Original citation:

Permanent WRAP URL:
http://wrap.warwick.ac.uk/67990

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher’s statement:
"This is the peer reviewed version of the following article: which has been published in final form at http://dx.doi.org/10.1111/1462-2920.12896 This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher’s version. Please see the ‘permanent WRAP url’ above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
**XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments**

Martin Taubert¹, Carolina Grob¹, Alexandra M. Howat¹, Oliver J. Burns², Joanna L. Dixon³, Yin Chen⁴, and J. Colin Murrell¹

¹School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK
²School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK
³Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK
⁴School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK

Corresponding author:
J. Colin Murrell
School of Environmental Sciences
University of East Anglia
Norwich, NR4 7TJ, UK
E mail: j.c.murrell@uea.ac.uk
Tel: +44 1603 592959 (Office); Fax: +44 1603 591327

Running title: “xoxF in costal marine environments”
Abstract

The \textit{xoxF} gene, encoding a pyrroloquinoline quinone-dependent methanol dehydrogenase, is found in all known proteobacterial methylotrophs. In several newly discovered methylotrophs, XoxF is the active methanol dehydrogenase, catalysing the oxidation of methanol to formaldehyde. Apart from that, its potential role in methylotrophy and carbon cycling is unknown. So far, the diversity of \textit{xoxF} in the environment has received little attention. We designed PCR primer sets targeting clades of the \textit{xoxF} gene, and used 454 pyrosequencing of PCR amplicons obtained from DNA of four coastal marine environments for a unique assessment of the diversity of \textit{xoxF} in these habitats. Phylogenetic analysis of the data obtained revealed a high diversity of \textit{xoxF} genes from two of the investigated clades, and substantial differences in sequence composition between environments. Sequences were classified as being related to a wide range of both methylotrophs and non-methylotrophs from Alpha-, Beta- and Gammaproteobacteria. The most prominent sequences detected were related to the family Rhodobacteraceae, the genus \textit{Methylotenera} and the OM43 clade of Methylophilales, and are thus related to organisms that employ XoxF for methanol oxidation. Furthermore, our analyses revealed a high degree of so far undescribed sequences, suggesting a high number of unknown species in these habitats.

Introduction

Methylotrophs are organisms which can use reduced organic compounds with no carbon-carbon bonds, such as methane, methanol, or methylamine, as their sole source of carbon and energy (Anthony, 1982; Chistoserdova, 2011). There are over 200 described species of methylotrophs belonging mostly to the Alpha-, Beta-, and Gammaproteobacteria, but also to Verrucomicrobia, Bacteroidetes, Firmicutes, and Actinobacteria (Madhaiyan et al., 2010; Kolb and Stacheter, 2013). The majority of methylotrophs are aerobic bacteria, and almost all are able to utilise methanol (Anthony, 1982; Chistoserdova, 2011). The catalysis of methanol to formaldehyde requires a methanol dehydrogenase (MDH). Cultured gram negative methylotrophs usually use a periplasmic pyrroloquinoline quinone (PQQ)-dependent MDH, which is an \(\alpha_2\beta_2\) tetramer of MxaF and Mxal (Anthony, 1986; Duine et al., 1986).
The classic MDH encoded by mxaF and mxaI is widespread in bacteria that grow on methanol and was shown initially in *Methylobacterium* sp. M27 (previously named *Pseudomonas* sp. M27) (Anthony and Zatman, 1964, 1965). It was later extensively studied in a close relative, *Methylobacterium extorquens* AM1, a key model bacterium for methylotrophy (Nunn and Lidstrom, 1986). Another PQQ-dependent dehydrogenase, MDH2, encodes an MDH in some organisms such as *Methylloversatilis* sp. (initially classified as Burkholderiales bacteria), but appears to be much less widespread in the environment (Kalyuzhnaya et al., 2008a). In the past two decades, a homolog of the mxaF gene, xoxF, has been implicated in one-carbon compound (C1) metabolism (Harms et al., 1996; Chistoserdova and Lidstrom, 1997). In contrast to the calcium containing MxaF, XoxF seems to require rare earth elements (REE) like lanthanum or cerium for activity (Keltjens et al., 2014). XoxF is present in all known gram-negative methylotrophs to date (Chistoserdova et al., 2009; Chistoserdova, 2011). Several organisms not described as methylotrophs, such as certain Rhizobiales and Burkholderiales, as well as some Aquificales and Acidobacteria, also have xoxF present in their genomes (Chistoserdova, 2011). Phylogenetic analysis has revealed the presence of five distinct clades of the xoxF gene, named xoxF1 - 5, and many organisms contain several different copies of it (Chistoserdova, 2011; Keltjens et al., 2014).

Some methylotrophs contain only xoxF and no other MDH encoding gene, as for instance *Rhodobacter* sp. (Wilson et al., 2008), *Beggiatoa alba* (Jewell et al., 2008) and the methanotroph *Methylacidiphilum fumariolicum* SolV (Pol et al., 2014). Likewise, *Methylotenera mobilis*, a major species that oxidises methanol in freshwater lake sediment (Kalyuohnaya et al., 2009), contains two xoxF4 genes but no mxaF. The genome of the methylotroph *Methylphilales bacterium* HTCC2181 also contains xoxF4 as the only putative MDH. Strain HTCC2181 is a representative of the OM43 clade, and one of the most abundant marine Betaproteobacteria that uses methanol as growth substrate and energy source (Giovanoni et al., 2008). XoxF deletion mutants of *M. mobilis* (xoxF4) and *R. sphaeroides* (xoxF5) no longer metabolised methanol, strongly suggesting a role for XoxF as the functional MDH in these organisms (Wilson et al., 2008; Mustakhimov et al., 2013). In *M. fumariolicum* SolV, XoxF2 catalyses the oxidation of methanol to formate instead of formaldehyde (Pol et al., 2014). This organism lacks any other pathway for formaldehyde oxidation, and fixes carbon dioxide via the Calvin-Benson-Bassham cycle (Khadem et al., 2011).

Although its true role is not fully understood, xoxF along with its product has been shown to be more abundant than its mxaF counterpart in different environments (Kalyuzhnaya et al., 2008b). For example, XoxF
is highly abundant in the phyllosphere of soybean, clover and Arabidopsis (Delmotte et al., 2009). Moreover, high expression of XoxF-like proteins has been found in coastal oceanic microbial plankton (Sowell et al., 2011). Transcriptomics studies further suggested varying roles for different xoxF homologues in Methylophilacea (Vorobev et al., 2013). Its presence in every known methylotroph and across a range of environments suggests a high ecological importance. This is especially true in coastal marine and other aquatic habitats, where different methylotrophs employing XoxF for the oxidation of methanol have been found, and where REE are available from sediments or coastal runoff (Elderfield et al., 1990).

This study is the first targeted approach to investigate xoxF diversity in marine environments. XoxF-specific PCR primer sets were designed and used in combination with 454 amplicon pyrosequencing to obtain xoxF gene sequence datasets from environmental DNA. Phylogenetic analysis of xoxF gene sequences retrieved from four different marine environments was performed with reference to a new xoxF gene database in order to affiliate them to putative methylotrophs.

**Results & Discussion**

*Classification of genomic xoxF and primer design*

A comprehensive database of xoxF gene sequences was made from bacterial genomes available in the NCBI nucleotide database. The sequences were aligned and clustered based on phylogenetic analysis to detect regions suitable for primer design. A total of 388 genes of putative PQQ-dependent dehydrogenase from 101 methylotrophic and non-methylotrophic bacterial organisms were investigated, resulting in 147 xoxF gene sequences being identified. The phylogenetic analysis resolved xoxF into five clades (see Figure 1, Supplementary Figure 1 for full tree), as described previously (Chistoserdova, 2011; Keltjens et al., 2014). Some xoxF genes were outliers that did not fall in a particular clade. As xoxF sequences from the different clades show low identity, it was not possible to design a set of PCR primers that could amplify all xoxF genes. Hence, clade-specific primer sets were designed (see Table 1). Sequences were considered to be detected by a primer set if at most one mismatch per primer was present.

*Characterisation of xoxF clades and theoretical primer coverage*

Only a limited number of genes were found in the NCBI nucleotide database belonging to the xoxF1 and xoxF2 clades (6 and 5 genes, respectively). Genes of the xoxF1 clade were found in Xanthomonas species.
(Gammaproteobacteria), some Beijerinckiaceae (Alphaproteobacteria) and the methanotroph Candidatus
*Methylomirabilis oxyfera* (Ettwig et al., 2010). *Xanthomonas* species are typical terrestrial plant pathogens and
thus are not expected to play a major role in marine environments. The Beijerinckiaceae include
methylo trophs which usually also contain *xoxF3* and *xoxF5*. Genes of the *xoxF2* clade have been found in
*Methylacidiphilum* species (Verrucomicrobia) that have been isolated from extreme environments such as
volcanic mudpots (Dunfield et al., 2007; Islam et al., 2008; Pol et al., 2014). The XoxF2 of *Methylacidiphilum
fumariolicum* SolV has been shown to catalyse the oxidation of methanol to formate (Pol et al., 2014).
Furthermore, *xoxF2* genes are present in the thermophilic bacterium *Hydrogenobacter thermophilus* TK-6 and
also in Candidatus *M. oxyfera*, isolated from an anoxic enrichment culture obtained from a Dutch drainage
ditch. Thus, the clades *xoxF1* and *xoxF2* are probably only of limited interest for this study of the marine
environment. The primer sets nevertheless cover all known members of the respective clades with the
exception of the *H. thermophilus* TK-6 gene, which is somewhat divergent from other *xoxF2*.
The diverse *xoxF3* clade is dominated by alphaproteobacterial sequences, mostly from Rhizobiales, but also
contains *xoxF* from Betaproteobacteria (*Variovorax paradoxus*, *Methylobacillus flagellatus*) and
Gammaproteobacteria (*Methylobacter marinus*) and a *xoxF* gene from Candidatus *Solibacter usitatus*
(Actiobacteria). Many of these organisms are known methylo trophs which might play a role in C1 cycling in
marine environments, and most of them additionally contain *xoxF* from clades 4 and 5. The primer set covers
all 14 *xoxF3* genes in the used *xoxF* database. A more dissimilar *xoxF* from the methylo troph *Methylosinus
trichosporium* OB3b (Alphaproteobacteria, see Figure 1) is not covered, however, this organism also possesses
several *xoxF5* genes which are covered by the respective primer set (see below). A comparable number of
sequences was found in clade *xoxF4* (17 genes), which is specific for the family Methylophilaceae, also
encompassing beta proteobacterium KB13 and M ethylophilales bacterium HTCC2181 of the OM43 clade.
Many of these species are known methylo trophs and are commonly found in coastal and fresh water
environments, and XoxF4 has been shown to be the only functional methanol dehydrogenase in some of them
(Rappe et al., 2000; Giovannoni et al., 2008; Kalyuzhnaya et al., 2009), indicating a potentially important role
for *xoxF4* in C1 metabolism. The majority of *xoxF* sequences analysed here belongs to clade *xoxF5* (102 genes).
It contains various Alpha-, Beta- and Gammaproteobacteria, including many marine methylo trophs, such as
*Methylophaga* sp. (Neufeld et al., 2007; Neufeld et al., 2008) and *Methylotersatilis* sp. (Kalyuzhnaya et al.,
2008a). The relatedness of *xoxF5* sequences does in most cases follow 16S rRNA gene phylogeny. However,
the similarity between alpha- and betaproteobacterial sequences is relatively high, which leads to a less robust classification on class and order level than on lower taxonomic levels, as indicated by lower bootstrap values (see Supplementary Figure 1). The presence of multiple, divergent gene copies in *Methylocella* and *Xanthobacter* might be a hint at occurrences of horizontal gene transfer. In some organisms, such as *Rhodobacter sphaeroides* and *Beggiatoa alba*, XoxF5 is present as the only functional methanol dehydrogenase (Jewell et al., 2008; Wilson et al., 2008). In *Methylobacterium extorquens* AM1, xoxF5 is required for expression of the methanol dehydrogenase MxaF1 (Skovran et al., 2011). Given the high diversity and widespread appearance of xoxF5, the function of these genes cannot be generalised, and involvement of the gene in further processes cannot be excluded.

All known methylotrophs that possess the classic methanol dehydrogenase gene *mxaF* additionally have at least one xoxF, typically from clade xoxF4 or xoxF5. Genes of these two clades also have been previously detected in metagenomic sequences obtained from marine samples (Gilbert et al., 2010). As xoxF4 and xoxF5 are most likely to play a role in marine C1 cycling, and all representative xoxF genes with a demonstrated function in C1 metabolism belong to these clades (with the exception of the xoxF2 of *Methylacidiphilum fumariolicum* SolV), they are the focus of this study. All xoxF4 and xoxF5 sequences are covered by the respective primer sets, with the exception of the xoxF of extremophilic *Acidiphilium* species (Alphaproteobacteria), which are only distantly related to xoxF5 (see Figure 1), and not expected to play a role in marine environments.

**Testing of primers with genomic DNA from reference strains**

To confirm specificity of the designed primer sets, PCR assays with genomic DNA of reference strains containing different xoxF genes were performed. For xoxF1 and xoxF3, DNA from *Methylocella silvestris* BL2, *Methylokerula stellata* AR4 and *Methyllobacillus flagellatus* KT (xoxF3 only) was used. For xoxF4, DNA from *M. flagellatus* KT and *Methylotenera mobilis* JLW8 was used. For xoxF5, several reference strains were available: *Methylosinus trichosporium* OB3b, *M. silvestris* BL2, *Sagittula stellata* E-37, *Methylococcus capsulatus* Bath, *Methylphaga marina* DSM 5689, *M. stellata* AR4 and *Roseobacter denitrificans* OCh 114. Unfortunately, no reference strains were available that possess xoxF2 genes. Interestingly, the primer set was used to retrieve a xoxF2 sequence related to Verrucomicrobia from DNA of a soil enrichment culture (unpublished data). PCR products with reference DNA were obtained for xoxF1 and xoxF3 to xoxF5 (see Supplementary Figure 2).
identity of all PCR products was confirmed by Sanger sequencing of clone libraries. Little or no cross-specificity or unspecific products were observed. The only exceptions were the xoxF1 primer set, which produced non-specific bands with some of the strains used, and the xoxF4 primer set, which also amplified several xoxF5 genes. Products of the latter were clearly distinguishable from genuine xoxF4 amplicons on agarose gels due to a smaller size, as xoxF5 genes have an 84 bp deletion compared to xoxF4 in the region targeted by the xoxF4 primer set. Reinvestigation of the primer binding sites revealed that 21 and 12 of the 102 xoxF5 genes had one or no mismatch with the forward and reverse primer, respectively. The cross-specificity could not be narrowed down to a particular group of xoxF5 sequences. On average, both forward and reverse primer had 2.4 mismatches per xoxF5 gene as opposed to 0.24 (fwd) and 0.47 (rev) mismatches per xoxF4 gene. As no alternative regions conserved in all available xoxF4 genes were found, the primer set was further used to test whether the cross-specificity would be of relevance when investigating environmental DNA.

Detection of xoxF in environmental DNA by PCR assays

To test the PCR primer sets as an assay for xoxF diversity in environmental habitats, water samples were collected in four coastal marine environments around the UK, including the Western Channel Observatory Station L4 (L4; salinity ~35, surface sample, water column depth ~50 m); Stiffkey Salt Marsh (SM; salinity ~30, sample taken from aqueous layer above sediment, high turbidity due to sediment resuspension), Cromer Beach (CB; surface sample, water column only a few meters deep) and offshore of Lowestoft (LO; bottom of a water column only a few meters deep). It has previously been shown that methanol concentrations in surface seawater at L4 and across the Atlantic ocean range between 34-97 nM (Beale et al., 2015). Algeal growth and decay, atmospheric influx, precipitation and methane oxidation (most likely in SM) have been discussed as potential sources of methanol, but their contribution to the overall methanol budget still has to be elucidated (Felix et al., 2014; Beale et al., 2015). Methylotrophic bacteria have been shown to actively take up this methanol and using it as carbon and energy sources at rates of 2-146 nmol l⁻¹ d⁻¹ (Dixon et al., 2011). L4, SM, CB and LO sites were therefore chosen as representatives of different marine environments with high potential for XoxF-mediated methanol oxidation.

DNA extracted from these samples was used as template for PCR reactions. Sanger sequencing was done to verify gene identity, and a subset of samples was selected for analysis by 454 pyrosequencing to investigate diversity of xoxF genes. PCR products were obtained with primer sets targeting xoxF5 and xoxF4, with the
latter being detected in L4 and SM but not in CB or LO (see Supplementary Figure 3). XoxF4 PCR products showed the expected size, and in some samples, a weaker, additional band at a slightly lower size, corresponding to that of the xoxF5 gene, was visible. Sanger sequencing confirmed the presence of xoxF4, with only a low abundance of xoxF5 products. Interestingly, with CB and LO DNA, where no xoxF4 product was obtained, also no amplification of xoxF5 was observed with the xoxF4 primer set. As xoxF5 genes were detected in these samples with the xoxF5 primer set, it can be excluded that the presence of xoxF5 might prevent amplification of the xoxF4 genes. Thus, the cross-specificity observed on reference DNA was confirmed, but was not considered to be a major problem for this study. PCR assays targeting xoxF1, xoxF2 and xoxF3 only produced very faint bands with environmental DNA (see Supplementary Figure 3), and sequencing of these bands did not reveal any xoxF products. Hence, only xoxF4 and xoxF5 genes seem to have been present in sufficient abundance in the investigated marine samples to be detected by PCR. The corresponding amplicons were selected for 454 pyrosequencing.

454 Pyrosequencing of xoxF4 and xoxF5 amplicons

For analysis of xoxF4 and xoxF5 diversity, sequences were extracted from raw 454 amplicon pyrosequencing data, quality filtered, trimmed and binned to OTUs. Verification of gene identity and phylogenetic analysis was performed by alignment with reference sequences followed by construction of neighbor joining (NJ) and maximum-likelihood (ML) phylogenetic trees. OTUs that did not belong to the respectively targeted clade were discarded from further analysis. For xoxF5 amplicons, this involved less than 1% of the sequences. For xoxF4 amplicons, which showed cross-specificity to xoxF5 as described above, less than 10% of the sequences were excluded for the L4 amplicon, but almost 40% of the sequences for the SM amplicon. The majority of the excluded sequences were xoxF5, with some additional mxaF sequences. Based on the results obtained using reference DNA, this problem was expected, but the remaining xoxF4 sequences still provided a satisfactory basis for further analysis. An overview of the number of obtained sequences and OTUs is given in Supplementary Table 1.

Phylogenetic analysis of xoxF4 genes

Genes of the xoxF4 clade are specific to the family Methylophilaceae of the Betaproteobacteria. XoxF4 was only detected in DNA from L4 and SM, possibly indicating the absence or very low abundance of Methylophilaceae in the other two environments investigated. Overall, xoxF4 diversity was relatively low: one
major phylogenetic group of xoxF4 sequences was detected in each of the environments, with only a few other xoxF4 genes being present (Figure 2). In DNA from L4, this major group was most closely related to xoxF4 of Methylophilales bacterium HTCC2181 of the OM43 clade. Members of the OM43 clade are known to be abundant methylotrophs in coastal waters (Rappe et al., 2000), and related XoxF4 proteins have been detected in metaproteomes of coastal surface waters (Sowell et al., 2011; Williams et al., 2012). In DNA from SM, the major phylogenetic group was represented by three OTUs that were most closely related to xoxF4 from Methylophilera sp. Different Methylophilera mobilis strains have recently been shown to have a highly diverse physiology, with some possessing the mxaF gene and using MxaFI as methanol dehydrogenase, others lacking this gene and employing XoxF (Mustakhimov et al., 2013; Beck et al., 2014). In environmental studies (Kalyuzhnaya et al., 2008b) and in microcosm experiments (Beck et al., 2013), Methylophilaceae populations were dominated by those only possessing xoxF. Interestingly, xoxF4 transcripts related to Methylophilera sp. were previously also found in metatranscriptomes from L4 (Gilbert et al., 2010). The xoxF4 sequences recovered from L4 and SM are highly similar to genes from organisms that employ XoxF for methanol oxidation, which suggests a possible role for xoxF4 in C1 metabolism in these environments. However, the additional presence of mxaF in the organisms detected cannot be excluded, as only relatedness to, but not identity with the reference sequences can be assessed.

**Phylogenetic analysis of xoxF5 genes**

The clade xoxF5 comprises the majority of known xoxF sequences and is present in a wide range of Proteobacteria. A high diversity of xoxF5 OTUs, covering almost all major phylogenetic groups, was observed (see Supplementary Figure 4 and 5). Large differences between xoxF5 gene distributions in the investigated marine environments at the class level were detected (see Figure 3a). While in DNA samples from L4 and SM, xoxF5 sequences from Alphaproteobacteria were by far the dominant OTUs, LO and CB samples revealed a more diverse distribution of xoxF5 genes from the Alpha-, Beta- and Gammaproteobacteria. Further differences were observed at lower taxonomic levels (see Figure 3b-d), but some subgroups of xoxF5 sequences were also present in all four environments. This includes sequences that were similar to xoxF5 of bacteria of the family Rhodobacteraceae, which contains methylotrophs such as *Rhodobacter*, *Roseobacter*, *Roseovarius* and *Sagittula* (Gonzalez et al., 1997; Barber and Donohue, 1998). Most of these are able to metabolise methanol or other C1 compounds despite not containing the mxaFI genes encoding the classic
methanol dehydrogenase (Wilson et al., 2008; Boden et al., 2011). In metatranscriptomes from coastal water of the North Sea, members of the Rhodobacteraceae showed high metabolic activity levels during algal bloom, indicating an important ecological role (Klindworth et al., 2014). However, very little has been reported on expression of xoxF5 in coastal environments.

In DNA samples from L4, xoxF5 related to genes of another methylotroph that lacks mxaFI, Beggiatoa sp. (Jewell et al., 2008), was detected. Other major groups of xoxF5 sequences showed similarity to xoxF genes of the genera Azoarcus (LO) and Azospirillum (CB), for which no methylotrophic representatives have been described so far. Additionally, xoxF sequences related to genes of methylotrophs that use the classical MxaFI for methanol oxidation were also detected: in L4, xoxF5 related to Methylocystaceae sequences were found, while in SM, a xoxF5 related to Methylophaga sp. sequences was detected. In CB, xoxF5 were found that were similar to Hyphomicrobium sp. genes. This genus has previously been identified in marine environments (Dixon et al., 2013) and includes representatives that possess MxaFI.

Several xoxF5 OTUs could not be classified, with increasing numbers at lower taxonomic levels. This was especially the case in DNA samples from CB and SM. The classification of the sequences obtained strongly depends on the availability of corresponding reference genomes. If identities to the reference sequences were low, OTUs could only be classified at higher taxonomic levels. Though reproducible with both NJ and ML clustering, the classification of xoxF5 sequences is not very reliable, as indicated by the low bootstrap values in the phylogenetic trees (see Supplementary Figure 4 and 5). Nevertheless, the detection of unclassified xoxF genes indicates the presence of novel bacterial families that have no sequenced representatives, and demonstrates the high, yet uncovered diversity and distribution of this gene in coastal marine environments.

**Conclusions**

It was previously shown that methylotrophs relying on XoxF for methanol oxidation are highly abundant in different marine environments (Giovannoni et al., 2008; Wilson et al., 2008; Lapidus et al., 2011; Mustakhimov et al., 2013). The present investigation revealed an exceedingly high diversity and widespread appearance of xoxF genes in coastal marine habitats. Several groups of the detected xoxF sequences were highly similar to genes of those organisms where xoxF has been previously implicated in C1 metabolism. Other abundant OTUs were related to organisms thus far not described as methylotrophs, suggesting either different functional roles for xoxF outside of methylotrophy or an underestimation of the metabolic potential of these organisms.
Moreover, a high number of OTUs obtained could not be classified, especially at lower taxonomic levels, due to the lack of reference sequences available. Thus, xoxF is widely distributed in microbial genomes, presumably also in various organisms where yet no genome sequences are available. Although the actual function(s) and ecological implications of xoxF genes in the marine environment remain unclear, the results presented here suggest a widespread role in methanol cycling. The differences observed between the four coastal sites indicate that the prevailing environmental conditions could be important in determining the observed diversity of the xoxF gene. The different detected organisms are potentially occupying various ecological niches: They might be associated with other (micro)organisms, or their activity might be subjected to seasonal differences in the environmental conditions. It is also unknown if all discovered xoxF genes encode functional methanol dehydrogenases. There is evidence that XoxF is also involved in other processes in C1 metabolism or has regulatory functions (Skovran et al., 2011; Pol et al., 2014). Further efforts are needed to establish xoxF function(s) in the environment. Also the requirement of xoxF for Lanthanides, which are typically only present in the pico- to nanomolar range in surface water, and thus might require dedicated uptake systems to be available for the microorganisms, warrants further investigation (Keltjens et al., 2014).

The PCR primer sets designed amplify a wide range of different xoxF4 and xoxF5 sequences. However, we cannot exclude that bias is introduced by this PCR-based approach or that genes present in low abundance are missed. Nevertheless, these new xoxF PCR assays have provided highly relevant data about the diversity of xoxF in marine environments and thus present a valuable tool for further investigations on the distribution and significance of xoxF.

**Experimental Procedures**

*xoxF database construction and primer design*

A comprehensive database of xoxF gene sequences was built by investigating genomes and shotgun genomes of methylotrophs and non-methylotrophs within the nucleotide database of the National Center for Biotechnology Information (NCBI, Bethesda MD, USA; http://www.ncbi.nlm.nih.gov/nuccore) for genes encoding PQQ-dependent dehydrogenases. For closely related organisms, only a few representatives were included. Candidate genes were identified using the Basic Local Alignment Search Tool (BLAST) (Altschul et al.,
1997) with reference sequences of known genes for \( \text{xoxF} \) and other PQQ-dependent dehydrogenases as queries. Some partial \( \text{xoxF} \) genes from organisms of interest were also included in the database if no full length sequence was available.

Sequences were translated to amino acids and aligned in MEGA (v6.06) (Tamura et al., 2013) using the MUSCLE algorithm (Edgar, 2004). Phylogenetic analysis was carried out at the nucleic acid level for aligned sequences. Phylogenetic trees were constructed by neighbour-joining (NJ) and maximum likelihood (ML) clustering methods, using the maximum composite likelihood method (Tamura et al., 2004) and the Tamura-Nei model (Tamura and Nei, 1993) to infer evolutionary distances, respectively. To provide confidence estimates for tree topology, the Bootstrap method with 500 replications was used. For missing data/gaps, pairwise deletion and partial deletion with 95 % cutoff was selected for NJ and ML trees, respectively. Phylogenetic classification was compared between both methods and genes were grouped into several clades based on this comparison.

Aligned \( \text{xoxF} \) DNA sequences were inspected for conserved regions using MEGA. Consensus sequences of conserved regions were used for primer design, allowing a maximum of one mismatch to a particular gene sequence, a maximum of four degenerated bases per primer, and at least two of the outmost five nucleotides on each side of a primer being G/C. Primer candidates were further analysed to exclude hairpin formation, self- and cross-complementarity, using the tools Multiple primer analyzer (http://www.thermoscientificbio.com/webtools/multipleprimer/) and OligoCalc (Kibbe, 2007) (http://www.basic.northwestern.edu/biotools/oligocalc.html).

**Environmental sampling and DNA extraction**

Environmental samples from four different marine and coastal sites were used in this study: (L4) Surface water from the Western Channel Observatory station L4 (50°15.0’N; 4°13.0’W) off the coast of Plymouth, UK; (SM) Brackish water from the Stiffkey Salt Marshes at the Northern coast of Norfolk, UK, (52°57’44”N 0°55’27”E), a tidal mud flat environment; (CB) Surface water, 100 m offshore of Cromer Beach on the Northern coast of Norfolk, UK, (52°56’02”N 1°18’04”E); (LO) sea water immediately offshore of the Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK, (52°27’32”N 1°44’23”E). L4, SM and CB were collected in November 2012, LO was collected in January 2013. Approximately 5 to 10 l of water from each environment
were filtered through a 0.22 µm Sterivex™ filter (Merck Millipore, Darmstadt, Germany) and frozen within 24 h of sampling.

DNA was extracted from Sterivex filters using a modified version of the protocol published in (Neufeld et al., 2007). 1.6 ml of SET buffer (0.75 M sucrose, 40 mM EDTA, 50 mM Tris-HCl pH 9) and 0.2 ml 10 % (w/v) SDS were added and the filter was incubated with rotation in a hybridization oven (Hybaid, Waltham, MA, USA) at 55°C for 2 h. Lysates were withdrawn with 5-ml syringes and the filters again incubated with 1 ml of SET buffer and 0.15 ml of SDS solution for 30 min as described. Both lysates were combined in a 15-ml organic solvent resistant tube. Two extractions with phenol:chloroform:isoamyl alcohol (25:24:1) and one with chloroform:isoamyl alcohol (24:1) were performed, using 2 ml of organic solvent each. Finally, 100 µg glycogen (Roche, Basel, Switzerland), 1 ml of 7.5 M ammonium acetate and 8 ml of pure ethanol were added to the aqueous phase, and DNA was precipitated overnight at -20°C. Samples were centrifuged for 30 min at 4 500 x g and the nucleic acid pellets were washed twice with 80 % (v/v) ethanol, dried for 15 min at room temperature, and resuspended in 50 µl of Nuclease-free water. Quality and quantity of the DNA was checked using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Waltham, MA, USA) and a 1 % (w/v) agarose gel.

Extraction of DNA from reference strains was done using a modified version of this protocol, starting with cell pellets instead of Sterivex filters.

**Primer testing with reference strains**

Five candidate sets of xoxF primers targeting different clades were tested for specificity by PCR using genomic DNA of the following bacterial strains known to possess one or several xoxF genes: Methylocella silvestris BL2 (Chen et al., 2010), Methylosinus trichosporium OB3b (Stein et al., 2010), Sagittula stellata E-37 (Gonzalez et al., 1997), Roseobacter denitrificans OCh 114 (Swingley et al., 2007), Methylococcus capsulatus Bath (Ward et al., 2004), Methylophaga marina DSM 5689 (Janvier et al., 1985), Methyloferula stellata AR4 (Vorobev et al., 2011), Methylobacillus flagellatus KT (Chistoserdova et al., 2007) and Methyloferula mobilis JLY8 (Lapidus et al., 2011). Touchdown PCR protocols were used as follows: for xoxF2, xoxF3 and xoxF5, an initial step at 94 °C for 5 min was followed by 11 cycles of 1 min at 94 °C (denaturation), 62 °C to 52 °C, decreasing by 1 °C per cycle, (annealing) and 72 °C (extension) each. This was followed by 25 cycles of 1 min at 94 °C, 52 °C and 72 °C each and a final extension for 10 min at 72 °C. For xoxF1 and xoxF4, a different protocol was used, with the annealing temperature set to 58 °C – 48 °C for the first 11 cycles and to 48 °C for the remaining 25 cycles, and
the addition of 5% DMSO (final concentration) to the reactions in the case of xoxF1. Otherwise the protocol was identical to the one described above. Clone libraries were constructed using the PCR products, and five clones for each strain and primer set were randomly selected for Sanger sequencing (Source BioScience, Nottingham, UK) to check specificity of PCR assays.

Amplification and sequencing of xoxF from environmental DNA

PCR was performed on the extracted environmental DNA as described above. In cases where multiple abundant product bands were observed on a 1% (w/v) agarose gel, the band of the correct size was excised and purified using the GeneJET Gel Extraction Kit (Thermo Scientific, Waltham, MA, USA). Clone libraries were constructed, and 10 random clones from each amplicon were selected for Sanger sequencing as described above, to verify gene identity. These sequences were not used for assessment of gene diversity. For 454 pyrosequencing, amplicons were purified using the GeneJET PCR Purification Kit, followed by quantity and quality control as described above. Attachment of 454 primers and barcodes in a proprietary 4-cycle PCR reaction and subsequent sequencing using 454 pyrosequencing technology on a GS FLX Titanium system (454 Life Sciences, Branford, CT, USA) was done at MR DNA (Molecular Research LP, Shallowater TX, USA, http://www.mrdnalab.com/).

Analysis of xoxF amplicon pyrosequencing data

Pyrosequencing datasets were analysed using the software packages mothur (Schloss et al., 2009) and USEARCH (Edgar, 2013). Mothur was used to extract flowgrams from raw *.sff data files. Flowgrams with less than 450 usable flows were removed, the remaining flowgrams were cut to 720 flows. Flowgrams were denoised and translated to nucleic acid sequences. Sequences with errors in the barcode or primer region were removed, as well as sequences with ambiguous bases or homopolymer runs > 6 bp. Sequences were demultiplexed, barcodes and forward primers removed. Sequences were filtered by length, allowing only sequences between 350 and 550 bp for xoxF4 and between 350 and 390 bp for xoxF5. USEARCH was used for OTU binning (with a 90% identity threshold), chimera removal and singleton removal. The most abundant sequences of each OTU were chosen as representative.

OTUs obtained were aligned in MEGA and phylogenetic analysis was performed as described above. For verification of sequence identity, NJ trees were constructed including a selection of reference sequences from the different clades of PQQ-dependent dehydrogenases. OTU sequences that did not belong to the targeted
oxF clade were removed from the alignment. In a second step, sequences were trimmed to a common length, also removing reverse primer binding regions, and NJ and ML trees were constructed as described above, exclusively including all reference sequences from the targeted clade. Each OTU was classified using the taxonomic identity of the closest reference sequence in both trees, or, if equally related to multiple reference sequences, the lowest common taxonomic level was chosen, i.e., the lowest common branching point in both trees.

**Nucleotide sequence accession numbers**

Nucleotide sequences from 454 amplicon pyrosequencing obtained in this study were deposited in the GenBank nucleotide sequence database under accession numbers KM657613 - KM657640 (L4, xoxF5), KM657589 - KM657602 (L4, xoxF4), KM660746 - KM660788 (SM, xoxF5), KM657603 - KM657612 (SM, xoxF4), KM657493 - KM657573 (CB, xoxF5) and KM660726 - KM660745 (LO, xoxF5). Raw data from 454 amplicon pyrosequencing has been deposited in the Sequence Read Archive (SRA) of NCBI under accession numbers SRR1584508, SRR1584509, SRR1584511 - SRR1584513, and SRR1584515.

**Acknowledgements**

This project is funded by the Gordon and Betty Moore Foundation Marine Microbiology Initiative Grant GBMF3303 to J. Colin Murrell and Yin Chen and through the Earth and Life Systems Alliance, Norwich Research Park, Norwich, UK. We thank the crew of the R.V. Plymouth Quest and scientists sampling on behalf of the Western Channel Observatory (WCO, www.westernchannelobservatory.org.uk). The sampling of station L4 was funded by NERC national capability funding to Plymouth Marine Laboratory for the WCO. We thank Martin Johnson for providing samples from the Centre for Environment, Fisheries and Aquaculture Science (Cefas) in Lowestoft, Peter Dunfield and Angela Smirnova for providing genomic DNA of Methyloferula stellata AR4, Marina Kalyuzhnaya for providing cultures of Methylobacillus flagellatus KT and Methylotenera mobilis JLW8 and Michael Macey for testing of the xoxF2 primers on soil enrichments.

**Conflict of Interest Statement**

The Authors declare no conflict of interest with this manuscript.


**Figure 1** Phylogenetic relationship between the different clades of \( \text{xoxF} \) genes, \( \text{mxaF} \) genes and genes encoding other PQQ-dependent dehydrogenases. Full gene sequences were derived from the NCBI nucleotide database. The tree was constructed using the neighbour-joining method for clustering and the maximum composite likelihood method for computing evolutionary distances. Numbers at branches are bootstrap values of 500 replicates. Scale bar: 1 nucleotide substitution per 10 nucleotides. Major phylogenetic groups within the \( \text{xoxF} \) clades are: \( \text{xoxF1} \) Xanthomonas and Beijerinckiaeae, \( \text{xoxF2} \) Verrucomicrobia, \( \text{xoxF3} \) Rhizobiales, some Beta- and Gammaproteobacteria, \( \text{xoxF4} \) Methylophilaceae, \( \text{xoxF5} \) various Alpha-, Beta- and Gammaproteobacteria.

**Figure 2** Phylogenetic classification of \( \text{xoxF4} \) OTUs from 454 amplicon pyrosequencing obtained from (a) Western Channel Observatory Station L4 and (b) Stiffkey Salt Marsh. Absolute abundance of sequences in each OTU is given as “size”. The total number of sequences is 5,168 and 1,462, respectively. Multiple \( \text{xoxF} \) gene copies in reference strains are numbered in parenthesis. The trees were constructed using the neighbour-joining method for clustering and the maximum composite likelihood method for computing evolutionary distances. Numbers at branches are bootstrap values of 500 replicates. Scale bars: 5 nucleotide substitution per 100 nucleotides. Trees constructed with the maximum likelihood method showed a virtually identical relationship between the sequences and thus are not shown.

**Figure 3** Phylogenetic classification of \( \text{xoxF5} \) sequences retrieved by 454 amplicon pyrosequencing. Abundance of taxonomic groups in the investigated environments is shown at (a) family and (b) genus level. The “unclassified” category contains all sequences that were unclassified at the previous taxonomic level. Data was derived from samples collected at the Western Channel Observatory Station L4 (L4), Stiffkey Salt Marsh (SM), Cromer Beach (CB) and offshore of Lowestoft (LO).