquitous in marine, soil and gut microbiomes, suggesting phosphonate cycling is widespread in nature. In contrast, the uptake of these molecules is comparatively understudied, with only two characterised ABC transport systems confirmed, both of which are linked solely to P-acquisition[28,42,43]. The abundance of these Pi-sensitive transporters in marine systems is not equivalent to the abundance of catabolic genes[20], especially those (*phnWAY*) recently shown to be involved in Pi-insensitive catabolism[22]. This would suggest our knowledge of the microbial uptake of phosphonates, particularly 2AEP, is incomplete. Using transporter expression as a proxy for the cycling of specific nutrients has significantly advanced our understanding of in situ biogeochemical cycling[39,40,65]. This molecular approach helped resolve the biogenesis of the climate-active gas methane in oxygenated surface waters, driven through the uptake and degradation of methylphosphonate[21,38]. Thus, a gap in mechanistic knowledge on 2AEP metabolism impairs our ability to survey the in situ

cycling of reduced organophosphorus compounds, especially when high-resolution separation of such compounds is difficult[9,66]. Here, identification of Pi-insensitive 2AEP transporters allowed us to develop molecular markers to investigate the cycling of 2AEP on a global scale and compare these with previously characterised Pi-sensitive markers.

To date, whilst several ABC transporters show preference for methylphosphonate and phosphite[53,67], or bind a range of phosphonates with low micromolar or lower K_d[43], no ABC transporter showing a strong preference for 2AEP has been identified. Here, we revealed AepX appears to be highly specific for 2AEP and has substantially lower affinity for methylphosphonate or ethylphosphonate than PhnD[43]. The occurrence of *aepXVW* adjacent to putative phosphonate catabolic genes, and the characterised PbfA[37], does suggest some degree of promiscuous binding, albeit likely to related aminophosphonates. Thus, the molecular mechanisms governing the specificity of AepX towards aminophosphonates warrant further investigation.

Genomic and biochemical analyses have revealed 2AEP and the alkylphosphonates, methylphosphonate and hydroxyethylphosphanate are ubiquitously synthesised in the marine environment in relatively large quantitities[1,3,9,25,66]. However, several studies have shown 2AEP is not detected as a significant component of 'semi-labile' DOM whilst alkylphosphonates tend to accumulate[16,17,68]. Collectively, this would suggest 2AEP is more susceptible to microbial mineralisation[22] and thus shorter residence times. Here, we reveal genes for the Pi-insensitive uptake (*aepX*) and catabolism (*phnA*) are expressed at significantly higher levels across the global ocean than the Pi-repressible *phnD* and *phnJ*, providing a clear mechanism for this phenomenon. Our data also suggest 2AEP is preferentially mineralised independently of both N and P status, explaining why phosphonates are metabolised in regions where Pi or ammonium concentrations are high enough to repress C-P lyase-expression[20,21,69]. Together, this adds further weight to the notion that phosphonates are rapidly cycled between reduced and oxidised forms and are a source of regenerated Pi throughout the water column[9,15,22].

Substrate inducible expression of catabolic genes targeting organic N molecules, irrespective of nutrient status, has previously been shown to drive mineralisation of N and cross feed into surrounding microbial cells[70–73]. Indeed, ammonium mineralisation may also occur if 2AEP, (*R*)-1-hydroxy-2-aminoethylphosphonate or (*N*)-trimethyl-2-aminoethylphosphonate are also used as carbon and energy sources, similar to methylamines and quaternary amines[70,71]. In agreement with Chin et al.[22], our proteomic data for *S. stellulata* and in situ environmental data strongly suggests PhnA and AepX are highly synthesised in a substrate-inducible manner that would facilitate the remineralisation of labile inorganic N and P. Even if ammonium concentration does play a role in the occurrence and regulation of 2AEP degradation genes (i.e. 2AEP is primarily a N source), our combined data clearly demonstrates the potential for the in situ mineralisation of semi-recalcitrant DOP into labile Pi, a mechanism which is important for maintaining biological production in Pi-deplete regions of the ocean[74,75]. It should be noted that the *Pseudomonas aepR* located adjacent to *aepP* is most similar to marine *aepR*; this could explain why AepP was more abundant in the presence of 2AEP (Fig. 1d), although the differences in growth rate and protein abundance suggest substrate induction is not the sole mechanism of regulation in *P. putida* BIRD-1.

In summary, this study identified three 2AEP transporters in marine and terrestrial bacteria capable of Pi-insensitive 2AEP catabolism. The substrate-binding protein for one of these, AepX, is the most abundant and highly transcribed of the known phosphonate binding proteins in seawater. Therefore, 2AEP mineralisation may represent a major process in the marine organic N and P cycles and a significant source of regenerated labile Pi for oceanic production. This conclusion is strengthened by two key observations: (1) *aepX* transcription is not repressed by standing stock concentrations of Pi, and (2) extracellular Pi export occurs during Pi-insensitive 2AEP metabolism. Thus, we provide further evidence for the role of low molecular weight phosphonates acting as a P currency between autotrophic and heterotrophic microbes[9].

Methods

Bacterial strains and growth conditions. All strains used in this work were axenic. *Pseudomonas* strains were maintained on Luria Bertani (LB) agar (1.5% w/v) medium at 30 °C. *Stappia stellulata* and the Roseobacter strains were maintained on Marine Broth agar (1.5% w/v) medium at 30 °C. *Pseudomonas* mutants and complemented strains were maintained on similar plates containing the appropriate antibiotic. For all growth and proteomics experiments cultures were grown in an adapted minimal A medium[44] using Na-Succinate (20 mM) as the sole carbon source and, where applicable, 10 mM NH_4Cl was added as the sole N source. 2AEP and KH_2PO_4 were added to a final concentration of 100 µM or 1.5 mM as specified in the text and figure legends. *Pseudomonas* strains were pre-cultured in minimal medium A containing 100 µM Pi and 1.5 mM NH_4Cl to ensure adequate growth while minimising the potential for carryover of residual nutrients into experimental cultures. Culture experiments were performed using a FLUOStar Omega 96-well plate reader using Sarstedt 96-well plates incubated at 30°C, shaking at 200 rpm. For proteomics where 2AEP is used as an N source, *S. stellulata* was grown in modified defined marine ammonium mineral salt (MAMS) media lacking ammonium and with HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) replacing the Pi buffer[71]. For all other growth and proteomics experiments these marine bacteria were grown in Sea Salts media[76], with Pi buffer replaced by HEPES buffer. The method used for quantifying extracellular Pi is detailed in the Supplementary Materials.

Generation and complementation of *Pseudomonas* mutants. Mutants were generated and complemented via the protocols outlined in[44,77], detailed descriptions of which are outlined in the Supplementary Materials. A full list of strains, plasmids, and primers used in this study is presented in Supplementary Table 4.

Proteomics preparation and analysis. To identify proteins involved in 2AEP uptake and catabolism in *S. stellulata*, total protein ($n = 3$ for each treatment), was retrieved by sampling cell cultures (OD_{540} 0.8–1.0) and pelleting cells (centrifugation at 16200 x g for 5 mins). Cell lysis was achieved via boiling in 100 µl lithium dodecyl sulphate (LDS) buffer (Expedeon) prior to loading 20 µl onto a 4-20% Bis-Tris sodium dodecyl sulphate (SDS) precast gel (Expedeon). For enrichment of the membrane protein fraction of the *P. putida* BIRD-1 Δ*aepXVW* mutant, we adapted the methods outlined in[44]. A full description of this protocol is outlined in the Supplementary Materials. Gel sections were de-stained with 50 mM ammonium bicarbonate in 50% (v/v) ethanol, dehydrated with 100% ethanol, reduced and alkylated with Tris-2-carboxyethylphosphine (TCEP) and iodoacetamide (IAA), washed with 50 mM ammonium bicarbonate in 50% (v/v) ethanol and dehydrated with 100% ethanol prior to overnight digestion with trypsin. Peptides were extracted and analysed using an Orbitrap Fusion Ultimate 3000 RSLCNano System (Thermo Scientific) in electrospray ionisation mode at the Warwick Proteomics Research Technology Platform.

Resulting tandem mass spectrometry (MS/MS) files were searched against the relevant protein sequence database (*P. putida* BIRD-1, NC_017530.1, *S. stellulata*, GCF_000423715.1) using MaxQuant[78] with default settings and quantification was achieved using Label Free Quantification (LFQ). The proteomics analysis software Perseus (1.6.12)[79] was used to identify differentially expressed proteins based on LFQ values, using a False Discovery Rate (FDR) of 0.01. Identified proteins were retained if they were present in at least two biological replicates within a treatment. Missing (N/A) values were imputed from a normal distribution using the default parameters. Differential expression was identified by two-sample Student's *t*-test, using an s0 constant of 0.1, or ANOVA, where appropriate.

Ligand binding affinity of recombinant *S. stellulata* AepX. Recombinant AepX was over-produced and purified to homogeneity using methods described in[53]. Binding affinity was determined by microscale thermophoresis using a Monolith NT.115 instrument (NanoTemper Technologies GmbT, Germany) following the protocols described in[53,54]. A detailed protocol is outlined in the Supplementary Materials.

Bioinformatics analyses of *aepXVW/aepP*. ORFs encoding AepX homologs were identified using the IMG/JGI (https://img.jgi.doe.gov/) 'Customised Homolog Display' search tool. Strains containing homologs were identified (cut-off values:

e^{-70}, min. identity 40%), preferentially from type strains. In addition, representative strains from soil and marine environments were added to this list (see Supplementary Table 2). Protein sequences were aligned using ClustalOmega[80] and profile Hidden Markov Models (pHMMs) were constructed from these sequences using the hmmbuild function of hmmer 3.3 (http://hmmer.org). The previously characterised *E. coli* K-12 PhnD[81] and the SAR11 clade isolate *Pelagibacter* sp. HTCC7211 PhnD[21] showed surprisingly low sequence homology (BLAST ID 28.46%, query coverage 76%, e-value 2e-25). We therefore, developed two pHMMS for PhnD to reflect this. There was no overlap in environmental sequences retrieved from each search using either pHMM. Therefore, abundance counts for each PhnD form were combined together as a collective PhnD group. These pHMMs were used to search the TARA ocean metagenome (OM-RGC_v2_metaG) and metatranscriptome (OM-RGC_v2_metaT) via the Ocean Gene Atlas web interface[55], using a stringency of $1E^{-80}$. Sequence abundances were expressed as the average percentage of genomes containing a copy by dividing the percentage of total mapped reads by the median abundance (as a percentage of total mapped reads) of 10 single-copy marker genes[82] for both MG and MT. The pHMMs were used to search the soil MG via a hmmsearch[83] using the same stringency as above. Similarly, abundances were calculated as the average percentage of genomes containing a copy as above.

Phylogenetic analyses were performed using IQ-TREE 2[84] using the following parameters: -m TEST -bb 1000 -alrt 1000. Evolutionary relationships were inferred by maximum-likelihood analysis, and visualised using the Interactive Tree of Life (iTOL) v5.6.3 online platform (https://itol.embl.de/)[85].

Statistical analysis. Unless specified above, all statistical analysis was performed using R (version 4.02)[86], within the RStudio programme (version 1.3).

Reproducibility. All growth experiments were performed a minimum of two times. Proteomics experiments were performed once.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The single-culture proteomic data generated in this study have been deposited in the ProteomeXchange Consortium via the PRoteomics IDEntifications (PRIDE) database under accession code PXD026804 [10.6019/PXD026804]. The TARA Oceans metagenomic and metatranscriptomic data used in this study are available in the European Nucleotide Archive database under accession code PRJEB7988.

Received: 10 December 2020; Accepted: 29 June 2021;
Published online: 27 July 2021

References

1. Villarreal-Chiu, J. F., Quinn, J. P. & McGrath, J. W. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Front Microbiol* **3**, 19 (2012).
2. Mukhamedova, K. S. & Glushenkova, A. I. Natural Phosphonolipids. *Chem. Nat. Compd.* **36**, 329–341 (2000).
3. Yu, X. et al. Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci. USA* **110**, 20759–20764 (2013).
4. Ju, K.-S., Doroghazi, J. R. & Metcalf, W. W. Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 345–356 (2014).
5. Peck, S. C. & van der Donk, W. A. Phosphonate biosynthesis and catabolism: a treasure trove of unusual enzymology. *Curr. Opin. Chem. Biol.* **17**, 580–588 (2013).
6. Metcalf, W. W. et al. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* **337**, 1104–1107 (2012).
7. Born, D. A. et al. Structural basis for methylphosphonate biosynthesis. *Science* **358**, 1336 (2017).
8. Hildebrand, R. The Role of Phosphonates in Living Systems. (CRC Press, 1983).
9. Van Mooy, B. A. S. et al. Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science* **348**, 783 (2015).
10. Kolowith, L. C., Ingall, E. D. & Benner, R. Composition and cycling of marine organic phosphorus. *Limnol. Oceanogr.* **46**, 309–320 (2001).
11. Clark, L. L., Ingall, E. D. & Benner, R. Marine phosphorus is selectively remineralized. *Nature* **393**, 426–426 (1998).
12. Turner, B. L., Baxter, R., Mahieu, N., Sjögersten, S. & Whitton, B. A. Phosphorus compounds in subarctic Fennoscandian soils at the mountain birch (*Betula pubescens*)—tundra ecotone. *Soil Biol. Biochem* **36**, 815–823 (2004).
13. Tate, K. R. & Newman, R. H. Phosphorus fractions of a climosequence of soils in New Zealand tussock grassland. *Soil Biol. Biochem* **14**, 191–196 (1982).
14. Cade-Menun, B. J., Navaratnam, J. A. & Walbridge, M. R. Characterizing dissolved and particulate phosphorus in water with 31P nuclear magnetic resonance spectroscopy. *Environ. Sci. Technol.* **40**, 7874–7880 (2006).
15. Clark, L. L., Ingall, E. D. & Benner, R. Marine organic phosphorus cycling; novel insights from nuclear magnetic resonance. *Am. J. Sci.* **299**, 724–737 (1999).
16. Repeta, D. J. et al. Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nat. Geosci.* **9**, 884–887 (2016).
17. Sosa, O. A. et al. Phosphonate cycling supports methane and ethylene supersaturation in the phosphate-depleted western North Atlantic Ocean. *Limnol. Oceanogr.* **65**, 2443–2459 (2020).
18. Dyhrman, S. T., Benitez-Nelson, C. R., Orchard, E. D., Haley, S. T. & Pellechia, P. J. A microbial source of phosphonates in oligotrophic marine systems. *Nat. Geosci.* **2**, 696–699 (2009).
19. Acker, M. et al. Phosphonate production by marine microbes: exploring new sources and potential function. bioRxiv, 2020.2011.2004.368217, https://doi.org/10.1101/2020.11.04.368217 (2020).
20. Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D. & Karl, D. M. Phosphate-limited ocean regions select for bacterial populations enriched in the carbon-phosphorus lyase pathway for phosphonate degradation. *Environ. Microbiol* **21**, 2402–2414 (2019).
21. Carini, P., White, A. E., Campbell, E. O. & Giovannoni, S. J. Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nat. Comms* **5**, 4346 (2014).
22. Chin, J. P., Quinn, J. P. & McGrath, J. W. Phosphate insensitive aminophosphonate mineralisation within oceanic nutrient cycles. *ISME J.* **12**, 973–980 (2018).
23. Ternan, N. G. & Quinn, J. P. Phosphate starvation-independent 2-aminoethylphosphonic acid biodegradation in a newly isolated strain of *Pseudomonas putida*, NG2. *Syst. Appl Microbiol* **21**, 346–352 (1998).
24. White, A. K. & Metcalf, W. W. Microbial metabolism of reduced phosphorus compounds. *Annu Rev. Microbiol* **61**, 379–400 (2007).
25. Martinez, A., Tyson, G. W. & Delong, E. F. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environ. Microbiol* **12**, 222–238 (2010).
26. Wanner, B. L. Molecular genetics of carbon-phosphorus bond cleavage in bacteria. *Biodegradation* **5**, 175–184 (1994).
27. Kononova, S. V. & Nesmeyanova, M. A. Phosphonates and their degradation by microorganisms. *Biochem. (Mosc.)* **67**, 184–195 (2002).
28. Jiang, W., Metcalf, W. W., Lee, K. S. & Wanner, B. L. Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of *Salmonella typhimurium* LT2. *J. Bacteriol.* **177**, 6411 (1995).
29. Kim, A. D. et al. The 2-aminoethylphosphonate-specific transaminase of the 2-aminoethylphosphonate degradation pathway. *J. Bacteriol.* **184**, 4134 (2002).
30. Cooley, N. A. et al. Phosphonoacetate biosynthesis: In vitro detection of a novel $NADP^+$-dependent phosphonoacetaldehyde-oxidizing activity in cell-extracts of a marine Roseobacter. *Microbiology* **80**, 335–340 (2011).
31. Borisova, S. A. et al. Genetic and biochemical characterization of a pathway for the degradation of 2-aminoethylphosphonate in *Sinorhizobium meliloti* 1021. *J. Biol. Chem.* **286**, 22283–22290 (2011).
32. Agarwal, V. et al. Structure and function of phosphonoacetaldehyde dehydrogenase: The missing link in phosphonoacetate formation. *Chem. Biol.* **21**, 125–135 (2014).
33. McSorley, F. R. et al. PhnY and PhnZ comprise a new oxidative pathway for enzymatic cleavage of a carbon-phosphorus bond. *J. Am. Chem. Soc.* **134**, 8364–8367 (2012).
34. van Staalduinen, L. M. et al. Crystal structure of PhnZ in complex with substrate reveals a di-iron oxygenase mechanism for catabolism of organophosphonates. *Proc. Natl Acad. Sci. USA* **111**, 5171–5176 (2014).
35. Sosa, O. A., Casey, J. R. & Karl, D. M. Methylphosphonate oxidation in *Prochlorococcus* strain MIT9301 supports phosphate acquisition, formate excretion, and carbon assimilation into purines. *Appl Environ. Microbiol* **85**, e00289–e00219 (2019).
36. Gama, S. R. et al. An oxidative pathway for microbial utilization of methylphosphonic acid as a phosphate source. *ACS Chem. Biol.* **14**, 735–741 (2019).
37. Zangelmi, E. et al. Discovery of a new, recurrent enzyme in bacterial phosphonate degradation: (*R*)-1-hydroxy-2-aminoethylphosphonate ammonia-lyase. *Biochemistry* **60**, 1214–1225 (2021).
38. Sowell, S. M. et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3**, 93–105 (2008).

39. Ottesen, E. A. et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl Acad. Sci. USA* **110**, E488–E497 (2013).
40. Lidbury, I., Murrell, J. C. & Chen, Y. Trimethylamine *N*-oxide metabolism by abundant marine heterotrophic bacteria. *Proc. Natl Acad. Sci. USA* **111**, 2710–2715 (2014).
41. Mauchline, T. H. et al. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc. Natl Acad. Sci. USA* **103**, 17933–17938 (2006).
42. Alicea, I. et al. Structure of the *Escherichia coli* phosphonate binding protein PhnD and rationally optimized phosphonate biosensors. *J. Mol. Biol.* **414**, 356–369 (2011).
43. Rizk, S. S., Cuneo, M. J. & Hellinga, H. W. Identification of cognate ligands for the *Escherichia coli phnD* protein product and engineering of a reagentless fluorescent biosensor for phosphonates. *Protein Sci.* **15**, 1745–1751 (2006).
44. Lidbury, I. D. et al. Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. *Environ. Microbiol* **18**, 3535–3549 (2016).
45. Lemieux, M. J., Huang, Y. & Wang, D. N. Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from *E. coli*. *J. Electron Microsc* **54**, i43–i46 (2005). Suppl 1.
46. Elvin, C. M., Hardy, C. M. & Rosenberg, H. Pi exchange mediated by the GlpT-dependent sn-glycerol-3-phosphate transport system in *Escherichia coli*. *J. Bacteriol.* **161**, 1054–1058 (1985).
47. Enkavi, G. & Tajkhorshid, E. Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT. *Biochemistry* **49**, 1105–1114 (2010).
48. Law, C. J. et al. Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *J. Mol. Biol.* **378**, 828–839 (2008).
49. Law, C. J., Enkavi, G., Wang, D.-N. & Tajkhorshid, E. Structural basis of substrate selectivity in the glycerol-3-phosphate: phosphate antiporter GlpT. *Biophysical J.* **97**, 1346–1353 (2009).
50. Moradi, M., Enkavi, G. & Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nat. Comms* **6**, 8393 (2015).
51. Kulakova, A. N. et al. Structural and functional analysis of the phosphonoacetate hydrolase (*phnA*) gene region in *Pseudomonas fluorescens* 23F. *J. Bacteriol.* **183**, 3268 (2001).
52. Kulakova, A. N. et al. Expression of the phosphonoalanine-degradative gene cluster from *Variovorax* sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate. *FEMS Microbiol Lett.* **292**, 100–106 (2009).
53. Bisson, C. et al. The molecular basis of phosphite and hypophosphite recognition by ABC-transporters. *Nat. Comms* **8**, 1746 (2017).
54. Adams, N. B. P., Robertson, A. J., Hunter, C. N., Hitchcock, A. & Bisson, C. Phosphite binding by the HtxB periplasmic binding protein depends on the protonation state of the ligand. *Sci. Rep.* **9**, 10231–10231 (2019).
55. Villar, E. et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res* **46**, W289–W295 (2018).
56. Rajakovich, L. J. et al. A new microbial pathway for organophosphonate degradation catalyzed by two previously misannotated non-heme-iron oxygenases. *Biochemistry* **58**, 1627–1647 (2019).
57. Smith, A. F. et al. Elucidation of glutamine lipid biosynthesis in marine bacteria reveals its importance under phosphorus deplete growth in *Rhodobacteraceae*. *ISME J.* **13**, 39–49 (2019).
58. Chin, J. P., McGrath, J. W. & Quinn, J. P. Microbial transformations in phosphonate biosynthesis and catabolism, and their importance in nutrient cycling. *Curr. Opin. Chem. Biol.* **31**, 50–57 (2016).
59. White, A. K. & Metcalf, W. W. Two C-P lyase operons in *Pseudomonas stutzeri* and their roles in the oxidation of phosphonates, phosphite, and hypophosphite. *J. Bacteriol.* **186**, 4730–4739 (2004).
60. Ermakova, I. T. et al. Organophosphonates utilization by soil strains of *Ochrobactrum anthropi* and *Achromobacter* sp. *Arch. Microbiol* **199**, 665–675 (2017).
61. Hartley, L. E., Kaakoush, N. O., Ford, J. L., Korolik, V. & Mendz, G. L. Characterisation of *Campylobacter jejuni* genes potentially involved in phosphonate degradation. *Gut Pathog.* **1**, 13 (2009).
62. Imazu, K. et al. Enhanced utilization of phosphonate and phosphite by *Klebsiella aerogenes*. *Appl Environ. Microbiol* **64**, 3754–3758 (1998).
63. Mendz, G. L., Mégraud, F. & Korolik, V. Phosphonate catabolism by *Campylobacter* spp. *Arch. Microbiol* **183**, 113–120 (2005).
64. Metcalf, W. W. & Wanner, B. L. Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*. *Gene* **129**, 27–32 (1993).
65. Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **521**, 208–212 (2015).
66. Frischkorn, K. R. et al. *Trichodesmium* physiological ecology and phosphate reduction in the western tropical South Pacific. *Biogeosciences* **15**, 5761–5778 (2018).
67. Feingersch, R. et al. Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *ISME J.* **6**, 827–834 (2012).
68. Young, C. L. & Ingall, E. D. Marine dissolved organic phosphorus composition: insights from samples recovered using combined electrodialysis/reverse osmosis. *Aquat. Geochem* **16**, 563–574 (2010).
69. Benitez-Nelson, C. R., O'Neill, L., Kolowith, L. C., Pellechia, P. & Thunell, R. Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin. *Limnol. Oceanogr.* **49**, 1593–1604 (2004).
70. Lidbury, I. D. E. A., Murrell, J. C. & Chen, Y. Trimethylamine and trimethylamine *N*-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling. *ISME J.* **9**, 760–769 (2015).
71. Lidbury, I., Kimberley, G., Scanlan, D. J., Murrell, J. C. & Chen, Y. Comparative genomics and mutagenesis analyses of choline metabolism in the marine *Roseobacter* clade. *Environ. Microbiol* **17**, 5048–5062 (2015).
72. Zecher, K., Hayes, K. R. & Philipp, B. Evidence of interdomain ammonium cross-feeding from methylamine- and glycine betaine-degrading *Rhodobacteraceae* to diatoms as a widespread interaction in the marine phycosphere. *Front Microbiol* **11**, 2431 (2020).
73. Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J. & Scanlan, D. J. Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nat. Microbiol* **2**, 17100 (2017).
74. Dyhrman, S. T. et al. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* **439**, 68–71 (2006).
75. Karl, D. M. et al. Aerobic production of methane in the sea. *Nat. Geosci.* **1**, 473–478 (2008).
76. Chen, Y. Comparative genomics of methylated amine utilization by marine Roseobacter clade bacteria and development of functional gene markers (*tmm*, *gmaS*). *Environ. Microbiol* **14**, 2308–2322 (2012).
77. Lidbury, I. D. E. A. et al. Identification of extracellular glycerophosphodiesterases in *Pseudomonas* and their role in soil organic phosphorus remineralisation. *Sci. Rep.* **7**, 2179 (2017).
78. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
79. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
80. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636–W641 (2019).
81. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29–W37 (2011).
82. Wackett, L. P., Wanner, B. L., Venditti, C. P. & Walsh, C. T. Involvement of the phosphate regulon and the *psiD* locus in carbon-phosphorus lyase activity of *Escherichia coli* K-12. *J. Bacteriol.* **169**, 1753–1756 (1987).
83. Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Comms* **10**, 1014 (2019).
84. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
85. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256–W259 (2019).
86. R. Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2020).

Acknowledgements

We thank the Warwick Proteomics Research Facility, namely Dr. Cleidiane Zampronio for her assistance in generating and processing the mass-spectrometry data. This study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under project codes BB/L026074/1, BB/T009152/1 and NE/S013539/1 linked to The Soil and Rhizosphere Interactions for Sustainable Agri-ecosystems (SARISA) programme and a Discovery Fellowship (IL) and NERC Environmental 'Omics Synthesis Grant (IL and EW), respectively.

Author contributions

A.M. and I.L. wrote the manuscript with comprehensive feedback and guidance from Y.C. and D.S. A.M. performed the experimental work, and A.M. and I.L. analysed the data, except for the binding affinity assays, performed and analysed by N.A., W.C., A.H. and I.L. I.L. and A.M. conceptualised and developed the research, respectively. A.B. provided input and support for proteomic data generation and analyses. All authors contributed to revisions of the manuscript.

Competing interests

The authors declare no competing interests.

101

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-24646-z.

Correspondence and requests for materials should be addressed to I.D.E.A.L.

Peer review information *Nature Communications* thanks P. Dreux Chappell, Juan Villarreal Chiu, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

The ISME Journal (2021) 15:2440–2453
https://doi.org/10.1038/s41396-021-00933-x

ARTICLE

A novel class of sulfur-containing aminolipids widespread in marine roseobacters

Alastair F. Smith[1] · Eleonora Silvano[1] · Orsola Päuker[2] · Richard Guillonneau[1] · Mussa Quareshy[1] · Andrew Murphy[1] · Michaela A. Mausz[1] · Rachel Stirrup[1] · Branko Rihtman[1] · Maria Aguilo-Ferretjans[1] · Joost Brandsma[3] · Jörn Petersen[2] · David J. Scanlan[1] · Yin Chen[1]

Received: 11 November 2020 / Revised: 7 February 2021 / Accepted: 8 February 2021 / Published online: 9 March 2021
© The Author(s) 2021. This article is published with open access

Abstract

Marine roseobacter group bacteria are numerically abundant and ecologically important players in ocean ecosystems. These bacteria are capable of modifying their membrane lipid composition in response to environmental change. Remarkably, a variety of lipids are produced in these bacteria, including phosphorus-containing glycerophospholipids and several amino acid-containing aminolipids such as ornithine lipids and glutamine lipids. Here, we present the identification and characterization of a novel sulfur-containing aminolipid (SAL) in roseobacters. Using high resolution accurate mass spectrometry, a SAL was found in the lipid extract of *Ruegeria pomeroyi* DSS-3 and *Phaeobacter inhibens* DSM 17395. Using comparative genomics, transposon mutagenesis and targeted gene knockout, we identified a gene encoding a putative lyso-lipid acyltransferase, designated *salA*, which is essential for the biosynthesis of this SAL. Multiple sequence analysis and structural modeling suggest that SalA is a novel member of the lysophosphatidic acid acyltransferase (LPAAT) family, the prototype of which is the PlsC acyltransferase responsible for the biosynthesis of the phospholipid phosphatidic acid. SAL appears to play a key role in biofilm formation in roseobacters. *salA* is widely distributed in *Tara* Oceans metagenomes and actively expressed in *Tara* Oceans metatranscriptomes. Our results raise the importance of sulfur-containing membrane aminolipids in marine bacteria.

Introduction

Bacterial lipids are the key constituent segregating cellular components from the external environment. Bacterial lipids are highly diverse yet there is currently little understanding of the benefits that this diversity provides [1]. Glycerophospholipids are by far the best studied lipids in bacteria and a key branch point for glycerophospholipid biosynthesis is phosphatidic acid (PA), from which a variety of lipids, including phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylcholine (PC), diacylglycerol (DAG), and triacylglycerol (TAG), can be made through either the cytidine diphosphate (CDP)-diacylglycerol (DAG) pathway or the Kennedy pathway [2]. PA biosynthesis in bacteria is carried out by a membrane-attached acyltransferase PlsC, the founding member of the large lysophosphatidic acid acyltransferase (LPAAT) family [3, 4].

Aside from phospholipids, the study of bacterial lipid diversity is currently hampered by a lack of knowledge of both the chemical structures of many of these lipids and the identity of genes involved in their synthesis. These have severely hindered our understanding of lipid diversity and their physiological function in bacteria. Once the chemical structure of a lipid is known, analytical strategies can then be devised to detect the lipid in both the natural environment and cell cultures [5]. This can also help to direct studies into the biosynthesis of the lipid, knowledge of which can provide a clearer idea of the likely distribution of the lipid

These authors contributed equally: Alastair F. Smith, Eleonora Silvano

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41396-021-00933-x.

✉ Yin Chen
y.chen.25@warwick.ac.uk

[1] School of Life Sciences, University of Warwick, Coventry, UK

[2] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

[3] Southampton General Hospital, University of Southampton, Southampton, UK

amongst various bacterial classes. A group of poorly studied bacterial lipids are the aminolipids, of which only ornithine lipids have been detected in diverse cultured bacteria since the 1960s [6]. However, it was not until the genes involved in its biosynthesis were elucidated that it became clear how widespread the capacity to produce ornithine lipid really was [7, 8]. Similarly, Sebastian et al. [9] found several uncharacterized aminolipids in marine heterotrophic bacteria one of which was recently determined as a glutamine-containing aminolipid, often found in the marine roseobacter group [10]. Both ornithine and glutamine lipids play a key role in the adaptation of cosmopolitan marine bacteria (e.g., the marine SAR11 clade and the roseobacter group) to oligotrophic environments [9–11].

In this study, we report the characterization and chemical structure of a novel sulfur-containing aminolipid using high resolution-accurate mass spectrometry from the marine roseobacter group. This newly identified lipid represents a novel class of sulfur-containing lipids with an aminosulfonate head group. Furthermore, we describe a novel acyltransferase enzyme (SalA), part of the LPAAT family, that is responsible for the biosynthesis of this sulfonolipid. This sulfonolipid appears widespread within the roseobacter group that are key players in marine biogeochemical cycles and important for biofilm formation. Furthermore, the *salA* gene is abundant and actively transcribed in marine surface microbial assemblages.

Materials and methods

Bacterial strains and cultivation

All marine bacteria used in this study were cultivated using either marine broth medium (BD Difco™ 2216), ½YTSS medium containing yeast extract (2 g/L), tryptone (1.25 g/L), and sea salts (20 g/L, Sigma-Aldrich) or a defined marine ammonium mineral salts (MAMS) medium [10]. The MAMS medium contained 30 g/L NaCl, 10 mM glucose, 1 mM K_2HPO_4, 0.75–7.5 mM NH_4Cl, 10 mM HEPES buffer (pH 7.6), 1.36 mM $CaCl_2$, 0.98 mM $MgSO_4$, 7.2 µM $FeCl_2$, 84 µM Na_2MoO_4, 370 nM $ZnCl_2$, 510 nM $MnCl_2$, 97 nM H_3BO_3, 1.1 µM $CoCl_2$, 12 nM $CuCl_2$, 100 nM $NiCl_2$, 30 nM thiamine, 160 nM nicotinic acid, 97 nM pyridoxine, 73 nM aminobenzoic acid, 53 nM riboflavin, 84 nM pantothenate, 4.1 nM biotin, 1.5 nM cyanocobalamin, and 11 nM folic acid. All cultures were grown at 30 °C aerobically in a shaker (150 r.p.m) unless stated otherwise.

Intact polar lipid analysis

Lipid extraction from bacterial cultures was carried out using the modified Folch extraction protocol as described previously [10, 12]. Briefly 1 mL culture of OD_{540} ~ 1.0 was collected by centrifugation. Total lipids were then extracted using methanol-chloroform, dried under nitrogen gas and the pellet re-suspended in 1 mL solvent (95% (v/v) liquid chromatography-mass spectrometry (LC-MS) grade acetonitrile and 5% 10 mM ammonium acetate pH 9.2 in water). These lipids were then analysed by LC-MS using a Dionex 3400RS HPLC with a HILIC BEH amide XP column (2.5 µm, 3.0 × 150 mm, Waters) coupled with an amaZon SL ion trap MS (Bruker) via electrospray ionization (ESI) in both positive (+ve) and negative (−ve) ionization mode. Samples were run on a 15 min gradient from 95% (v/v) acetonitrile/5% (w/v) ammonium acetate (in water, 10 mM, pH 9.2) to 70% (v/v) acetonitrile/30% (w/v) ammonium acetate (in water, 10 mM, pH 9.2), followed by 5 min of isocratic run 70% acetonitrile/30% ammonium acetate with 10 min equilibration between samples. The flow rate was maintained at 150 µL min^{-1} and the column temperature at 30 °C. The injection volume was 5 µL for each run; the ionization was done in both positive and negative mode. Drying conditions were the same for both modes (8 L min^{-1} drying gas at 300 °C and nebulizing gas pressure of 15 psi). The end cap voltage was 4500 V in positive mode and 3500 V in negative mode, both with 500 V offset. Data analysis was carried out using the Bruker Compass software package. Unless stated otherwise, base peak chromatographs were presented with *m/z* range from 400 to 1000.

High resolution MS identification and fragmentation was carried out using either a quadrupole-time-of-flight MS (Q-TOF, Waters Synapt G2-Si) or an Orbitrap Fusion (Thermo Fisher Scientific) by direct infusion and collision induced dissociation (CID). For the Orbitrap Fusion, the resolution was set at 120 K with CID for MS^n. A TriVersa Nanomate nanospray source (Advion, NY) was used and the flow rate was at 300 nL min^{-1}. The voltage was set at 1.4 kV and the gas pressure was 0.3 psi. Sheath and sweep gas were set to zero and the cone voltage was 2100 V and the mass range was from 50 to 1000 Da. MS data were analyzed using Xcalibur (Thermo Fisher Scientific). For Q-TOF, samples were injected through a Universal NanoFlow Sprayer (Waters) by direct infusion at 200–300 nL min^{-1} and the cone voltage was 30 V in negative mode ESI. Mass range was set from 50 to 1000 Da and data analyses ware carried out in MassLynx (Waters). The most abundant peak in the negative ion spectrum corresponding to the SAL lipid (*m/z* 656.6) was selected for MS^n fragmentation. Spectra were obtained in profile mode and smoothed using a moving mean. Background correction using a linear baseline was applied with a 40% noise cut-off. For accurate mass determination, the centroid of each peak was used. The peak corresponding to $C17H33COO^-$ (*m/z* 281.2480, an 18:1 fatty acid carboxylate anion) was used as a lock mass. Calculation of candidate elemental formulae from the

accurate mass considered formulae containing C0–100, H0-100, N0-100, S0-4, and P0-1. A conservative mass error of 100 ppm was assumed.

Marker-exchange mutagenesis

Marker-exchange mutagenesis was carried out as described previously using a suicide vector pK18*mobsacB* [10]. Briefly, DNA fragments corresponding to an upstream element and a downstream element that flank the target gene were amplified by PCR using high-fidelity *Phusion* DNA polymerase. A Gm-resistance cassette was amplified from plasmid p34S-Gm [10, 13]. These fragments, together with the linearized pK18*mobsacB* vector were then assembled through Gibson cloning and transformed into competent *Escherichia coli* DH5α cells. The engineered suicide vector was then extracted from *E. coli* DH5α and transformed into the conjugation donor strain *E. coli* S17.1 *λpir* before conjugating into *Ruegeria pomeroyi* DSS-3 as described previously [10]. Transconjugants were then selected on defined MAMS medium containing gentamycin (Gm, 10 μg mL^{-1}). All mutants were confirmed by PCR using the confirmation primers (Supplementary Table 1) and subsequent Sanger sequencing.

Transposon library of *Phaeobacter inhibens* DSM 17395

A library of 5500 transposon mutants of *Phaeobacter inhibens* DSM 17395, which was established at the DSMZ, served as a basis to identify genes involved in the biosynthesis of the novel SAL lipid. Transposon mutagenesis was performed with the EZ-Tn5<R6Kγori/Kan-2> Tnp Transposome kit (Epicentre, Illumina, CA, USA) and the insertion site of all mutants was determined via arbitrary PCR [14]. Transposon mutants were streaked out three times to eliminate attached wild type cells. The absence of wild type cells and the presence of the 65 kb plasmid were validated as described previously [14–16]. The transposon integration site of each mutant was also confirmed via sequencing of the amplification PCR product, and stable maintenance of all three extrachromosomal elements was validated via diagnostic PCR [17].

The transposon mutant #1036 of *P. inhibens* DSM 17395 (PGA1_c01210) unable to produce the SAL lipid was complemented using the *salA* homolog of *Ruegeria pomeroyi* DSS-3 (locus tag SPO0716) and *P. inhibens* DSM 17395 (locus tag PGA1_c01210). Complementation was carried out by PCR amplification of the *salA* homologs together with a constitutive promoter (~250 bp upstream of the *aacC1* gene from plasmid p34S-GM [13]), which was then cloned into the broad host range vector pBBR1MCS and transformed into the *salA* mutant of *P. inhibens* DSM 17395 by conjugation as described previously [9, 10]. The complemented mutants were cultivated using marine broth medium and cells were harvested for lipidomics analysis as described above.

Biofilm assays

To grow biofilms of *Phaeobacter inhibens* DSM 17395 and the *salA* mutant, post-exponential grown bacterial cells were washed and diluted in fresh marine broth medium and inoculated at an $OD_{590\ nm}$ of 0.2 into 24-well plates (Corning Incorporated Costar®, New York, NY, USA) containing a sterilized glass coverslip into each well. At each time point (3, 24, and 48 h), biofilms were washed to remove non-adherent bacteria and fixed using formalin 3.5% (v/v) for 20 min. Bacteria were stained using DAPI (5 μg• mL^{-1}, Sigma-Aldrich, Darmstadt, Germany) and coverslips were mounted with a drop of Mowiol antifade before observation using confocal laser scanning microscopy (CLSM) (Zeiss LSM 880, Göttingen, Germany). The biovolume and the average thickness of the biofilms were determined using COMSTAT software developed in MATLAB R2017a (MathWorks, Natick, MA, United States) as described previously [18, 19]. To test for statistically significant differences between the wild-type strain and the *salA* mutant, a *t*-test was performed using SPSS 13.0 (IBM, Armonk, NY, USA).

A crystal violet biofilm assay was also performed which was adapted from Guillonneau et al. [18]. Bacterial biofilms were developed in 96-well microtiter plates (Greiner Bio-One, Kremsmünster, Austria) with bacteria in the post-exponential growth phase using marine broth medium. Cells were diluted to a final $OD_{590\ nm} = 0.1$ into each well ($n = 4$ for both the wild type and the *salA* mutant) and grown in static conditions at 30 °C. At each time point (3, 24, 48, and 72 h) samples were washed three times with fresh marine broth medium and dried for 30 min at 50 °C. Biofilms were then stained for 15 min with 200 μL crystal violet 0.01% (w/v), rinsed three times with phosphate-buffered saline and dried for 10 min. Biofilm quantification was performed by releasing the stain from the biofilm using absolute ethanol for 10 min at 30 °C with gentle shaking. The absorbance of the crystal violet in solution was measured at 595 nm. The final absorbance of each sample was calculated by subtracting the blank (i.e., marine broth medium only treated with crystal violet, $n = 4$).

Bioinformatics analysis

Phylogenetic analysis of 16S rRNA genes from *Rhodobacteraceae* was carried out using the full length 16S rRNA gene retrieved from the Integrated Microbial Genomes (IMG) database (https://img.jgi.doe.gov/). Sequence

alignment of 16S rRNA genes and LPAAT genes (also retrieved from IMG) were performed using Muscle and phylogenetic analyses were performed with MEGA7.0 [20] with 500 bootstrap replicates. Sequence alignment was visualized using JalView [21].

To search for SalA homologs in the *Tara* metagenome/metatranscriptomics datasets, we used the Ocean Gene Atlas (OGA) database OM_RGCv2_metaG (metagenomics) and OM-RGCv2_metaT (metatranscriptomics) with *e*-value cut-off of e^{-40} [22]. Abundance was normalized as a percentage of the median mapped read abundance of genes/transcripts of ten prokaryotic single-copy marker genes [23]. Taxonomic distribution of homologs was displayed using Krona in the OGA interface.

The genomes of the marine roseobacters used in this study were downloaded from the NCBI database. These comprised nine strains that were found to produce SAL and two strains (*Stappia stellulata* DSM 5886 and *Dinoroseobacter shibae* DFL12) that did not. In order to identify genes potentially involved in SAL synthesis, each gene from the 11 genomes was assigned to an orthologous group using the eggNOG mapper [24]. This program conducts a BLAST search of each sequence against the eggNOG database [25] of orthologous genes, with the query sequence being annotated with the same orthologous group as the best BLAST hit. Orthologous groups that were present in the genomes of all SAL-producing strains but absent from the genomes of *S. stellulata* and *D. shibae* were considered to be potentially involved in SAL synthesis.

Abundance data of SalA homologs from four depths derived from the *Tara* metagenomics/metatranscriptomics datasets were tested for normal distribution using a Shapiro–Wilks test. Significant differences between depths was tested for using a Kruskal–Wallis test followed by a post-hoc Dunn's test using Holm's correction for multiple comparisons. All statistical analysis was performed in RStudio (version 1.3) using R (version 4.02).

In silico homology modeling and docking studies for SalA

A SalA homology model was generated using the Phyre2 protein folding prediction server [26], and the lyso-SAL lipid was drawn in MarvinSketch (v19.10.0, 2019, ChemAxon for Mac) and exported as a Mol SDF format file. The homology model was built using the structure of the lysophosphatidic acid acyltransferase PlsC (PDB code 5KYM [4]). The SalA protein model was then imported into Flare (v3.0, Cresset) for docking the lyso-SAL substrate and energy minimized with 2000 iterations with a cut off of 0.200 kcal/mol/A. The lyso lipid was imported as a ligand and energy minimized in Flare before being docked into the active site and the best scoring pose selected.

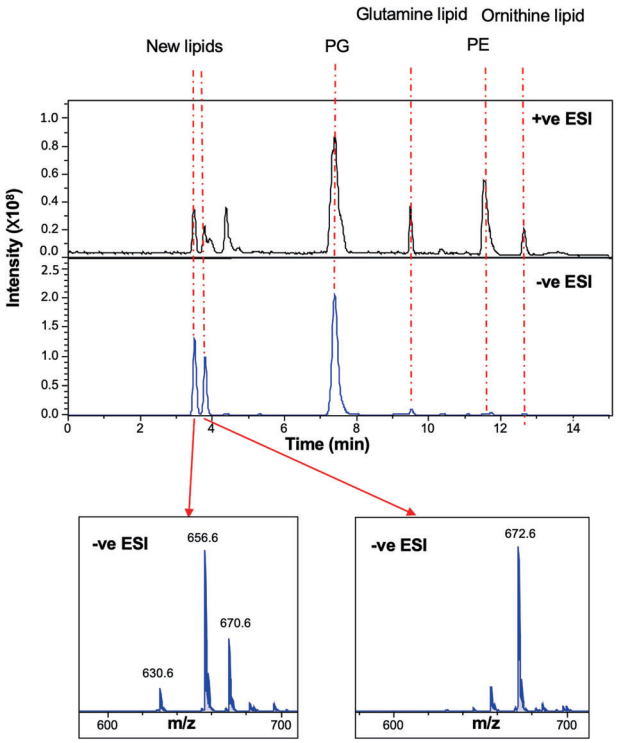


Fig. 1 Lipid profiles of *Ruegeria pomeroyi* DSS-3 cultivated in 1/2 YTSS medium. MS spectra were obtained in both positive (+ve) and negative (−ve) ionization mode using electrospray ionization (ESI) using an amaZon SL ion trap MS (Bruker). PG phosphatidylglycerol, PE phosphatidylethanolamine. New lipids with *m/z* of 656.6 and 672.6 in −ve ESI were eluted between 3 and 4 min from the liquid chromatography column.

Results

A new sulfur-containing aminolipid is found in *Ruegeria pomeroyi* DSS-3

During LC-MS analysis of lipid extracts from *Ruegeria pomeroyi* DSS-3 grown on ½ YTSS medium, two prominent peaks eluting around 3.5 min were found in both negative and positive ionization mode (Fig. 1). The most prominent ions in the two peaks had *m/z* values of 656.6 and 672.7 in the negative ionization mode, respectively. Other major lipids identified in this bacterium include two phospholipids, PG and PE and two aminolipids, ornithine lipid (OL) and glutamine lipid (QL) [10].

To elucidate the structure of the new lipids eluted at 3.5 min, the most intense species, at 656.4882 *m/z*, was selected for high resolution MS/MS analysis on a quadrupole-time of flight (Q-TOF) mass spectrometer (Fig. 2). At low collision energy (40 eV) the major species formed corresponded to a neutral loss of 282 mass units. This is consistent with the neutral loss of an 18:1 fatty acid. A second peak at *m/z* 281.2480 is likely the carboxylate anion of an 18:1 fatty acid. Further fragmentation, at higher

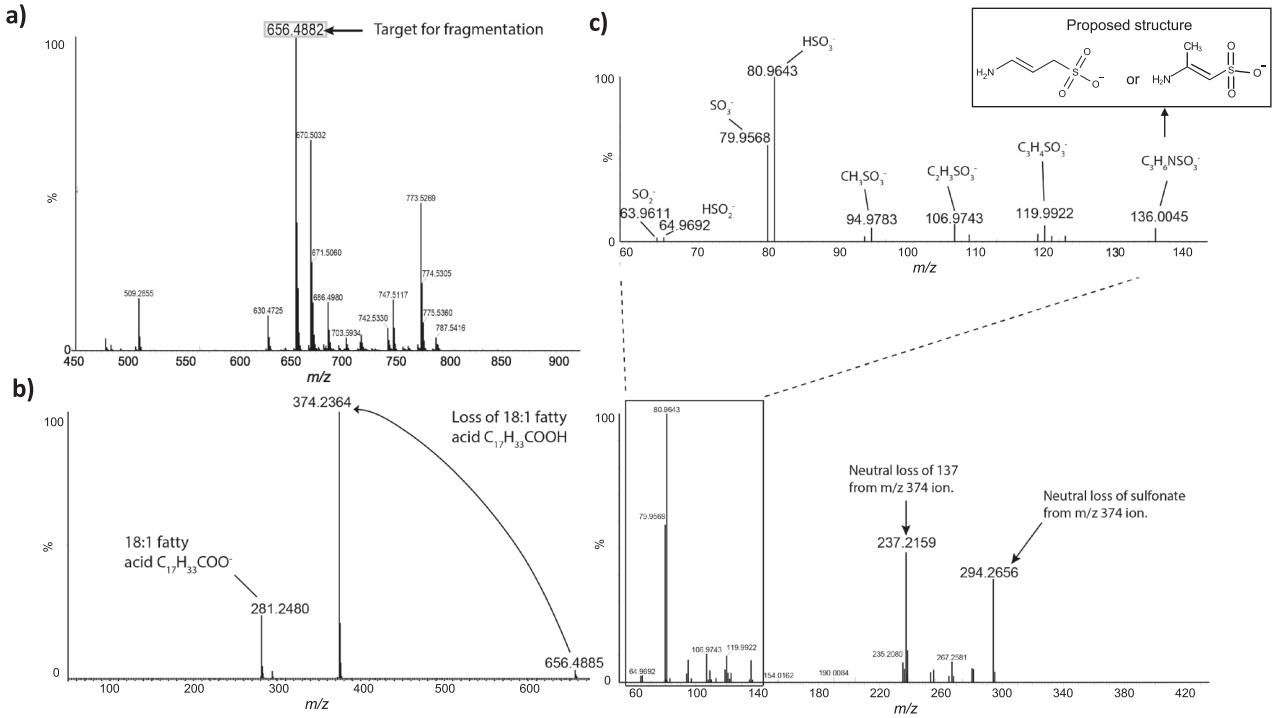


Fig. 2 A previously unidentified class of sulfur-containing aminolipids (SAL) is present in *Ruegeria pomeroyi* DSS-3. a Intact masses of *R. pomeroyi* lipids, measured using a high resolution, accurate mass quadrupole-time of flight (Q-TOF) mass spectrometer (Waters Synapt G2-Si) in negative ionization mode. The identity of the most abundant ion, highlighted, was unknown, so this ion was fragmented in order to elucidate its structure. **b** Fragmentation spectrum of *m/z* 656.4882 ion at 40 eV collision energy (MS^2). The most abundant species corresponds to the loss of a 18:1 fatty acid. **c** Fragmentation spectrum of *m/z* 374.2364 ion at 90 eV collision energy (MS^3). The lower spectrum shows a peak at 237.2159, consistent with the presence of a 16:0 fatty acid. The upper spectrum shows an expanded view of the spectrum in the *m/z* range below 140. Ions are annotated with their predicted elemental composition. Inset is the proposed structure of the 136.0045 *m/z* fragment (3-aminopropane sulfonic acid or 2-aminopropane sulfonic acid, respectively). *Ruegeria pomeroyi* DSS-3 was cultivated in 1/2 YTSS medium.

collision energies (up to 90 eV), yielded a major ion at *m/z* 237.2159. This ion likely corresponds to a 16:0 fatty acid present as a ketene, which would be consistent with the fragmentation scheme proposed for ornithine lipids and glutamine lipids [27]. These results therefore suggest a lipid class with a similar fatty acyl backbone structure to the aminolipids, such as ornithine and glutamine lipid [10]. The glutamine lipid (QL, $[M+H]^+$ *m/z* 719.7) and ornithine lipid (OL, $[M+H]^+$ *m/z* 705.7) was eluted at 9.5 and 12.5 min, respectively (Fig. 1). The formation of these novel lipids at ~3.5–4 min is not affected in the *olsA* or *glsB* mutants of *R. pomeroyi* DSS-3 (Supplementary Fig. S1). The *olsA* and *glsB* genes in *R. pomeroyi* DSS-3 were essential for the production of the nitrogen-containing ornithine/glutamine lipids [10].

Prominent peaks at 80 and 81 *m/z*, respectively, were apparent in the fragmentation spectrum obtained at 90 eV collision energy (Fig. 2c). The accurate masses of these ions were 79.9568 and 80.9643. Of the candidate formulae within 100 ppm of the measured mass, SO_3^- and HSO_3^- appear most plausible, with mass errors of 0.182 ppm and 4.194 ppm, respectively. A smaller peak doublet at *m/z* 63.9611 and 64.9692 was also present in the 90 eV spectrum. These masses are unambiguously assigned to SO_2^- (mass error 12.506 ppm) and HSO_2^- (mass error 8.08 ppm). Taken together, these results demonstrate the presence of a sulfonate group in the lipid. An ion at 136.0045 *m/z* corresponded to the deprotonated head group. The mass determined here is larger than that of deprotonated taurine (*m/z* 124). Since the head group includes a sulfonate (SO_3^-) group, the plausible formula most closely corresponding to the accurate mass is $C_3H_6NSO_3$ (Table 1). This is consistent with the structure being aminopropane sulfonic acid, although the position of the amino group cannot be

Table 1 Nominal and accurate masses of proposed head group fragments, with mass errors implied by the proposed formulae.

Nominal *m/z*	Accurate *m/z*	Formula	Mass error/ppm
136	136.0045	$C_3H_6NSO_3$	12.784
120	119.9922	$C_3H_4SO_3$	34.048
107	106.9743	$C_2H_3SO_3$	51.315
95	94.9783	CH_3SO_3	16.737

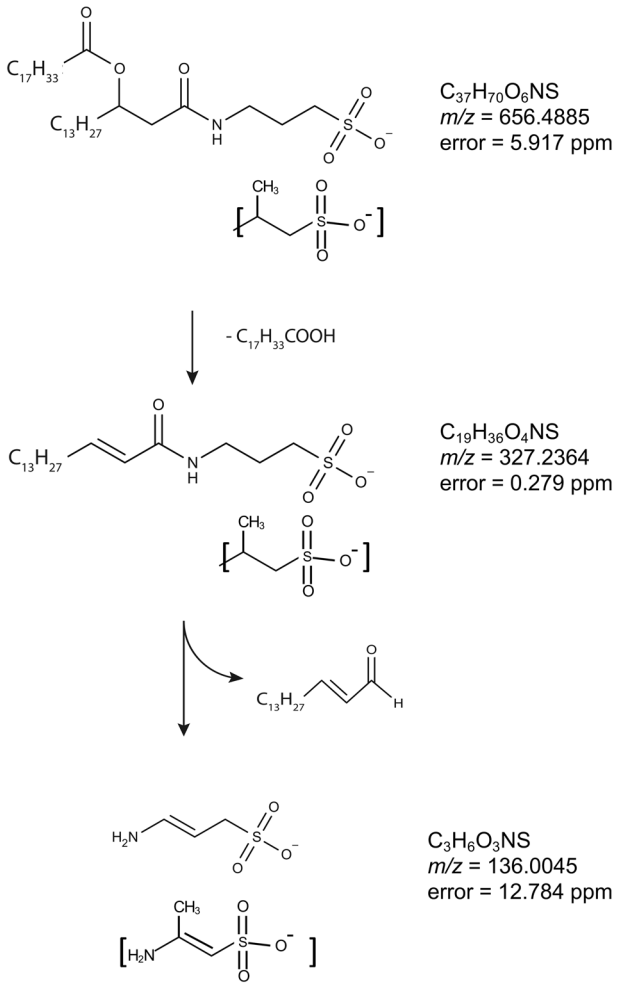


Fig. 3 Proposed fragmentation scheme for the SAL lipid with *m/z* 656.4885. Mass errors indicate the difference between the measured and theoretical masses of the proposed species. The aminolipid head group is consisted of an aminopropane sulfonic acid of 3-aminopropane sulfonic acid (a.k.a. homotaurine) or 2-aminopropane sulfonic acid (as indicated by square brackets).

unequivocally determined by mass spectrometry (Fig. 2). The proposed fragmentation scheme is presented in Fig. 3.

To further confirm the presence of an amino-group in the hydrophilic head of this SAL, we cultivated *Ruegeria pomeroyi* DSS-3 in a chemically defined marine ammonium mineral salts (MAMS) medium using ^{15}N-ammonium as the sole nitrogen source. Indeed, the ^{15}N-labeled SAL was readily observed in the lipid extract resulting in a shift of *m/z* from 656.4951 to 657.4903 (Supplementary Fig. S2a), whereas the non-nitrogen containing lipids, such as PG were not labeled by ^{15}N as expected (Supplementary Fig. S2b). The incorporation of the ^{15}N isotope into the head group of SAL was confirmed by MS^n (Supplementary Fig. S2c, d). We also performed the same MS^n analysis on the *m/z* 672.4875 species as well as the ^{15}N-labeled *m/z* 673.4852 species. Loss of 282 at MS^2 (672.4875→390.2317; 673.4852→391.2285) suggests the R_2 fatty acid was C18:1. Therefore, the data suggest that the lipid species eluted immediately after the *m/z* 656.6 species is likely a hydroxylated SAL, and the proposed fragmentation scheme is presented in Supplementary Fig. S3.

The sulfur-containing aminolipid is found in a range of marine roseobacters

To investigate the presence of SAL amongst roseobacters we selected 16 strains, in addition to *R. pomeroyi* DSS-3, to obtain a wide coverage of the roseobacter group including the model roseobacter bacterium *Phaeobacter inhibens* DSM 17395 (Fig. 4). The selected strains included *Stappia stellulata*, which recent phylogenetic studies indicate is not a member of the *Rhodobacteraceae* [28], which served as an outgroup. These strains were each grown in marine broth overnight, before cells were harvested for lipid analysis. SAL was detected in all the strains tested apart from *S. stellulata* and *Dinoroseobacter shibae* (Fig. 4a). The separation of these two strains from the remaining roseobacter sequences is in line with previous results showing *D. shibae* branching deeply within the *Rhodobacteraceae* phylogeny [29].

Comparative genomics to determine genes involved in SAL biosynthesis

We then conducted a comparative genomics investigation into the roseobacter strains whose lipid profiles had been analysed. We reasoned that synthesis of the SAL would require an *N*-acyltransferase activity to acylate aminopropane sulfonic acid, analogous to that mediated by OlsB and GlsB in the synthesis of ornithine and glutamine lipid [8, 10]. We investigated predicted *N*-acyltransferases that were present in all the strains that produced SAL in marine broth (the "producers"), while being absent from the strains that did not produce SAL (the "non-producers"). We assigned all the genomic sequences from the nine genome-sequenced producer strains and two non-producer strains to orthologous groups (OGs) using the eggNOG-mapper software [24], which provides a consistent pipeline for sequence annotation and OG assignment by comparison to the eggNOG database [24]. We identified a group of 1417 "core" genes which were present in the genomes of all SAL producer strains of which 1060 were also present in the two non-producers (Fig. 4b). Thirty-seven candidate genes are present in all SAL producer strains but not in the genomes of the non-producers (Fig. 4b), two of which (OG accession numbers 08UX5 and 05CDD) were annotated as being potential acyltransferases (Table 2). We therefore generated mutants in these two genes in the two model bacteria, *R. pomeroyi* DSS-3 and *P. inhibens* DSM 17395 and screened for the loss of SAL production. The 08UX5 mutant (locus SPO2471) of *R. pomeroyi* DSS-3 still produced SAL to the same level as the wild type (data not shown), suggesting that this gene is unlikely involved in

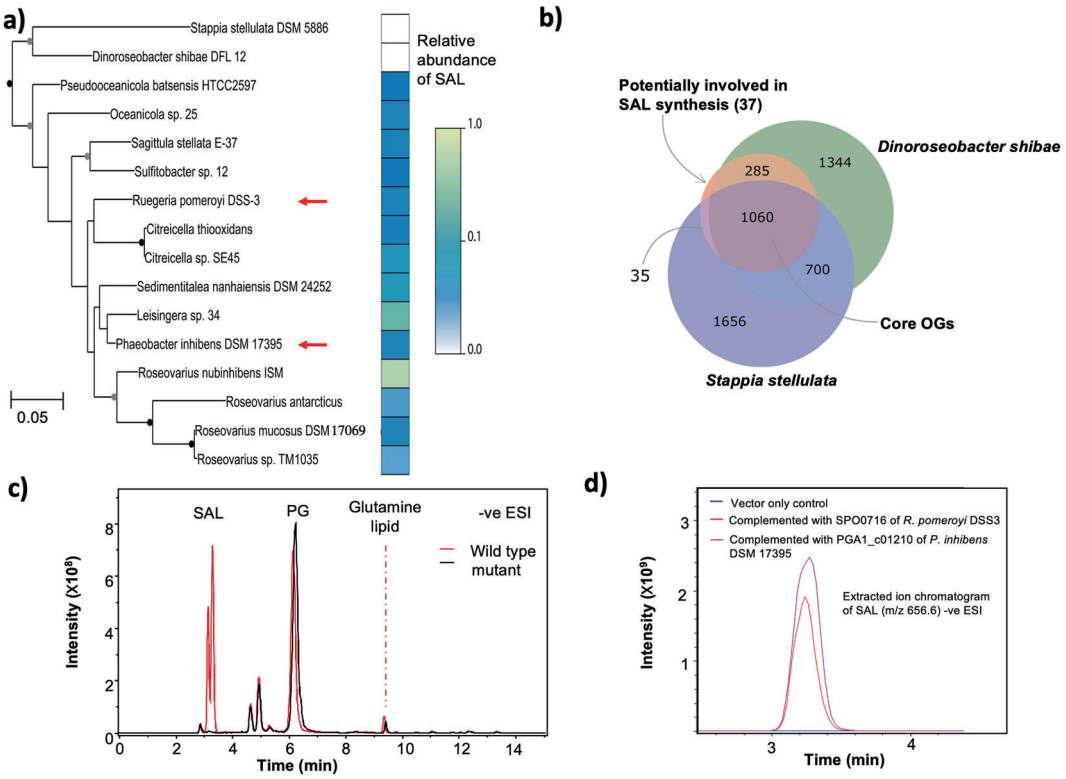


Fig. 4 The distribution of SALs amongst *Rhodobacteraceae* and identification of the *salA* gene. a Phylogeny of 16S rRNA gene sequences from selected *Rhodobacteraceae* plotted alongside information on the abundance of the SAL. Nodes with more than 50% bootstrap support are indicated with circles (gray: 50–69%; black: >70%). The red arrows point to the two bacteria which were used as models to investigate SAL biosynthesis. The *Stappia stellulata* 16S rRNA gene sequence was used as an outgroup. The abundance of SAL measured in each strain, as a proportion of the maximum abundance measured, is plotted alongside the phylogeny. One ml culture of OD_{540} ~ 1.0 was collected by centrifugation before lipid extraction, and abundances were calculated as the peak area of SAL species relative to a phosphatidylglycerol internal standard (0.25 µM diheptadecanoyl phosphatidylglycerol, C17/C17 PG). **b** Identification of two candidate acyltransferases to be investigated for involvement in SAL biosynthesis. The Venn diagram illustrates the logic used to identify the candidate genes. The central circle represents the 1417 orthologous groups (OGs) common to all SAL-producing strains. **c** LC-MS chromatogram showing the lipid profile of wild type and the *salA* mutant of *Phaeobacter inhibens* DSM 17395. The absence of SAL lipids in the mutant is evident in the black trace. **d** Extracted ion chromatogram (EIC) showing the complementation of the *salA* mutant of *Phaeobacter inhibens* DSM 17395 by either *salA* of *R. pomeroyi* DSS3 (SPO0716) or *P. inhibens* DSM 17395 (PGA1_c01210), which restores the production of SAL (*m/z* 656.6). The empty vector control is shown in blue. Bacterial cells were cultivated in marine broth medium. Ions were analysed in negative (−ve) ionization mode using electrospray ionization (ESI) using an amaZon SL ion trap MS (Bruker).

SAL formation. However, in the 05CDD mutant of *P. inhibens* DSM 17395 (locus tag PGA1_c01210), SAL formation is completely abolished, suggesting that this gene is indeed responsible for SAL biosynthesis (Fig. 4c). This gene is named *salA* hereafter. Indeed, when the mutant was complemented with either *salA* from *R. pomeroyi* DSS-3 (SPO0716) or *P. inhibens* DSM 17395 (PGA1_c01210), SAL production was restored (Fig. 4d).

SalA is a putative O-acetyltransferase-like protein with a recognized LPAAT (lysophosphatidic acid acyltransferase) domain. Amongst bacterial LPAAT-domain containing proteins, the best characterized examples are PlsC and OlsA, encoding enzymes responsible for the final step in the biosynthesis of the anionic phospholipid phosphatidic acid (PA) and the ornithine/glutamine-containing aminolipid, respectively [3, 10, 30]. The structure of PlsC has recently been solved, showing an in silico docked LPA lipid together with the fatty acid in an acyl carrier protein (ACP, [4]). Multiple sequence alignments of SalA, PlsC, and OlsA shows the presence of two conserved sequence motifs (Fig. 5), representing the catalytic center ($HX_{4/5}D$) and the substrate co-ordination center (FP[E/S]G[T/V]), respectively. Notably, both PlsC and OlsA have the conserved HX_4D motif whereas SalA has the HX_5D motif. Interestingly, the reported key Lys105 in PlsC, thought to be responsible for electrostatic interactions via its amide nitrogen backbone to the negatively-charged oxygen of the ACP-fatty acid intermediate, was replaced with Arg135 in SalA. The LPA phosphate head group is thought to be coordinated by Arg159 in PlsC. However, the sequence alignment shows a Val189 in SalA. In order to further investigate the implications of the sequence alignment, we obtained a homology model of SalA. The

Table 2 Candidate eggNOG orthologous groups (OGs) that are unique in the genomes of SAL-producers.

OG	Locus tag	COG category	eggNOG annotation
07TNN	SPO0365	S	Domain of unknown function (DUF1989)
05CDD	**SPO0716**	**S**	**Phospholipid glycerol acyltransferase**
01VSN	SPO0838	M	Bacterial sugar transferase
05MEE	SPO0968	L	Nudix hydrolase
05K39	SPO1109	S	Endonuclease exonuclease phosphatase
05NU0	SPO1287	S	Glyoxalase bleomycin resistance protein dioxygenase
08EFD	SPO1424	P	Sodium hydrogen exchanger
08K1F	SPO1437	I, Q	Dehydrogenase
05D82	SPO1743	E	Glutamate dehydrogenase
05S99	SPO1919	P	Tellurite resistance protein
090P8	SPO1963	G	Phosphoglycerate mutase
05N59	SPO2284	M	Transglycosylase
01QFF	SPO2332	P	Acriflavin resistance protein
08YCB	SPO2344	E	Sarcosine oxidase, gamma subunit
08UX5	**SPO2471**	**S**	**Acetyltransferase, (GNAT) family**
01QH6	SPO2697	C	CoA-binding domain protein
07QRG	SPO2741	L	DNA polymerase iii
05X3J	SPO2754	S	Uncharacterized protein
060TV	SPO2757	S	EF hand domain protein
01TDR	SPO2815	P	ABC transporter permease protein
08TCB	SPO3054	S	Uncharacterized protein
0808N	SPO3431	S	Domain of unknown function (DUF3576)
08RQB	SPO3439	I	Enoyl-coA hydratase isomerase family protein
08R76	SPO3444	M	3-deoxy-d-manno-octulosonic-acid transferase
07RFB	SPO3740	O	Catalyzes the reversible oxidation-reduction of methionine sulfoxide in proteins to methionine
6405	SPO0725	S	SH3 type 3
01RKE	SPO2816	P	ABC transporter permease protein
05DM9	SPO0035	S	Core-2 I-branching enzyme family protein
0627S	SPO_RS14095	S	Uncharacterized protein
07SF7	SPO2814	E	Peptide opine nickel uptake family ABC transporter periplasmic substrate-binding protein
05SH6	SPO3502	S	DUF3118 protein
08U9J	SPO3819	S	Type I secretion target repeat protein
05WTG	SPO2776	S	Uncharacterized protein
05Q6E	SPO3716	S	Uncharacterized protein
08T58	SPO3836	S	Mg^{2+} transport protein CorA
064YK	SPO1241	S	Uncharacterized protein
05U51	SPO0776	S	Thioesterase

The two predicted acyltransferases are highlighted in bold.

model shows the catalytic HX_5D motif to be structurally comparable to that of PlsC despite the additional residue, with the His109 and Asp115 adjacent to each other, analogous to that in PlsC (Supplementary Fig. S4). In silico docking of a lyso-SAL lipid molecule into the model demonstrated a possible pose for the lyso-lipid hydroxy group adjacent to His109 (Supplementary Fig. S4), with the Arg135 suggested to coordinate the sulfonate head group. The conformationally flexible alkyl chain group was able to adopt many configurations, but the polar head group was docked consistently in the same region. Overall, the data suggests a diversification in function of LPAAT family enzymes during evolution, with SalA representing a novel member of this group. The presence of this unique motif of HX_5D in SalA allowed us to determine the distribution of SAL-biosynthesis in environmental metagenomes and metatranscriptomes (see below).

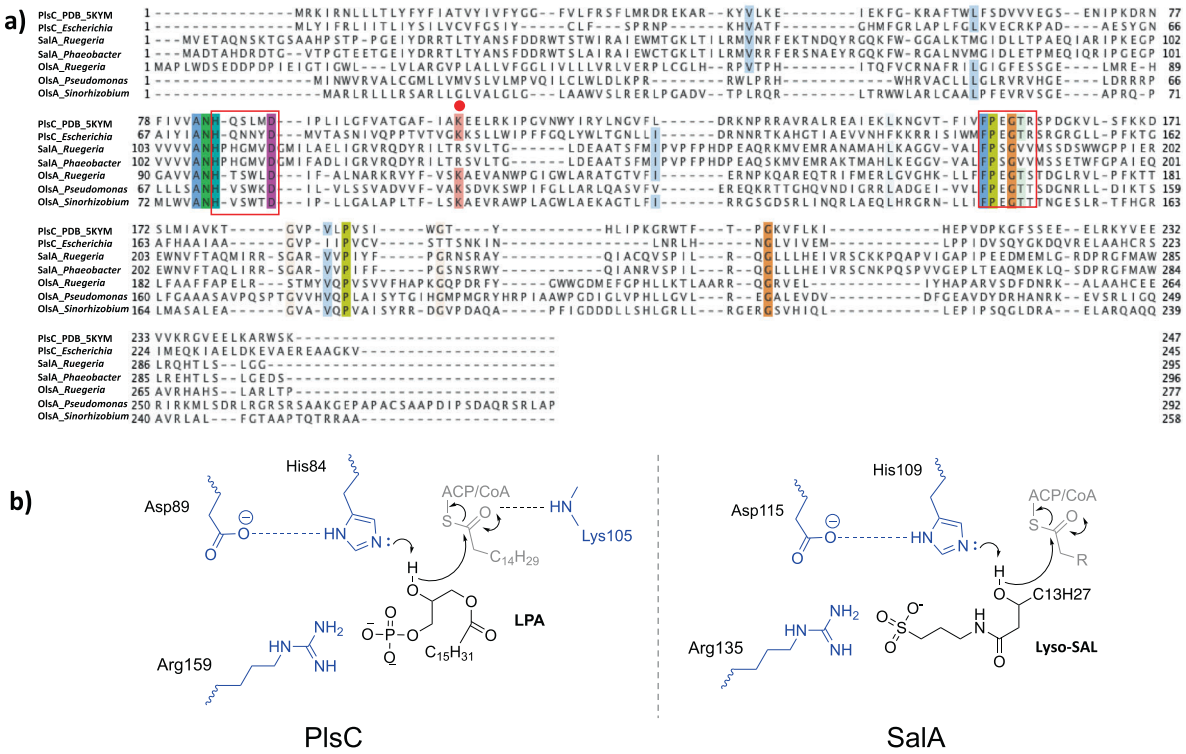


Fig. 5 SalA represents a new member of the lysophosphatidic acid acyltransferase (LPTAA) family. a Multiple sequence alignment of SalA from *Ruegeria pomeroyi* DSS-3 and *Phaeobacter inhibens* DSM 17395, lysophosphatidic acid acyltransferase PlsC from *Escherichia coli* and *Thermotoga maritima* (PDB code 5KYM [4]), and OlsA involved in ornithine lipid biosynthesis from *R. pomeroyi* DSS-3 (SPO1979 [10]), *Pseudomonas aeruginosa* (PO4351 [48]), and *Sinorhizobium meliloti* [44]. The two red boxes highlight the catalytic center (HX_nD) and the substrate binding site (FPXGXX), respectively. The red dot represents the key lysine 105 involved in catalysis in PlsC. **b** Proposed residues involved in PlsC and SalA catalysis. LPA lysophosphatidic acid.

SAL production in *Phaeobacter inhibens* DSM 17395 is involved in biofilm formation

We next investigated the role of SAL lipids in the physiology of roseobacters. The loss of SAL lipids had no clear role in the growth of the bacterium. Both wild type and the *salA* mutant of *Phaeobacter inhibens* DSM 17395 had comparable growth rates and reached similar final cell density in marine broth medium (Supplementary Fig. S5a). An important change of lifestyle for roseobacters is the switch from planktonic growth to biofilm formation, which triggers a particle-associated life strategy that is ecologically relevant for their survival in the natural environment [31]. It has been shown previously that many roseobacters including *Phaeobacter inhibens* DSM 17935 are able to form a biofilm, and a 65 kb plasmid in this bacterium was important for biofilm formation [15, 16]. Interestingly, we observed that the *salA* mutant has a significantly reduced ability to form biofilms when in contact with solid surfaces, such as glass (Fig. 6) and plastics (Supplementary Fig. S5b). Both the bioviolume on the glass surface as well as the thickness of the biofilm are significantly reduced in the *salA* mutant strain in the early phase of biofilm formation (3 h), and the latter stage (24 and 48 h) of biofilm maturation (Fig. 6). The 65 kb biofilm plasmid was confirmed to be present in the *salA* mutant (Supplementary Fig. S5c). Thus, the significant reduced ability of the *salA* mutant in biofilm formation suggests that this lipid may play a key role in roseobacters in their natural environment.

Distribution of the new acetyltransferase SalA in the *Tara* Ocean metagenomes and metatranscriptomes

To better understand the distribution of SAL in environmental microbial assemblages, we searched the *Tara* Ocean metagenomes and metatranscriptomics datasets using SalA (locus tag, SPO0716 of *R. pomeroyi* DSS-3) as the query. We experimentally determined the *e* value cut-off to be e^{-40} at which value it selectively retrieves LPAAT homologs belonging to SalA but not OlsA or PlsC. The environmental SalA homologs obtained from the *Tara* Oceans metagenome and metatranscriptomics dataset were aligned, and the key sequence motifs were manually examined. In particular, the HX_5D motif is strictly conserved in all SalA sequences retrieved from the *Tara* Oceans datasets providing strong support, that these environmental sequences are of the SalA but not PlsC nor OlsA clade. On average, between 2–4% of microbial cells are estimated to have the potential for SAL

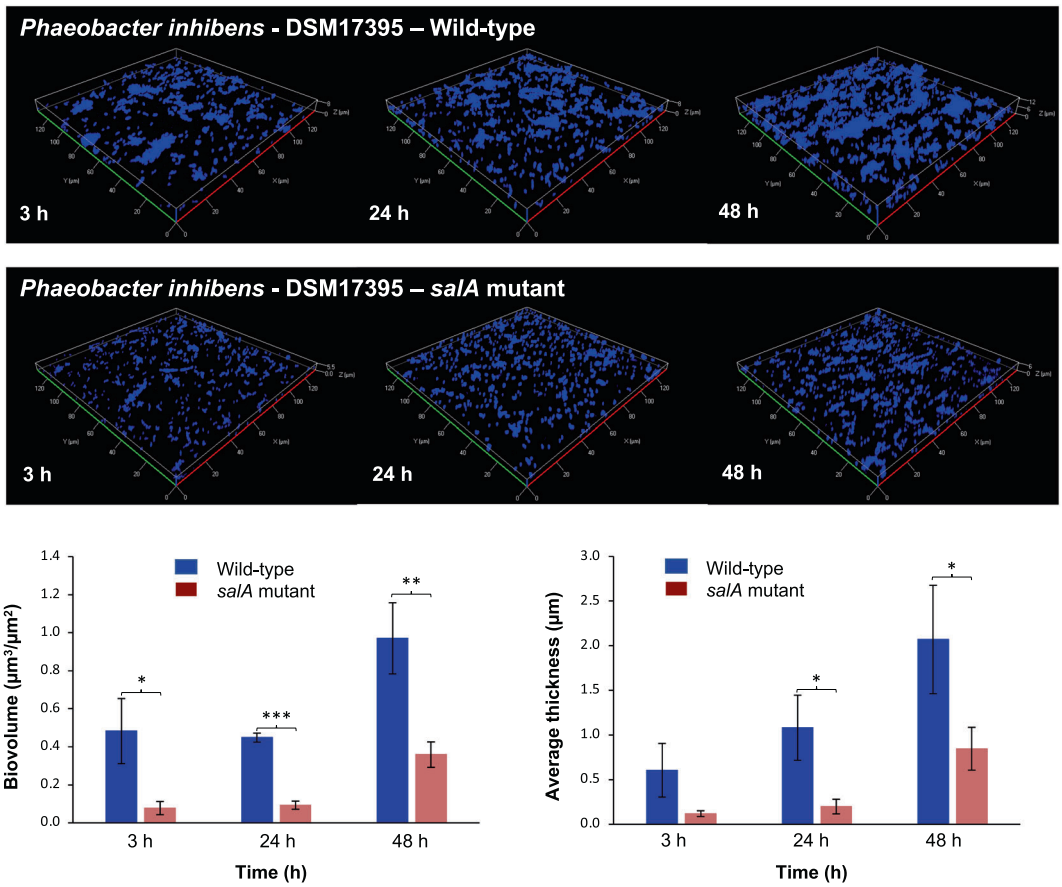


Fig. 6 Quantification of biofilm formation in *Phaeobacter inhibens* DSM 17395 wild type and the *salA* mutant. Biofilm was developed for 3 h, 24 h, and 48 h respectively. The biovolume and the thickness of the biofilm over time were quantified. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Bacterial cells were cultivated in marine broth medium.

biosynthesis; this is comparable to that of the *olsA* gene but somewhat lower than the *plcP* gene in the same dataset, suggesting SAL biosynthesis is less prevalant than the PlcP-mediated lipid remodeling pathway [9, 10]. This is likely due to the fact that SALs are primarily found in marine roseobacters but not in other dominant marine *Alphaproteobacteria*, such as the abundant bacterium *Pelagibacter ubique* of the SAR11 clade which are capable of PlcP-mediated lipid remodeling [9, 11]. Indeed, the majority (>85%) of the SalA sequences from the *Tara* Oceans dataset were classified as members of the *Rhodobacteraceae* in both *Tara* Oceans metagenomes and metatranscriptomes (Fig. 7), and a thorough search of 120-genome sequenced *Rhodobacteraceae* confirmed the wide occurrence of *salA* in all ten clades of the roseobacters (Supplementary Fig. S6 [32]).

Discussion

Here, we identify a novel aminolipid containing an aminopropane sulfonic acid head group that is widespread amongst marine roseobacters. The presence of a sulfonate group means this SAL lipid also falls into the broad category of sulfonolipids. The most abundant, and arguably one of the best studied lipids of this type, is sulfoquinovosyl diacylglycerol (SQDG), which is present in the membranes of most oxygenic phototrophs [33] as well as some heterotrophic bacteria [34]. SQDG likely plays a structural role in photosynthetic membranes, since crystal structures of photosystem proteins show specific binding of this lipid [35].

Other sulfolipids appear to elicit potent responses when certain organisms are exposed to them. Thus, a sulfolipid produced by zooplankton from a number of copepod species was found to induce toxin production in the dinoflagellate *Alexandrium minutum* [36], likely as a defense against predation. Conversely, a sulfonolipid produced by the *Bacteroidetes* bacterium *Algoriphagus machipongonensis* induced the development of multicellularity in a choanoflagellate [37]. Both examples suggest that sulfolipids are used by the sensing organism as a marker for the presence of another organism with which it interacts (either as a predator or as a symbiont). The fact that sulfolipids

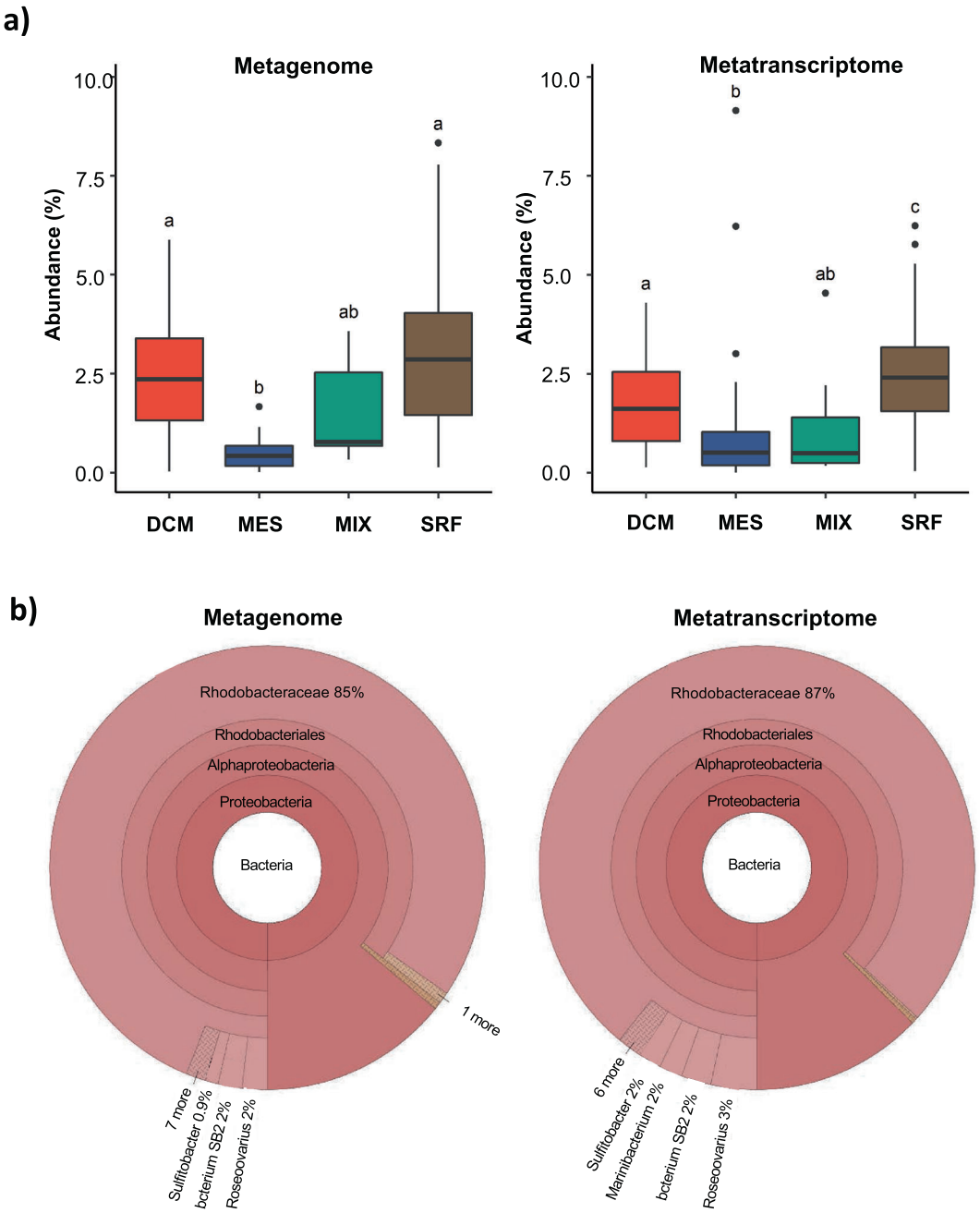


Fig. 7 The distribution of the *salA* gene in the *Tara* oceans database. a The Ocean Gene Atlas (OGA) database was searched using *salA* of *Ruegeria pomeroyi* DSS-3 (SPO0716) with an *e*-value cut-off of e^{-40} and the abundance was calculated as a percentage of median abundance of ten prokaryotic single-copy marker genes/transcripts [22]. **b** The taxonomic distribution of homologs was displayed using Krona in the OGA interphase. DCM deep chlorophyll maximum, SRF surface water, MES mesopelagic zone, MIX mixed layer. More than 85% of the SalA homologs are classified as marine roseobacters (*Rhodobacteraceae*) in metagenomes (left panel) and metatranscriptomes (right panel).

appear to be relatively rare across the tree of life likely makes them well suited to mediate such chemical interactions, where a high degree of specificity is required. Lipids similar to those produced by *A. machipongonensis* have been described in a number of *Bacteroidetes*, particularly amongst *Cytophaga* [38–40]. They tend to be localized to the outer membrane, and seem to play a role in the gliding motility of these organisms [41, 42]. The sulfonolipids from *Bacteroidetes* differ from those that we describe here in roseobacters in that they are composed of a base, termed capnine, similar to the sphingoid bases of sphingolipids, which may be *N*-acylated to form the full sulfonolipid [38]. In this way they are similar structurally to sphingolipids, whereas the SALs of the roseobacter group are more similar

to aminolipids such as ornithine lipid and glutamine lipid (Fig. 5). Whether the SAL lipid plays a role in interspecies interactions requires further work. However, we already observed that this lipid is involved in biofilm formation in *Phaeobacter inhibens* DMS17395 (Fig. 6), suggesting that formation of this SAL lipid may play an important role in the adaptation of marine roseobacters to a biofilm lifestyle.

A survey of the distribution of SAL among isolates from the roseobacters indicated that the ability to produce this lipid is widely distributed within the group. One strain, *D. shibae*, taxonomically the most basal of the strains examined, lacked any SAL under the conditions assessed, as did the outgroup strain *Stappia stellulata*. The absence of SAL in these strains suggests they lack the capacity to produce this lipid as the other roseobacters examined seem to produce SAL constitutively. However, it is possible that these strains have the capacity to produce SAL, but only do so under certain conditions. This pattern is observed for ornithine lipid, which is produced constitutively in some bacteria, such as *R. pomeroyi* DSS-3 [10], but in others is only produced as a response to P-depletion [11, 43]. Indeed, a close *salA* homolog was found in the genome of *D. shibae* (Dshi_0206), but it is absent in *S. stellulata*.

Although we have identified the LPAAT enzyme, SalA, involved in the last step of synthesis of this new sulfur-containing aminolipid, the key steps and genes involved in the synthesis of the lyso-SAL lipid remain to be determined. It is likely that SAL synthesis occurs in a manner analogous to that of ornithine and glutamine lipids. As such, 3-hydroxy fatty acids would be required as a substrate for the first step in SAL synthesis [44]. Such a hypothesis suggests that the aminopropane sulfonic acid moiety is also directly produced by the marine roseobacters since no exogenous supply was provided. The presence of 3-aminopropane sulfonic acid (a.k.a. homotaurine) has been documented in some red algae [45, 46] and unicellular green algae (prasinophytes such as *Ostreococcus* and *Micromonas*, [47]) but, to the best or our knowledge, never previously in bacteria. However, a hydroxylated form of 2-aminopropane sulfonic acid, cysteinolic acid, has been found in a variety of marine phytoplankton and heterotrophic bacteria, including *Ruegeria pomeroyi* DSS-3 although its biosynthetic pathway remains to be established [47]. Nevertheless, it is tempting to speculate that 2-aminopropane sulfonic acid is likely the hydrophilic head of the new SAL observed in these marine roseobacters, and this certainly warrants further investigation.

To sum up, this study describes a new class of lipid, which are an important component of the membranes of a number of marine *Rhodobacteraceae*. Comparative genomics of SAL-producing strains has identified a novel acyltransferase (SalA), which is involved in the production of this lipid. *salA* is widely distributed in marine microbial assemblages in the Oceans and actively expressed in *Tara* Oceans metatranscriptomes, and its functional role in addition to biofilm formation in these marine bacteria certainly warrants further investigation.

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 726116) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 34509606—TRR 51. We also thank the Natural Environment Research Council, UK for funding the PhD studentship awarded to AFS. We acknowledge Dr I. Hands-Portman and the Imaging Suite at the School of Life Sciences, University of Warwick for providing assistance with the use of confocal microscopy. We are grateful to Prof. J.D. Todd, Dr H. Schäfer, Dr A. Curson, Prof. M. Simon, and Dr H.-A. Giebel for many insightful discussions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Sohlenkamp C, Geiger O. Bacterial membrane lipids: diversity in structures and pathways. FEMS Microbiol Rev. 2016;40:133–59.
2. Parsons JB, Rock CO. Bacterial lipids: metabolism and membrane homeostasis. Prog Lipid Res. 2013;52:249–76.
3. Rottig A, Steinbuchel A. Acyltransferases in bacteria. Microbiol Mol Biol Rev. 2013;77:277–321.
4. Robertson RM, Yao J, Gajewski S, Kumar G, Martin EW, Rock CO, et al. A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. Nat Stru Mol Biol. 2017;24:666–71.
5. Moore EK, Hopmans EC, Rijpstra WI, Villanueva L, Damsté JS. Elucidation and identification of amino acid containing membrane lipids using liquid chromatography/high-resolution mass spectrometry. Rapid Commun Mass Spectrom. 2016;30:739–50.
6. Shively JM, Knoche HW. Isolation of an ornithine-containing lipid from *Thiobacillus thiooxidans*. J Bacteriol. 1969;98:829–30.
7. Vences-Guzmán MÁ, Guan Z, Escobedo-Hinojosa WI, Bermúdez-Barrientos JR, Geiger O, Sohlenkamp C. Discovery of a bifunctional acyltransferase responsible for ornithine lipid synthesis in *Serratia proteamaculans*. Environ Microbiol. 2015;17:1487–96.

8. Gao J-L, Weissenmayer B, Taylor AM, Thomas-Oates J, López IM, Geiger O. Identification of a gene required for the formation of lyso-ornithine lipid, an intermediate in the biosynthesis of ornithine-containing lipids. Mol Microbiol. 2004;53:1757–70.
9. Sebastián M, Smith AF, González JM, Fredricks HF, Van Mooy B, Koblížek M, et al. Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. ISME J. 2016;10:968–78.
10. Smith AF, Rihtman B, Stirrup R, Silvano E, Mausz MA, Scanlan DJ, et al. Elucidation of glutamine lipid biosynthesis in marine bacteria reveals its importance under phosphorus deplete growth in *Rhodobacteraceae*. ISME J. 2019;13:39–49.
11. Carini P, Van Mooy BAS, Thrash JC, White AE, Zhao Y, Campbell EO, et al. SAR11 lipid renovation in response to phosphate starvation. Proc Natl Acad Sci USA. 2015;112:7767–72.
12. Silvano E, Yang M, Wolterink M, Giebel H-A, Simon M, Scanlan DJ, et al. Lipidomic analysis of roseobacters of the pelagic RCA cluster and their response to phosphorus limitation. Front Microbiol. 2020;11:552135.
13. Dennis JJ, Zylstra GJ. Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes. Appl Environ Microbiol. 1998;64:2710–5.
14. Segev E, Wyche TP, Kim KH, Petersen J, Ellebrandt C, Vlamakis H, et al. Dynamic metabolic exchange governs a marine algal-bacterial interaction. eLife . 2016;5:pii: e17473. https://doi.org/10.7554/eLife.17473.
15. Michael V, Frank O, Bartling P, Scheuner C, Göker M, Brinkmann H, et al. Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters. ISME J. 2016;10:2498–513.
16. Frank O, Michael V, Päuker O, Boedeker C, Jogler C, Rohde M, et al. Plasmid curing and the loss of grip–the 65-kb replicon of *Phaeobacter inhibens* DSM 17395 is required for biofilm formation, motility and the colonization of marine algae. Syst Appl Microbiol. 2015;38:120–7.
17. Trautwein K, Will SE, Hulsch R, Maschmann U, Wiegmann K, Hensler M, et al. Native plasmids restrict growth of *Phaeobacter inhibens* DSM 17395: energetic costs of plasmids assessed by quantitative physiological analyses. Environ Microbiol. 2016;18:4817–29.
18. Guillonneau R, Baraquet C, Bazire A, Molmeret M. Multispecies biofilm development of marine bacteria implies complex relationships through competition and synergy and modification of matrix components. Front Microbiol. 2018;9:1960.
19. Heydorn A, Nielsen AT, Hentzer M, Sternberg C, Givskov M, Ersbøll BK, et al. Quantification of biofilm structures by the novel computer program COMSTAT. Microbiology. 2000;146:2395–407.
20. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.
21. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91.
22. Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. Nucleic Acids Res. 2018;46:W289–W295.
23. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10:1014.
24. Huerta-Cepas J, Forslund K, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2016a;34:2115–22.
25. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016b;44:D286–D293.
26. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845–58.
27. Zhang X, Ferguson-Miller SM, Reid GE. Characterization of ornithine and glutamine lipids extracted from cell membranes of *Rhodobacter sphaeroides*. J Am Soc Mass Spectrom. 2009;20:198–212.
28. Pujalte MJ, Lucena T, Ruvira MA, Arahal DR, Macian MC. The Family *Rhodobacteraceae*. In: Rosenberg E, editors. The Prokaryotes: alphaproteobacteria and betaproteobacteria. Springer; 2013. p. 440–512.
29. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, et al. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. ISME J. 2017;11:1483–99.
30. Korbes AP, Kulcheski FR, Margis R, Margis-Pinheiro M, Turchetto-Zolet AC. Molecular evolution of the lysophosphatidic acid acyltransferase (LPAAT) gene family. Mol Phylogen Evol. 2016;96:55–69.
31. Sule P, Belas R. A novel inducer of Roseobacter motility is also a disruptor of algal symbiosis. J Bacteriol. 2013;195:637–46.
32. Bartling P, Vollmers J, Petersen J. The first world swimming championships of roseobacters-Phylogenomic insights into an exceptional motility phenotype. Syst Appl Microbiol. 2018;41:544–54.
33. Frentzen M. Phosphatidylglycerol and sulfoquinovosyldiacylglycerol: anionic membrane lipids and phosphate regulation. Curr Opin Plant Biol. 2004;7:270–6.
34. Villanueva L, Bale N, Hopmans EC, Schouten S, Damste JSS. Diversity and distribution of a key sulpholipid biosynthetic gene in marine microbial assemblages. Environ Microbiol. 2013;16:774–87.
35. Umena Y, Kawakami K, Shen J-R, Kamiya N. Crystal structure of oxygen- evolving photosystem II at a resolution of 1.9 Å. Nature. 2011;473:55–60.
36. Selander E, Kubanek J, Hamberg M, Andersson MX, Cervin G, Pavia H. Predator lipids induce paralytic shellfish toxins in bloom-forming algae. Proc Natl Acad Sci USA. 2015;112:6395–6400.
37. Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, et al. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. eLife. 2012;1:e00013.
38. Godchaux W, Leadbetter ER. *Capnocytophaga* spp. contain sulfonolipids that are novel in procaryotes. J Bacteriol. 1980;144:592–602.
39. Godchaux W, Leadbetter ER. Sulfonolipids of gliding bacteria. J Biol Chem. 1984;259:2982–90.
40. Godchaux W, Leadbetter ER. Unusual sulfonolipids are characteristic of the *Cytophaga-Flexibacter* group. J Bacteriol. 1983;153:1238–46.
41. Abbanat DR, Leadbetter ER, Godchaux W, Escher A. Sulphonolipids are molecular determinants of gliding motility. Nature. 1986;324:367–9.
42. Godchaux W, Leadbetter ER. Sulfonolipids are localized in the outer membrane of the gliding bacterium *Cytophaga johnsonae*. Arch Microbiol. 1988;150:42–47.
43. Minnikin DE, Abdolrahimzadeh H. The replacement of phosphatidylethanolamine and acidic phospholipids by an ornithine-amide lipid and a minor phosphorus-free lipid in *Pseudomonas fluorescens* NCMB 129. FEBS Lett. 1974;43:257–60.

44. Weissenmayer B, Gao J-L, Lopez-Lara IM, Geiger O. Identification of a gene required for the biosynthesis of ornithine-derived lipids. Mol Microbiol. 2002;45:721–33.
45. Ito K, Miyazawa K, Matsumoto F. Amino acid composition of the ethanolic extractives from 31 species of marine red algae. Hiroshima Daigaku Sui Chikusangakubu Kiyo. 1977;16:77–90.
46. Miyasawa K, Ito K, Matsumoto F. Occurrence of (+)-2-hydroxy-3-aminopropanesulfonic acid and 3-aminopropanesulfonic acid in a red alga, *Grateloupia livida*. Nippon Suisan Gakkaishi. 1970;36:109–14.
47. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. Nat Microbiol. 2019;4:1706–15.
48. Lewenza S, Falsafi R, Bains M, Rohs P, Stupak J, Sprott GD, et al. The *olsA* gene mediates the synthesis of an ornithine lipid in *Pseudomonas aeruginosa* during growth under phosphate-limiting conditions, but is not involved in antimicrobial peptide susceptibility. FEMS Microbiol Lett 2011;320:95–102.

The ISME Journal (2021) 15:3303–3314
https://doi.org/10.1038/s41396-021-01008-7

ARTICLE

Phosphorus stress induces the synthesis of novel glycolipids in *Pseudomonas aeruginosa* that confer protection against a last-resort antibiotic

Rebekah A. Jones[1,2] · Holly Shropshire[2] · Caimeng Zhao[3] · Andrew Murphy[2] · Ian Lidbury[2,4] · Tao Wei[3] · David J. Scanlan[2] · Yin Chen[2]

Received: 16 February 2021 / Revised: 20 April 2021 / Accepted: 6 May 2021 / Published online: 24 May 2021
© The Author(s) 2021. This article is published with open access

Abstract

Pseudomonas aeruginosa is a nosocomial pathogen with a prevalence in immunocompromised individuals and is particularly abundant in the lung microbiome of cystic fibrosis patients. A clinically important adaptation for bacterial pathogens during infection is their ability to survive and proliferate under phosphorus-limited growth conditions. Here, we demonstrate that *P. aeruginosa* adapts to P-limitation by substituting membrane glycerophospholipids with sugar-containing glycolipids through a lipid renovation pathway involving a phospholipase and two glycosyltransferases. Combining bacterial genetics and multi-omics (proteomics, lipidomics and metatranscriptomic analyses), we show that the surrogate glycolipids monoglucosyldiacylglycerol and glucuronic acid-diacylglycerol are synthesised through the action of a new phospholipase (PA3219) and two glycosyltransferases (PA3218 and PA0842). Comparative genomic analyses revealed that this pathway is strictly conserved in all *P. aeruginosa* strains isolated from a range of clinical and environmental settings and actively expressed in the metatranscriptome of cystic fibrosis patients. Importantly, this phospholipid-to-glycolipid transition comes with significant ecophysiological consequence in terms of antibiotic sensitivity. Mutants defective in glycolipid synthesis survive poorly when challenged with polymyxin B, a last-resort antibiotic for treating multi-drug resistant *P. aeruginosa*. Thus, we demonstrate an intriguing link between adaptation to environmental stress (nutrient availability) and antibiotic resistance, mediated through membrane lipid renovation that is an important new facet in our understanding of the ecophysiology of this bacterium in the lung microbiome of cystic fibrosis patients.

Introduction

Pseudomonas aeruginosa is a significant nosocomial pathogen in intensive care units causing pneumonia, surgical wound site infections and sepsis [1, 2]. It is now recognised as a leading cause of morbidity and mortality in chronically infected cystic fibrosis (CF) patients and immunocompromised individuals due to the surge of carbapenem-resistant strains, a key group of first line antibiotics for treating *P. aeruginosa* infections [3]. For these drug-resistant *P. aeruginosa* strains, a viable but not ideal treatment option are polymyxins, considered to be last resort antibiotics. Although polymyxins are active against *P. aeruginosa*, their use was originally discontinued due to concerns over toxicity [4]. Indeed, *P. aeruginosa* has started to develop mechanisms of resistance to polymyxins due to an increase in their use globally. These primarily include modifications to the lipopolysaccharide (LPS) layer of the outer membrane through the addition of 4-amino-4-deoxy-

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41396-021-01008-7.

✉ Yin Chen
y.chen.25@warwick.ac.uk

[1] MRC Doctoral Training Partnership, University of Warwick, CV4 7AL Coventry, UK

[2] School of Life Sciences, University of Warwick, CV4 7AL Coventry, UK

[3] School of Food and Biological Engineering, Zhengzhou University of Light Industry, 450000 Zhengzhou, China

[4] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

L-arabinose (L-Ara4N) or phosphoethanolamine (pEtN) [5, 6]. These changes perturb the electrostatic interaction between cationic polymyxins and the normally negatively charged LPS.

Glycerophospholipids, such as phosphatidylglycerol (PG) and phosphatidylethanolamine (PE), are the major lipids forming the membrane lipid bilayer in bacteria, archaea, and eukaryotes [7–11]. They play a fundamental role in the evolution of the cell and it is widely accepted that the last universal common ancestor possessed a phospholipid membrane [12, 13]. Although it is uncertain why evolution selected glycerophospholipids as the building blocks for maintaining cellular membranes [13], it is known that organisms can alter their membrane lipid composition in response to nutrient stress or environmental changes [7, 14]. Previous studies have firmly established the link between nutrient stress, particularly phosphorus availability, and the expression of a variety of virulence factors in *P. aeruginosa* [15–19]. However, it is unclear whether adaptation to phosphorus limitation in this bacterium causes a change in membrane lipid composition and, if so, whether lipid remodelling comes with unforeseen ecophysiological consequences. Using a synthesis of multi-omics approaches, here we show that *P. aeruginosa* produces surrogate glycolipids to replace phospholipids in response to phosphorus limitation. These glycoglycerolipids have not previously been reported in *P. aeruginosa*. This lipid renovation pathway is strictly conserved in all *P. aeruginosa* strains isolated from a range of clinical settings and actively expressed in the metatranscriptome of CF patients. Importantly, such a phospholipid-to-glycolipid transition comes with a significant consequence in antibiotic sensitivity, in that glycolipids confer protection when challenged with the antimicrobial peptide polymyxin B. As such, glycolipid-mediated resistance to polymyxin B represents a new resistance mechanism that is quite different from the previously documented modification of LPS [5, 6]. This work highlights how the physiological adaptation of *P. aeruginosa* to phosphorus limitation can mediate a physiological response that may have profound implications for the survival of the bacteria in the lung microbiome.

Materials and methods

Cultivation of *P. aeruginosa* and mutants

P. aeruginosa strain PAO1 was obtained from the DSMZ culture collection (Germany) and routinely cultured in lysogeny broth (LB). A defined medium previously outlined for *Pseudomonas* species to control phosphate levels was also used [20]. This modified minimal media A comprised: Na-succinate 20 mM, NaCl 200 mg L^{-1}, NH_4Cl 450 mg L^{-1}, $CaCl_2$ 200 mg L^{-1}, KCl 200 mg L^{-1}, $MgCl_2$ 450 mg L^{-1}, with trace metals $FeCl_2$ 10 mg L^{-1} and $MnCl_2$ 10 mg L^{-1}, with 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer used at pH 7. Na_2HPO_4 was then added to a final concentration of 50 µM (low P) or 1 mM (high P). An intermediate phosphate source of 400 µM Na_2HPO_4 was used for overnight cultures in some experiments to prevent any excess storage of phosphate that could hamper results. All components were filter sterilised using 0.22 µm pore-size filters, and made up using double deionised H_2O. Mutants were obtained from the *P. aeruginosa* strain PAO1 transposon mutant library at the University of Washington, and confirmed using PCR and subsequent sequencing.

Alkaline phosphatase assay

Alkaline phosphatase activity was monitored as a measure of Pi stress. Liquid *P. aeruginosa* culture samples were incubated with 10 mM *para*-nitrophenol phosphate (*p*NPP) to a final concentration of 1 mM. Yellow-*p*NP supernatant was measured in triplicate at 407 nm (BioRad iMark microplate reader). Readings were normalised using both a Tris-only incubation control and further by bacterial density (optical density reading at 600 nm (OD_{600})).

Over-expression of Agt1 and Agt2 in *E. coli*

P. aeruginosa genes PA3218 (*agt1*) and PA0842 (*agt2*) were codon optimised for *Escherichia coli* and chemically synthesised (GenScript) into plasmid pET-28a(+). *E. coli* BLR(DE3) competent cells were thawed for 5 min before incubation with 10 ng pET-28a_Agt plasmid, and placed on ice for 5 min. Cells were then subjected to heat shock at 42 °C for 30 s, placed back on ice for 2 min. Recovery SOC media was added, with samples incubated at 37 °C shaking, for 1 h. Transformed cells were then plated onto kanamycin-LB agar and grown overnight at 37 °C. To harvest cells for lipid extraction, single colonies were picked to grow in small volume LB-Kan to 0.6 OD_{600} before induction with 0.4 mM IPTG overnight at 25 °C. 1 mL samples were then pelleted at 10,000 × *g* for 5 min. Pellets were stored at −80 °C until lipid extraction and subsequent analysis on HPLC-MS.

To purify the Agt1 and Agt2 proteins from recombinant *E. coli*, IPTG was added to a final concentration of 0.5 mM once the cultures reached an OD_{600} of 0.6. After a further 12 h of growth at 30 °C, cells were harvested by centrifugation and resuspended in buffer A containing 50 mM Tris-HCl, pH 7.9, 50 mM NaCl. Cells were disrupted by sonication and 1% (w/v) triton X-100 was then added and the cells were then incubated for 2.5 h at 4 °C. The cells were then centrifuged at 12,000 × *g* for 20 min, and the soluble fraction was loaded onto a nickel column (GE Healthcare, USA) pre-equilibrated with buffer A. The

recombinant Agt1 and Agt2 enzymes were eluted with an elution buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 300 mM imidazole) and dialysed overnight into buffer A to remove imidazole. For further purification, the samples were dialysed overnight into buffer B containing 50 mM Tris-HCl, pH 7.9, 200 mM NaCl, concentrated by ultrafiltration using a 30-kDa membrane (Millipore), and loaded onto a Superdex 200 (16/60) gel filtration column (GE Healthcare, USA), which was pre-equilibrated with buffer B (50 mM Tris-HCl, pH 7.9, 200 mM NaCl). The fraction size was 0.5 ml, and the flowrate was 0.5 ml/min. Purified protein was analysed by SDS-PAGE, and protein concentrations were determined using the Bradford assay.

Membrane lipid extraction and HPLC-MS analysis

Intact polar membrane lipids were extracted using a modified version of the typically used Folch extraction method [21, 22]. Liquid *P. aeruginosa* cultures growing in high and low phosphate modified minimal medium A were sampled after 8 h, collecting the equivalent of 0.5 OD_{600} into a 2 mL glass chromacol vial (Thermo Scientific), pelleted at 4 °C, 4000 rpm for 15 min. For lipid extraction, a ratio of 500:300:1000 µL of methanol:water:chloroform (all LC-MS grade) was used. The lipid fraction was collected from the lower phase using a glass Pasteur pipette. This chloroform extract was then dried under a stream of nitrogen (Techne sample concentrator) and resuspended in 1 mL 95% (v/v) acetonitrile (HPLC grade): 5% (w/v) ammonium acetate (10 mM, pH 9.2) for analysis. Extracted lipid samples were analysed using an UltiMate 3000 HPLC (Thermo Scientific) system coupled to an AmazonSL quadrupole ion trap (Bruker) mass spectrometer (MS), using electrospray ionisation. Hydrophilic interaction chromatography using a BEH amide XP column (Waters) was utilised to separate lipid classes based on their head group [23]. The column chamber was maintained at 30 °C and the samples passed through at a 150 µL min^{-1} flow rate. The mobile phase of acetonitrile:ammonium acetate (pH 9.2) was used to elute the sample in a 15 min per sample gradient, from 95 to 28% ammonium acetate. The lipid d17:1/12:0 sphingosylphosphoethanolamine (Sigma-Aldrich, 50 nM) was added to the samples and used as an internal standard. Tandem MS (or MS^n) was used to fragment the intact lipids for identification. The data were analysed using the Bruker Compass software package (DataAnalysis and QuantAnalysis).

Enzyme activity assays

The glycosyltransferase activity of Agt1 and Agt2 was measured using uridine diphosphate (UDP)-glucose or UDP-glucuronic acid and 0.1 mM C16:0/C18:1 diacylglycerol (DAG) as the substrate. 2.0 µM purified enzyme was used in 10 mM Tricine/KOH buffer, pH 8.5 with 2 mM dithiothreitol. The resulting mixture (500 µl) was incubated at 30 °C for 60 min with constant shaking at 200 rpm. The lipid products were extracted using the Folch method as described above. The lipid extracts were further analysed by LC-MS for the identification of MGDG/GADG though MS^n fragmentation and for the quantification of DAG against standards. The K_m and V_{max} values were calculated using Michaelis–Menten plots with various concentrations of UDP-sugars (0.1 to 1.0 mM) in three replicates.

Antibiotic sensitivity assays

P. aeruginosa cultures were grown to an OD_{600} of 0.6 in high or low phosphate minimal media A (see above). Cultures were then diluted 1:100 in prewarmed minimal media A containing 4 µg mL^{-1} polymyxin B sulfate (Sigma). Samples were incubated at 37 °C, 180 rpm, and assayed for survivors at specified time points by serial dilution plating onto LB. *E. coli* cultures containing pET-28a-Agt1 or pET-28a-Agt2 were grown to an OD_{600} of 0.6 in LB broth, and the expression of Agt1 and Agt2 was induced by incubation with 0.4 mM IPTG overnight at 25 °C. A negative control of *E. coli* containing the pET-28a vector only was also set up. Overnight cultures were diluted 1:100 in prewarmed LB broth containing 0.4 mM IPTG and 20 µg mL^{-1} polymyxin B sulfate (Sigma). Samples were incubated at 37 °C, 180 rpm, and assayed for survivors at specified time points by serial dilution plating onto LB agar + kanamycin 25 µg mL^{-1}.

Comparative proteomic analysis

P. aeruginosa PAO1 WT (1 mM phosphate, 50 µM phosphate) and PlcP mutant (50 µM phosphate) cell pellets in three biological replicates were resuspended in LDS (lithium dodecyl sulfate) sample buffer containing 1% β-mercaptoethanol before lysing at 95 °C and vortexing. 30 µL of each sample were run on NuPAGE 10% Bis-Tris protein gel (Invitrogen) for a short time before staining with SafeStain (Thermo Fisher) and excising the whole protein band. In-gel proteins were de-stained using 50% (v/v) ethanol, 50 mM ammonium bicarbonate (ABC), before being reduced and alkylated for 5 min at 70 °C using 10 mM TCEP (tris(2-carboxyethyl)phosphine) and 40 mM CAA (2-chloroacetamide), respectively. After washing with 50% (v/v) ethanol 50 mM ABC, peptides were lysed overnight using trypsin. Finally, peptides were extracted by sonication in a water bath (10 min at room temperature), concentrated using a Speed-Vac (50 min) and resuspended in 2.5% acetonitrile 0.05% formic acid. Extracted peptides were analysed by nanoLC-ESI-MS/MS using the Ultimate 3000/Orbitrap Fusion instrumentation (Thermo Scientific). The UniProt proteome for *P. aeruginosa* strain PAO1 was

used for peptide analysis. Further data analysis was carried out using MaxQuant and Perseus software as described previously; peptides without triplicate measures were filtered out [24].

Phylogenomics and metatranscriptomics analyses

The protein sequences of PA3219, PA3218 and PA0842 were used to search genome sequences of *Pseudomonas* clades in the JGI IMG genome portal (https://img.jgi.doe.gov/). Note that the PA3218 protein is incorrectly annotated in the genome of PAO1. The putative glycosyltransferase located immediately downstream of PA3219 was manually inspected by aligning to the corresponding gene (PA14_22600) in the genome of *P. aeruginosa* PA14. To identify PA3218 in misannotated *P. aeruginosa* genomes, the nucleotide sequence immediately downstream of PA3219 was aligned with *agt1* in strain PA14, (locus tag PA14_22600). The phylogeny of *Pseudomonas* clades was determined using the nucleotide sequences of six housekeeping genes (*rpoB, rpoD, dnaE, recA, atpD, gyrB*) retrieved from each genome using IQ-Tree with the parameters -m TEST -bb 1000 -alrt 1000. The most suitable model was chosen by the software. Evolutionary distances were inferred using maximum-likelihood analysis. Relationships were visualised using the online platform the Interactive Tree of Life viewer (https://itol.embl.de/). The conserved Pho box sequence was predicted using the MEME server [25].

The metatranscriptomics datasets of sputum samples obtained from a CF patient 7-days (SRX5145606) and 8-days (SRX5145605) before death [26], CF patient G (SRR6833349) from Denmark [27] and a patient (SRR6833340) with chronic wound infection [27] were retrieved from the short reads archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). The reads were downloaded using fastq-dump and mapped using the BBMap aligner as described previously [28]. Briefly, the SRA reads were mapped to the genome sequence of *P. aeruginosa* PAO1 using a stringent cut-off of minid = 0.97. Relative abundance data were compared using RPKM (reads per kilobase of transcript, per million mapped reads).

Results and discussion

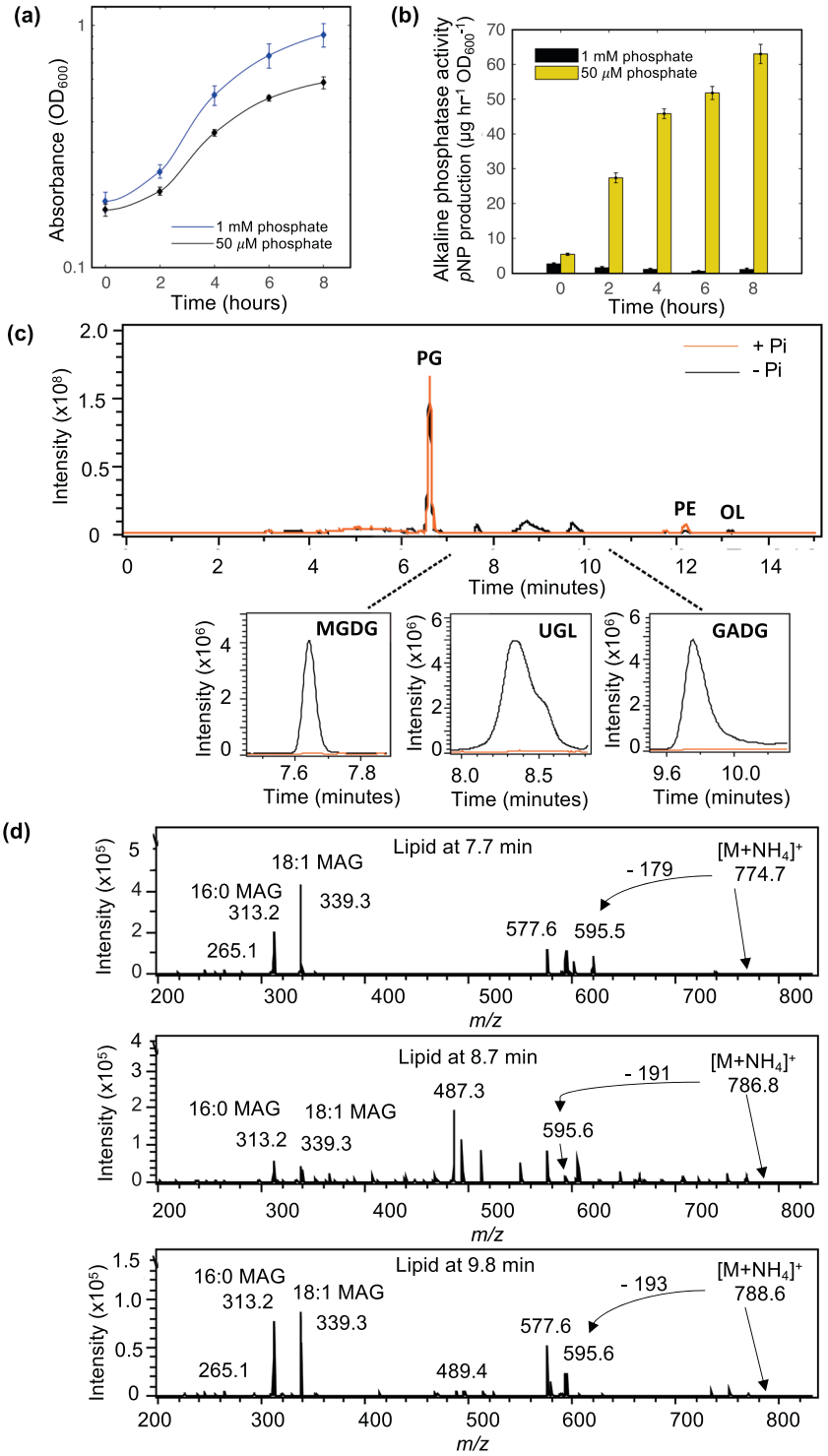
P. aeruginosa produces novel glycolipids in response to Pi stress

To determine changes in the membrane lipidome in response to P-stress, the model *P. aeruginosa* strain PAO1 was grown in minimal medium under high (1 mM) or low Pi (50 µM) conditions (Fig. 1a). The latter condition elicited strong alkaline phosphatase activity, measured through the liberation of *para*-nitrophenol (*p*NP) from *p*NPP (Fig. 1b), this being a strong indication that cells were P-stressed. Analysis of membrane lipid profiles using high-performance liquid chromatography coupled to mass spectrometry (HPLC-MS) revealed the presence of several new lipids under Pi stress conditions (Fig. 1c). Thus, during Pi-replete growth (1 mM phosphate), the lipidome is dominated by two glycerophospholipids: PG (eluted at 6.8 min) and PE (eluted at 12.2 min). During Pi-stress a lipid species with mass to charge ratio (*m/z*) of 623 and 649 were also found, with MS fragmentation resulting in a 131 *m/z* peak, a diagnostic ion for the amino-acid containing ornithine lipid. This is consistent with previous reports of ornithine lipids in the *P. aeruginosa* membrane in response to Pi stress [29, 30].

Further to ornithine lipids, three unknown lipids eluting at 7.7, 8.7 and 9.8 min, were only present under Pi stress conditions (Fig. 1c). Using several rounds of MS fragmentation (MS^n), with a quadrupole ion trap MS, fragmentation patterns characteristic of glycolipids were found for all three peaks. For each peak of interest, the most predominant lipid masses of 774.7, 786.8 and 788.6 *m/z* were analysed by MS^n in positive ionisation mode (Fig. 1d). In each case, an initial head group was lost leaving a significant signal of 595.6 *m/z*, the mass of the glycolipid building block diacylglycerol (DAG). Further fragmentation leads to the loss of either fatty acyl chain from DAG, leaving monoacylglycerols of 313.2 and 339.3 *m/z*. Two monoacylglycerols with different masses are produced as a result of the original lipid containing 16:0 and 18:1 fatty acids (313.2 and 339.3 *m/z* monoacylglycerols, respectively). To further elucidate the identity of the peaks, a search for a neutral loss of a polar head group was carried out. Thus, the intact masses of 774.7 and 788.6 *m/z* in positive ionisation mode leads to the loss of a head group of −179 and −193 *m/z*, which corresponds to a hexose- and a glucuronate- group, respectively (Fig. 1d), suggesting the occurrence of novel monoglucosyldiacylglycerol (MGDG) and glucuronic acid diacylglycerol (GADG) glycolipids in *P. aeruginosa*. The third glycolipid peak at 8.7 min remains an unknown lipid with intact mass of 786.8 *m/z* (hereafter designated as a putative unknown glycolipid, UGL). Together, these data confirm the production of new glycolipids in *P. aeruginosa* in response to Pi stress.

Comparative proteomics uncover the lipid renovation pathway in *P. aeruginosa*

To determine the proteomic response of *P. aeruginosa* to phosphorus limitation, and identify the genes involved in glycolipid formation, strain PAO1 was cultivated under high and low Pi conditions for 8 h and the cellular proteome then analysed. A total of 2844 proteins were detected, 175



of which were found to be differentially regulated by Pi availability (Fig. 2a, Table S1). In line with previous transcriptomic studies of strain PAO1 [18], major phosphorus acquisition mechanisms were highly expressed under Pi stress conditions, *e.g.* the Pi-specific transporter PstSCAB, the two-component regulator PhoBR (Table S1) [31].

Comparative proteomics also identified several genes which are likely important for membrane lipid remodelling (Fig. 2b) including PA3219 (4.6-fold increase under Pi-deplete conditions, FDR < 0.01), encoding a putative phospholipase C protein, and PA0842 (4-fold increase under Pi-deplete conditions, FDR < 0.01), encoding a

◀ **Fig. 1 Lipidomics analysis uncovers novel glycolipid formation in *Pseudomonas aeruginosa* strain PAO1 in response to phosphorus limitation. a** Growth of strain PAO1 WT in minimal medium A containing 1 mM phosphate (+Pi, blue) or 50 µM phosphate (−Pi, black) over 12 h. Data are the average of three independent replicates. **b** Liberation of *para*-nitrophenol (*p*NP) from *para*-nitrophenol phosphate (*p*NPP) through alkaline phosphatase activity, under Pi-replete (1 mM, black) and Pi-deplete (50 µM, yellow) conditions. Error bars represent the standard deviation of three independent replicates. **c** Representative chromatograms in negative ionisation mode of the *P. aeruginosa* lipidome when grown under phosphorus stress (−Pi, black) compared to growth under phosphorus sufficient conditions (+Pi, orange). PG phosphatidylglycerol, PE phosphatidylethanolamine, OL ornithine lipids. Lower panel: extracted ion chromatograms of three new glycolipid species in *P. aeruginosa* which are only produced during Pi-limitation (black, 1 mM; orange, 50 µM). MGDG monoglucosyldiacylglycerol, GADG glucuronic acid-diacylglycerol and UGL unconfirmed glycolipid. **d** Mass spectrometry fragmentation spectra of three glycolipid species present under Pi stress in *P. aeruginosa*, at retention times of 7.7 (*m/z* 774.7), 8.7 (*m/z* 786.7) and 9.8 (*m/z* 788.6) minutes, respectively. Each spectrum depicts an intact lipid mass with an ammonium (NH_4^+) adduct exhibiting neutral loss of a head group, yielding diacylglycerol (DAG) (595 *m/z*). Further fragmentation yields monoacylglycerols (MAG) with C16:0 or C18:1 fatty acyl chains.

putative glycosyltransferase (Fig. 2a). PA3219 has 47% protein sequence identity to PlcP from *Phaeobacter* sp. MED193 and 46% identity to PlcP from *Sinorhizobium meliloti* [26–28]. In these bacteria, PlcP is essential in the lipid remodelling pathway for the formation of the diacylglycerol (DAG) backbone, representing the essential intermediate for the production of glycolipids [32, 33]. In *P. aeruginosa* PAO1, PA3219 appears to form an operon with PA3218, a putative glycosyltransferase likely under the control of the PhoBR two-component system, as a highly conserved Pho box sequence was recognisable in the promoter region (Fig. 2c). PA3218 (hereafter referred to as Agt1) has 41% protein sequence identity to the Agt of *Phaeobacter* sp. MED193. PA0842 showed 35% identity to the Agt of *Phaeobacter* sp. MED193 and a Pho box sequence is also found in its promoter region. This corroborates the finding that the PA0842 protein (hereafter referred to as Agt2) was significantly upregulated under Pi-deplete conditions (Fig. 2a). In summary, comparative proteomic analysis suggests that *P. aeruginosa* PAO1 adopts this PlcP-Agt lipid remodelling pathway for the production of glycolipids in response to Pi-stress (Fig. 2b).

The PlcP-Agt mediated lipid renovation pathway is strictly conserved in *P. aeruginosa* and actively transcribed in the metatranscriptomes of cystic fibrosis patients

To uncover how widespread this predicted PlcP-Agt lipid remodelling pathway is amongst the genus *Pseudomonas*, including *P. aeruginosa* strains, we conducted a thorough comparative genomics analysis of these lipid renovating loci. PlcP-Agt is strictly conserved in all 770 genome-sequenced *P. aeruginosa* strains in the IMG/M database, including all three-previously recognised *P. aeruginosa* lineages [34, 35], group 1 represented by strain PAO1, group 2 represented by strain PA14 and group 3 represented by strain PA7 (Fig. 3, Table S2). Indeed, this remodelling pathway is prevalent in many *Pseudomonas* groups, including the plant pathogen *P. syringae*. To investigate whether the PlcP-Agt lipid remodelling pathway is involved in host-pathogen interactions, we analysed metatranscriptomic datasets from CF patients, where *P. aeruginosa* is known to be prevalent in the fatal exacerbation period before patient death [26]. To the best of our knowledge, only two studies have reported the metatranscriptome of the bacterial community present in CF sputum [26, 27]. Indeed, *phoBR* and *pstS* are amongst the most highly expressed genes, confirming previous observations that *P. aeruginosa* is Pi-limited during human airway epithelia infection [36, 37]. Interestingly, the alkaline phosphatase *phoA* [38] was highly expressed in sputum from CF patients but not from wound samples which was also dominated by *P. aeruginosa*. Importantly, the transcripts of *P. aeruginosa agt1/plcP/agt2* are highly expressed in CF sputum during the fatal exacerbation period before death (Fig. 2d). Therefore, our phylogenomic and metatranscriptomic analyses suggest that not only is the PlcP-Agt lipid remodelling pathway strictly conserved and prevalent in *P. aeruginosa*, but also the corresponding genes are also highly expressed during CF patient infection, suggesting a potential role for lipid renovation in host-pathogen interactions.

Experimental validation of the lipid renovation pathway for glycolipid formation in *P. aeruginosa*

To validate the function of these two putative glycosyltransferases (Agt1, Agt2) in the formation of glycolipids, we synthesised the codon-optimised genes (PA3218 and PA0842, respectively) for recombinant expression in *E. coli*. The total lipidomes from the recombinant *E. coli* strains were then analysed by HPLC-MS to determine the presence of glycolipids in a gain-of-function assay. Expression of *P. aeruginosa* Agt1 (PA3218) was sufficient for the production of MGDG (eluted at 7.7 min) in *E. coli*, confirmed through MS^n fragmentation (Fig. 4a). No UGL nor GADG was observed in the lipidome of this Agt1-overexpressing *E. coli* strain. Expressing Agt2 (PA0842) from *P. aeruginosa* in *E. coli* was sufficient for the accumulation of the GADG glycolipid (eluted at 9.8 min), also confirmed through the MS^n fragmentation pattern (Fig. 4b). Equally, no UGL nor MGDG was observed in the Agt2-overexpressing *E. coli* strain. Production of these glycolipids was not observed in the same strain of *E. coli*

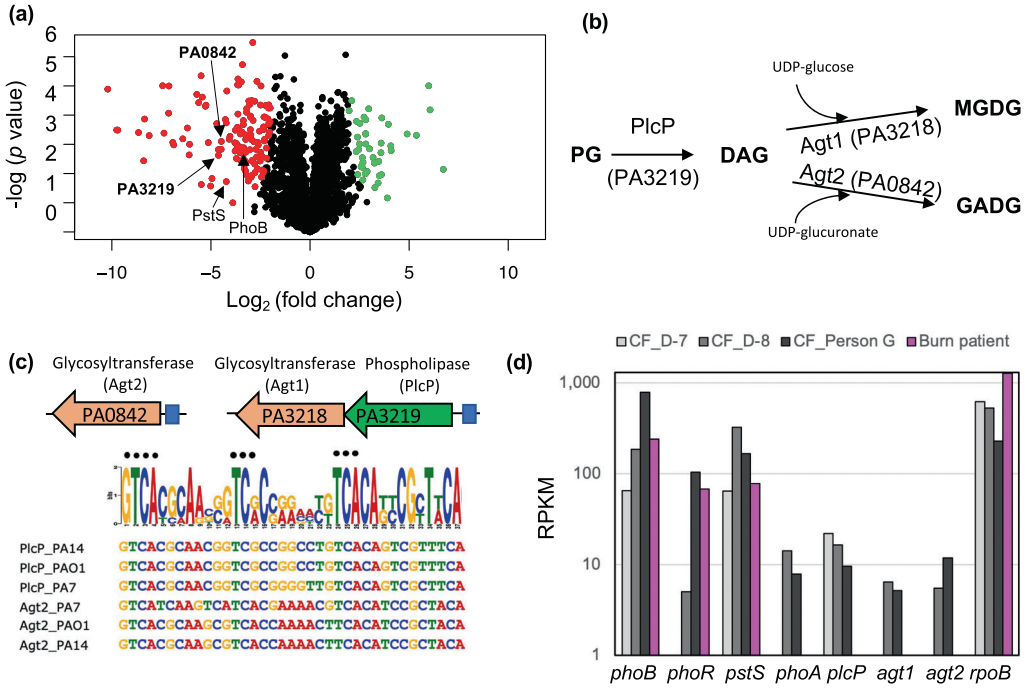


Fig. 2 Comparative multi-omic analyses for the identification of the PlcP-Agt pathway responsible for glycolipid formation in *Pseudomonas aeruginosa* strain PAO1. a Volcano plot depicting differentially expressed proteins when comparing Pi-replete and Pi-deplete conditions. Significantly upregulated proteins when under Pi stress are shown in red (left), and those that are significantly upregulated when Pi is sufficient are in green (right). Significance was accepted when the false discovery rate (FDR) was <0.05, and a fold change ≥2. **b** The proposed pathway for lipid remodelling through the PlcP-Agt pathway. PlcP degrades membrane phospholipids such as PG, to generate diacylglycerol (DAG) intermediates for the formation of MGDG and GADG through the activity of glycosyltransferases, using either UDP-glucose or UDP-glucuronate as the co-substrate [41]. **c** Genomic organisation of predicted lipid remodelling genes in *P. aeruginosa*. Glycosyltransferases (orange) PA3218 (Agt1) and PA0842 (Agt2) are predicted to be involved in glycolipid synthesis. PA3219 is predicted to be PlcP in *P. aeruginosa*. Predicted Pho box sequences in the promoter regions (represented in blue boxes) of each glycosyltransferase operon from *P. aeruginosa* strains representing the PAO1 clade, the PA7 clade and the PA14 clade are shown. The black dots represent residues which are conserved in the Pho box consensus CTGTCATNNNNCTGTCAT [42]. **d** Metatranscriptomic analysis of PlcP-Agt lipid remodelling genes in sputum samples from a cystic fibrosis patient 7-days (CF_D-7) and 8-days (CF_D-8) before death [26] and a Danish CF patient (CF_Person G) [27] as well as a wound sample from a burns patient from the USA (Burn patient) [27]. Relative abundance is expressed as RPKM (reads per kilobase of transcript, per million mapped reads). PhoA (PA3296) encodes an alkaline phosphatase [38]. The list of RPKM abundance of individual genes of *P. aeruginosa* PAO1 is shown in Table S4.

transformed with an empty vector control (pET28a). It is therefore likely that UGL production is carried out by another glycosyltransferase, the identity of which remains to be discovered.

To confirm the role of Agt1 and Agt2 in the production of MGDG and GADG, we purified Agt1 and Agt2 from recombinant *E. coli* (Fig. 4c) and carried out enzyme assays using UDP-glucose and UDP-glucuronic acid as the sugar donor and DAG as the acceptor. Agt1 can only accept UDP-glucose as the substrate with an affinity of $K_m = 298.1 \pm 9.5$ μM (Fig. 4c, middle panel) and produced MGDG as expected (Fig. 4d, left panel). Similarly, Agt2 can use UDP-glucuronic acid as the substrate ($K_m = 373.0 \pm 12.9$ μM (Fig. 4c, right panel), producing the GADG lipid (Fig. 4d, right panel). Interestingly, the purified Agt2 enzyme can also use UDP-glucose to some extent with a K_m of 480 μM (data not shown) although the corresponding lipid MGDG was not observed in the lipid extract from the lipidome of the recombinant host *E. coli* (Fig. 4b).

To further confirm the role of these genes in *P. aeruginosa* glycolipid biosynthesis we analysed the lipidomes of mutants in *ΔplcP*, *Δagt1* and *Δagt2* in strain PAO1 (Fig. 5a, b). Differences were analysed by searching for the intact masses of the glycolipids MGDG and GADG: 774.6 and 788.6 *m/z* in positive ionisation mode with an ammonium adduct, respectively. As expected, under Pi stress MGDG is no longer produced in the *Δagt1* mutant and similarly GADG is no longer produced in the *Δagt2* mutant (Fig. 5a). In the *ΔplcP* mutant, no MGDG was found and the GADG lipid was significantly reduced but not entirely abolished (Fig. 5b). The small amount of GADG produced in the *ΔplcP* mutant suggests that an alternative supply of DAG (independent of the degradation of phospholipids by PlcP) is available in this mutant. Nevertheless, lipidome analyses

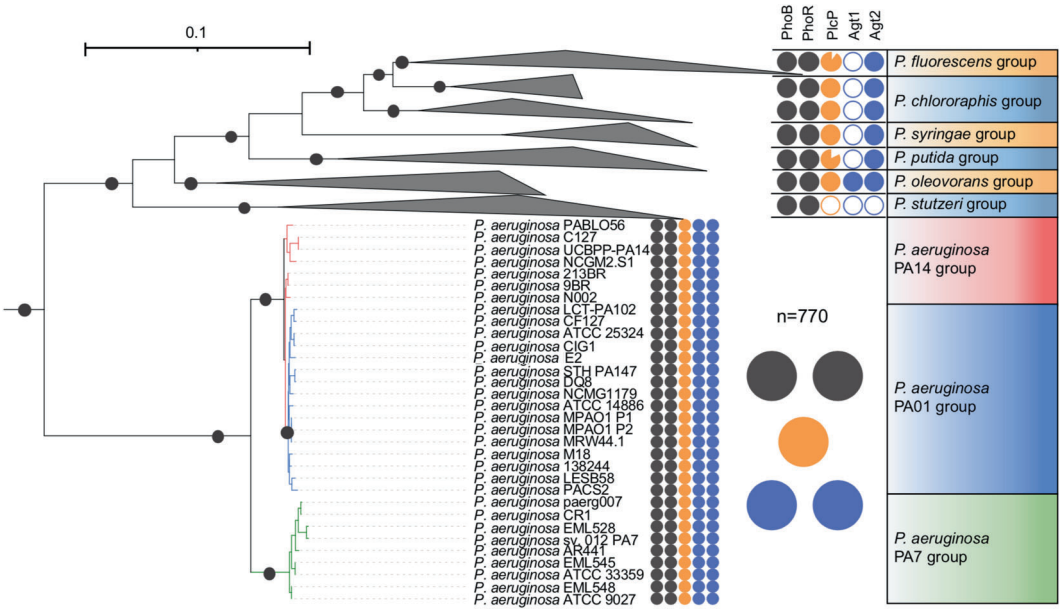


Fig. 3 Occurrence of *plcP-agt* genes in major *Pseudomonas* groups. The phylogeny of *Pseudomonas* clades was determined using the nucleotide sequences of six housekeeping genes (*rpoB, rpoD, dnaE, recA, atpD, gyrB*) retrieved from each genome using IQ-Tree [43]. The filled colour indicates the presence of the genes in the genomes whereas a blank indicates the absence of the corresponding gene in the genomes. The two-component system PhoBR (black circles) is found in all genomes and the PlcP-Agt1/Agt2 are strictly conserved in all 770 genome-sequenced *P. aeruginosa* strains that form three clades represented by strain PA14, PA01 and PA7, respectively. Bootstrap values >75% are shown.

of the *ΔplcP*, *Δagt1* and *Δagt2* mutants strongly supports the key role of this PlcP-Agt pathway (Fig. 2a) in lipid renovation in *P. aeruginosa*.

The protective role of glycolipids to antibiotic resistance in *Pseudomonas aeruginosa*

To assess whether growth of the glycolipid-deficient mutants (*ΔplcP*, *Δagt1*, *Δagt2*) was affected by Pi stress, we grew the mutants in the defined minimal medium under high and low Pi conditions. However, no significant difference in growth rates was found (Fig. S1). The presence of glycolipids in the membrane may, however, have a profound impact on the functioning of the membrane during Pi stress. For example, PG is an anionic lipid with net negative charges whereas MGDG has a neutral charged sugar group. Although a PG-to-GADG substitution may not necessarily change membrane charge [23], it may affect membrane curvature and the packing density of lipids. Thus, subsequent knock-on effects in membrane function might be expected [10]. We therefore set out to investigate whether membrane lipid composition may have an impact on antibiotic resistance in *P. aeruginosa*. As cationic antimicrobial peptides directly interact with bacterial cell membranes, we focused on the impact of lipid remodelling on the killing activity of polymyxin B. We conducted the analyses under P-deplete conditions, since Pi-stress is clinically important, already known to induce the expression of virulence factors [15, 17, 18, 30], and our own analysis confirmed that an array of genes involved in phosphate acquisition and lipid remodelling in *P. aeruginosa* are indeed highly expressed in sputum samples from the lung microbiome of CF patients (Fig. 2d). Polymyxins represent the drug-of-last resort for effectively treating carbapenem-resistant *P. aeruginosa* infections [3, 39].

To test the sensitivity of the mutants in the PlcP-Agt pathway to polymyxin B, we compared WT and mutants using kill curve analyses as the typically used disk diffusion method does not work efficiently for cationic antimicrobials [40]. Indeed, there was a significant decrease in the survival of all three PAO1 glycolipid synthesis mutants (*ΔplcP*, *Δagt1* and *Δagt2*) compared to the wild type when challenged with polymyxin B, suggesting a protective role of glycolipids in polymyxin B resistance (Fig. 5c). Such a protective role of glycolipids in polymyxin B resistance was not observed for other antibiotics, including ciprofloxacin, gentamicin, ceftazidime and meropenem (data not shown). *P. aeruginosa* is known to enhance its resistance to polymyxins through decoration of its LPS layer using either 4-amino-4-deoxy-L-arabinose (L-Ara4N) by *arnB* [5], or the addition of phosphoethanolamine (pEtN) by *eptA* [6]. It is thought that these changes perturb the electrostatic interaction between the cationic polymyxin B and the normally negatively charged LPS. To investigate whether these mechanisms play a role in the glycolipid-deficient mutants, we conducted a comparative proteomics analysis of the *ΔplcP* mutant and WT under

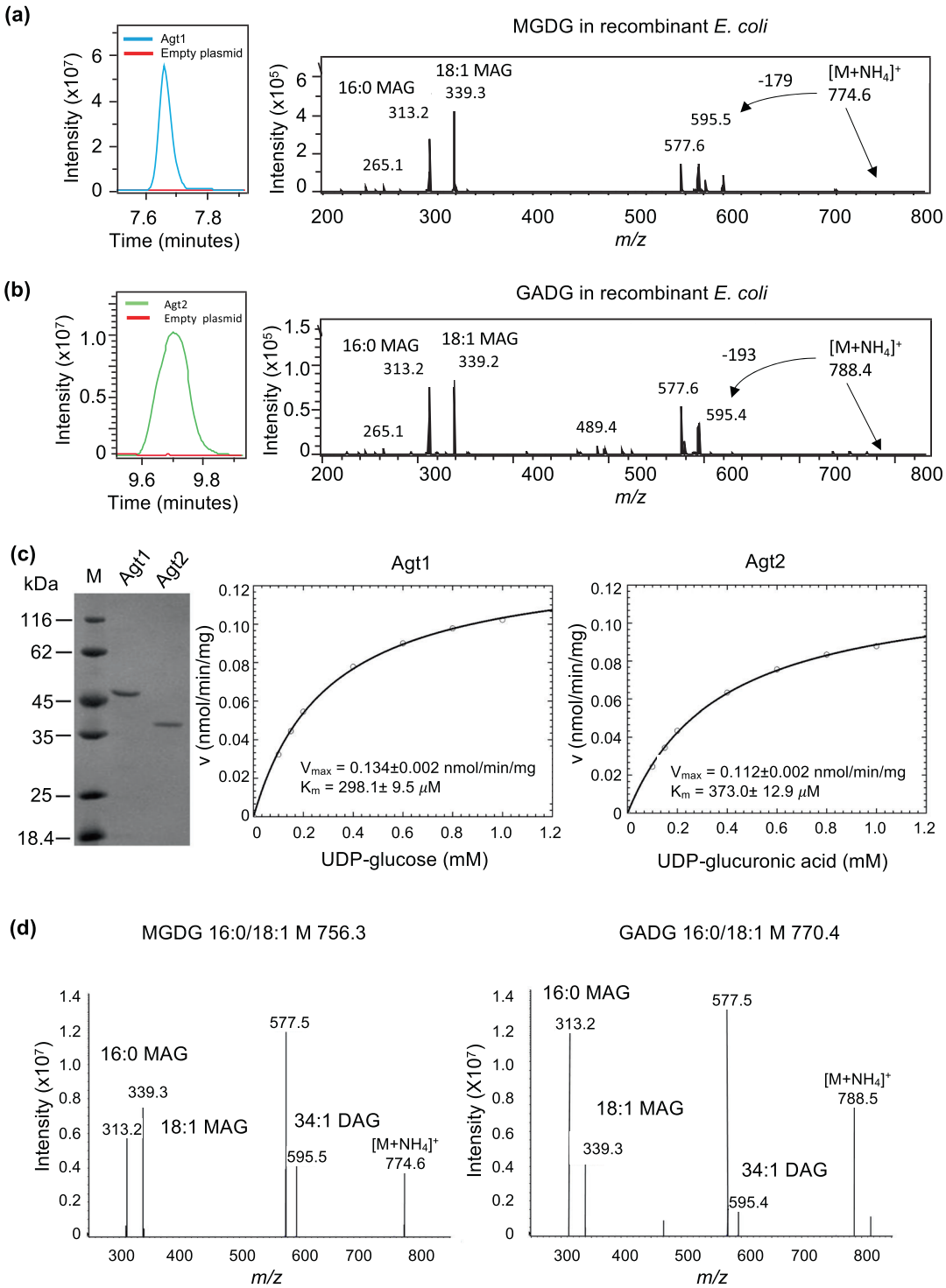


Fig. 4 Characterisation of glycolipid formation from recombinant Agt1 and Agt2. a Extracted ion chromatogram of the MGDG lipid from recombinant *E. coli* expressing Agt1. An empty vector control is also shown (red line). The identity of MGDG is further validated using mass spectrometry fragmentation showing the neutral loss of 179 corresponding to the loss of glucose and the formation of monoacylglycerols (MAG) with C16:0 or C18:1 (*m/z* 313.2, 339.3). **b** Extracted ion chromatogram of the GADG lipid from recombinant *E. coli* expressing Agt2. An empty vector control is also shown (red line). The identity of GADG is further validated using mass spectrometry fragmentation showing the neutral loss of 193 corresponding to the loss of glucose and the formation of monoacylglycerols (MAG) with C16:0 or C18:1 (*m/z* 313.2, 339.2). **c** Purified Agt1 and Agt2 protein from recombinant *E. coli* (left panel) and Michaelis Menten kinetics of Agt1 towards UDP-glucose (middle panel) and Agt2 towards UDP-glucuronic acid (right panel) as substrate, respectively. **d** Mass spectrometry identification of MGDG and GADG produced from purified Agt1 and Agt2 using DAG and UDP-glucose and UDP-glucuronic acid as the substrate, respectively.

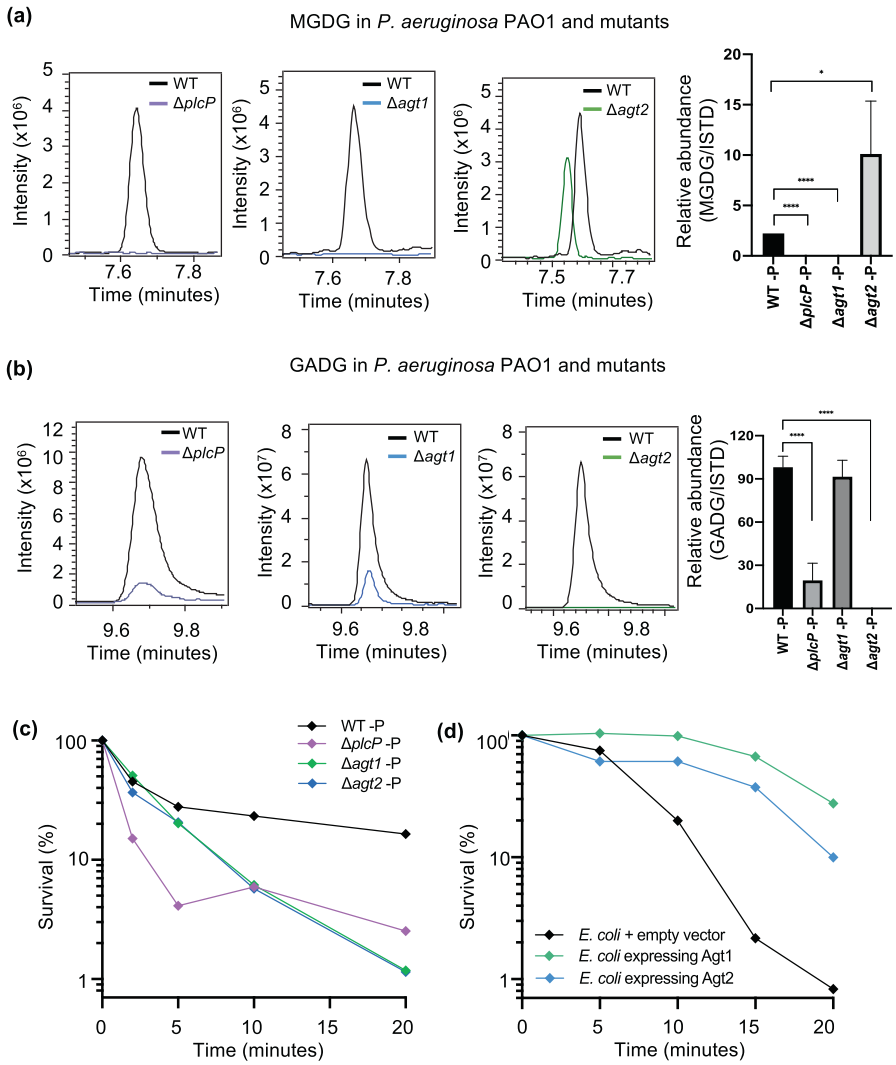


Fig. 5 Glycolipid formation in *P. aeruginosa* and mutants under Pi stress, showing a protective role of glycolipids to polymyxin B. Relative abundance of the glycolipid MGDG (**a**) and GADG (**b**) in *P. aeruginosa* mutants *plcP* (purple trace), *agt1* (blue trace) and *agt2* (green, trace) compared to the wild type (WT). Cells were cultivated under low Pi conditions (50 µM) and a representative extracted ion chromatogram of MGDG/GADG is shown between the WT, (black trace) and each mutant. The right most panel shows the abundance of MGDG or GADG calculated relative to an internal lipid standard d17:1/12:0 sphingosylphosphoethanolamine (Sigma-Aldrich) in the wild-type and mutant strains of *P. aeruginosa*. Values are calculated from three biological replicates and the error bars denote standard deviation. MGDG, monoglucosyldiacylglycerol, GADG, glucuronic acid-diacylglycerol. **c** Survival of glycolipid remodelling mutants Δ*plcP* (purple), Δ*agt1* (green) and Δ*agt2* (blue) when challenged with 4 µg mL^{-1} polymyxin B compared to WT under Pi stress (black). All experiments were conducted under Pi stress conditions and the results are the average of three biological replicates; error bars denote standard deviation. **d** Survival of glycolipid producing *Escherichia coli* when challenged with 20 µg mL^{-1} polymyxin B. All experiments were conducted in three replicates and error bars denote standard deviation. Black, *E. coli* containing the empty vector pET28a; green, *E. coli* containing plasmid pET28a-Agt1; blue, *E. coli* containing plasmid pET28a-Agt2.

Pi-deplete conditions, which revealed only a small number of differentially expressed proteins (Table S3). The majority of these differentially expressed proteins are uncharacterised. However, importantly, LPS modification enzymes previously found to confer antimicrobial peptide resistance, such as ArnB and EptA, were not differentially expressed between the WT and *ΔplcP* mutant. Therefore, our data suggests that it is the glycolipids that are the major contributor to increased polymyxin B resistance, which constitutes a new biological mechanism for polymyxin resistance. To this end, we tested the resistance to polymyxin B of recombinant *E. coli* strains overexpressing *P. aeruginosa* Agt1 and Agt2, that produce MGDG and GADG, respectively (Fig. 4a, b). Indeed, in this gain-of-function assay, both Agt1 and Agt2-overexpressing *E. coli* strains had enhanced resistance to polymyxin B compared to the empty vector control (Fig. 5d), supporting the protective role of these glycolipids to antimicrobial peptides.

To conclude, we present here the discovery of novel glycolipids produced in *P. aeruginosa* during adapation to phosphorus stress. This lipid renovation pathway is strictly conserved in all *P. aeruginosa* isolates to date and highly expressed in the metatranscriptome of CF patients, suggesting a key role of lipid remodelling in the ecophysiology of this bacterium. Interestingly, lipid remodelling as a response to survive phosphorus stress in turn comes with trade-offs in terms of antibiotic resistance; these glycolipids may protect the bacterium from insult by cationic antimicrobial peptides, highlighting a new resistance mechanism to polymyxin B which has been previously overlooked. It remains to be seen whether the altered susceptibility to polymyxin B is the sole trade-off following lipid remodelling of phospholipids to glycolipids. After all, evolution appears to have selected phospholipids as the dominant lipids in the last universal common ancestor [12].

Acknowledgements This work was funded by an MRC Doctoral Training Partnership studentship in Interdisciplinary Biomedical Research (MR/J003964/1) awarded to RAJ and by a Royal Society International Exchanges 2017 Cost Share (China) award (IEC\NSFC \170213; grant agreement no. 170213). AM and YC are supported by a European Research Council (ERC) award under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 726116). We also thank the Proteomics Research Technology Platform, University of Warwick, UK for their contribution.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Murray TS, Egan M, Kazmierczak BI. *Pseudomonas aeruginosa* chronic colonization in cystic fibrosis patients. Curr Opin Pediatr. 2007;19:83–8.
2. Gaynes R, Edwards JR, System NNIS. Overview of nosocomial infections caused by Gram-negative bacilli. Clin Infect Dis. 2005;41:848–54.
3. Hawkey PM, Livermore DM. Carbapenem antibiotics for serious infections. BMJ. 2012;344:e3236.
4. Landman D, Georgescu C, Martin DA, Quale J. Polymyxins revisited. Clin Microbiol Rev. 2008;21:449–65.
5. Chung ES, Lee JY, Rhee JY, Ko KS. Colistin resistance in *Pseudomonas aeruginosa* that is not linked to *arnB*. J Med Microbiol. 2017;66:833–41.
6. Nowicki EM, O'Brien JP, Brodbelt JS, Trent MS. Extracellular zinc induces phosphoethanolamine addition to *Pseudomonas aeruginosa* lipid A via the ColRS two-component system. Mol Microbiol. 2015;97:166–78.
7. Parsons JB, Rock CO. Bacterial lipids: metabolism and membrane homeostasis. Prog Lipid Res. 2013;52:249–76.
8. Zhang Y-M, Rock CO. Membrane lipid homeostasis in bacteria. Nat Rev Microbiol. 2008;6:222–33.
9. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol. 2008;9:112–24.
10. Harayama T, Riezman H. Understanding the diversity of membrane lipid composition. Nat Rev Mol Cell Biol. 2018;19:281–96.
11. May KL, Silhavy TJ. Making a membrane on the other side of the wall. Biochim et Biophy Acta. 2017;1862:1386–93.
12. Lombard J, López-García P, Moreira D. The early evolution of lipid membranes and the three domains of life. Nat Rev Microbiol. 2012;10:507–15.
13. Peretó J, López-García P, Moreira D. Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci. 2004;29:469–77.
14. Sahonero-Canavesi DX, López-Lara IM, Geiger O. Membrane lipid degradation and lipid cycles in microbes. Aerob Utilization Hydrocarb Oils Lipids. 2019. https://doi.org/10.1007/978-3-319-50418-6_38.
15. Lamarche MG, Wanner BL, Crépin S, Harel J. The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. FEMS Microbiol Rev. 2008;32:461–73.
16. Long J, Zaborina O, Holbrook C, Zaborin A, Alverdy J. Depletion of intestinal phosphate after operative injury activates the virulence of *P. aeruginosa* causing lethal gut-derived sepsis. Surgery. 2008;144:189–97.
17. Francis VI, Stevenson EC, Porter SL. Two-component systems required for virulence in *Pseudomonas aeruginosa.* FEMS Microbiol. Lett. 2017;364. https://doi.org/10.1093/femsle/fnx104.
18. Bains M, Fernández L, Hancock REW. Phosphate starvation promotes swarming motility and cytotoxicity of *Pseudomonas aeruginosa*. Appl Environ Microbiol. 2012;78:6762–8.
19. Son MS, Matthews WJ, Kang Y, Nguyen DT, Hoang TT. In vivo evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. Infect Immun. 2007;75:5313–24.
20. Lidbury IDEA, Murphy ARJ, Scanlan DJ, Bending GD, Jones AME, Moore JD, et al. Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. Environ Microbiol. 2016;18:3535–49.
21. Sebastián M, Smith AF, González JM, Fredricks HF, Van Mooy B, Koblížek M, et al. Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. ISME J. 2016;10:968–78.
22. Smith AF, Rihtman B, Stirrup R, Silvano E, Mausz MA, Scanlan DJ, et al. Elucidation of glutamine lipid biosynthesis in marine bacteria reveals its importance under phosphorus deplete growth in *Rhodobacteraceae*. ISME J. 2019;13:39–49.
23. Diercks H, Semeniuk A, Gisch N, Moll H, Duda KA, Hölzl G. Accumulation of novel glycolipids and ornithine lipids in *Mesorhizobium loti* under phosphate deprivation. J Bacteriol. 2015;197:497–509.
24. Shropshire H, Jones RA, Aguilo-Ferretjans MM, Scanlan DJ, Chen Y. Proteomics insights into the *Burkholderia cenocepacia*

phosphorus stress response. Environ Microbiol. 2021. https://doi.org/10.1111/1462-2920.15451.
25. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, Californias: AAAI Press; 1994. p. 28–36.
26. Cobián Güemes AG, Lim YW, Quinn RA, Conrad DJ, Benler S, Maughan H, et al. Cystic fibrosis rapid response: translating multi-omics data into clinically relevant information. mBio. 2019;16:e00431-19.
27. Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH, Kirketerp-Moller K, et al. *Pseudomonas aeruginosa* trascriptome during human infection. Proc Nat Acad Sci USA. 2018;115:E5125–34.
28. Jones HJ, Krober E, Stephenson J, Mausz MA, Jameson E, Millard A, et al. A new family of uncultivated bacteria involved in methanogenesis from the ubiquitous osmolyte glycine betaine in coastal saltmarsh sediments. Microbiome. 2019;7:120.
29. Kim SK, Park SJ, Li XH, Choi YS, Im DS, Lee JH. Bacterial ornithine lipid, a surrogate membrane lipid under phosphate-limiting conditions, plays important roles in bacterial persistence and interaction with host. Environ Microbiol. 2018;20:3992–4008.
30. Lewenza S, Falsafi R, Bains M, Rohs P, Stupak J, Sprott GD, et al. The *olsA* gene mediates the synthesis of an ornithine lipid in *Pseudomonas aeruginosa* during growth under phosphate-limiting conditions, but is not involved in antimicrobial peptide susceptibility. FEMS Microbiol Lett. 2011;320:95–102.
31. Wilton M, Halverson TW, Charron-Mazenod L, Parkins MD, Lewnza S. Secreted phosphatase and deoxyribonuclease are required by *Pseudomonas aeruginosa* to defend against neutrophil extracellular traps. Infect Immun. 2018;86:e00403–18.
32. Wei T, Quareshy M, Zhang YZ, Scanlan DJ, Chen Y. Manganese is essential for PlcP metallophosphoesterase activity involved in lipid remodeling in abundant marine heterotrophic bacteria. Appl Environ Microbiol 2018;84:e01109–18.
33. Zavaleta-Pastor, M, Sohlenkamp, C, Gao, J-L, Guan, Z, Zaheer, R, Finan, TM, et al. *Sinorhizobium meliloti* phospholipase C required for lipid remodeling during phosphorus limitation. Proc Natl Acad Sci USA. 2010;107:302–7.
34. Freschi L, Jeukens J, Kukavica-Ibrulj I, Boyle B, Dupont MJ, Laroche J, et al. Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. Front Microbiol. 2015;29:1036.
35. Ozer EA, Nnah E, Didelot X, Whitaker RJ, Hauser AR. The population structure of *Pseudomonas aeruginosa* is characterized by genetic isolation of *exoU*+ and *exoS*+ lineages. Genome Biol Evol. 2019;11:1780–96.
36. Chugani SA, Greenberg EP. The influence of human respiratory epithelia on *Pseudomonas aeruginosa* gene expression. Microb Pathog. 2007;42:29–35.
37. Frisk A, Schurr JR, Wang G, Bertucci DC, Marrero L, Hwang SH, et al. Transcriptome analysis of *Pseudomonas aeruginosa* after interaction with human airway epithelial cells. Infect Immun. 2004;72:5433–8.
38. Filloux A, Bally M, Soscia C, Murgier M, Lazdunski A. Phosphate regulation in *Pseudomonas aeruginosa*: cloning of the alkaline phosphatase gene and identification of *phoB* and *phoR*-like gnes. Mol Gen Genet. 1988;212:510–3.
39. Poirel L, Jayol A, Nordmann P. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. Clin Microbiol Rev. 2017;30:557–96.
40. Ezadi F, Ardebili A, Mirnead R. Antimicrobial susceptibility testing for polymyxins: challenges, issues, and recommendations. J Clin Microbiol. 2019;57:e01390–18.
41. Semeniuk A, Sohlenkamp C, Duda K, Hölzl G. A bifunctional glycosyltransferase from *Agrobacterium tumefaciens* synthesizes monoglucosyl and glucuronosyl diacylglycerol under phosphate deprivation. J Biol Chem. 2014;289:10104–14.
42. Monds RD, Newell PD, Schwartzman JA, O'Toole GA. Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. Appl Environ Microbiol. 2006;72:1910–24.
43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.

Supplementary material

Niche-adaptation in plant-associated Bacteroidetes favours specialisation in organic phosphorus mineralisation

Ian D.E.A. Lidbury[1,2*+], Chiara Borsetto[1+], Andrew R. J. Murphy[1], Andrew Bottrill[1], Alex Jones[1], Gary D. Bending[1], John P. Hammond[2], Yin Chen[1], Elizabeth M. H. Wellington[1], David J. Scanlan[1]

[1] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, UK

[2] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

[3] School of Agriculture, Policy, and Development, University of Reading, Earley Gate, Whiteknights, Reading, UK

[+]Both are lead authors

***Corresponding author: I.lidbury@sheffield.ac.uk**

145

Supplementary Methods

Transcriptomic analysis of *F. johnsoniae*

RNA was extracted using the RNeasy Mini Kit (QIAGEN) following the manufacturer's instructions. Briefly, equal amounts of cells (1-2 mL culture) according to OD_{600nm} were harvested after 16-20 h of growth for all the conditions tested (Controls, Low Pi, Low NH_4, Low-Fe). RNAprotect® Bacteria Reagent (QIAGEN) was added to the cells as per the manufacturer's instructions prior to centrifugation and flash-freezing to store the stabilised bacterial pellet at -80°C. After thawing the pellet, an enzymatic lysis and a proteinase K digestion step was performed and RNA was recovered using the RNA binding column provided in the kit. RNA concentration was measured using a Nanodrop (ThermoScientific) and genomic contamination was checked by agarose gel electrophoresis prior to DNase treatment when required. Removal of genomic DNA was performed with TURBO DNase (Invitrogen) as per the manufacturer's instructions. All RNA samples were run on the 2100 Bioanalyzer (Agilent) to obtain RNA Integrity Number (RIN) and final concentration. RNA-seq library preparations, sequencing and standard bioinformatics analysis were performed by Novogene according to the company pipeline. For each RNA sample provided, a 250-300 bp insert strand specific library with rRNA removal (Ribo-ZeroTM Magnetic Kit) was prepared and 1 GB of raw sequence data obtained. (Bioproject accession PRJNA635152). For review purposes, the following link can be used to access the data (https://dataview.ncbi.nlm.nih.gov/object/PRJNA635152?reviewer=9nav0b3m2iboav1fiqfkhsvgk4). Differential gene expression results were independently verified using the DEGUST online platform (http://degust.erc.monash.edu/).

Proteomic analysis of *F. johnsoniae*

For each condition tested, three biological replicates were extracted and analysed. For each sample, 25 mL culture was harvested by centrifugation at 10000 g for 15 min at 4°C. Pellets were recovered, flash-frozen and stored at -80°C until further processing. For the whole cell protein extraction, each cell pellet was thawed on ice and resuspended in 2-2.5 mL of 20 mM Tris-HCl pH 7.8. Cells were lysed using a French press (3 cycles/ samples at 1000 psi) and kept on ice during the whole process. For each lysate 1 mL was centrifuged at 14000 rpm 4°C for 15 min and 0.5 mL of the supernatant was recovered and used for protein quantification using the Bradford reagent (Alfa Aesar) as per the manufacturer's instructions. 10 µg protein for each sample were mixed with dithiothreitol (DTT) and RunBlue lithium dodecyl sulphate (LDS) loading buffer (Expedeon), heated at 80°C for 10 min and loaded for SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in a precast RunBlue SDS gel 4-20 % (Expedeon). SDS-PAGE was performed with RunBlue SDS Running Buffer (TEO-Tricine) 1X (Expedeon) at 140 V for either 40 min or 10 min. Gels were stained with Instant Blue (Expedeon). The long run gels were visually checked for consistency between replicates while the short run gels were used for the proteomic run.

For the exoproteomes, cells were pelleted at 4,000 rpm 4ºC in a Beckman benchtop centrifuge prior to gentle hand filtration through 0.45 μm and 0.22 μm pore size PDVF membranes. The remaining supernatant was frozen at -20ºC until further analysis. Culture supernatants were thawed overnight at 4ºC and exoproteins were concentrated following the Trichloroacetic acid precipitation protocol described by Christie-Oleza et al., (2012)[1]. Protein pellets were dissolved in 60 μL Sample Buffer (Expedeon, UK) containing the RunBlue© DTT Reducer (Expedeon, UK) following the manufacturer's guidelines. Exoproteomes were visualised using SDS-PAGE (Expedeon, UK) following the manufacturer's recommended settings. For protein identification a short run (~2 min) was performed to create a single gel band containing the entire exoproteome, as previously described by Christie-Oleza et al., (2012). In-gel reduction was performed prior to trypsin digestion and subsequent clean up as previously described by[1]. Samples were analysed by means of nanoLC-ESI-MS/MS using an Ultimate 3000 LC system (Dionex-LC Packings) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific, USA) using a 60 min LC separation on a 25 cm column and settings as previously described[2]. Quantification, statistical analyses and data visualisation of exoproteomes was carried out in Perseus[3]. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRoteomics IDEntifications (PRIDE) partner repository with the dataset identifier PXD014380 and 10.6019/PXD014380.

Protein fractionation of the soluble *F. johnsoniae* cellular proteome

To extract and enrich/purify proteins with phosphatase activity 500 ml of culture was grown. To obtain whole cell protein extract, cells were pelleted at 12000 rpm, and resuspended in 4 ml loading buffer (20 mM Tris HCl buffer, pH7.4, containing 250 mM NaCl), prior to bursting via French Press. This resulting protein milieu was centrifuged at 25000 rpm for 20 min at 4ºC using a Beckman-Coulter JA25.50 fixed angle rotor to remove any residual cellular debris. Size-exclusion chromatography was used to isolate protein fractions with phosphatase activity. Both size-exclusion and anion-exchange chromatography (SEC and AIEC respectively) were performed using an ÄKTAPURIFIER UPC 10 (Cytiva). SEC used a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) with 120 mL bed volume. Flow rate was 0.5 ml/min for four hours (1x column volume), using the loading buffer as an isocratic buffer. 1ml fractions were collected between 40-120 ml, and phosphatase activity was assayed using PNPP degradation and A405 absorption in a Wallac Victor 96-well plate reader. Fractions making up the two definite phosphatase-active peaks (peak 1 and peak 2, see Fig S6) were taken forward for further purification efforts.

The inherent property of phosphatases to bind the phosphonate moiety, which they cannot cleave was utilised, to further purify phosphatases from the active fractions following SEC. 5 ml EconoColumn (BioRad) gravity columns were loaded with 5 ml L-Histidyldiazobenzylphosphonic acid agarose (Sigma-Aldrich). Briefly, the column was equilibrated with 5x volume loading buffer, protein was loaded onto the column and left to bind for 1 hour. The column was then washed with 5x volume loading buffer to remove unbound protein, and bound protein was then eluted in 1x volume elution buffer 1 (loading buffer + 1M $NaPO_4$). For peak 2, phosphatase activity of fractions was again assayed using PNPP degradation. Peak 1 failed to fractionate correctly using this method.

147

For peak 1, AIEC used a Mono Q 10/100 GL (Sigma-Aldrich) with 8 ml bed volume. Flow rate was 2 ml/min. Protein in loading buffer was loaded onto the column, then the column was washed with 5x volume loading buffer before a 0-50% gradient of elution buffer 2 (20 mM Tris HCl pH7.4, 1M NaCl) was applied over 40 mins (10x volume). This 50% ratio was held for a further 1x column volume before residual protein was washed off with 100% elution buffer 2. Phosphatase activity was again assayed using PNPP degradation.

References

1. Christie-Oleza, J.A. & Armengaud, J. In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case study. *Marine Drugs* **8**, 2223-2239 (2010).
2. Christie-Oleza, J.A., Armengaud, J., Guerin, P. & Scanlan, D.J. Functional distinctness in the exoproteomes of marine *Synechococcus*. *Environmental Microbiology* **17**, 3781-3794 (2015).
3. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731 (2016).
4. Zhu, Y. *et al.* Genetic analyses unravel the crucial role of a horizontally acquired alginate lyase for brown algal biomass degradation by *Zobellia galactanivorans*. *Environmental Microbiology* **19**, 2164-2181 (2017).
5. Kovach, M.E. *et al.* Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175-176 (1995).
6. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* (2016).
7. Kappelmann, L. *et al.* Polysaccharide utilization loci of North Sea Flavobacteria as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *The ISME Journal* **13**, 76-91 (2019).

148

Supplementary Tables Legend (see Supplementary tables excel file,)

Table S1. Transcriptomic analysis of nutrient-limited DSM2064

Table S2. Whole-cell proteomic analysis of nutrient-limited DSM2064

Table S3. Detailed exoproteomic analysis of DSM2064

Table S4. Proteomic analysis of DSM2064 fractionated cellular proteome with strong phosphatase activity (peak 1). Separation of the active fraction was performed using Anion-Exchange Chromatography

Table S5. Proteomic analysis of DSM2064 fractionated cellular proteome with strong phosphatase activity (peak 2). Separation of the active fraction was performed using phosphonate-affinity Chromatography

Table S6. Genome characteristics of the *Flavobacterium* isolates used in this study

Table S7. Detailed exoproteomic analysis of F52

Table S8. Detailed exoproteomic analysis of LOA-5

Table S9. Detailed exoproteomic analysis of OSR001

Table S10. Detailed exoproteomic analysis of OSR003

Table S11. Detailed exoproteomic analysis of OSR004

Table S12. Detailed exoproteomic analysis of OSR005

Table S13. Detailed exoproteomic analysis of TSA_1_4_3

Table S14. Comparative genomics of ~100 *Flavobacterium* strains

Table S15. Detailed taxon list of Bacteroidetes isolates (n=468)

Table S16. Corresponding amino acid sequences for all locus tags within the OSR001 genome

Table S17. Corresponding amino acid sequences for all locus tags within the OSR002 genome

Table S18. Corresponding amino acid sequences for all locus tags within the OSR003 genome

Table S19. Corresponding amino acid sequences for all locus tags within the OSR004 genome

Table S20. Corresponding amino acid sequences for all locus tags within the OSR005 genome

Table S21. Corresponding amino acid sequences for all locus tags within the TSA_1_4_3 genome

Table S22. Corresponding amino acid sequences for all locus tags within the LOA-5 genome

149

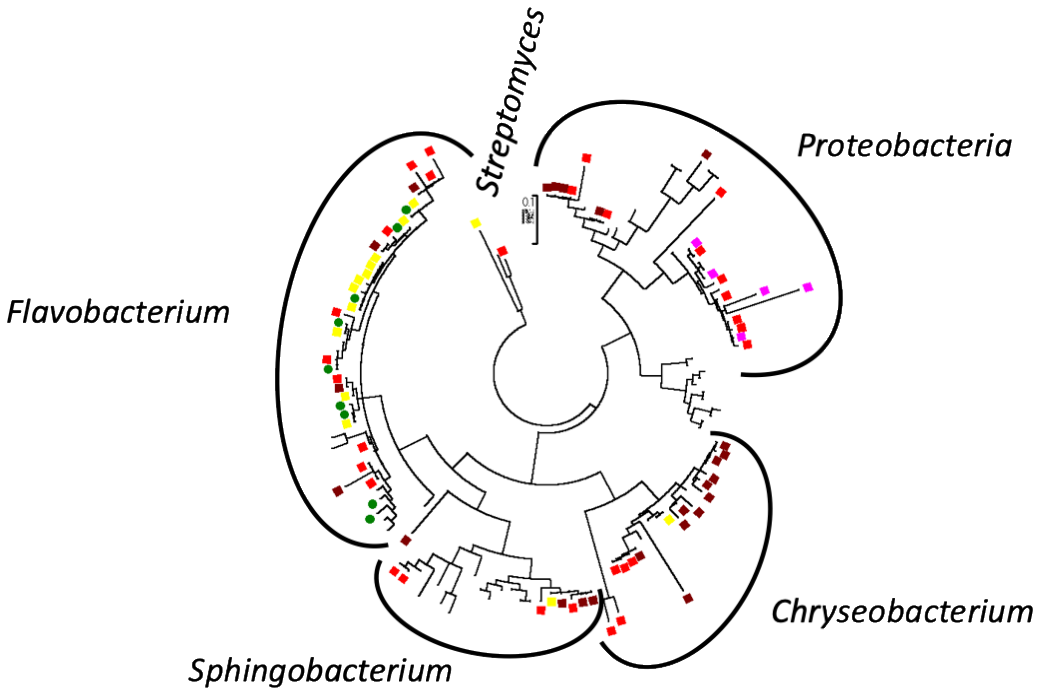


Figure S1. Taxonomy of rhizobacterial strains isolated from field-grown *Brassica napus* using a semi-selective Bacteroidetes medium based on the 16S rRNA gene marker. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [1]. The tree with the highest log likelihood (-10495.28) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 134 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions with less than 75% site coverage were eliminated. That is, fewer than 25% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 433 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [2]. Green circles represent strains used in this study. Coloured squares denote the various sampling efforts.

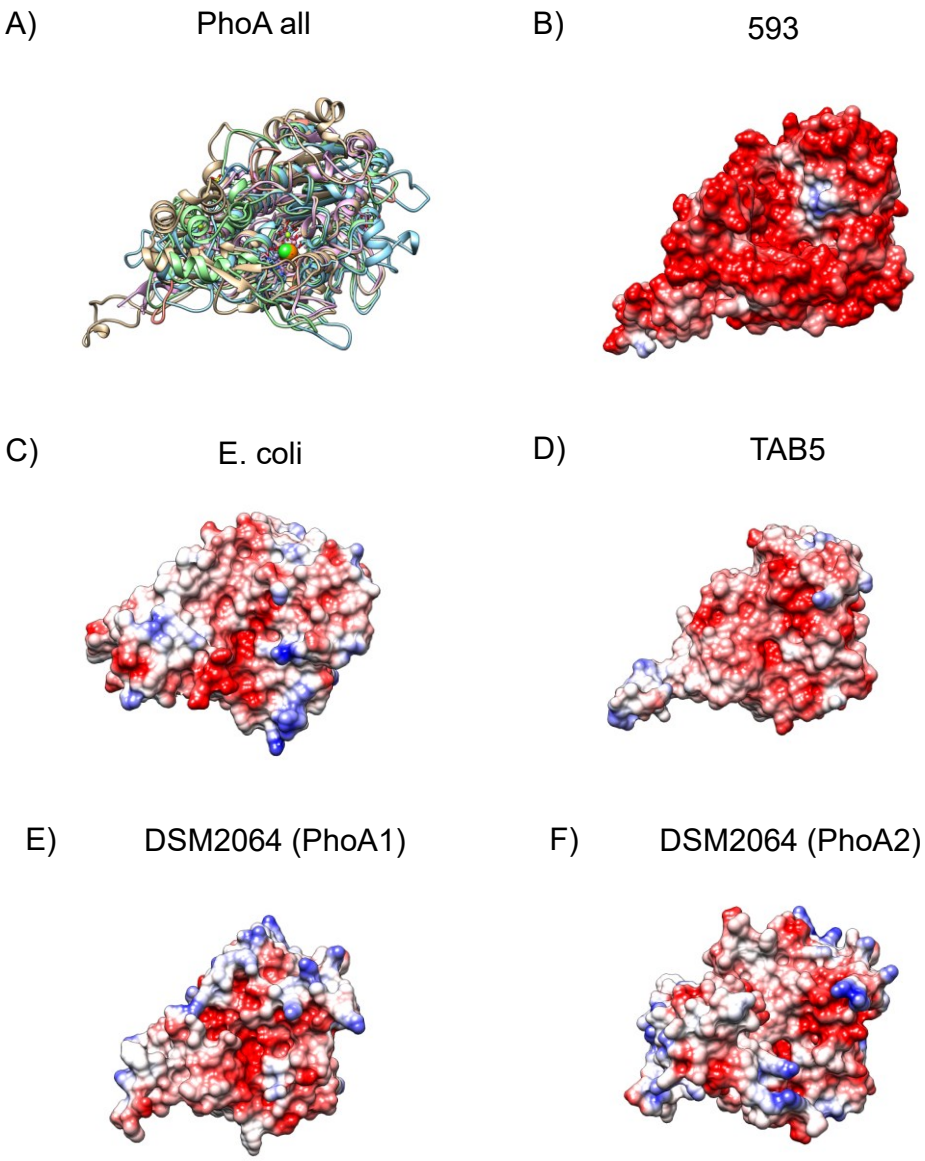


Figure S2. Structural homology modelling of DSM2064 PhoA1 and PhoA2. (A) Superimposition of 3D models of *Halomonas* sp. 593 PhoA (cream, 3WBH) *E. coli* PhoA (sky blue, 1Y6V), antarctic bacterium TAB5 PhoA (light pink, 2IUC), DSM2064 PhoA2 (light green), DSM2064 PhoA1 (pale red). **(B-F)** Surface electrostatic potential for each model was calculated using the Coulombic scaling method. PhoA1 and PhoA PDB files were generated using PHYRE 2 and models were further analysed using CHIMERA (UCSF) V 1.14.

151

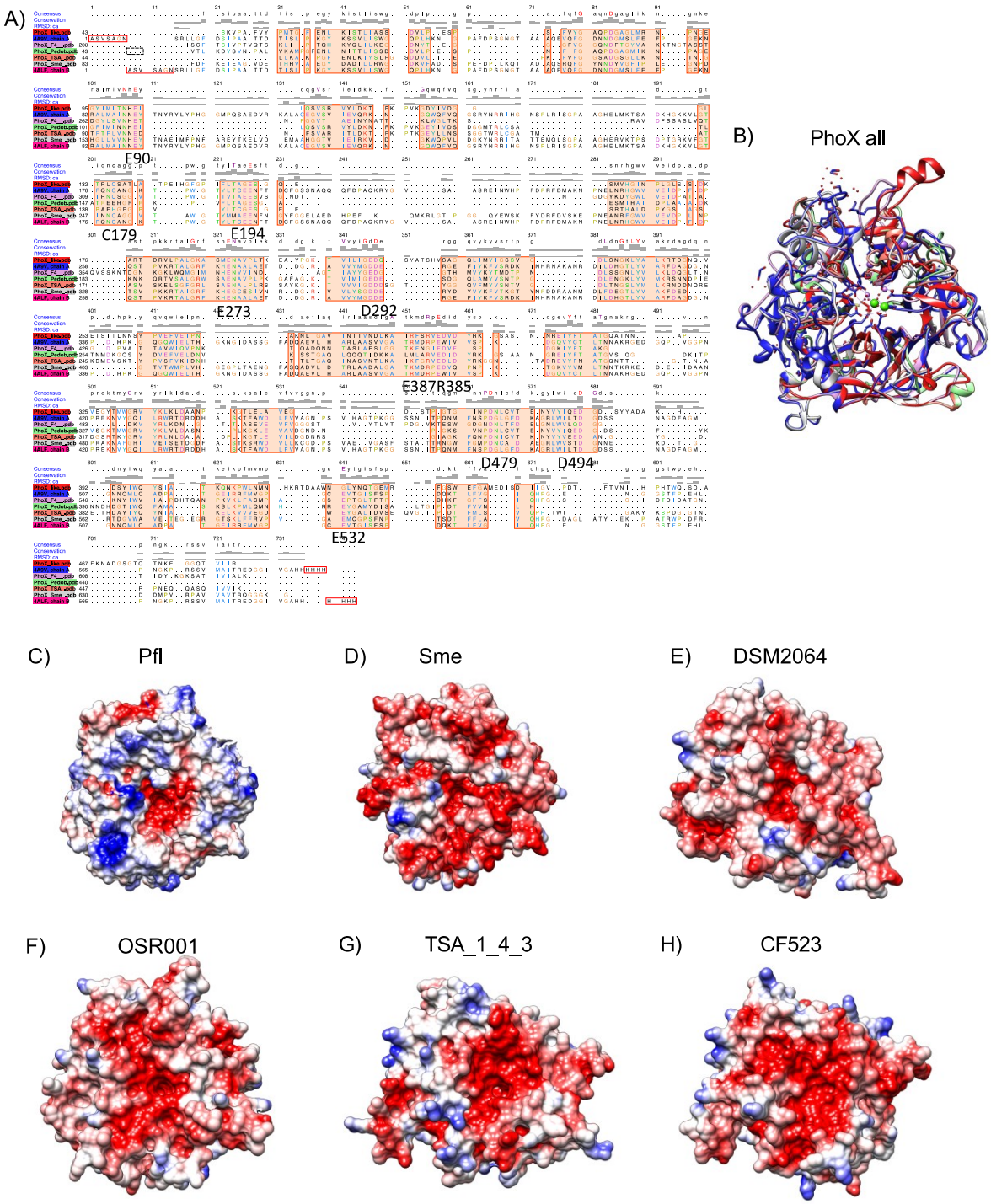


Figure S3. Structural homology modelling of Bacteroidetes PhoX. A) Multi-alignment based on structural superimposition of *P. fluorescens* PhoX (4A9V/4ALF, blue/magenta) with DSM2064 PhoX (red), OSR001 PhoX (light pink), *Pedobacter* sp. CF523 (green), TSA_1_4_3 PhoX (orange), *Sinorhizobium meliloti* SM11 (grey). The 10 key residues are also highlighted. **(B)** Superimposition of 3D models of *P. fluorescens* PhoX (Pfl) with *S. meliloti* SM11 (Sme) DSM2064 PhoX , OSR001 PhoX, *Pedobacter* sp. CF523 (CF523), TSA_1_4_3 PhoX (orange). **(C-F)** Surface electrostatic potential for each model was calculated using the Coulombic scaling method. *Flavobacterium* and *Sme* PhoX PDB files were generated using PHYRE 2 and models were further analysed using CHIMERA (UCSF) V 1.14.

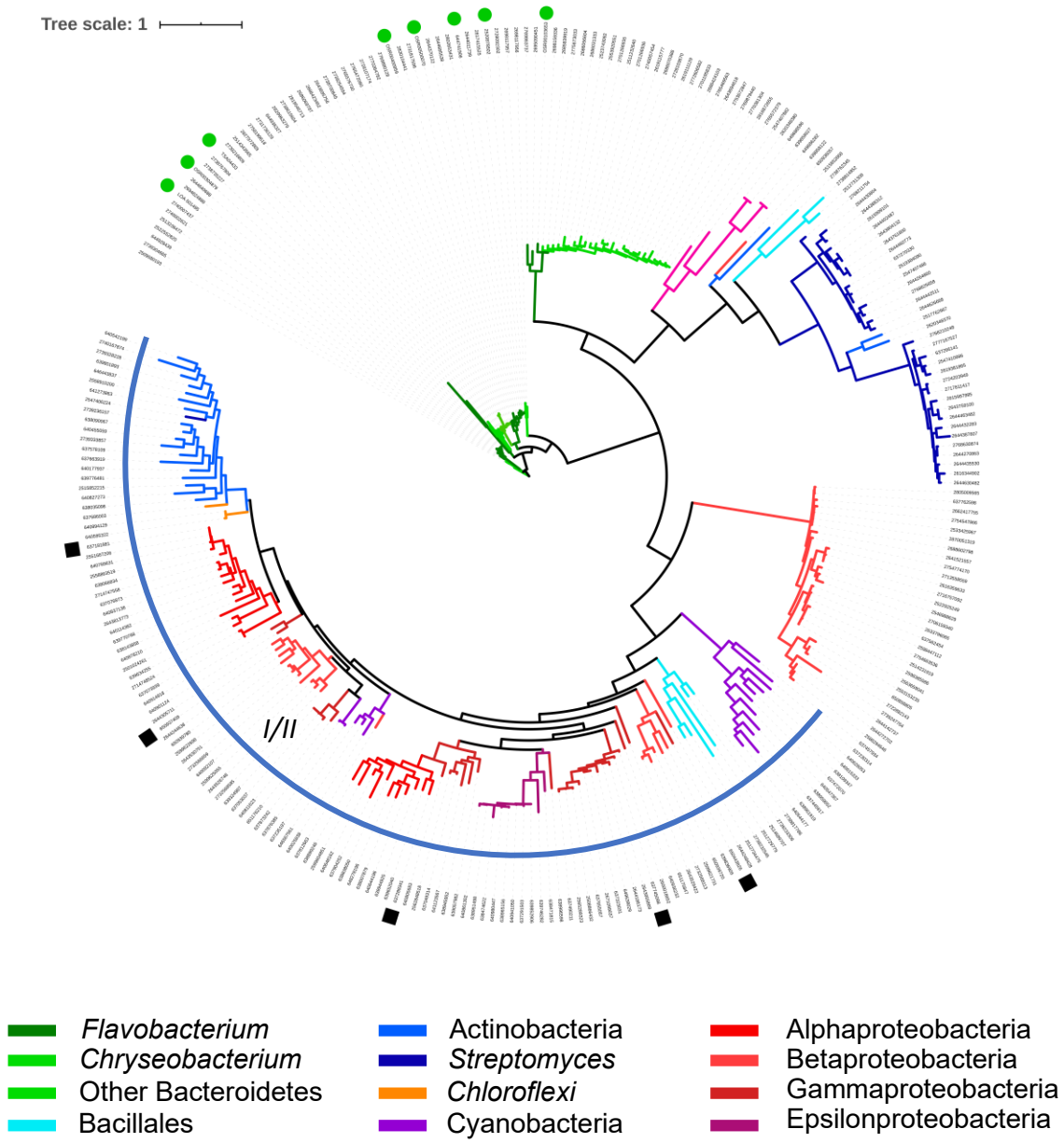


Figure S4. Phylogenetic comparison of PhoX found across all bacterial taxa. Tree topology and branch lengths were calculated by maximum likelihood using the WAG+I+G4 model of evolution for 277 amino acid sequences based on sites in lQ-TREE software. Trees were visualized and annotated using the Interactive Tree of Life (ITOL) online server. The blue truncated ring (I/II) represents PhoX clades I and II. Black squares represent structurally or genetically characterised PhoX homologs. Green circles represent PhoX homologs found in the *Flavobacterium* isolates used in this study. IMG accession numbers for each homologs are given at the end of each leaf.

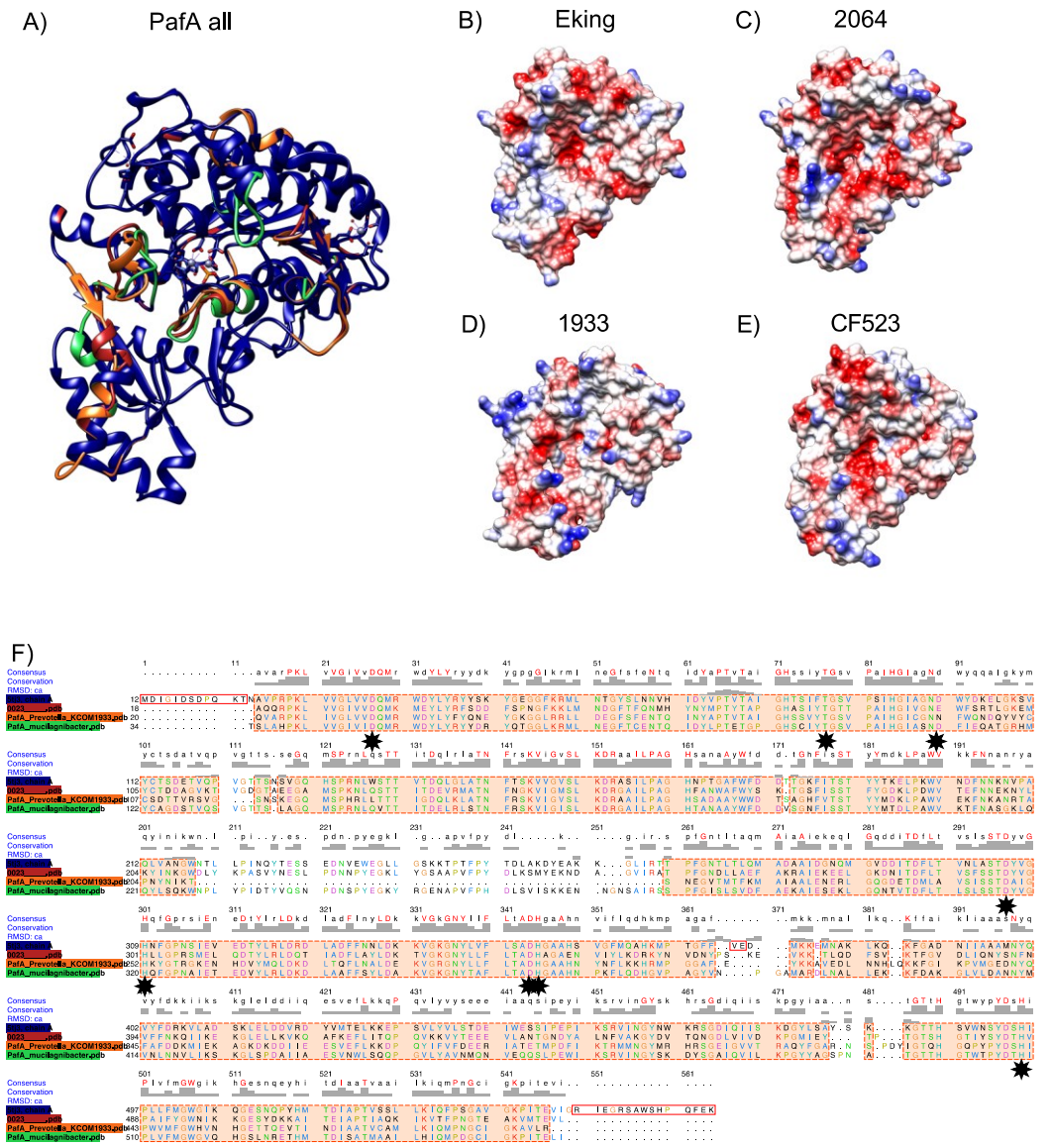


Figure S5. Structural homology modelling of Bacteroidetes PafA. (A) Superimposition of 3D models of *Elizabethkingia meningosepticum* PafA (Eking) with DSM2064 PafA (2064), *Prevotella intermedia* KCOM 1933 (1933), *Mucilaginibacter polytrichastri* DSM 26907 (CF523). **(D-E)** Surface electrostatic potential for each model was calculated using the Coulombic scaling method. **F)** Multi-alignment based on structural superimposition of Eking PafA (navy blue, 5TJ3), DSM2064 PafA (red), 1933 PafA (orange), CF523 PafA (green). The 10 key residues are also highlighted. Bacteroidetes PafA PDB files were generated using PHYRE 2 and models were further analysed using CHIMERA (UCSF) V 1.14.

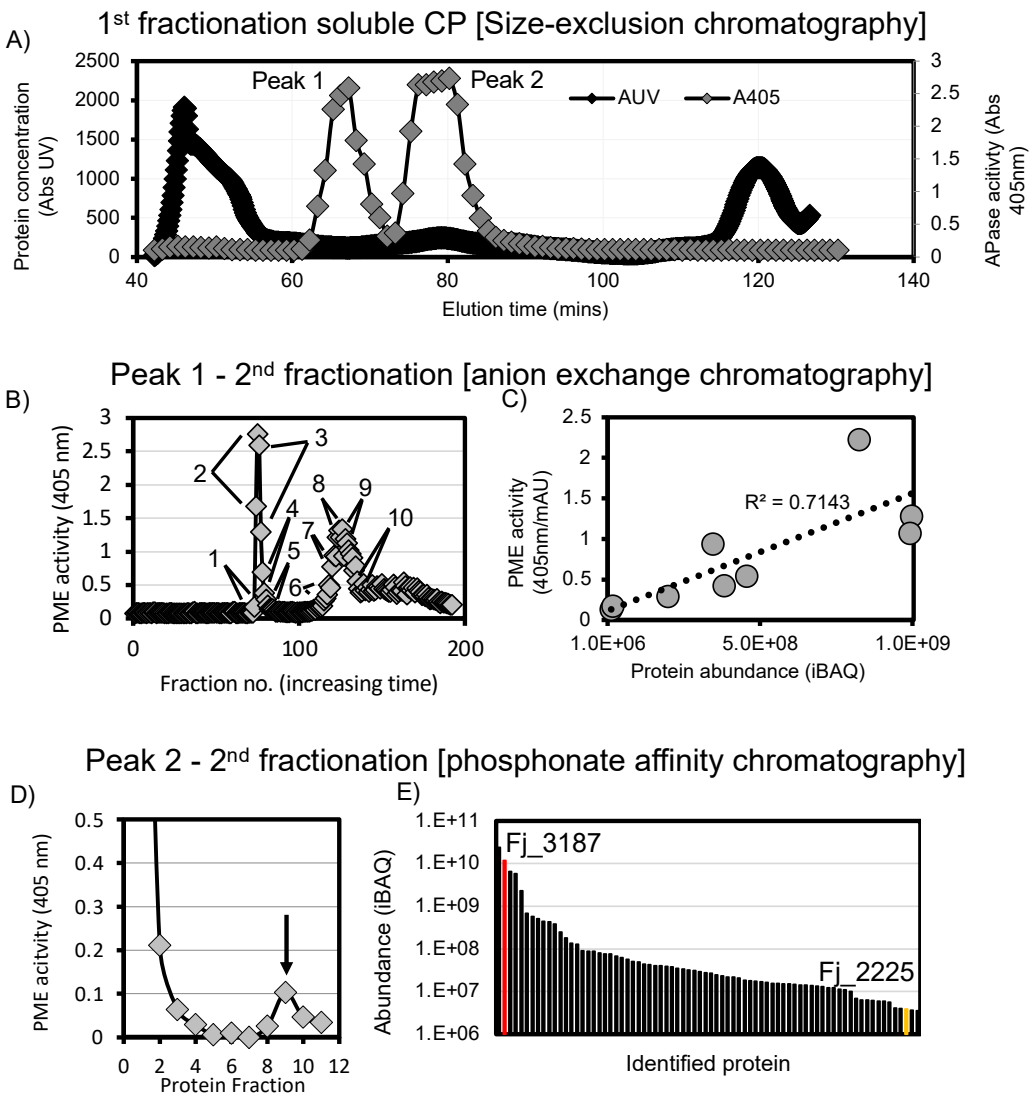


Figure S6. Biochemical analysis of DSM2064 cellular proteome (CP) (A) Size-exclusion chromatogram of low Pi-grown DSM2064 cellular proteome (soluble fraction only) followed by screening for phosphomonoesterase (PME) activity. **B)** PME activity screen for protein fractions separated by anion exchange (Na^{2+}) chromatography. Protein fractions from Peak 1 were used as the starting material for the second fractionation step. To obtain enough protein, two fractions (see arrows) were merged together prior to peptide identification. Two peaks for PME activity were detected across the fractions. **(C)** Linear regression of PME activity predicted from Fj_0023 (PafA) protein abundance. Dots represent observed data points and the dotted line represents the linear regression model. PME activity plotted represents the average activity of two fractions. PafA abundance correlated with PME activity across both peaks, suggesting this enzyme is responsible for the biochemical activity **(D)** PME activity screen for protein fractions separated by phosphonate (L-Histidyldiazobenzylphosphonic acid-Agarose). Protein fractions from peak 2 were used as the material for the second fractionation step. The majority of protein failed to bind to the column and eluted in the flow through (highest PME activity). The arrow denotes the fraction sent for peptide identification **(E)** Abundance (iBAQ) of the individual proteins detected in the fraction 9 (arrowed on panel D), Fj_3187 refers to PhoA2 and Fj_2225 refers to one of the inducible periplasmic Apases described in Table 1.

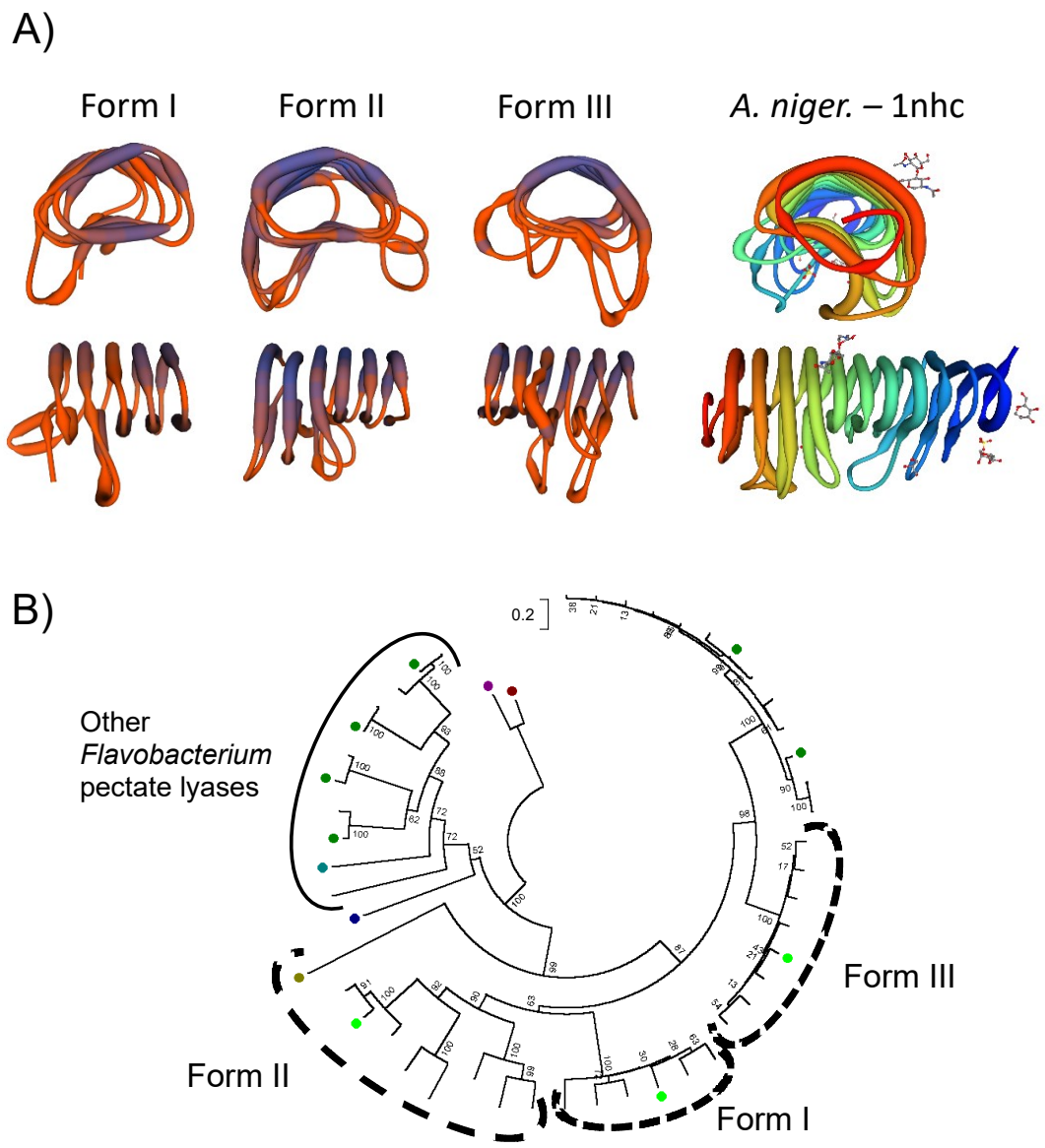


Figure S7. Analysis of high-expressed Pi-inducible hypothetical lipoproteins, FI-FIII A) Structural modelling of the abundantly secreted Pi-sensitive hypothetical exoproteins FI, FII and FIII. SWISS homology modelling revealed that these hypothetical exoproteins share structural homology with pectate lyases/characterised polygalacturonases found in Bacteria and Fungi. **(B)** Maximum-likelihood tree displaying the phylogeny of low-Pi inducible hypothetical exoproteins FI, FII, and FIII (light green represents DSM2064). The previously characterised classical polygalacturonases (Maroon, Dark Blue and Mustard) and other predicted polygalacturonases found in DSM2064 and F52 (Dark green) were also added to the analysis.

Transporter characterisation reveals aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle

Andrew R. J. Murphy[1], David J. Scanlan[1], Yin Chen[1], Nathan B. P. Adams[2,3], William A. Cadman[2], Andrew Bottrill[1], Gary Bending[1], John P. Hammond[4], Andrew Hitchcock[2], Elizabeth M. H. Wellington[1], Ian D.E.A. Lidbury[1,5*]

[1]School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, UK

[2]Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK

[3]Nanotemper Technologies GmbH, Flößergasse 4, 81369, Munich, Germany

[4]School of Agriculture, Policy, and Development, University of Reading, Earley Gate, Whiteknights, Reading, UK

[5]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

***Corresponding author: I.lidbury@sheffield.ac.uk**

Table of contents:

Supplementary Methods

Generation and complementation of *Pseudomonas* mutants

Regions of genomic DNA at the 5' and 3' end of each deletion target were amplified, along with in some cases the gentamicin resistance cassette from p34S-Gm[1]. Other mutants were constructed to be marker-less, and so lacked this resistance cassette. DNA fragments were ligated into linearised pk18mob*sacB*[2] using the HiFi DNA Assembly Kit (New England Biolabs, Hitchin, UK) according to the manufacturer's instructions. For the construction of plasmids for complementation, genes and their corresponding promoters were cloned into linearised pBBR1McS-km again using the HiFi DNA Assembly Kit.

Plasmids were introduced into *Escherichia coli* S17.1 by electroporation and mobilized into *Pseudomonas* by conjugation. Transconjugants were selected with gentamicin (50 µg ml^{-1}) or kanamycin (50 µg ml^{-1}), and chloramphenicol (10 µg ml^{-1}) was used for counter-selection. Single crossover transconjugants were identified by polymerase chain reaction (PCR), and double crossover mutants selected via plating on LB containing either gentamicin or no antibiotic, and 10% (w/v) sucrose. Homologous recombination was confirmed by PCR and Sanger sequencing. Complemented mutants were selected using kanamycin with chloramphenicol counter-selection.

Enrichment of membrane-associated protein fraction in *P. putida* BIRD-1

40 ml cells were grown to an OD_{600} of 0.8-1.0. A volume equivalent to 5 OD_{600} units was centrifuged at 2000 x *g* for 10 mins, resuspended in 1 ml 50 mM Tris-HCl, pH 7.6, and centrifuged at 2000 x *g* for 10 mins. Pellets were resuspended in 500 µl 200 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.6 and incubated for 30 mins at 30°C with gentle shaking. Cells were cooled on ice for 5 mins and incubated at room temperature for 15 mins, then centrifuged at 8000 x *g* for 10 mins at 4°C. Pellets were washed in 1 ml 50 mM Tris-HCl, pH 7.6, and again centrifuged at 8000 x *g* for 10 mins at 4°C. Pellets were resuspended in 0.5 ml 50 mM Tris-HCl, pH 7.6, and sonicated for 30 s twice, on ice. The solution was centrifuged at 2000 x *g* for 15 mins at 4°C, and supernatants were then ultracentrifuged at 120000 x *g* for 45 mins at 4°C. Following this, the supernatant was discarded and the pellet resuspended in 50 µl LDS buffer (Expedeon) prior to loading 20 µl onto a 4-20% Bis-Tris (2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol) SDS precast gel (Expedeon).

Detection of Pi efflux

Culture aliquots were centrifuged at 16200 x *g* for 5 mins, and phosphate concentrations in the supernatants were determined according to the method of Chen *et al*.[3] using the modification developed by Christianson and Dunham[4]. Briefly, supernatant was added 1:1 to a 2:1:1:1 solution of dH_2O:6N sulfuric acid:2.5% w/v ammonium molybdate:10% w/v ascorbic acid, incubated at 37 °C for 90 mins, and absorbance was read at 820 nm.

Production, purification and characterisation of recombinant *S. stellulata* AepX

The *S. stellulata* DSM 5886 *aepX* gene lacking the sequence encoding the predicted N-terminal signal peptide (28 aa) and stop codon was amplified from *S. stellulata* DSM 5886 genomic DNA using primer pair pET21:AepXSs_Fwd and pET21:AepXSs_Rev (Supplementary table 4). The resulting product was inserted between the NdeI and XhoI sites of pET21a(+) (Novagen) using the HiFi DNA Assembly Kit (New England Biolabs, Hitchin, UK) according to the manufacturer's instructions. Plasmid

pET21a::*aepX*^{Ss} was sequence verified by automated DNA sequencing (Eurofins, Germany) and introduced to *E. coli* BL21(DE3) for over-production of recombinant AepX with a C-terminal hexahistidine tag.

E. coli BL21(DE3) harbouring pET21a::*aepX*^{Ss} was grown in LB broth with 100 µg mL^{-1} ampicillin in a culture volume of 1 L contained within a 2 L baffled flask. Following inoculation from a 5 mL overnight starter culture, cells were grown at 37 ^{o}C and 250 rpm shaking until they reached an OD_{600} of ~0.6, at which point isopropyl β-D-1-thiogalactopyranoside was added to final concentration of 0.4 mM and the culture was incubated shaking (250 rpm) overnight at 18 ^{o}C. Cells were harvested by centrifugation at 5000 x *g* for 15 mins at 4 ^{o}C and resuspended in ~50 mL of 25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer pH 7.5 containing 500 mM NaCl and 5 mM imidazole (binding buffer). A small spatula of DNase I was added to the suspension and cells were broken by sonication on ice (6 × 30 s bursts with 30 s intervals between) prior to removal of insoluble debris by centrifugation at 56000 x *g* for 20 min at 4 ˚C. Proteins were purified by immobilised Ni-affinity chromatography on a 5 ml Chelating SepharoseTM Fast Flow resin column (GE Healthcare, UK). Bound protein was washed with binding buffer containing 20 mM imidazole and then eluted with 25 mM HEPES pH 7.5 containing 100 mM NaCl and 400 mM imidazole. Eluted protein was concentrated to a volume of 2 mL in a Vivaspin centrifugal concentrator and AepX was further purified on a 22 mL Superdex S200 Increase gel filtration column (GE Healthcare) in 20 mM HEPES pH 7.5 containing 200 mM NaCl. AepX eluted as a single peak and was adjudged pure by SDS-PAGE analysis. Pooled peak fractions were stored at 4 ˚C.

The binding affinity of AepX for phosphonate ligands was determined using Microscale Thermophoresis (MST) with a Monolith NT.115 instrument (NanoTemper Technologies, Germany). AepX (100 nM) was labelled with 2nd Generation RED-tris-NTA dye (Nanotemper Technologies, Germany) in 20 mM HEPES pH 7.5 with 200 mM NaCl and 0.005% (v/v) Tween-20 (MST buffer) according to the manufacturer's instructions. Ligand stocks (purchased from Sigma) were prepared at appropriate concentrations in MST buffer. A 10 µl, 16 point, 2-fold ligand dilution series was prepared, and subsequently mixed with 10 µl of labelled protein. 4 µL of each protein-ligand mixture was loaded into Monolith NT.115 Premium Capillaries (Nanotemper Technologies, Germany) and thermophoresis was measured at 22 ˚C for 22 s with 60 % LED power and High MST power. Data from at least two independent titrations were combined and analysed using the MO.Affinity Analysis software version 2.3 (Nanotemper Technologies, Germany), using the early thermophoresis data at 1 – 2.5 seconds in the experiment for analysis. Isotherms were fitted to a single binding site model (Equation 1) where [l] is concentration of ligand and data plotted using Igor Pro version 8.04 (Wavemetrics Inc., USA).

$$f(\mathrm{l}) = \mathrm{Unbound} + \frac{(\mathrm{Bound}-\mathrm{Unbound})\times([\mathrm{l}]+[\mathrm{protein}]+K_d-\sqrt{([\mathrm{l}]+[\mathrm{protein}]+K_d)^2-4\times[\mathrm{l}]\times[\mathrm{protein}]}}{2\times[\mathrm{protein}]}$$

Equation 1. Single binding site model used to determine dissociation constants.

References

1. Dennis, J.J. and G.J. Zylstra, *Plasposons: Modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes.* Applied and Environmental Microbiology, 1998. **64**(7): p. 2710-2715.
2. Schäfer, A., et al., *Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum.* Gene, 1994. **145**(1): p. 69-73.
3. Chen, P.S., T.Y. Toribara, and H. Warner, *Microdetermination of Phosphorus.* Analytical Chemistry, 1956. **28**(11): p. 1756-1758.
4. Christianson, C. and M. Dunham. *Phosphate Assay*. 2005; Available from: https://dunham.gs.washington.edu/MDphosphateassay.htm.

Supplementary Figures

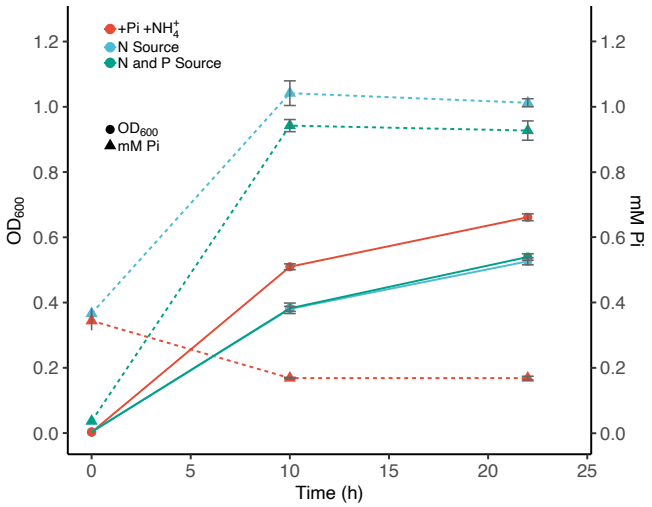


Figure S1. Growth of *Pseudomonas putida* BIRD-1 using 2AEP as the sole N or sole N and sole P source. Pi efflux (triangles and dashed lines) and growth (circles and solid lines) on 2AEP as the sole N (blue) or sole N and sole P source in *P. putida* BIRD-1 (green). A positive control was also established where orthophosphate (Pi) and ammonium (NH_4^+) were the sole P and N sources respectively (red). All conditions were performed in triplicate; error bars represent the standard deviation of the mean.

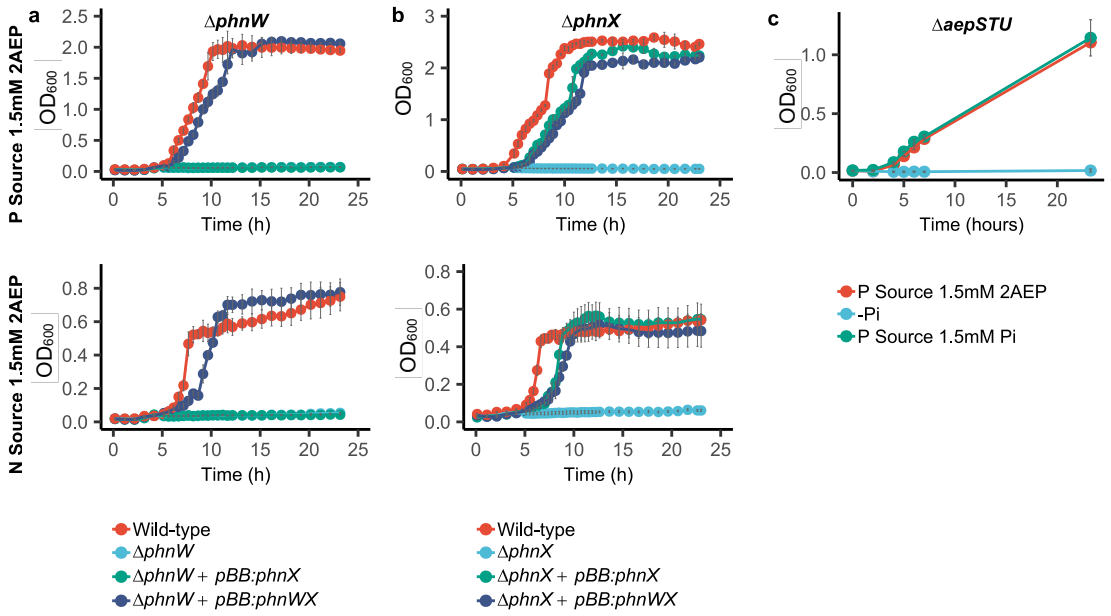


Figure S2 Growth curves of WT and mutant strains of *Pseudomonas putida* BIRD-1 on 2AEP as sole N or P source. (**a**) Shows WT, Δ*phnW* and Δ*phnW in trans* complemented with *phnW(X)* grown with 1.5 mM 2AEP as the sole P source (upper panel) or N source (lower panel). (**b**) Shows WT, Δ*phnX* and Δ*phnX in trans* complemented with *phnWX* grown with 1.5 mM 2AEP as the sole P source (upper panel) or N source (lower panel). (**c**) Shows Δ*aepSTU* grown in the absence of Pi or with 1.5 mM 2AEP or Pi as sole P source. Error bars show standard deviations of the mean from four biological replicates.

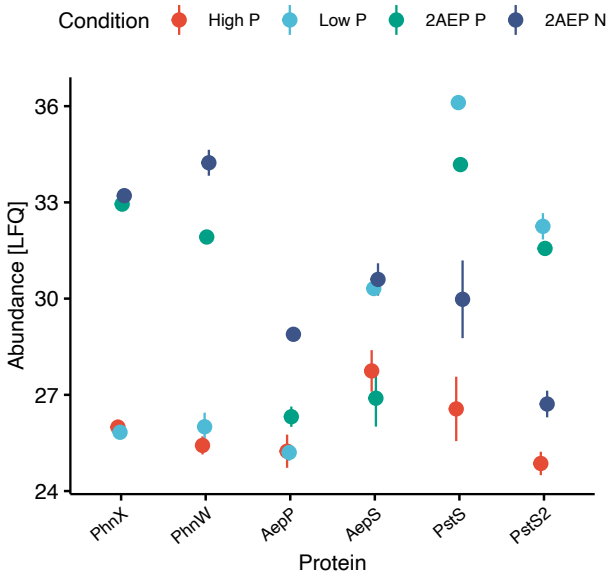


Figure S3. Assessment of *Pseudomonas putida* BIRD-1 membrane-associate protein fraction. Quantification (LFQ_{Log2} values) of PhnWX, candidate 2AEP transport proteins and phosphate binding proteins (PstS) in Δ*aepXVW*::*gm* in response to growth on 2AEP (n=3) as sole P or N source. High and low P controls used orthophosphate (Pi) as the sole P source. Error bars show standard deviation of the mean.

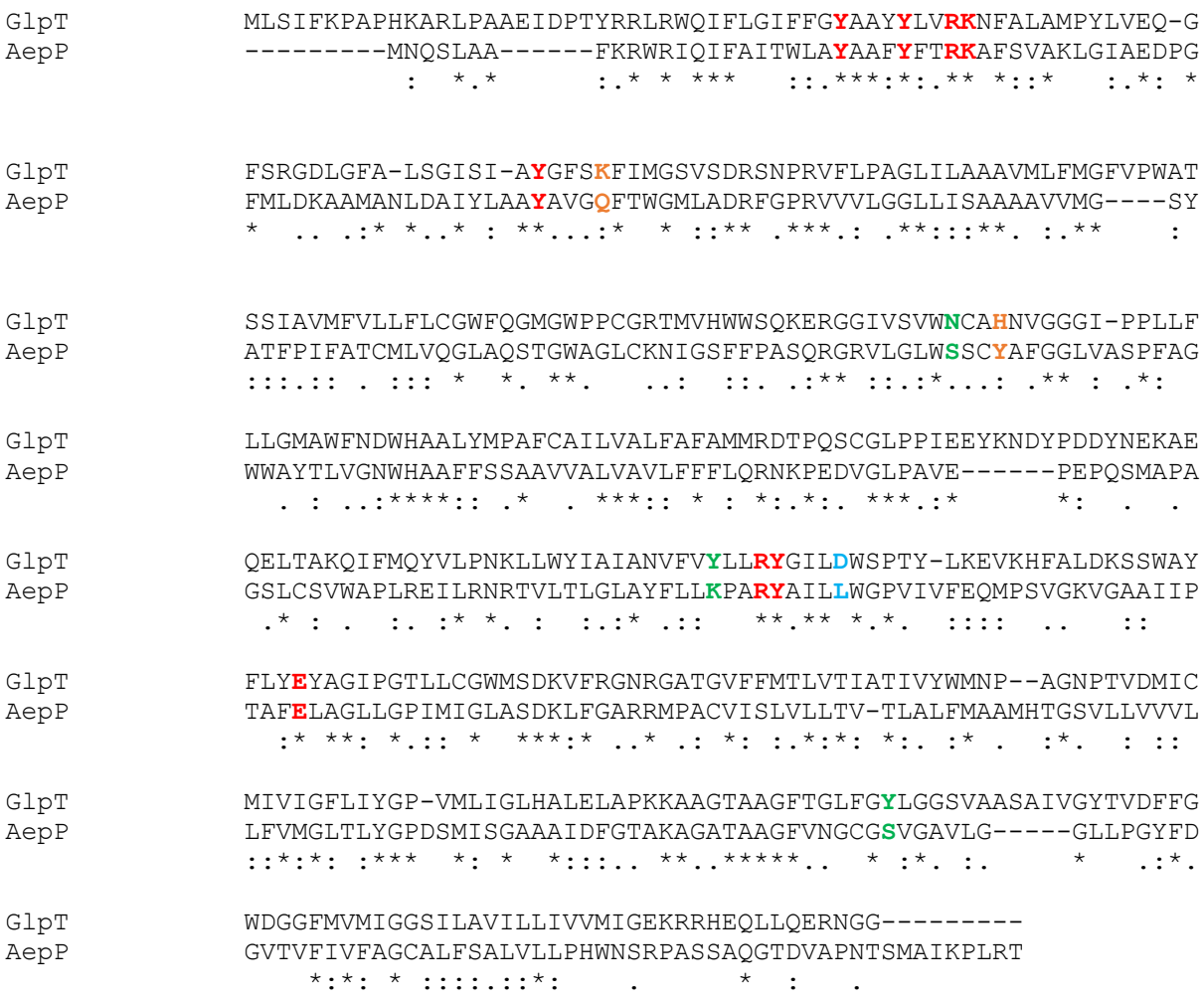


Figure S4. ClustalOmega alignment of GlpT, the glycerol-3-phosphate/phosphate (G3P/Pi) antiporter, and AepP protein sequences. Residues that have been shown to bind with the phosphate moiety of G3P, and/or Pi are shown in red if identical, in orange if related, and in blue if unrelated. Residues involved in binding the glycerol moiety of G3P are shown in green.

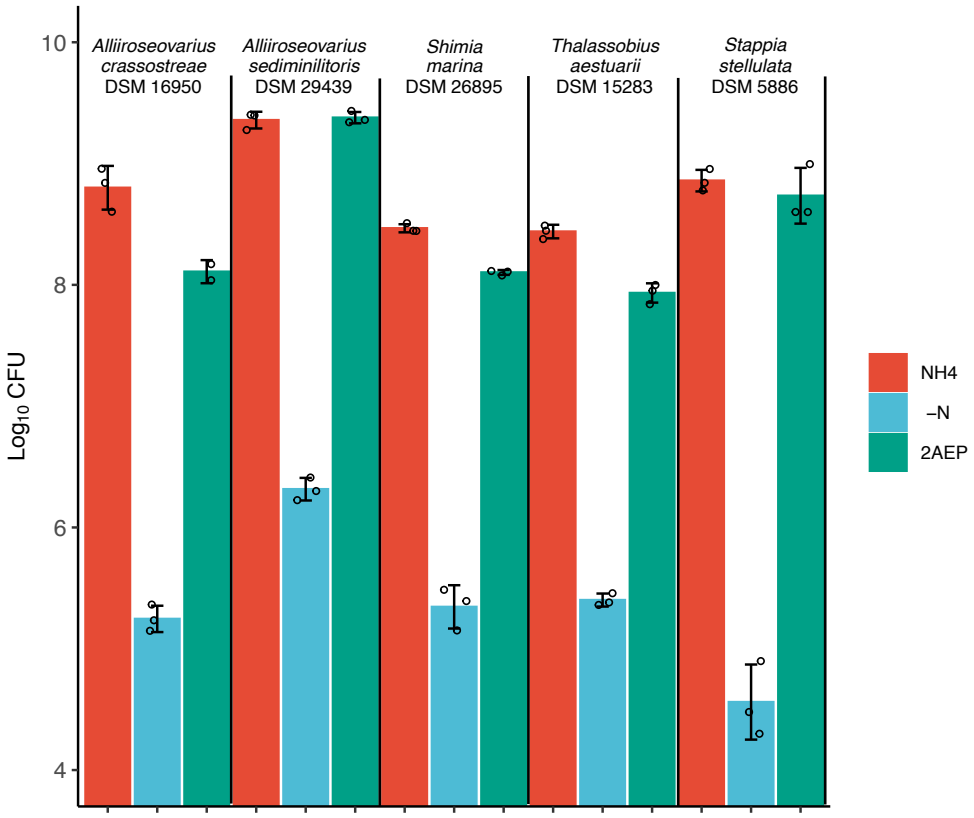


Figure S5. Enumeration of *Roseobacter* strains grown on 2AEP as a sole nitrogen source. Colony forming units were obtained for various *Roseobacter* strains grown on either 1.5 mM NH_4 or 1.5 mM 2AEP in addition to a negative control containing no exogenous nitrogen source. Cultures were sampled between 36-60 hours growth. Results presented are the mean of triplicate cultures. Error bars denote standard deviation.

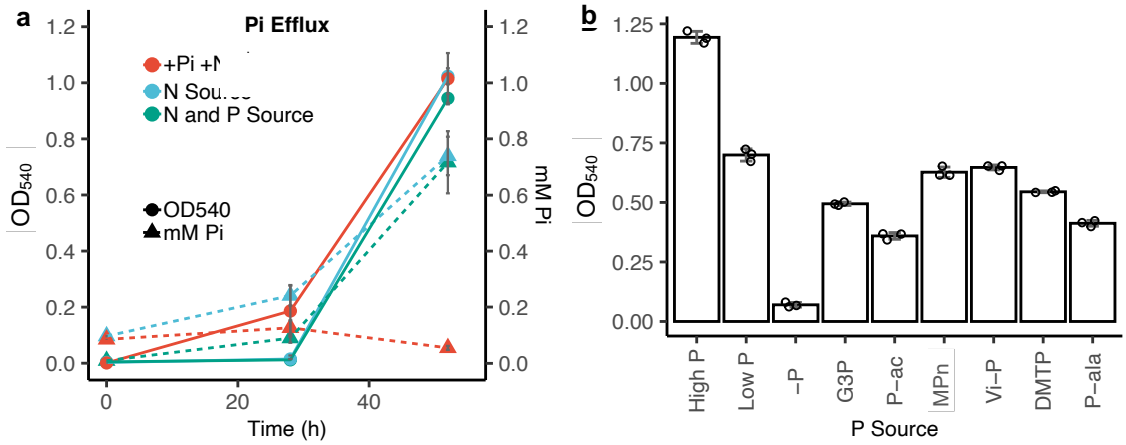


Figure S6. Growth of *Stappia stellulata* DSM5886 on various phosphonate compounds. (**a**) Pi efflux (triangles and dashed lines) and growth (circles and solid lines) on 2AEP as the sole N source (blue), or sole N and P source (green). A positive control was also established where orthophosphate (Pi) and ammonium (NH_4^+) were the sole P and N sources respectively (red). All conditions were performed in triplicate; error bars represent the standard deviation of the mean. (**b**) Growth on various P sources. High P = 1 mM Pi, Low P = 0.1 mM Pi, -P = No P Control, G3P = Glycerol-3-phosphate, P-ac = Phosphonoacetate, MPn = Methylphosphonate, Vi-P = Vinylphosphonate, DMTP = Dimethylthiophosphonate, P-ala = Phosphonoalanine. All organic P sources are 0.1 mM. Values are the mean of three replicates; error bars are the standard deviation of the mean.

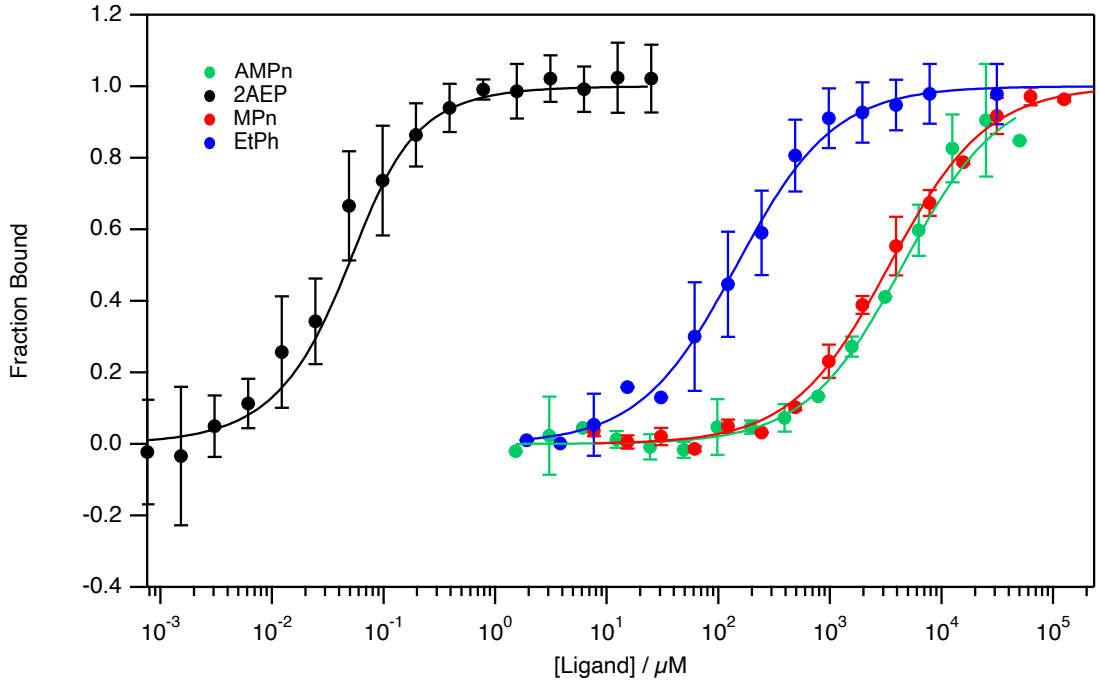


Figure S7. Binding affinities for selected phosphonates by the *Stappia stellulata* AepX substrate binding protein determined by microscale thermophoresis. Results presented are the mean of n=3 (2AEP) or n=2 (MPn, methylphosphonate; EtPh, ethylphosphonate; AMP, aminomethylphosohonate) titrations and error bars denote standard deviation. The fraction of ligand bound to protein was calculated using the MO.affinity analysis software (Nanotemper Technologies, Garmany) by converting Fnorm values.

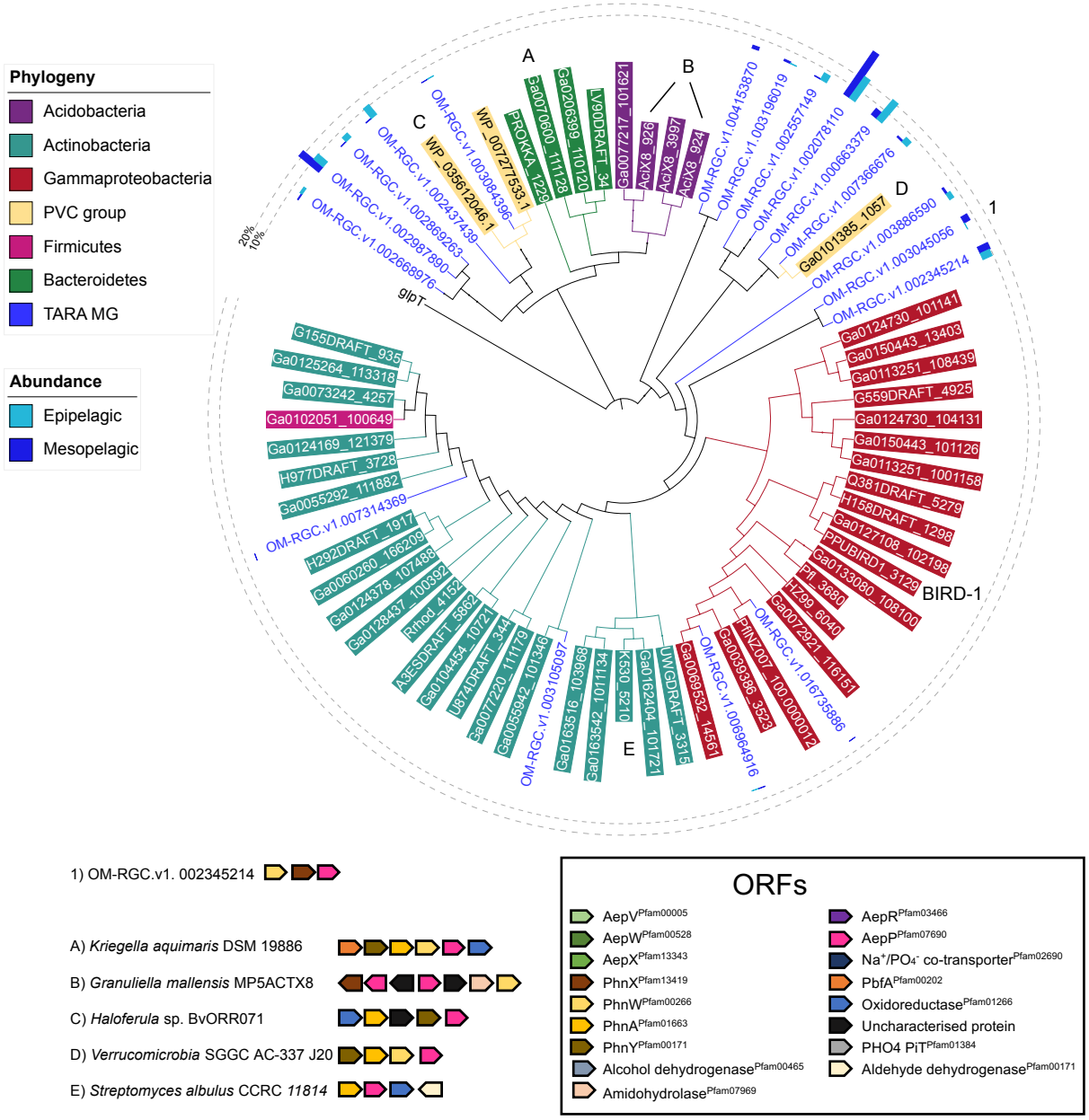


Figure S8. Phylogenetic and genomic analyses of AepP in marine and terrestrial bacteria. The genetic neighbourhood for selected *aepP* homologs is presented adjacent to the tree. Numbers indicate environmental OTUs and letters indicate isolates or MAGs/SAGs. Tree topology and branch lengths were calculated by maximum likelihood using the LG+F+I+G4 model of evolution for amino acid sequences based on 301 sites in IQ-TREE software. A consensus tree was generated using 1000 bootstraps. Branches representing isolates or MAGs/SAGs are colour-coded based on their phylogenetic affiliation (see legends). Branches and identifiers for representative environmental OTU sequences (clustered at 0.8) retrieved from the TARA Oceans database (blue) are also highlighted. The outer ring denotes the relative abundance of environmental OTUs using the same colour scheme; relative abundance values of 10% (dashed line) and 20% (filled line) are shown for scale; *P. putida* BIRD-1 AepP is labelled.

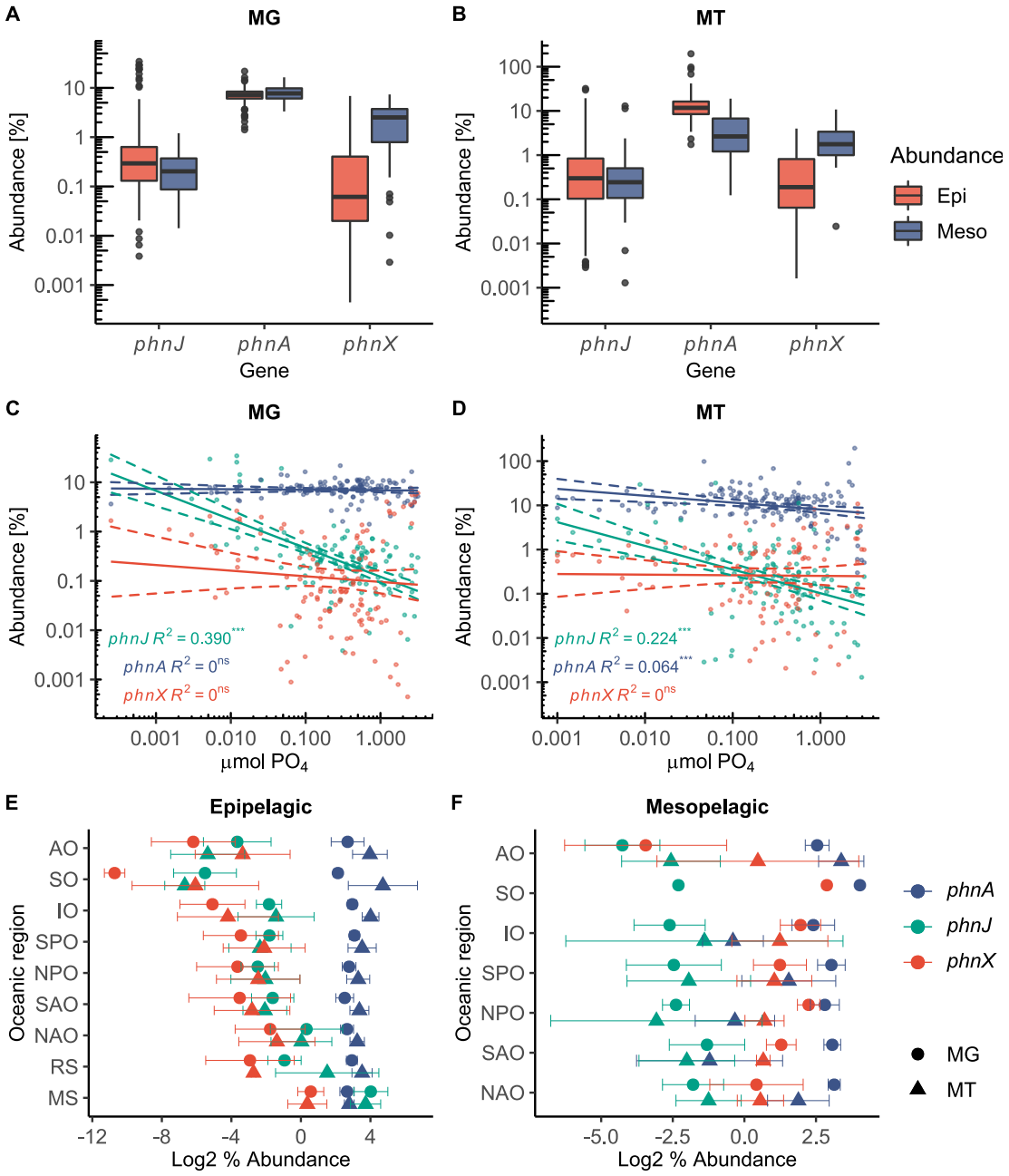


Figure S9. Metagenomic (A) and metatranscriptomic (B) abundance of marker genes for 2AEP degradation within the TARA ocean dataset. Data is split between epipelagic (Epi – Red) and mesopelagic (Meso – Blue) waters. Abundance is normalised as a percentage of the median abundance of 10 single copy marker genes within either OM-RGCv2+G (MG) or OM-RGCv2+T (MT). MG (**A**) epipelagic: *phnJ n* = 137, *phnA n* = 137, *phnX n* = 126, mesopelagic: *phnJ n* = 43, *phnA n* = 43, *phnX n* = 43, where *n* equals the number of biologically independent sampling sites where the genes were located. MT (**B**) epipelagic: *phnJ n* = 133, *phnA n* = 154, *phnX n* = 104, mesopelagic: *phnJ n* = 31, *phnA n* = 33, *phnX n* = 30, where *n* equals the number of biologically independent sampling sites where the transcripts were located. The relationship between the standing stock Pi concentration and degradation marker abundance, analysed by linear regression of Log_{10} Pi concentration and Log_{10} gene/transcript abundance, in the MG (**C**) (*phnJ* R^2 = 0.390, p = $7.73e^{-16}$, *phnA* R^2 = -0.006, p = 0.636, *phnX* R^2 = 0.000, p = 0.333), and MT (**D**) (*phnJ* R^2 = 0.224, p = $3.55e^{-9}$, *phnA* R^2 = 0.064, p = $8.15e^{-4}$, *phnX* R^2 = -0.009, p = 0.893) R^2 values are shown. *** = p<0.001, ns =non-significant. Log2 abundance of

phosphonate degradation marker genes within metagenome (circles) and metatranscriptome (triangles) in Epipelagic (**E**) and Mesopelagic (**F**) waters within various oceanic regions. Circles/triangles denote mean Log_2 abundance values, error bars denote standard deviation of the mean. *phnA* (blue), *phnJ* (green), *phnX* (red). AO = Arctic Ocean, SO = Southern Ocean, IO = Indian Ocean, SPO = South Pacific Ocean, NPO = North Pacific Ocean, SAO = South Atlantic Ocean, NAO = North Atlantic Ocean, RS = Red Sea, MS = Mediterranean Sea. Epipelagic (**E**) *phnJ* AO MG *n* = 29, MT *n* = 14, SO MG *n* = 3, MT *n* = 4, IO MG *n* = 21, MT *n* = 19, SPO MG *n* = 25, MT *n* = 34, NPO MG *n* = 11, MT *n* = 20, SAO MG *n* = 14, MT *n* = 15, NAO MG *n* = 16, MT *n* = 17, RS MG *n* = 6, MT *n* = 3, MS MG *n* = 12, MT *n* = 7, *phnA* AO MG *n* = 29, MT *n* = 28, SO MG *n* = 3, MT *n* = 8, IO MG *n* = 21, MT *n* = 19, SPO MG *n* = 25, MT *n* = 35, NPO MG *n* = 11, MT *n* = 20, SAO MG *n* = 14, MT *n* = 17, NAO MG *n* = 16, MT *n* = 17, RS MG *n* = 6, MT *n* = 3, MS MG *n* = 12, MT *n* = 7, *phnX* AO MG *n* = 21, MT *n* = 11, SO MG *n* = 2, MT *n* = 2, IO MG *n* = 20, MT *n* = 9, SPO MG *n* = 25, MT *n* = 26, NPO MG *n* = 11, MT *n* = 19, SAO MG *n* = 13, MT *n* = 13, NAO MG *n* = 16, MT *n* = 16, RS MG *n* = 6, MT *n* = 2, MS MG *n* = 12, MT *n* = 6. Mesopelagic (**F**) *phnJ* AO MG *n* = 9, MT *n* = 7, SO MG *n* = 1, MT *n* = 0, IO MG *n* = 6, MT *n* = 4, SPO MG *n* = 9, MT *n* = 9, NPO MG *n* = 5, MT *n* = 5, SAO MG *n* = 5, MT *n* = 2, NAO MG *n* = 8, MT *n* = 5, *phnA* AO MG *n* = 9, MT *n* = 7, SO MG *n* = 1, MT *n* = 0, IO MG *n* = 6, MT *n* = 4, SPO MG *n* = 9, MT *n* = 9, NPO MG *n* = 5, MT *n* = 5, SAO MG *n* = 5, MT *n* = 2, NAO MG *n* = 8, MT *n* = 6, MT *n* = 7, *phnX* AO MG *n* = 9, MT *n* = 7, SO MG *n* = 1, MT *n* = 0, IO MG *n* = 6, MT *n* = 4, SPO MG *n* = 9, MT *n* = 9, NPO MG *n* = 5, MT *n* = 5, SAO MG *n* = 5, MT *n* = 2, NAO MG *n* = 8, MT *n* = 5. In **a**-**b** data are represented as boxplots, where the middle line is the median and the upper and lower hinges correspond to the first and third quartiles. The upper whisker extends from the upper hinge to the largest value that is no more than 1.5×IQR (inter-quartile range) from the upper hinge, and the lower whisker extends from the lower hinge to the smallest value that is no further than 1.5×IQR from the lower hinge. Data beyond the ends of the whiskers are outlying points that are plotted individually.

169

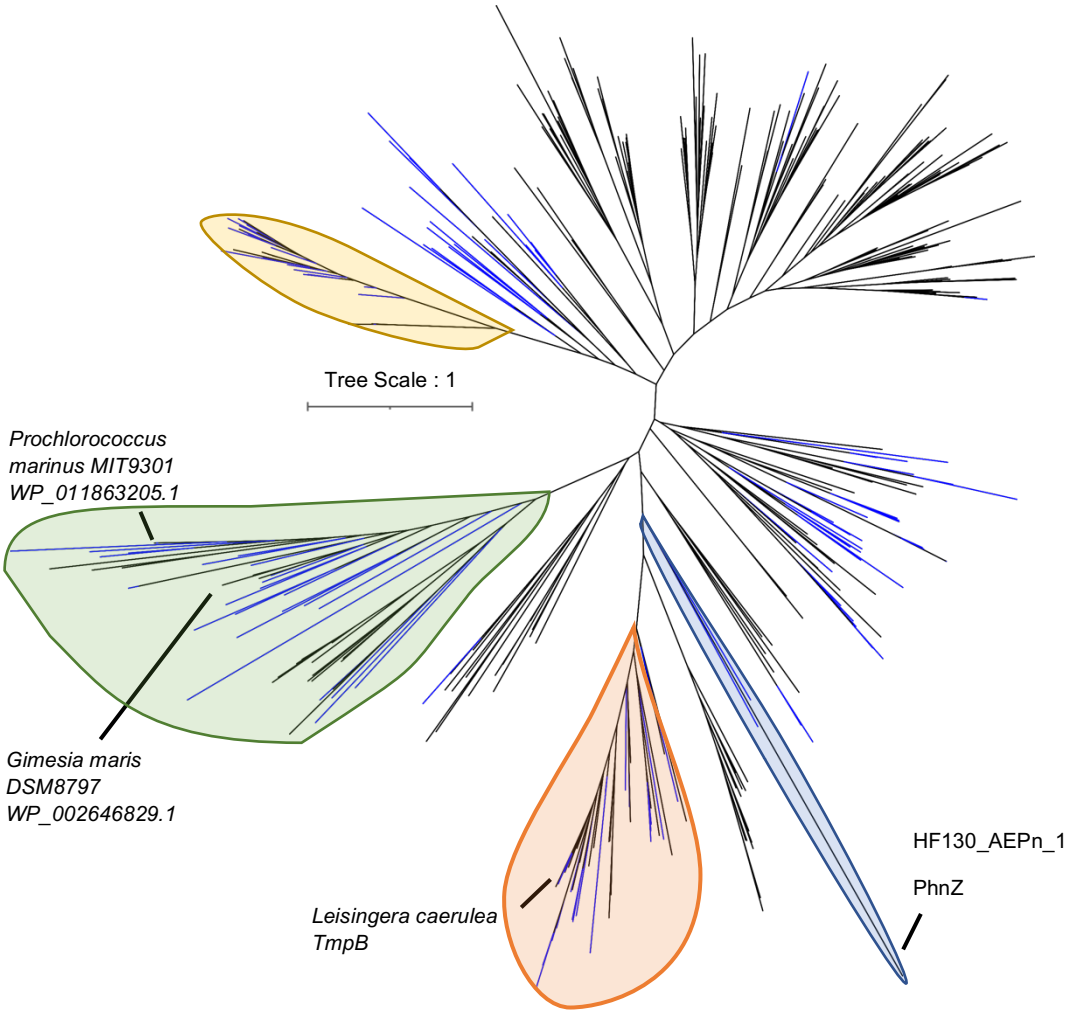


Figure S10. Phylogenetic tree of PhnZ homologs. Environmental sequences are shown in blue, Refseq sequences are shown in black. Clades of interest are highlighted based on examination of the gene neighbourhoods of the Refseq PhnZ homologs contained within. The (hydroxy-)methylphosphonate specific PhnZ clade is highlighted in green, the (N)-trimethyl-2-aminophosphonate specific TmpB clade is highlighted in orange, and the 2AEP specific clade is highlighted in blue. Many of the remaining abundant sequences fall within the clade highlighted in gold, but the genetic context of PhnZ sequences from Refseq organisms within is not conserved, and no function has been demonstrated.

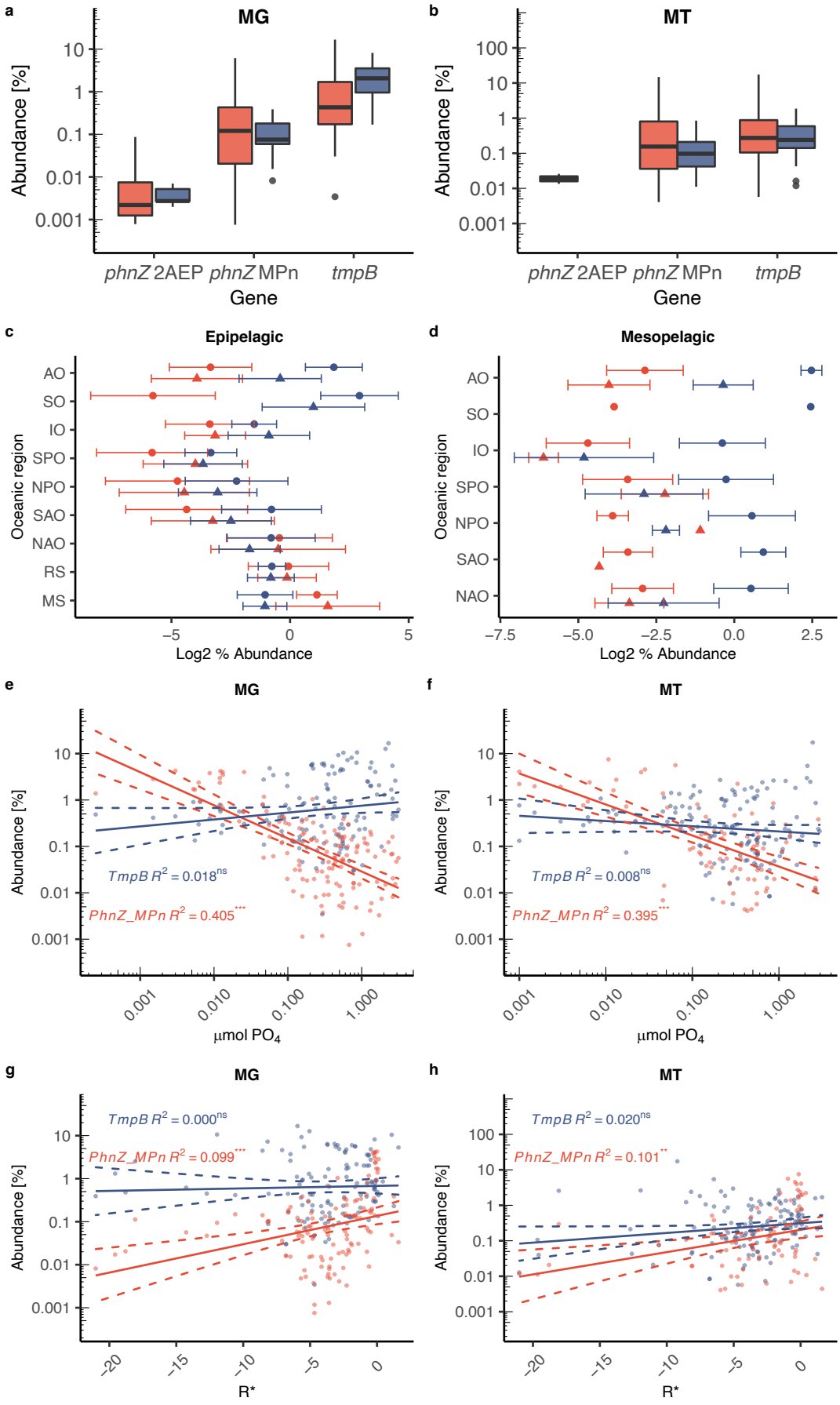

a
MG
Abundance [%]
10
1
0.1
0.01
0.001
phnZ 2AEP
phnZ MPn
tmpB
Gene
b
MT
Abundance [%]
100
10
1
0.1
0.01
0.001
phnZ 2AEP
phnZ MPn
tmpB
Gene
c
Epipelagic
Oceanic region
AO
SO
IO
SPO
NPO
SAO
NAO
RS
MS
−5
0
5
Log2 % Abundance
d
Mesopelagic
Oceanic region
AO
SO
IO
SPO
NPO
SAO
NAO
−7.5
−5.0
−2.5
0.0
2.5
Log2 % Abundance
e
MG
Abundance [%]
TmpB R^2 = 0.018^{ns}
PhnZ_MPn R^2 = 0.405***
0.001
0.010
0.100
1.000
µmol PO_4
f
MT
Abundance [%]
TmpB R^2 = 0.008^{ns}
PhnZ_MPn R^2 = 0.395***
0.001
0.010
0.100
1.000
µmol PO_4
g
MG
Abundance [%]
TmpB R^2 = 0.000^{ns}
PhnZ_MPn R^2 = 0.099***
−20
−15
−10
−5
0
R*
h
MT
Abundance [%]
TmpB R^2 = 0.020^{ns}
PhnZ_MPn R^2 = 0.101**
−20
−15
−10
−5
0
R*

Figure S11. PhnZ abundance in the TARA oceans dataset. Metagenomic (**a**) and metatranscriptomic (**b**) abundance of three *phnZ* clades (*phnZ* 2AEP = 2AEP specific, *phnZ* MPn = methlyphosphonate specific, *tmpB* = (*N*)-trimethyl-2-aminoethylphosphonate specific) within the TARA ocean dataset. Data is split between epipelagic (red), and mesopelagic (blue) waters. Abundance is normalised as a percentage of the median abundance of 10 single copy marker genes within either OM-RGCv2+G (MG) or OM-RGCv2+T (MT). MG (**a**) epipelagic: *phnZ* 2AEP *n* = 4, *phnZ* MPn *n* = 134, *tmpB n* = 137, mesopelagic: *phnZ* 2AEP *n* = 5, *phnZ* MPn *n* = 43, *tmpB n* = 43, where *n* equals the number of biologically independent sampling sites where the genes were located. MT (**b**) epipelagic: *phnZ* 2AEP *n* = 2, *phnZ* MPn *n* = 82, *tmpB n* = 148, mesopelagic: *phnZ* 2AEP *n* = 2, *phnZ* MPn *n* = 19, *tmpB n* = 24, where *n* equals the number of biologically independent sampling sites where the transcripts were located. Abundance (Log2 % abundance relative to median abundance of 10 single copy core genes) of *phnD*, *aepX*, *aepP* in MG (circles) and MT (triangles) in epipelagic (**c**) and mesopelagic (**d**) waters, split by oceanic region. *phnZ*_MPn (red), *tmpB* (blue). AO = Arctic Ocean, SO = Southern Ocean, IO = Indian Ocean, SPO = South Pacific Ocean, NPO = North Pacific Ocean, SAO = South Atlantic Ocean, NAO = North Atlantic Ocean, RS = Red Sea, MS = Mediterranean Sea. Circles/triangles are mean values of Log_2 abundance, error bars represent standard deviation of the mean. Epipelagic (**c**) *phnZ* MPn AO MG *n* = 27, MT *n* = 26, SO MG *n* = 3, MT *n* = 8, IO MG *n* = 21, MT *n* = 18, SPO MG *n* = 24, MT *n* = 35, NPO MG *n* = 11, MT *n* = 20, SAO MG *n* = 14, MT *n* = 17, NAO MG *n* = 16, MT *n* = 17, RS MG *n* = 6, MT *n* = 3, MS MG *n* = 12, MT *n* = 7, *tmpB* AO MG *n* = 29, MT *n* = 28, SO MG *n* = 3, MT *n* = 8, IO MG *n* = 21, MT *n* = 19, SPO MG *n* = 25, MT *n* = 35, NPO MG *n* = 11, MT *n* = 20, SAO MG *n* = 14, MT *n* = 17, NAO MG *n* = 16, MT *n* = 17, RS MG *n* = 6, MT *n* = 3, MS MG *n* = 12, MT *n* = 7. Mesopelagic (**d**) *phnZ* MPn AO MG *n* = 9, MT *n* = 7, SO MG *n* = 1, MT *n* = 0, IO MG *n* = 6, MT *n* = 4, SPO MG *n* = 9, MT *n* = 9, NPO MG *n* = 5, MT *n* = 5, SAO MG *n* = 5, MT *n* = 2, NAO MG *n* = 8, MT *n* = 6, MT *n* = 7, *tmpB* AO MG *n* = 9, MT *n* = 7, SO MG *n* = 1, MT *n* = 0, IO MG *n* = 6, MT *n* = 4, SPO MG *n* = 9, MT *n* = 9, NPO MG *n* = 5, MT *n* = 5, SAO MG *n* = 5, MT *n* = 2, NAO MG *n* = 8, MT *n* = 6, MT *n* = 7. The relationship between the standing stock Pi concentration and *phnZ* abundance, analysed by linear regression of Log_{10} Pi concentration and Log_{10} gene/transcript abundance, in the MG (**e**) (*phnZ* MPn R^2 = 0.405, p = $2.354e^{-16}$, *tmpB* R^2 = 0.018, p = 0.066), and MT (**f**) (*phnZ* MPn R^2 = 0.395, p = $2.707e^{-10}$, *tmpB* R^2 = 0.008, p = 0.1426) R^2 values are shown. *** = p<0.001, ns =non-significant. The relationship between R*, a measure of N vs P limitation defined as the sum of standing stock nitrate plus nitrite concentration minus 16x standing stock Pi concentration, and transporter abundance, analysed by linear regression of R* and Log_{10} gene/transcript abundance, in the MG (**g**) (*phnZ* MPn R^2 = 0.099, p = $2.8e^{-4}$, *tmpB* R^2 = -0.007, p = 0.7158) and MT (**h**) (*phnZ* MPn R^2 = 0.101, p = $3.057e^{-3}$, *tmpB* R^2 = 0.021, p = 0.051). *phnZ*_MPn (red), *tmpB* (blue), ns = not significant, ** = p < 0.01, *** = p < 0.001. 95% confidence intervals are shown by dashed lines. In **a**-**b** data are represented as boxplots, where the middle line is the median and the upper and lower hinges correspond to the first and third quartiles. The upper whisker extends from the upper hinge to the largest value that is no more than 1.5×IQR (inter-quartile range) from the upper hinge, and the lower whisker extends from the lower hinge to the smallest value that is no further than 1.5×IQR from the lower hinge. Data beyond the ends of the whiskers are outlying points that are plotted individually.

172

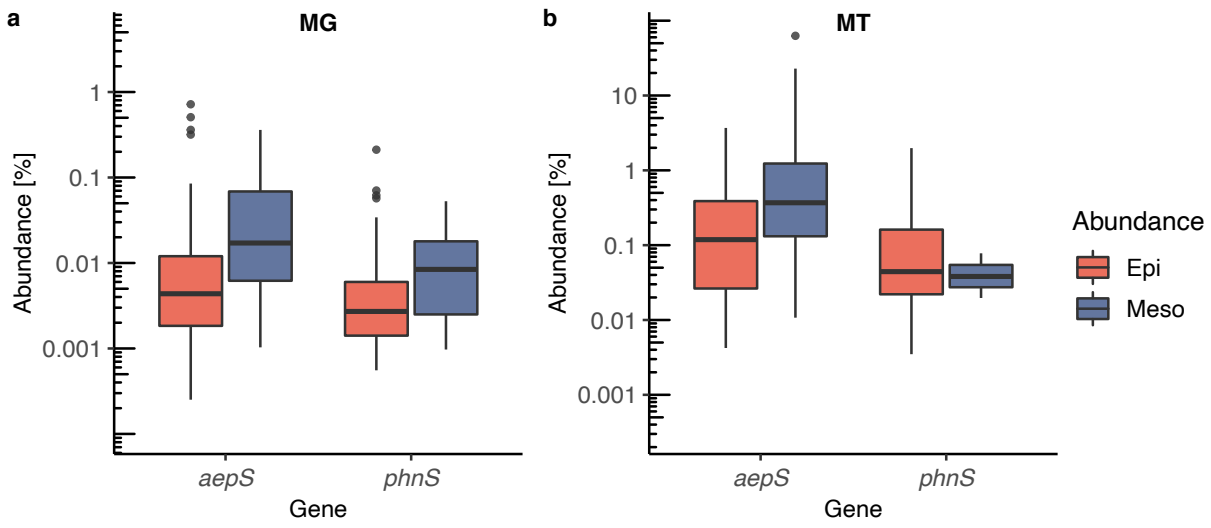


Figure S12. Metagenomic (a) and metatranscriptomic (b) abundance of the alternate 2AEP ABC transporter substrate binding proteins AepS and PhnS within the TARA oceans dataset. Data is split between epipelagic (Epi – Red), and mesopelagic (Meso – Blue) waters. Abundance is normalised as a percentage of the median abundance of 10 single copy marker genes within either OM-RGCv2+G (MG) or OM-RGCv2+T (MT). MG (**a**) epipelagic: *aepS n* = 88, *phnS n* = 48, mesopelagic: *aepS n* = 35, *phnS n* = 29, where *n* equals the number of biologically independent sampling sites where the genes were located. MT (**b**) epipelagic: *aepS n* = 91, *phnS n* = 25, mesopelagic: *aepS n* = 26, *phnS n* = 3, where *n* equals the number of biologically independent sampling sites where the transcripts were located In **a**-**b** data are represented as boxplots, where the middle line is the median and the upper and lower hinges correspond to the first and third quartiles. The upper whisker extends from the upper hinge to the largest value that is no more than 1.5×IQR (inter-quartile range) from the upper hinge, and the lower whisker extends from the lower hinge to the smallest value that is no further than 1.5×IQR from the lower hinge. Data beyond the ends of the whiskers are outlying points that are plotted individually.

Supplementary Tables

Table S1 Final growth yields (mean OD_{540} ± standard deviation) of *Stappia stellulata* (DSM 5886), *Alliiroseovarius crassostreae* (DSM 16950), *Alliiroseovarius sediminilitoris* (DSM 29439), *Thalassobius aestuarii* (DSM 15283), and *Shimia marina* (DSM26895) on 2AEP as sole N or P source with corresponding negative and positive controls. All conditions were performed in triplicate.

Organism	+Pi	+NH_4	-Pi	-NH_4	2AEP (P Source)	2AEP (N Source)
DSM 16950	0.90±0.018	0.776±0.044	0.082±0.012	0.018±0.011	1.050±0.018	0.535±0.039
DSM 29439	1.100±0.037	0.257±0.004	0.108±0.003	0.026±0.007	0.748±0.019	0.239±0.008
DSM 26895	1.045±0.020	0.397±0.035	0.157±0.019	0.045±0.008	0.827±0.013	0.347±0.004
DSM 15283	1.003±0.030	0.549±0.041	0.179±0.052	0.059±0.000	0.869±0.094	0.487±0.020
DSM 5886	1.01±0.07	0.680±0.016	0.07±0.008	0.026±0.004	0.980±0.005	0.816±0.080

174

Table S2 Phosphonate transport and catabolism gene complement of strains used in this paper. Strains used for growth curves and/or proteomics are highlighted in bold. + = present, - = absent.

Organism	PhnWAY	PhnWX	PhnJ	PhnY*Z	AepVWX	PhnCDE	PhnSTUV	AepP
***Alliiroseovarius crassostreae* DSM 16950**	+	-	-	-	+	-	-	-
***Aliiroseovarius sediminilitoris* DSM 29439**	+	-	-	-	+	-	-	-
***Shimia marina* DSM 26895**	+	-	+	-	+	+	-	-
***Thalassobius aestuarii* DSM 15283**	+	-	+	-	+	+	-	-
***Stappia stellulata* DSM 5886**	+	-	+	-	+	+	-	-
***Pseudomonas putida* BIRD-1**	-	+	-	-	+	-	-	+
***Pseudomonas flourescens* SBW25**	-	+	+	-	+	+	-	-
Chitinibacter tainanensis DSM 15459	-	+	-	-	+	-	-	-
Paraburkholderia insulsa LMG 28183	+	-	+	-	+	+	+	-
Alicycliphilus denitrificans BC	+	-	-	-	+	-	-	-
Oceanimonas smirnovii ATCC BAA-899	-	+	-	-	+	-	-	-
Burkholderia cepacia RB-39	+	-	-	-	+	-	+	-
Ruegeria faecimaris DSM 28009	+	-	+	-	+	+	-	-
Collimonas fungivorans Ter331	+	-	-	-	+	+	-	-
Burkholderia cepacia GG4	+	-	-	-	+	-	+	-
Burkholderia territorii MSMB1499	+	-	-	-	+	-	+	-
Aeromonas hydrophila NF1	-	+	-	-	+	-	-	-
Plesiomonas shigelloides GN7	-	+	-	-	+	-	-	-
Variovorax sp. GV051	+	+	-	-	+	-	-	-
Sinorhizobium medicae WSM419	+	-	+	-	+	+	-	-
Sinorhizobium meliloti 1021	+	-	+	-	+	+	-	-
Terasakiella pusilla DSM 6293	+	+	-	-	+	-	-	-
Vibrio alginolyticus NBRC 15630	-	+	-	-	+	-	-	-

175

Vibrio diabolicus CNCM I-1629	-	+	-	-	+	-	-	-
Vibrio cyclitrophicus 1F97	-	+	-	-	+	-	-	-
Arthrobacter sp. YC-RL1	+	-	-	-	-	-	-	+
Streptomyces albulus CCRC 11814	+	-	-	-	-	-	+	+
Acidobacteriaceae bacterium S15	-	+	-	-	-	-	-	+
Roseobacter sp. MED193	-	-	+	-	-	+	-	-
Falsirhodobacter sp. alg1	-	-	+	-	-	+	-	-
Rhodococcus rhodnii LMG 5362	-	-	+	-	-	-	-	+
Rhodococcus fascians A22b	-	-	-	-	-	-	-	+
Ruegeria pomeroyi DSS-3	-	-	+	-	-	+	-	-
Streptomyces sulphureus L180	-	-	-	-	-	-	-	+

Table S4 List of primers used in this paper.

Primer	Sequence	Plasmid	Used For
AepSTU ArmA_fwd	ATTCGAGCTCGGTACCCGGGCTCGAGGTCACAGTCGATC	pkmob*sacB-aepSTU*	Cloning region A of PPUBIRD1_3891-3895 (*aepSTU*)
AepSTU ArmA_rev	GGTCATGTTAGCGACCTTGTACTTGCCTTTTTC	pkmob*sacB-aepSTU*	Cloning region A of PPUBIRD1_3891-3895 (*aepSTU*)
AepSTU ArmB_fwd	ACAAGGTCGCGAAGGTCAAGGCCGACGAAA	pkmob*sacB-aepSTU*	Cloning region B of PPUBIRD1_3891-3895 (*aepSTU*)
AepSTU ArmB_rev	TAAAACGACGGCCAGTGCCAACACTGTGCGATGTAGGAGC	pkmob*sacB-aepSTU*	Cloning region B of PPUBIRD1_3891-3895 (*aepSTU*)
PhnWm-ArmA_fwd	TACGAATTCGAGCTCGGTACCCGGGAGCTACCTCCAGGCGCCC	pkmob*sacB-phnW*	Cloning region A of PPUBIRD1_3442 (*phnW*)
PhnWm-ArmA_rev	ACCACACCGATACCCCAGGAGCCCCAGTCCA	pkmob*sacB-phnW*	Cloning region A of PPUBIRD1_3442 (*phnW*)
PhnWm-ArmB_fwd	TGGGGCTCCTGGGGTATCGGTGTGGTCGGGG	pkmob*sacB-phnW*	Cloning region B of PPUBIRD1_3442 (*phnW*)
PhnWm-ArmB_rev	CGTTGTAAAACGACGGCCAGTGCCACCGATCTTCAGGCCGCCT	pkmob*sacB-phnW*	Cloning region B of PPUBIRD1_3442 (*phnW*)
PhnX_ArmA_fwd	TACGAATTCGAGCTCGGTACCCGGGCACTGCACGAAGCACTGC	pkmob*sacB-phnX-gm*	Cloning region A of PPUBIRD1_3443 (*phnX*)
PhnX_ArmA_rev	CTCTAGAGTCGACATCTGGGTGGGAGCGAATG	pkmob*sacB-phnX-gm*	Cloning region A of PPUBIRD1_3443 (*phnX*)
PhnX_Gent_fwd	CTCCCACCCAGATGTCGACTCTAGAGGATCCCCGG	pkmob*sacB-phnX-gm*	Cloning of *gm* cassette for Δ*phnX:gm* knockout
PhnX_Gent_rev	GCCGGCAAACAGTTTGGCCGCGGCGTTGTGA	pkmob*sacB-phnX-gm*	Cloning of *gm* cassette for Δ*phnX:gm* knockout
PhnX_ArmB_fwd	ACGCCGCGGCCAAACTGTTTGCCGGCTCCCG	pkmob*sacB-phnX-gm*	Cloning region B of PPUBIRD1_3443 (*phnX*)
PhnX_ArmB_rev	CGACGGCCAGTGCCAAGCTTGCATGGGGCCGGTTGGCCAATTTC	pkmob*sacB-phnX-gm*	Cloning region B of PPUBIRD1_3443 (*phnX*)

AepP_ArmA_fwd	ATTCGAGCTCGGTACCCGGGGAACAGC ACCAGCAGGATG	pkmob*sacB-aepP*	Cloning region A of PPUBIRD1_3219 (*aepP*)
AepP_ArmA_rev	CCGTGACGCCCTTGTCGAGCATGAAAC CTG	pkmob*sacB-aepP*	Cloning region A of PPUBIRD1_3219 (*aepP*)
AepP_ArmB_fwd	GCTCGACAAGGGCGTCACGGTGTTCA TCG	pkmob*sacB-aepP*	Cloning region B of PPUBIRD1_3219 (*aepP*)
AepP_ArmB_rev	TAAAACGACGGCCAGTGCCACGGGCC TTTGACCAGCGG	pkmob*SacB-aepP*	Cloning region B of PPUBIRD1_3219 (*aepP*)
pBB:AepP_fwd	GCTGCAGGAATTCGATATCAGCTGCCG TGCACGGCCAC	pBB:*aepP*-km	Cloning of PPUBIRD1_3219 (*aepP*) for complementation
pBB:AepP_rev	TACCGGGCCCCCCCTCGAGGTTAAGTA CGCAGTGGCTT GATTGCCATGC	pBB:*aepP*-km	Cloning of PPUBIRD1_3219 (*aepP*) for complementation
pBB:AepXVW^{BIRD}_fwd	GCTGCAGGAATTCGATATCACGGCGCC AGGCCAGGTTG	pBB:*aepXVW*^{BIRD}-km	Cloning of PPUBIRD1_4925-4927 (*aepXVW*) for complementation
pBB:AepXVW^{BIRD}_rev	TACCGGGCCCCCCCTCGAGGTTACTGC TGGGCCACCTTCTCC	pBB:*aepXVW*^{BIRD}-km	Cloning of PPUBIRD1_4925-4927 (*aepXVW*) for complementation
pBB:PhnWX_fwd	CTATAGGGCGAATTGGAGCTGAGCATG GCCTGCGCTTC	pBB:*phnWX*-km	Cloning of PPUBIRD1_3442-3443 (*phnWX*) for complementation
pBB:PhnWX_rev	CGAATTCCTGCAGCCCGGGGCGTTCTA GCTCAGGTTTGTCG	pBB:*phnWX*-km	Cloning of PPUBIRD1_3442-3443 (*phnWX*) for complementation
pBB:PhnW^{prom}_fwd	CTATAGGGCGAATTGGAGCTGAGCATG GCCTGCGCTTC	pBB:*phnW*^{prom}*X*-km	Cloning of PPUBIRD1_3443 (*phnX*) under the control of the *phnW* promoter for complementation
pBB:PhnW^{prom}_rev	TGTAGTTCATTCCGTTATTCCTCACA AGCGG	pBB:*phnW*^{prom}*X*-km	Cloning of PPUBIRD1_3443 (*phnX*) under the control of the *phnW* promoter for complementation
pBB:PhnX_fwd	GAATAACGGAATGAACTACAACA ACCCC	pBB:*phnW*^{prom}*X*-km	Cloning of PPUBIRD1_3443 (*phnX*) under the control of the *phnW* promoter for complementation

pBB:PhnX_rev	CGAATTCCTGCAGCCCGGGGCGTTCTAGCTCAGGTTTG	pBB:*phnW*^{prom}*X*-km	Cloning of PPUBIRD1_3443 (*phnX*) under the control of the *phnW* promoter for complementation
pBB:AepX^{prom}_fwd	CCACCGCGGTGGCGGCCGCTGATTTCTGCTGGCTTCTTCG	pBB:*aepX*^{BIRDprom}*:* *aepXVW*^{Stappia}-km	Cloning of G572DRAFT_0252-0254 (*aepXVW Stappia stellulata*) under the control of the PPUBIRD1_4927 (*aepXVW*) promoter for complementation
pBB:AepX^{prom}_rev	CCGTTCCCATGGCATGGCCTCATCGAAG	pBB:*aepX*^{BIRDprom}*:* *aepXVW*^{Stappia}-km	Cloning of G572DRAFT_0252-0254 (*aepXVW Stappia stellulata*) under the control of the PPUBIRD1_4927 (*aepXVW*) promoter for complementation
pBB:AepXVW^{Stappia}_fwd	AGGCCATGCCATGGGAACGGTATCAAAGACTTTCGGAC	pBB:*aepX*^{BIRDprom}*:* *aepXVW*^{Stappia}-km	Cloning of G572DRAFT_0252-0254 (*aepXVW Stappia stellulata*) under the control of the PPUBIRD1_4927 (*aepXVW*) promoter for complementation
pBB:AepXVW^{Stappia}_rev	CGAATTCCTGCAGCCCGGGGTCAGCCCGGATCGCGCTG	pBB:*aepX*^{BIRDprom}*:* *aepXVW*^{Stappia}-km	Cloning of G572DRAFT_0252-0254 (*aepXVW Stappia stellulata*) under the control of the PPUBIRD1_4927 (*aepXVW*) promoter for complementation
pET21:AepXSs_Fwd	ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGCTTCGGCGCGGACTTCGA	pET21a:*aepXSs*-amp	Cloning *S. stellulata aepX* into the pET21a expression vector
pET21:AepXSs_Rev	TTTAACTTTAAGAAGGAGATCTACATATGCGCGAGACAATCACCGTC	pET21a:*aepXSs*-amp	Cloning *S. stellulata aepX* into the pET21a expression vector

Appendix II

JCM
Journals.ASM.org

Performance of a Noninvasive Test for Detecting *Mycobacterium bovis* Shedding in European Badger (*Meles meles*) Populations

Hayley C. King,[a,b] Andrew Murphy,[a] Phillip James,[a] Emma Travis,[a] David Porter,[a] Jason Sawyer,[c] Jennifer Cork,[c] Richard J. Delahay,[d] William Gaze,[e] Orin Courtenay,[a,b] Elizabeth M. Wellington[a]

University of Warwick, School of Life Sciences, Gibbet Hill Campus, Coventry, United Kingdom[a]; Warwick Infectious Disease Epidemiology Research, University of Warwick, Coventry, United Kingdom[b]; Animal and Plant Health Agency Weybridge, Addlestone, Surrey, United Kingdom[c]; National Wildlife Management Centre, Animal and Plant Health Agency, Woodchester Park, Gloucestershire, United Kingdom[d]; European Centre for Environmental and Human Health, University of Exeter Medical School, Truro, Cornwall, United Kingdom[e]

The incidence of *Mycobacterium bovis*, the causative agent of bovine tuberculosis, in cattle herds in the United Kingdom is increasing, resulting in substantial economic losses. The European badger (*Meles meles*) is implicated as a wildlife reservoir and is the subject of control measures aimed at reducing the incidence of infection in cattle populations. Understanding the epidemiology of *M. bovis* in badger populations is essential for directing control interventions and understanding disease spread; however, accurate diagnosis in live animals is challenging and currently uses invasive methods. Here we present a noninvasive diagnostic procedure and sampling regimen using field sampling of latrines and detection of *M. bovis* with quantitative PCR tests, the results of which strongly correlate with the results of immunoassays in the field at the social group level. This method allows *M. bovis* infections in badger populations to be monitored without trapping and provides additional information on the quantities of bacterial DNA shed. Therefore, our approach may provide valuable insights into the epidemiology of bovine tuberculosis in badger populations and inform disease control interventions.

Mycobacterium bovis infection in wildlife is an issue of growing importance worldwide, with infections found in a range of species, including buffalo in Africa (1), wild boar in Spain (2), brushtail possums in New Zealand (3), and European badgers in the United Kingdom (4) and the Republic of Ireland (5). In the United Kingdom and the Republic of Ireland, badgers are involved in the transmission of tuberculosis (TB) to cattle (6–8). The incidence of *M. bovis* in cattle herds in the United Kingdom has been increasing for over 30 years (9), resulting in substantial economic losses (10). Once infected, badgers may intermittently shed *M. bovis* cells in sputum, feces, and urine (4), creating an environmental source of potential infection for other badgers and cattle (11, 12). *M. bovis* DNA has been shown to survive outside the host for up to 21 months, and 16S rRNA has been detected in badger setts and latrines (13). In addition, studies have found a 2.5% positivity rate when culturing from badger feces (14), and *M. bovis* has been cultured from cattle feces several months after excretion (15). Furthermore, *M. bovis* that had persisted in soil for over 12 months was able to colonize mice (16). This indicates that at least a proportion of *M. bovis* cells shed in badger feces can remain viable in the environment. Monitoring *M. bovis* infections in badger populations is important for understanding the location and spread of disease and directing control efforts. TB control interventions targeted at badgers are currently based on culling, vaccination, and farm biosecurity (17).

Accurate diagnosis of *M. bovis* infections in live animals is challenging yet essential in order to understand the epidemiology of the disease and its onward spread. Currently, infections in live badgers can be monitored through trapping and diagnosis with immunoassays (gamma interferon [IFN-γ] assay [18] and the Brock Stat-Pak assay [7]) and culture (19). Culture of clinical samples (sputum and feces) has low sensitivity of 8% and specificity of 100% (20). Furthermore, infected badgers may only intermittently excrete *M. bovis*, and culture from noninvasive sources such as feces is challenging due to decontamination methods. Immunoassays are more sensitive than culture but can be affected by animal age and duration of infection (21). The IFN-γ assay involves stimulating lymphocytes in whole blood and then detecting IFN-γ by a sandwich enzyme-linked immunosorbent assay (ELISA) (18). This method has sensitivity values ranging from 57 to 85% in badgers, with lower sensitivity in cubs than adults, and specificity values ranging from 93 to 98% (18, 20, 21). The Stat-Pak assay is a lateral flow serum antibody test with sensitivity values ranging from 50 to 78% (21) in badgers, with reduced sensitivity in the earlier stages of disease, compared to advanced infections, and specificity values ranging from 93 to 97% (20, 21). As no individual test is sufficiently sensitive or specific enough to use alone for diagnosis, the combined application of the IFN-γ and Stat-Pak assays has been recommended (20). Diagnosis through immunoassay and culture of clinical samples also requires that badgers be trapped, which is costly, logistically challenging, and likely to result in limited sample sizes.

Received 22 March 2015 **Returned for modification** 13 April 2015
Accepted 27 April 2015

Accepted manuscript posted online 3 June 2015

Citation King HC, Murphy A, James P, Travis E, Porter D, Sawyer J, Cork J, Delahay RJ, Gaze W, Courtenay O, Wellington EM. 2015. Performance of a noninvasive test for detecting *Mycobacterium bovis* shedding in European badger (*Meles meles*) populations. J Clin Microbiol 53:2316–2323. doi:10.1128/JCM.00762-15.

Editor: A. B. Onderdonk

Address correspondence to Hayley C. King, hayley.king@warwick.ac.uk.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00762-15.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.00762-15

The authors have paid a fee to allow immediate free access to this article.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

A noninvasive assay for sampling badger populations may offer an opportunity to gather information on the spatiotemporal distribution of *M. bovis* in badger populations over a larger area, more easily and cost-effectively, than by trapping. A noninvasive approach would also remove the ethical concerns related to trapping. Furthermore, immunoassays provide information on prior exposure but do not identify shedding status. Our study was focused on quantifying fecal shedding of *M. bovis* among badgers at the social group level, using an existing quantitative PCR (qPCR) assay developed in our group that quantifies *M. bovis* genome copy number (22, 23). We aimed (i) to determine the correspondence between immunoassay results and fecal qPCR assay results for *M. bovis* infections in live-trapped badgers and (ii) to establish an optimal fecal sampling regimen to maximize detection of shedding in badger populations.

MATERIALS AND METHODS

Sampling and trapping. Fresh fecal samples were obtained from latrines associated with 12 badger social groups in Woodchester Park (Gloucestershire, England) in 2012 and 2013. In 2012, intensive sampling took place during the two periods of peak badger latrine activity (spring and autumn), when up to 10 unique fecal samples were obtained from each social group on alternate days for 10 days. Feces were collected from latrines within the vicinity of the main sett of each social group. In each season, starting 2 days after trapping operations took place in that location, 10 unique fecal samples per day were taken from the latrines identified near each main sett, on two nonconsecutive days. For the purpose of this study, March through May was classified as spring, June through August as summer, September through November as autumn, and December through February as winter.

To determine the relative performance of the qPCR assay, we compared results to those from immunoassays and culture for clinical samples obtained during routine trapping and sampling of the 12 targeted badger social groups. Sputum samples were collected by aspiration of both the esophagus and the trachea using catheters. Collected samples are then flushed into physiological saline. Sputum and feces were cultured on solid medium (24) and identified as *M. bovis* by typical colony morphology followed by spoligotyping. Each social group of badgers was subjected to one trapping event per season, and trapping took place over 2 consecutive days. Badgers were trapped using baited cage traps placed around the main sett of each social group, and individual animals were identified using a unique tattoo applied at the first capture event. Trapped badgers from each of the 12 social groups were tested with the BrockTB Stat-Pak assay, the IFN-γ assay, and culture of clinical samples. To establish the relative sensitivity and specificity of the fecal qPCR assay, compared to immunoassays, at an individual animal level, we collected fecal samples (following administration of an enema) from badgers trapped and tested throughout the Woodchester Park study area. An individual or a social group was deemed *M. bovis* positive if at least one diagnostic test or culture from a clinical sample was positive. All work was approved by the University of Warwick and the Food and Environment Research Agency Ethical Review Committee and was carried out under a license granted by the Home Office under the 1986 Animals (Scientific Procedures) Act.

DNA extraction and qPCR testing. Total community DNA was extracted from 0.1 g (± 0.003 g) of feces using the Fast DNA spin kit for soil (MP Biomedicals), following the manufacturer's instructions. *M. bovis* was detected and quantified using a qPCR assay that targets the RD4 deletion region unique to the *M. bovis* genome. An initial qPCR screen of each sample was performed using an ABI 7500 Fast qPCR system (ABI) with two technical replicates of each sample. Positive controls (8.5×10^2 genome equivalents) and negative controls were also present in duplicate on each plate. PCRs were set up using 900 nM levels of each primer (RD4F, 5′-TGTGAATTCATACAAGCCGTAGTCG-3′; RD4R, 5′-CCCG TAGCGTTACTGAGAAATTGC-3′), 250 nM TaqMan probe (6-carboxyfluorescein [6FAM]-AGCGCAACACTCTTGGAGTGGCCTAC-tetramethylrhodamine [TMR]), 1 mg/ml bovine serum albumin (BSA), 12.5 μl of Environmental Master Mix 2.0 (ABI), and 10 μl of template, made up to 25 μl with molecular biology-grade water (Sigma-Aldrich). PCR cycling conditions were 50°C for 2 min followed by 95°C for 10 min and then 40 cycles of 95°C for 15 s and 58°C for 1 min. Samples exhibiting amplification in one or more technical replicates were taken on to full quantification using three technical replicates per sample, under the same conditions. If one or more of the technical replicates of the quantification assay exhibited amplification, then the sample was deemed positive for *M. bovis*. Serial dilutions of *M. bovis* BCG Danish 1331 genomic DNA were used as standards for this quantification. A previously described inhibition control assay (23) was used to detect the possibility of false-negative results due to inhibition. Each extracted sample was screened as a singlet; if the threshold cycle difference (ΔC_T) was greater than 2.5, then the sample was rescreened as a doublet. If the average ΔC_T was greater than 2.5, then the sample was reextracted from frozen fecal aliquots; if not, then the sample was considered uninhibited The number of *M. bovis* genome equivalents was quantified independently by qPCR at the University of Warwick and the Animal and Plant Health Agency (APHA) Weybridge.

Statistical analysis. All data analyses were performed using the statistical program R. Binomial and gaussian generalized linear models (GLMs) were used to determine differences in fecal sample positivity (as a binary variable) and *M. bovis* genome equivalents shed between social groups and seasons. All GLMs were carried out with the old oak group as the baseline social group, because it had the lowest prevalence of positive fecal samples, and winter as the baseline season against which all other social groups and seasons were compared. One- and two-way analysis of variance (ANOVA) was used to determine differences among social groups in sample numbers and proportions of trapped animals that were positive. Spearman's ranks were calculated to determine whether there was correspondence in the rank order of social groups based on the prevalence estimated by live testing and fecal qPCR assays.

Sensitivity, specificity, positive predictive value (PPV), and negative predicative value (NPV) were calculated for individual animals, as well at the social group level. The confidence intervals (CIs) for these values were calculated using the Wilson score interval. For analysis of data for individuals, a positive result was defined as a positive result from any diagnostic test or culture with clinical samples from a trap event. For analysis on a social group level, a social group was deemed positive if any diagnostic test or culture was positive for any trap event within the group.

False-positive rates were calculated using 68 known negative fecal samples obtained from captive badgers at the APHA, which were routinely tested for bovine TB using the IFN-γ assay, and 49 water samples. Negative samples were prepared in a double-blind manner and randomly introduced into the experiment at both laboratories. As this qPCR assay is highly specific for *M. bovis* (25), all false-positive findings are expected to result from contamination introduced in the laboratory; therefore, negative controls were included at every stage of DNA extraction and quantification.

RESULTS

Correspondence between immunoassays and fecal qPCR assays for trapped badgers. Routine badger trapping at Woodchester Park took place prior to the collection of feces from latrines. Trapped badgers were tested for *M. bovis* infection with immunoassays (IFN-γ and Stat-Pak assays) and culture of clinical samples and qPCR assays of fecal samples. In total, there were 120 trapping events, with 50% found to be positive by any test. The Stat-Pak assay identified a greater number of positive samples than did either the IFN-γ assay or the qPCR assay, which identified similar numbers of positive samples (Fig. 1). No culture-positive results were obtained from feces or sputum samples from trapped badgers. The correlation between tests was low, as follows: Stat-Pak

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

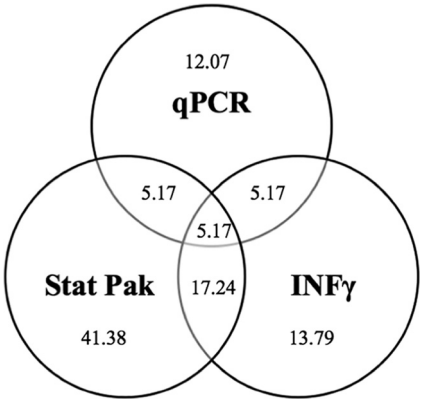


FIG 1 Percentage agreement between positive test results. There were 120 total trappings, with 60 positive trappings for which at least one test gave a positive result.

assay and IFN-γ assay, $r = 0.27$ ($P < 0.05$); Stat-Pak assay and qPCR assay, $r = 0.11$ ($P > 0.05$); IFN-γ assay and qPCR assay, $r = 0.20$ ($P < 0.05$).

As there is no gold standard for diagnosing infections in badgers, sensitivity, specificity, PPV, and NPV calculations were carried out using the Stat-Pak assay and the IFN-γ assay separately as the gold standard and with the two tests combined. The sensitivity of the qPCR assay in comparison with the Stat-Pak and IFN-γ assays, separately or combined, was low, ranging from 14 to 25%, whereas the sensitivity of the Stat-Pak and IFN-γ assays in comparison with one another was higher, at 32% and 59%, respectively (Table 1). The lower sensitivity of qPCR was expected because it is a measure of shedding, rather than infection, and infected badgers may shed *M. bovis* intermittently or not at all.

The relative specificity of qPCR was high, ranging from 91 to 93%, in comparison with the Stat-Pak and IFN-γ assays, which had specificities of 86 and 67%, respectively, in comparison with one another. The PPVs for qPCR ranged from 43 to 64% and from 33 to 59% in comparison with the Stat-Pak and IFN-γ assays, respectively. The high relative specificity of qPCR is due to the definitive detection of the DNA target, which is unique to *M. bovis*.

The NPVs ranged from 54 to 83% for qPCR and from 67 to 86% for the Stat-Pak and IFN-γ assays in comparison with one another. The NPV was lower overall for qPCR than for the immunoassays as the former detects shedding, rather than infected animals, which results in some positive animals being missed.

Historically positive trapping events. As badger populations in Woodchester Park have been extensively studied for over 20 years, trapping information is available for several years prior to this study. In the 120 trapping events discussed above, the captured animal was historically positive by at least one test on 57% of the occasions. Of these historically positive captures, 29% were positive by the Stat-Pak assay only, 25% by the IFN-γ assay only, 43% by the Stat-Pak and IFN-γ assays, and 3% by culture of clinical samples, the IFN-γ assay, and the Stat-Pak assay (see Fig. S1 in the supplemental material). Contemporary (2012) trapping indicated that, in 53% of cases (32 cases), the animal had at least one historical positive test result; 24% were positive by the Stat-Pak assay only, 14% by the IFN-γ assay only, 59% by the Stat-Pak and IFN-γ assays, and 3% by culture and Stat-Pak and IFN-γ assays, but none was positive by fecal culture alone (see Fig. S1 in the supplemental material).

The majority (93%) of animals that were historically Stat-Pak assay positive were also positive with the Stat-Pak assay during contemporary testing. Also, 65% of animals that were historically IFN-γ assay positive were also positive with the IFN-γ assay during contemporary testing (see Fig. S2 in the supplemental material). No captured animals that were historically culture positive were positive by contemporary culture; however, they were all positive by fecal qPCR, indicating that they were still shedding *M. bovis*. Forty-five percent of trapped badgers that were historically positive by both the Stat-Pak and IFN-γ assays were also positive by these tests during contemporary testing. Past diagnostic test results were pooled for all capture events for each badger; therefore, IFN-γ and Stat-Pak assays might have been positive at different capture events rather than simultaneously.

Of the 16 trap events that were positive by qPCR, 12 (75%) were historically positive by at least one of the live tests. Both of the historically culture-positive animals were positive by qPCR.

Contemporary seasonal trapping. Badgers were trapped seasonally throughout 2012, which coincided with the collection of fecal samples from latrines. Trapped badgers were routinely tested using the Stat-Pak assay, the IFN-γ assay, and culture of clinical samples. The numbers of badgers trapped were highly variable among the social groups (5 to 18 animals per group) and seasons (see Tables S1 and S2 in the supplemental material). The numbers of badgers caught per sampling day were also highly variable between seasons, with 2 to 28 badgers being trapped on a given day (see Table S2 in the supplemental material). Greater numbers of badgers were trapped in spring ($t = 4.731$, $P < 0.001$) and summer ($t = 2.880$, $P < 0.05$) than in autumn or winter ($F_{3,44} = 9.421$, $P < 0.001$). There were no differences in the numbers of badgers caught per social group throughout the year ($F_{11,36} = 1,272$, $P > 0.05$) or in the percentages of positive badgers trapped per season ($F_{3,44} = 0.8523$, $P > 0.05$). However, there was significant variation in the percentages of positive animals (as estimated by immunoassay) per social group across the whole year ($F_{11,36} = 3.635$, $P < 0.001$), with the Honeywell ($t = 2.563$, $P < 0.05$), nettle ($t = 2.357$, $P < 0.05$), and septic tank ($t = 2.457$, $P < 0.05$) groups having larger proportions of test-positive badgers than the other groups.

TABLE 1 Sensitivity and specificity of *M. bovis* diagnostics for trapped badgers with Stat-Pak and IFN-γ assays individually and combined as gold standards against qPCR and with Stat-Pak and IFN-γ assays as gold standards against each other

Test(s)	Sensitivity (95% CI) (%)	Specificity (95% CI) (%)	PPV (95% CI) (%)	NPV (95% CI) (%)
Stat-Pak	15.00 (9.41–23.06)	92.19 (85.66–96.07)	54.55 (45.03–83.08)	63.44 (53.91–72.03)
IFN-γ	25.00 (18.08–33.48)	91.40 (84.55–94.97)	42.86 (35.27–52.83)	82.52 (74.71–88.30)
Stat-Pak and IFN-γ	14.29 (8.86–22.24)	92.59 (85.94–96.23)	63.64 (57.81–74.93)	54.35 (44.84–63.56)
Stat-Pak as true positive (IFN-γ)	32.50 (23.89–41.47)	85.71 (77.76–91.14)	59.09 (49.53–68.01)	66.67 (57.21–74.96)
IFN-γ as true positive (Stat-Pak)	59.09 (49.53–68.01)	66.67 (57.21–74.96)	32.50 (24.30–41.94)	85.71 (77.76–91.14)

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

TABLE 2 Odds of finding *M. bovis*-positive samples each season, with winter as the baseline

Season	Odds ratio (95% CI)	*P*
Spring	1.76 (0.84–3.66)	0.13
Summer	2.72 (1.31–5.64)	0.007
Autumn	1.97 (0.96–4.04)	0.06
Winter	1	

In this study, badger sex was not related to the likelihood of yielding a positive test result (female, odds ratio, 1; male, odds ratio, 0.86 [95% CI, 0.43 to 1.73]; $P > 0.05$). When diagnostic tests were examined individually, neither season nor badger sex was related to the likelihood of a positive Stat-Pak or IFN-γ test result.

Seasonal and social group differences in latrine fecal sampling. The total numbers of fecal samples collected varied between social groups, from 76 to 175 samples across the year (see Table S3 in the supplemental material). On average, more samples were collected per sampling day in the spring, with a mean of 51 samples per day, than in other seasons, which ranged from 23 to 38 samples per day. On each sampling day, the aim was to collect 10 fresh fecal samples; spring and summer sample numbers averaged 9 and 8 samples per day, respectively, with 6 samples per day being collected in winter.

The odds ratios for finding a positive fecal sample were equal across all seasons except for summer, when there was a significantly higher probability (Table 2). There was a significant difference in the number of *M. bovis* genome equivalents shed over the year, with significantly greater numbers of cells being detected in summer and autumn than in winter and spring, although there were no more positive samples in autumn than in winter or spring (Table 2). There was no correlation between the number of badgers trapped and the number of fecal samples collected in the same season per social group ($r = 0.18$, $P > 0.05$) or between the percentage of positive badgers trapped and the percentage of positive fecal samples per social group by season ($r = 0.22$, $P > 0.05$).

Fecal qPCR replication at two centers. A total of 1,090 samples (67% of all samples collected) were subjected to DNA extraction at both the University of Warwick and the APHA Weybridge. Of these, 13% (140 samples) were found to be positive, of which 32% (45 samples) were positive only at the University of Warwick, 29% (41 samples) were positive only at the APHA, and 39% (54 samples) were positive at both centers. There was no statistical difference in whether a sample was positive at the University of Warwick or at the APHA (McNemar test, $\chi^2_{1,1,090} = 0.165$, $P > 0.05$). There was a significant difference in *M. bovis* genome equivalents in samples that were positive at both centers, with fewer genome equivalents being found in samples analyzed at the University of Warwick ($\beta = -2.53$, $P < 0.01$). Although 61% of samples were identified as positive at only one center, the rankings of the social groups (in terms of the proportions of positive samples) were strongly correlated (Spearman's $\rho = 0.750$, $P < 0.05$).

Correspondence between live testing with immunoassays and culture and fecal qPCR testing from latrines. During 2012, 10 of the 12 social groups targeted for the present study were positive by the Stat-Pak and IFN-γ assays, whereas qPCR testing of fecal samples obtained from latrines identified all 12 to be positive (Table 3). The numbers of positive social groups in each season were consistently higher using qPCR testing of fecal samples from latrines, compared with immunoassay results from live-captured animals, for both longitudinal and intensive sampling regimens (Table 3). The Colliers Wood and Wych Elm social groups were negative by contemporary immunoassays for live animals but were found to be positive in three and four seasons, respectively, by qPCR (Table 3). The largest discrepancy between the two diagnostic approaches was in the spring, when live-trapping diagnostics identified 58% fewer social groups as positive, compared to qPCR testing of fecal specimens from latrines. The smallest difference was in the winter, with 8% fewer social groups being identified by immunoassays and culture of clinical samples than by qPCR testing of fecal samples from latrines.

The social groups were ranked according to the proportions of positive test results, using results from live-trapping diagnostics and qPCR testing of fecal samples from latrines. The proportions

TABLE 3 Summary of trapping and fecal qPCR positivity for each social group by season

	Spring			Summer			Autumn			Winter		
Group[a]	Trap positive	qPCR positive	qPCR prevalence (%)	Trap positive	qPCR positive	qPCR prevalence (%)	Trap positive	qPCR positive	qPCR prevalence (%)	Trap positive	qPCR positive	qPCR prevalence (%)
Nettle	−	+	29.17	+	+	50.00	+	+	53.13	−	+	100.00
West	+	+	20.00	+	+	19.44	−	−	0	+	+	7.69
Honeywell	−	+	10.20	+	+	33.33	+	+	4.88	+	−	0
Septic tank	−	−	0	+	+	5.71	+	+	29.63	−	+	15.00
Top	−	+	26.09	+	+	20.00	−	−	0	−	−	0
Wych elm	−	+	15.00	−	+	13.33	−	+	11.11	−	+	20.00
Beech	+	+	8.00	+	+	8.11	+	+	10.45	−	+	11.11
Woodrush	−	+	2.63	+	+	6.06	−	+	11.11	+	−	0
Colliers Wood	−	+	2.56	−	+	5.71	−	+	12.24	−	−	0
Yew	−	+	2.04	+	+	15.63	+	+	3.28	+	+	5.26
Kennel	−	−	0	+	+	21.05	−	−	0	−	−	0
Old oak	−	−	0	+	+	8.57	+	+	2.22	+	−	0

[a] The overall positive proportions were as follows: spring, trap positive, 16.67%; qPCR positive, 75.00%; qPCR prevalence, 9.64%; summer, trap positive, 83.33%; qPCR positive, 100.00%; qPCR prevalence, 17.25%; autumn, trap positive, 50.00%; qPCR positive, 75.00%; qPCR prevalence, 11.50%; winter, trap positive, 41.67%; qPCR positive, 50.00%; qPCR prevalence, 13.26%.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

TABLE 4 Sensitivity and specificity of seasonal qPCR results in comparison with all-year trapping data, using Stat-Pak assay, IFN-γ assay, and both Stat-Pak and IFN-γ assays as gold standards

Season and measure	% (95% CI)		
	Stat-Pak assay	IFN-γ assay	Stat-Pak and IFN-γ assays
Spring			
Sensitivity	80.00 (51.73–93.72)	83.33 (55.19– 95.30)	83.33 (55.19–95.30)
Specificity	0.00 (0.00–24.25)	16.67 (4.70–44.81)	16.67 (4.70–44.81)
PPV	80.00 (51.73–93.72)	50.00 (25.38–74.62)	50.00 (25.38–74.62)
NPV	0.00 (0.00–24.25)	50.00 (25.38–74.62)	50.00 (25.38–74.62)
Summer			
Sensitivity	100 (75.75–1)	100 (75.75–100.00)	100 (75.75–100.00)
Specificity	0.00 (0.00–24.25)	0.00 (0.00–24.25)	0.00 (0.00–24.25)
PPV	83.33 (55.19– 95.30)	50.00 (25.38–74.62)	50.00 (25.38–74.62)
NPV	0.00 (0.00–24.25)	0.00 (0.00–24.25)	0.00 (0.00–24.25)
Autumn			
Sensitivity	70.00 (42.07–88.23)	66.67 (39.07–86.19)	66.67 (39.07–86.19)
Specificity	0.00 (0.00–24.25)	16.67 (4.70–44.81)	16.67 (4.70–44.81)
PPV	77.78 (49.49–92.60)	44.44 (2128–70.29)	44.44 (2128–70.29)
NPV	0.00 (0.00–24.25)	33.33 (13.81–60.93)	33.33 (13.81–60.93)
Winter			
Sensitivity	50.00 (25.38–74.62)	66.67 (39.07–86.19)	66.67 (39.07–86.19)
Specificity	100 (75.75–100.00)	66.67 (39.07–86.19)	66.67 (39.07–86.19)
PPV	100 (75.75–100.00)	66.67 (39.07–86.19)	66.67 (39.07–86.19)
NPV	16.67 (4.70–44.81)	66.67 (39.07–86.19)	66.67 (39.07–86.19)

of test-positive samples per group estimated by qPCR testing of fecal samples collected from latrines in June and from live-trapping diagnostics in the summer were highly correlated (Spearman's $\rho = 0.87$, $P < 0.001$). Live-trapping diagnostic results for the whole year correlated strongly with qPCR results for fecal samples collected in June (Spearman's $\rho = 0.71$, $P < 0.05$), with results of fecal sampling for the whole year (Spearman's $\rho = 0.70$, $P < 0.05$), and with results from all longitudinal fecal sampling (Spearman's $\rho = 0.62$, $P < 0.05$). Ranking of social groups on the basis of qPCR results alone correlated well with contemporaneous ranking based on immunoassay and culture results.

Sensitivity and specificity of qPCR tests and immunoassays. Sensitivity, specificity, PPV, and NPV were calculated at the social group level for seasonal qPCR testing of feces from latrines in comparison with the results of live diagnostic tests for a whole year. The group-level sensitivity of qPCR testing was seasonally variable but consistently high, with the highest value being noted in summer, i.e., 100% sensitivity, in comparison with immunoassay results, and the lowest in winter, i.e., 50% (Table 4). Group-level specificity was also highly variable, ranging from 0 to 100%, in comparison with immunoassays. The lower bound of this range is due to no social groups being found negative by both live testing and qPCR. The social groups targeted in this study were all chosen because they were historically positive in live testing; therefore, it is expected that few if any would be negative by both approaches. The PPV of qPCR ranged from 44% to 100%, relative to live testing, as this test has a low false-positive rate. The NPV of qPCR ranged from 0 to 67%. This wide range may be due to the inability of the test to detect the presence of infection in the absence of shedding; therefore, if social groups contain animals that are infected but not shedding *M. bovis* in feces, then they will not be identified by qPCR testing (Table 4).

Comparison of intensive and longitudinal fecal sampling. No differences were observed between intensive and longitudinal sampling in the numbers of *M. bovis* cells shed or the numbers of positive samples, with accounting for season, and this did not vary with social group. When intensive sampling was divided into 2-day sampling periods, to mirror longitudinal sampling, no differences were observed between sampling blocks within intensive periods, in terms of the numbers of positive samples overall or per social group. In the autumn, both intensive sampling and cross-sectional sampling were carried out, with no difference between the sampling approaches in the odds of finding positive samples in each social group.

Level of fecal sampling required to detect positive social groups. Random resampling of fecal samples determined the sampling intensity required to identify, with 95% certainty, positive groups with various proportions of positive samples across a year. The number of samples required varied between 5 for the group with the most positive samples (nettle group) and up to 50 for the group with the least (yew group) (Fig. 2). In the present study, up to 20 fecal samples were collected over 2 days, but more intensive sampling would have been required to collect more. Our results indicate that the number of samples required varies with the season, with the fewest samples being required in early summer. However, sampling in only one season may not detect all positive social groups. In this study, up to 10 social groups were identified as positive by qPCR in a single sampling period. More accurate assessments of the shedding status of a group would require sampling across a whole year.

False-positive rate for fecal qPCR testing. Two (2.9%) of the 68 negative fecal samples tested positive and one (2.1%) of the 46 water samples tested positive, giving a false-positive rate of 2.6%.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

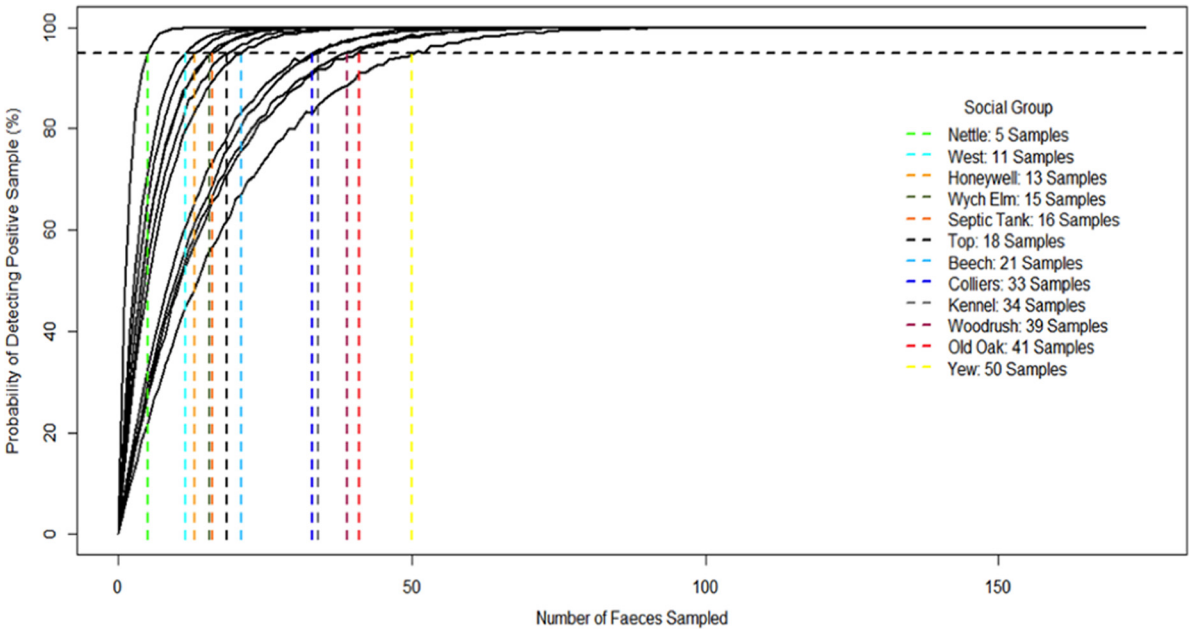


FIG 2 Numbers of fecal samples required to detect a positive social group across a year.

DISCUSSION

The results presented here indicate that qPCR testing of fecal samples from latrines is likely to be as sensitive or more so than live testing in detecting *M. bovis* in badger populations. Therefore, this method provides an alternative or complement to immunoassays and culture of clinical samples, which currently are the only measures of *M. bovis* infections in badger populations that do not involve postmortem examinations but are themselves limited in performance.

When trapped badgers were tested with immunoassays, culture of clinical samples, and qPCR testing of voided feces, there was low correspondence between test results within individuals, as noted in other studies (18). Culture of clinical samples did not produce any positive results, whereas qPCR identified 28% of all immunoassay-positive cases on an individual badger level. In the field, however, where qPCR testing was conducted on feces from latrines, the ranking of social group shedding status inferred from qPCR results correlated strongly with prevalence estimates based on immunoassay results, thus demonstrating the correspondence between approaches. As qPCR quantifies *M. bovis* genome equivalents, it provides a relative measure of the levels of shedding among social groups. Studies in other hosts have found that animals shedding the highest levels of pathogen are responsible for large proportions of transmission events (26, 27). While this has not been explicitly investigated for *M. bovis* in badgers, heterogeneity in individual- and group-level shedding may warrant further research.

Although there was strong correlation between group rankings based on the two diagnostic approaches, immunoassays consistently identified fewer positive social groups than did qPCR testing of fecal samples from latrines per season and across the year. qPCR tests also identified greater differences in the proportions of positive results among social groups than did immunoassays. Notably, the prevalence rates of infections in the west and old oak social groups estimated by immunoassays were similar, but the two groups differed widely with regard to the results of fecal qPCR tests, with the west group being highly positive and the old oak group having the fewest positive fecal samples.

The greater odds of identifying a positive fecal sample from a latrine in summer suggests that this would be the optimal time to sample badger populations. This approach maximized the number of positive samples with the lowest possible sampling intensity. Because the number of fecal samples collected did not differ between seasons, the greater odds of detecting positive samples in summer are due to an increase in positive samples, rather than a greater abundance of fecal samples during this period. The optimal fecal sampling regimen would involve collecting 10 fresh samples per day on 2 nonconsecutive days in early summer, which would detect the top 83.34% of shedding social groups, including those excreting the most *M. bovis* genome equivalents into the environment. This sampling regimen provided qPCR test data indicating relatively high sensitivity and specificity, compared with live testing, and also had greater odds of finding positive samples, compared with all other seasons. Some social groups required more than 20 samples to be collected over a year to detect fecal shedding. If the aim of sampling is to identify groups with the largest proportions of positive samples, then sampling in early summer only may be adequate. In order to identify all positive social groups (as determined by qPCR) in this study, however, two separate sampling sessions would be required, once in early summer and once in late summer, which would be the most cost-effective method for detecting all social groups. The requirement for two sampling sessions to detect all positive groups is most likely due to the intermittent nature of *M. bovis* shedding and the fact that a wider window of sampling is needed to obtain fecal samples from a large proportion of animals in each social group. While the false-positive rate of this qPCR assay is low, the probability of obtaining false-positive results increases if large numbers of samples are tested; therefore, we suggest that positive fecal samples be retested to maintain a low false-positive rate.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

In addition to being equally or more sensitive than live-trapping diagnostics, the qPCR assay with latrine samples benefits from being noninvasive and less logistically challenging than live trapping and testing. Our study has identified the potential value of qPCR testing of fecal samples collected from latrines for monitoring *M. bovis* shedding in badger populations at the group level. This may prove to be a valuable adjunct to trapping and live testing in field studies to investigate the epidemiology of *M. bovis* spread in badger populations. However, the approach could be implemented as an alternative to capture and testing when the cost of the latter may be prohibitive for monitoring disease risks over relatively large areas. For example, qPCR testing of latrine fecal samples could be applied at the edges of the areas in which TB is currently endemic in the United Kingdom or throughout high-risk areas, in order to provide spatial information on relative levels of environmental contamination, which may facilitate monitoring of spread and targeting of management. Although our study focuses on badgers, the same approach to noninvasive sampling has the potential to be applied to other pathogens or other wildlife and disease systems, particularly those involving elusive host species or settings in which capture and live testing are challenging.

ACKNOWLEDGMENTS

We acknowledge funding from the Department for Environment, Food, and Rural Affairs. H.C.K. was in receipt of a Biotechnology and Biological Sciences Research Council doctoral training grant studentship. We are also grateful to the APHA field team at Woodchester Park for support during fieldwork and to the Department for Environment, Food, and Rural Affairs for funding the long-term study.

H.C.K. was involved in sample collection and processing, undertook statistical analysis, prepared tables and figures, and wrote the manuscript with assistance from E.M.W. and R.J.D. A.M. collected and processed samples and undertook statistical analysis. P.J. collected and processed samples. E.T. advised on the project and data analysis. D.P. collected and processed samples. J.S. coordinated and oversaw processing of samples at the Animal Health and Veterinary Laboratories Agency (AHVLA) Weybridge. J.C. processed samples at AHVLA Weybridge. R.J.D. contributed to the study design, reporting, and data evaluation and organized and oversaw the fieldwork at AHVLA Woodchester. W.G. was involved in designing the original experiment. O.C., in collaboration with E.M.W., was involved in the project design. E.M.W., in collaboration with O.C., devised the experimental setup, supervised all field and practical work, and was responsible for biosafety.

We declare no competing financial interests.

REFERENCES

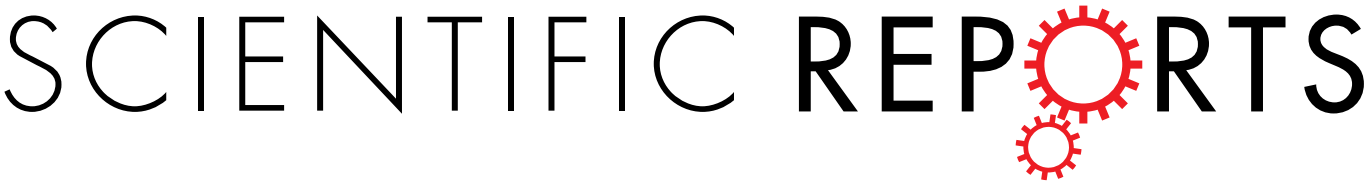
1. **Renwick AR, White PCL, Bengis RG.** 2007. Bovine tuberculosis in southern African wildlife: a multi-species host-pathogen system. Epidemiol Infect **135:**529–540. http://dx.doi.org/10.1017/S0950268806007205.
2. **Aranaz A, de Juan L, Montero N, Sanchez C, Galka M, Delso C, Alvarez J, Romero B, Bezos J, Vela AI, Briones V, Mateos A, Domínguez L.** 2004. Bovine tuberculosis (*Mycobacterium bovis*) in wildlife in Spain. J Clin Microbiol **42:**2602–2608. http://dx.doi.org/10.1128/JCM.42.6.2602-2608.2004.
3. **Coleman JD, Cooker MM.** 2001. *Mycobacterium bovis* infection in wildlife in New Zealand. Tuberculosis (Edinb) **81:**191–202. http://dx.doi.org/10.1054/tube.2001.0291.
4. **Clifton-Hadley RS, Wilesmith JW, Stuart FA.** 1993. *Mycobacterium bovis* in the European badger (*Meles meles*): epidemiological findings in tuberculous badgers from a naturally infected population. Epidemiol Infect **111:**9–19. http://dx.doi.org/10.1017/S0950268800056624.
5. **Gormley E, Collins JD.** 2000. The development of wildlife control strategies for eradication of tuberculosis in cattle in Ireland. Tuber Lung Dis **80:**229–236. http://dx.doi.org/10.1054/tuld.2000.0250.
6. **Donnelly CA, Woodroffe R, Cox DR, Bourne J, Gettinby G, Le Fevre AM, McInerney JP, Morrison WI.** 2003. Impact of localized badger culling on tuberculosis incidence in British cattle. Nature **426:**834–837. http://dx.doi.org/10.1038/nature02192.
7. **Chambers MA, Crawshaw T, Waterhouse S, Delahay R, Hewinson RG, Lyashchenko KP.** 2008. Validation of the BrockTB Stat-Pak assay for detection of tuberculosis in Eurasian badgers (*Meles meles*) and influence of disease severity on diagnostic accuracy. J Clin Microbiol **46:**1498–1500. http://dx.doi.org/10.1128/JCM.02117-07.
8. **Aznar I, McGrath G, Murphy D, Corner LAL, Gormley E, Frankena K, More SJ, Martin W, O'Keeffe J, De Jong MCM.** 2011. Trial design to estimate the effect of vaccination on tuberculosis incidence in badgers. Vet Microbiol **151:**104–111. http://dx.doi.org/10.1016/j.vetmic.2011.02.032.
9. **Department for Environment, Food, and Rural Affairs.** 2014. Monthly publication of national statistics on the incidence of tuberculosis (TB) in cattle to end of April 2014 for Great Britain. Department for Environment, Food, and Rural Affairs, London, United Kingdom. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343248/bovinetb-statsnotice-16jul14.pdf.
10. **Department for Environment, Food, and Rural Affairs.** 2013. Bovine tuberculosis evidence plan. Department for Environment, Food, and Rural Affairs, London, United Kingdom. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221077/pb13909-evidence plan-bovine-tuberculosis.pdf.
11. **Hutchings M, Harris S.** 1999. Quantifying the risks of TB infection to cattle posed by badger excreta. Epidemiol Infect **122:**167–173. http://dx.doi.org/10.1017/S0950268898001897.
12. **Benahm PF, Broom DM.** 1991. Responses of dairy cows to badger urine and faeces on pasture with reference to bovine tuberculosis transmission. Br Vet J **147:**517–532.
13. **Young JS, Gormley E, Wellington EMH.** 2005. Molecular detection of *Mycobacterium bovis* and *Mycobacterium bovis* BCG (Pasteur) in soil. Appl Environ Microbiol **71:**1946–1952. http://dx.doi.org/10.1128/AEM.71.4.1946-1952.2005.
14. **Wilesmith JW, Sayers PE, Bode R, Pritchard DG, Stuart FA, Brewer JI, Hillman GD.** 1986. Tuberculosis in East Sussex. II. Aspects of badger ecology and surveillance for tuberculosis in badger populations (1976–1984). J Hyg (Lond) **97:**11–26.
15. **Courtenay O, Wellington EMH.** 2008. *Mycobacterium bovis* in the environment: towards our understanding of its biology. Cattle Pract **16:**122–126.
16. **Ghodbane R, Mba Medie F, Lepidi H, Nappez C, Drancourt M.** 2014. Long-term survival of tuberculosis complex mycobacteria in soil. Microbiology **160:**496–501. http://dx.doi.org/10.1099/mic.0.073379-0.
17. **Department for Environment, Food, and Rural Affairs.** 2014. The strategy for achieving officially bovine tuberculosis free status for England. Department for Environment, Food, and Rural Affairs, London, United Kingdom. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300447/pb14088-bovine-tb-strategy-140328.pdf.
18. **Dalley D, Davé D, Lesellier S, Palmer S, Crawshaw T, Hewinson RG, Chambers M.** 2008. Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (*Meles meles*). Tuberculosis (Edinb) **88:**235–243. http://dx.doi.org/10.1016/j.tube.2007.11.001.
19. **Delahay RJ, Langton S, Smith GC, Clifton-Hadley RS, Cheeseman CL.** 2000. The spatio-temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density badger population. J Anim Ecol **69:**428–441. http://dx.doi.org/10.1046/j.1365-2656.2000.00406.x.
20. **Drewe JA, Tomlinson AJ, Walker NJ, Delahay RJ.** 2010. Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. PLoS One **5:**e11196. http://dx.doi.org/10.1371/journal.pone.0011196.
21. **Chambers MA, Waterhouse S, Lyashchenko K, Delahay R, Sayers R, Hewinson RG.** 2009. Performance of TB immunodiagnostic tests in Eurasian badgers (*Meles meles*) of different ages and the influence of duration of infection on serological sensitivity. BMC Vet Res **5:**42. http://dx.doi.org/10.1186/1746-6148-5-42.
22. **Travis ER, Gaze WH, Pontiroli A, Sweeney FP, Porter D, Mason S, Keeling MJC, Jones RM, Sawyer J, Aranaz A, Rizaldos EC, Cork J, Delahay RJ, Wilson GJ, Hewinson RG, Courtenay O, Wellington EMH.** 2011. An inter-laboratory validation of a real time PCR assay to measure host excretion of bacterial pathogens, particularly of *Mycobacterium bovis*. PLoS One **6:**e27369. http://dx.doi.org/10.1371/journal.pone.0027369.
23. **Pontiroli A, Travis ER, Sweeney FP, Porter D, Gaze WH, Mason S, Hibberd V, Holden J, Courtenay O, Wellington EMH.** 2011. Pathogen quantitation in complex matrices: a multi-operator comparison of DNA

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

extraction methods with a novel assessment of PCR inhibition. PLoS One **6:**e17916. http://dx.doi.org/10.1371/journal.pone.0017916.

24. **Gallagher J, Horwill DM.** 1977. A selective oleic acid albumin agar medium for the cultivation of *Mycobacterium bovis*. J Hyg (Lond) **79:**155–160. http://dx.doi.org/10.1017/S0022172400052943.
25. **King HC, Murphy A, Travis E, James P, Porter D, Hung Y-J, Sawyer J, Cork J, Delahay RJ, Gaze WH, Courtenay O, Wellington EMH.** The variability and seasonality of the environmental reservoir of *Mycobacterium bovis* shed by European badgers. Sci Rep, in press.
26. **Matthews L, Low JC, Gally DL, Pearce MC, Mellor DJ, Heesterbeek JAP, Chase-Topping M, Naylor SW, Shaw DJ, Reid SWJ, Gunn GJ, Woolhouse MEJ.** 2006. Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. Proc Natl Acad Sci U S A **103:**547–552. http://dx.doi.org/10.1073/pnas.0503776103.
27. **Matthews L, McKendrick IJ, Ternent H, Gunn GJ, Synge B, Woolhouse MEJ.** 2006. Super-shedding cattle and the transmission dynamics of *Escherichia coli* O157. Epidemiol Infect **134:**131–142. http://dx.doi.org/10.1017/S0950268805004590.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.



OPEN

The variability and seasonality of the environmental reservoir of *Mycobacterium bovis* shed by wild European badgers

Received: 17 November 2014
Accepted: 02 April 2015
Published: 06 August 2015

Hayley C. King[1], Andrew Murphy[1], Phillip James[1], Emma Travis[1], David Porter[1], Yu-Jiun Hung[1], Jason Sawyer[2], Jennifer Cork[2], Richard J. Delahay[3], William Gaze[4], Orin Courtenay[1,5] & Elizabeth M. Wellington[1]

The incidence of *Mycobacterium bovis*, the causative agent of bovine tuberculosis, has been increasing in UK cattle herds resulting in substantial economic losses. The European badger (*Meles meles*) is implicated as a wildlife reservoir of infection. One likely route of transmission to cattle is through exposure to infected badger urine and faeces. The relative importance of the environment in transmission remains unknown, in part due to the lack of information on the distribution and magnitude of environmental reservoirs. Here we identify potential infection hotspots in the badger population and quantify the heterogeneity in bacterial load; with infected badgers shedding between $1 \times 10^3 - 4 \times 10^5$ *M. bovis* cells g^{-1} of faeces, creating a substantial and seasonally variable environmental reservoir. Our findings highlight the potential importance of monitoring environmental reservoirs of *M. bovis* which may constitute a component of disease spread that is currently overlooked and yet may be responsible for a proportion of transmission amongst badgers and onwards to cattle.

The incidence of *Mycobacterium bovis* in cattle herds in Great Britain (GB) has increased from 0.01% in 1979[1] to 4.8% in 2012[2]. Control and compensation has cost the taxpayer £500 million over the past decade and this is predicted to increase to £1 billion over the next 10 years if further geographical spread is observed[3]; making bovine tuberculosis one of the most economically important animal health problems in the UK farming industry[4].

The European badger is a wildlife reservoir involved in the transmission of *M. bovis* to cattle in the UK and RoI[5,6]. Once infected, badgers may intermittently shed *M. bovis* cells in sputum, faeces and urine[7]. One likely route of transmission to cattle is through exposure to infected badger urine and faeces. Although the location and extent of environmental *M. bovis* has not been previously quantified, indirect contact with infected faeces and urine may be an important exposure pathway[8]. *M. bovis* genomic DNA can survive outside the host for up to 21 months[9] and cells have been shown to be viable by culture from mice fed soil in which *M. bovis* had been persisting for months[10]. The survival of shed *M. bovis* cells is likely to vary in space and time in relation to local environmental conditions and the distribution of infectious badgers. Understanding patterns in environmental contamination (defined as the presence of

[1]University of Warwick, School of Life Sciences, Gibbet Hill Campus, Coventry, CV4 7AL. [2]Animal and Plant Health Agency, Weybridge, Woodham lane, New Haw, Addlestone, Surrey, KT15 3NB. [3]National Wildlife Management Centre, Animal and Plant Health Agency, Woodchester Park, Nympsfield, Gloucestershire, GL10 3UJ, UK. [4]European Centre for Environmental and Human Health, University of Exeter Medical School, Knowledge Spa, Royal Cornwall Hospital, Truro, Cornwall, TR1 3HD. [5]Warwick Infectious Disease Epidemiology Research (WIDER), University of Warwick, Coventry, CV4 7AL. Correspondence and requests for materials should be addressed to H.C.K. (email: Hayley.King@Warwick.ac.uk)

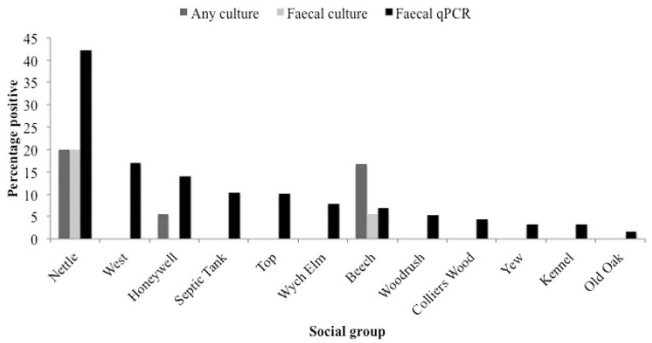


Figure 1. Percentage positive badgers per social group determined by any culture positive (tracheal or faecal) or faecal culture compared with positives by faecal qPCR. Data aggregated across the entire year.

Social group	Percentage positive faeces by qPCR	Cumulative genome equivalents in faeces	Percentage positive IFNγ	Percentage positive Stat Pak	Percentage positive IFNγ or Stat Pak
Nettle	42.2	1.08×10^6	60	100	100
West	16.9	1.48×10^6	33.3	20	40
Honeywell	13.9	4.08×10^5	50	50	66.7
Septic Tank	10.3	4.57×10^5	40	20	60
Top	10.1	9.00×10^5	20	90	90
Wych Elm	7.8	4.19×10^4	0	0	0
Beech	6.9	4.98×10^5	41.2	29.4	44.4
Woodrush	5.3	1.45×10^5	0	33.3	33.3
Colliers Wood	4.3	8.83×10^4	0	0	0
Yew	3.3	2.25×10^4	0	40	40
Kennel	3.2	2.76×10^4	0	20	20
Old Oak	1.6	2.56×10^5	0	31.3	31.3

Table 1. Summary of *M. bovis* genome equivalents counts by social group from faecal field sampling and immunoassay testing results on trapped badgers.

M. bovis genome equivalents in the environment) could aid in the design of more effective interventions, currently based on culling and vaccination strategies.

The availability of a method to quantify relative levels of environmental contamination with *M. bovis* could open up possibilities for monitoring spatial and temporal variation in risk and may help direct the implementation of disease control interventions. Currently the only means of measuring levels of infection in badger populations is through trapping and testing with BrockTB Stat Pak® (Stat Pak)[11], Interferon gamma (IFNγ)[12] and culture of clinical samples[13]. Cultivation, particularly from faecal material, has low sensitivity and is qualitative (Fig. 1). A qPCR method for non-invasive environmental monitoring of shedding was developed in our group[14,15]. This qPCR assay quantifies faecal shedding, a measure that correlates strongly with the level of infection within a social group as measured by immunoassay (Spearman's rho = 0.92, $p < 0.001$)[16]. The only other non-invasive method for monitoring infection in badger populations is culture of faecal material, which is particularly insensitive (Fig. 1). Using this optimised qPCR assay we are able to report on the spatio-temporal reservoir of *M. bovis* from badger faecal shedding in a natural population over the course of a year. Badgers defecate in latrines within or at the edges of their territories[17] and hence they can be used to identify a defined population of animals.

Methods

Sampling and Trapping. Fresh faecal samples were obtained from latrines associated with 12 badger social groups (Table 1) in Woodchester Park Gloucestershire through 2012 and 2013. Two intensive sampling periods of two weeks each were undertaken during the period of peak badger latrine activity in the spring and autumn of 2012 where up to 10 unique faecal samples were obtained from latrines associated with each social group on alternate days. Faecal samples were taken from latrines in closest proximity to

the main sett of each social group. A second sampling regime was undertaken over a year long period where up to 10 unique fresh, faecal samples were taken from latrines associated with each social group per day over two non-consecutive days in each season, starting two days after trapping operations took place in that location. For the purpose of this study March—May was classified as spring, June—August as summer, September—November as autumn and December—February as winter.

Each of the 12 badger social groups in the study was trapped four times throughout the year, once per season, with variable numbers of animals trapped between groups and seasons (Table S1). Badgers were trapped using baited cage traps placed around the main setts of each social group and identified using a unique tattoo applied at the first capture of that animal. Trapped badgers from each of the 12 social groups were tested by BrockTB Stat-Pak®, IFNγ and culture of clinical samples. All experimental protocols were approved by the University of Warwick and the Food and Environment Research Agency Ethical Review Committee and carried out in accordance with the approved guidelines and under the license granted by the Home Office under the 1986 Animal (Scientific Procedures) Act.

DNA Extraction and qPCR. Total community DNA was extracted from 0.1 g (+/−0.003 g) of faeces using the Fast DNA spin kit for soil (MP Biomedicals) following the manufacturer's instructions. *M. bovis* was detected and quantified using a qPCR assay which targets the RD4 deletion region unique to the *M. bovis* genome (Specificity data Table S2). An initial qPCR screen of each sample was performed using an ABI 7500 Fast qPCR machine (ABI) with two technical replicates of each sample. Positive controls (8.5×10^5 genome equivalents) and negative controls were also present in duplicate on each plate. PCR reactions were set up using 900 nM of each primer (RD4F $^{5'}$TGTGAATTCATACAAGCCGTAGTCG$^{3'}$, RD4R $^{5'}$CCCGTAGCGTTACTGAGAAATTGC$^{3'}$), 250 nM of Taqman probe (6FAM-AGCGCAACACTCTTGGAGTGGCCTAC—TMR), 1 mg ml^{-1} bovine serum albumen (BSA), 12.5 μl of Environmental Mastermix 2.0 (ABI), 10 μl of template and made up to 25 μl with molecular grade water (Sigma Aldrich). PCR cycling conditions were 50 °C for 2 min followed by 95 °C for 10 min then 40 cycles of 95 °C for 15 sec and 58 °C for 1 min. Samples exhibiting amplification in one or more technical replicates were taken on to full quantification using three technical replicates per sample under the same conditions. If one or more of the technical replicates of the quantification assay exhibited amplification the sample was deemed positive for *M. bovis*. Serial dilutions of *M. bovis* BCG Danish 1331 genomic DNA were used as standards for this quantification. An inhibition control assay previously described[14] was used to detect the possibility of false negative results due to inhibition. Where significant inhibition was detected DNA was re-extracted from frozen aliquots and qPCR assays were repeated. The number of *M. bovis* genome equivalents was quantified independently by qPCR at The University of Warwick and APHA Weybridge (Supplementary Figure 1).

Data Analysis. All data analysis was performed using the statistical program R. Logistic regression with social group (Old Oak) as the baseline was used to determine whether the number of positive samples varied amongst social groups throughout the year. Binomial generalised linear models (GLM) were performed to determine differences in *M. bovis* cells numbers shed between groups and between seasons. For spring two sampling days per social group were chosen to represent cross sectional sampling. Variability within groups was determined by calculating the median, upper and lower quartiles and range for each soil group.

The probability of detecting a false positive rate was 2%, calculated using known negative faecal samples obtained from captive badgers at APHA which were routinely tested for bTB using IFNγ. Negative samples were double blinded and randomly introduced into the experiment at both centres. The probability of detecting x false positive for a given number of samples was calculated using equation 1 where $p(x)$ = the probability of exactly x false positives, f = the false positive rate, n = the number of samples and x = the number of false positives.

$$p(x) = \frac{n!}{(n-x)!x!} \times (1-f)^{n-x} \times f^x \quad (1)$$

The number of confirmatory re-extractions (e) needed to result in the probability (p) of exactly x false positives was calculated using equation 2.

$$p(x) = \left(\frac{n!}{(n-x)!x!} \times (1-f)^{n-x} \times f^x \right) \times (1-f)^e \quad (2)$$

Results

Infection levels within social groups. During the study, 53.6% of trapped badgers were *M. bovis* positive by Stat-Pak, IFNγ or culture. By qPCR faecal samples from every social group examined were found to be positive (Fig. 2). Although the percentage of infected faecal samples varied considerably (Table 1, Table S3), the numbers of *M. bovis* genome equivalents per faecal sample also varied widely ranged from 1×10^3 to 4×10^5 per gram of faeces (Table 1).

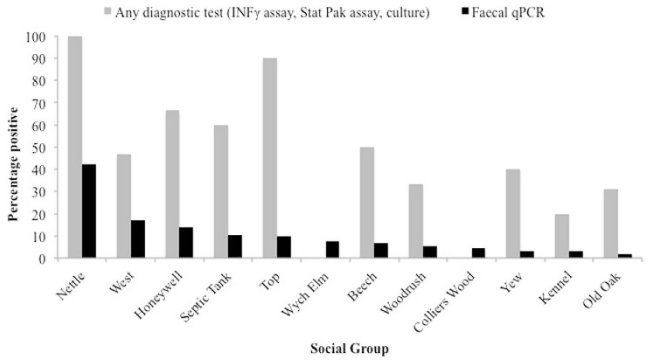


Figure 2. The percentage of badgers positive by any diagnostic tests compared to the percentage of positive faecal samples by qPCR per social group.

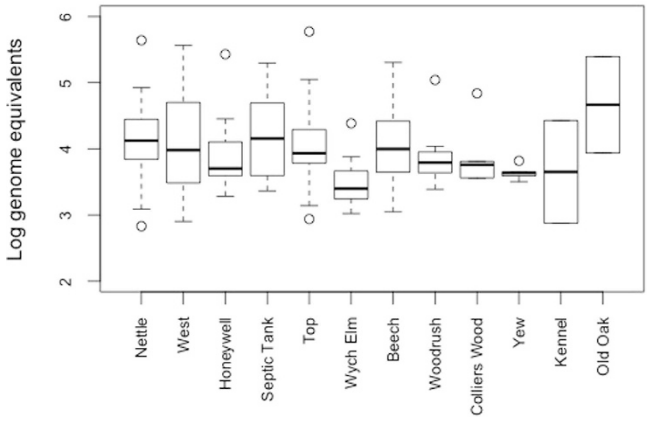


Figure 3. Distribution of *M. bovis* genome equivalents in positive samples by social group.

Significant variability in genome equivalents was identified both within and between social groups (Fig. 3) with social groups Nettle, Top, Septic Tank and West shedding more cells over the year than the other social groups (Table 1). Social groups with a high percentage of positive samples consistently shed amongst the highest cumulative numbers of *M. bovis* cells during the year (Table 1). Social group Old Oak was exceptional as it has one of the highest cumulative *M. bovis* genome equivalent values yet had the lowest percentage of positive samples in the study (Table 1). This distribution is consistent with the presence of a relatively small number of animals shedding large amounts of bacteria in some groups. However, as we could not assign faecal samples to individuals we cannot discount within-individual variation in shedding from being responsible for this observation. Hence the need for further research into heterogeneity in transmission risks amongst individual badgers.

Seasonal variability in *M. bovis* shedding. Overall a significantly greater number of *M. bovis* genome equivalents were shed in summer than in any other season. There were substantial seasonal differences in the cumulative number of *M. bovis* equivalents detected per social group (Fig. 4) with different groups identified as the largest contributors to the environmental pool of *M. bovis* throughout the year. Although summer had the highest number of genome equivalents overall, Septic Tank shed fewer cells in summer compared to other seasons and Top and shed more cells in spring. Nettle also shed fewer *M. bovis* genome equivalents in spring compared with the rest of the year. However, five social groups (Nettle, West, Honeywell, Septic Tank, and Top) were identified as having consistently high proportions of positive faeces and relatively large quantities of *M. bovis* bacilli shed (Table 1). This corresponds to immunoassay tests carried out on trapped badgers, which also identified these five groups as the most heavily infected (Table 1). Although there is strong correspondence between immunoassay and qPCR results there are some discrepancies, in particular Nettle and Top are 100% and 90% positive by immunoassay yet there was a large difference in the percentage of positive faecal samples with 42.2% and 10.0% respectively.

Figure 4. The cumulative *M. bovis* genome equivalents shed by each social group per season. Created in R version 3.0.2 using the packages ggplots 2[31] and ggmaps[32]. The scales for all graphs are identical.

Discussion

Detection of *M. bovis* by qPCR allows the presence of faecal shedding and hence infectious badgers to be established non invasively and raises the possibility of identifying infectious social groups. Unlike standard diagnostic tests the qPCR approach also quantifies levels of *M. bovis* shedding, providing opportunities to assess spatio-temporal variations in the environmental distribution of this potential source of infection for cattle, badgers and other wild mammals. Environmental transmission is likely to be a complex mixture of a number of factors including the infectious load of *M. bovis* in faeces and urine and changes in these reservoirs over time, proximity to cattle pasture, the frequency and type of contact cattle have with badger excrement and the age of faecal samples. The application of qPCR to further understand the epidemiology and transmission dynamics of bovine tuberculosis may be an important component in managing the advancing frontier between endemic and non-endemic cattle infection, and to inform transmission models (e.g. Brooks-Pollock *et al.* (2014)).

The heterogeneities observed in this study between social groups and the consistency with which five groups were identified as highly infected and shedding, suggesting that interventions targeted at particular high risk populations could have a larger impact than random and blanket control strategies[18]. However, we are mindful that any perturbation of badger populations could result in increased rather than decreased transmission[19,20]. The observed discrepancies in the percentage of positive faecal samples for social groups with similar prevalences of infection by immunoassay highlights the need for further work to establish the causes of these differences. Whilst heterogeneity in transmission is a well-known phenomenon, this study is one of the few empirical studies which have attempted to demonstrate the extent of this variability[21]. Although this study does not assess the viability of *M. bovis* in faeces, previous work has identified the presence of *M. bovis* 16S rRNA in soil[9] and badger setts and latrines[22]. In addition, studies have had a culture success rate of 2.5% from badger faecal samples[23] and *M. bovis* has been cultured from cattle faeces several months after excretion[24]. This indicates that at least a proportion of *M. bovis* cells shed in badger faeces can remain viable in the environment; however, further research is required to determine potential survival and transmissibility of *M. bovis* in environmental samples.

Whilst the focus in the UK and RoI is on badgers, other wildlife hosts are present[25,26]; however, little is currently known of their contribution to environmental reservoirs and their relative importance for transmission to cattle[25]. Issues controlling *M. bovis* are not confined to the UK and RoI. Worldwide there are problems with *M. bovis* in buffalo and lions in South Africa[27], possums in New Zealand[28], white tailed deer in America[29] and wild boar in Spain[30]. This non-invasive qPCR assay can be employed to detect shedding in other systems and samples types including milk, water and clinical tissues, is possible using this method. Whilst controlling and monitoring *M. bovis* in wildlife populations remains a challenge, non-invasive monitoring of environmental contamination may open up opportunities to identify spatio-temporal heterogeneity in disease risks and hence contribute to the development of suitable approaches fro disease control in livestock.

References

1. Garnett, B. T., Roper, T. J. & Delahay, R. J. Use of cattle troughs by badgers (*Meles meles*) A potential route for the transmission of bovine tuberculosis (*Mycobacterium bovis*) to cattle. *Appl. Anim. Behav. Sci.* **80,** 1–8 (2003).
2. Department for Environment Food and Rural Affairs, *Monthly publication of National Statistics on the Incidence of Tuberculosis (TB) in Cattle to end of April 2014 for Great Britain*, (2014) Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343248/bovinetb-statsnotice-16jul14.pdf (Accessed: 3rd August 2014).
3. Department for Environment Food and Rural Affiars, *Bovine Tuberculosis Evidence Plan*, (2013) Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221077/pb13909-evidenceplan-bovine-tuberculosis.pdf (Accessed: 23rd of April 2014).
4. Reynolds, D. A review of tuberculosis science and policy in Great Britain. *Vet. Microbiol.* **112,** 119–26 (2006).

5. Donnelly, C. A. *et al.* Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature*. **426,** 834–837 (2003).
6. Aznar, I. *et al.* Trial design to estimate the effect of vaccination on tuberculosis incidence in badgers. *Vet. Microbiol.* **151,** 104–11 (2011).
7. Clifton-Hadley, R. S., Wilesmith, J. W. & Stuart, F. A. *Mycobacterium bovis* in the European badger (*Meles meles*): epidemiological Findings in tuberculosis badgers from a naturally infected population. *Epidemiol. Infect.* **111,** 9–19 (1993).
8. Tolhurst, B. A., Delahay, R. J., Walker, N. J., Ward, A. I. & Roper, T. J. Behaviour of badgers (*Meles meles*) in farm buildings : Opportunities for the transmission of *Mycobacterium bovis* to cattle? *Appl. Anim. Behav. Sci.* **117,** 103–113 (2009).
9. Young, J. S., Gormley, E. & Wellington, E. M. H. Molecular detection of *Mycobacterium bovis* and *Mycobacterium bovis* BCG (Pasteur) in soil. *Appl. Environ. Microbiol.* **71,** 1946–52 (2005).
10. Ghodbane, R., Mba Medie, F., Lepidi, H., Nappez, C. & Drancourt, M. Long-term survival of tuberculosis complex mycobacteria in soil. *Microbiology* **160,** 496–501 (2014).
11. Chambers, M. A. *et al.* Validation of the BrockTB stat-pak assay for detection of tuberculosis in Eurasian badgers (*Meles meles*) and influence of disease severity on diagnostic accuracy. *J. Clin. Microbiol.* **46,** 1498–500 (2008).
12. Dalley, D. J. *et al.* Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (*Meles meles*). *Tuberculosis* **88,** 235–243 (2008).
13. Delahay, R. J., Langton, S., Smith, G. C., Clifton-Hadley, R. S. & Cheeseman, C. L. The spatio-temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density badger population. *J. Anim. Ecol.* **69,** 428–441 (2000).
14. Pontiroli, A. *et al.* Pathogen quantitation in complex matrices: a multi-operator comparison of DNA extraction methods with a novel assessment of PCR inhibition. *PloS one* **6,** e17916 (2011).
15. Travis, E. R. *et al.* An inter-laboratory validation of a real time PCR assay to measure host excretion of bacterial pathogens, particularly of *Mycobacterium bovis*. *PloS one* **6,** e27369 (2011).
16. Wellington, E. M., Courtenay, O. Badgers and bovine TB: how can environmental microbiology help? *Microbiol. Today* **41,** 143-144 (2014).
17. Roper, T. *et al.* Territorial Marking with Faeces in Badgers (*Meles meles*): a Comparison of Boundary and Hinterland Latrine Use. *Behaviour* **127,** 289 – 307 (1993).
18. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438,** 355–9 (2005).
19. Tuyttens, F. A. M., Delahay, R. J., Macdonald, D. W., Cheeseman, C. L., Long, B. & Donnelly, C. A. Spatial perturbation caused by a badger (*Meles meles*) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (*Mycobacterium bovis*). *J. Anim. Ecol.* **69,** 815–828 (2000).
20. Woodroffe, R. *et al.* Effects of culling in badger Meles meles spatial organization: Implications for the control of bovine tuberculosis. *J. Appl. Ecol.* **43,** 1–10 (2006).
21. Matthews, L. *et al.* Heterogeneous shedding of Escherichia coli O157 in cattle and its implications for control. *P. Natl. Acad. Sci. USA*. **103,** 547–52 (2006).
22. Courtenay, O. *et al.* Is *Mycobacterium bovis* in the environment important for the persistence of bovine tuberculosis? *Biol. Lett*ers. **2,** 460 – 462 (2006)
23. Wilesmith, J. W. *et al.* Tuberculosis in East Sussex II. Aspects of badger ecology and surveillance for tuberculosis in badger populations (1976–1984). *J. Hyg. Camb.* **97,** 11 – 26 (1986).
24. Courtenay, O., Wellington, E. M. H. *Mycobacterium bovis* in the environment: towards our understanding of its biology. *Cattle Pract.* **16,** 122–126 (2008).
25. Delahay, R. J., De Leeuw, a. N. S., Barlow, a. M., Clifton-Hadley, R. S. & Cheeseman, C. L. The Status of *Mycobacterium bovis* Infection in UK Wild Mammals: A Review. *Vet. J.* **164,** 90–105 (2002).
26. Mathews, F. *et al.* Bovine tuberculosis (*Mycobacterium bovis*) in British farmland wildlife : the importance to agriculture. *Proc Biol Sci.* **7,** 357 – 365 (2006).
27. Renwick, R. White, P. C. L. & Bengis, R. G. Bovine tuberculosis in southern African wildlife: a multi-species host-pathogen system. *Epidemiol. Infect.* **135,** 529–40 (2007).
28. Coleman, J. D. & Cooker, M. M. *Mycobacterium bovis* infection in wildlife in New Zealand. *Tuberculosis.* **81,** 191–202 (2001).
29. Miller, R. S. & Sweeney, S. J. *Mycobacterium bovis* (bovine tuberculosis) infection in North American wildlife: current status and opportunities for mitigation of risks of further infection in wildlife populations. *Epidemiol. Infect.* **141,** 1357–70 (2013).
30. Aranaz, A. *et al.* Bovine Tuberculosis (*Mycobacterium bovis*) in Wildlife in Spain. *J. Clin. Microbiol.* **42,** 2602–2608 (2004).
31. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, New York, 2009).
32. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2. *R J.* **5,** 114–161 (2013).

Acknowledgments

We acknowledge funding from Defra, H.C.K. was in receipt of a BBSRC DTG studentship and E.M.W. and O.C. acknowledge support from BBSRC for collaboration with Eamonn Gormley, UCD. We are also grateful to the APHA field team at Woodchester Park for support during fieldwork, and to Defra who fund the long-term study. We acknowledge the work of Frank Sweeney and Victoria Hibbard in producing specificity data. We are grateful to Deidre Hollingsworth for critical reading of the manuscript.

Author Contributions

H.C.K. was involved with sample collection and processing, undertook statistical analysis, prepared tables and figures and wrote the manuscript with assistance from E.M.W. P.J. collected and processed samples and undertook statistical analysis. A.M. collected and processed samples and undertook statistical analysis. E.T. undertook statistical analysis and advised on the project. D.P. collected and processed samples. Y.H. produced figures for the manuscript. J.S. coordinated and oversaw processing of samples at A.P.H.A. Weybridge. J.C. processed samples at APHA Weybridge. R.J.D. contributed to the study design, reporting and data evaluation, organised and oversaw the fieldwork at APHA Woodchester. W.G. was involved with designing the original experiment. O.C. in collaboration with E.M.W. was involved in the project design. E.M.W. in collaboration with O.C. devised the experimental set up and supervised all field and practical work and was responsible for biosafety.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: King, H. C. *et al.* The variability and seasonality of the environmental reservoir of *Mycobacterium bovis* shed by wild European badgers. *Sci. Rep.* **5**, 12318; doi: 10.1038/srep12318 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

Received: 13 October 2016 | Revised: 1 February 2017 | Accepted: 16 February 2017

DOI: 10.1002/mbo3.474

ORIGINAL RESEARCH

WILEY MicrobiologyOpen Open Access

The 'known' genetic potential for microbial communities to degrade organic phosphorus is reduced in low-pH soils

Ian D. E. A. Lidbury[1] | Tandra Fraser[2] | Andrew R. J. Murphy[1] | David J. Scanlan[1] | Gary D. Bending[1] | Alexandra M. E. Jones[1] | Jonathan D. Moore[3] | Andrew Goodall[2] | Mark Tibbett[2] | John P. Hammond[2,4] | Elizabeth M. H. Wellington[1]

[1]School of Life Sciences, University of Warwick, Coventry, West Midlands, United Kingdom

[2]School of Agriculture, Policy, and Development, University of Reading, Whiteknights, United Kingdom

[3]The Genome Analysis Centre, Norwich Research Park, Norwich, United Kingdom

[4]Southern Cross Plant Science, Southern Cross University, Lismore, Australia

Correspondence
Ian D. E. A. Lidbury, School of Life Sciences, University of Warwick, Coventry, West Midlands, United Kingdom.
Email: I.Lidbury@warwick.ac.uk

Funding information
Biotechnology and Biological Sciences Research Council (BBSRC), Grant/Award Number: BB/L026074/1

Abstract

In soil, bioavailable inorganic orthophosphate is found at low concentrations and thus limits biological growth. To overcome this phosphorus scarcity, plants and bacteria secrete numerous enzymes, namely acid and alkaline phosphatases, which cleave orthophosphate from various organic phosphorus substrates. Using profile hidden Markov modeling approaches, we investigated the abundance of various non specific phosphatases, both acid and alkaline, in metagenomes retrieved from soils with contrasting pH regimes. This analysis uncovered a marked reduction in the abundance and diversity of various alkaline phosphatases in low-pH soils that was not counterbalanced by an increase in acid phosphatases. Furthermore, it was also discovered that only half of the bacterial strains from different phyla deposited in the Integrated Microbial Genomes database harbor alkaline phosphatases. Taken together, our data suggests that these 'phosphatase lacking' isolates likely increase in low-pH soils and future research should ascertain how these bacteria overcome phosphorus scarcity.

KEYWORDS
Acid phosphatase, alkaline phosphatase, metagenomics, microbial community, microbial diversity, soil

1 | INTRODUCTION

In soil, total phosphorus (P) concentrations typically vary between 100 and 3,000 mg kg^{-1} (Hedley et al. 1995; Mengel 1997), of which 30%–60% is in the form of complex organic P esters (White & Hammond, 2008). However, plants can only acquire simple orthophosphate (Pi), which is frequently found at low concentrations in soils (<10 μmol L^{-1}) (White & Hammond, 2008). Therefore, P is often a limiting nutrient for global crop production and subsequently inorganic rock phosphate is applied in large quantities with deleterious economic and environmental consequences (Vance, Uhde-Stone, & Allan, 2003; Cordell, Drangert, & White, 2009; White & Hammond, 2008, 2009).

Microorganisms can have a beneficial effect on crop production, partly through the liberation of unavailable P in the rhizosphere and surrounding soil (Rodríguez & Fraga, 1999). Various studies have revealed that upon depletion of extracellular Pi bacteria can undergo a regulatory and physiological response resulting in the secretion of various exoproteins involved in liberating and binding Pi (Hirani, Suzuki, Murata, Hayashi, & Eaton-Rye, 2001; Rittmann, Sorger-Herrmann, & Wendisch, 2005; Monds, Newell, Schwartzman, & O'Toole, 2006; Antelmann, Scharf, & Hecker, 2000; Lidbury et al., 2016). Through this process, numerous reports have revealed that various phosphatases play an important role in the bioavailability of P in soils (Nannipieri, Newell, Giagnoni, Landi, & Renella, 2011). Usually, the most heavily

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2017 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

secreted exoenzymes are alkaline phosphatases (APases), which cleave Pi from a plethora of complex organic P monoesters and diesters. The latter compounds account for the bulk (up to 90%) of organic P in soils (Condron, Turner, & Cade-Menun, 2005; Lidbury et al., 2016; Santos-Beneit, 2015; Zaheer, Morton, Proudfoot, Yakunin, & Finan, 2009). APases encompass a wide genetic diversity including PhoA (Roy, Ghosh, & Das, 1982), PhoX (Sebastian & Ammerman, 2009) and PhoD (Eder, Shi, Jensen, Yamane, & Hulett, 1996) types. APases are secreted through various mechanisms, including the twin-arginine translocation (TAT) pathway (Putker et al., 2013) or sec pathway (PhoA) (Angelini et al., 2001). Mutagenesis of *phoX* almost entirely abolishes *para*-nitrophenyl phosphate degradation capacity (Monds et al., 2006), the substrate most frequently used to determine in situ soil phosphatase activity. Similarly, both PhoD and PhoA have been shown to be highly active against *para*-nitrophenyl phosphate (Rodriguez et al., 2014; Yang & Metcalf, 2004). APases are thought to be partly responsible for the 'P-fertilisation' effect seen when inoculating soils with plant growth-promoting bacteria (PGPB) (Condron et al., 2005; Rodríguez & Fraga, 1999).

2 | EXPERIMENTAL PROCEDURES

2.1 | Bioinformatic analyses

For each protein analyzed, between 50 and 80 sequences from a range of phylogenetically distinct soil bacteria were downloaded from the Integrated Microbial Genomes database (IMG/JGI) (https://img.jgi.doe.gov/). To identify these homologs, a combination of BLASTP and function searches (IMG search option) using the conserved PFAM domains for each protein were performed. Downloaded sequences were aligned using MUSCLE and HMM profiles were constructed using the hmmbuild function in HMMER (Eddy, 1998).

The eight metagenomes were downloaded from the EBI metagenomics portal under the project code ERP001068 (Title: Functional diversity of soil microbes across environmental gradients). The eight sample IDs for each metagenome are as follows: ERS078132, ERS078133, ERS078134, ERS078135, ERS078136, ERS078137, ERS078138, ERS078139 corresponding to CS1, CS179, CS864, CS922, CS78, CS251, CS511, CS1053 sites of the UK countryside survey, respectively. Information regarding the taxonomic profile of each site (based on the 16S rRNA gene) can be obtained directly from the EBI metagenomics portal. For each site, both FASTA files for the 'predicted CDS' and 'predicted CDS without annotation' were downloaded.

Each site was screened for the abundance of each functional gene using the hmmsearch option in HMMER and using easel to generate output files. For determining the abundance of a given functional gene at each site, the method used by Howard, Sun, Biers, and Moran (2008) was adapted. Briefly, the raw number of hits at each site was normalized to RecA by taking the ratio of the length (amino acid) of each given functional protein against the length (amino acid) of RecA. Four single copy housekeeping genes, RpoB, AtpB, GyrB, and SucD were also used. The mean number of counts retrieved for each enzyme was used to determine the 'average genome equivalent' for each site. Next, the ratio of the number of hits related to phosphorus-scavenging enzymes against each genome equivalent was calculated to provide the number of bacteria containing a given enzyme at each site.

To assess the diversity of alkaline phosphatases and GyrB, manually curated databases were established for each gene by downloading all the known homologs present in the IMG/JGI database. A BLASTP search was performed using a relatively relaxed stringency (e-10). To confirm the validity of the hits retrieved from hmmsearches, individual sequences retrieved by hmmsearches were used as queries in BLASTP searches using the nr database located on the National Centre for Biotechnology Information server (www.ncbi.nlm.nih.gov/BLAST.cgi). Higher taxonomic ranks were retrieved using the NCBI taxonomy tool. The data were visualized using the Krona tools package (Ondov et al. 2011).

3 | RESULTS AND DISCUSSION

To determine the 'genetic potential' of microbial communities to remineralize organic forms of P, we screened eight metagenomes (https://www.ebi.ac.uk/metagenomics/) from geographically distinct locations across the UK landscape (Griffiths et al., 2011). Four sites were considered low-pH soils (4.12–4.37), and four high-pH soils (8.04–8.46). Profile hidden Markov modeling (HMM) searches (Eddy, 1998) were conducted for each APase (PhoX, PhoD, PhoA), the nonspecific class A, class B, class C acid phosphatases (ACPases) (Gandhi & Chandra, 2012; Thaller, Berlutti, Schippa, Lombardi, & Rossolini, 1994), a number of other P-scavenging enzymes and several housekeeping genes (see Table S1). The number of bacteria containing P-scavenging genes was determined according to Howard et al. (2008). Based on 16S rRNA gene diversity, there were no significant changes (*T*-test, $p > .06$) in the dominant phyla (Actinobacteria, Proteobacteria, Firmicutes, Acidobacteria) between low-pH and high-pH soils (Figure S1). The diversity of GyrB at the phylum and class levels was also comparable (Figure S4) between low- and high-pH soils. Furthermore, there were no major differences in the broad functional capabilities (based on GO terms) between the high-pH and low-pH sites (Figure S2).

However, there was a significant reduction (*T*-test, unpaired, $p < .05$) in the various APases (PhoX, PhoD, PhoA) detected in low-pH soils compared to high-pH soils. In high-pH soils, 47%, 56%, and 20% of bacteria possessed genes encoding PhoX, PhoD, and PhoA, respectively, whereas in low-pH soils, only 3%, 7% and 1% of bacteria contained these genes, respectively (Figure 1A). Unexpectedly, the number of bacteria containing class A and C ACPases was similar (*T*-test, unpaired, $p > .05$) between all sites while the number of bacteria containing class B ACPases actually decreased (*T*-test, unpaired, $p = .046$) in low-pH soils (Figure 1A). Besides the nucleotidase, UshA (Rittmann et al., 2005), which also showed a significant increase in high-pH soils, the abundance of several other enzymes involved in organic P scavenging (phytases, phosphodiesterases) did not change (*T*-test, unpaired, $p > .05$) between low-pH and high-pH soils. The diversity of the three APases was comparable to that of GyrB

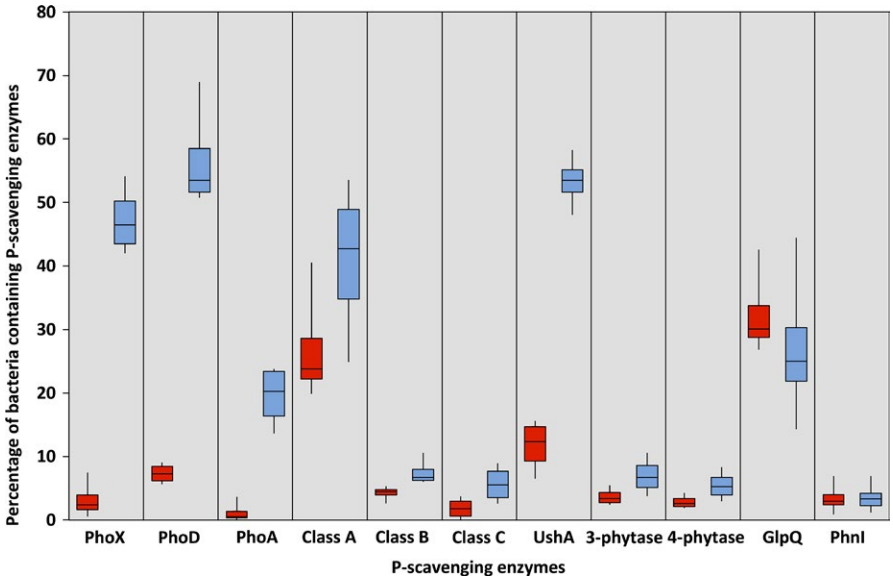


FIGURE 1 The abundance of various P-scavenging enzymes detected in the eight metagenomes retrieved from soils with contrasting pH values (red, low pH; blue, high pH). The percentage of bacteria containing phosphatases was calculated assuming a copy number of one per cell. A reduction in the three alkaline phosphatases was observed in low-pH soils while there was no concurrent increase in acid phosphatases Abbreviations: PhoX, PhoD, PhoA, alkakine phosphatases; classA/B/C, nonspecific acid phosphatases; UshA, 5′ nucleotidase, GlpQ, glycerolphosphodiester phosphodiesterase; PhnI, carbon-phosphorus lyase complex subunit I

and the 16S rRNA gene (Figure 2 and S3). For example, the majority of APase sequences (63%–73%) were related to the Proteobacteria, Actinobacteria, and Firmicutes.

Based on our curated databases for APases, the majority of sequences found in the soils analyzed here had a mean amino acid identity of 69% to those found in the Integrated Microbial Genomes database at the Joint Genome Institute (IMG/JGI). However, due to a lack of biochemical and biophysical data on the majority of APases found in less-studied bacteria that make up our curated database, for example, members of the *Betaproteobacteria* or *Bacteroidetes*, it remains to be determined whether all of the sequences related to the various APases represent *bona fide* APases. Future work should ascertain whether or

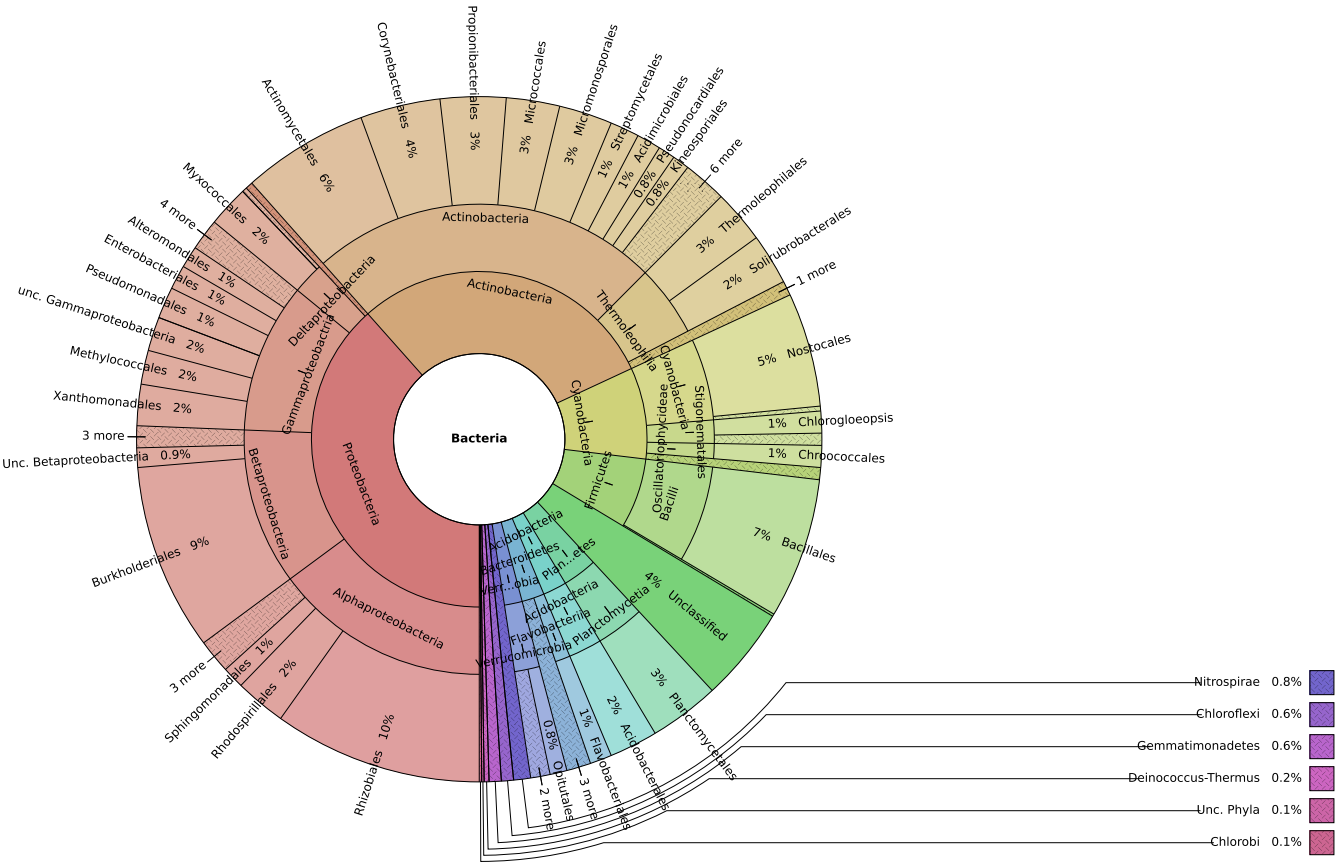


FIGURE 2 Combined taxonomic diversity of the three alkaline phosphatases (PhoX, PhoD, PhoA) in the metagenomes retrieved from all four high-pH sites (CS922, CS78, CS251, CS511). The chart was generated using the KronaTools software package (Ondov et al., 2011)

not predicted APases found in genomic databases, having significant sequence divergence from the relatively few characterized APases, function in a similar manner to those studied to date.

To determine the number of sequenced bacterial genomes possessing APases, we scrutinized all the genomes (status, 'finished') deposited in the IMG/JGI for the presence of each APase. It was discovered that approximately half of the strains related to the Proteobacteria (1510/2244), Actinobacteria (257/494), Firmicutes (471/1052), and Cyanobacteria (66/99) possess at least one of the three APases (PhoX, PhoA, PhoD). Furthermore, only 165/258 genomes (status, finished) tagged with 'soil' and 37/45 genomes (status, finished) tagged with 'rhizome/rhizoplane' under the filter 'ecosystem type' harbor one of the three APases, revealing that a number of bacteria exist that do not possess the typical phosphatases associated with overcoming Pi limitation (Lidbury et al., 2016; Santos-Beneit, 2015).

Together, this analysis revealed that low-pH soils have a marked reduction in their 'known genetic potential' to remineralize organic P. In a previous study, using a targeted amplicon approach, soil pH was also shown to be a driver of *phoD* diversity (Ragot, Kertesz, & Bünemann, 2015). Whether this reduction in genetic potential found in low-pH soils equates to a reduction in P remineralization capabilities warrants further investigation. Although APases have a pH optimum between 9 and 11, PhoX showed activity against 79 phosphomonoesters at pH 7.5 (Zaheer et al., 2009) and PhoD functions well at pH 8 (Rodriguez et al., 2014). However, it is likely that these promiscuous enzymes do not function in soils at such low-pH values. Considering that the pH optimum of ACPases is much lower (pH 4.8–7) (Rossolini et al., 1998), it was surprising that there was no observed increase in the genes encoding these enzymes in low-pH soils.

Low-pH soils may select for microbial communities containing bacteria lacking 'known' APases (PhoX, PhoD, PhoA), but which possess as-yet-unidentified phosphatases, rather than a total loss of non-specific phosphatases from bacterial genomes. This hypothesis is indirectly supported by evidence that numerous bacterial strains deposited in genome banks lack characterized APases. Given the fact that soils harbor a tremendous amount of genetic diversity (Torsvik & Øvreås, 2002) and that multiple Phyla inhabit the rhizosphere (Bulgarelli, Schlaeppi, Spaepen, van Themaat, & Schulze-Lefert, 2013; Bulgarelli et al., 2012), our knowledge of the P-scavenging abilities of soil bacteria remains relatively poor. Clearly our understanding of the microbial response to Pi limitation is limited to studies focusing on easily culturable strains. Future studies should focus on APase-lacking isolates to determine their P mineralization capabilities and thus their capacity to overcome P scarcity in soils. This should include the isolation and cultivation of various bacterial strains from low-pH soils to determine if they still elicit phosphatase activity or conversely, are reliant solely on various forms of inorganic P.

ACKNOWLEDGEMENTS

This research was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under grant BB/L026074/1.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Angelini, S., Moreno, R., Gouffi, K., Santini, C., Yamagishi, A., Berenguer, J., & Wu, L. (2001). Export of *Thermus thermophilus* alkaline phosphatase via the twin-arginine translocation pathway in *Escherichia coli*. *FEBS Letters*, *506*, 103–107.

Antelmann, H., Scharf, C., & Hecker, M. (2000). Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis. *Journal of Bacteriology*, *182*, 4478–4490.

Bulgarelli, D., Rott, M., Schlaeppi, K., Loren, V., van Themaat, E., Ahmadinejad, N., … Schulkze-Leftert, P. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature*, *488*, 91–95.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology*, *64*, 807–838.

Condron, L. M., Turner, B. L., & Cade-Menun, B. J. (2005). Chemistry and dynamics of soil organic phosphorus. In J. T. Sims, & A. N. Sharpley (Eds.), *Phosphorus: agriculture and the environment* (pp. 87–121). Soil Science Society of America Inc, Madison, WI: American Society of Agronomy.

Cordell, D., Drangert, J.-O., & White, S. (2009). The story of phosphorus: Global food security and food for thought. *Global Environmental Change*, *19*, 292–305.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, *14*, 755–763.

Eder, S., Shi, L., Jensen, K., Yamane, K., & Hulett, F. M. (1996). A *Bacillus subtilis* secreted phosphodiesterase/alkaline phosphatase is the product of a Pho regulon gene, *phoD*. *Microbiology*, *142*, 2041–2047.

Gandhi, N. U., & Chandra, S. B. (2012). A comparative analysis of three classes of bacterial non-specific acid phosphatases and archaeal phosphoesterases: Evolutionary perspective. *Acta Informatica Medica*, *20*, 167–173.

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., & Whiteley, A. S. (2011). The bacterial biogeography of British soils. *Environmental Microbiology*, *13*, 1642–1654.

Hirani, T. A., Suzuki, I., Murata, N., Hayashi, H., & Eaton-Rye, J. J. (2001). Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in *Synechocystis* sp. PCC 6803. *Plant Molecular Biology*, *45*, 133–144.

Hedley, M. J., Mortvedt, J. J., Bolan, N. S., & Syers, J. K. (1995). Phosphorus Fertility Management In Agroecosystems. In H. Tiessen (Ed.), *Phosphorus in the Global Environment: Transfers, Cycles and Management* (pp. 59–92). USA: Wiley & Sons Ltd.

Howard, E. C., Sun, S., Biers, E. J., & Moran, M. A. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental Microbiology*, *10*, 2397–2410.

Lidbury, I. D. E. A., Murphy, A. R. J., Scanlan, D. J., Bending, G. D., Jones, A. M. E., Moore, J. D., … Wellington, E. M. H. (2016). Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. *Environmental Microbiology*, *18*, 3535–3549. doi: 10.1111/1462-2920.13390

Mengel, K. (1997). Agronomic measures for better utilization of soil and fertilizer phosphates. *European Journal of Agronomy*, *7*, 221–233.

Monds, R. D., Newell, P. D., Schwartzman, J. A., & O'Toole, G. A. (2006). Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. *Applied and Environment Microbiology*, *72*, 1910–1924.

Nannipieri, P., Newell, P. D., Giagnoni, L., & Landi, L. Renella, G. (2011). Role of phosphatase enzymes in soil. *in*E. K. Bünemann*(Eds.), et al.* ,

Phosphorus in action, *soil biology* (pp. 215–243) Berlin Heidelberg, Springer-Verlag.

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a web browser. *BMC Informatics*, *12*, 385. doi: 10.1186/1471-2105-12-385

Putker, F., Tommassen-van Boxtel, R., Stork, M., Rodríguez-Herva, J. J., Koster, M., & Tommassen, J. (2013). The type II secretion system (Xcp) of *Pseudomonas putida* is active and involved in the secretion of phosphatases. *Environmental Microbiology*, *15*, 2658–2671.

Ragot, S. A., Kertesz, M. A., & Bünemann, E. K. (2015). *phoD* alkaline phosphatase gene diversity in soil. *Applied and Environment Microbiology*, *81*, 7281–7289.

Rittmann, D., Sorger-Herrmann, U., & Wendisch, V. F. (2005). Phosphate starvation-inducible gene *ushA* encodes a 5′ nucleotidase required for growth of *Corynebacterium glutamicum* on media with nucleotides as the phosphorus source. *Applied and Environment Microbiology*, *71*, 4339–4344.

Rodríguez, H., Fraga, R. (1999). Phosphate solubilizing bacteria and their role in plant growth promotion. *Advanced Biotech*, *17*, 319–339.

Rodriguez, F., Lillington, J., Johnson, S., Timmel, C. R., Lea, S. M., & Berks, B. C. (2014). Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. *Journal of Biological Chemistry*, *289*, 30889–30899.

Rossolini, M. G., Schippa, S., Riccio, L. M., Berlutti, F., Macaskie, E. L., & Thaller, C. M. (1998). Bacterial nonspecific acid phosphohydrolases: Physiology, evolution and use as tools in microbial biotechnology. *Cellular and Molecular Life Sciences*, *54*, 833–850.

Roy, N. K., Ghosh, R. K., & Das, J. (1982). Monomeric alkaline phosphatase of *Vibrio cholerae*. *Journal of Bacteriology*, *150*, 1033–1039.

Santos-Beneit, F. (2015). The Pho regulon: A huge regulatory network in bacteria. *Frontiers in Microbiology*, *6*, 402.

Sebastian, M., & Ammerman, J. W. (2009). The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME Journal*, *3*, 563–572.

Thaller, M. C., Berlutti, F., Schippa, S., Lombardi, G., & Rossolini, G. M. (1994). Characterization and sequence of PhoC, the principal phosphate-irrepressible acid phosphatase of *Morganella morganii*. *Microbiology*, *140*, 1341–1350.

Torsvik, V., & Øvreås, L. (2002). Microbial diversity and function in soil: From genes to ecosystems. *Current Opinion in Microbiology*, *5*, 240–245.

Vance, C. P., Uhde-Stone, C., & Allan, D. L. (2003). Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource. *New Phytologist*, *157*, 423–447.

White, P. J., & Hammond, J. P. (2008). Phosphorus nutrition of terrestrial plants. In P. J. White, & J. P. Hammond (Eds.), *The ecophysiology of plant-phosphorus interactions* (pp. 51–81). Dordrecht: Springer, Netherlands.

White, P. J., & Hammond, J. P. (2009). The sources of phosphorus in the waters of Great Britain. *Journal of Environmental Quality*, *38*, 13–26.

Yang, K., & Metcalf, W. W. (2004). A new activity for an old enzyme: *Escherichia coli* bacterial alkaline phosphatase is a phosphite-dependent hydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 7919–7924.

Zaheer, R., Morton, R., Proudfoot, M., Yakunin, A., & Finan, T. M. (2009). Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. *Environmental Microbiology*, *11*, 1572–1587.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lidbury IDEA, Fraser T, Murphy ARJ, et al. The 'known' genetic potential for microbial communities to degrade organic phosphorus is reduced in low-pH soils. *MicrobiologyOpen*. 2017;6:e474. https://doi.org/10.1002/mbo3.474

Gut Pathogens

RESEARCH Open Access

Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels

Mohammad Tauqeer Alam[1*], Gregory C. A. Amos[2,6], Andrew R. J. Murphy[2], Simon Murch[1], Elizabeth M. H. Wellington[2*] and Ramesh P. Arasaradnam[1,3,4,5]

Abstract

Background: Inflammatory bowel disease (IBD), is a debilitating group of chronic diseases including Crohn's Disease (CD) and ulcerative colitis (UC), which causes inflammation of the gut and affects millions of people worldwide. At different taxonomic levels, the structure of the gut microbiota is significantly altered in IBD patients compared to that of healthy individuals. However, it is unclear how these IBD-affected bacterial groups are related to other common bacteria in the gut, and how they are connected across different disease conditions at the global scale.

Results: In this study, using faecal samples from patients with IBD, we show through diversity analysis of the microbial community structure based on the 16S rRNA gene that the gut microbiome of IBD patients is less diverse compared to healthy individuals. Furthermore, we have identified which bacterial groups change in abundance in both CD and UC compared to healthy controls. A substantial imbalance was observed across four major bacterial phyla including Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria, which together constitute > 98% of the gut microbiota. Next, we reconstructed a bacterial family co-abundance network based on the correlation of abundance profiles obtained from the public gut microbiome data of > 22,000 samples of faecal and gut biopsies taken from both diseased and healthy individuals. The data was compiled using the EBI metagenomics database (Mitchell et al. in Nucleic Acids Res 46:D726–D735, 2018). By mapping IBD-altered bacterial families to the network, we show that the bacterial families which exhibit an increased abundance in IBD conditions are not well connected to other groups, implying that these families generally do not coexist together with common gut organisms. Whereas, the bacterial families whose abundance is reduced or did not change in IBD conditions compared to healthy conditions are very well connected to other bacterial groups, suggesting they are highly important groups of bacteria in the gut that can coexist with other bacteria across a range of conditions.

Conclusions: IBD patients exhibited a less diverse gut microbiome compared to healthy individuals. Bacterial groups which changed in IBD patients were found to be groups which do not co-exist well with common commensal gut bacteria, whereas bacterial groups which did not change in patients with IBD were found to commonly co-exist with commensal gut microbiota. This gives a potential insight into the dynamics of the gut microbiota in patients with IBD.

Keywords: Inflammatory bowel disease, Crohn's disease, Ulcerative colitis, Gut microbiota, Microbial imbalance

Introduction

Inflammatory bowel disease (IBD), a group of chronic intestinal disorders including Crohn's disease (CD) and ulcerative colitis (UC), causes inflammation of the gut and affects millions of people worldwide [1–4]. Both CD and UC diseases are differentiated by their location and levels of inflammation in the gastrointestinal (GI) tract.

*Correspondence: T.Alam@warwick.ac.uk; E.M.H.Wellington@warwick.ac.uk
[1] Warwick Medical School, University of Warwick, Coventry, UK
[2] School of Life Sciences, University of Warwick, Coventry, UK
Full list of author information is available at the end of the article

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

UC mostly involves inflammation to the rectum and colon, whereas CD most often affects the terminal ileum and colon though in some cases it can affect any part of the GI tract [2, 5]. Currently, there is no full cure for IBD, but different treatments such as taking steroids, immunosuppressants, liquid diet or surgery can help in reducing the symptoms [5]. To date, the exact cause of IBD is not understood, however, a combination of genetic variants, environmental factors, deregulated host immune system, and gut microbiota dysbiosis is associated with IBD [6–16].

More than 215 IBD-associated loci have been identified so far from various genome-wide association studies (GWAS) [7]. It has been reported that a large percentage (~30%) of these loci are common between CD and UC, showing involvement of common biological processes in both conditions [14, 17]. Moreover, these IBD-associated loci are mostly involved with immune system deregulation, a process which the gut microbiome has also been implicated in [18]. The gut microbiota, which has a complex community of a hundred trillion bacterial and archaeal cells comprising more than a thousand species, provides benefits to the host such as short-chain fatty acids (SCFA) and amino acids, metabolism of undigested carbohydrate, and stimulation of the immune system [19, 20]. In patients with IBD, the structure and the composition of the gut microbiota is severely altered compared to that of a healthy condition [10, 13, 18, 21]. Previous work has reported imbalances in IBD patients for the Firmicutes and Bacteroidetes at the phylum level, and Ruminococcaceae, Veillonellaceae, Christensenellaceae, Bacteroidaceae, and Rikenellaceae at the family level. However, there is a large degree of variability across studies, with many reporting contradictory findings. In particular, it is unclear what the relationship is between microbial groups when there is inflammation of the gut epithelium during various diseased states. There is furthermore, a general knowledge gap in establishing the relationships between microbial groups across different disease conditions.

The aim of the current work was to investigate the relationships between changing microbial groups in IBD. In particular, we wanted to understand which microbial groups differ during IBD, and how these groups differ in co-abundance patterns across a variety of diseases at the global scale. To do this, we initially investigated the gut microbial imbalance, at different taxonomic levels for healthy volunteers and CD and UC patients. We next reconstructed a network of the co-abundance patterns of different bacterial groups using publicly available data from a variety of studies at a global scale. Our results indicate that the bacterial groups which increase in abundance during IBD are specific to both CD and UC conditions. In comparison, bacterial groups which did not change in abundance during different disease states are well connected in global networks, giving us a better understanding of the dynamics of the microbiome in both health and disease.

Results and discussion

Gut microbial richness in IBD patients

We collected faecal samples from 30 individuals (9 CD patients, 11 UC patients and 10 healthy volunteers) and performed 16S rRNA taxonomic profiling to understand changes in community structure during disease with resulting data analysed using the DADA2 pipeline. Amplicon sequence variants (ASVs) were used as a measure of diversity. As previously reported, species richness in the gut of IBD patients (both CD and UC) was lower than that of healthy volunteers [10, 13]. Moreover, within the IBD patients groups, the gut of CD patients exhibit substantially lower species richness than that of UC patients (Fig. 1a). As a measure of diversity, we identified a total of 2261 ASVs, of which 81% belong to the Firmicutes, 9.25% Bacteroidetes, 5.13% Proteobacteria, and 3.14% ASVs are from Actinobacteria. Combined, these four phyla constitute more than 98% of the total identified ASVs [22, 23]. To understand how this compared to other studies, we investigated the global microbial species-richness in the human gut across >22,000 samples from 113 different studies from a variety of conditions (EBI metagenomics accession numbers [1] in Additional file 1: Table S1). We have considered only the known species in each study and made a unique list of gut bacterial species across studies. Similar to our experimental observations, the four phyla including Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria comprised >94% of the total diversity in the gut, with the Firmicutes being the most species-rich phylum across conditions (Fig. 1b). However, the number of species identified as belonging to the Firmicute Phyla in our dataset (81% of all identified species from 20 IBD and 10 healthy condition samples) is substantially higher than what is usually reported at a global level (38.18%). As the observation comes from 2/3 of IBD patient and 1/3 of healthy control samples, this increased level of Firmicutes species-richness was attributed to the high number of IBD samples. Analysis of global studies for species-richness of Proteobacteria demonstrated this phyla usually accounts for 31.18% of all species, however this was substantially lower in our dataset (5.13%). Finally, the richness of Bacteroidetes was also reduced in our dataset compared to the global datasets (9.24% in our study compared to 14.35% globally). We find a similar observation when considering all OTUs from individual EBI gut microbiome studies (Additional file 2: Figure S1).

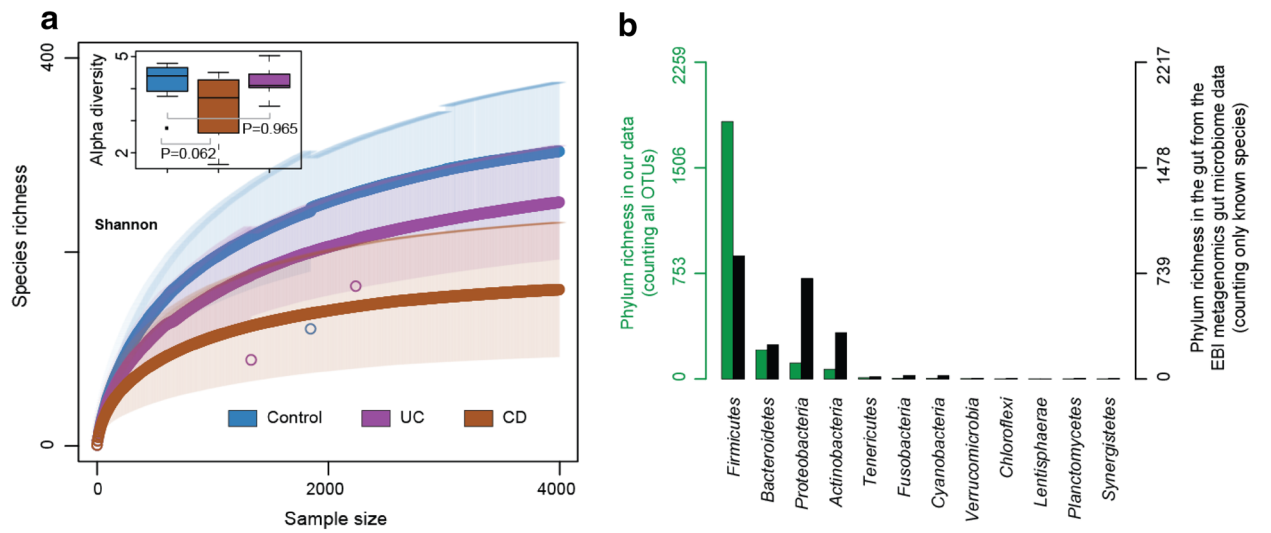


Fig. 1 Microbial diversity and richness. **a** Species richness is substantially less in CD patient samples compared to the healthy control and UC patient samples. Shannon alpha diversity plot demonstrate that CD patients samples are less diverse compared to the healthy control and UC patient samples (inset figure). **b** Phylum level richness in the gut microbiota from our samples (left side Y-axis shown in green) compared to the global gut microbial species richness obtained across more than 20 K samples from a variety of conditions (right side Y-axis shown in black)

After identifying the differences at different taxonomic levels for each disease condition, for future work it is crucial to understand the reasons for such dysbiosis and whether they are causative or consequential of disease. Studies suggest that metabolic dependency [24] and nutritional preferences [25] between microorganisms are driving forces in microbial community formation. For example, metabolic cooperation between bacteria is crucial to microbial assemblages and changes to this could cause shifts across the whole community. For future work, it would be interesting to investigate the microbial metabolic interactions during disease and how this compares to a healthy gut.

The gut microbial abundance at different taxonomic levels in IBD patients

Compared to the healthy controls, both IBD patient groups (CD and UC patients) demonstrated strong microbial imbalance at different taxonomic levels (Fig. 2). At the phylum level, both IBD conditions exhibit an increased abundance of Firmicutes and Actinobacteria, relative to the controls. In particular, the abundance of this phyla during UC, was far greater than CD or healthy controls. For the two other dominant bacterial phyla, the Proteobacteria and Bacteroidetes, the abundance profiles varied across disease conditions (Fig. 2a). In CD, the abundance of Bacteroidetes, which is often associated with a healthy gut, was deceased 2.4-fold, whereas the abundance of Proteobacteria, a phyla associated including wide variety of pathogens, was increased 3.8 fold. Interestingly, for patients with UC the abundance of Proteobacteria was decreased (3.4 fold) relative to controls and there was no significant differences in levels Bacteroidetes [26]. Several studies have reported the gut microbial imbalances for IBD, however, the imbalance at the level of different phylum is variable across studies [10, 13, 18, 21, 27, 28]. This could likely be a reflection on the lack of standardisation across microbiome techniques, or perhaps due to a heterogeneity in the microbiome associated with disease.

We further investigated how different taxonomic levels belonging to each of the main four phyla, Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria, were changed during IBD. For Firmicutes, the most abundant phylum in the gut in all conditions, we observed four classes and nine different families which changed in abundance relative to healthy controls. For CD patients, the abundance of two classes including Clostridia and Erysipelotrichia, was reduced, and three families including Ruminococcaceae, Christensenellaceae and Erysipelotrichaceae were reduced relative to healthy controls. The level of two other two classes such as Negativicutes and Bacilli (obligately aerobic) and five families including Veillonellaceae, Lactobacillaceae, Acidaminococcaceae, Streptococcaceae and Peptostreptococcaceae was increased, similar to the imbalance in their parent phylum Firmicutes. Interestingly, Lachnospiraceae, the most abundant Firmicutes family, was at a similar level to the control. For UC patients, the abundance of the Firmicute classes Clostridia, Negativicutes and Bacilli, and

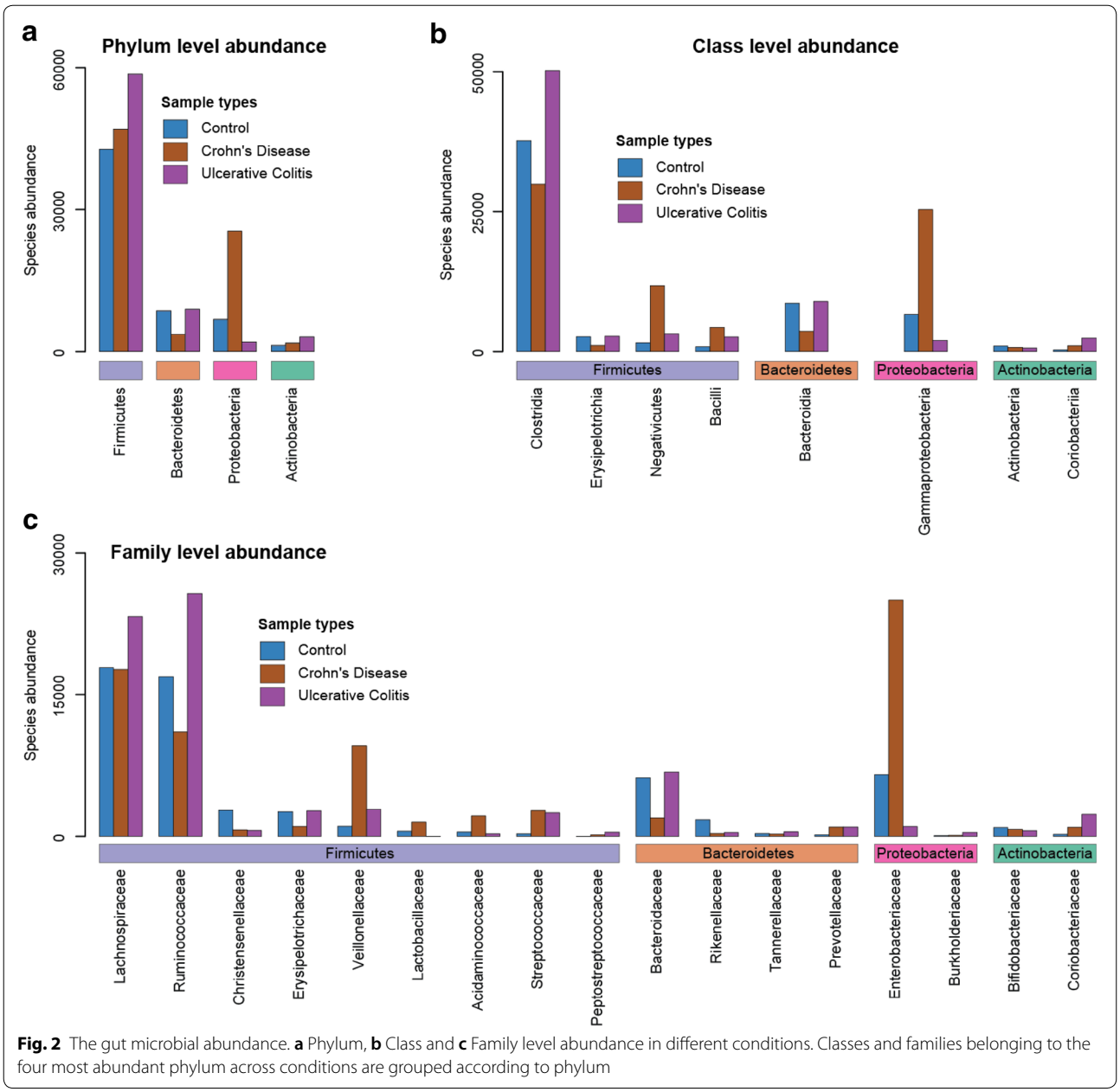


Fig. 2 The gut microbial abundance. **a** Phylum, **b** Class and **c** Family level abundance in different conditions. Classes and families belonging to the four most abundant phylum across conditions are grouped according to phylum

Firmicute families Ruminococcaceae, Lachnospiraceae, Veillonellaceae, Streptococcaceae and Peptostreptococcaceae were increased. The Erysipelotrichia class and Erysipelotrichaceae family were the same as the controls, whereas, three families including Acidaminococcaceae, Christensenellaceae and Lactobacillaceae, were reduced in abundance. For Bacteroidetes, which is the only reduced phylum in CD, we observed a reduced abundance in the Bactersoidia class and Bacteroidia families Bacteroidaceae and Rikenellaceae. The Prevotellaceae family, in particular, was increased in CD patients. For UC patients, the only families to change of the Bacteroidetes was the Rikenellaceae and Tannerellaceae families which were decreased in abundance, and the Prevotellaceae which increased in abundance as with CD. Finally, for the Proteobacteria phylum, we observed an imbalance in Enterobacteriaceae and Burkholderiaceae families, with the abundance level of Enterobacteriaceae increased in CD patients and decreased in UC patients compared to the controls. Burkholderiaceae abundance was increased for both CD and UC patients. Finally, for the Actinobacteria phylum, the abundance level of both the

class Coriobacteriia and family Coriobacteriaceae was increased in both IBD conditions relative to the controls, whereas the class Actinobacteria and family Bifidobacteriaceae was reduced [13, 29]. In summary, we demonstrate that multiple families of a class, classes of a phyla differ between both the IBD conditions, and between each IBD condition and healthy control. This suggests that changes in one bacterial family has consequences for others. To investigate this further, we used co-occurrence network analysis to establish patterns of how bacterial groups increase and decrease across global studies.

Global co-abundance in the gut of different bacterial families

The gut microbiota abundance profiles from > 22,000 samples across a variety of conditions from 113 different studies was analysed to explore how different bacterial groups change across global studies. Using the Pearson's correlation test, we built a network of significantly co-abundant (Pearson's correlation coefficient > 0.3 and P value < 1e−10) bacterial families across a range of conditions collected from global studies (Fig. 3ai). We observed that the majority of bacterial families in the network belonged to the phyla Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes (Fig. 3aii). In the global gut bacterial family–family co-abundance network (Fig. 3ai), family nodes of 45% of the total connections are from the same phylum, compared to a random network of the same size where percentage connections were much lower (~ 30%) (Fig. 3aiii). This suggests that groups from the same phylum, which likely have similar metabolic requirements, are likely to change abundance as a collective. To understand the family level microbial imbalance during IBD, we further considered a subnetwork of the global family co-abundance network, where at least one family node was from the most abundant bacterial families in either CD, UC or healthy subjects. In this subnetwork, we highlighted bacterial families which were changed during CD or UC, compared to the healthy (Fig. 3b, c).

In CD, seven bacterial families including Coriobacteriaceae, Prevotellaceae, Burkholderiaceae, Veillonellaceae, Streptococcaceae, Pseudomonadaceae and Acidaminococcaceae have increased abundance compared to the healthy controls (Fig. 2c), however, only two families including Prevotellaceae and Veillonellaceae are connected in the global network (Fig. 3b). In contrast, the level of five other families including Erysipelotrichaceae, Christensenellaceae, Ruminococcaceae, Bacteroidaceae and Rikenellaceae, were well connected in the global network (Fig. 3b) and had reduced abundance in CD (Fig. 2c). Similarly, for UC, families with an increased abundance in UC were less well connected on the global scale (Fig. 3c). This suggests that bacterial groups which increase in abundance during IBD, are not typically associated with the healthy gut microbiome, nor do they commonly co-exist with commensals observed in the healthy gut. Furthermore, families which had increased abundance levels in healthy conditions compared to CD are very well connected, suggesting microbes in the gut of healthy individuals exist as a co-operative microbial assemblage. In particular, the connection between families such as Bacteroidaceae and Ruminococcaceae in the co-abundance network indicates that they may coexist together in across conditions, potentially due to similarities in physiology or the presence of metabolically cooperating species. For future work, it would be highly interesting to examine species of these families and investigate the relationships between these organisms.

Conclusions

In summary, our analysis demonstrates that IBD patients (both CD and UC) and healthy volunteers have reduced species richness, and imbalances in families, classes, and phyla, relative to healthy volunteers. Four bacterial phyla including Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria comprised > 98% of the species in this study. To understand how bacteria assemblages depend on co-operation, we reconstructed a large co-abundance network based on the public gut microbiome data of > 22,000 samples. From this we demonstrated that the bacterial families which have an increased abundance level in IBD conditions are not well connected to other bacterial groups in the global family co-abundance network. This suggests that these bacteria do not co-exist with healthy gut microbial commensals and supports the concept that healthy assemblages are dependent on metabolic co-operation, due to the high connectivity of bacterial groups found in healthy conditions across > 22,000 samples.

Methodology

Sample collection and DNA extraction

Samples were collected from patients in standard 300 ml sterilin tubes and frozen immediately in −80 °C. Patients were asked to produce the first-morning sample for consistency and to avoid alcohol the previous 24 h. Samples were thawed and DNA was extracted using FastDNA Spin Kit for Soil (MPBiomedicals) [30] as per the manufacturer's instructions.

16S rRNA sequencing

454 pyrosequencing using 16S universal eubacterial primers 27F and 534R [31] was performed by Molecular Research (MRDNA), Shallowater, Texas, using an adapted protocol developed in [32]. Number of reads per

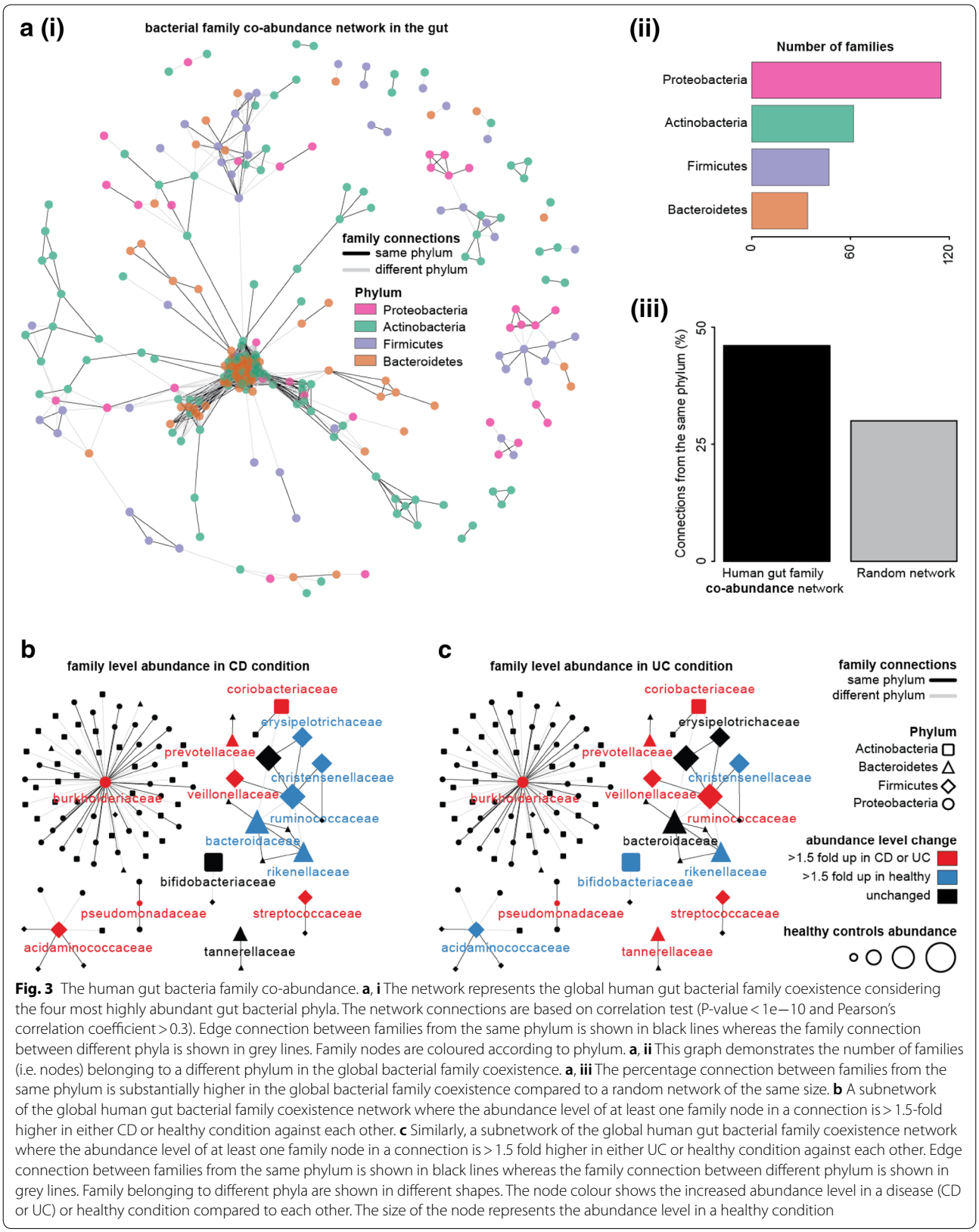


Fig. 3 The human gut bacteria family co-abundance. **a**, **i** The network represents the global human gut bacterial family coexistence considering the four most highly abundant gut bacterial phyla. The network connections are based on correlation test (P-value < 1e−10 and Pearson's correlation coefficient > 0.3). Edge connection between families from the same phylum is shown in black lines whereas the family connection between different phyla is shown in grey lines. Family nodes are coloured according to phylum. **a**, **ii** This graph demonstrates the number of families (i.e. nodes) belonging to a different phylum in the global bacterial family coexistence. **a**, **iii** The percentage connection between families from the same phylum is substantially higher in the global bacterial family coexistence compared to a random network of the same size. **b** A subnetwork of the global human gut bacterial family coexistence network where the abundance level of at least one family node in a connection is > 1.5-fold higher in either CD or healthy condition against each other. **c** Similarly, a subnetwork of the global human gut bacterial family coexistence network where the abundance level of at least one family node in a connection is > 1.5 fold higher in either UC or healthy condition against each other. Edge connection between families from the same phylum is shown in black lines whereas the family connection between different phylum is shown in grey lines. Family belonging to different phyla are shown in different shapes. The node colour shows the increased abundance level in a disease (CD or UC) or healthy condition compared to each other. The size of the node represents the abundance level in a healthy condition

sample ranged from 6936 to 100,972, with an average of 38,931 reads per sample.

Bioinformatic analysis of 16S rRNA sequencing data

16S rRNA high-throughput sequencing data was analysed by following the workflow from Callahan et al. [33]. Quality checking, filtering and trimming of fastq files were performed by functions from the dada2 package in R [34]. After filtering the reads, high-resolution Amplicon Sequence Variants (ASVs) were inferred using dada function [34]. ASVs are a higher-resolution analogue of the traditional OTUs. Chimeric sequences were removed and taxonomy assigned to ASVs based on the naive Bayesian classifier method with silva_nr_v132_train_set.fa as the training set [34]. Species-richness and alpha diversity (Shannon) were analysed by plot_richness function from the phyloseq package in R [35]. To make the rarefaction species richness curve, 'rarecurve' function from the vegan package [36] in R was used.

Statistical analysis: bacterial family co-abundance network based on microbiome data

Taxonomic assignments, containing a detailed taxonomy and abundance data of OTUs or ASVs in samples, of 113 gut microbiome studies, covering more than 22,000 samples, were downloaded from the EBI metagenomics database [1]. The data were then parsed and tables containing bacterial abundance from different phyla, classes and families were generated. Abundance at the level of phylum was then used for the global gut microbial abundance. The family-level bacterial abundance was used to construct the bacterial family–family coexistence network. For each pair of bacterial families, Pearson's correlation test was performed. Family nodes were connected when P-value $< 1e{-}10$ and Pearson's correlation coefficient > 0.3.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13099-019-0341-6.

Additional file 1: Table S1. List of gut microbiome studies from EBI metagenomics database.

Additional file 2: Figure S1. Phylum level richness in the gut microbiota from our samples (shown in green) compared to the gut microbial species richness obtained across studies, each with more than 100 OTUs (total 81 study), from a variety of conditions (shown in black).

Acknowledgements
We thank BBSRC (BBSRC BB/L027801/1) and FAPIC (EU H2020 Proposal Number: SEP-210176939 2015-2020) for funding the project. MTA was Funded by the Warwick medical school.

Authors' contributions
The study was conceived by RPA and EMHW and they secured the funding. RPA and SM recruited the patients. Sequencing was performed by GCAA and ARJM. MTA performed the data analysis. The first draft of the manuscript was written by MTA. All authors read and approved the final manuscript.

Funding
EMHW and GCAA received financial support from BBSRC (BBSRC BB/L027801/1), Warwick Medical School and AM from FAPIC—Fast Assay for Pathogen Identification and Characterisation EU H2020 Proposal Number: SEP-210176939 2015-2020.

Availability of data and materials
Sequence data of the study has been submitted to European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/submit/sra/#studies) under the Accession Number PRJEB33711

Ethics approval and consent to participate
Patients were recruited as part of the Famished Multicentre Study (Warwickshire local research Ethics No: 09/H1211/38).

Consent for publication
Not required.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Warwick Medical School, University of Warwick, Coventry, UK. [2] School of Life Sciences, University of Warwick, Coventry, UK. [3] Department of Gastroenterology, University Hospitals Coventry & Warwickshire NHS Trust, Clifford Bridge Road, Coventry CV2 2DX, UK. [4] School of Life Sciences, University of Coventry, Coventry, UK. [5] Faculty of Life Science, University of Leicester, Leicester, UK. [6] Present Address: G.C.A.A National Institute for Biological Standards and Control (NIBSC), Potters Bar, UK.

Received: 19 September 2019 Accepted: 12 December 2019
Published online: 04 January 2020

References

1. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, et al. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res. 2018;46:D726–35.
2. Abraham C, Cho JH. Inflammatory bowel disease. N Engl J Med. 2009;361:2066–78.
3. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol. 2015;12:720–7.
4. Mozdiak E, O'Malley J, Arasaradnam R. Inflammatory bowel disease. BMJ. 2015. h4416.
5. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. 2012;380:1590–605.
6. McGovern DPB, Kugathasan S, Cho JH. Genetics of inflammatory bowel diseases. Gastroenterology. 2015;149(1163–76):e2.
7. Luo Y, de Lange KM, Jostins L, Moutsianas L, Randall J, Kennedy NA, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. Nat Genet. 2017;49:186–92.
8. Molodecky NA, Kaplan GG. Environmental risk factors for inflammatory bowel disease. Gastroenterol Hepatol. 2010;6:339–46.
9. Ananthakrishnan AN. Environmental risk factors for inflammatory bowel disease. Gastroenterol Hepatol. 2013;9:367–74.
10. Sheehan D, Moran C, Shanahan F. The microbiota in inflammatory bowel disease. J Gastroenterol. 2015;50:495–507.
11. Goldsmith JR, Sartor RB. The role of diet on intestinal microbiota metabolism: downstream impacts on host immune function and health, and therapeutic implications. J Gastroenterol. 2014;49:785–98.
12. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47:979–86.

13. Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. Clin J Gastroenterol. 2018;11:1–10.
14. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307–17.
15. Lee SH, Kwon JE, Cho M-L. Immunological pathogenesis of inflammatory bowel disease. Intest Res. 2018;16:26–42.
16. Shaw KA, Cutler DJ, Okou D, Dodd A, Aronow BJ, Haberman Y, et al. Genetic variants and pathways implicated in a pediatric inflammatory bowel disease cohort. Genes Immun. 2019;20:131–42.
17. Ek WE, D'Amato M, Halfvarson J. The history of genetics in inflammatory bowel disease. Ann Gastroenterol Hepatol. 2014;27:294–303.
18. Belkaid Y, Hand TW. Role of the Microbiota in Immunity and Inflammation. Cell. 2014;157:121–41.
19. Nishida AH, Ochman H. A great-ape view of the gut microbiome. Nat Rev Genet. 2019;20:195–206.
20. Cani PD. Human gut microbiome: hopes, threats and promises. Gut. 2018;67:1716–25.
21. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4:293–305.
22. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308:1635–8.
23. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science. 2008;320:1647–51.
24. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci USA. 2015;112:6449–54.
25. Tramontano M, Andrejev S, Pruteanu M, Klünemann M, Kuhn M, Galardini M, et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat Microbiol. 2018;3:514–22.
26. Gophna U, Sommerfeld K, Gophna S, Doolittle WF, van Zanten SJO. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. J Clin Microbiol. 2006;44:4136–41.
27. Vester-Andersen MK, Mirsepasi-Lauridsen HC, Prosberg MV, Mortensen CO, Träger C, Skovsen K, et al. Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. Sci Rep. 2019;9:13473.
28. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci USA. 2007;104:13780–5.
29. LeBlanc JG, Laiño JE, del Valle MJ, Vannini V, van Sinderen D, Taranto MP, et al. B-group vitamin production by lactic acid bacteria—current knowledge and potential applications. J Appl Microbiol. 2011;111:1297–309.
30. King HC, Murphy A, James P, Travis E, Porter D, Sawyer J, et al. Performance of a Noninvasive test for detecting *Mycobacterium bovis* shedding in European Badger (*Meles meles*) populations. J Clin Microbiol. 2015;53:2316–23.
31. Tamaki H, Wright CL, Li X, Lin Q, Hwang C, Wang S, et al. Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. PLoS ONE. 2011;6:e25263.
32. Dowd SE, Sun Y, Wolcott RD, Domingo A, Carroll JA. Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: bacterial diversity in the ileum of newly weaned Salmonella-infected pigs. Foodborne Pathog Dis. 2008;5:459–72.
33. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor workflow for Microbiome Data analysis: from raw reads to community analyses. F1000 Res. 2016;5:1492.
34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
35. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 2013;8:e61217.
36. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2018. https://CRAN.R-project.org/package=vegan.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

CLINICAL VETERINARY MICROBIOLOGY

Evaluation of a Fecal Shedding Test To Detect Badger Social Groups Infected with *Mycobacterium bovis*

Andrew R. J. Murphy,[a] Emma R. Travis,[a] Victoria Hibberd,[a] David Porter,[a] Elizabeth M. H. Wellington[a]

[a]School of Life Sciences, University of Warwick, Coventry, United Kingdom

ABSTRACT Bovine tuberculosis (bTB) is an economically important disease affecting the cattle industry in England and Wales. bTB, caused by *Mycobacterium bovis*, also causes disease in the Eurasian badger (*Meles meles*), a secondary maintenance host. Disease transmission between these two species is bidirectional. Infected badgers shed *M. bovis* in their feces. The Animal and Plant Health Agency (APHA) of the United Kingdom organized a comparative trial to determine the performance of tests in detecting *M. bovis* in badger feces for the Department for Environment, Food, and Rural Affairs (DEFRA). Here, we assessed the performance of the existing Warwick Fast24-qPCR test and its modified version based on a high-throughput DNA extraction method (Fast96-qPCR). We found Fast24-qPCR to have a sensitivity of 96.7% (95% confidence interval [CI], 94.5 to 99%; $n = 244$) and a specificity of 99% (95% CI, 97.8 to 100%; $n = 292$). Fast96-qPCR requires further optimization. Determining the disease status of badger social groups requires multiple tests per group. Therefore, to increase specificity further, we independently repeated the Fast24-qPCR test on positive samples, increasing stringency by requiring a second positive result. Fast24-qPCR with repeat testing had a sensitivity of 87.3% (95% CI, 83.1 to 91.5%; $n = 244$), and a specificity of 100% (95% CI, 100 to 100; $n = 201$) on an individual-sample level. At the social-group level, this repeat testing gives Fast24-qPCR high herd specificity, while testing multiple samples per group provides high herd sensitivity. With Fast24-qPCR, we provide a social-group-level test with sufficient specificity and sensitivity to monitor shedding in badgers via latrine sampling, delivering a potentially valuable tool to measure the impacts of bTB control measures.

KEYWORDS badger, bovine tuberculosis, disease reservoir, *Mycobacterium bovis*, cattle, environmental microbiology, molecular epidemiology

Bovine tuberculosis (bTB) caused by *Mycobacterium bovis* is an economically important disease that is estimated to cost £100 million to the taxpayer per year (1) in the United Kingdom. Prevalence of the disease in cattle herds has increased from 0.49% in 1979 (1) to 5.3% in England and 5.6% in Wales in 2019 (2), though it is not evenly distributed and is concentrated in the high-risk area of southwestern England and the surrounding edge area, in addition to southwestern and eastern Wales. In Britain, the Eurasian badger (*Meles meles*) is considered a secondary maintenance host of bTB (3), with an estimated bTB prevalence of 24.2% (4) within the high-risk area of England. Although cattle-to-cattle aerosol transmission is considered the predominant route of infection (5), badgers are known to transmit disease to cattle (6, 7) and are estimated to contribute up to 52% of individual cases within areas of endemicity (8), inclusive of subsequent cattle-cattle transmission. Cattle are also implicated in the transmission of bTB to badgers, as delays in removing infected cattle have been shown to increase bTB prevalence in badgers (7).

The route of transmission between badgers and cattle has not been demonstrated. However, several lines of evidence suggest an important role for transmission occurring

Citation Murphy ARJ, Travis ER, Hibberd V, Porter D, Wellington EMH. 2021. Evaluation of a fecal shedding test to detect badger social groups infected with *Mycobacterium bovis*. J Clin Microbiol 59:e01226-20. https://doi.org/10.1128/JCM.01226-20.

Editor Brad Fenwick, University of Tennessee at Knoxville

Copyright © 2020 Murphy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Andrew R. J. Murphy, a.r.j.murphy@warwick.ac.uk.

Received 4 June 2020
Returned for modification 3 July 2020
Accepted 2 October 2020
Accepted manuscript posted online 14 October 2020
Published 17 December 2020

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

through contamination of the environment. Direct contact between badgers and cattle is a rare event (9, 10), and cattle-cattle and badger-badger transmission rates are also low (11–13), despite high levels of intraspecies contact in these social animals. This suggests that the levels of direct contact are not high enough to explain the levels of interspecific transmission. Although badgers avoid direct contact with cattle (14), they actively favor cattle pasture for foraging, suggesting common occurrence of shared environment. *M. bovis* has been shown to persist in the environment in a number of studies (15–17), and models of bTB transmission in cattle have also suggested a substantial role for the environment (18) as a reservoir and route of transmission.

Various studies have also implicated the environment as a vector for bidirectional *M. bovis* transmission between cattle and badgers. Badgers sampled from a natural population shed *M. bovis* cells in sputum, urine, and feces (19, 20), indicating that contamination of pasture with feces and urine creates a potential source of infection (21). Furthermore, cattle do not avoid areas contaminated with badger urine and will graze at badger latrines given sufficient competition for fresh pasture (22). Similarly, the feces of infected cattle contained viable *M. bovis* (23, 24), though cattle were not shown to be infected from pastures contaminated with these feces, and, as these experiments were performed in the 1930s, it cannot be assumed that disease progression occurs in the same fashion under modern bTB testing regimens. However, at the point of detection by the single intradermal comparative tuberculin test (SICCT), the test used in the United Kingdom to detect bTB in cattle (2), the disease has progressed to the extent that 55.5% of positive animals have visible lesions at slaughter (25), compared to a background rate of 0.63/1,000 (0.063%) in negative animals (26). The spreading of slurry is also considered a risk factor for bTB breakdowns (27). Finally, badgers are known to forage under cattle dung (28), and earthworms have been demonstrated to spread *M. bovis* BCG from spiked cattle feces to surrounding soil (29).

A number of assays have been developed to diagnose bTB infection in badgers. These include immunoassays such as gamma interferon (IFN-γ) assays (4) and the BrockTB Stat-Pak assay (4), as well as culture from clinical samples (19, 30, 31). Sensitivity (Se) estimates of IFN-γ assays range from 52 to 85% in adult badgers (4, 31–33)—estimates for cubs are lower—compared with a range of 49 to 78% (depending on the severity of disease) for Stat-Pak (34). Specificity (Sp) estimates for IFN-γ range from 88 to 94% (4, 31, 32), compared with 93 to 97% for Stat-Pak (31, 34). Culture is very insensitive (8%), though it is considered to have near perfect specificity (99.8%) (31). These tests are not considered sensitive or specific enough to use alone and also necessitate the live trapping of animals, which requires intensive effort to achieve high coverage and is expensive (35, 36). It has been suggested that diagnosis should be performed at the social-group level, using IFN-γ and Stat-Pak assays in parallel (33). However, in order to maintain sufficient specificity at the group level, this approach requires a threshold of 2 badgers with positive tests to accurately identify a social group as infected. Therefore, this approach requires substantial trapping coverage (50%) and is unlikely to identify social groups with only one positive animal. The prevalence of *M. bovis* shedding in badger feces correlates well with prevalence of infection, as determined by IFN-γ and Stat-Pak assays of contemporaneous trapped badgers at a social-group level (37), though social groups with similar prevalence of infection showed heterogeneity in prevalence of shedding (38).

A culture-independent quantitative PCR (qPCR) test for detecting *M. bovis* DNA in environmental samples, including badger feces, was developed at Warwick University (39, 40). Pathogen detection in feces was thus pioneered by the use of qPCR on DNA extracted from feces and soil. It was possible to culture *M. bovis* from a proportion of qPCR positive badger feces (20), indicating pathogen viability; however, not all qPCR-positive feces provided culture-positive data due to the low sensitivity of culture from fecal samples. Quantification of the level of *M. bovis* genome equivalents in badger feces is likely to be a good proxy for the status of shedding through sputum, and thus transmission of infection through the respiratory route and biting; because lesions in the gut of badgers are extremely rare (41), the presence of *M. bovis* DNA in feces likely

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

occurs via the passage of infected lung discharge through the gastrointestinal tract. Indeed, the detection of *M. bovis* DNA in both the trachea and feces of infected badgers correlates with severity of disease (42). The test also has the advantage of being noninvasive, with the potential to provide greater coverage of the population than trapping-based methods.

In this study, we assessed the diagnostic sensitivity and specificity of Fast24-qPCR, our existing DNA extraction and qPCR method, and Fast96-qPCR, a novel high-throughput DNA extraction method. Fast96-qPCR uses the same qPCR but a high-throughput version of the DNA extraction based on the same chemistry as Fast24-qPCR. We demonstrate that Fast24-qPCR provides a noninvasive method to detect bTB-infected badger social groups through latrine sampling with a high degree of social-group-level specificity and sensitivity. This will provide a valuable tool to enable monitoring of badger social-group bTB status through *M. bovis* shedding in badger feces and, by extension, the effects of bTB control measures.

MATERIALS AND METHODS

Production of the panel. In order to determine the diagnostic sensitivity and specificity of the tests involved in this comparative study, a panel consisting of spiked positive samples ($n = 245$), known negative samples ($n = 205$), and putative positive samples ($n = 119$) from 12 badger social groups containing badgers known to be positive by serological testing was prepared by the Animal and Plant Health Agency (APHA). APHA required minimum acceptable thresholds of social-group-level sensitivity and specificity (50% and 80%, respectively), on behalf of the Department for Environment, Farming and Rural Affairs (DEFRA). These thresholds are based on the assumption of a 10% within-herd prevalence (in badgers) and 10 samples analyzed per social group, and as such, they require a sample-level (diagnostic) sensitivity of 50% and specificity of 98%. The numbers of known positive and negative feces in the panels were thus calculated to be able to establish these thresholds, with a 5% margin of error within 95% confidence intervals in the case of sensitivity and a 2% margin of error within 95% confidence intervals in the case of specificity (43). This required minimums of 246 known positive samples and 188 known negative samples. All study participants were blinded to the status of all samples within this panel until the completion of testing. APHA is compliant with the Animals (Scientific Procedures) Act 1986, and in addition, all experiments involving animals are both reviewed and approved as well as subjected to retrospective analysis by an ethics committee composed of veterinarians, animal care staff, a biostatistician, scientists, and lay members of the community. The production of the panel is as follows.

Feces ($n = 50$) were collected from badgers of known negative status at APHA Weybridge ($n = 25$) and APHA York ($n = 12$) and also from wild latrines at APHA Woodchester ($n = 10$) and latrines in regions of the country where bovine TB is not endemic in cattle ($n = 3$). Fecal samples were collected from a variety of sources in order to account for factors such as fecal consistency and presence of inhibitors, as any test used on wild samples must be robust to these factors. APHA personnel conducted the sampling and prepared spiked samples as follows: positive fecal samples ($n = 245$) were prepared by spiking 150 g pooled from the above sources with 20 ml of buffer containing known quantities (10^5 to 10^1 CFU/g) of *M. bovis* 2122/97. To ensure that spiked samples were homogenous, the feces were mixed with a spatula for a period of 5 min. Aliquots of 1 g were then frozen at −20°C in 2-ml screw-cap microcentrifuge tubes. Five 1-g aliquots were taken from each spiked 150-g fecal sample; thus, there were 5 technical replicates of 49 biological replicates. Full details of the protocol used can be found in the DEFRA report (44). The dilution series chosen was based on previous research within our group at Warwick University (37, 38), and concentrations were determined based on CFU count.

Negative feces ($n = 205$) were aliquoted in a separate laboratory to prevent contamination; feces were mixed with a spatula for 5 min, and aliquoting was performed as described above. Five of these negatives were prepared by spiking feces as described above with dilution buffer only.

Putative positive samples were taken from latrines connected to 12 historically positive social groups at APHA Woodchester. These social groups had at least one culture-positive result and/or four positive IFN-γ or BrockTB StatPak test results on trapped animals within the 2-year period preceding the sampling. Though this does not guarantee that the social groups still contained infected animals at the time of sampling, it was considered a reasonable assumption. The aim was to sample 10 scats per social group; however, only 9 were available from one social group. Therefore, the panel contained 119 putative positive samples. Approximately 30 g of each scat was mixed with a spatula for 5 min. One 1-g aliquot was taken from each unique scat to make up the panel, as described above.

Samples were deidentified by APHA, and two replicates of the panel (one for each extraction method) were sent to Warwick University using the appropriate secure transport procedures (WHO shipping number UN 2814).

In addition to the deidentified samples in the panel, known negative feces ($n = 88$) were added at Warwick University. These feces were also collected from badgers of known negative status at APHA Weybridge. This was to provide our own internal indicator of test performance. The composition of the panel is presented in Table 1.

DNA extraction from badger feces. DNA was extracted from the badger feces using two methods, the existing Fast24-qPCR extraction and the new Fast96-qPCR extraction. For Fast24-qPCR, total com-

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

TABLE 1 Composition of the panel used in this study[a]

Spiked *M. bovis* concn (CFU/g)	No. of samples
570,666.7	5
114,000	25
57,066.67	25
11,400	25
5,706.67	20
1,140	25
570.67	25
285.33	30
142.67	25
71.33	15
35.67	15
17.83	10
0	5
Negative (as part of original panel)	200
Negative (added at Warwick)	88 (Fast24-qPCR), 24 (Fast96-qPCR)
Putative positive	119

[a]The four lowest spiked concentrations (17.83 to 142.67 CFU/g) were used to determine diagnostic sensitivity at low concentration (DSeLC).

munity DNA was extracted from 0.1 g (±0.005 g) of feces using the FastDNA spin kit for soil (SKU 116560200-CF; MP Biomedicals) per the manufacturer's instructions. A modified ribolysis step involving two sequential homogenization steps of 40 s at 6,000 rpm separated by a 30-s pause was performed using a Precellys 24 homogenizer (P000669-PR240-A; Bertin Instruments) as previously reported (37). DNA was extracted from the Fast24-qPCR panel twice in parallel by two operators.

For the Fast96-qPCR extraction, total community DNA was extracted from 0.1 g of feces using the FastDNA 96 soil microbe DNA kit (SKU 119696200; MP Biomedicals) with some alterations from the manufacturer's instructions, detailed as follows. Ribolysis took place in lysing matrix E tubes (SKU 116914050-CF; MP Biomedicals) containing 400 μl lysis buffer and 100 μl sterile molecular-grade distilled H_2O (dH_2O). Samples were ribolysed using a FastPrep-96 instrument (SKU 116010500; MP Biomedicals) at 1,600 rpm for 60 s and centrifuged at 16,110 $\times$ *g* for 10 min. The supernatant was transferred to a 96-well deep-well plate, and DNA extraction then continued as per the manufacturer's instructions. All subsequent centrifugation steps were performed in an Eppendorf 5810R instrument using the A-2-DWP-AT rotor at 3,486 $\times$ *g*. DNA was extracted from the Fast96 panel once.

qPCR testing. The RD4 qPCR assay was used as described previously (37). Briefly, samples were screened using duplicate qPCR assays, and those with a positive result in either replicate were subjected to full quantification in triplicate. A serial dilution of genomic DNA from *M. bovis* BCG Danish 1331 was used as the standard. If one or more replicates showed amplification in the quantification assay, then samples were deemed positive; otherwise, samples were deemed negative. Assays were performed for inhibition of the qPCR using a previously described inhibition assay (40) according to the previously described protocol (37), in order to detect possible false negatives. Briefly, an inhibition control was previously designed with an exotic probe target (green fluorescent protein [GFP]) flanked by DNA complementary to the RD4 primers. A known concentration of this target was added to all samples; if inhibitory compounds were present in the sample, qPCR of the inhibition control target was impacted in comparison with the negative control (equivalent volume of dH_2O). This was quantified by comparison of the threshold C_T of each sample to a negative control, with the difference in C_T values referred to as ΔC_T. Samples were screened in singlet, and if C_T differed by $>$2.5 from that of the negative control, the sample was rescreened in duplicate. If the average C_T of the duplicates differed by $>$2.5 from the negative control, the sample was considered inhibited. Inhibited, nonpositive samples were excluded from analysis. All qPCRs were performed in an ABI 7500 Fast qPCR system (4351106; Thermo Fisher Scientific), using 10 μl of DNA as the template. qPCR protocols were identical for both extraction methods.

Statistical analysis. The study was designed to assess the performance of multiple tests in detecting *M. bovis* in badger feces. Performance was measured in terms of diagnostic sensitivity and specificity at the individual sample level (DSe and DSp). Based on discrepancies between observed genome equivalents within the panel and in naturally infected samples (Fig. 1), we also considered sensitivity within the samples spiked with the four lowest concentrations of *M. bovis* separately (DSeLC) (Table 2). The quantity of *M. bovis* genome equivalents from these four lowest-spiked samples were most similar to the quantities found in positive wild badger feces in previous work (37, 38). For Fast24-qPCR and Fast96-qPCR, sensitivity is dependent on DNA concentration, and estimating sensitivity from samples that contain *M. bovis* cell concentrations higher than those found in natural positive feces will produce a biased overestimate of true sensitivity. Despite the intention for spiked cell concentrations to cover a range similar to that found in naturally infected samples, it was clear that many of the spiked samples had substantially higher cell concentrations; hence the need for this subanalysis. We ascribe this discrepancy to the difference between cell number as measured by genome equivalents and that determined by CFU. Statistical comparisons of DSe with DSeLC were one tailed, as, *a priori*, we anticipated sensitivity to be lower at lower spiked *M. bovis* concentrations.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

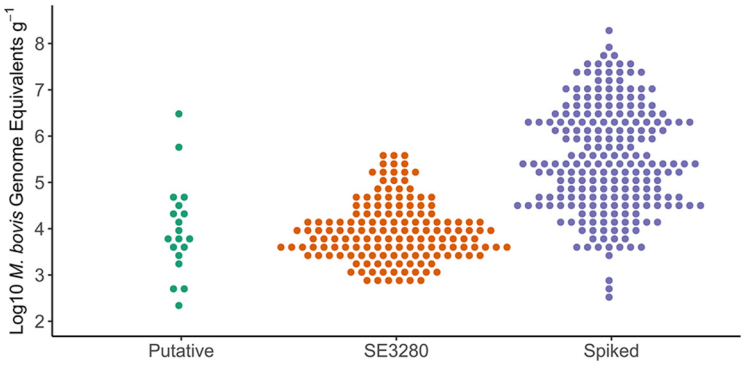


FIG 1 Log_{10} distribution of *M. bovis* genome equivalents obtained by qPCR of positive samples and comparison to a standard curve. Putative and spiked samples are from this study; SE3280 samples are from a previous DEFRA project (SE3280) reported by King et al. (37, 38). Putative and SE3280 samples are taken from natural infected populations. One-way analysis of variance (ANOVA) shows a significant difference between the means of the three populations ($P < 0.001$). Bonferroni-corrected two-tailed Welch's *t* test shows significant differences between SE3280 and spiked samples ($P < 0.001$) and between putative and spiked samples ($P < 0.001$) but not between SE3280 and putative samples ($P = 0.90$).

At the level of the social group, test performance was estimated by calculating herd sensitivity and specificity (HSe and HSp, respectively). These epidemiological terms refer to the ability of the tests to accurately identify positive social groups ("herds") in the case of herd sensitivity and to accurately identify negative social groups in the case of herd specificity. As it is difficult to link feces to the animal which excreted them, positive test results can infer positivity only at the social-group level. The performance of a test at a social-group level is dependent on its sensitivity and specificity at an individual level, the number of samples tested (*n*), the true within-social-group prevalence (TP), and the threshold value of individual positives used to classify the social group as positive. Herd specificity is dependent on diagnostic specificity and the number of samples tested, and herd sensitivity (based on the binomial distribution) can be calculated from apparent within-social-group prevalence (AP), which is calculated as follows (45):

$$\mathrm{AP} = \mathrm{DSe} \times \mathrm{TP} + \left[(1 - \mathrm{DSp}) \times (1 - \mathrm{TP})\right] \quad (1)$$

$$\mathrm{HSe} = 1 - (1 - \mathrm{AP}^n) \quad (2)$$

$$\mathrm{HSp} = 1 - \mathrm{DSp}^n \quad (3)$$

The variance of apparent prevalence [Var(AP)] can be estimated as follows (46):

$$\mathrm{Var(AP)} \approx \mathrm{TP}^2 \times \frac{\mathrm{DSe} \times (1 - \mathrm{DSe})}{N} + (1 - \mathrm{TP})^2 \times \frac{\mathrm{DSp} \times (1 - \mathrm{DSp})}{M} \quad (4)$$

where DSe is estimated from *N* known positives and DSp is estimated from *M* known negatives. From this, 95% confidence intervals (CI) for apparent within-group prevalence (AP) can be calculated as follows:

$$95\%\mathrm{CI} \approx \mathrm{AP} \pm 1.96\sqrt{\mathrm{Var(AP)}} \quad (5)$$

The posttest probability at the social-group level, or the subjective probability of the presence of infection within a social group, can calculated as herd positive predictive value (HPPV) and herd negative predictive value (HNPV) from these estimates of social-group-level sensitivity and specificity, as follows (45):

$$\mathrm{HPPV} = \frac{\mathrm{HSe} \times \mathrm{HP}}{\mathrm{HSe} \times \mathrm{HP} + \left[(1 - \mathrm{HSp}) \times (1 - \mathrm{HP})\right]} \quad (6)$$

TABLE 2 DSe, DSeLC, and DSp

Test	Value (%) (95% CI, *n*)[a]		
	DSe	DSeLC	DSp
Fast24-qPCR (1st operator)	96.7ab* (94.5–99.0, 244)	89.2* (81.7–96.7, 65)	99.0 (97.8–100, 292)
Fast24-qPCR (2nd operator)	89.8a (86.0–93.6, 245)	87.7 (79.7–95.7, 65)	96.9 (95.0–98.9, 293)
Fast96-qPCR	88.4b* (84.3–92.4, 241)	75.0* (64.5–85.5, 64)	97.0 (95.0–98.9, 231)

[a]The letters "a" and "b" indicate pairwise comparisons within DSe: $P < 0.05$ by two-tailed Fisher's exact test (Bonferroni corrected). Asterisks indicate pairwise comparisons between DSe and DSeLC: $P < 0.05$ by one-tailed Fisher's exact test (Bonferroni corrected).

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

$$\mathrm{HNPV} = \frac{\mathrm{HSp} \times (1 - \mathrm{HP})}{[\mathrm{HSp} \times (1 - \mathrm{HP})] + [(1 - \mathrm{HSe}) \times \mathrm{HP}]} \quad (7)$$

where herd prevalence (HP) is the proportion of social groups that contain individuals with disease. For the purposes of our analyses, we modeled *n* up to 20 as, based on our field experience, this represented a reasonable upper limit for unique samples taken over two sampling events (38). In addition, the range of HP (0.05 to 0.2) was chosen based on the range of prevalence within badger social groups (37, 38).

We modeled the effects of requiring two independent DNA extractions and qPCR tests to both give positive results (serial testing [47]) in order to assign a sample as positive. The equations for repeat diagnostic sensitivity and specificity (DSe^R and DSp^R, respectively) are as follows:

$$\mathrm{DSe}^R = \mathrm{DSe}^2 \quad (8)$$

$$\mathrm{DSp}^R = 1 - (1 - \mathrm{DSp})^2 \quad (9)$$

This is possible if samples are split and stored at the point of sampling or when introduced into the laboratory. False-positive results are likely to be the result of contamination with DNA extracted from other positive samples, most likely during the DNA extraction process, as our qPCR is 100% specific for *M. bovis* DNA (38). As such, reextraction from a second aliquot of feces would give an independent result.

We modeled the effects of using repeat extractions using Fast24-qPCR, as it was more sensitive and more specific (though not to a statistically significant degree).

Statistical analysis was performed in RStudio (48) using R (49). Graphics were created using ggplot2 (50).

RESULTS

Sensitivity and specificity at the sample level. DNA was extracted from two replicates of a deidentified panel of badger feces containing known positive and negative samples using the Fast24 (panel 1) and Fast96 (panel 2) extraction methods, prior to qPCR screening and quantification. These panels' identification was restored by APHA when all data had been collected. The performance of the two operators using the Fast24-qPCR method differed significantly in terms of DSe (Bonferroni-corrected $P < 0.01$; two-tailed Fisher's exact test) (Table 2). DSeLC and DSp were not significantly different between the two operators, though both were lower for operator 2 (Op2). Operator 1 (Op1) possessed the most experience with the technique at the time of the study, which may explain the discrepancy.

DSe was significantly higher for the Fast24-qPCR method (Op1) than the Fast96-qPCR method (Bonferroni-corrected $P < 0.01$; two-tailed Fisher's exact test, respectively) (Table 2), but this was not the case for Op2. DSeLC was not significantly different between Fast24-qPCR (Op1 versus Op2) or between Fast24-qPCR (either operator) and Fast96-qPCR. These comparisons of subsamples are comparatively statistically underpowered, however, though DSeLC was similar for both operators of Fast24-qPCR. Fast24-qPCR (Op1) and Fast96-qPCR DSe was significantly higher than DSeLC ($P < 0.01$ and $P < 0.05$, respectively; one-tailed Fisher's exact test). This was not the case for Op2. Diagnostic specificity (DSp) was not significantly different between the two methods or between operators. For both methods, DSe meets the minimum threshold (20%) established prior to the study. In terms of DSp, Fast24-qPCR as performed by Op1 meets the minimum threshold (98%), though its 95% CI did drop below it; however, this is not the case for Op2, though the minimum threshold is within the 95% CI and the difference between operators is not statistically significant. Fast96-qPCR does not meet the threshold, though again the threshold is within the 95% CI.

Sensitivity and specificity at the social-group level. Group-level sensitivity (HSe) increased with number of samples and HP, while group-level specificity (HSp) decreased with number of samples for both Fast24-qPCR and Fast96-qPCR (Fig. 2A and B). To make this trade-off more favorable, two amendments were modeled, based on DSp and DSe from Op1. The first was serial testing, i.e., requiring independent, confirmatory re-extraction and qPCR retest of each positive feces to assign positive status to a sample. This substantially increased HSp while moderately decreasing HSe for both Fast24-qPCR and Fast96-qPCR (Fig. 2C and D). The second amendment, increasing the threshold of positive samples required to assign positive status to a social group from one to two, also increased HSp, but this had the effect of sharply decreasing HSe (Fig. S1).

The Fast24-qPCR data sets were also analyzed to model the effects of serial testing using both operators' data sets as independent repeats (Table 3) in order to compare

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

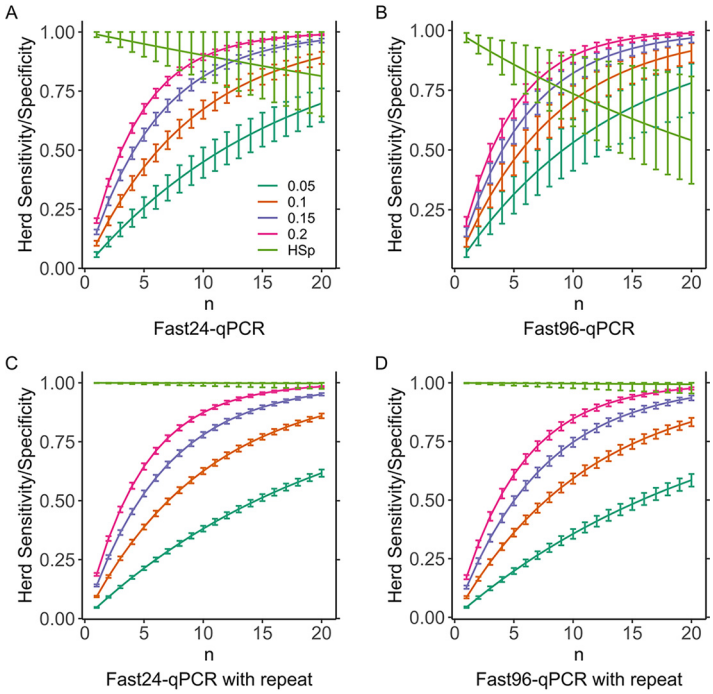


FIG 2 Relationship between herd sensitivity (HSe) and herd specificity (HSp) and the number of samples tested (*n*). A range of herd prevalences (HP) are modeled. HSe at a HP of 0.05 is shown in dark green, 0.1 in orange, 0.15 in purple, and 0.2 in pink, and HSp is shown in light green. (A) Fast24-qPCR; (B) Fast96-pPCR; (C) Fast24-qPCR with repeated positives; (D) Fast96-qPCR with repeated positives. For both Fast24-qPCR (A) and Fast96-qPCR (B), HSe increased with *n*; however, this came at the expense of HSp, which decreases with *n*. This can be alleviated by repeat testing of positives, which decreases the decline in HSp with *n* while maintaining HSe in both Fast24-qPCR (C) and Fast96-qPCR (D).

these to estimations based on DSe and DSp from Op1. When compared to the original DSe and DSp data (Table 2), sensitivity decreased (93.6% compared to 96.7%), while specificity increased (99.99% compared to 99.0%). DSe^R for the combined Fast24-qPCR data set was lower than estimated via equation 8 (87.3% compared to 93.6%; $P < 0.05$; two-tailed Fisher's exact test), which was explained by the lower DSe for Op2. $DSeLC^R$ and DSp^R for this combined data set are similar to the values estimated by equations 8 and 9 (78.5% compared to 79.6% and 100% compared to 99.99%, respectively). Repeat testing resulted in a substantially reduced decline in HSp caused by increasing sample number, thus allowing HSe to be increased without compromising HSp despite the reduction in DSe caused by repeat testing.

Predictive values at the social-group level. At low levels of HP, both Fast24-qPCR and Fast96-qPCR have low HPPV but high HNPV (Fig. 2). HPPV is increased by testing 20 samples, with a threshold of 2 positives required to determine herd level infection, and is increased still further by requiring positive repeats in serial testing of positive samples (Fig. 3), which has a minimal effect on HNPV. These figures assume a within-herd shedding prevalence of 10%. The figures are based on estimates of HSe determined from DSeLC, though the values are similar if DSe is used. HPPV increases with

TABLE 3 DSe, DSeLC, and DSp with estimated effects of repeat testing and as measured by combining Fast24-qPCR panel results from both operators

	Value (95% CI, *n*)[a]		
Test	**DSe (%)**	**DSeLC (%)**	**DSp (%)**
Fast24-qPCR with repeat	93.6a (89.3–97.9)	79.6 (66.7–93.6)	99.99 (99.95–100)
Fast96-qPCR with repeat	85.5 (79.7–91.5)	66.9 (52.6–82.8)	99.97 (99.88–100)
Fast24-qPCR both operators	87.3a (83.1–91.5, 244)	78.5 (68.5–88.5, 65)	100 (100, 201)

[a]The letter "a" indicates pairwise comparison: $P < 0.05$ by two-tailed Fisher's exact test.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

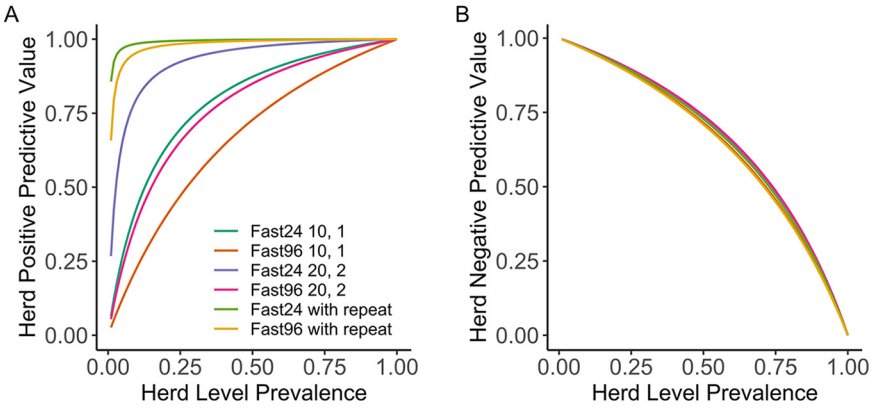


FIG 3 Relationship between herd positive predictive value (HPPV) (A), herd negative predictive value (HNPV) (B), and herd prevalence (HP)—the proportion of herds that are positive—for a variety of testing modalities. "Fast24-qPCR 10, 1" indicates Fast24-qPCR with 10 samples, with 1 positive sample required to designate the herd positive. Considering Fast24-qPCR 10, 1 and Fast96 10, 1 as the baseline, HPPV is improved by doubling both the number of samples tested, and the number of positive samples required, but not by as much as requiring the repeat testing of positive samples (10, 1 with repeat) (A). HNPV shows a similar relationship for all testing modalities (B).

sample number and HP, both with and without retesting of positives (Fig. S2). HNPV increases with sample number but decreases with HP, and the effect of retesting of positives on this is minimal (Fig. S3).

DISCUSSION

The sensitivity and specificity of two tests that detect *M. bovis* in badger feces were estimated. The results presented here show that, on a per-sample basis, both Fast24-qPCR and Fast96-qPCR have diagnostic specificity that is similar or superior to that of existing trapping-based immunological tests. All tests met the threshold criteria for diagnostic sensitivity proposed in advance. Neither test consistently met the threshold criteria for diagnostic specificity, with only one operator of the Fast24-qPCR method meeting this threshold. However, we estimated that serial testing of positives would substantially increase specificity, and by combining the data sets from both operators, we showed 100% specificity. Such repeats show that Fast24-qPCR can be applied to multiple fecal samples from a social group in order to maximize group-level sensitivity without compromising group-level specificity. We predict that similar repeat tests could also improve the specificity of Fast96, though this remains to be demonstrated.

The Fast24-qPCR method, as performed by the first operator, was significantly more sensitive than the Fast96-qPCR, though this was not the case for the second operator, possibly due to the difference in experience between the two operators. Regardless, both extraction methods displayed high levels of sensitivity for the spiked samples analyzed in this study. However, for sensitivity, comparisons with other diagnostic tests should be applied with caution, as the two methods detect shedding, in contrast to immunological tests, which detect immune status. It is likely that there are more animals that are exposed to and show immunological responses to *M. bovis* than there are animals that actively shed the bacteria in their feces, and thus, a lower herd prevalence is expected for fecal testing than immunological testing, as previously shown by our laboratory (37). However, animals that are shedding may be both more infectious and more likely to spread infection via environmental contamination than seropositive animals that are not shedding. For this reason, and due to the difficulty of linking badger feces to individual animals, the tests cannot be used to determine the infection status of individuals, unless feces are taken directly from trapped animals.

For feces collected from latrines, the tests can be applied at a social-group level. To achieve adequate social-group-level sensitivity (HSe) requires the testing of multiple feces;

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

however, this comes at the cost of decreasing group level specificity (HSp). This is overcome by the retesting of any positive samples using a previously stored aliquot of the same feces. This allows up to 20 fecal samples, approximately the upper limit for the quantity of unique samples that can be collected on two sampling trips, to be tested with low false-positive rates at the social-group level. Buzdugan et al. (33) have modeled HSe and HSp based on the parallel use of Stat-Pak and IFN-γ assays on samples from trapped badgers—i.e., both tests are used, and the animal is assigned positive status if either test is positive (47). Serial testing resulted in too-low DSe (30% at the individual-animal level) (33); therefore, to achieve the highest HSe and HSp, Buzdugan et al. (33) modeled the effects of parallel testing of animals, with a threshold requiring that two animals test positive for a social group to be considered bTB positive. Assuming that 50% of animals within a social group of 15 are trapped and tested, this results in an HSp of 91%, with an HSe of ~60% at 25% prevalence of infection (33), though an HSp of >95% is also reported if 40% of the social group is trapped and tested. Serial testing of positive samples with Fast24-qPCR therefore shows a higher HSp than a trapping-based strategy. Given that initial data suggest that fecal qPCR can identify different animals than the IFN-γ and BrockTB StatPak assays (37), Fast24-qPCR could therefore be used to complement the immunological testing model described by Buzdugan et al. (33).

Fast24-qPCR is a noninvasive sampling method that can detect the shedding of *M. bovis* in badger feces at the level of the social group. When performed with retesting of positives, it has very high specificity and high sensitivity at the social-group level. In comparison to Fast24-qPCR, Fast96-qPCR increases the throughput of samples, but at the expense of reduced sensitivity and specificity. The reduction in specificity can likely be alleviated substantially with an independent retest of positive fecal samples using the Fast24-qPCR DNA extraction method, allowing testing of higher numbers of feces per social group, leading to higher sensitivity at the social-group level while maintaining high herd positive predictive value. While retesting of positives does increase the expense of the test, it need be applied only to the proportion of samples that are positive, which, based on previous research, we estimate to be in the range of 5 to 15% within regions where *M. bovis* is endemic. Given sufficient sampling effort, Fast24-qPCR therefore provides a social-group-level test that is capable of measuring the impacts of interventions designed to reduce the spread of bTB from badgers to cattle and vice versa, by accurately measuring the shedding of *M. bovis* into the environment.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

We acknowledge the staff at APHA for their experimental design and preparation of the panels used in this study.

This study was funded by DEFRA under the grant "Detection of *Mycobacterium bovis* in badger feces—an interlaboratory comparison study," DEFRA RRD project SE3289 2014-2015. E.M.H.W. and E.R.T. also acknowledge funding from BBSRC BB/N004655/1.

We declare no conflicts of interest.

REFERENCES

1. Department for Environment Food & Rural Affairs. 2014. The strategy for achieving officially bovine tuberculosis free status for England. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300447/pb14088-bovine-tb-strategy-140328.pdf.
2. Department for Environment Food & Rural Affairs. 2020. Quarterly publication of national statistics on the incidence and prevalence of tuberculosis (TB) in cattle in Great Britain—to end June 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871583/bovinetb-statsnotice-Q4-quarterly-11mar20.pdf.
3. Cheeseman CL, Wilesmith JW, Stuart FA. 1989. Tuberculosis: the disease and its epidemiology in the badger, a review. Epidemiol Infect 103:113–125. https://doi.org/10.1017/S0950268800030417.
4. Dalley D, Davé D, Lesellier S, Palmer S, Crawshaw T, Hewinson RG, Chambers M. 2008. Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (*Meles meles*). Tuberculosis (Edinb) 88:235–243. https://doi.org/10.1016/j.tube.2007.11.001.
5. Menzies FD, Neill SD. 2000. Cattle-to-cattle transmission of bovine tuberculosis. Vet J 160:92–106. https://doi.org/10.1016/S1090-0233(00)90482-9.
6. Donnelly CA, Woodroffe R, Cox DR, Bourne J, Gettinby G, Le Fevre AM, McInerney JP, Morrison WI. 2003. Impact of localized badger culling on

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

tuberculosis incidence in British cattle. Nature 426:834–837. https://doi.org/10.1038/nature02192.
7. Woodroffe R, Donnelly CA, Jenkins HE, Johnston WT, Cox DR, Bourne FJ, Cheeseman CL, Delahay RJ, Clifton-Hadley RS, Gettinby G, Gilks P, Hewinson RG, McInerney JP, Morrison WI. 2006. Culling and cattle controls influence tuberculosis risk for badgers. Proc Natl Acad Sci U S A 103:14713–14717. https://doi.org/10.1073/pnas.0606251103.
8. Donnelly CA, Nouvellet P. 2013. The contribution of badgers to confirmed tuberculosis in cattle in high-incidence areas in England. PLoS Curr 5. https://doi.org/10.1371/currents.outbreaks.097a904d3f3619db2fe78d24bc776098.
9. Drewe JA, O'Connor HM, Weber N, McDonald RA, Delahay RJ. 2013. Patterns of direct and indirect contact between cattle and badgers naturally infected with tuberculosis. Epidemiol Infect 141:1467–1475. https://doi.org/10.1017/S0950268813000691.
10. Bohm M, Hutchings MR, White PC. 2009. Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. PLoS One 4:e5016. https://doi.org/10.1371/journal.pone.0005016.
11. Conlan AJ, McKinley TJ, Karolemeas K, Pollock EB, Goodchild AV, Mitchell AP, Birch CP, Clifton-Hadley RS, Wood JL. 2012. Estimating the hidden burden of bovine tuberculosis in Great Britain. PLoS Comput Biol 8:e1002730. https://doi.org/10.1371/journal.pcbi.1002730.
12. Delahay RJ, Walker N, Smith GC, Wilkinson D, Clifton-Hadley RS, Cheeseman CL, Tomlinson AJ, Chambers MA. 2013. Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high-density badger (*Meles meles*) population. Epidemiol Infect 141:1445–1456. https://doi.org/10.1017/S0950268813000721.
13. O'Hare A, Orton RJ, Bessell PR, Kao RR. 2014. Estimating epidemiological parameters for bovine tuberculosis in British cattle using a Bayesian partial-likelihood approach. Proc Biol Sci 281:20140248. https://doi.org/10.1098/rspb.2014.0248.
14. Woodroffe R, Donnelly CA, Ham C, Jackson SYB, Moyes K, Chapman K, Stratton NG, Cartwright SJ. 2016. Badgers prefer cattle pasture but avoid cattle: implications for bovine tuberculosis control. Ecol Lett 19:1201–1208. https://doi.org/10.1111/ele.12654.
15. Fine AE, Bolin CA, Gardiner JC, Kaneene JB. 2011. A study of the persistence of *Mycobacterium bovis* in the environment under natural weather conditions in Michigan, USA. Vet Med Int 2011:765430. https://doi.org/10.4061/2011/765430.
16. Ghodbane R, Mba Medie F, Lepidi H, Nappez C, Drancourt M. 2014. Long-term survival of tuberculosis complex mycobacteria in soil. Microbiology (Reading) 160:496–501. https://doi.org/10.1099/mic.0.073379-0.
17. Sweeney FP, Courtenay O, Hibberd V, Hewinson RG, Reilly LA, Gaze WH, Wellington EM. 2007. Environmental monitoring of *Mycobacterium bovis* in badger feces and badger sett soil by real-time PCR, as confirmed by immunofluorescence, immunocapture, and cultivation. Appl Environ Microbiol 73:7471–7473. https://doi.org/10.1128/AEM.00978-07.
18. Brooks-Pollock E, Roberts GO, Keeling MJ. 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature 511:228–231. https://doi.org/10.1038/nature13529.
19. Clifton-Hadley RS, Wilesmith JW, Stuart FA. 1993. *Mycobacterium bovis* in the European badger (*Meles meles*): epidemiological findings in tuberculous badgers from a naturally infected population. Epidemiol Infect 111:9–19. https://doi.org/10.1017/S0950268800056624.
20. Wilesmith JW, Sayers PE, Bode R, Pritchard DG, Stuart FA, Brewer JI, Hillman GD. 1986. Tuberculosis in East Sussex. II. Aspects of badger ecology and surveillance for tuberculosis in badger populations (1976–1984). J Hyg (Lond) 97:11–26. https://doi.org/10.1017/s0022172400064317.
21. Hutchings MR, Harris S. 1999. Quantifying the risks of TB infection to cattle posed by badger excreta. Epidemiol Infect 122:167–173. https://doi.org/10.1017/S0950268898001897.
22. Hutchings MR, Harris S. 1997. Effects of farm management practices on cattle grazing behaviour and the potential for transmission of bovine tuberculosis from badgers to cattle. Vet J 153:149–162. https://doi.org/10.1016/S1090-0233(97)80035-4.
23. Williams RS, Hoy WA. 1930. The viability of B. tuberculosis (Bovinus) on pasture land, in stored faeces and in liquid manure. Epidemiol Infect 30:413–419. https://doi.org/10.1017/S0022172400010561.
24. Maddock ECG. 1936. Experiments on the infectivity for healthy calves of bovine tubercle bacilli discharged in dung upon pasture. Part I. From tubercular calves fed with emulsions of tubercle bacilli 1934–5. Part II. From tubercular cows passing tubercle bacilli in their dung 1935–6. Epidemiology & Infection 36:594–601. https://doi.org/10.1017/S0022172400043953.
25. Liebana E, Johnson L, Gough J, Durr P, Jahans K, Clifton-Hadley R, Spencer Y, Hewinson RG, Downs SH. 2008. Pathology of naturally occurring bovine tuberculosis in England and Wales. Vet J 176:354–360. https://doi.org/10.1016/j.tvjl.2007.07.001.
26. Chaintarli K, Upton P. 2018. Analysis of bovine tuberculosis surveillance at routine slaughter of cattle in Great Britain 2013–2016. Animal and Plant Health Agency, Addleston, United Kingdom.
27. Griffin JM, Hahesy T, Lynch K, Salman MD, McCarthy J, Hurley T. 1993. The association of cattle husbandry practices, environmental factors and farmer characteristics with the occurence [*sic*] of chronic bovine tuberculosis in dairy herds in the Republic of Ireland. Prev Vet Med 17:145–160. https://doi.org/10.1016/0167-5877(93)90025-O.
28. Kruuk H, Parish T, Brown CAJ, Carrera J. 1979. Use of pasture by the European badger (*Meles meles*). J Appl Ecol 16:453–459. https://doi.org/10.2307/2402521.
29. Barbier E, Chantemesse B, Rochelet M, Fayolle L, Bollache L, Boschiroli ML, Hartmann A. 2016. Rapid dissemination of *Mycobacterium bovis* from cattle dung to soil by the earthworm *Lumbricus terrestris*. Vet Microbiol 186:1–7. https://doi.org/10.1016/j.vetmic.2016.01.025.
30. Delahay RJ, Langton S, Smith GC, Clifton-Hadley RS, Cheeseman CL. 2000. The spatio-temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density badger population. J Anim Ecol 69:428–441. https://doi.org/10.1046/j.1365-2656.2000.00406.x.
31. Drewe JA, Tomlinson AJ, Walker NJ, Delahay RJ. 2010. Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. PLoS One 5:e11196. https://doi.org/10.1371/journal.pone.0011196.
32. Chambers MA, Waterhouse S, Lyashchenko K, Delahay R, Sayers R, Hewinson RG. 2009. Performance of TB immunodiagnostic tests in Eurasian badgers (*Meles meles*) of different ages and the influence of duration of infection on serological sensitivity. BMC Vet Res 5:42–47. https://doi.org/10.1186/1746-6148-5-42.
33. Buzdugan SN, Chambers MA, Delahay RJ, Drewe JA. 2016. Diagnosis of tuberculosis in groups of badgers: an exploration of the impact of trapping efficiency, infection prevalence and the use of multiple tests. Epidemiol Infect 144:1717–1727. https://doi.org/10.1017/S0950268815003210.
34. Chambers MA, Crawshaw T, Waterhouse S, Delahay R, Hewinson RG, Lyashchenko KP. 2008. Validation of the BrockTB Stat-Pak assay for detection of tuberculosis in Eurasian badgers (*Meles meles*) and influence of disease severity on diagnostic accuracy. J Clin Microbiol 46:1498–1500. https://doi.org/10.1128/JCM.02117-07.
35. Woodroffe R, Gilks P, Johnston WT, Le Fevre AM, Cox DR, Donnelly CA, Bourne FJ, Cheeseman CL, Gettinby G, McInerney JP, Morrison WI. 2007. Effects of culling on badger abundance: implications for tuberculosis control. J Zool 274:28–37. https://doi.org/10.1111/j.1469-7998.2007.00353.x.
36. Smith GC, Cheeseman CL. 2007. Efficacy of trapping during the initial proactive culls in the randomised badger culling trial. Vet Rec 160:723–726. https://doi.org/10.1136/vr.160.21.723.
37. King HC, Murphy A, James P, Travis E, Porter D, Sawyer J, Cork J, Delahay RJ, Gaze W, Courtenay O, Wellington EM. 2015. Performance of a non-invasive test for detecting *Mycobacterium bovis* shedding in European badger (*Meles meles*) populations. J Clin Microbiol 53:2316–2323. https://doi.org/10.1128/JCM.00762-15.
38. King HC, Murphy A, James P, Travis E, Porter D, Hung Y, Sawyer J, Cork J, Delahay RJ, Gaze W, Courtenay O, Wellington EM. 2015. The variability and seasonality of the environmental reservoir of *Mycobacterium bovis* shed by wild European badgers. Sci Rep 5:12318. https://doi.org/10.1038/srep12318.
39. Travis ER, Gaze WH, Pontiroli A, Sweeney FP, Porter D, Mason S, Keeling MJC, Jones RM, Sawyer J, Aranaz A, Rizaldos EC, Cork J, Delahay RJ, Wilson GJ, Hewinson RG, Courtenay O, Wellington EMH. 2011. An inter-laboratory validation of a real time PCR assay to measure host excretion of bacterial pathogens, particularly of *Mycobacterium bovis*. PLoS One 6:e27369. https://doi.org/10.1371/journal.pone.0027369.
40. Pontiroli A, Travis ER, Sweeney FP, Porter D, Gaze WH, Mason S, Hibberd V, Holden J, Courtenay O, Wellington EM. 2011. Pathogen quantitation in complex matrices: a multi-operator comparison of DNA extraction methods with a novel assessment of PCR inhibition. PLoS One 6:e17916. https://doi.org/10.1371/journal.pone.0017916.
41. Gallagher J, Clifton-Hadley RS. 2000. Tuberculosis in badgers; a review of the disease and its significance for other animals. Res Vet Sci 69:203–217. https://doi.org/10.1053/rvsc.2000.0422.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

42. Wellington EMH, Courtenay O. 2014. Badgers and bovine TB: how can environmental microbiology help? Microbiol Today 41:143–144.
43. Jacobson RH. 1998. Validation of serological assays for diagnosis of infectious diseases. Rev Sci Tech 17:469–526. https://doi.org/10.20506/rst.17.2.1119.
44. DEFRA. 2015. A study to comparatively assess diagnostic methods for detection of *M. bovis* in badger faeces. Final report. https://tinyurl.com/yy4lycr6
45. Martin SW, Shoukri M, Thorburn MA. 1992. Evaluating the health status of herds based on tests applied to individuals. Prev Vet Med 14:33–43. https://doi.org/10.1016/0167-5877(92)90082-Q.
46. Christensen J, Gardner IA. 2000. Herd-level interpretation of test results for epidemiologic studies of animal diseases. Prev Vet Med 45:83–106. https://doi.org/10.1016/s0167-5877(00)00118-5.
47. Fletcher RW, Fletcher SW. 2005. Clinical epidemiology: the essentials, 4th ed, Lippicott Williams & Wilkins, Baltimore, MD.
48. RStudio Team. 2020. RStudio: integrated development for R. RStudio, PBC, Boston, MA.
49. R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
50. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.

Downloaded from https://journals.asm.org/journal/jcm on 08 June 2022 by 137.205.238.49.

AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

MECHANISMS OF RESISTANCE

Mechanisms Involved in the Active Secretion of CTX-M-15 β-Lactamase by Pathogenic *Escherichia coli* ST131

Severine Rangama,[a] Ian D. E. A. Lidbury,[a,b] Jennifer M. Holden,[a,c] Chiara Borsetto,[a] Andrew R. J. Murphy,[a] Peter M. Hawkey,[d] Elizabeth M. H. Wellington[a]

[a]School of Life Sciences, University of Warwick, Coventry, United Kingdom
[b]Department of Animal and Plant Science, The University of Sheffield, Sheffield, United Kingdom
[c]Micropathology Ltd., University of Warwick Science Park, Coventry, United Kingdom
[d]Institute of Microbiology and Infection, University of Birmingham, Edgbaston, United Kingdom

ABSTRACT Infections caused by antimicrobial-resistant bacterial pathogens are fast becoming an important global health issue. Strains of *Escherichia coli* are common causal agents of urinary tract infection and can carry multiple resistance genes. This includes the gene $bla_{CTX\text{-}M\text{-}15}$, which encodes an extended-spectrum beta-lactamase (ESBL). While studying antimicrobial resistance (AMR) in the environment, we isolated several strains of *E. coli* ST131 downstream of a wastewater treatment plan (WWTP) in a local river. These isolates were surviving in the river sediment, and characterization proved that a multiresistant phenotype was evident. Here, we show that *E. coli* strain 48 (river isolate ST131) provided a protective effect against a third-generation cephalosporin (cefotaxime) for susceptible *E. coli* strain 33 (river isolate ST3576) through secretion of a functional ESBL into the growth medium. Furthermore, extracellular ESBL activity was stable for at least 24 h after secretion. Proteomic and molecular genetic analyses identified CTX-M-15 as the major secreted ESBL responsible for the observed protective effect. In contrast to previous studies, outer membrane vesicles (OMVs) were not the route for CTX-M-15 secretion. Indeed, mutation of the type I secretion system led to a significant reduction in the growth of the ESBL-producing strain as well as a significantly reduced ability to confer protective effect. We speculate that CTX-M-15 secretion, mediated through active secretion using molecular machinery, provides a public goods service by facilitating the survival of otherwise susceptible bacteria in the presence of cefotaxime.

KEYWORDS *Escherichia coli*, antibiotic resistance genes, beta-lactamases, enzyme secretion, secretion systems

Pathogenic strains of *Escherichia coli* producing CTX-M β-lactamases have recently emerged worldwide and now present the most common type of extended-spectrum β-lactamase (ESBL) enzymes in *Enterobacteriaceae* (1–5). Limited treatment is available for patients infected with these *E. coli* strains, which presents severe challenges to health care (1, 4, 6–9). The global emergence of CTX-M-producing *E. coli* is driven by the rapid dissemination of the gene $bla_{CTX\text{-}M}$ located on highly mobilizable elements, such as plasmids and transposons (2, 10). Over 172 variants of CTX-M have been identified, which cluster into five groups, CTX-M-1, -2, -8, -9, and -25 groups (11). $bla_{CTX\text{-}M\text{-}15}$ belongs to group 1 and is the predominant variant in the human population globally, including the United Kingdom (2, 7).

Clinical studies have suggested that secretion of hydrolytic enzymes, such as β-lactamases, irreversibly inactivate antibiotics outside the cell and protect both the producer and otherwise susceptible bacteria in close proximity (12). One proposed mechanism for the secretion of ESBLs is the formation and likely stochastic release of outer

Citation Rangama S, Lidbury IDEA, Holden JM, Borsetto C, Murphy ARJ, Hawkey PM, Wellington EMH. 2021. Mechanisms involved in the active secretion of CTX-M-15 β-lactamase by pathogenic *Escherichia coli* ST131. Antimicrob Agents Chemother 65:e00663-21. https://doi.org/10.1128/AAC.00663-21.

Copyright © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Severine Rangama, m.rangama@warwick.ac.uk, or Elizabeth M. H. Wellington, e.m.h.wellington@warwick.ac.uk.

Received 31 March 2021
Returned for modification 10 May 2021
Accepted 19 July 2021
Accepted manuscript posted online 26 July 2021
Published 17 September 2021

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

membrane vesicles (OMVs), which are common in Gram-negative bacteria (13). This secretory process eliminates the need for bacterial contact or complex molecular architectures at the cell wall-periplasm interface typically required for long-distance dissemination of extracellular proteins (14). The packaging of β-lactamases into OMVs has been demonstrated in *Pseudomonas aeruginosa* by microscopy and enzymatic studies (15). In addition, the release of OMVs containing various antibiotic-related proteins from a drug-resistant *E. coli* isolate facilitated the survival of various susceptible bacteria in the presence of β-lactam antibiotics (16). However, in this study, the relative contribution of ESBLs compared to other antibiotic-related proteins, such as bacterial transporter systems, was not conclusively determined. Thus, the mechanism of CTX-M variants, such as CTX-M-15, in providing a protective effect remains uncertain.

In Gram-negative bacteria, secretion of extracellular enzymes is achieved through either a one- or two-step process. In the two-step process, initial translocation across the cytoplasmic membrane to the periplasm is achieved through two main pathways: the twin-arginine (Tat) or the general secretory (Sec) pathway. While the Tat pathway translocates a small number of folded proteins across the cytoplasmic membrane, the majority of proteins are translocated in an unfolded state using the Sec pathway (17). The second translocation event across the outer membrane is coordinated by various specialized export systems classified as type II and type V secretion systems (T2SS and T5SS, respectively). The one-step process, performed by type 1 (T1SS), type III (T3SS), type IV (T4SS), and type VI (T6SS) secretion systems, translocates proteins directly from the cytoplasm to the extracellular milieu, bypassing the periplasmic space (18, 19). The architecture of the T1SS is closely related to the secretion system of the multidrug efflux pumps, called resistance nodulation division (RND), which secretes most antibacterial molecules out of the cells, contributing to antibiotic resistance (20). In contrast, the T2SS, including components called the general secretion pathway (Gsp), ensures the transport of hydrolyzing enzymes and toxins (21–23). Secretion of extracellular enzymes is often thought of as a public goods service, as they can provide an auxiliary function to bacteria otherwise lacking a given phenotype, for example, the degradation of recalcitrant organic phosphorus or carbohydrate polymers (24–27).

In this study, our aim was to further investigate β-lactamase resistance in *E. coli* ST131 isolated from a UK river system downstream of a waste water treatment plant (WWTP) to establish the mechanism of enzyme secretion (28–33). We report that this strain provided a protective effect to a susceptible *E. coli* river isolate against cefotaxime. Further investigations demonstrated that CTX-M-15 was secreted and a role for T1SS was established.

RESULTS

Phenotypic and genotypic testing of the two isolates. The resistance profile of two environmental *E. coli* strains, strains 33 and 48, isolated from a UK river system against five antibiotics were determined. Strain 33 showed no phenotypic resistance to any of the antibiotics tested; in contrast, strain 48 was resistant to three of the five tested antibiotics (see Table S4 in the supplemental material). Whole-genome sequencing of strains 33 and 48 revealed the presence of three β-lactamase genes of clinical relevance in strain 48 only, $bla_{\text{TEM-1}}$, $bla_{\text{OXA-1}}$, and $bla_{\text{CTX-M-15}}$ with CTX-M, the main type of ESBLs. PCR and sequencing confirmed the presence of the three β-lactamase genes in strain 48.

Strain 48 constitutively expresses a secreted β-lactamase. The two *E. coli* river isolates, 33 and 48, were challenged with cefotaxime (8 μg/ml, 16 μg/ml, 32 μg/ml, and 64 μg/ml); growth of strain 33 was inhibited by all cefotaxime concentrations, while strain 48 was completely resistant (Fig. 1A). Previous genomic analyses revealed strain 48 is predicted to possess three ESBLs, encoded by $bla_{\text{TEM-1}}$, $bla_{\text{OXA-1}}$, and $bla_{\text{CTX-M-15}}$. Biochemical analyses revealed that strain 48 possessed extracellular β-lactamase activity in either the presence or absence of cefotaxime (Fig. 1B), although the greatest activity was observed during growth on the highest concentration of this antibiotic. As expected, *E. coli* strain 33, which was sensitive to the β-lactams, showed no secreted β-lactamase activity (Fig. 1B).

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

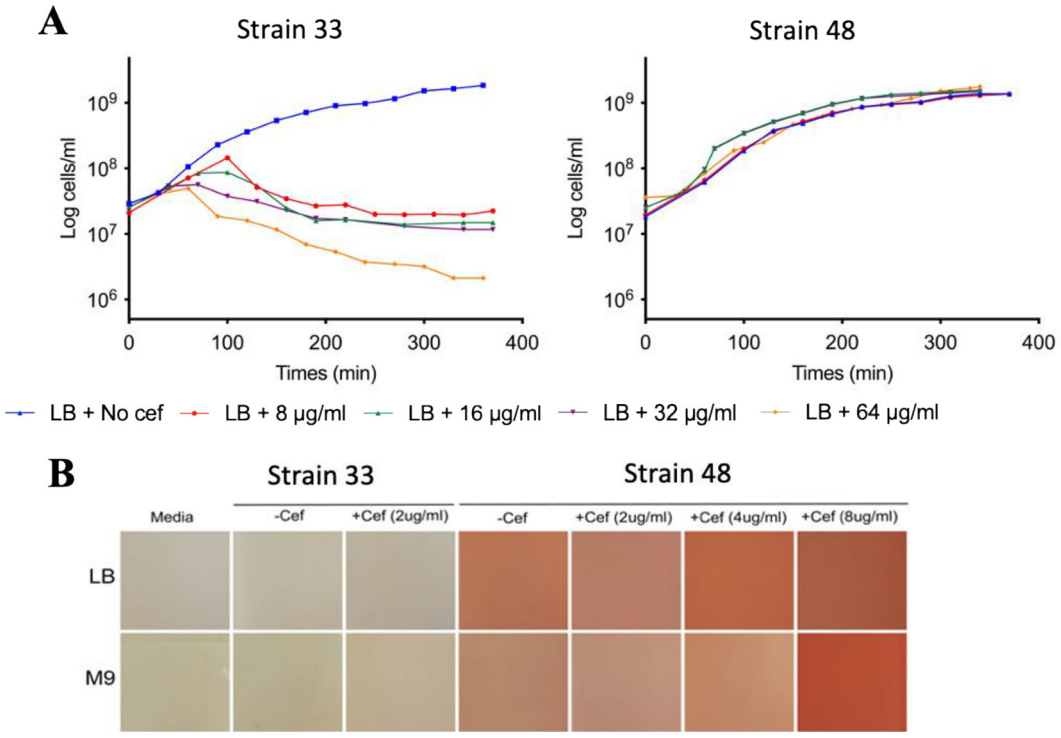


FIG 1 Secretion of β-lactamases by strain 48. (A) Growth of *E. coli* strains 33 and 48 cultivated in LB medium in various cefotaxime (cef) concentrations, 8 μg/ml (in red), 16 μg/ml (in green), 32 μg/ml (in purple), and 64 μg/ml (in yellow). (B) Nitrocefin assay with various concentrations of cefotaxime (2 μg/ml, 4 μg/ml, and 8 μg/ml) in the presence of *E. coli* strain 33 or strain 48.

CTX-M-15 is responsible for conferring cefotaxime resistance. To identify which of the three ESBLs was responsible for β-lactamase activity in strain 48, the proteome of this bacterium, partitioned into cellular (CP) and extracellular (XP) fractions, was analyzed. Cells were grown in either the presence (8 μg/ml) or absence of cefotaxime. Both TEM and CTX-M-15 β-lactamases were identified in the CP and XP; however, the relative abundance of these in the CP was ~10-fold lower than their relative abundance in the XP (Fig. 2A). Exoproteomics identified 1,845 proteins across all treatments, the majority of which represented a long tail of very-low-abundance proteins (<0.1%). Notably, CTX-M-15 was the third most abundant protein in the XP of strain 48, in either the presence (1.7%) or absence (2.48%) of cefotaxime (Fig. 2A). The abundance of TEM in the XP was lower (absence, 0.38%; presence 0.6%). Constitutive expression of CTX-M-15 was confirmed by reverse transcription-quantitative PCR (RT-qPCR), which showed no significant difference in $bla_{\text{CTX-M-15}}$ transcription in either the presence or absence of antibiotic (see Fig. S1 in the supplemental material).

To confirm if CTX-M-15 was responsible for conferring cefotaxime resistance, $bla_{\text{CTX-M-15}}$, bla_{TEM}, and bla_{OXA} from strain 48 were separately cloned into the expression vector pGEM-T. Plasmids were mobilized into a susceptible host, the commercial strain *E. coli* JM109, resulting in the strains JM109-OXA, JM109-TEM, and JM109-CTX-M-15. An empty vector control was also mobilized into JM109, creating the strain JM109-pGEM-T (Fig. 2B). Only JM109-CTX-M-15 grew in the presence of cefotaxime, confirming $bla_{\text{CTX-M-15}}$ was essential for resistance to cefotaxime (Fig. 2C).

CTX-M-15 secretion provides protection to susceptible cells. Strain 33 was susceptible to cefotaxime, so to determine if secreted CTX-M-15 from strain 48 could complement a susceptible strain, strain 33 was grown in both the presence and absence of cefotaxime in a conditioned medium (CM) used for growing strain 48. Strain 33 grew in CM in the presence of cefotaxime, demonstrating that strain 48 secreted sufficient

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

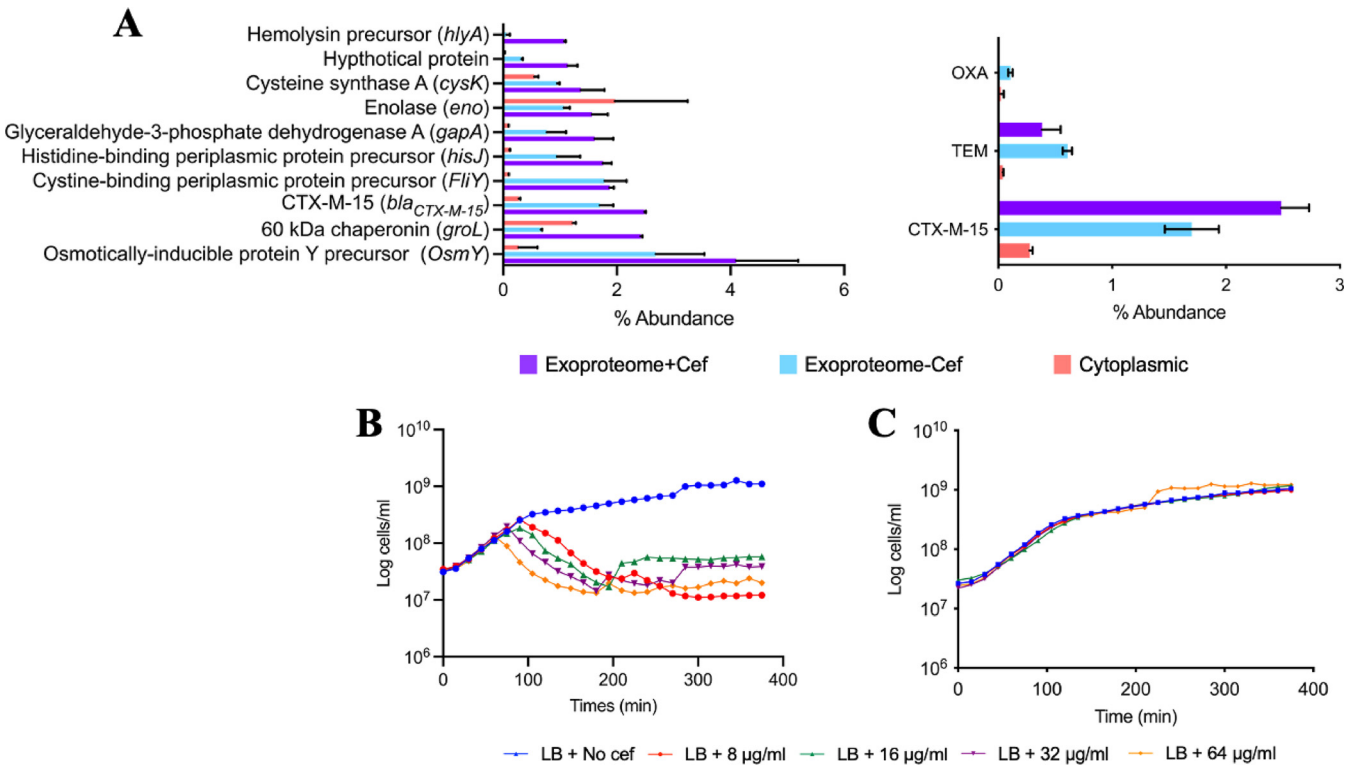


FIG 2 Identification of CTX-M-15 as the major secreted ESBL. (A) Top 10 most abundant proteins found in the exoproteome of strain 48 in the presence of cefotaxime. Protein abundance was evaluated by tandem mass spectrometry spectral counts, which correlated linearly with the protein abundance. Error bars indicate means from three replicates. Abundance of the three ESBLs proteins is shown. (B and C) Growth of JM109 empty plasmid (in blue) (B) and JM109 CTX-M-15 (in blue) (C) cultivated in LB medium in various cefotaxime concentrations, 8 μg/ml (in red), 16 μg/ml (in green), 32 μg/ml (in purple), and 64 μg/ml (in yellow).

quantities of CTX-M-15 into the medium to degrade the antibiotic and prevent inhibition of the otherwise susceptible strain 33 (Fig. 3A).

The protection given by the secreted CTX-M-15 of strain 48 was confirmed with the engineered JM109 strains. Only CM from JM109-CTX-M-15 facilitated the growth of strain 33, while CM from TEM- and OXA-producing strains did not (Fig. 3B). Finally, we tested the stability of secreted CTX-M-15 by storing CM from JM109-CTX-M-15 at 4°C for 24 h, 48 h, and 72 h prior to inoculation with strain 33. Again, strain 33 grew in the presence of cefotaxime (Fig. S2). Together, this demonstrated that CTX-M-15 is functionally stable after secretion outside the cell and can provide protection to otherwise susceptible bacterial strains.

The T1SS is involved in the secretion of CTX-M-15. To identify a mechanism of secretion for this CTX-M-15, we first investigated the role of OMVs, which have previously been reported to express ESBL activity (16). To remove OMVs from the supernatant, CM obtained from strain 48 was additionally filtered through a 0.02-μm filter. Filtration did not affect the ESBL activity of the supernatant, and susceptible strain 33 still grew in the presence of cefotaxime when grown in conditioned medium demonstrating that OMVs are not the main mechanism for CTX-M-15 secretion (Fig. S3). These data demonstrated that CTX-M-15 is fully functional in the extracellular medium.

In agreement with their abundance in the XP of strain 48, *in silico* prediction revealed both CTX-M-15 and TEM contained the SP1 leader sequence required for translocation across the cytoplasmic membrane by the Sec pathway (Fig. S4 and Table S5). Strain 48 was predicted to contain four known secretion systems, T1SS, T2SS, T4SS, and T5SS, that therefore were candidates for CTX-M-15 secretion. T1SS and T2SS have potential to be involved in the secretion of hydrolytic enzymes, such as ESBLs. Mutant *E. coli* strains defective for genes required for either T1 secretion (*tolC*) or T2 secretion (*gspD*) were obtained from the Keio database, as was the parental wild type, *E. coli* BW

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

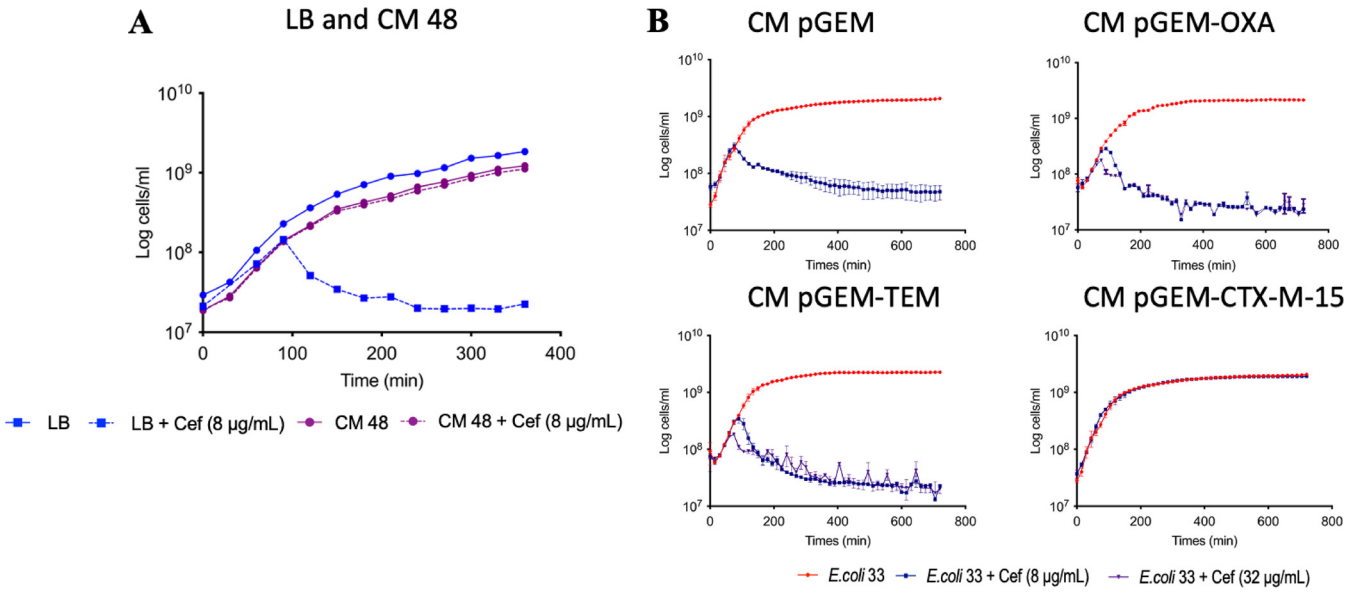


FIG 3 Protective effect of conditioned medium obtained from strain 48 growth chambers. (A) Growth of *E. coli* strain 33 (in blue) cultivated in fresh LB medium compared to growth in CM of strain 48 (in purple). (B) Growth of strain 33 in CM of pGEM, CM of pGEM-OXA, CM of pGEM-TEM, CM of pGEM-CTX-M-15 in absence (in red), and presence of 8 μg/ml (in blue) and 32 μg/ml (in purple) cefotaxime.

25113. pGEM-T-CTX-M-15, conferring cefotaxime resistance, was mobilized into all three strains (*E. coli* BW+, *E. coli* Δ*tolC*+, and *E. coli* Δ*gspD*+ strains), which again were subjected to growth in the presence of cefotaxime. In addition, all three strains, *E. coli* BW, *E. coli* Δ*tolC*, and *E. coli* Δ*gspD* strains, were transformed with an empty pGEM-T vector as controls. As expected, these three control strains failed to grow in the presence of cefotaxime (Fig. S5A). *E. coli* BW+, Δ*tolC*+, and Δ*gspD*+ strains all grew on various concentrations of cefotaxime, but mutation of either T1SS or T2SS significantly inhibited their growth rates (Fig. 4A and Fig. S5B). For the Δ*gspD* mutant, one cefotaxime concentration (8 μg/ml) showed a significant ($P < 0.01$) reduction in growth rate, while the Δ*tolC*+ strain displayed a significantly ($P < 0.05$) lower growth rate when challenged with all three concentrations of cefotaxime (Fig. 4A). To determine if this partial growth inhibition in either mutant was due to an inhibition by CTX-M-15 extracellular secretion, the growth of strain 33 on CM obtained from *E. coli* BW+, *E. coli* Δ*tolC*+, and Δ*gspD*+ cultures was monitored in the presence or absence of cefotaxime. Similar to the growth of both secretion mutants, the growth rate of strain 33 was significantly reduced by the presence of cefotaxime when cultured in CM obtained from either the Δ*tolC*+ or Δ*gspD*+ strain relative to the BW wild-type CM (Fig. 4B and Fig. S5C). Mutation of Δ*tolC* (T1SS) resulted in a greater sensitivity to cefotaxime for either the producer or the susceptible strain, indicating this secretion system is involved in, but not essential for, CTX-M-15 secretion. Given we also observed a smaller, albeit nonsignificant, effect when mutating *gspD* (T2SS), it is likely that both systems can facilitate CTX-M-15 secretion, with T1SS being the most important. Together, these data suggest secretion of CTX-M-15 is not a passive process and is facilitated by common secretion systems present in widespread bacteria.

DISCUSSION

Secretion of ESBLs into the surrounding environment may reduce any damage to the cell wall by preventing entry of β-lactam into the periplasm, where they would be degraded by enzymes such as OXA or TEM. It is also likely to be an important ecological trait, enabling otherwise susceptible bacteria to survive long enough to allow mobilization of plasmid-encoded *bla* genes through conjugation (34, 35). While strain 48 possessed three annotated ESBLs, the data presented here prove that CTX-M-15 was the major secreted ESBL conferring resistance and providing a protective effect. This

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

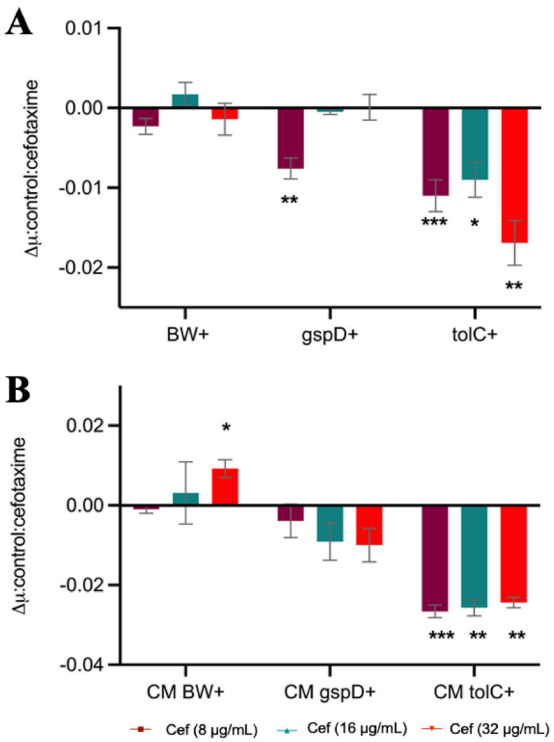


FIG 4 Potential role of T1SS in the secretion of CTX-M-15. (A) Growth rate (μ) of *E. coli* BW+, *E. coli* Δ*gspD*+, and *E. coli* Δ*tolC*+ strains in the presence of 8 μg/ml (in magenta), 16 μg/ml (in green), and 32 μg/ml (in red) cefotaxime. (B) Growth rate of strain 33 cultivated in CM obtained from *E. coli* BW+ (CM BW+), *E. coli* Δ*gspD*+ (CM Δ*gspD*+), and *E. coli* Δ*tolC*+ (CM Δ*tolC*+) in the presence of 8 μg/ml (in magenta), 16 μg/ml (in green), and 32 μg/ml (in red) cefotaxime. Graphs show the difference between the growth rate in the presence and the absence of antibiotic. The value indicates the means ± standard deviations from three biological replicates. *t* test was used to determine significance. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

was further confirmed by exoproteomics, which revealed that CTX-M-15 was the third most abundant protein in the XP. The significant secretion of this resistance enzyme may provide improved protection against the antibiotic and explain how it became selected and transferred onto plasmids that were rapidly disseminated (1, 35–37).

After the gentle filtration of the supernatant, our data demonstrated CTX-M-15 remains fully functional and confers full protection to a susceptible strain. While the presence of ESBLs in OMVs has been linked to extracellular degradation of β-lactam antibiotics (15, 16, 38), to the best of our knowledge, this is the first-time secretion of an individual ESBL (i.e., CTX-M-15) has been linked to T1SS and proven to confer protection to susceptible bacteria. Our data clearly demonstrated CTX-M-15 was actively secreted into the extracellular milieu with no evidence that OMVs played a role. The metallo-β-lactamase NDM-1, which is the most widespread carbapenemase worldwide, has a lipobox proximal to its SP1 leader sequence that enables anchoring to the outer membrane and subsequent export in OMVs (38), a process known as lipidation. Removal of this lipobox inhibited NDM-1 export via OMVs, and the enzyme accumulated in the periplasm. Lipidation only occurs in a few ESBLs, such as BRO-1 (from the human pathogen *Moraxella catarrhalis*) and PenA (from *Burkholderia pseudomallei*) (39, 40), and does not include CTX-M-15, which may explain the lack of OMV involvement. While OMVs were implicated in the secretion and extracellular functioning of a nonlipidated serine β-lactamase CTX-M-1 (16), in our study, removal of vesicles by membrane filtration (0.02 μm) from the culture supernatant did not reduce the efficacy of CTX-M-15 to confer a protective effect, suggesting secretion occurred through an alternative mechanism. OMVs may play a large role in protecting ESBL integrity when inside a host, where biological fluids are likely to provide harsher conditions for

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

enzyme activity. It should be noted that Gram-positive ESBLs are secreted and are stable to external attack; therefore, CTX-M-15 may also be resistant to harsher environmental conditions (41–43).

In contrast, we discovered that mutation of key genes required for either type I or type II secretion inhibited the efficacy of CTX-M-15-induced protection, both to the producer and the susceptible strains. Interestingly, our data proved that T1SS played a role in the direct secretion of CTX-M-15 despite the fact that this ESBL contains a leader sequence for localization in the periplasm. The differential effect of mutants on CTX-M-15-induced protection suggests a more direct role for the T1SS. The T2SS may also play a role in secretion of CTX-M-15, although such a small amount of proteins in the XP could be explained by spontaneous release of the periplasmic proteins. Indeed, out of the top five most abundant proteins in the XP, OsmY and two ligand binding proteins were also predicted to be periplasmic (44, 45) but were still found in the culture supernatant. Exoproteomics often captures a range of periplasmic proteins, especially ligand binding proteins (46–48), which suggests that the outer membrane is leaky. The mechanism of secretion for OsmY has not been studied, but it was used in biotechnology to deliver proteins into the medium via C-terminal fusion (49, 50). Reports concerning secretion in *E. coli* remain elusive mainly because nonpathogenic laboratory strains generally express a small amount of proteins in the culture medium (51, 52). Identifying the causal mechanism for CTX-M-15 secretion could help develop novel therapeutic drugs to block secretion, as we have proven it is essential for resistance under exposure to cefotaxime.

We have strong evidence based on the presence of enzyme, bioinformatics, and mutant studies that T1SS and not T2SS is responsible for the secretion of all CTX-M-15 in the exoproteome, but further work is required to consolidate this observation and fully establish the secretion pathway for CTX-M-15. It is feasible that secretion of CTX-M-15 represents an evolutionary advantage, as no damage would occur to the cell wall if the antibiotic is disabled outside the cell, in opposition to hydrolysis in the periplasm. The role of environmental contamination in the transmission of *Enterobacteriaceae* and, in particular, *E. coli* ST131 is increasingly recognized. However, factors influencing duration of survival in the environment have not yet been extensively studied.

MATERIALS AND METHODS

Bacterial strains and growth medium. Bacterial strains used in this study are listed in the supplemental information (see Table S1). Environmental *E. coli* strains were both isolated from the River Sowe, Coventry, United Kingdom, namely, *E. coli* ST3576 O8:H7 (strain 33) and *E. coli* ST131 O25:H4 (strain 48). Commercial laboratory strains of *E. coli* JM109, *E. coli* BW2511, *E. coli* JW5503, and *E. coli* JW5707 were also used. Cells were routinely grown in lysogeny broth (LB) liquid (10 g/liter tryptone, 5 g/liter yeast extract, 10 g/liter sodium chloride) or LB agar medium (addition of 15 g/liter agar). The following antibiotics were supplemented when required: 8 μg/ml cefotaxime, 100 μg/ml ampicillin, and 5 μg/ml kanamycin. Additionally, culture medium for JM109 was supplemented with isopropyl β-D-thiogalactosidase (IPTG) (0.1 M) and X-galactosidase (20 mg/ml) to induce expression of recombinant CTX-M-15. Cells were incubated at 37°C with either shaking (200 rpm) or static conditions.

Antimicrobial phenotypic screening. Oxid antibiotic discs were used to determine phenotypic resistance profiles. Strains 33 and 48 were streaked on LB agar plates, and discs containing either 25 μg ampicillin, 5 μg cefotaxime, 10 μg imipenem, 30 μg chloramphenicol, or 8 μg erythromycin were added on top of the plates. All of the plates were incubated overnight at 37°C.

ESBL genotypic screening. Single colonies of strains 33 and 48 were picked and individually inoculated into 10 ml LB and incubated overnight at 37°C with shaking at 150 rpm. Cultures were then centrifuged at (1,500 rpm for 10 min) and supernatant discarded. Pellets were resuspended in 500 μl phosphate-buffered saline and used for DNA extraction using the MPBio FastDNA spin kit by following the manufacturer's guidelines. Specific primers for amplification of ESBL genes, $bla_{\text{CTX-M-15}}$, bla_{TEM}, and bla_{OXA}, were designed from the Illumina sequencing done previously (53) (Table S2). PCRs were done using 12.5 μl Master Mix 2× (Promega), 1.25 μl dimethyl sulfoxide, 0.8 μM forward primer, 0.8 μM reverse primer, 2 μl DNA template, and distilled H_2O for a final PCR volume of 25 μl. PCR was performed at an initial denaturation temperature of 95°C for 5 min, followed by 34 cycles of denaturation at 95°C for 30 s, annealing temperature at 55 to 66°C (depending on the primer set) for 30 s, and extension for 1 min 50 s. A final extension was performed at 72°C for 5 min.

Antibiotic resistance screening. Growth curves were implemented in 96-well plates with 200 μl culture per well and incubated at 37°C with shaking (200 rpm) in a microplate reader (POLARstar Omega; BMG Labtech). As inoculum, overnight starter cultures of each bacterial strain (5 ml) were

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

diluted to an initial concentration of 3×10^7 cells/ml. Culture media were supplemented with 0, 8, 16, 32, or 64 μg/ml cefotaxime (VWR International Ltd.). Cell proliferation was determined by measuring the optical density at 600 nm for 8 or 12 h every 15 min. Each condition was set up in triplicate. Exponential growth rates were calculated for the growth of *E. coli* BW+, *E. coli* Δ*tolC*+, and *E. coli* Δ*gspD*+ strains and for the protective effect on strain 33 by using the formula $P(t) = P_0e^{rt}$, where $P(t)$ is the amount of cell number at time t, P_0 the initial cell number, r the growth rate, and t the number of periods (54). Two-sample t test was performed to compare the significance of the growth rate differences.

Generation of conditioned medium. Strain 48, or the engineered laboratory *E. coli* strains harboring CTX-M-15, were grown in the presence of cefotaxime (8 μg/ml). After overnight growth, cells were removed by pelleting (3,228 $\times$ *g* for 15 min). Supernatant was carefully filtered through a 0.22-μm membrane (Fisher Scientific) to avoid cell lysis. Conditioned medium (CM) was diluted 1:1 (vol/vol) with fresh LB medium and supplemented with various concentrations of cefotaxime. Overnight cultures of susceptible strain 33 were inoculated (1%, vol/vol) in the conditioned medium and grown as described above.

Cloning of the ESBLs, $bla_{\text{CTX-M-15}}$, bla_{TEM}, and bla_{OXA}. A full list of primers used in this study is presented in Table S2, and the genes $bla_{\text{CTX-M-15}}$, $bla_{\text{TEM-1}}$, and $bla_{\text{OXA-1}}$ were cloned from strain 48 into the cloning vector pGEM-T easy (Promega, UK) using the HiFi assembly kit (New England, Biolabs). The newly constructed plasmids pGEM-CTX-M-15, pGEM-TEM, and pGEM-OXA, and control empty-vector pGEM-T, were transformed into *E. coli* JM109.

Detection of β-lactamase activity by the nitrocefin assay. β-Lactamase activity was assessed by colorimetric assay using the chromogenic cephalosporin compound nitrocefin (Thermo Scientific). Strain 33 and strain 48 were inoculated in LB or M9 minimal medium (33.9 g/liter Na_2HPO_4, 15 g/liter KH_2PO_4, 5 g/liter NH_4Cl, 2.5 g/liter NaCl, 20% glucose, 1 M $MgSO_4$, 1 M $CaCl_2$) supplemented with 0, 2, 4, or 8 μg/ml cefotaxime. Strains were grown at 37°C until mid-exponential phase, and supernatant was collected by first removing cells (4,000 rpm for 15 min) and then gentle filtration through a 0.22-μm membrane (Fisher Scientific) to prevent cell lysis and remove intact cells. Supernatants were incubated with 15 μg/ml nitrocefin (stock concentration, 500 μg/ml) at room temperature (~22°C) for 30 min.

Determination of $bla_{\text{CTX-M-15}}$ transcription in strain 48. Strain 48 was grown at 37°C in LB supplemented with 0 or 8 μg/ml cefotaxime. Diluted cultures were grown at 37°C with shaking (200 rpm) to exponential phase before RNA extraction. A detailed protocol for extraction, reverse transcription, and quantitative PCR can be found in the supplemental material.

Preparation of exoproteome, total proteome samples, and liquid chromatography tandem mass spectrometry analysis. Exoproteomes and total proteomes of strain 48 were prepared by adapting the protocol described in Christie-Oleza and Armengaud (47). Briefly, Strataclean beads (Agilent) were used to isolate proteins instead of trichloroacetic acid (TCA) precipitation. A detailed procedure is provided in the supplemental material.

Peptide identification and comparative proteomic analysis. A custom database was made with the genome of strain 48 by using Prokka v1.14.5 for annotation (59), and MASCOT was used to assign peptide to protein by using the custom database. Identified proteins were further analyzed using Scaffold (55) (protein threshold, 99.9%; minimum peptide 2, peptide threshold, 80%). The normalized spectral abundance factor (NSAF) (56) was calculated for each protein to compare the abundance for all proteins. Two-sample t test was used to determine if the presence of antibiotic significantly impacted protein abundance.

***In silico* prediction of protein localization and secretion pathways.** Analysis was done on the SignalP 5.0 server (http://www.cbs.dtu.dk/services/SignalP-5.0/) and enabled the prediction of the presence and the location of cleavage sites in the three β-lactamase proteins CTX-M-15, TEM, and OXA using the Fasta sequences generated in-house (see the supplemental material) (57). The TXSScan web tool (https://galaxy.pasteur.fr/) (58) was used for prediction of the presence of secretion systems in strain 48 using the genome of *E. coli* 48.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.4 MB.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the Natural Environment Research Council (grant NE/N019857/1). S.R. was supported by a CENTA NERC studentship.

We declare that we have no conflict of interest.

REFERENCES

1. Bevan ER, Jones AM, Hawkey PM. 2017. Global epidemiology of CTX-M beta-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother 72:2145–2155. https://doi.org/10.1093/jac/dkx146.
2. Canton R, Coque TM. 2006. The CTX-M beta-lactamase pandemic. Curr Opin Microbiol 9:466–475. https://doi.org/10.1016/j.mib.2006.08.011.
3. Rossolini GM, D'Andrea MM, Mugnaioli C. 2008. The spread of CTX-M-type extended-spectrum beta-lactamases. Clin Microbiol Infect 14(Suppl 1):33–41. https://doi.org/10.1111/j.1469-0691.2007.01867.x.
4. Coque TM, Baquero F, Canton R. 2008. Increasing prevalence of ESBL-producing Enterobacteriaceae in Europe. Euro Surveill 13:19–44.

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

5. Pitout JD, DeVinney R. 2017. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. F1000Res 6:F1000 Faculty Rev-195. [PMC] https://doi.org/10.12688/f1000research.10609.1.
6. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, Baquero F, Canton R, Nordmann P. 2008. Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. Emerg Infect Dis 14:195–200. https://doi.org/10.3201/eid1402.070350.
7. Hawkey PM, Jones AM. 2009. The changing epidemiology of resistance. J Antimicrob Chemother 64(Suppl 1):i3–i10. https://doi.org/10.1093/jac/dkp256.
8. Bush K, Jacoby GA. 2010. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother 54:969–976. https://doi.org/10.1128/AAC.01009-09.
9. Peirano G, Pitout JD. 2010. Molecular epidemiology of Escherichia coli producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. Int J Antimicrob Agents 35:316–321. https://doi.org/10.1016/j.ijantimicag.2009.11.003.
10. Woodford N, Carattoli A, Karisik E, Underwood A, Ellington MJ, Livermore DM. 2009. Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major Escherichia coli lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. Antimicrob Agents Chemother 53:4472–4482. https://doi.org/10.1128/AAC.00688-09.
11. Ramadan AA, Abdelaziz NA, Amin MA, Aziz RK. 2019. Novel blaCTX-M variants and genotype-phenotype correlations among clinical isolates of extended spectrum beta lactamase-producing Escherichia coli. Sci Rep 9:4224. https://doi.org/10.1038/s41598-019-39730-0.
12. Brook I. 2004. Beta-lactamase-producing bacteria in mixed infections. Clin Microbiol Infect 10:777–784. https://doi.org/10.1111/j.1198-743X.2004.00962.x.
13. Mashburn-Warren LM, Whiteley M. 2006. Special delivery: vesicle trafficking in prokaryotes. Mol Microbiol 61:839–846. https://doi.org/10.1111/j.1365-2958.2006.05272.x.
14. Kulp A, Kuehn MJ. 2010. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. Annu Rev Microbiol 64:163–184. https://doi.org/10.1146/annurev.micro.091208.073413.
15. Ciofu O, Beveridge T, Walther-Rasmussen J, Hoiby N, Kadurugamuwa J. 2000. Chromosomal beta-lactamase is packaged into membrane vesicles and secreted from Pseudomonas aeruginosa. J Antimicrob Chemother 45:9–13. https://doi.org/10.1093/jac/45.1.9.
16. Kim SW, Park SB, Im SP, Lee JS, Jung JW, Gong TW, Lazarte JMS, Kim J, Seo JS, Kim JH, Song JW, Jung HS, Kim GJ, Lee YJ, Lim SK, Jung TS. 2018. Outer membrane vesicles from beta-lactam-resistant Escherichia coli enable the survival of beta-lactam-susceptible E. coli in the presence of beta-lactam antibiotics. Sci Rep 8:5402. https://doi.org/10.1038/s41598-018-23656-0.
17. Robinson C, Matos CF, Beck D, Ren C, Lawrence J, Vasisht N, Mendel S. 2011. Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. Biochim Biophys Acta 1808:876–884. https://doi.org/10.1016/j.bbamem.2010.11.023.
18. Green ER, Mecsas J. 2016. Bacterial secretion systems: an overview. Microbiol Spectr 4:10.1128/microbiolspec.VMBF-0012-2015. [PMC] https://doi.org/10.1128/microbiolspec.VMBF-0012-2015.
19. Tsirigotaki A, De Geyter J, Sostaric N, Economou A, Karamanou S. 2017. Protein export through the bacterial Sec pathway. Nat Rev Microbiol 15:21–36. https://doi.org/10.1038/nrmicro.2016.161.
20. Piddock LJV. 2006. Clinically relevant bacterial chromosomally encoded multi-drug resistance efflux pumps. Clin Microbiol Rev 19:382–402. https://doi.org/10.1128/CMR.19.2.382-402.2006.
21. Kulkarni R, Dhakal BK, Slechta ES, Kurtz Z, Mulvey MA, Thanassi DG. 2009. Roles of putative type II secretion and type IV pilus systems in the virulence of uropathogenic Escherichia coli. PLoS One 4:e4752. https://doi.org/10.1371/journal.pone.0004752.
22. Tauschek M, Gorrell RJ, Strugnell RA, Robins-Browne RM. 2002. Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of Escherichia coli. Proc Natl Acad Sci U S A 99:7066–7071. https://doi.org/10.1073/pnas.092152899.
23. Nivaskumar M, Francetic O. 2014. Type II secretion system: a magic beanstalk or a protein escalator. Biochim Biophys Acta 1843:1568–1577. https://doi.org/10.1016/j.bbamcr.2013.12.020.
24. Smith P, Schuster M. 2019. Public goods and cheating in microbes. Curr Biol 29:R442–R447. https://doi.org/10.1016/j.cub.2019.03.001.
25. Lidbury I, Murphy ARJ, Fraser TD, Bending GD, Jones AME, Moore JD, Goodall A, Tibbett M, Hammond JP, Scanlan DJ, Wellington EMH. 2017. Identification of extracellular glycerophosphodiesterases in Pseudomonas and their role in soil organic phosphorus remineralisation. Sci Rep 7:2179. https://doi.org/10.1038/s41598-017-02327-6.
26. Reintjes G, Arnosti C, Fuchs BM, Amann R. 2017. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J 11:1640–1650. https://doi.org/10.1038/ismej.2017.26.
27. Enke TN, Datta MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, Pascual-Garcia A, Cordero OX. 2019. Modular assembly of polysaccharide-degrading marine microbial communities. Curr Biol 29:1528–1535. https://doi.org/10.1016/j.cub.2019.03.047.
28. Whitmer GR, Moorthy G, Arshad M. 2019. The pandemic Escherichia coli sequence type 131 strain is acquired even in the absence of antibiotic exposure. PLoS Pathog 15:e1008162. https://doi.org/10.1371/journal.ppat.1008162.
29. Nicolas-Chanoine MH, Bertrand X, Madec JY. 2014. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev 27:543–574. https://doi.org/10.1128/CMR.00125-13.
30. Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ, Upton M. 2012. Population structure, virulence potential and antibiotic susceptibility of uropathogenic Escherichia coli from Northwest England. J Antimicrob Chemother 67:346–356. https://doi.org/10.1093/jac/dkr451.
31. Pitout JD. 2012. Extraintestinal pathogenic Escherichia coli: a combination of virulence with antibiotic resistance. Front Microbiol 3:9. https://doi.org/10.3389/fmicb.2012.00009.
32. Wiles TJ, Kulesus RR, Mulvey MA. 2008. Origins and virulence mechanisms of uropathogenic Escherichia coli. Exp Mol Pathol 85:11–19. https://doi.org/10.1016/j.yexmp.2008.03.007.
33. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, Choroszy-Krol I. 2019. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: recent reports. Gut Pathog 11:10. https://doi.org/10.1186/s13099-019-0290-0.
34. De La Cruz F, Davies J. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol 8:128–133. https://doi.org/10.1016/s0966-842x(00)01703-0.
35. Amos GC, Hawkey PM, Gaze WH, Wellington EM. 2014. Waste water effluent contributes to the dissemination of CTX-M-15 in the natural environment. J Antimicrob Chemother 69:1785–1791. https://doi.org/10.1093/jac/dku079.
36. Amos GCA, Ploumakis S, Zhang L, Hawkey PM, Gaze WH, Wellington EMH. 2018. The widespread dissemination of integrons throughout bacterial communities in a riverine system. ISME J 12:681–691. https://doi.org/10.1038/s41396-017-0030-8.
37. Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum beta-lactamases in the community: toward the globalization of CTX-M. Clin Microbiol Rev 26:744–758. https://doi.org/10.1128/CMR.00023-13.
38. Gonzalez LJ, Bahr G, Nakashige TG, Nolan EM, Bonomo RA, Vila AJ. 2016. Membrane anchoring stabilizes and favors secretion of New Delhi metallo-beta-lactamase. Nat Chem Biol 12:516–522. https://doi.org/10.1038/nchembio.2083.
39. Bootsma HJ, Van Dijk H, Verhoef J, Fleer A, Mooi FR. 1996. Molecular characterization of the BRO beta-lactamase of Moraxella (Branhamella) catarrhalis. Antimicrob Agents Chemother 40:966–972. https://doi.org/10.1128/AAC.40.4.966.
40. Randall LB, Dobos K, Papp-Wallace KM, Bonomo RA, Schweizer HP. 2015. Membrane-bound PenA beta-lactamase of Burkholderia pseudomallei. Antimicrob Agents Chemother 60:1509–1514. https://doi.org/10.1128/AAC.02444-15.
41. Anne J, Economou A, Bernaerts K. 2017. Protein secretion in Gram-positive bacteria: from multiple pathways to biotechnology. Curr Top Microbiol Immunol 404:267–308. https://doi.org/10.1007/82_2016_49.
42. Lee J, Lee EY, Kim SH, Kim DK, Park KS, Kim KP, Kim YK, Roh TY, Gho YS. 2013. Staphylococcus aureus extracellular vesicles carry biologically active beta-lactamase. Antimicrob Agents Chemother 57:2589–2595. https://doi.org/10.1128/AAC.00522-12.
43. Liu Y, Defourny KAY, Smid EJ, Abee T. 2018. Gram-positive bacterial extracellular vesicles and their impact on health and disease. Front Microbiol 9:1502. https://doi.org/10.3389/fmicb.2018.01502.
44. Yim HH, Villarejo M. 1992. osmY, a new hyperosmotically inducible gene, encodes a periplasmic protein in Escherichia coli. J Bacteriol 174:3637–3644. https://doi.org/10.1128/jb.174.11.3637-3644.1992.
45. Oh BH, Kang CH, De Bondt H, Kim SH, Nikaido K, Joshi AK, Ferro-Luzzi Ames G. 1995. The bacterial periplasmic histidine-binding protein. J Biol Chem 270:16097–14143. https://doi.org/10.1016/S0021-9258(17)41754-6.

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

46. Lidbury I, Krober E, Zhang Z, Zhu Y, Murrell JC, Chen Y, Schafer H. 2016. A mechanism for bacterial transformation of dimethylsulfide to dimethylsulfoxide: a missing link in the marine organic sulfur cycle. Environ Microbiol 18:2754–2766. https://doi.org/10.1111/1462-2920.13354.
47. Christie-Oleza JA, Armengaud J. 2010. In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the Ruegeria pomeroyi DSS-3 case-study. Mar Drugs 8:2223–2239. https://doi.org/10.3390/md8082223.
48. Christie-Oleza JA, Armengaud J. 2015. Proteomics of the Roseobacter clade, a window to the marine microbiology landscape. Proteomics 15:3928–3942. https://doi.org/10.1002/pmic.201500222.
49. Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Lee TS, Tullman-Ercek D, Voigt CA, Simmons BA, Keasling JD. 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. Proc Natl Acad Sci U S A 108:19949–19954. https://doi.org/10.1073/pnas.1106958108.
50. Qian ZG, Xia XX, Choi JH, Lee SY. 2008. Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in Escherichia coli. Biotechnol Bioeng 101:587–601. https://doi.org/10.1002/bit.21898.
51. Papanikou E, Karamanou S, Economou A. 2007. Bacterial protein secretion through the translocase nanomachine. Nat Rev Microbiol 5:839–851. https://doi.org/10.1038/nrmicro1771.
52. Kotzsch A, Vernet E, Hammarström M, Berthelsen J, Weigelt J, Gräslund S, Sundström M. 2011. A secretory system for bacterial production of high-profile protein targets. Protein Sci 20:597–609. https://doi.org/10.1002/pro.593.
53. Hill G. 2016. Investigating WWTP impact on antibiotic resistance within UK river systems. Doctoral Dissertation University of Warwick, Coventry, UK.
54. Hall BG, Acar H, Nandipati A, Barlow M. 2014. Growth rates made easy. Mol Biol Evol 31:232–238. https://doi.org/10.1093/molbev/mst187.
55. Searle BC. 2010. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. Proteomics 10:1265–1269. https://doi.org/10.1002/pmic.200900437.
56. Zhu W, Smith JW, Huang CM. 2010. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010:840518. https://doi.org/10.1155/2010/840518.
57. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.
58. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
59. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Downloaded from https://journals.asm.org/journal/aac on 08 June 2022 by 137.205.238.49.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

The Proteobacterial Methanotroph *Methylosinus trichosporium* OB3b Remodels Membrane Lipids in Response to Phosphate Limitation

Julie Scanlan,[a] Richard Guillonneau,[a] Mark R. Cunningham,[b] Sahanara Najmin,[a] Michaela A. Mausz,[a] Andrew Murphy,[a] Leanne L. Murray,[b] Limei Zhang,[a,c] Deepak Kumaresan,[b] Yin Chen[a]

[a]School of Life Sciences, University of Warwick, Coventry, UK
[b]School of Biological Sciences, Queen's University Belfast, Belfast, UK
[c]Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China

ABSTRACT Methane is a potent greenhouse gas in the atmosphere, and its concentration has continued to increase in recent decades. Aerobic methanotrophs, bacteria that use methane as the sole carbon source, are an important biological sink for methane, and they are widely distributed in the natural environment. However, relatively little is known on how methanotroph activity is regulated by nutrients, particularly phosphorus (P). P is the principal nutrient constraining plant and microbial productivity in many ecosystems, ranging from agricultural land to the open ocean. Using a model methanotrophic bacterium, *Methylosinus trichosporium* OB3b, we demonstrate here that this bacterium can produce P-free glycolipids to replace membrane phospholipids in response to P limitation. The formation of the glycolipid monoglucuronic acid diacylglycerol requires *plcP-agt* genes since the *plcP-agt* mutant is unable to produce this glycolipid. This *plcP-agt*-mediated lipid remodeling pathway appears to be important for *M. trichosporium* OB3b to cope with P stress, and the mutant grew significantly slower under P limitation. Interestingly, comparative genomics analysis shows that the ability to perform lipid remodeling appears to be a conserved trait in proteobacterial methanotrophs; indeed, *plcP* is found in all proteobacterial methanotroph genomes, and *plcP* transcripts from methanotrophs are readily detectable in metatranscriptomics data sets. Together, our study provides new insights into the adaptation to P limitation in this ecologically important group of bacteria.

IMPORTANCE Methane is a potent greenhouse gas in the atmosphere, and its concentration has continued to increase steadily in recent decades. In the natural environment, bacteria known as methanotrophs help mitigate methane emissions at no cost to human beings. However, relatively little is known regarding how methane oxidation activity in methanotrophs is regulated by soil nutrients, particularly phosphorus. Here, we show that methanotrophs can modify their membrane in response to phosphorus limitation and that the ability to change membrane lipids is important for methanotroph activity. Genome and metatranscriptome analyses suggest that such an adaptation strategy appears to be strictly conserved in all proteobacterial methanotrophs and is used by these bacteria in the natural environment. Together, our study provides a plausible molecular mechanism for better understanding the role of phosphorus on methane oxidation in the natural environment.

KEYWORDS *Methylosinus*, lipid remodeling, methanotroph

Methane is a powerful greenhouse gas, and its atmospheric concentrations have increased 2.5-fold since preindustrial times (1). Worryingly, methane concentrations in the atmosphere have continued to rise at a rate of ~7.6 ppb each year over

Editor Stephen J. Giovannoni, Oregon State University

Copyright © 2022 Scanlan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Yin Chen, y.chen.25@warwick.ac.uk.

The authors declare no conflict of interest.

Received 27 January 2022
Accepted 25 April 2022

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

the past decade (1, 2). Both human activities and natural environment contribute to methane emissions; however, oxidation of methane by microbes is the only known biological sink for this greenhouse gas. It is estimated that microbial methane uptake by soil comprises ~30 Tg yr^{-1}, accounting for 5% of annual methane sink globally (1, 2). Methane-oxidizing microbes are widely distributed in the environment, from agricultural soils to natural wetlands and from coastal sediments to deep-sea methane seeps (reviewed in references 3 and 4). As such, aerobic methanotrophs play an important role in mitigating methane emissions before it is released into the atmosphere.

Phylogenetically, aerobic methanotrophs have been found in *Alphaproteobacteria* (*Methylocystaceae* and *Beijerinckiaceae*), *Gammaproteobacteria* (*Methylococcaceae* and *Methylothermaceae*), and *Verrucomicrobia* (*Methylacidiphilaceae*) (3, 5). Traditionally, these aerobic methanotrophs have been classified as type I (*Methylococcaceae* and *Methylothermaceae*), type II (*Methylocystaceae* and *Beijerinckiaceae*), and type III (*Methylacidiphilaceae*) based on contrasting characteristics in physiology, morphology, and membrane ultrastructures, although this rather simplified classification system for aerobic methanotrophs struggles to keep up with the pace of the ever-expanding discovery of novel methanotrophs in diverse habitats (6).

A variety of factors are known to impact methanotroph activities in the environment, including methane and oxygen concentrations, trace metals (e.g., copper which is required for the activity of the membrane-bound particulate methane monooxygenase [pMMO]), temperature, and salinity, as well as changes in land use (reviewed in reference 4). An important, yet largely overlooked, macronutrient for methanotrophs is phosphorus (P). P is a key nutrient for all forms of life, being a major constituent of nucleic acids and membrane lipids. Indeed, phospholipid fatty acid has long been used as an important biomarker for the identification of methanotrophs and quantification of their activity in response to environmental changes (7). However, conflicting results in the literature exist regarding the role of P in methanotroph activity (8). For example, in Dutch drainage ditches, a strong positive correlation between methane oxidation potential and phosphate concentrations has been observed (8). Similarly, in natural forest and paddy soils, phosphate addition significantly enhanced methane oxidation activities (9, 10). However, other reports found no impact or inhibition of P addition on methane oxidation in microcosm experiments (11, 12). Since P is an essential nutrient for all life forms, it is perhaps not surprising that a variety of strategies are adopted by microbes to cope with P limitation, including reduction of the cellular P quota, recycling of P-containing molecules, and an increase in organic and inorganic P uptake (reviewed in reference 13). As such, a mechanistic understanding of how P affects methanotroph activity is needed in order to better understand the role of P on methane oxidation dynamics in the natural environment.

Here, we report the identification of glycoglycerolipids replacing membrane phospholipids in response to P limitation in the model methanotrophic bacterium *Methylosinus trichosporium* OB3b (14). Such a membrane lipid remodeling process involves the *plcP* and *agt* genes, which appear to be common in proteobacterial methanotrophs. Interestingly, lipid remodeling is important for the bacterium to grow on methane at low P concentrations. Our study thus provides a mechanistic role for how P availability affects methane oxidation.

RESULTS

***M. trichosporium* OB3b remodel membrane lipids in response to P limitation.** Intact phospholipids in *M. trichosporium* OB3b have previously been analyzed, and it is known that its lipidome is dominated by phospholipids, primarily phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) and its methylated derivatives monomethyl-PE (MMPE) and dimethyl-PE (DMPE) (15, 16). Indeed, when the strain was cultivated under the nitrate mineral salts (NMS) medium, which has 4 mM phosphate, only PG and PE and its methylated derivatives were found. PG is represented by a major species eluted at ~6.1 min with a mass-to-charge ratio (*m/z*) of 773.6 ($C_{36:2}$) in the negative

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

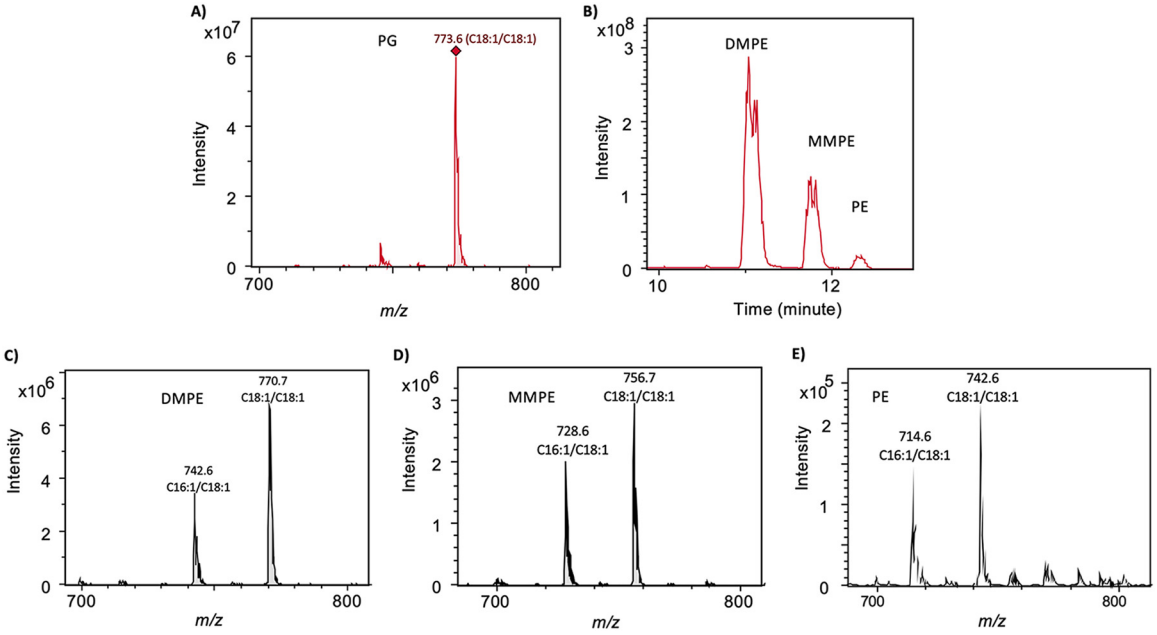


FIG 1 Formation of phospholipids in *Methylosinus trichosporium* OB3b cultivated in 1× nitrate mineral salts (NMS) medium supplemented with 4 mM phosphate. (A) The phospholipid PG is dominated by one species with an *m/z* of 773.6 that is composed of two monounsaturated fatty acids ($C_{18:1}/C_{18:1}$). (B to E) The phospholipid phosphatidylethanolamine (PE) and its methylated derivatives (monomethyl phosphatidylethanolamine, MMPE; dimethyl phosphatidylethanolamine, DMPE) were eluted consecutively between 10 and 13 min, all of which were dominated by two species composed of two monounsaturated fatty acids of $C_{16:1}/C_{18:1}$ and $C_{18:1}/C_{18:1}$, respectively).

mode ionization (Fig. 1A). When fragmented by MS^n, this PG species leads to the formation of $C_{18:1}$ fatty acid (see Fig. S1 in the supplemental material), which was common in type II methanotrophs (7). Thus, the bacterium synthesizes a dominant PG lipid consisting of the two monounsaturated fatty acids $C_{18:1}$ and $C_{18:1}$ ($C_{18:1}/C_{18:1}$). PE, MMPE, and DMPE were eluted between 11 and 13 min (Fig. 1B). PE and its derivatives had characteristic neutral losses of 141 (PE), 155 (MMPE), and 169 (DMPE), respectively, in the positive ionization mode (see Fig. S2). Each of these lipids are represented by two dominant species that are composed of $C_{16:1}/C_{18:1}$ and $C_{18:1}/C_{18:1}$ fatty acids, respectively (Fig. 1C to E). Thus, as with PG, PE and its derivatives also appear to be dominated by monounsaturated fatty acid, primarily $C_{18:1}$.

To determine the impact of phosphate concentrations on membrane lipids in this methanotroph, we replaced phosphate buffer in the NMS medium with 10 mM HEPES buffer (pH 6.8) and added phosphate back to the medium at a range of concentrations from 10 μM to 4 mM. Diversity of lipids from this bacterium remain largely unchanged when phosphate was supplemented between 50 μM to 4 mM (see Fig. S3), and its lipidome is predominated by PG, PE, and its derivatives (i.e., MMPE and DMPE). However, at 10 μM phosphate in the medium, the intensity of these phospholipids significantly diminished, and new lipids were synthesized (Fig. 2A and B). At approximately 7 to 7.5 min, two major lipids were found with *m/z* values of 800.6 and 772.6, respectively (Fig. 2A). These lipids produce a characteristic neutral loss of 179 in the positive fragmentation by MS^n (Fig. 2C), consistent with the monoglucosyldiacylglycerol (MGDG) lipid which was identified previously in other bacteria (see, for example, references 17 and 18). At approximately 9.5 to 10 min, two lipids with *m/z* values of 814.6 and 786.6 were found, which showed a neutral loss of 193 when fragmented in the positive mode (Fig. 2B and D). This is consistent with glucuronic acid diacylglycerol (GADG) lipids that appear to be common surrogate lipids in response to P limitation in other bacteria (17–19). Interestingly, the methanotroph also produced amino-acid-containing ornithine lipids (20), which were eluted at approximately 13 to 13.5 min

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

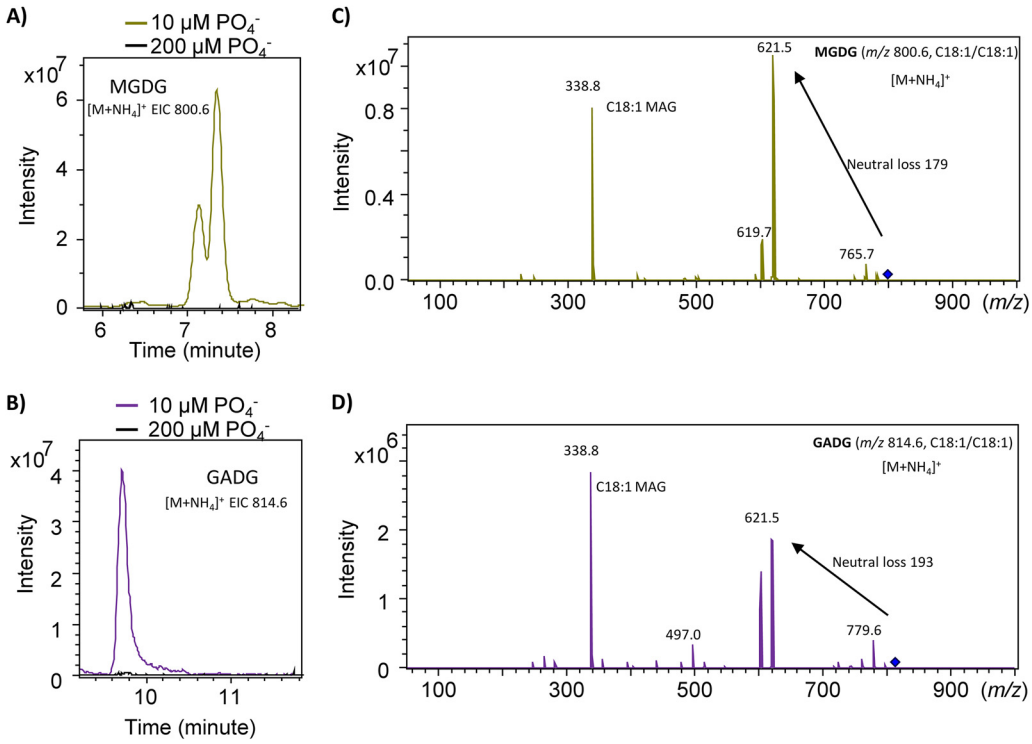


FIG 2 Formation of the new glyceroglycolipids monoglucosyldiacylglycerol (MGDG) and glucuronic acid diacylglycerol (GADG) in response to phosphate limitation in *Methylosinus trichosporium* OB3b. (A) Formation of the MGDG lipid (extracted ion chromatogram, EIC of the *m/z* 800.6 species, width ± 0.2) showing a characteristic neutral loss of 179 of the hexose (C). (B) Formation of the GADG lipid (extracted ion chromatogram [EIC] of the *m/z* 814.6 species, width ± 0.2) showing a characteristic neutral loss of 193 of the hexuronic acid (D). MS^n fragmentation of MGDG and GADG (showing as an ammonium adduct $[M+NH_4]^+$ in the positive mode ionization) produces a monoacylglycerol (MAG) of $C_{18:1}$ (C and D).

with characteristic product ions of ornithine lipids (see Fig. S4). Thus, in response to low phosphate in the medium, this methanotrophic bacterium is capable of remodeling its membrane lipids using several surrogate lipids. To the best of our knowledge, these non-P surrogate lipids have not been reported in methanotrophs previously.

The *plcP-agt* genes are involved in glycolipids formation in *M. trichosporium* OB3b. Formation of MGDG and GADG glycolipids requires a phospholipase C-type protein PlcP and a glycosyltransferase Agt and this PlcP-Agt pathway has been studied in detail in several bacteria, including the marine roseobacters (18, 19), *Pseudomonas aeruginosa* (17) and *Agrobacterium tumefaciens* (21). In the genome of strain OB3b, a putative *plcP-agt* cluster was found (Fig. 3). PlcP and Agt of strain OB3b had 51 and 43% sequence identity to the characterized counterparts from *Pelagibacter ubique* HTCC7211, respectively (22, 23). Multiple sequence alignment and phylogenetic analyses of PlcP (Fig. 3D) and Agt (Fig. 3E) from strain OB3b show that they are closely related to the corresponding proteins that are known to be involved in lipid remodeling (17–19, 22, 24). Indeed, key residues that are known to be essential for PlcP and Agt activities are strictly conserved (see Fig. S5), suggesting that PlcP-Agt may encode a functional lipid remodeling pathway in this methanotroph (Fig. 2).

To confirm the role of *plcP-agt* genes in lipid remodeling in strain OB3b, a knockout mutant was generated using marker exchange mutagenesis. Lipidomics analyses showed that the mutant was unable to produce GADG anymore (Fig. 3), whereas production of MGDG (Fig. 3) and OL (not shown) was not affected. It is therefore likely that an as-yet-unidentified glycosyltransferase may be responsible for the synthesis of MGDG in strain OB3b. Together, the data suggest that the PlcP-Agt pathway in *M. trichosporium* OB3b is responsible for GADG biosynthesis during lipid remodeling in response to P limitation.

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

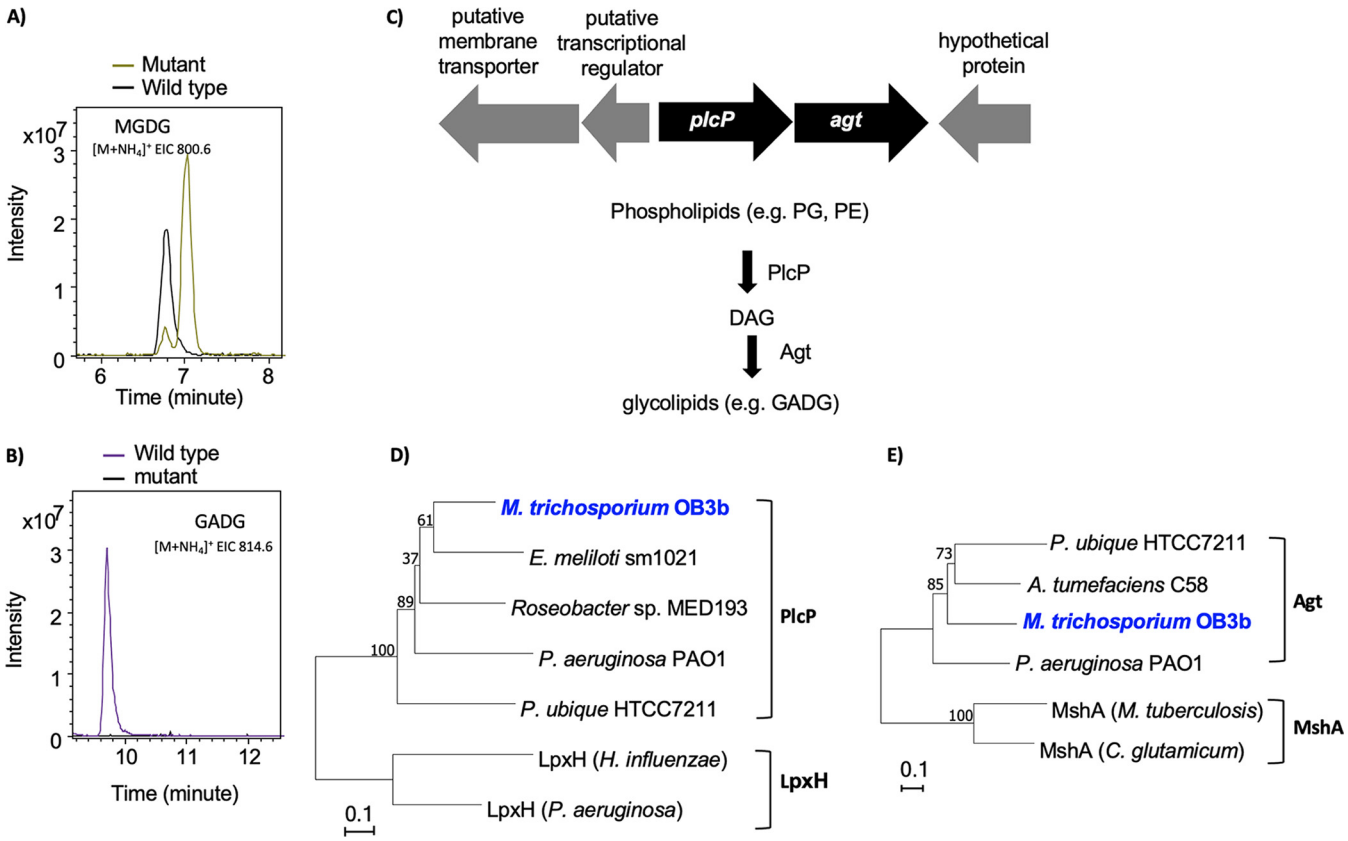


FIG 3 The *plcP-agt* operon in *Methylosinus trichosporium* OB3b is responsible for the formation of the GADG glycolipids. (A and B) Formation of the glycolipids MGDG (A) and GADG (B) in the wild type, but GADG is not produced in the *plcP-agt* deletion mutant (B). (C) *plcP-agt* genes and their neighborhood and the proposed pathway for lipid remodeling involving PlcP and Agt. DAG, diacylglycerol. (D) Neighbor-joining phylogenetic tree showing PlcP of strain OB3b, together with characterized PlcP from other bacteria, including *Ensifer meliloti* (24), *Roseobacter* sp. MED193 (18), *Pseudomonas aeruginosa* (17), and *Pelagibacter ubique* (19). LpxH encoding an enzyme involved in lipid A biosynthesis is used as an outgroup (49, 50). (E) Neighbor-joining phylogenetic tree showing Agt of strain OB3b, together with characterized GT4-group glycosyltransferases, including Agt from *Pelagibacter ubique* (19, 23), *Agrobacterium tumefaciens* (21), and *Pseudomonas aeruginosa* (17). MshA involved in mycothiol biosynthesis is used as an outgroup (51, 52).

The *plcP-agt* mutant had significant growth defect in phosphorus limitation. To determine the role of lipid remodeling in the growth of *M. trichosporium* OB3b on methane, we cultivated the wild-type strain and the mutant in the modified NMS medium supplied with high (200 μM) and low (5 μM) phosphate and monitored the growth of both strains on methane as the sole carbon source. The data presented in Fig. 4A clearly demonstrate that the ability for lipid remodeling is important for the bacterium to better adapt to phosphate limitation, and no growth defect was observed when the cells were cultivated at higher phosphate concentrations. While no difference in methane oxidation rate was observed for the wild type and the mutant cultivated in the modified NMS medium supplied with high phosphate, the mutant had significantly reduced methane oxidation rate under low-phosphate conditions (Fig. 4B and C). Interestingly, thin-section transmission electron microscopy (TEM) showed that the intracytoplasmic membrane of the mutant cultivated under low-phosphate conditions displayed a noticeable defect in membrane packing, and larger gaps in intracytoplasmic membranes were observed (Fig. 4D), suggesting that the activity of the membrane-bound particulate methane monooxygenase (pMMO) is likely to be affected in the mutant.

Wide occurrence of lipid remodeling pathway in methanotrophs. Having established that PlcP-mediated lipid remodeling is important for this methanotrophic bacterium to better adapt to low-phosphate growth conditions, we used comparative genomics to better understand the distribution of lipid remodeling in genome-sequenced aerobic methanotrophs. We built a profile hidden Markov model using PlcP sequences from characterized bacteria and searched genomes of methanotrophs in the NCBI and JGI-IMG

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

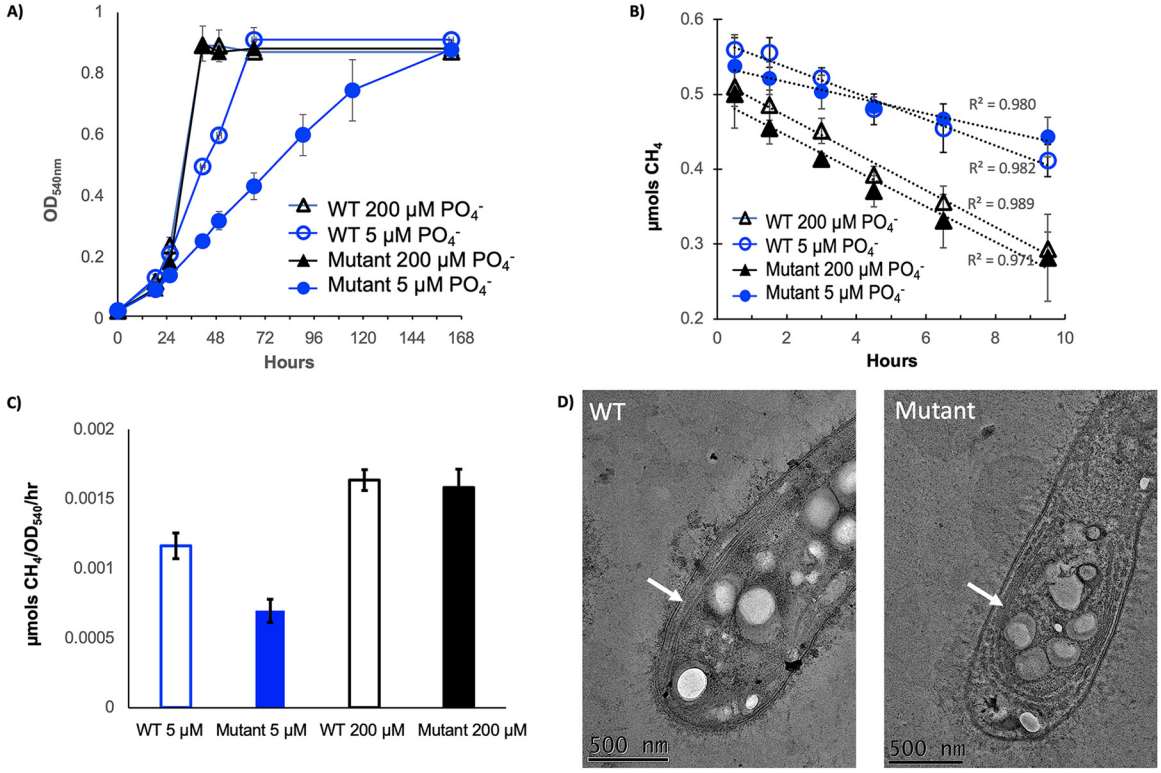


FIG 4 The *plcP-agt* mutant had significant growth defect when cultivated in low-phosphate medium. (A) Growth of the wild type and the mutant under P-replete and P-deplete conditions. All cultures were grown in P-replete medium with 4 mM phosphate and washed once in the medium with no phosphate before inoculating into either P-deplete (5 μM phosphate) or P-replete medium (200 μM phosphate). (B and C) Methane oxidation kinetics (B) and oxidation rates (C) of wild-type and mutant cultivated under P-replete and P-deplete conditions. (D) Thin-section TEM showed significant changes in intracytoplasmic membrane (indicated by white arrows).

databases (as of November 2021). We mapped out the presence or absence of PlcP in methanotrophs on a phylogenomics tree established using 140 aerobic methanotroph genomes from the RefSeq database and three metagenome assembled genomes of the so-called high-affinity atmospheric methane oxidizers of the Upland Soil Cluster α (USCα) and USCγ. Our data suggest that PlcP appears to be ubiquitous in all proteobacterial methanotrophs, including both type I and type II, whereas the ability to produce ornithine lipids using the OlsB/A pathway is found in some but not all methanotrophs (Fig. 5). Interestingly, PlcP is not found in verrucomicrobial methanotrophs, acidophiles that are isolated from hot and acidic geothermal habitats such as volcano muds (5). In all type II methanotrophs and the type Ib clade, *plcP* is usually located immediately upstream of *agt*, whereas the gene neighborhood of *plcP* in other type I methanotrophs varied significantly (Fig. 6). Another interesting observation is that in the genomes of the USCα and USCγ (25, 26), a similar *plcP-agt* gene cluster was also found, suggesting that these so-called high-affinity methanotrophs may also be capable of lipid remodeling in response to P limitation in the soil.

Identification of methanotroph *plcP* transcripts in metatranscriptomics data sets. To gain further insights into *plcP*-mediated lipid remodeling in environmental samples and to determine whether *plcP* transcripts of methanotrophs can be readily detectable, we analyzed recently published metatranscriptomics data sets from Lake Washington (27). This data set contains comprehensive metatranscriptomics sequencing of microbial communities from the lake sediments and how they responded to oxygen tension in a 14-week period. Furthermore, genome assembly was available, allowing exploitation of gene neighborhood for synteny. We detected 209 PlcP homologs in this data set (Fig. 7), the majority of which were classified as *Methylobacter tundripaludum* (75%) which is known to be the dominant methanotroph in the lake sediment (27). In scaffolds that were assembled from these metatranscriptomics data sets,

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

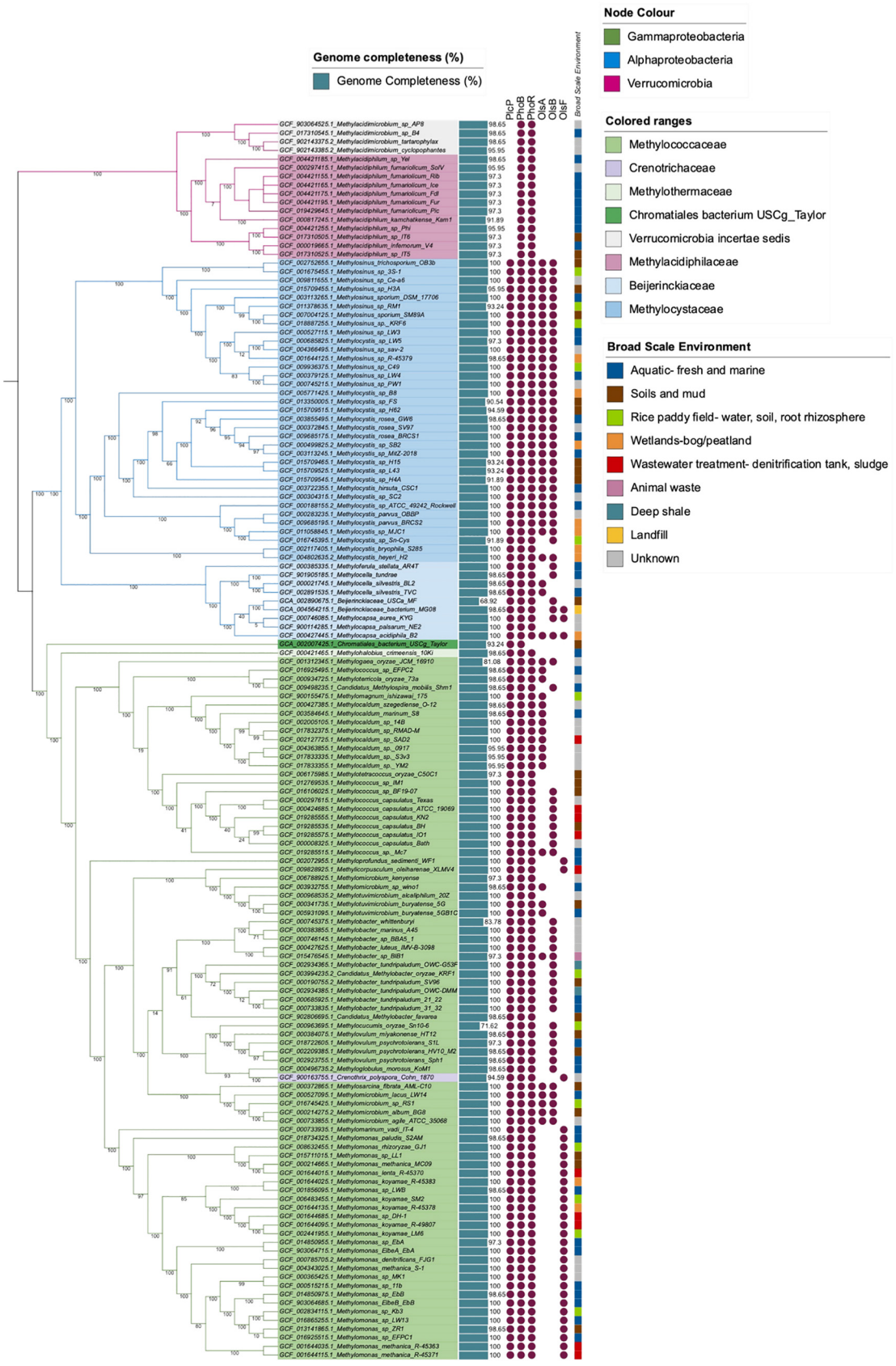


FIG 5 Analysis of lipid remodeling in methanotroph genomes. A phylogenomic tree shows 140 methanotroph genomes and 3 MAGs mapped to the presence or absence of genes involved in *plcP*-mediated lipid remodeling. The tree nodes are colored

(Continued on next page)

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

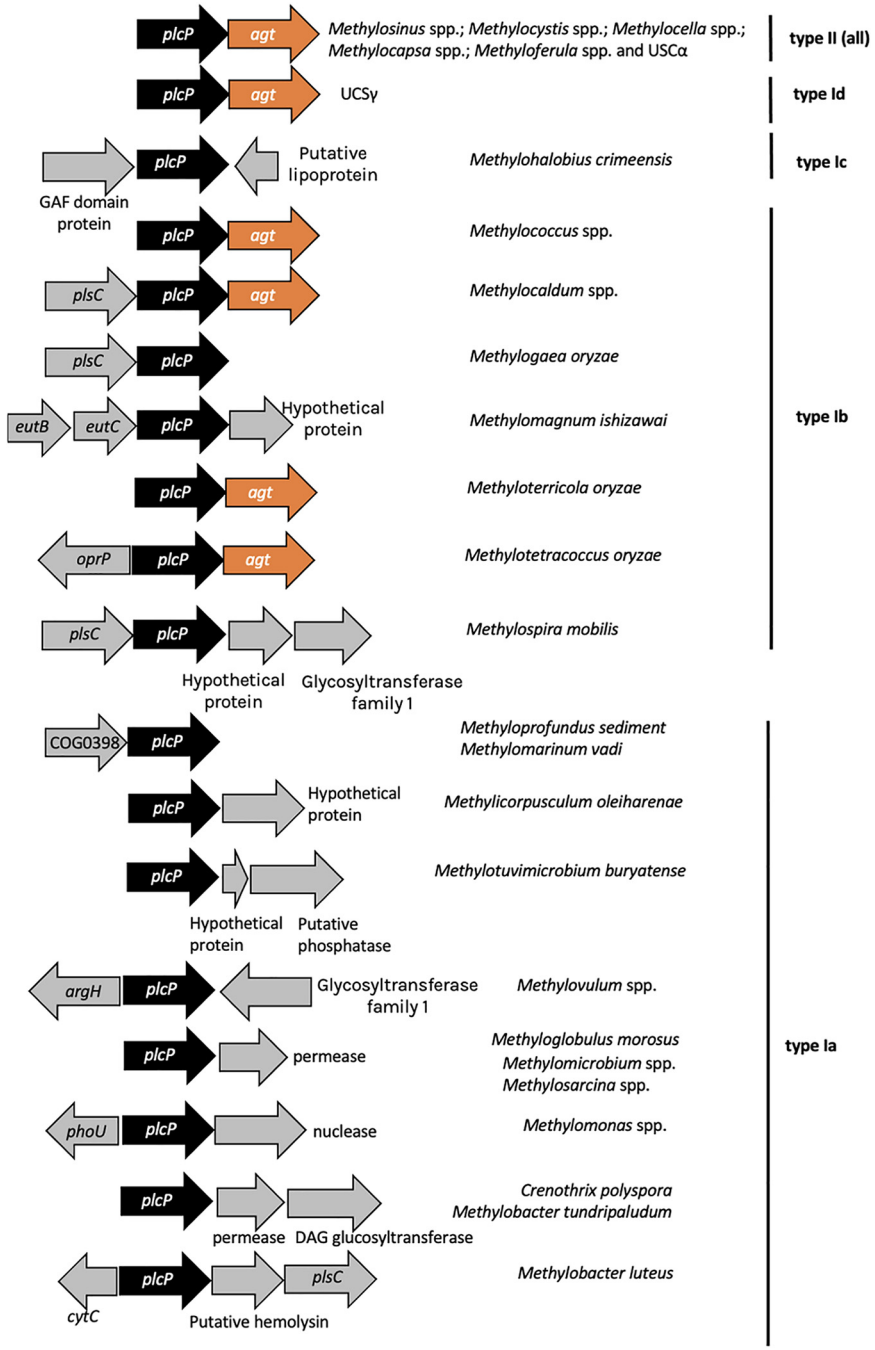


FIG 6 Comparative genomics of PlcP-pathway in methanotrophs showing the neighborhood of *plcP* in type I and type II proteobacterial methanotrophs, including high-affinity methane oxidizers in MAGs.

plcP retrieved from these environmental samples showed conserved gene synteny to that of the genome of *Methylobacter tundripaludum* and a DAG glucosyltransferase involved in glycolipid production is also found in the environmental scaffold (28). In addition to *Methylobacter* spp., several PlcP homologs were classified as from methanotrophs (e.g., *Methylomicrobium*, *Methylomonas*, or an uncultured *Methylococcaceae* bacte-

FIG 5 Legend (Continued)
according to phylum or class (if *Proteobacteria*), the label color ranges indicate bacterial families, and the genome completeness (%) is shown as a blue bar. The broad-scale environment is displayed as a color strip. A purple dot represents a candidate gene has been detected. Bootstrap (1,000 bootstrap alignments applied) values are displayed on the nodes as percentages.

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

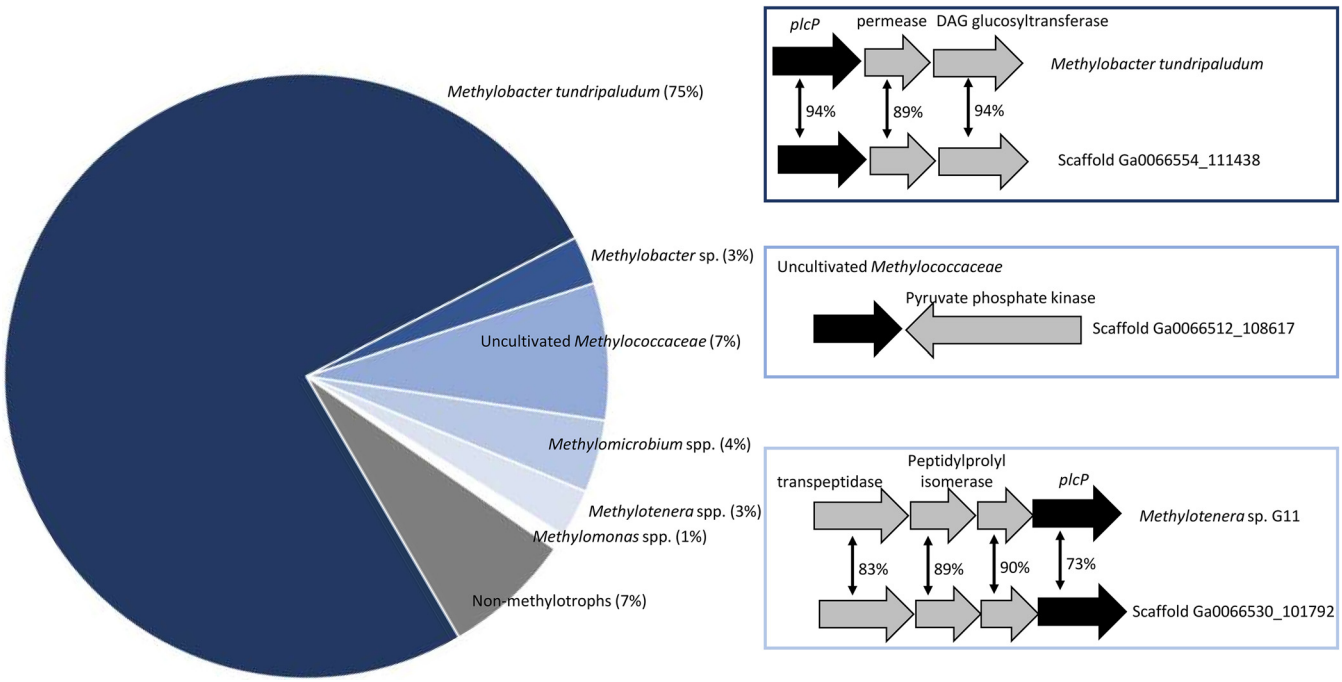


FIG 7 Analysis of metatranscriptomic data sets from Lake Washington sediments. The majority of the PlcP homologs were classified as being from methanotrophs, although PlcP from other bacteria were also found, including the methylotroph *Methylotenera* (3%) and other heterotrophic bacteria (7%). The gene neighborhoods for *plcP* assembled from metatranscriptomics data were analyzed and compared to the closest homologs in genome sequenced bacteria. Bidirectional arrows show protein sequence identity.

rium). Although many of these *plcP* genes were not assembled into contigs, the *plcP* of an uncultured *Methylococcaceae* bacterium was found in Scaffold Ga0066512_108617, and the two genes in this scaffold showed 90 and 92% identity, respectively, to that of an uncultivated *Methylococcaceae* strain from a drinking water metagenome (accession numbers NOU12518.1 and NOU12800.1). Furthermore, *plcP* transcripts from the methylotroph *Methylotenera*, which is also known to be abundant in these lake sediments (29), were detected (Fig. 7).

DISCUSSION

Here, we demonstrate for the first time that methanotrophs can modulate their membrane lipid compositions in response to the changes of phosphate levels. Formation of new glycolipids appears crucial for the activity of methanotrophs since the mutant lacking the lipid remodeling pathway grew much slower during phosphate limitation. Interestingly, this PlcP-mediated lipid remodeling pathway appears common in proteobacterial methanotrophs of both type I and type II clades but not in the verrucomicrobial clade acidophilic methanotrophs (5).

Inspection of the gene neighborhood provides some interesting insights into the differentiation of proteobacterial methanotrophs. All type II methanotrophs, including *Methylosinus* and *Methylocystis* of the *Methylocystaceae* family and *Methylocella*, *Methylocapsa*, and *Methyloferula* of the *Beijerinckiaceae* family have the glycosyltransferase *agt* gene immediately downstream of the *plcP* gene, suggesting that these type II methanotrophs likely replace membrane phospholipids with glycolipids during phosphate limitation. The same gene organization was also found in the genome sequence of the upland soil clusters USCα and USCγ (25, 26), which can oxidize methane at atmospheric concentrations (~1.8 ppm). However, in type I methanotrophs, the gene neighborhood of *plcP* varied considerably. Although *agt* is commonly located next to *plcP* in *Methylococcus* and *Methylocaldum* spp., no *agt* was found immediately downstream or upstream of type Ia methanotrophs. Instead, in several genomes of type Ia methanotrophs (e.g., *Crenothrix* and *Methylobacter tundripaludum*), another putative

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

DAG glycosyltransferase is found. Interestingly, *plcP* transcripts from *Methylobacter tundripaludum* inhabiting lake sediments can be readily detectable, and its gene neighborhood is highly conserved (Fig. 7). In *Bacillus subtilis*, this DAG glycosyltransferase is involved in the production of a series of glycolipids consisting of mono-, di-, tri-, and tetraglycosyldiacylglycerol (28). However, whether these rather complex lipids can be produced in methanotrophs in the natural environment awaits further experimental validation.

In addition to the glycolipids, strain OB3b also produced an ornithine-containing aminolipid in response to phosphate limitation (see Fig. S3). In proteobacteria, ornithine lipid biosynthesis is carried out by either OlsA/OlsB, encoding a *O*-acyltransferase and a *N*-acyltransferase, respectively, or a bifunctional acyltransferase OlsF (20, 30). The genome of strain OB3b appears to have the OlsA (locus tag Ga0263880_11897)/OlsB (locus tag Ga0263880_112024) pathway for ornithine lipid biosynthesis, although their role in ornithine lipid biosynthesis awaits further experimental validation. Our data suggest that an ornithine lipid is overproduced in P limitation in strain OB3b suggesting that it is involved in substituting membrane phospholipids during P starvation. However, the ecophysiological role of this aminolipid in the membrane of this methanotroph remains to be established.

Many previous studies have focused on the regulation of methanotroph activities in the environment, and it has been firmly established that, among many intrinsically interlinked abiotic factors, substrate availability for methanotrophs, i.e., methane and oxygen, is one of the most influential factors for controlling methanotroph populations in natural ecosystems (see, for example, references 4 and 27). A comprehensive metatranscriptomics analysis of the microbial populations involved in methane cycling in Lake Washington provided new insight into how different methanotroph ecotypes (*Methylobacter* versus *Methylosarcina*) respond to oxygen levels (27, 31). However, studies on how methane oxidation is affected by macronutrients, particularly P, are not always consistent. While many studies have shown that methane oxidation potential can be significantly enhanced by phosphate addition (reviewed in reference 8), others showed no impact of P or indeed an opposite trend, which can be at least partially explained by the fact that methanotrophs were outcompeted by cohabiting microbes (11). To this end, we propose that whether P addition will enhance methanotroph activities in environmental samples will likely be driven by whether or not methanotrophs are truly P limited in the natural environment. Given that *plcP* is conserved in all proteobacterial methanotrophs and that *plcP*-mediated lipid remodeling is induced by P limitation, we propose that *plcP* transcription is likely a reliable biomarker to determine whether P limitation occurs in proteobacterial methanotroph populations in the natural environment. As such, future experiments aiming to link P and methane oxidation activities can benefit from a thorough investigation of *plcP* transcription in methanotrophs.

In summary, we show here that the model proteobacterial methanotroph *M. trichosporium* OB3b can produce glycolipids using the *plcP*-mediated lipid remodeling pathway, which appears to be important for the growth of the bacterium in response to P limitation. The wide occurrence of *plcP* in proteobacterial methanotrophs and the readiness for detection of *plcP* transcripts in metatranscriptomics beg further investigation of its role in methanotroph ecophysiology in the natural environment.

MATERIALS AND METHODS

Cultivation of *M. trichosporium* OB3b and its mutant. *M. trichosporium* was typically cultivated in the NMS medium (32) supplied with 4 mM phosphate (Na_2HPO_4-KH_2PO_4) as the buffer. To investigate the impact of phosphate concentration on the growth of the bacterium, a modified NMS medium was used where 10 mM HEPES buffer (pH 6.8) was used to replace the phosphate buffer. Phosphate was added to a range of concentrations from 5 μM to 4 mM, and copper was added to a final concentration of 5 μM. Cell growth was monitored by recording the optical density at 540 nm (OD_{540}). All growth experiments were set up in triplicate using 250-mL QuickFit flasks containing 50 mL of medium, with an inoculum size of 1 to 2% (vol/vol). The flasks were sealed with Suba seals, with methane added to the headspace in a ratio of 1:5 (methane:air) incubated at 30°C in a shaker (150 rpm).

Generation of a *plcP-agt* knockout mutant. The Δ*plcP-agt* mutant was created through homologous recombination using the suicide plasmid pK18mobSacB (33). Two homologous regions flanking

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

the upstream and downstream of the *plcP-agt* gene cluster and a gentamicin (Gm)-resistant gene cassette originated from plasmid p34S-Gm (34) were amplified by PCR and cloned into the pK18mobSacB plasmid using Gibson assembly. The resulting plasmid was mobilized to *M. trichosporium* OB3b through conjugation using *Escherichia coli* S17.1 λ*pir* as the donor. Transconjugants were selected for on NMS agar plates supplemented with Gm at 2.5 μg/mL, and the mutant was confirmed by PCR amplification and Sanger sequencing. PCR primers used for Gibson cloning and validation of the mutant are presented in Table S1 in the supplemental material.

Complementation of the *plcP-agt* mutant. The *plcP-agt* mutant was complemented by Gibson cloning of a 2,250-bp fragment containing the *plcP-agt* from strain OB3b into the vector pBBR1 carrying kanamycin (Km) resistance (35). This plasmid was mobilized into the existing *plcP-agt* mutant through conjugation as described previously, and transconjugants were selected on NMS plates containing Km (12.5 μg/mL). The complementation was confirmed by PCR amplification using primers plcP-F and Agt R, along with Agt F comp and pK18KmF primers (see Table S1).

Methane oxidation kinetics. Wild-type and mutant OB3b (30 mL, OD_{540} ~ 0.5) from P-replete (200 μM phosphate) and P-deplete (5 μM phosphate) medium were harvested from exponentially grown cultures by centrifugation, washed once, and resuspended in 10 mL of the same medium. Resuspended cultures were then put in a 125- mL serum vial, and 0.01% (vol/vol) methane was added. The sample was then left at room temperate (~25°C) to measure methane consumption by gas chromatography (Agilent 6890) with a flame ionization detector fitted with a silica capillary column, as described previously (36). Three biological replicates were used for both wild-type and mutant, and the methane concentrations in the headspace of the serum vials were measured at 0.5, 1.5, 3, 4.5, 6.5, and 9.5 h. Each measurement was carried out twice by injecting 200 μL of gas using a 500-μL Hamilton syringe. Methane oxidation rate was calculated using linear regression analysis in Excel.

Lipid extraction and lipidomic analyses. Lipid extraction from *M. trichosporium* OB3b and the Δ*plcP-agt* mutant was carried out in three replicates using a modified Folch extraction method, as described previously (37). Briefly, 1 mL of bacterial culture (OD_{540} = 0.5) was pelleted by centrifugation, and chloroform (1 mL), Milli-Q water (0.3 mL), and methanol (0.5 mL) were then added. The chloroform phase containing the lipids was dried under nitrogen before resuspension in 0.5 to 1 mL of solvent (0.05 mL of 10 mM ammonium acetate in water [pH 9.2] and 0.95 mL acetonitrile). The lipid standard d17:1/12:0 sphingosylphosphoethanolamine (SPE; Avanti Polar Lipids) was added to the samples to a final concentration of 500 nM. Bacterial lipids were separated by a Dionex 3400RS HPLC using an XBridge BEH amide XP column (2.5 μm 3.0 × 150 mm; Waters) on a 15-min gradient from 95% (vol/vol) acetonitrile/5% (wt/vol) ammonium acetate (10 mM [pH 9.2]) to 70% (vol/vol) acetonitrile/30% (wt/vol) ammonium acetate (10 mM [pH 9.2]). The flow rate was 150 μL min^{-1}, and the column temperature was 30°C. Ionization and MS fragmentation of lipids were carried out in both positive and negative modes. The drying conditions were 8 L min^{-1} drying gas at 300°C, nebulizing gas pressure of 15 lb/in^2, and the end cap voltage was 4,500 V in the positive mode and 3,500 V in the negative mode, both with a 500-V offset. Data analyses were carried out using the Bruker Compass Software with DataAnalysis for peak identification and lipid MS^n fragmentation and QuantAnalysis for lipid quantification against the internal standard SPE. The retention times of phospholipids PE (~12.2 min) and PG (~6.1 min) and the glycolipid MGDG (~7.1 min) from strain OB3b were confirmed by running authentic lipid standards obtained from Avanti Polar Lipids, Inc. ($C_{34:0}$ PG, 830456P; $C_{34:0}$ PE, 830756P; $C_{34:1}$ MGDG 840522P). The identity of GADG lipid in strain OB3b was further validated by high-resolution MS^n using an Orbitrap mass spectrometer (see Fig. S6).

Transmission electron microscopy. Cells of the exponentially growing cultures were collected by centrifugation and fixed with 1% (wt/vol) glutaraldehyde in the NMS medium for 1 h at 4°C. Bacterial pellets were then postfixed with 1% (wt/vol) osmium tetroxide (OsO_4) in water for 1 h at 4°C. Samples were dehydrated using an increasing ethanol concentration (50 to 100% [vol/vol]) before embedding into low-viscosity resin (Agar Scientific, UK). Ultrathin sections (70 nm) were stained with 2% (wt/vol) uranyl acetate, followed by 3% (wt/vol) lead citrate. The wild-type specimen and the mutant samples cultivated in high and low phosphate were then examined with a transmission electron microscope (Jeol, Tokyo, Japan) operating at 200 kV, equipped with a Gatan OneView IS detector (Gatan Ametek, USA) at the University of Warwick Advanced Bioimaging Research Technology Platform.

Genomic data acquisition. A total of 143 known aerobic methanotroph genomes belonging to phyla *Proteobacteria* and *Verrucomicrobia*, available from the National Centre for Biotechnology Information (NCBI) database, were used in this study and included 140 genomes from the RefSeq genome database (38), and three metagenome-assembled genomes (MAGs) representing atmospheric methane oxidizers; *Beijerinckiaceae* bacterium USCα_MF (GCA_002890675.1) (25) and *Chromatiales* bacterium USCg_Taylor (GCA_002007425.1) (26) from the GenBank assembly database (the genome accession numbers are available in Table S2). Metadata from the NCBI was used to identify the source environment and was categorized into broad-scale environments (i.e., aquatic [both fresh and marine], soils and mud, rice paddies, wetlands, wastewater treatment, animal waste, deep shale, and unknown [genomes with no information on the environment]).

Identification of genes and gene-specific phylogenetic trees. Genomes were downloaded and concatenated to generate a local methanotroph genomes database using DIAMOND (v0.9.14.115) (39). Validated protein sequences (i.e., *plcP* [WP_024829183.1], lyso-ornithine lipid:acyl-ACP *O*-acyltransferase [*olsA*; AGB69899.1], ornithine:acyl-ACP *N*-acyltransferase [*olsB*; AGB69885.1], bifunctional acyltransferase [*olsF*, A0A6N6NHX3], phosphate response regulator transcription factor [*phoB*; HAI5817470.1], and phosphate regulon sensor protein [*phoR*; P08400]) were used as query sequence to perform Basic Local Alignment Search Tool (BLASTX) searches (1e–10) to detect the presence of genes of interest (39). Homologs retrieved from BLASTX results were translated to amino acid sequences and used in

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

phylogenetic tree construction. Amino acid sequences from methanotroph genomes alongside with other related homologs were used to perform multiple sequence alignment using ClustalO (v 1.2.3) (40). Maximum-likelihood phylogenetic trees were constructed from the multiple sequence alignments using the IQ-Tree2 (v2.1.2) (41) based on the Jones-Taylor-Thornton (JTT) matrix-based model. Bootstrap analysis was performed with 1,000 replicates to provide confidence estimates for phylogenetic tree topologies. Interactive tree of life (ITOL V4) was used for tree visualization annotation with labels colored via taxonomy at the family level (42).

Phylogenomic analysis. Phylogenomic relatedness of the methanotroph genomes and MAGs was inferred using 74 single-copy marker genes (specific to the bacterial domain) via the GToTree (v1.5.52) analysis pipeline (43). For each set of protein sequences retrieved using the HMMER3 tool (v3.3.2) using predefined model cutoffs (44), multiple protein sequence alignments were produced using MUSCLE (v3.8.1551) with default settings (45). Automated trimming was performed on the alignments using trimAL (v1.4.rev15) (46). A phylogenetic tree was generated by the concatenated trimmed alignment using FastTree2 (v2.1.10) (47). Taxonomy was assigned using TaxonKit (48). The output newick tree file from GToTree was uploaded to the iTOL platform for further annotation (42). Gene presence-absence data were mapped to the phylogenomic tree to visually represent the detection of *plcP*, *phoB/phoR*, *olsA*, *olsB*, and *olsF* genes in the genomes.

Metatranscriptomics. Assembled metatranscriptomes of Lake Washington microbial communities (under the study name Freshwater Sediment Methanotrophic Microbial Communities from Lake Washington under Simulated Oxygen Tension) (27) were analyzed for the presence of PlcP using BLASTP searches carried out at the JGI-IMG site (https://img.jgi.doe.gov). BLASTP searches were carried out with the PlcP of strain OB3b as the query sequence with an e value cutoff of 1e–40. These retrieved 209 PlcP homologs sequences. To further classify the phylogenies of these environmental PlcP homologous, they were then downloaded from IMG, and their identities were classified using a BLASTP search against the NCBI nonredundant protein sequences database (nr). The gene neighborhood of *plcP* from each scaffold was manually inspected using the IMG "gene neighborhoods" function.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, DOCX file, 0.1 MB.

FIG S2, DOCX file, 0.4 MB.

FIG S3, DOCX file, 0.1 MB.

FIG S4, DOCX file, 0.1 MB.

FIG S5, DOCX file, 1.5 MB.

FIG S6, DOCX file, 0.3 MB.

TABLE S1, DOCX file, 0.02 MB.

TABLE S2, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We are grateful for a joint Royal Society/National Natural Science Foundation of China (RS-NSFC) Newton Advanced Fellowship Award (NAF/R1/180191, 41861130357). J.S., R.G., and M.A.M. were funded by a European Research Council (ERC) award under the European Union's Horizon 2020 research and innovation program (grant agreement 726116). M.R.C. was funded by a Department for Economy PhD studentship.

We thank E. Silvano for help with liquid chromatography-mass spectrometry, the Advanced Bioimaging Research Technology Platform for training on TEM to R.G., and the Proteomics Research Technology Platform at the University of Warwick for assistance with the Orbitrap Fusion mass spectrometer. We acknowledge the Midlands Regional Cryo-EM Facility, hosted at the Warwick Advanced Bioimaging Research Technology Platform, for the use of the JEOL 2100Plus, supported by a Medical Research Council award (MC_PC_17136).

We declare that we have no conflicts of interest.

REFERENCES

1. IPCC. 2021. AR6 Climate Change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, Geneva, Switzerland. https://www.ipcc.ch/report/ar6/wg1/.
2. Jackson RB, Saunois M, Bousquet P, Canadell JG, Poulter B, Stavert AR, Bergamaschi P, Niwa Y, Segers A, Tsuruta A. 2020. Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources. Environ Res Lett 15:e071002. https://doi.org/10.1088/1748-9326/ab9ed2.
3. Knief C. 2015. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. Front Microbiol 6:1346. https://doi.org/10.3389/fmicb.2015.01346.
4. Bodelier PL, Pérez G, Veraart AJ, Krause SM. 2019. Methanotroph ecology, environmental distribution, and functioning. *In* Methanotrophs, p 1–38. Springer, Cham, Switzerland.
5. Schmitz RA, Peeters SH, Versantvoort W, Picone N, Pol A, Jetten MSM, Op den Camp HJM. 2021. Verrucomicrobial methanotrophs: ecophysiology

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

of metabolically versatile acidophiles. FEMS Microbiol Rev 45:fuab007. https://doi.org/10.1093/femsre/fuab007.
6. Pandit PS, Hoppert M, Rahalkar MC. 2018. Description of '*Candidatus* Methylocucumis oryzae', a novel type I methanotroph with large cells and pale pink colour, isolated from an Indian rice field. Antonie Van Leeuwenhoek 111:2473–2484. https://doi.org/10.1007/s10482-018-1136-3.
7. Bodelier PLE, Gillisen MJB, Hordijk K, Damste JSS, Rijpstra WIC, Geenevasen JA, Dunfield PF. 2009. A reanalysis of phospholipid fatty acids as ecological biomarkers for methanotrophic bacteria. ISME J 3:606–617. https://doi.org/10.1038/ismej.2009.6.
8. Veraart AJ, Steenbergh AK, Ho A, Kim SY, Bodelier PLE. 2015. Beyond nitrogen: the importance of phosphorus for CH_4 oxidation in soils and sediments. Geoderma 259-260:337–346. https://doi.org/10.1016/j.geoderma.2015.03.025.
9. Zhang T, Zhu W, Mo J, Liu L, Dong S. 2011. Increased phosphorus availability mitigates the inhibition of nitrogen deposition on CH_4 uptake in an old-growth tropical forest, southern China. Biogeosciences 8:2805–2813. https://doi.org/10.5194/bg-8-2805-2011.
10. Lu Y, Wassmann R, Neue H, Huang C. 1999. Impact of phosphorus supply on root exudation, aerenchyma formation and methane emission of rice plants. Biogeochemistry 47:203–218. https://doi.org/10.1007/BF00994923.
11. Alam MS, Xia W, Jia Z. 2014. Methane and ammonia oxidations Interact in paddy soils. Int J Agric Biol 16:365–370.
12. Zheng Y, Zhang L-M, He J-Z. 2013. Immediate effects of nitrogen, phosphorus, and potassium amendments on the methanotrophic activity and abundance in a Chinese paddy soil under short-term incubation experiment. J Soils Sediments 13:189–196. https://doi.org/10.1007/s11368-012-0601-2.
13. Merchant SS, Helmann JD. 2012. Elemental economy: microbial strategies for optimizing growth in the face of nutrient limitation. Adv Microb Physiol 60:91–210. https://doi.org/10.1016/B978-0-12-398264-3.00002-4.
14. Stein LY, Yoon S, Semrau JD, Dispirito AA, Crombie A, Murrell JC, Vuilleumier S, Kalyuzhnaya MG, Op den Camp HJM, Bringel F, Bruce D, Cheng J-F, Copeland A, Goodwin L, Han S, Hauser L, Jetten MSM, Lajus A, Land ML, Lapidus A, Lucas S, Médigue C, Pitluck S, Woyke T, Zeytun A, Klotz MG. 2010. Genome sequence of the obligate methanotroph *Methylosinus trichosporium* strain OB3b. J Bacteriol 192:6497–6498. https://doi.org/10.1128/JB.01144-10.
15. Weaver TJ, Patrick MA, Dugan PR. 1975. Whole-cell and membrane lipids of the methylotrophic bacterium *Methylosinus trichosporium*. J Bacteriol 124:602–605. https://doi.org/10.1128/jb.124.2.602-605.1975.
16. Fang J, Barcelona MJ, Semrau JD. 2000. Characterization of methanotrophic bacteria on the basis of intact phospholipid profiles. FEMS Microbiol Lett 189:67–72. https://doi.org/10.1111/j.1574-6968.2000.tb09207.x.
17. Jones RA, Shropshire H, Zhao C, Murphy A, Lidbury I, Wei T, Scanlan DJ, Chen Y. 2021. Phosphorus stress induces the synthesis of novel glycolipids in *Pseudomonas aeruginosa* that confer protection against a last-resort antibiotic. ISME J 15:3303–3314. https://doi.org/10.1038/s41396-021-01008-7.
18. Sebastián M, Smith AF, González JM, Fredricks HF, Van Mooy B, Koblížek M, Brandsma J, Koster G, Mestre M, Mostajir B, Pitta P, Postle AD, Sánchez P, Gasol JM, Scanlan DJ, Chen Y. 2016. Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. ISME J 10:968–978. https://doi.org/10.1038/ismej.2015.172.
19. Carini P, Van Mooy BA, Thrash JC, White A, Zhao Y, Campbell EO, Fredricks HF, Giovannoni SJ. 2015. SAR11 lipid renovation in response to phosphate starvation. Proc Natl Acad Sci U S A 112:7767–7772. https://doi.org/10.1073/pnas.1505034112.
20. Vences-Guzmán MÁ, Geiger O, Sohlenkamp C. 2012. Ornithine lipids and their structural modifications: from A to E and beyond. FEMS Microbiol Lett 335:1–10. https://doi.org/10.1111/j.1574-6968.2012.02623.x.
21. Semeniuk A, Sohlenkamp C, Duda K, Hölzl G. 2014. A bifunctional glycosyltransferase from *Agrobacterium tumefaciens* synthesizes monoglucosyl and glucuronosyl diacylglycerol under phosphate deprivation. J Biol Chem 289:10104–10114. https://doi.org/10.1074/jbc.M113.519298.
22. Wei T, Quareshy M, Zhang YZ, Scanlan DJ, Chen Y. 2018. Manganese is essential for PlcP metallophosphoesterase activity involved in lipid remodeling in abundant marine heterotrophic bacteria. Appl Environ Microbiol 84:e01109-18. https://doi.org/10.1128/AEM.01109-18.
23. Wei T, Zhao C, Quareshy M, Wu N, Huang S, Zhao Y, Yang P, Mao D, Chen Y. 2021. A glycolipid glycosyltransferase with broad substrate specificity from the marine bacterium "*Candidatus* Pelagibacter sp." strain HTCC7211. Appl Environ Microbiol 87:e0032621. https://doi.org/10.1128/AEM.00326-21.
24. Zavaleta-Pastor M, Sohlenkamp C, Gao JL, Guan Z, Zaheer R, Finan TM, Raetz CR, López-Lara IM, Geiger O. 2010. *Sinorhizobium meliloti* phospholipase C required for lipid remodeling during phosphorus limitation. Proc Natl Acad Sci U S A 107:302–307. https://doi.org/10.1073/pnas.0912930107.
25. Pratscher J, Vollmers J, Wiegand S, Dumont MG, Kaster AK. 2018. Unravelling the identity, metabolic potential and global biogeography of the atmospheric methane-oxidizing upland soil cluster alpha. Environ Microbiol 20:1016–1029. https://doi.org/10.1111/1462-2920.14036.
26. Edwards CR, Onstott TC, Miller JM, Wiggins JB, Wang W, Lee CK, Cary SC, Pointing SB, Lau MCY. 2017. Draft genome sequence of uncultured upland soil cluster *Gammaproteobacteria* gives molecular insights into high-affinity methanotrophy. Genome Announc 5:e00047-17. https://doi.org/10.1128/genomeA.00047-17.
27. Zheng Y, Wang H, Yu Z, Haroon F, Hernandez ME, Chistoserdova L. 2020. Metagenomic insight into environmentally challenged methane-fed microbial communities. Microorganisms 8:1614. https://doi.org/10.3390/microorganisms8101614.
28. Jorasch P, Wolter FP, Zähringer U, Heinz E. 1998. A UDP glucosyltransferase from *Bacillus subtilis* successively transfers up to four glucose residues to 1,2-diacylglycerol: expression of *ypfP* in *Escherichia coli* and structural analysis of its reaction products. Mol Microbiol 29:419–430. https://doi.org/10.1046/j.1365-2958.1998.00930.x.
29. Chistoserdova L. 2011. Methylotrophy in a lake: from metagenomics to single-organism physiology. Appl Environ Microbiol 77:4705–4711. https://doi.org/10.1128/AEM.00314-11.
30. Vences-Guzmán MÁ, Guan Z, Escobedo-Hinojosa WI, Bermúdez-Barrientos JR, Geiger O, Sohlenkamp C. 2015. Discovery of a bifunctional acyltransferase responsible for ornithine lipid synthesis in *Serratia proteamaculans*. Environ Microbiol 17:1487–1496. https://doi.org/10.1111/1462-2920.12562.
31. Hernandez ME, Beck DA, Lidstrom ME, Chistoserdova L. 2015. Oxygen availability is a major factor in determining the composition of microbial communities involved in methane oxidation. PeerJ 3:e801. https://doi.org/10.7717/peerj.801.
32. Whittenbury R, Phillips KC, Wilkinson JF. 1970. Enrichment, isolation and some properties of methane-utilizing bacteria. J Gen Microbiol 61:205–218. https://doi.org/10.1099/00221287-61-2-205.
33. Schäfer A, Tauch A, Jager W, Kalinowski J, Thierbach G, Puhler A. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145:69–73. https://doi.org/10.1016/0378-1119(94)90324-7.
34. Dennis JJ, Zylstra GJ. 1998. Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Appl Environ Microbiol 64:2710–2715. https://doi.org/10.1128/AEM.64.7.2710-2715.1998.
35. Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop IIR, Peterson KM. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1 MCS carrying different antibiotic- resistance cassettes. Gene 166:175–176. https://doi.org/10.1016/0378-1119(95)00584-1.
36. Chen Y, Dumont MG, McNamara NP, Chamberlain PM, Bodrossy L, Stralis-Pavese N, Murrell JC. 2008. Diversity of the active methanotrophic community in acidic peatlands as assessed by mRNA and SIP-PLFA analyses. Environ Microbiol 10:446–459. https://doi.org/10.1111/j.1462-2920.2007.01466.x.
37. Silvano E, Yang M, Wolterink M, Giebel HA, Simon M, Scanlan DJ, Zhao Y, Chen Y. 2020. Lipidomic analysis of roseobacters of the pelagic RCA cluster and their response to phosphorus limitation. Front Microbiol 11:552135. https://doi.org/10.3389/fmicb.2020.552135.
38. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189.
39. Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366–368. https://doi.org/10.1038/s41592-021-01101-x.
40. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.
41. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
42. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
43. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164. https://doi.org/10.1093/bioinformatics/btz188.
44. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
45. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
46. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
47. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximate maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
48. Shen W, Ren H. 2021. TaxonKit: a practical and efficient NCBI taxonomy toolkit. J Genet Genomics 48:844–850. https://doi.org/10.1016/j.jgg.2021.03.006.
49. Cho J, Lee M, Cochrane CS, Webster CG, Fenton BA, Zhao J, Hong J, Zhou P. 2020. Structural basis of the UDP-diacylglucosamine pyrophosphohydrolase LpxH inhibition by sulfonyl piperazine antibiotics. Proc Natl Acad Sci U S A 117:4109–4116. https://doi.org/10.1073/pnas.1912876117.
50. Okada C, Wakabayashi H, Kobayashi M, Shinoda A, Tanaka I, Yao M. 2016. Crystal structures of the UDP-diacylglucosamine pyrophosphohydrase LpxH from *Pseudomonas aeruginosa*. Sci Rep 6:32822. https://doi.org/10.1038/srep32822.
51. Newton GL, Koledin T, Gorovitz B, Rawat M, Fahey RC, Av-Gay Y. 2003. The glycosyltransferase gene encoding the enzyme catalyzing the first step of mycothiol biosynthesis (*mshA*). J Bacteriol 185:3476–3479. https://doi.org/10.1128/JB.185.11.3476-3479.2003.
52. Vetting MW, Frantom PA, Blanchard JS. 2008. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*: substrate-assisted catalysis. J Biol Chem 283:15834–15844. https://doi.org/10.1074/jbc.M801017200.

Downloaded from https://journals.asm.org/journal/mbio on 08 June 2022 by 137.205.238.49.

Check for updates

ORIGINAL RESEARCH

Mechanistic insights into the key marine dimethylsulfoniopropionate synthesis enzyme DsyB/DSYB

Chun-Yang Li[1,2,3,#], Jason C. Crack[4,#], Simone Newton-Payne[5,#], Andrew R. J. Murphy[6], Xiu-Lan Chen[2,3], Benjamin J. Pinchbeck[5], Shun Zhou[1,5], Beth T. Williams[5], Ming Peng[2], Xiao-Hua Zhang[1], Yin Chen[6], Nick E. Le Brun[4,*], Jonathan D. Todd[1,5,*], and Yu-Zhong Zhang[1,2,3,*]

Edited by Ning-Yi Zhou, Shanghai Jiao Tong University, China; Received April 6, 2022; Accepted May 17, 2022; Published online xxx

Abstract

Marine algae and bacteria produce approximately eight billion tonnes of the organosulfur molecule dimethylsulfoniopropionate (DMSP) in Earth's surface oceans annually. DMSP is an antistress compound and, once released into the environment, a major nutrient, signaling molecule, and source of climate-active gases. The methionine transamination pathway for DMSP synthesis is used by most known DMSP-producing algae and bacteria. The *S*-directed *S*-adenosylmethionine (SAM)-dependent 4-methylthio-2-hydroxybutyrate (MTHB) *S*-methyltransferase, encoded by the *dsyB*/*DSYB* gene, is the key enzyme of this pathway, generating *S*-adenosylhomocysteine (SAH) and 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB). *DsyB*/*DSYB*, present in most haptophyte and dinoflagellate algae with the highest known intracellular DMSP concentrations, is shown to be far more abundant and transcribed in marine environments than any other known *S*-methyltransferase gene in DMSP synthesis pathways. Furthermore, we demonstrate *in vitro* activity of the bacterial DsyB enzyme from *Nisaea denitrificans* and provide its crystal structure in complex with SAM and SAH-MTHB, which together provide the first important mechanistic insights into a DMSP synthesis enzyme. Structural and mutational analyses imply that DsyB adopts a proximity and desolvation mechanism for the methyl transfer reaction. Sequence analysis suggests that this mechanism may be common to all bacterial DsyB enzymes and also, importantly, eukaryotic DSYB enzymes from e.g., algae that are the major DMSP producers in Earth's surface oceans.

Keywords: catalytic mechanism; DMSP synthesis; marine sulfur cycle; *S*-methyltransferase

Impact statement

Dimethylsulfoniopropionate (DMSP) is one of Earth's most abundant organosulfur-containing molecules which influence marine nutrient cycling, chemotaxis, atmospheric chemistry, and potentially the climate. This study provides the first structural and mechanistic understanding of the key DMSP synthesis enzyme in marine bacteria (DsyB) and algae (DSYB) that are responsible for the annual production of ~8 billion tonnes of DMSP. DsyB is the first DMSP synthesis enzyme to be analyzed at the structural and mechanistic level. This study defines the functional characteristics of the DsyB/DSYB enzyme family that has a central role in driving the marine microbial cycling of organosulfur.

[1]Frontiers Science Center for Deep Ocean Multispheres and Earth System, College of Marine Life Sciences, Ocean University of China, Qingdao, China. [2]State Key Laboratory of Microbial Technology, Marine Biotechnology Research Center, Shandong University, Qingdao, China. [3]Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology, Qingdao, Shandong, China. [4]School of Chemistry, Centre for Molecular and Structural Biochemistry, University of East Anglia, Norwich Research Park, Norwich, UK. [5]School of Biological Sciences, University of East Anglia, Norwich, UK. [6]School of Life Sciences, University of Warwick, Coventry, UK.

*** Correspondence**: Yu-Zhong Zhang, zhangyz@sdu.edu.cn; Jonathan D. Todd, Jonathan.Todd@uea.ac.uk; Nick E. Le Brun, N.Le-brun@uea.ac.uk

[#]Chun-Yang Li, Jason C. Crack, and Simone Newton-Payne contributed equally to this study.

DOI: 10.1002/mlf2.12030

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Approximately eight billion tonnes of the compatible solute dimethylsulfoniopropionate (DMSP) are produced annually in Earth's surface waters[1], constituting up to 10% of surface ocean organic carbon[2]. Many marine algae, bacteria, corals, and some plants produce DMSP[3] for its proposed functions as for example, a compatible solute[4], grazing deterrent[5], antioxidant[6], and protectant against hydrostatic pressure[7]. Furthermore, DMSP is a major nutrient for marine microorganisms, and a precursor for climate-active volatiles such as dimethyl sulfide (DMS)[3,8,9]. DMSP was thought to be mainly produced by marine algae in Earth's surface oceans, but recent studies suggest that bacteria, particularly in marine sediment, are also important DMSP producers[3,10–12].

Three pathways for DMSP synthesis have been proposed based on the identification of intermediates and enzyme activities in various model DMSP producers: a methylation pathway in some plants and bacteria, a transamination pathway in algae and bacteria, and a decarboxylation pathway in onedinoflagellate[10,12–15] (Figure 1). Of these, the transamination pathway is thought to be the most important in marine environments as it functions in the majority of DMSP-producing algae (spanning dinoflagellates, haptophytes, and diatoms) and bacteria[10,12,14]. The committed enzyme of the transamination pathway (Figure 1) is an *S*-adenosylmethionine (SAM)-dependent 4-methylthio-2-hydroxybutyrate (MTHB) *S*-methyltransferase that yields 4-dimethylsulfonio-2-hydroxybutyrate (DMSHB)[13,14,16]. Recently, the key MTHB *S*-methyltransferase enzyme 'DsyB', was identified in many DMSP-producing marine *Alphaproteobacteria*[10]. Enzymes, termed DSYB, with ~33% amino acid identity to bacterial DsyB enzymes and that have SAM-dependent MTHB*S*-methyltransferase activity are found in many eukaryotes including most DMSP-producing dinoflagellates, haptophytes, corals, and ~20% of diatoms[12]. This is consistent with the detection of DMSHB and its oxidative decarboxylation to DMSP in some prymnesiophytes, diatoms, and prasinophytes[14]. *dsyB/DSYB* genes are robust indicators of an organism's potential to produce DMSP[10,12]. The centric diatom *Thalassiosira pseudonana*, which lacks DSYB, contains an isoform MTHB *S*-methyltransferase enzyme termed TpMMT, but this

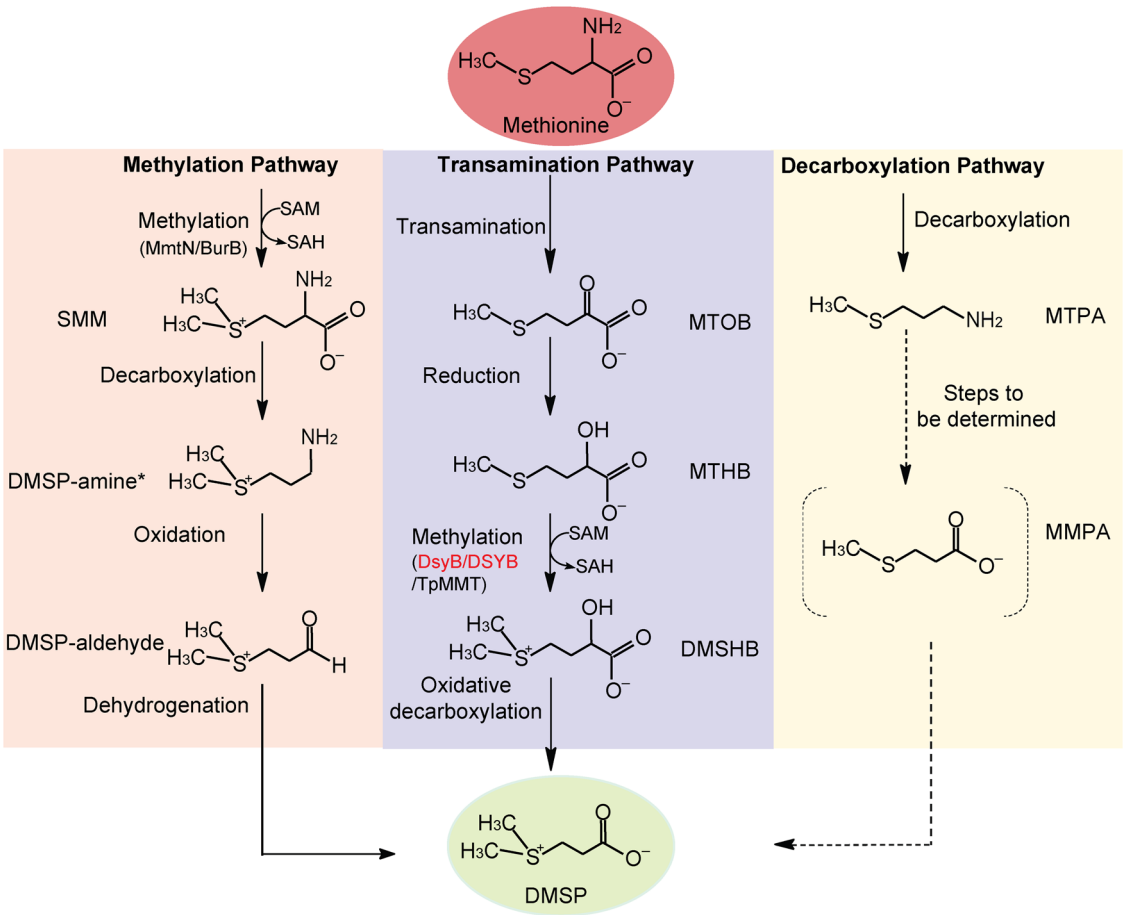


Figure 1. Predicted DMSP biosynthesis pathways[10]. Different pathways are shown in different colors. Enzymes of interest in this study (DsyB/DSYB) are in red. *SMM is converted to DMSP-aldehyde directly in *Wollastonia*. Dotted lines represent unconfirmed steps of the decarboxylation pathway. DMSHB, 4-dimethylsulfonio-2-hydroxybutyrate; DMSP, dimethylsulfoniopropionate; MMPA, methylmercaptopropionate; MTHB, 4-methylthio-2-hydroxybutyrate; MTOB, 4-methylthio-2-oxobutyrate; MTPA, 3-methylthiopropylamine; SAH, *S*-adenosylhomocysteine; SAM, *S*-adenosylmethionine; SMM, *S*-methylmethionine.

enzyme has not been studied in any other organism[17]. Published and new analysis here (see below) shows that *DsyB*/*DSYB* genes are far more abundant in known DMSP-producing microorganisms (phytoplankton and bacteria) and in the marine environmental metagenome and metatranscriptome datasets than other identified DMSP synthesis genes[10–12]. Furthermore, acquisition of *dsyB* is sufficient to enable some organisms to produce DMSP[10]. Together, these data suggest that the transamination pathway using DsyB/DSYB enzymes is the most important marine DMSP synthesis pathway.

DSYB and DsyB belong to the SAM-dependent methyltransferase (SAM-MT) family[10,12]. SAM is the second most widely used enzyme-substrate after ATP and is involved in many important biological processes[18]. SAM-MTs are categorized based on the methyl-accepting atom, usually *O*, *N*, *C*, or *S*[19]. The majority (54%) of known SAM-MTs are *O*-directed, whereas only 3% are *S*-directed[19]. SAM-MTs, which catalyze transmethylation via S_N2 nucleophilic substitution[20,21], have evolved three distinct mechanisms: proximity and desolvation (PD), general acid/base-mediated catalysis, and a metal-dependent mechanism[19]. There are no reported protein crystallographic studies on any DMSP synthesis enzyme and, thus, the mechanism of *S*-directed SAM-MT in DMSP synthesis pathways, for example, via DsyB/DSYB or TpMMT, are unknown.

Here, we investigate *Nisaea denitrificans* DR41_21, a marine *Alphaproteobacterium*[22] predicted to produce DMSP, and characterize its DsyB enzyme. X-ray crystallography and mutational analyses are employed to establish the DsyB structure and predict its interaction with SAM and MTHB substrates and reaction mechanism. Furthermore, sequence alignment and structural analysis are used to infer mechanistic similarities between bacterial DsyB and algal DSYB enzymes. We also probe marine microorganisms, metagenomes, and metatranscriptomes for the presence of DsyB/DSYB and other key SAM-MT in DMSP synthesis pathways to investigate the importance of these proteins in the global oceans. Our results provide the first insights into the mechanism of global DMSP production via the most abundant known DMSP synthesis enzymes.

RESULTS AND DISCUSSION

N. denitrificans DR41_21 is a DMSP-producing bacterium

Isolated from coastal Mediterranean Sea surface waters, *N. denitrificans* DR41_21 (DSM 18348) is a marine *Alphaproteobacterium* of the *Rhodospirillaceae* family which was not previously known to produce DMSP[22]. *N. denitrificans* contains a DsyB enzyme, 337 amino acid residues in length with 59% identity to *Labrenzia aggregata* DsyB and is thus predicted to make DMSP[10]. Indeed, cloned *N. denitrificans dsyB* conferred onto *Rhizobium*, a heterologous host that lacks DsyB and makes no DMSP, MTHB *S*-methyltransferase activity. Furthermore, *N. denitrificans dsyB* fully restored the production and accumulation of DMSP (105 ± 3.4 pmol DMSP μg protein^{-1}) of an *L. aggregata dsyB*$^-$ deletion mutant, which produces and accumulates no DMSP (wild-type *L. aggregata* produces 99.8 ± 1.2 pmol DMSP μg protein^{-1})[10]. *N. denitrificans* itself produced DMSP when grown in the absence of methylated sulfur compounds, and DMSP production and *dsyB* transcription were enhanced by increased salinity and by nitrogen starvation (Figure 2A,B). This study further confirms that the presence of *dsyB* and its transcription in a bacterium reports the ability of the strain to produce DMSP and the levels it makes, respectively.

In vitro characterization of *N. denitrificans* DsyB

As shown above and in Curson et al.[10], DsyB has MTHB *S*-methyltransferase activity when expressed in alphaproteobacterial hosts. However, *L. aggregata* DsyB and *Chrysochromulina* DSYB enzymes[12] had no detectable MTHB *S*-methyltransferase activity when expressed in *Escherichia coli*. The same was generally true of the recombinant *N. denitrificans* DsyB enzyme purified from *E. coli*, although variable MTHB *S*-methyltransferase activity was observed *in vitro* using MTHB and SAM as substrates (see Materials and Methods section). The reason for this lack of activity upon isolation is unknown; one possibility is that the enzyme requires an essential co-factor or modification that was provided by an algal or alphaproteobacterial host but not by *E. coli*[12]. This hypothesis was initially supported by the fact that the addition of heat-denatured cell lysate fractions (from a PD10 desalting column) liberated from the *L. aggregata dsyB*$^-$ deletion mutant, which produces no DMSP, recovered *N. denitrificans* DsyB MTHB *S*-methyltransferase activity. Similar complementation was shown with the addition of heat-killed *Prymnesium parvum* extracts to DSYB in ref.[12]. The activated DsyB protein was shown to have K_m and V_{max} values of 0.14 mM and 365 nmol min^{-1} mg protein^{-1}, respectively, for MTHB (Figure 2C), which were similar to those previously established for *P. parvum* DSYB (0.09 mM and 294 nmol min^{-1} mg protein^{-1}) in Curson et al.[12] The activated DsyB had a K_m of 0.16 mM and V_{max} 368.9 nmol min^{-1} mg protein^{-1} for the cosubstrate SAM (Figure 2D), which were also similar to those obtained with *P. parvum* DSYB (0.06 mM and 303 nmol min^{-1} mg protein^{-1}) in ref.[12].

Liquid chromatography with mass spectrometry (LC-MS) and/or native mass spectrometry were used in an attempt to identify the activation factor in the *L. aggregata dsyB*$^-$ extracts (Figure S1). A prominent peak at 37,084 Da was observed in the LC-MS spectrum for both the as-isolated and activated samples, which corresponds to DsyB with its N-terminal Met residue cleaved (commonly observed for proteins overexpressed in *E. coli*)[23]. A lower intensity peak at +131 Da, corresponding to the full-length protein (37,215 Da), was also observed in the as-isolated sample, indicating that the Met cleavage was not complete (Figure S1A). There was an additional minor peak at +269 Da of unknown origin in the activated sample (Figure S1B). Under nondenaturing conditions, both

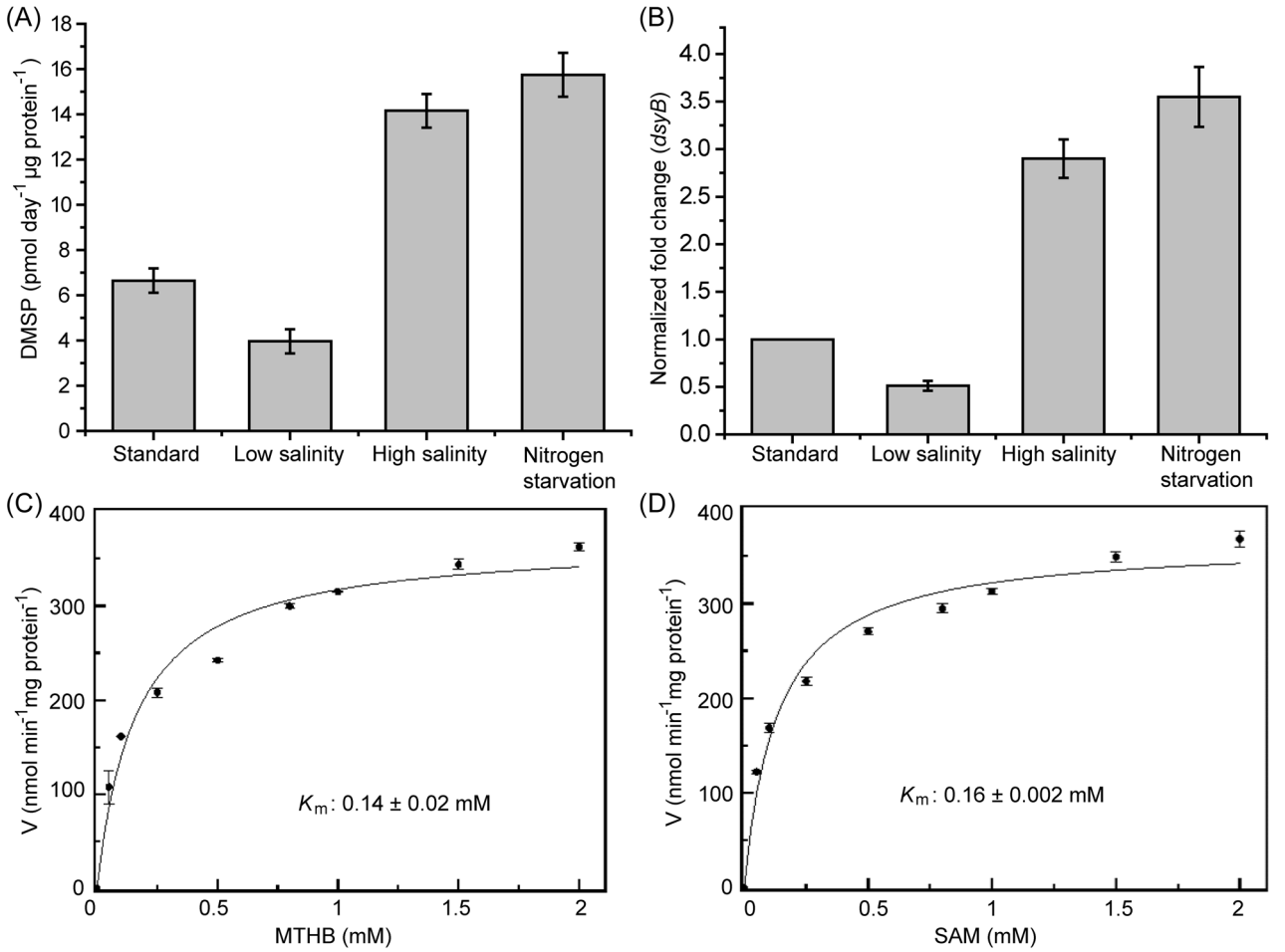


Figure 2. Characterization of *Nisaea dentrificans* DMSP production, *dsyB* transcription, and the DsyB enzyme. *N. dentrificans* DMSP production (A) and *dsyB* transcription (B) observed under different conditions. Standard conditions were MBM medium at 35 PSU with 10 mM NH_4Cl compared to low salinity (5 PSU), high salinity (50 PSU), and nitrogen starvation conditions (where cells in standard MBM were resuspended in standard MBM with no added NH_4Cl). (C) A nonlinear fit curve for MTHB methylation by DsyB. Initial rates were determined with 1.97 µM DsyB (50 mM Tris-HCl, 100 mM NaCl, pH 8.0) and 0–2 mM MTHB. K_m was 0.14 ± 0.02 mM. (D) A nonlinear fit curve for SAM demethylation by DsyB. Initial rates were determined with 1.97 µM DsyB and 0–2 mM SAM in the same reaction buffer. K_m was 0.16 ± 0.002 mM. The error bar represents the standard deviation of triplicate experiments. MBM, marine basal minimal; PSU, practical salinity unit.

monomeric and dimeric forms of DsyB were detected in the as-isolated sample, a feature commonly observed in non-denaturing mass spectra of solution dimers[24–26]. In the monomeric region, the main protein peaks (due to cleaved and noncleaved proteins) were again observed, but, in addition, a number of adduct species were present in the spectrum. Two of these, at +36 and +98 Da, correspond to chloride and (most likely) phosphate adducts. An additional adduct at +63 Da was also observed, possibly due to metal ion binding. In general, the spectrum of the dimeric form of DsyB was less well-resolved, but the main protein peak (at 74,168 Da) along with chloride and possible metal ion adducts were all detected (Figure S2). The non-denaturing mass spectrum of the activated monomeric DsyB (Figure S1B) revealed a number of adducts, including those most likely due to chloride, phosphate, and metal ion binding (all common within the as-isolated DsyB), along with an additional adduct at +122 Da (and at +244 Da), which is likely due to Tris buffer. Thus, we have no data to support there being a cofactor or modification of DsyB caused by the addition of the heat-killed *L. aggregata* *dsyB*$^-$ extracts to as-isolated DsyB, and further work is required to understand the variable nature of DsyB activity (see below).

The association of metal ions with DsyB was investigated further. Inductively coupled plasma mass spectrometry (ICP-MS) analysis revealed variable metal ion content with some preparations of as-isolated DsyB containing up to 0.85 Cu per protein, with other metals such as Ni (up to 0.5 per protein), Zn (0.4) and Fe (0.14) also detected. However, there was no correlation between metal ion content and activity of as-isolated samples.

Despite mostly lacking consistent *in vitro* MTHB *S*-methyltransferase activity (see below), native MS showed that the as-isolated *N. denitrificans* DsyB enzyme binds to SAM. The deconvoluted mass spectrum of a DsyB sample under nondenaturing conditions and containing 25 equivalents of SAM contained a peak in the DsyB dimer

region at +870 Da (predicted mass of a $(DsyB)_2$-$(SAM)_2$ is 74,966 Da), indicative of a $(DsyB)_2$-$(SAM\text{-}Cl)_2$ adduct, that was not observed in the absence of SAM (Figure S3A). Evidence for SAM binding was also apparent in the monomer region, though the presence of chloride adducts spreads out the intensity in this region (Figure S3B). Evidence for an MTHB-bound form of DsyB was also observed (although, again, the presence of chloride adducts spreads out

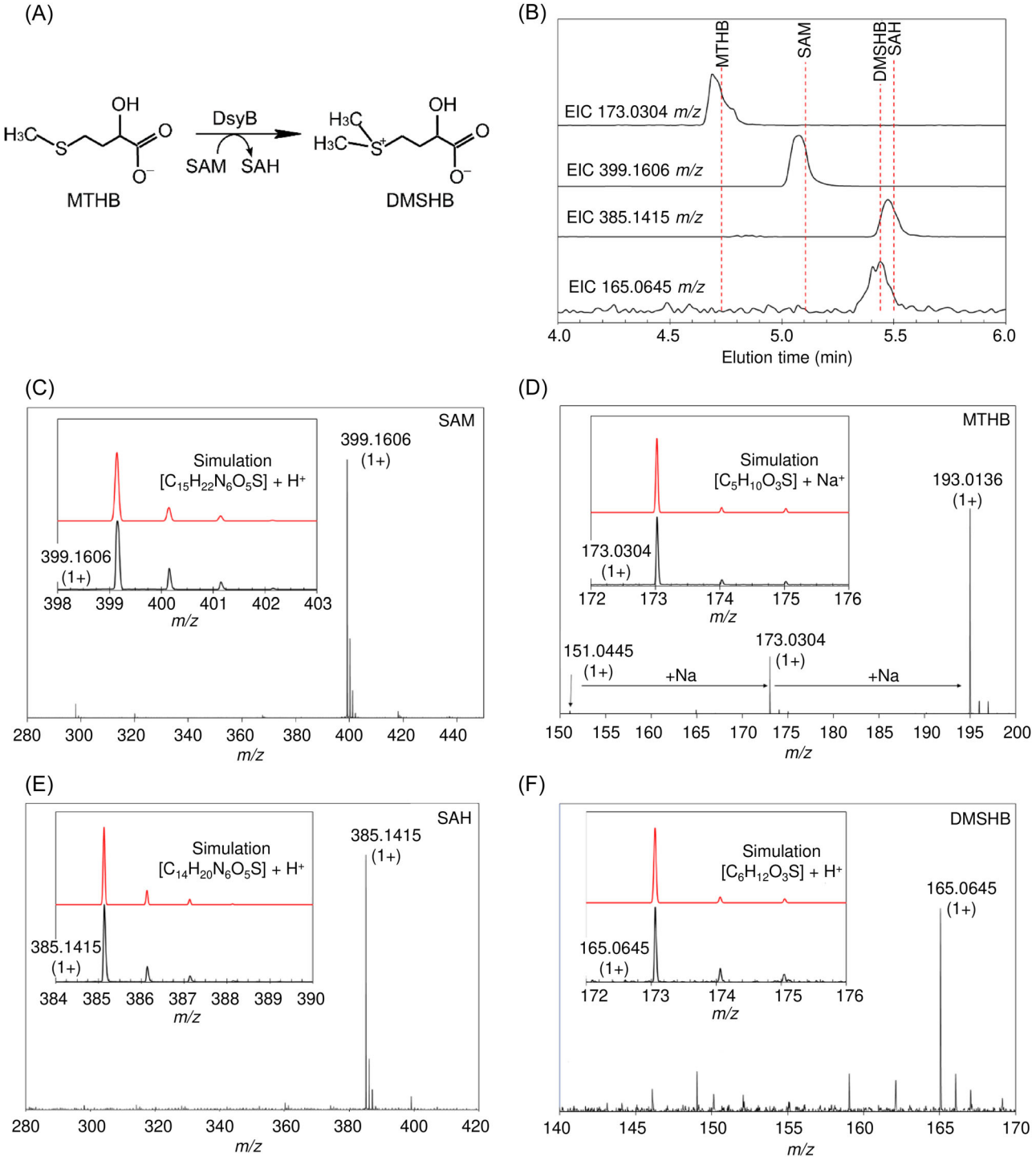


Figure 3. Identification of DsyB substrates and products by LC-MS. (A) The chemical equation for MTHB *S*-methylation into DMSHB. (B) Extracted ion chromatograms for MTHB, SAM, DMSHB, and SAH. Mass spectrometry data were analyzed to extract ion counts as a function of elution volume for the *m/z* ions indicated, which correspond to the substrates and products of the DsyB-catalyzed reaction. The red broken lines indicate elution volumes of the molecules when run as standards. (C–F) Mass spectra recorded for the eluted species, as indicated. Insets are spectra over a narrower *m/z* range (black lines) along with simulated spectra (showing the isotope distribution, red line) for each molecule, providing clear confirmation of identity. LC-MS, liquid chromatography with mass spectrometry.

intensity; Figure S3B). Although care is needed in interpreting intensities of peaks in the nondenaturing mass spectrum, the low intensity of the SAM- and MTHB-bound forms of DsyB suggests relatively low affinities when these substrates are present individually.

Small molecule hydrophilic interaction liquid chromatography (HILIC)-MS analysis of reactions following the addition of MTHB to DsyB-SAM resulted in the detection of substrates SAM and MTHB, and products DMSHB and SAH (Figure 3). Nondenaturing mass spectrometry of similarly generated samples resulted in the loss of SAM- and MTHB-bound forms of DsyB (Figure S3B). Together, the data are consistent with DsyB being a SAM-dependent MTHB *S*-methyltransferase.

Overall structure of DsyB

To analyze the catalytic mechanism of DsyB, we solved the crystal structures of complexes of DsyB with SAM and with SAH-MTHB. The crystal structure of the DsyB-SAM complex was determined by the single-wavelength anomalous dispersion (SAD) method using a selenomethionine derivative (Se derivative; Table S1).

Crystals of the DsyB-SAM complex belonged to the $P2_12_12_1$ space group, with four molecules arranged as a tetramer in the asymmetric unit. Each DsyB molecule contains two domains, an N-terminal domain (N-domain, Met1-Ala125) and a C-terminal domain (C-domain, Thr126-Glu337), which can be seen binding to the SAM molecule (Figure 4A). The DsyB C-domain contains seven β-strands surrounded by six α-helices, which together adopt the typical Rossmann-like α/β fold of Class I SAM-MTs (Figure 4A). Structural analysis showed that two DsyB monomers are tightly intertwined, mainly through interactions of residues from the N-domains of two adjacent monomers (Figure 4A). Analysis of DsyB using the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) predicted the DsyB dimer to be stable in solution. Indeed, gel filtration analysis indicated that DsyB is likely a dimer in solution (Figure 4B), consistent with the nondenaturing mass spectrometry data above (Figures S2 and S3A). These results indicate that DsyB

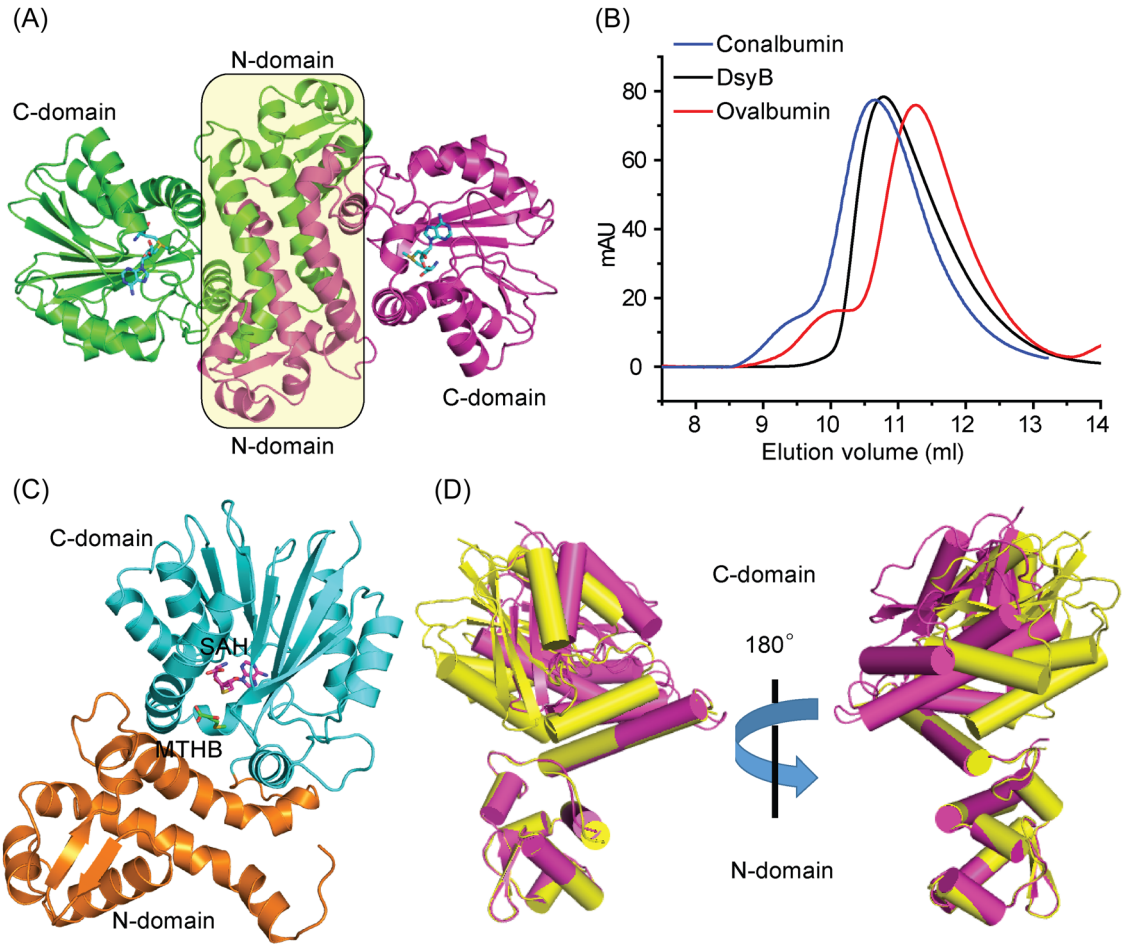


Figure 4. Overall structure analysis of DsyB. (A) Ribbon representation of DsyB dimer. Each DsyB monomer contains an N-domain and a C-domain. SAM molecules are shown as sticks colored in cyan. (B) Analysis of the form of DsyB in solution by gel filtration. Conalbumin (molecular mass = 75,000 Da; GE Healthcare) and ovalbumin (molecular mass = 43,000 Da; GE Healthcare) were used as markers. The predicted molecular mass of DsyB monomer is 37,215 Da. (C) The overall structure of DsyB-SAH-MTHB complex. The SAH molecule (colored in purple) and the MTHB molecule (colored in green) are represented as sticks. (D) Superimposed structures of DsyB with (colored in purple) and without (colored in yellow) binding the SAM molecule.

functions as a dimer in the same way as other SAM-MTs, whose N-domains are also responsible for dimerization[19,27,28]. ICP-MS and LC-MS analyses showed that the as-isolated DsyB contained variable metals. However, in the crystal structure of DsyB-SAM complex, no explicit electron density associated with metals was observed, suggesting that the binding site of metals may be not specific in DsyB.

The crystals of the DsyB-SAH-MTHB complex belong to the $P2_1$ space group (Table S1) and the resulting structure has a similar overall structure to that of the DsyB-SAM complex (Figure 4C). Interestingly, in this case the MTHB cosubstrate molecule is located between the C-domain and the N-domain of one DsyB monomer (Figure 4C).

The conformational change of DsyB in binding SAM

During the structural refinement of the DsyB-SAM complex, we found that three monomers (Chain A, B, and C) of the asymmetric unit contained SAM molecules. The structures of these three monomers are similar, with root mean square deviations (RMSDs) of ~0.5 Å between any two monomers. The Chain D of the DsyB-SAM complex is not bound to a SAM molecule. The conformation of Chain D is different to the other monomers bound to SAM, with a RMSD of ~2.4 Å between Chain D and Chain A. Moreover, residues Asp123 to Tyr143 in Chain D exhibited weak electron density, suggesting that this region is highly flexible. By superposing molecules of Chain A and Chain D, we observed that the N-domains of Chain A and Chain D are almost completely aligned, whereas the C-domain rotates ~10° as a rigid body (Figure 4D). These structural differences indicate that DsyB possesses two conformations: an "open" form and a "closed" form. Although DsyB can bind SAM and MTHB individually (Figure S3B), the binding of SAM triggers the conformational change of DsyB from the "open" form to the "closed" form, shrinking the cavity between the N-domain and the C-domain of DsyB and possibly promoting the subsequent binding of MTHB.

The structure of the DsyB-SAM complex is similar to that of the *Streptosporangium sibiricum* SibL protein (PDB code: 4U1Q), a *C*-directed Class I SAM-MT, with an RMSD of ~1.3 Å between these two structures. SAM binding also triggers the conformational change of SibL from an "open" form to a "closed" form to complete the formation of a binding site for its methyl acceptor 3-hydroxykynurenine[27]. Similar conformational changes have also been observed in other *C*/*O*-directed SAM-MTs, despite their low sequence identities[19,28].

Binding sites of SAM and MTHB

The SAM molecule within the DsyB-SAM complex (Figure 4A) is bound mainly by hydrogen bonds with residues in the DsyB C-terminal domain. DsyB residues Asp223 and Ala224 participate in binding the adenine ring of SAM; Asp196 forms hydrogen bonds with the ribose moiety of SAM; and Ser150, Gly173, and Ser239 interacts with the terminal amino acid moiety of SAM (Figure 5A). A similar binding mode was observed between the same DsyB C-terminal residues for SAH in the DsyB-SAH-MTHB complex.

In the DsyB-SAH-MTHB complex, the electron density of the MTHB molecule is relatively poor and the distance between SAH and MTHB is more than 6 Å, which is too far to enable the methyl transfer reaction. We speculate that the position of MTHB observed in the structure is not the exact location of MTHB when the reaction occurs under physiological conditions, and that the observed structure represents a state where the MTHB molecule has not completely entered into the active site. Nevertheless, the location of MTHB clearly implies its initial binding site (Figure 5B), and several residues likely involved in the binding of MTHB were identified, including Tyr97 and Gln101 from the N-domain and Tyr129, Tyr142, Gln146, and His291 from the C-domain of DsyB (Figure 5C).

To determine the importance of the Tyr97, Gln101, Tyr129, Tyr142, Gln146, and His291 in DsyB binding to MTHB, we performed site-directed substitutions of these residues and assayed the *in vivo* MTHB *S*-methyltransferase activity of the resultant variant DsyB derivatives in *R. leguminosarum*. *In vivo* assays were performed in *Rhizobium* because of the sensitivity of *in vitro* assays, see above. Site-directed mutations of Tyr97, Gln101, Tyr129, Tyr142, Gln146 or His291 severely decreased the enzymatic activity of DsyB (Figure 5D), indicating the potentially important roles of these residues in binding MTHB. In particular, the activity of the Tyr97Ala and Tyr142Ala variants was completely abolished (Figure 5D). However, Tyr97Phe and Tyr142Phe variants maintained >30% residual activity (Figure 5D), suggesting that the elimination of activity of Tyr97Ala and Tyr142Ala substitutions may be caused by the replacement of the aromatic side chain. This mutational analysis suggests that residues likely involved in the binding of MTHB are not essential for catalysis.

Three distinct catalytic mechanisms have been reported for SAM-MTs, including the proximity and desolvation mechanism, the general acid/base-mediated mechanism, and the metal-dependent mechanism[19]. Structural and biochemical analyses indicate that the activity of DsyB is neither metal-dependent nor catalytic residue-dependent, but is likely driven by the proximity effect. The DsyB enzyme likely enables favorable orientations of MTHB and SAM molecules that bring the sulfur atom of MTHB in close proximity to the methyl group of SAM.

The catalytic mechanism of DsyB

Based on our structural and biochemical results, we propose DsyB first binds a SAM molecule to generate a conformational change from "open" to "closed" state, which may promote the binding of MTHB (Figure 6A). When an MTHB molecule enters the active site, DsyB brings the sulfur atom of MTHB close enough to the methyl group of SAM to allow nucleophilic attack on the methyl group of SAM (Figure 6B). Subsequently, the generated DMSHB and SAH are released, and DsyB can rebind a SAM molecule from the intracellular environment in preparation for the next reaction.

Ecm18, another bacterial *S*-directed SAM-MT, converts disulfide in triostin A to the thioacetal linkage in the peptide

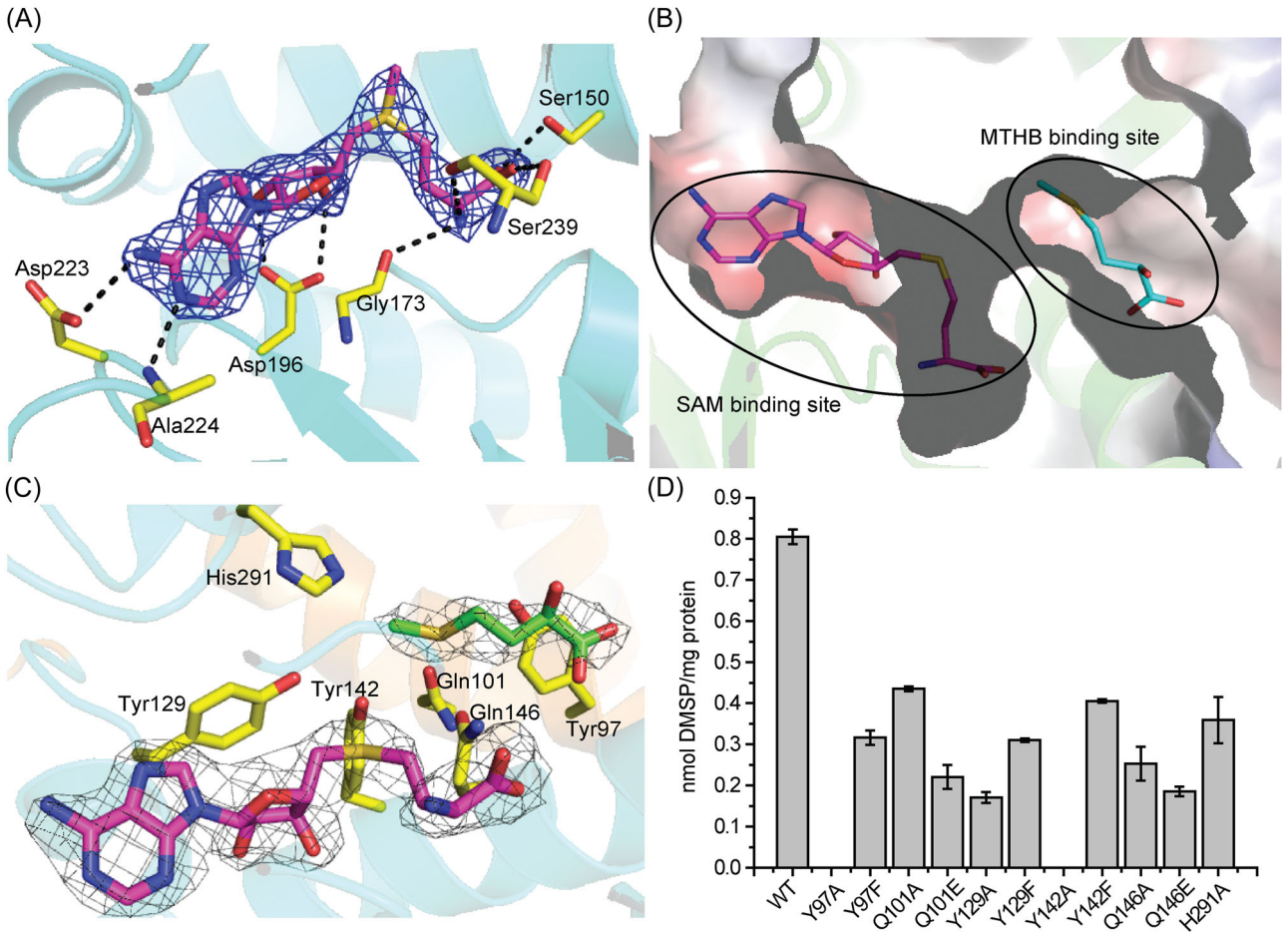


Figure 5. Analyses of residues of DsyB involved in binding SAM and MTHB. (A) Interactions between DsyB residues and SAM. SAM is colored in purple. The possible hydrogen bonds are represented by dashed lines. The $2F_o–F_c$ density for SAM is contoured in blue at 1.5σ. (B) Electrostatic surface of the crystal structure of DsyB. The SAM binding site and the MTHB binding site can be clearly identified. (C) The binding site of MTHB. Residues of DsyB that may participate in binding MTHB are shown in yellow. The $2F_o–F_c$ density for SAH (colored in purple) is contoured in grey at 1.5σ. The $2F_o–F_c$ density for MTHB (colored in green) is contoured in grey at 1.0σ. (D) Enzymatic activities of WT DsyB and site-directed mutants. The error bar represents the standard deviation of triplicate experiments. WT, wild-type.

antibiotic echinomycin through two stages, the methylation of one sulfur atom of the disulfide and the rearrangement of the methylated disulfide to form the thioacetal[29]. Thus, as we predict for DsyB, Ecm18 also uses the PD mechanism for its methyl transfer reaction[29]. Thiopurine *S*-methyltransferase (TPMT) is a murine *S*-directed SAM-MT that methylates 6-mercaptopurine[30]. Unlike DsyB, TPMT does not contain an N-terminal domain likely involved in dimerization, as its N-terminus only constitutes 40 residues[30]. Although Arg147 and Arg221 are possible participants in 6-mercaptopurine deprotonation, the modest decrease in the enzymatic activities of the corresponding mutants suggests that TPMT may possess the PD strategy for catalysis[30]. *Catharanthus roseus* CrSMT1 is another *S*-directed SAM-MT that methylates a broad range of substrates including benzene thiol and furfuryl thiol[31]. Homology modeling suggests that CrSMT1 contains an N-domain for dimerization[31], which is similar to DsyB. However, CrSMT1 is thought to use a histidine residue as a general base to deprotonate the thiol group of the substrate[31]. Thus, although the *S*-directed SAM-MTs only constitute a small portion of the reported SAM-MTs[19], their domain structures and catalytic mechanisms appear diverse.

Universality of the catalytic mechanism of DsyB

The majority of bacteria containing DsyB are *Rhodobacterales*, which are abundant in marine environments, but this enzyme is also found in some *Rhizobiales* and *Rhodospirillales* (including *N. dentrificans*)[10,32]. To investigate the ubiquity of the DsyB catalytic mechanism, we performed sequence alignment of DsyB proteins from different *Rhodobacterales*, *Rhizobiales*, and *Rhodospirillales* bacteria (Figure S4). Most residues involved in initial MTHB binding (Tyr97, Gln101, Tyr129, Tyr142, Gln146 and His291) and SAM binding (Ser150, Gly173, Asp196, Asp223, Ala224 and Ser239) are highly conserved in DsyB proteins from these marine bacteria, indicating that mechanistic insight gained here for *N. denitrificans* DsyB has universal significance in bacteria containing DsyB.

Eukaryotic DSYB, which may originate from bacterial DsyB, is a key enzyme for DMSP synthesis in many

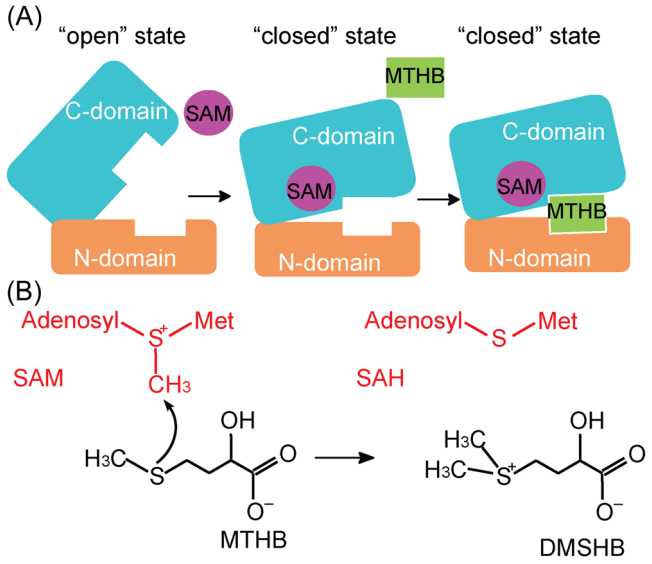


Figure 6. A proposed catalytic mechanism of DsyB. (A) The schematic diagrams of the DsyB conformational change triggered by the binding of SAM. (B) The sulfur atom of MTHB attacks on the methyl group of SAM to generate DMSHB and SAH. The MTHB molecule and DMSHB molecule are shown in black. And the SAM molecule and the SAH molecule are shown in red.

phytoplankton, such as marine haptophytes, dinoflagellates, and some diatoms[12]. DSYB shares ~33% sequence identity to DsyB, and we predicted the structure of DSYB from *Chrysochromulina tobin* CCMP291 by homologous modeling using SWISS-MODEL (https://swissmodel.expasy.org/)[33]. Structural alignment of DSYB and DsyB indicated that residues involved in binding MTHB are perfectly superposed (Figure S5). Moreover, sequence alignment of DsyB and DSYB from different marine algae showed that residues which play important roles in DsyB are highly conserved in different DSYB proteins (Figure S6), suggesting that DSYB proteins adopt a similar catalytic mechanism to DsyB.

DsyB/DSYB are the most abundant and transcribed *S*-methyltransferase enzymes of known DMSP synthesis pathways in marine microorganisms and environments

Having the identity of the key *S*-methyltransferases in diverse DMSP synthesis pathways (DsyB/DSYB and TpMMT in the bacterial and algal transamination pathway, and MmtN and BurB in bacterial methylation pathways; Figure 1), we carefully analyzed their presence in marine microorganisms and their abundance and transcript levels in published global metagenome and metatranscriptome datasets to quantify the potential environmental importance of these pathways.

Of the known DMSP synthesis enzymes, DsyB is by far the most abundant in sequenced and/or isolated bacteria (65.8% of cultured DMSP-producing isolates; Table S2)[7,10,11,34,35]. DsyB is mostly found in alphaproteobacterial *Rhodobacterales*, *Rhizobiales*, and *Rhodospirillales*, but is also sporadically found in, for example, an actinobacterial *Ponticoccus*isolate[7], and in some *Betaproteobacteria* and *Bacteroidetes* metagenome-assembled genomes[36]. MmtN has much fewer (14.4% of cultured DMSP-producing isolates), but equally diverse, host bacteria, being found in *Alphaproteobacteria*, *Gammaproteobacteria*, and *Actinobacteria* (Table S2). Finally, BurB is confined to very closely related *Burkholderia* spp. that likely uses DMSP as an intermediate in toxin production[34].

This hierarchy of DMSP synthesis gene abundance in bacteria (DsyB > MmtN >BurB) was mirrored in marine environmental data. In the *Tara* Oceans prokaryotic database, both *dsyB* and *mmtN* were found throughout the water column (Figure 7A), but no close homologs of *BurB* (e-value: <1e−40) were detected. This is consistent with BurB-mediated DMSP production in *Burkholderia* spp., possibly for toxin production, not being an important process in marine systems. *DsyB* was significantly more abundant than *mmtN* in both the metagenomic (median abundance 0.141% vs. 0.00376%) (Kruskal–Wallis $\chi^2 = 83.781$, $p < 0.001$) and metatranscriptomic (median abundance 0.2% vs. 0.0364%) (Kruskal–Wallis $\chi^2 = 33.64$, $p < 0.001$) *Tara* Oceans datasets (Figure 7A). Additionally, *dsyB* was found at 172 and 153 sampling sites (treating each depth as a separate site) in the metagenomes and metatranscriptomes, respectively, whereas *mmtN* was found at only 74 and 63 sites, respectively. Given this, our analysis of median abundance overestimates the contribution of *mmtN* to DMSP production in the global ocean. We, therefore, determined the relative abundance of *dsyB:mmtN* across depths at each sampling site in both metagenomes and metatranscriptomes (Figure S7). *DsyB* was more abundant at almost all sites in the metagenomes, and was more highly expressed across most, though there were a number of locations in the South Atlantic and South Pacific where *mmtN* was dominant (Figure S7). Taxonomic examination of both *dsyB* and *mmtN* sequences in the *Tara* database (Figure 7B) showed that both genes were exclusively from *Alphaproteobacteria*, primarily within the Orders *Rhodobacterales* and *Rhizobiales* for *dsyB*, and the genus *Thalassospira* for *mmtN* (Figure 7B). These data highlight DsyB as the most abundant, transcribed, and likely, important of the known bacterial DMSP synthesis enzymes in marine waters, which likely plays a significant role in the global production of DMSP.

Moving to eukaryotic DMSP synthesis, we carefully analyzed available transcriptome data from marine eukaryotes in the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP)[37]. The TpMMT MTHB *S*-methyltransferase has only been characterized in the centric diatom *T. pseudonana* CCMP1335, and close homologs (~70% protein identity) with the same singular domain structure only exist in 17/82 diatom transcriptomes (seven of which also contain DSYB), and no other phytoplankton (Tables S3 and S4)[38–53]. The next most homologous TpMMT-like proteins, present in for example, *Thalassiosira oceanica* (EJK59074) and *Fistulifera solaris* (GAX25165) that are more diverse (the methyltransferase domain being <50% identical to TpMMT), contain extra protein domains and, thus, are much larger proteins whose function is unknown. These TpMMT-like proteins cannot be considered as functional MTHB *S*-methyltransferase enzymes

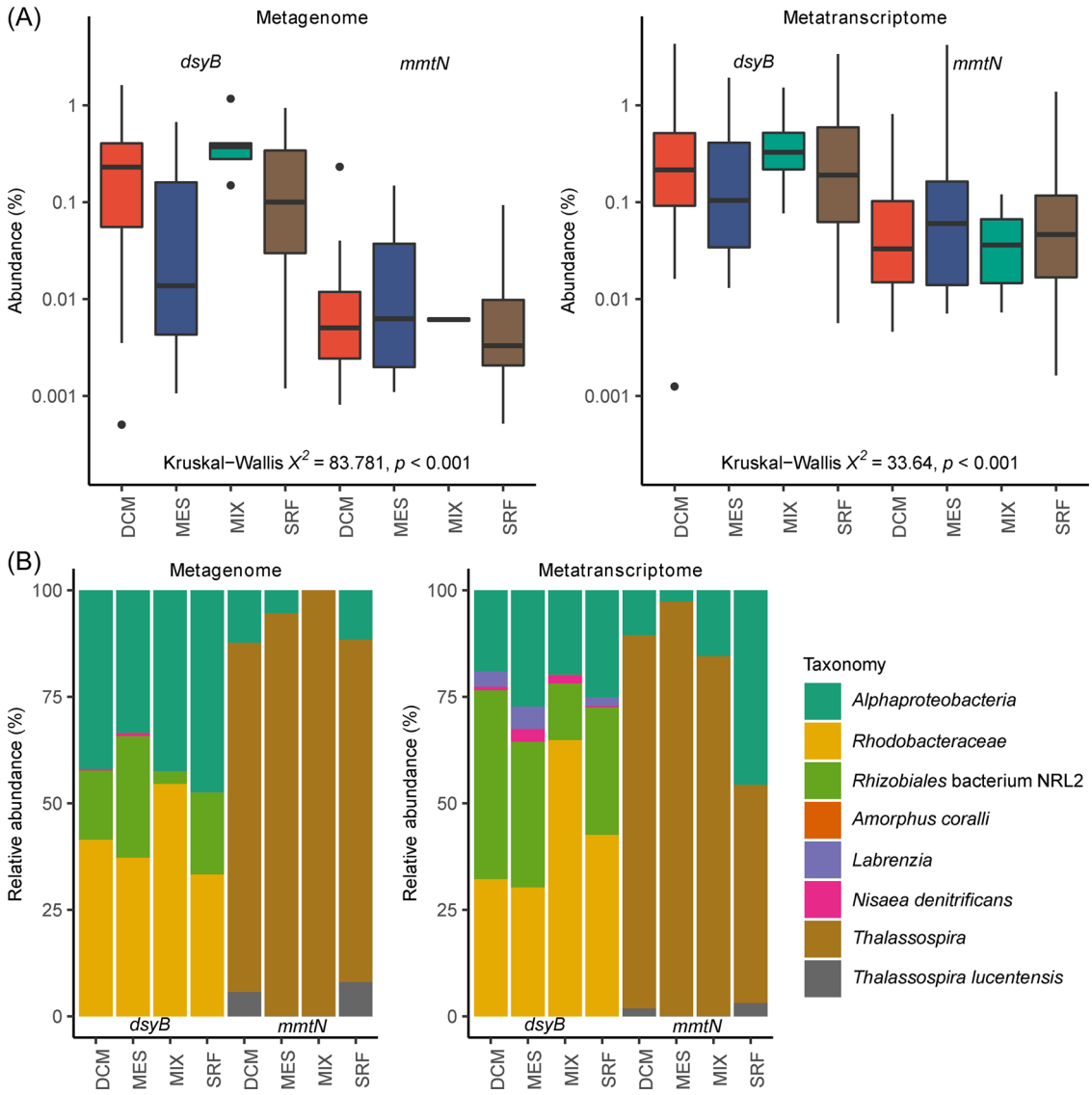


Figure 7. Analysis of MATOU for *dsyB* and *mmtN*. (A) Normalized abundance of *dsyB* and *mmtN* in *Tara* metagenomes and metatranscriptomes, by sampling depth. Abundances are normalized as a percentage of the median gene or transcript abundance of 10 single-copy marker genes. Box plots show median values (central black line), and lower and upper hinges correspond to the first and third quartiles of the data. Kruskal–Wallis χ^2 values for comparisons of median abundances between *dsyB* and *mmtN* (across all depths combined) are shown. (B) Taxonomic assignment and relative abundance (as a percentage) of *dsyB* and *mmtN* sequences in the *Tara* metagenomes and metatranscriptomes. Taxa designated *Alphaproteobacteria* lack further taxonomic resolution. DCM, deep chlorophyll maximum layer; MATOU, Marine Atlas of Tara Ocean Unigenes; MES, mesopelagic layer; MIX, epipelagic wind mixed layer; SRF, surface water layer.

and are omitted from this study. In contrast, DSYB is found in the transcriptomes of 47/61 dinoflagellates and 24/30 haptophytes, organisms known to produce the highest levels of DMSP per cell (>50 mM)[39,46]. Furthermore, 15/82 diatom transcriptomes, typically known to produce lower cellular DMSP levels (generally < 50 mM)[39], and some *Ochrophyta*, *Cnidaria*, and *Cilophora* transcriptomes also contained DSYB. These data show DSYB to be the most abundant and widespread DMSP synthesis enzyme known in eukaryotic DMSP-producing organisms.

Within the eukaryotic Marine Atlas of Tara Ocean Unigenes (MATOU), we found both *DSYB* and *TpMMT* within epipelagic (surface water layer [SRF] and deep chlorophyll maximum [DCM]) waters. Initial examination showed *DSYB* to be more abundant in ≤3 µm fractions than in larger fractions (Figure S8). Data from these smaller ≤3 µm fractions that likely contain picoeukaryotes (CoP) were considered together. Likewise, fractions with a minimum filter size of ≥3 µm that likely exclude picoeukaryotes (ExP) were also considered together. Abundance was not significantly different between SRF and DCM sampling depths for either CoP or ExP *DSYB* (Kruskal–Wallis $\chi^2 = 0.113$, $p = 0.74$, and Kruskal–Wallis $\chi^2 = 0.004$, $p = 0.95$, respectively), or for CoP or ExP *TpMMT* (Kruskal–Wallis $\chi^2 = 0.102$, $p = 0.75$, and Kruskal–Wallis $\chi^2 = 0.194$, $p = 0.66$, respectively), and as such these sampling depths were combined for the

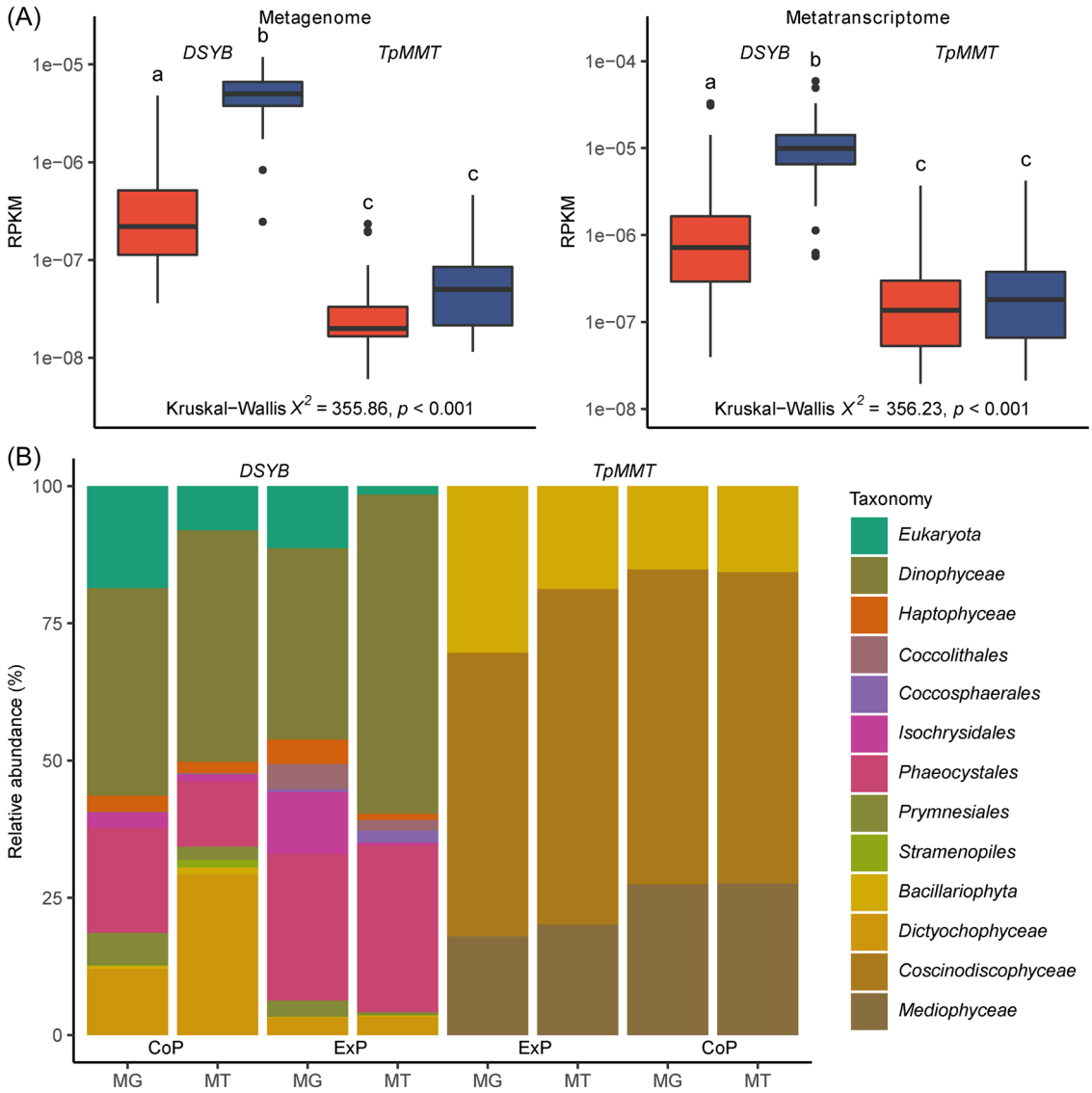


Figure 8. Analysis of MATOU for *DSYB* and *TpMMT*. (A) Normalized abundance of *DSYB* and *TpMMT* in MATOU metagenome and metatranscriptomes, by fractions containing picoeukaryotes (CoP) (i.e., with a minimum filter size of <3 μm) (blue) and fractions excluding eukaryotes (ExP) (i.e., with a minimum filter size of ≥3 μm) (red). DCM and SRF depths are combined for the purposes of this analysis. Abundances are normalized as reads per kilobase per million mapped reads (RPKM). Box plots show median values (central black line), and lower and upper hinges correspond to the first and third quartiles of the data. Kruskal–Wallis χ^2 values for comparisons between *DSYB*/*TpMMT* and fraction abundance are shown. Letters denote gene or transcript/fraction combinations that are significantly different ($p < 0.05$) by post-hoc Dunn's test, using Holm's correction. (B) Taxonomic assignment and relative abundance (as a percentage) of *DSYB* and *TpMMT* sequences in the MATOU metagenome (MG) and metatranscriptome (MT). Taxonomy is reported as Phylum (if available) for *DSYB* and as Class (if available) for *TpMMT*. Taxa designated *Eukaryota* lack further taxonomic resolution.

purposes of comparative analyses between *DSYB* and *TpMMT* abundance.

Metagenomes derived from the MATOU data set revealed that *DSYB* was significantly more abundant than *TpMMT* in both CoP and ExP fractions (Figure 8A; *DSYB* vs. *TpMMT* CoP median abundance 4.99e−06 vs. 5.01e−08 RPKM, post-hoc Dunn's test $z = 16.22$, $p < 0.001$, ExP median abundance 2.2e−07 vs. 1.99e−08 RPKM, post-hoc Dunn's test $z = 6.97$, $p < 0.001$). Similarly, *DYSB* was significantly more abundant than *TpMMT* in MATOU derived metatranscriptomes (Figure 8A; *DSYB* vs. *TpMMT* CoP median abundance 9.89e−06 vs. 1.81e−07 RPKM, post-hoc Dunn's test $z = 15.16$, $p < .001$, ExP median abundance 7.21e−07 vs. 1.36e−07 RPKM, post-hoc Dunn's test $z = 7.33$). *DSYB* was also significantly more abundant within the CoP fraction than the ExP fraction in the metagenome (Figure 8A; median abundance 4.99e−06 vs. 2.2e−07, post-hoc Dunn's test $z = 11.30$, $p < 0.001$), and in the metatranscriptome (Figure 8A; median abundance 9.89e−06 vs. 7.21e−07, post-hoc Dunn's test $z = 12.20$, $p < 0.001$). In contrast, *TpMMT* abundance was not significantly different between CoP and ExP fractions in either the metagenome

(Figure 8A; median abundance 5.01e−08 vs. 1.99e−08, post-hoc Dunn's test $z = 1.61$, $p = 0.11$) or the metatranscriptome (Figure 8A; median abundance 1.81e−07 vs. 1.36e−07, post-hoc Dunn's test $z = 0.92$, $p = 0.35$). Again, these analyses likely overestimated the abundance of *DSYB* in the ExP fraction, because, in the metagenome, *DSYB* was detected at 138/140 CoP fraction sites, but was only found at 178/272 ExP fraction sites (χ^2 [1, $N = 412$] = 56.77, $p < 0.001$). *TpMMT* abundance was also likely overestimated as *TpMMT* was detected at 90/140 CoP fraction sites, and at 39/272 ExP fraction sites (χ^2 [1, $N = 412$] = 107.21, $p < 0.001$). As such, *DSYB* was detected significantly more frequently in CoP (χ^2 [1, $N = 280$] = 54.41, $p < 0.001$) and ExP (χ^2 [1, $N = 544$] = 148.12, $p < 0.001$) fraction sites than *TpMMT*. Similarly, within the metatranscriptome, *DSYB* was detected at 139/140 CoP fraction sites and at 251/272 ExP fraction sites (χ^2 [1, $N = 412$] = 898, $p < 0.01$). *TpMMT* was detected at 94/140 CoP fraction sites, and at 99/272 ExP fraction sites (χ^2 [1, $N = 412$] = 35.09, $p < 0.001$). Again, *DSYB* was detected at significantly more CoP (χ^2 [1, $N = 280$] = 51.78, $p < 0.001$) and ExP (χ^2 [1, $N = 544$] = 185.10, $p < 0.001$) fraction sites than *TpMMT* in the metatranscriptome data. Given the greater abundance of *DSYB* over *TpMMT* in the environmental data, and that the majority of environmental *DSYB* sequences are likely from dinoflagellates and/or haptophytes, known to be high producers[39,46], of DMSP compared to *TpMMT* in the generally low-producing diatoms[39], DSYB is currently the most important known DMSP synthesis enzyme (Figure 8B).

To conclude, DMSP is an abundant and ecologically important organosulfur compound. DsyB/DSYB enzymes catalyze the committed *S*-methylation of MTHB to generate DMSHB, which is the key step of the transamination pathway for DMSP synthesis in most bacteria and algae[10–12]. Furthermore, DsyB/DSYB enzymes are present in the most prodigious DMSP-producing haptophyte and dinoflagellate phytoplankton, and represent the most abundant and transcribed *S*-methylase genes of known DMSP synthesis pathways in marine waters. In this study, we solved the first crystal structures of bacterial DsyB-SAM and DsyB-SAH-MTHB complexes and demonstrated the conversion of SAM and MTHB into SAH and DMSHB. Based on structural and mutational analyses, the catalytic mechanism of DsyB is proposed, and has universal significance in bacteria containing DsyB, and in marine algae containing DSYB. Our results provide novel insights into DMSP synthesis, shedding light on the global sulfur cycling.

MATERIALS AND METHODS

Bacterial strains and growth conditions

E. coli BL21 (DE3) was grown in lysogeny broth (LB) medium at 37°C. *R. leguminosarum* J391 was grown in TY[54] complete medium or Y[54] minimal medium (10 mM NH_4Cl as nitrogen source) at 28°C. *L. aggregata* J571 (*dsyB*$^-$) was grown in YTSS[55] complete medium or marine basal minimal (MBM)[56] medium (0.5 mM NH_4Cl as nitrogen source) at 30°C. *N. denitrificans* DR41_21 (DSM 18348), purchased from DSMZ, Germany, was cultured in the Difco 2216 medium (http://www.dsmz.de/) or MBM medium at 30°C under different salinity and nitrogen levels for differential DMSP production experiments. Standard conditions were 10 mM NH_4Cl and 35 practical salinity units (PSU) compared to 5 PSU (low salinity) or 50 PSU (high salinity). Cultures were sampled for DMSP and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) work in exponential-phase growth (after ~7 days). For nitrogen starvation experiments, exponential phase cells grown under standard conditions were harvested and incubated overnight in standard MBM media with no added NH_4Cl. Where necessary, 10 mM succinate was used as a carbon source and antibiotics were added at the following concentrations: gentamicin (20 μg ml^{-1}), streptomycin (400 μg ml^{-1}), kanamycin (20 μg ml^{-1}), ampicillin (100 μg ml^{-1}), spectinomycin (200 μg ml^{-1}).

General *in vivo* and *in vitro* genetic manipulations

Plasmids (Table S5) were transferred to *E. coli* by transformation, and *R. leguminosarum* J391 or *L. aggregata* J571 by conjugation using the helper plasmid pRK2013[57]. Routine restriction digestions and ligations for cloning were performed essentially as in ref.[58]. The oligonucleotide primers used in this study were synthesized by Eurofins Genomics (Table S6). Sequencing of plasmids and PCR products was performed by Eurofins Genomics.

The *dsyB* gene was amplified from *N. denitrificans* genomic DNA and cloned into the pET22b (Novagen) for the expression of DsyB with a C-terminal His-tag or into pLMB509 for expression in *Rhizobium* and *Labrenzia*[59]. Amino acid substitution mutations in DsyB were generated using QuickChange® mutagenesis kit (Agilent Technologies) and the primers in Table S6. All site-directed mutant (SDM) variant plasmids were verified by DNA sequencing.

Reverse transcription-quantitative polymerase chain reaction

RNA was isolated from 100 ml *N. denitrificans* cultures using RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol with some modifications. On-Column DNase digestion was performed with RNase-Free DNase Set (Qiagen). Reverse transcription of 1 μg of DNA-free RNA per sample was done using the QuantiTect Reverse Transcription Kit (Qiagen). PCR on RNA and complementary DNA (cDNA) samples confirmed that RNA samples were DNA-free.

Primers for RT-qPCR for *N. dentrificans dsyB* and housekeeping genes *recA* and *gyrB*, were designed using Primer5[60] (Table S6). The qPCR experiments were performed on a StepOnePlus instrument (Applied Biosystems). Quantification was performed using the SYBR® Green Jump Start™ Taq ReadyMix™ (Sigma-Aldrich) following the manufacturer's instructions. Reactions (20 μl) contained 2 μl cDNA and 0.8 μl primers (10 μM), with an annealing

temperature of 55°C. For each condition and gene, the cycle threshold (C_t) values of triplicate technical and biological replicates were averaged. Relative expression levels were determined with efficiency correction[61]. *dsyB* transcription was displayed as normalized fold change to the standard condition.

DsyB enzyme assays *in vivo*

To measure MTHB *S*-methyltransferase activity from pLMB509 clones (and SDM derivatives) in *R. leguminosarum* J391, cultures were grown overnight in TY complete medium. Then 1 ml of culture was centrifuged at 20,000*g* for 2 min, resuspended in the same volume of Y medium and then diluted 1:100 into 5 ml Y with 0.5 mM DL-MTHB (55875; Sigma-Aldrich), 10 mM taurine (to induce expression; T0625; Sigma-Aldrich) and gentamicin, and incubated for 60 h at 28°C before sampled for gas chromatography (GC) analysis to determine the amount of DMSP product (see below).

To measure MTHB *S*-methyltransferase activity from pLMB509 clones expressing the *dsyB* gene in the *L. aggregata dsyB*$^-$ mutant strain J571, cultures were grown overnight in YTSS complete medium. Following incubation, 1 ml of culture was then centrifuged at 20,000*g* for 2 min, resuspended in the same volume of MBM medium and then diluted 1:50 into 5 ml MBM with 10 mM taurine (to induce expression; Sigma-Aldrich), gentamicin and rifampicin, and incubated for 24 h at 30°C. DL-MTHB (0.5 mM) was added as substrate to the cultures and these were incubated for 4 h at 30°C before sampled for GC and protein estimation by the Bradford assay.

To measure DMSHB/DMSP in *Rhizobium* or *Labrenzia* assay mixtures, 200 µl of culture was added to a 2 ml glass serum vial, then 100 µl 10 M NaOH was added and vials were crimped with polytetrafluoroethylene/rubber crimp caps immediately. Vials were incubated at 80°C for 10 min (to capture DMSHB as well as DMSP) and then for 24 h at room temperature in the dark before being monitored by GC assay. All GC assays involved measurement of headspace DMS using a flame photometric detector (Agilent 7890A GC fitted with a 7693 autosampler) and an HP-INNOWax 30 m × 0.320 mm capillary column (Agilent Technologies J&W Scientific). Calibration curves were produced by alkaline lysis of DMSP standards in water. The detection limit for headspace DMS from DMSP was 0.015 nmol and from DMSHB was 0.3 nmol. DsyB activity is expressed as nmol DMSHB/DMSP min^{-1} mg $protein^{-1}$. Protein concentrations were determined using the Bradford method (BioRad). Control assays of *Rhizobium* or *Labrenzia* J571 containing pLMB509 were carried out, as above, and gave no detectable DsyB activity.

Protein expression and purification

E. coli BL21 (DE3) containing pET22b::*dsyB* clones were cultured in an LB medium containing ampicillin at 37°C. At mid-exponential growth (OD_{600} 0.5–0.7), 0.5 mM isopropyl β-*D*-1-thiogalactopyranoside was added and the cells were incubated at 20°C for 16 h. Cells were harvested by centrifugation (20 min, 7500*g*, 4°C), washed, and resuspended in 25 mM Tris-HCl, pH 8.0, 150 mM NaCl. Cell lysis was performed by three passages through a French Press (16,000 psi), unbroken cells and cell debris were removed by centrifugation (30 min, 5500*g*, 4°C) and the supernatant was recovered and subjected to centrifugation (60 min, 185,000*g*, 4°C) to pellet the membrane fraction. The soluble cell lysate was applied to a slurry of Ni-NTA resin (Qiagen) at a 3:1 ratio for 90 min with shaking at 4°C. The lysate/slurry mix was loaded into Econo-Pac polypropylene columns, washed with 50 mM Tris-HCl, 250 mM NaCl, 20 mM imidazole, pH 8.0, and DsyB was eluted in 5 × 1 ml fractions using 50 mM Tris-HCl, 250 mM NaCl, 250 mM imidazole, pH 8.0. Fractions containing DsyB were concentrated and buffer exchanged into 2 ml of 50 mM Tris-HCl, 100 mM NaCl, pH 8.0 and applied to a Superdex 200 10/300GL preparative grade gel filtration column (Cytiva). The purified protein was flash-frozen in liquid nitrogen and stored at −80°C until required.

DsyB enzyme assays *in vitro*

L. aggregata dsyB$^-$ J571[10] was grown to late exponential phase in MBM. Cell lysates were prepared by centrifuging 100 ml of culture for 10 min at 2500*g*. The pellet was washed and resuspended with 20 ml 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 150 mM NaCl, pH 7.5 before cell lysis via French press (16,000 psi). The cell lysate was heat-treated at 80°C for 10 min to denature proteins, then applied to a 10 ml PD10 column, eluted over 10 ml, and collected in 1 ml fractions. DsyB activity was monitored by performing *in vitro* enzyme assays with 50 µl of the separate heat-killed extract fractions, 1 mM SAM (Sigma-Aldrich), 1 mM DL-MTHB and 1.97 µM DsyB or no protein (control) in 400 µl total volume. Experiments were done as above with purified DsyB without addition of heat-killed extracts, but these gave no activity (data not shown). Reactions were incubated for 30 min at 28°C and then 800 µl of activated charcoal (38 mg ml^{-1} in 0.1 M acetic acid) was added to the samples and mixed. Samples were centrifuged at 14,000*g* for 15 min and the supernatant was retained. For GC analysis, 200 µl of the supernatant was added to a 2 ml vials with 10 M NaOH (100 µl) and was immediately crimped. Crimped vial was then heated to 80°C for 10 min and incubated in the dark at 22°C for 16 h. These samples were subsequently used for DMS quantification by GC analysis (as above). No DMS was produced from the no DsyB protein controls.

For kinetics analysis of *N. denitrificans* DsyB, the as-isolated protein was activated by addition of 400 µl heat-killed cell lysate fractions liberated from the *L. aggregata dsyB*$^-$ deletion mutant. K_m and V_{max} values were determined by nonlinear analysis using 1.97 µM DsyB and 0–2 mM SAM (fixed at 1 mM for DL-MTHB kinetic work), or 0–2 mM DL-MTHB (fixed at 1 mM for SAM kinetics work; Figure 2). The reaction mixture was incubated at 28°C for 30 min before detection of DMSHB as above. Origin version 8.5 was used to calculate K_m.

Mass spectrometry analysis

LC-MS was used to confirm the mass of intact (but denatured) DsyB, and also for the analysis of small molecules. For analysis of DsyB, denaturing LC-MS was conducted using a Bruker microQTof-QIII electrospray ionization time of flight mass spectrometer, operating in positive mode. The spectrometer was calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies). Samples were prepared by 10-fold dilution of ~100 µM DsyB protein solution with 2% (vol/vol) acetonitrile and 0.1% (vol/vol) formic acid to 0.5 ml. Samples were chromatographically separated by an UltiMate 3000 HPLC system (Dionex) fitted with a ProSwift reversed phase RP-1S column (4.6 × 50 mm; Dionex). Hystar (Bruker Daltonics) was used to coordinate mass spectrometer and high-performance liquid chromatography (HPLC) operations. Bound proteins were eluted using an isocratic gradient (2%–100% B) at a flow rate of 0.2 ml min^{-1} using the following solvents: Solvent A (water, 0.1% [vol/vol] formic acid); and Solvent B (acetonitrile, 0.1% [vol/vol] formic acid). The eluant was continuously infused into the source of the mass spectrometer operating with the following parameters: dry gas flow 8 l min^{-1}; nebulizer gas pressure 1.8 bar; dry gas 240°C; capillary voltage 4500 V; offset 500 V; collision RF 650 Vpp.

Mass spectrometric analysis of small molecule substrates and products in DsyB assay mixtures was performed using HILIC[62], which is particularly useful for the separation of small polar compounds such as MTHB or DMSHB. HILIC-MS[63] experiments were performed using the same mass spectrometer and HPLC system as above, but with the latter fitted with a Luna NH_2 column (2 × 100 mm) (Phenomenex). For HILIC chromatography, the following solvents were freshly prepared: Solvent A (95% [vol/vol] aqueous 5 mM ammonium formate, pH 3.75, 5% [vol/vol] acetonitrile); Solvent B (95% [vol/vol] acetonitrile, 5% [vol/vol] aqueous 100 mM ammonium formate, pH 3.75). Standard compounds (SAM, DL-MTHB, SAH [Sigma-Aldrich], DMSHB)[10] were used to calibrate the elution profile of the HILIC column. Samples were brought to 92% (vol/vol) acetonitrile and loaded onto a column pre-equilibrated with Solvent B. An optimized HILIC gradient was applied and compounds were eluted (0.6 ml min^{-1}) using the HILIC gradient between Solvent A and Solvent B, as previously described[10]. The eluant was continuously infused into the source of the mass spectrometer (optimized for 50–600 *m/z*) with the following parameters: dry gas flow 8.5 l min^{-1}; dry gas 200°C; nebulizer pressure 1.2 bar; capillary voltage 4500 V; offset 500 V; collision RF 400 Vpp. Each HILIC-MS run contained an internal sodium formate calibration segment at the end of the run.

Nondenaturing MS (often referred to as native MS), in which noncovalently bound protein–cofactor, simple protein–protein, or even multiprotein interactions are preserved[64,65], was used to investigate substrate binding and confirm the presence of dimeric DsyB. Before analysis, protein samples were exchanged into 250 mM ammonium acetate, pH 8.0, using Zeba spin (Thermo Scientific) or PD mini trap (Cytiva) desalting columns and infused (0.3 ml/h) directly into the ESI source of the Bruker microQTof-QIII mass spectrometer with the following parameters: dry gas flow 4 l min^{-1}; nebulizer gas pressure 0.8 bar; dry gas temperature 190°C; capillary voltage 3000 V; capillary offset 500 V; ion energy 8 eV; collisional RF 1500 Vpp; collision cell voltage 5 V; and, ion transmission range 1500–5500 *m/z*.

Processing, isotope pattern simulation and analysis of denaturing LC-MS, HILIC-MS, and nondenaturing MS data were carried out using Compass Data Analysis version 4.1. For denaturing LC-MS and nondenaturing MS, neutral mass spectra were generated using ESI compass Maximum Entropy deconvolution algorithm version 1.3. Proteins masses were reported from peak centroids representing the isotope average neutral mass and compared to predicted masses (Expasy)[66].

DsyB activity detection using HILIC-MS

Heat-killed J571 fractions that restored MTHB *S*-methyltransferase were added to pure DsyB as above to yield activated samples for analysis here. Samples of as-isolated DsyB prepared in HEPES buffer or activated DsyB, were immediately desalted (PD10; Cytiva) into 25 mM Tris, 100 mM NaCl, pH 8.0 before conducting HILIC or nondenaturing MS experiments. The methyltransferase activity of DsyB was measured using DL-MTHB and SAM as substrates, as previously described[17], with a slight modification. The reaction mixture (20 µl) consisted of 14 µl water, 2 µl of buffer (100 mM Tris-HCl, pH 7.5), 0.5 µl of 20 mM DL-MTHB, 1.5 µl of 32 mM SAM as cosubstrate, and 2 µl of enzyme solution (DsyB in range 7–30 µM depending on particular experiment). The reaction was incubated at 25°C, overnight and quenched by the addition of 230 µl acetonitrile. Samples were analyzed immediately by mass spectrometry.

Crystallization and data collection

The purified DsyB protein was concentrated to ~8 mg ml^{-1} in the buffer containing 100 mM NaCl and 10 mM Tris-HCl (pH 8.0). Initial crystallization trials for DsyB were performed using the sitting drop vapor diffusion method at 20°C. To obtain crystals of DsyB-SAM complex, the purified DsyB protein was mixed with 1 mM SAM at 4°C for 30 min. Diffraction-quality crystals of DsyB-SAM complex were obtained in hanging drops containing 0.1 M HEPES (pH 7.5), 0.2 M NaCl and 25% (wt/vol) polyethylene glycol (PEG) 3350 after 1-week incubation at 20°C. Crystals of the DsyB-SAM complex Se derivative were obtained in hanging drops containing 0.1 M Bis-Tris propane (pH 7.5), 0.2 M sodium acetate, and 20% (wt/vol) PEG 3350 after 1-week incubation at 20°C. To obtain crystals of DsyB-SAH-MTHB complex, the purified DsyB protein was mixed with SAH (1 mM) and DL-MTHB (1 mM) at 4°C for 30 min. Crystals of DsyB-SAH-MTHB complex were obtained in hanging drops containing 0.1 M Tris (pH 8.0), 0.2 M NaCl, and 20% PEG 4000 after 1-week

incubation at 20°C. X-ray diffraction data were collected on the BL18U1&BL19U1 beamlines at the Shanghai Synchrotron Radiation Facility. The initial diffraction data sets were processed by the HKL3000 program[67].

Structure determination and refinement

The crystals of DsyB-SAM complex belong to the $P2_12_12_1$ space group, while the crystals of DsyB-SAH-MTHB complex belong to the $P2_1$ space group. The structure of DsyB-SAM complex Se derivative was determined by SAD phasing. The crystal structures of DsyB-SAM complex and DsyB-SAH-MTHB complex were determined by molecular replacement using the CCP4 program Phaser[68] with the structure of the Se derivative as the search model. The refinements of these structures were performed using Coot[69] and *Phenix*[70]. All the structure figures were produced with the PyMOL (http://www.pymol.org/).

Analyses of DMSP synthesis genes in cultured microorganisms

The presence or absence of DMSP synthesis genes in 111 cultured DMSP-producing bacteria (published since the discovery of bacterial DMSP synthesis[7,10,11,34,35]), was analyzed (Table S2). This is based on previously published work that analyzed their sequenced genomes and/or used degenerate primers to detect the presence of *burB*, *dsyB*, and/or *mmtN*. Percentage abundances were calculated for all three DMSP synthesis genes within these cultured organisms, as well as those containing both *dsyB* and *mmtN*.

Eukaryotic transcriptomes from the MMETSP[37] (Tables S3 and S4) were analyzed for the presence of *DSYB* and/or *TpMMT* through tblastn searches against DSYB[12] and TpMMT[17] sequences whose enzyme activity had been previously demonstrated. These were manually curated to confirm identity (*e* value cutoff of $1e^{-30}$ for DSYB), although TpMMT has only been shown to be functional in *T. pseudonana*, we did not assume any sequences below 70% identity to *T. pseudonana* TpMMT to be functional. Strains confirmed to contain *DSYB* and/or *TpMMT* are listed in Table S3 and summarized in Table S4, alongside literature reporting the presence of DMSP synthesis in that particular strain (if tested).

Metagenome and metatranscriptome analyses

Verified sequences[7,10–12,71], were aligned using ClustalOmega[72], and profile hidden Markov models (hmms) of *dysB*, *DSYB*, and *mmtN* were constructed using the hmmbuild function of hmmer 3.3[73]. *Tara* metagenome (OM-RGC_v2_metaG/MATOU_v1_metaG) (prokaryotic/eukaryotic, respectively), and metatranscriptome (OM-RGC_v2_metaT/MATOU_v1_metaT) sequences together with their abundances and taxonomic assignations were downloaded from the Ocean Gene Atlas site[74] using an hmmsearch *e*-value threshold of 1e–70 (*dsyB*), 1e–80 (*DSYB*), or 1e–60 (*mmtN*). A blastp search (*e*-value threshold of 1e–80) was used for *TpMMT*, using the *T. pseudonana TpMMT* sequence as query. Environmental *dsyB*/*DSYB* sequences were aligned with *N. denitrificans dsyB* using ClustalOmega[72], and sequences that did not possess all six essential residues were excluded from further analysis. Environmental *TpMMT* sequences greater than 400 amino acids in length were also excluded from further analysis. Prokaryotic sequence abundances were normalized using the median abundance of 10 single-copy marker genes/transcripts[75]. This gave abundance as a percentage of single-copy gene abundance (equivalent to the percentage of cells containing a copy) in the metagenome, and transcription as a percentage of single-copy gene transcription in the metatranscriptome. These marker gene/transcript abundances were downloaded from the Ocean Gene Atlas using the hmm profiles developed by Milanese et al.[75] with an *e*-value threshold of 1e–80. The MATOU_v1_metaG (metagenomic) database featured few MIX and MES sampling sites (2 and 7, respectively), limiting the power of comparative analysis between sampling depths, thus, these sites were excluded from the analysis. Statistical analysis was performed in R (version 4.02) using RStudio.

ACKNOWLEDGMENTS

We thank the staff from BL18U1&BL19U1 beamlines of the National Facility for Protein Sciences Shanghai and Shanghai Synchrotron Radiation Facility, for assistance during data collection. This study was supported by: the National Science Foundation of China (Grant Nos. 91851205, 42076229, 31961133016), the National Key Research and Development Program of China (Grant No. 2021YFA0909600), the Fundamental Research Funds for the Central Universities (Grant Nos. 202172002, 202041011), the Major Scientific and Technological Innovation Project of Shandong Province (Grant No. 2019JZZY010817), the Program of Shandong for Taishan Scholars (Grant No. tspd20181203), the grant of Laboratory for Marine Biology and Biotechnology (OF2019NO02), Pilot National Laboratory for Marine Science and Technology (Qingdao), the United Kingdom's Natural and Environmental Research Council (NERC, NE/P012671/1, and NE/N002385/1), and the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC, BB/P006140/1).

AUTHOR CONTRIBUTIONS

Jonathan D. Todd and Yu-Zhong Zhang designed the research. Nick E. Le Brun and Yu-Zhong Zhang directed the research. Chun-Yang Li, Jason C. Crack, Simone Newton-Payne, Andrew R. J. Murphy, Benjamin J. Pinchbeck, and Shun Zhou performed the experiments. Xiu-Lan Chen, Beth T. Williams, Ming Peng, and Yin Chen helped in the data analysis. Chun-Yang Li, Jason C. Crack, Beth T. Williams, Andrew R. J. Murphy, and Xiu-Lan Chen wrote the manuscript. Xiao-Hua Zhang, Yin Chen, and Nick E. Le Brun edited the manuscript.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

DATA AVAILABILITY

The structures of DsyB-SAM complex and DsyB-SAH-MTHB complex have been deposited in the Protein Data Bank under the accession codes 7WDQ and 7WDW, respectively.

SUPPORTING INFORMATION

Additional Supporting Information for this article can be found online at https://doi.org/10.1002/mlf2.12030.

ORCID

Yu-Zhong Zhang http://orcid.org/0000-0002-2017-1005

REFERENCES

1 Gali M, Devred E, Levasseur M, Royer SJ, Babin M. A remote sensing algorithm for planktonic dimethylsulfoniopropionate (DMSP) and an analysis of global patterns. *Remote Sens Environ*. 2015; 171:171–84.

2 Simo R, Archer SD, Pedros-Alio C, Gilpin L, Stelfox-Widdicombe CE. Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnol Oceanogr*. 2002;47:53–61.

3 Zhang XH, Liu J, Liu J, Yang G, Xue CX, Curson ARJ, et al. Biogenic production of DMSP and its degradation to DMS-their roles in the global sulfur cycle. *Sci China Life Sci*. 2019;62:1296–319.

4 Stefels J. Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *J Sea Res*. 2000;43:183–97.

5 Wolfe GV, Steinke M. Grazing-activated production of dimethyl sulfide (DMS) by two clones of *Emiliania huxleyi*. *Limnol Oceanogr*. 1996;41:1151–60.

6 Sunda W, Kieber DJ, Kiene RP, Huntsman S. An antioxidant function for DMSP and DMS in marine algae. *Nature*. 2002;418:317–20.

7 Zheng Y, Wang J, Zhou S, Zhang Y, Liu J, Xue C-X, et al. Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments. *Nat Commun*. 2020;11:4658.

8 Vallina SM, Simó R. Strong relationship between DMS and the solar radiation dose over the global surface ocean. *Science*. 2007; 315:506–8.

9 Curson AR, Todd JD, Sullivan MJ, Johnston AW. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nat Rev Microbiol*. 2011;9:849–59.

10 Curson ARJ, Liu J, Bermejo Martínez A, Green RT, Chan Y, Carrión O, et al. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nat Microbiol*. 2017;2:17009.

11 Williams BT, Cowles K, Bermejo Martínez A, Curson ARJ, Zheng Y, Liu J, et al. Bacteria are important dimethylsulfoniopropionate producers in coastal sediments. *Nat Microbiol*. 2019;4:1815–25.

12 Curson ARJ, Williams BT, Pinchbeck BJ, Sims LP, Martínez AB, Rivera PPL, et al. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. *Nat Microbiol*. 2018;3:430–9.

13 Summers PS, Nolte KD, Cooper AJL, Borgeas H, Leustek T, Rhodes D, et al. Identification and stereospecificity of the first three enzymes of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte alga. *Plant Physiol*. 1998;116:369–78.

14 Gage DA, Rhodes D, Nolte KD, Hicks WA, Leustek T, Cooper AJ, et al. A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature*. 1997;387:891–4.

15 Rousseau H, Rousseau-Gueutin M, Dauvergne X, Boutte J, Simon G, Marnet N, et al. Evolution of DMSP (dimethylsulfoniopropionate) biosynthesis pathway: origin and phylogenetic distribution in polyploid Spartina (*Poaceae, Chloridoideae*). *Mol Phylogenet Evol*. 2017; 114:401–14.

16 Ito T, Asano Y, Tanaka Y, Takabe T. Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (*Chlorophyta*). *J Phycol*. 2011;47:517–23.

17 Kageyama H, Tanaka Y, Shibata A, Waditee-Sirisattha R, Takabe T. Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*: identification of a gene encoding MTHB-methyltransferase. *Arch Biochem Biophys*. 2018;645:100–6.

18 Cantoni GL. Biological methylation: selected aspects. *Annu Rev Biochem*. 1975;44:435–51.

19 Liscombe DK, Louie GV, Noel JP. Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat Prod Rep*. 2012;29:1238–50.

20 O'Hagan D, Schmidberger JW. Enzymes that catalyse S(N)2 reaction mechanisms. *Nat Prod Rep*. 2010;27:900–18.

21 Dunbar KL, Scharf DH, Litomska A, Hertweck C. Enzymatic carbon-sulfur bond formation in natural product biosynthesis. *Chem Rev*. 2017;117:5521–77.

22 Urios L, Michotey V, Intertaglia L, Lesongeur F, Lebaron P. *Nisaea denitrificans* gen. nov., sp nov and *Nisaea nitritireducens* sp nov., two novel members of the class Alphaproteobacteria from the Mediterranean Sea. *Int J Syst Evol Micr*. 2008;58:2336–41.

23 Meinnel T, Mechulam Y, Blanquet S. Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*. *Biochimie*. 1993;75:1061–75.

24 Pellicer Martinez MT, Crack JC, Stewart MY, Bradley JM, Svistunenko DA, Johnston AW, et al. Mechanisms of iron- and O_2-sensing by the [4Fe-4S] cluster of the global iron regulator RirA. *eLife*. 2019;8:e47804.

25 Volbeda A, Dodd EL, Darnault C, Crack JC, Renoux O, Hutchings MI, et al. Crystal structures of the NO sensor NsrR reveal how its iron-sulfur cluster modulates DNA binding. *Nat Commun*. 2017;8:15052.

26 Crack JC, Thomson AJ, Le Brun NE. Mass spectrometric identification of intermediates in the O_2-driven [4Fe-4S] to [2Fe-2S] cluster conversion in FNR. *Proc Natl Acad Sci USA*. 2017;114:E3215–E23.

27 Chen SC, Huang CH, Lai SJ, Liu JS, Fu PK, Tseng ST, et al. Structure and mechanism of an antibiotics-synthesizing 3-hydroxykynurenine *C*-methyltransferase. *Sci Rep*. 2015;5:10100.

28 Martin JL, McMillan FM. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol*. 2002;12:783–93.

29 Hotta K, Keegan RM, Ranganathan S, Fang M, Bibby J, Winn MD, et al. Conversion of a disulfide bond into a thioacetal group during echinomycin biosynthesis. *Angew Chem Int Edit*. 2014;53:824–8.

30 Peng Y, Feng Q, Wilk D, Adjei AA, Salavaggione OE, Weinshilboum RM, et al. Structural basis of substrate recognition in thiopurine *S*-methyltransferase. *Biochemistry*. 2008;47:6216–25.

31 Coiner H, Schröder G, Wehinger E, Liu CJ, Noel JP, Schwab W, et al. Methylation of sulfhydryl groups: a new function for a family of small molecule plant *O*-methyltransferases. *Plant J*. 2006;46:193–205.

32 Dang HY, Li TG, Chen MN, Huang GQ. Cross-ocean distribution of Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. *Appl Environ Microbiol*. 2008;74:52–60.

33 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296–W303.

34 Trottmann F, Ishida K, Franke J, Stanišić A, Ishida-Ito M, Kries H, et al. Sulfonium acids loaded onto an unusual thiotemplate assembly line construct the cyclopropanol warhead of a Burkholderia virulence factor. *Angew Chem Int Ed Engl*. 2020;59:13511–5.

35 Liu J, Zhang Y, Liu J, Zhong H, Williams BT, Zheng Y, et al. Bacterial dimethylsulfoniopropionate biosynthesis in the East China Sea. *Microorganisms*. 2021;9:657.

36 Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data*. 2018;5:170203.

37 Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 2014;12:e1001889.

38 Bucciarelli E, Ridame C, Sunda WG, Dimier-Hugueney C, Cheize M, Belviso S. Increased intracellular concentrations of DMSP and DMSO

in iron-limited oceanic phytoplankton *Thalassiosira oceanica* and *Trichodesmium erythraeum*. *Limnol Oceanogr*. 2013;58:1667–79.

39 Keller MD, Bellows WK, Guillard RRL. Dimethyl sulfide production in marine phytoplankton. In: Saltzman ES, Cooper WJ, editors. Biogenic sulfur in the environment. Washington, DC: ACS Publications; 1989. p. 167–82.

40 Keller MD, Korjeff-Bellows W. Physiological aspects of the production of dimeyhtlsulfoniopropionate (DMSP) by marine phytoplankton. In: Kiene RP, Visscher PT, Keller MD, Kirst GO, editors. Biological and environmental chemistry of DMSP and related sulfonium compounds. Boston, MA: Springer US; 1996. p. 131–42.

41 Meng A, Corre E, Probert I, Gutierrez-Rodriguez A, Siano R, Annamale A, et al. Analysis of the genomic basis of functional diversity in dinoflagellates using a transcriptome-based sequence similarity network. *Mol Ecol*. 2018;27:2365–80.

42 Caruana AMN, Le Gac M, Hervé F, Rovillon GA, Geffroy S, Malo F, et al. *Alexandrium pacificum* and *Alexandrium minutum*: harmful or environmentally friendly? *Mar Environ Res*. 2020;160:105014.

43 Wolfe GV, Strom SL, Holmes JL, Radzio T, Olson MB. Dimethylsulfoniopropionate cleavage by marine phytoplankton in response to mechanical, chemical, or dark stress. *J Phycol*. 2002;38:948–60.

44 Royer C, Borges AV, Lapeyra Martin J, Gypens N. Drivers of the variability of dimethylsulfoniopropionate (DMSP) and dimethylsulfoxide (DMSO) in the Southern North Sea. *Cont Shelf Res*. 2021;216:104360.

45 Caruana AMN, Steinke M, Turner SM, Malin G. Concentrations of dimethylsulphoniopropionate and activities of dimethylsulphide-producing enzymes in batch cultures of nine dinoflagellate species. *Biogeochemistry*. 2012;110:87–107.

46 Caruana AMN, Malin G. The variability in DMSP content and DMSP lyase activity in marine dinoflagellates. *Prog Oceanogr*. 2014;120:410–24.

47 Gutierrez-Rodriguez A, Pillet L, Biard T, Said-Ahmad W, Amrani A, Simó R, et al. Dimethylated sulfur compounds in symbiotic protists: a potentially significant source for marine DMS(P). *Limnol Oceanogr*. 2017;62:1139–54.

48 Yost DM, Mitchelmore CL. Dimethylsulfoniopropionate (DMSP) lyase activity in different strains of the symbiotic alga *Symbiodinium microadriaticum*. *Mar Ecol Prog Ser*. 2009;386:61–70.

49 Franklin DJ, Steinke M, Young J, Probert I, Malin G. Dimethylsulphoniopropionate (DMSP), DMSP-lyase activity (DLA) and dimethylsulphide (DMS) in 10 species of coccolithophore. *Mar Ecol Prog Ser*. 2010;410:13–23.

50 Steinke M, Wolfe GV, Kirst GO. Partial characterisation of dimethylsulfoniopropionate (DMSP) lyase isozymes in 6 strains of *Emiliania huxleyi*. *Mar Ecol Prog Ser*. 1998;175:215–25.

51 Spielmeyer A, Gebser B, Pohnert G. Dimethylsulfide sources from microalgae: improvement and application of a derivatization-based method for the determination of dimethylsulfoniopropionate and other zwitterionic osmolytes in phytoplankton. *Mar Chem*. 2011;124:48–56.

52 Kinsey JD, Kieber DJ, Neale PJ. Effects of iron limitation and UV radiation on *Phaeocystis antarctica*growth and dimethylsulfoniopropionate, dimethylsulfoxide and acrylate concentrations. *Environ Chem*. 2016;13:195–211.

53 Jean N, Bogé G, Jamet J-L, Richard S, Jamet D. Annual contribution of different plankton size classes to particulate dimethylsulfoniopropionate in a marine perturbed ecosystem. *J Mar Syst*. 2005;53:235–47.

54 Beringer JER. Factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol*. 1974;84:188–98.

55 Gonzalez JM, Whitman WB, Hodson RE, Moran MA. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. *Appl Environ Microbiol*. 1996;62:4433–40.

56 Baumann P, Baumann L. The prokaryotes: a handbook on habitats, isolation and identification of bacteria. 1st ed. Germany: Springer-Verlag Berlin Heidelberg; 1981.

57 Figurski DH, Helinski DR. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci USA*. 1979;76:1648–52.

58 Carrión O, Curson ARJ, Kumaresan D, Fu Y, Lang AS, Mercadé E, et al. A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nat Commun*. 2015;6:6579.

59 Tett AJ, Rudder SJ, Bourdes A, Karunakaran R, Poole PS. Regulatable vectors for environmental gene expression in Alphaproteobacteria. *Appl Environ Microbiol*. 2012;78:7137–40.

60 Lalitha S. Primer Premier 5. *Biotech Softw Internet Rep*. 2000;1:270–2.

61 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29:e45.

62 Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal Bioanal Chem*. 2012;402:231–47.

63 Tang DQ, Zou L, Yin XX, Ong CN. HILIC-MS for metabolomics: an attractive and complementary approach to RPLC-MS. *Mass Spectrom Rev*. 2016;35:574–600.

64 Heck AJR. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*. 2008;5:927–33.

65 Leney AC, Heck AJR. Native mass spectrometry: what is in the name? *J Am Soc Mass Spectr*. 2017;28:5–13.

66 Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531–52.

67 Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*. 2006;62:859–66.

68 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011;67:235–42.

69 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66:486–501.

70 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66:213–21.

71 Liao CS, Seebeck FP. *In vitro* reconstitution of bacterial DMSP biosynthesis. *Angew Chem Int Edit*. 2019;58:3553–6.

72 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.

73 Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37.

74 Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res*. 2018;46:W289–95.

75 Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun*. 2019;10:1014.

How to cite this article: Li C-Y, Crack JC, Newton-Payne S, Murphy ARJ, Chen X-L, Pinchbeck BJ, et al. Mechanistic insights into the key marine dimethylsulfoniopropionate synthesis enzyme DsyB/DSYB. *mLife*. 2022. doi:10.1002/mlf2.12030

© 2022 The Authors. *mLife* published by John Wiley & Sons Australia, Ltd. on behalf of Institute of Microbiology, Chinese Academy of Sciences.