environmental microbiology

Environmental Microbiology (2015) **17**(11), 4239–4252 doi:10.1111/1462-2920.12735

Genomic and proteomic characterization of two novel siphovirus infecting the sedentary facultative epibiont cyanobacterium *Acaryochloris marina*

Yi-Wah Chan,[1] Andrew D. Millard,[2] Peter J. Wheatley,[3] Antony B. Holmes,[4] Remus Mohr,[5] Anna L. Whitworth,[6] Nicholas H. Mann,[2] Anthony W. D. Larkum,[7] Wolfgang R. Hess,[8] David J. Scanlan[2] and Martha R. J. Clokie[9]*

[1]*Warwick Systems Biology Centre, Coventry, UK.*
[2]*School of Life Sciences,* [3]*Department of Physics,* [6]*Department of Chemistry, University of Warwick, Coventry, UK.*
[4]*Institute for Cancer Genetics, Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA.*
[5]*Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany.*
[7]*School of Biological Sciences, University of Sydney, Sydney, NSW, Australia.*
[8]*Institute of Biology III, University of Freiburg, Freiburg, Germany.*
[9]*Department of Infection, Immunity and Inflammation, University of Leicester, Leicester LE19HN, UK.*

Summary

Acaryochloris marina **is a symbiotic species of cyanobacteria that is capable of utilizing far-red light. We report the characterization of the phages A-HIS1 and A-HIS2, capable of infecting *Acaryochloris*. Morphological characterization of these phages places them in the family *Siphoviridae*. However, molecular characterization reveals that they do not show genetic similarity with any known siphoviruses. While the phages do show synteny between each other, the nucleotide identity between the phages is low at 45–67%, suggesting they diverged from each other some time ago. The greatest number of genes shared with another phage (a myovirus infecting marine *Synechococcus*) was four. Unlike most other cyanophages and in common with the *Siphoviridae* infecting *Synechococcus*, no photosynthesis-related genes were found in the genome. CRISPR (clustered regularly interspaced short palindromic repeats) spacers from the host *Acaryochloris* had partial matches to sequences found within the phages, which is the first time CRISPRs have been reported in a cyanobacterial/cyanophage system. The phages also encode a homologue of the proteobacterial RNase T. The potential function of RNase T in the mark-up or digestion of crRNA hints at a novel mechanism for evading the host CRISPR system.**

Received 16 October, 2014; revised 26 November, 2014; accepted 27 November, 2003. *For correspondence. E-mail mrjc1@le.ac.uk; Tel. 011 6252 2959; Fax 011 6252 5030.

Introduction

The sheer abundance of cyanophages and their inextricable relationship with their bacterial hosts has put them at the forefront of marine microbiology research (Suttle, 2007). Phages play key environmental roles in biogeochemical cycling, genetic diversity and bacterial evolution through lateral gene transfer (Suttle, 2005). It is also becoming increasingly apparent that they can influence host physiology (Lindell *et al.*, 2005; 2007; Clokie and Mann, 2006; Dammeyer *et al.*, 2008; Thompson *et al.*, 2011), most notably through the discovery of photosynthesis-related genes in phages that infect the marine picocyanobacteria *Synechococcus* and *Prochlorococcus* spp. (Mann *et al.*, 2003; Lindell *et al.*, 2004; Millard *et al.*, 2004). *Acaryochloris* is a widely distributed, niche unicellular marine cyanobacteria, known for its main photosynthetic pigment being chlorophyll *d* (Miyashita *et al.*, 1996; Larkum and Kuhl, 2005; Kashiyama *et al.*, 2008), and this is the first report of genomes of *Acaryochloris* phages.

To date, 58 cyanophages have complete genomes deposited in the NCBI database, of which six (S-CBS1, S-CBS2, S-CBS3, S-CBS4, P-SS2 and KBS2A) are published siphoviruses (Wang and Chen, 2008; Sullivan *et al.*, 2009; Huang *et al.*, 2012; Ponsero *et al.*, 2013). It is no coincidence that our ability to understand bacterial-phage systems has progressed concurrently with advancements in sequencing technologies and the availability of genomic resources. Of particular relevance to this study is that the genomes of *Acaryochloris* spp. (Swingley *et al.*, 2008;

Miller *et al*., 2011) and the cyanophage *Synechococcus* myovirus S-TIM5 recently became available (Sabehi *et al*., 2012), facilitating genomic comparisons with the phages isolated in this study.

Here then, we characterize two siphoviruses, A-HIS1 and A-HIS2, which infect *Acaryochloris marina* strain MBIC11017 (Miyashita *et al*., 1996; Larkum and Kuhl, 2005). We include morphological analysis, complete genome sequencing and proteomic analysis of each phage. A-HIS1 and A-HIS2 are the first sequenced phages of *Acaryochloris* spp., or indeed of any cyanobacterium that lives either symbiotically with or as an epibiont in biofilms on metazoans. In particular, we show that S-TIM5 and the *Acaryochloris* phages share certain genes, including the novel mitochondrial-like DNA polymerase, which we have previously described (Chan *et al*., 2011). We describe the presence of host CRISPR spacer-like sequences in these phages (only previously reported in thermophilic *Synechococcus* spp. originating from hot springs in Yellowstone National Park; Heidelberg *et al*., 2009), and show that while these *Acaryochloris* phages share many similarities and a common ancestor, the nucleotide and amino acid identity of their shared genes indicates that these phages have diverged from each other and should therefore be considered distinct from each other. Given these phages also contain many novel genes, they are an important addition to the growing genomic databases of phages not only as a resource for gene identification in other phages (or organisms), but also as a definitive reminder of the sheer genetic diversity that remains undiscovered in the oceans.

Results

Identification of phages A-HIS1 and A-HIS2

The cyanophages A-HIS1 and A-HIS2 isolated from the reef waters off Heron Island, Australia (Chan *et al*., 2011), were grouped with the family *Siphoviridae* of phages based on their morphology observed using transmission electron microscopy (TEM) since they have icosahedral capsid heads and long non-contractile tails (Fig. 1A and B). The capsids of both phages were ~ 60 nm in width, while the tail of A-HIS2 was ~ 179 nm compared with that of A-HIS1, which was ~ 124 nm.

Effect of phages on Acaryochloris *bacterial lawn development*

Phages A-HIS1 and A-HIS2 formed clear circular plaques on agarose-based bacterial lawns of *A. marina* MBIC11017 (Fig. 1C and D). Confluent lysis of *A. marina*

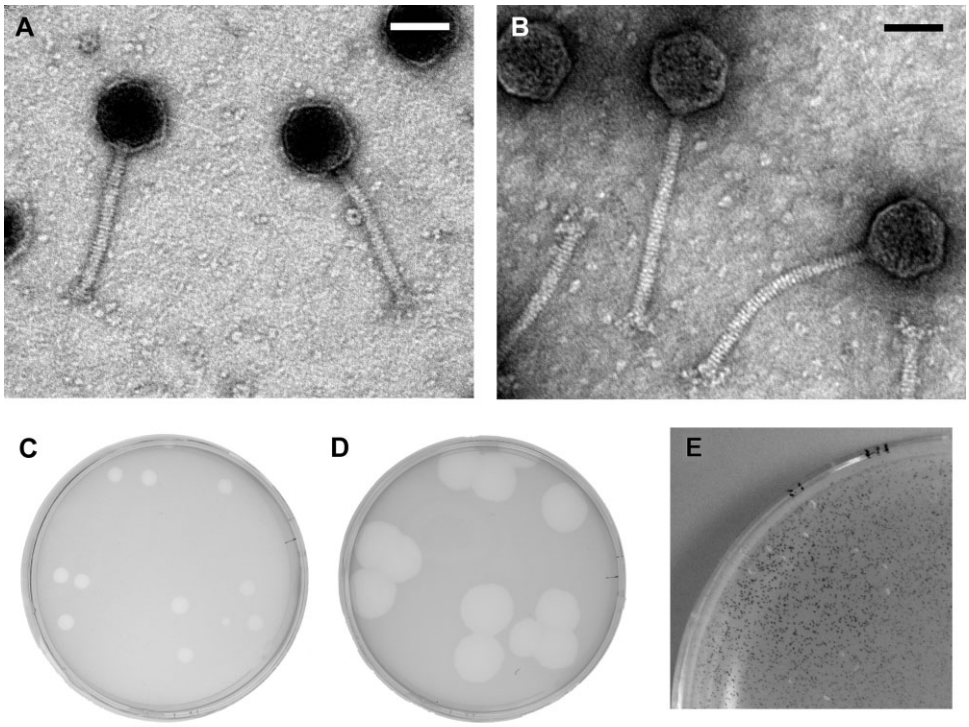

Fig. 1. Phage and plaque assay characteristics. Transmission electron micrographs of (A) A-HIS1 and (B) A-HIS2. Scale bar = 50 nm. The same plaque assay of phage A-HIS1 after (C) 4 and (D) 9 days. (E) Confluent lysis of a *A. marina* MBIC11017 lawn by phage A-HIS1, which yielded spontaneous colony growth 1–2 months after infection.

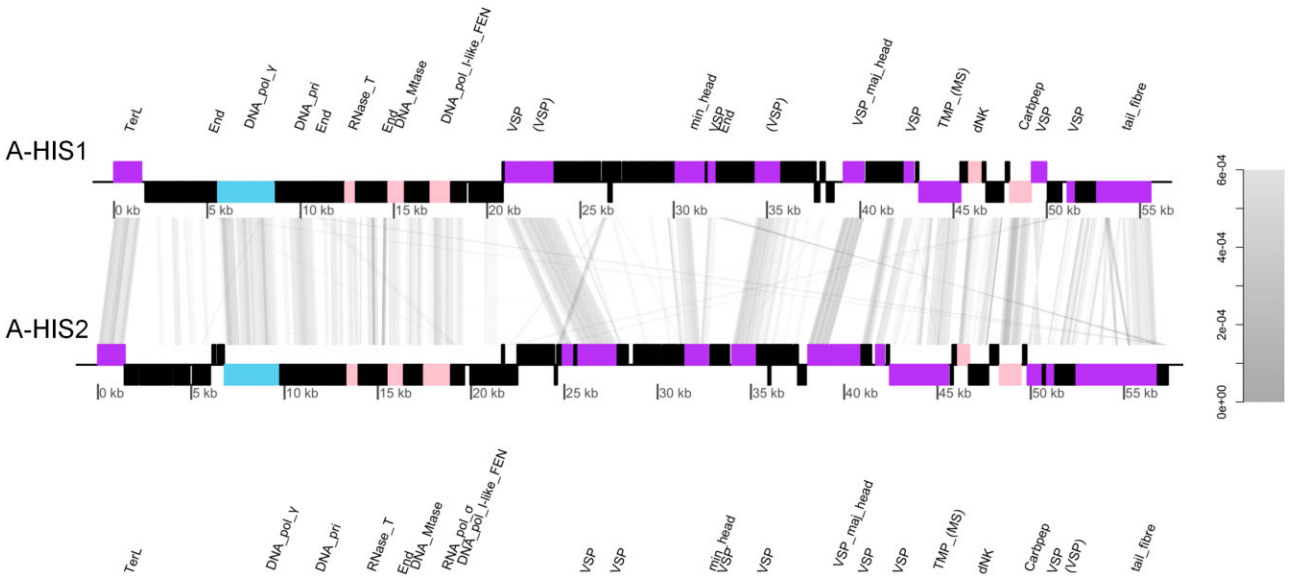

Fig. 2. Comparative genome map of A-HIS1 versus A-HIS2. Predicted ORFs are coloured according to the homology group of the *Acaryochloris* phage genes: purple, phage; pink, bacterial; blue, eukaryotic. Hypothetical ORFs are coloured in black where no annotation has been assigned based on database analyses. Grey lines represent matches using TBLASTX and the corresponding key denotes the associated e-values. A scale is included in kilobases. All ORFs shown encode predicted proteins. VSP indicates a virion structural protein, which was identified by mass spectrometry. (VSP) denotes an ORF categorized as a putative A-HIS1 structural protein based on amino acid sequence similarity to a corresponding ORF in A-HIS2 identified as a putative VSP by mass spectrometry and vice versa. (MS) denotes an ORF, which was identified by both BLASTP and mass spectrometry. See Table 1A and B for ORF details.

MBIC11017 lawns infected with either phage A-HIS1 or A-HIS2 was followed 1–2 months later by the development of spontaneous mutants producing dense colony growth (Fig. 1E). Subsequently, these spontaneous phage-resistant mutants of *A. marina* MBIC11017 were clonally isolated, and well assays showed them to be resistant to both phages A-HIS1 and A-HIS2.

Replication parameters of phages A-HIS1 and A-HIS2

One-step growth experiments were performed on both phages using an MOI (multiplicity of infection or phage : bacteria ratio) of 0.1 ($n = 3$, Fig. S1). The latent period (time from infection to lysis) was 5 h for both phages. The eclipse period (time for the first new viable phage to appear in the host after infection) for both phages was 3.25 h. There were differences in average burst sizes, with A-HIS1 having a burst size of ~ 6 (± 5) compared with ~ 25 (± 22) for A-HIS2, calculated by averaging the number of free phage per infected cell at 8 h and 10 h for both phages. These data showed there was large biological variation among the replicates.

Adsorption of phages to wild-type and phage-resistant Acaryochloris

Both phages A-HIS1 and A-HIS2 adsorbed to the host MBIC11017 after 5 min, with only ~ 10% and ~ 1% of unadsorbed phages remaining in the supernatant respectively (Fig. S2). Between 1 h and 3 h, a ~ 54% increase of unadsorbed A-HIS2 phage was observed compared with a ~ 42% increase of unadsorbed A-HIS1 phage. The same adsorption experiment was carried out with phages A-HIS1 and A-HIS2 against the phage-resistant strains A-HIS1R1 (Fig. S2B) and A-HIS2R1 (Fig. S2C), which resulted in adsorption of these phages to the resistant strains (Fig. S2), thus suggesting that the resistance was a result of a mechanism that was not receptor-based.

Bacteriophage genomes

The genome sizes of A-HIS1 and A-HIS2 are 55 653 bp and 57 391 bp respectively (Fig. 2). The phages have a similar average molecular G+C content of 47.1% and 47.2%, respectively, which is close to the average molecular G+C content of the *A. marina* MBIC11017 host strain at 47.0% (Swingley *et al.*, 2008).

GeneMark predicted 83 ORFs for each phage, and GLIMMER predicted 93 and 104 ORFs for A-HIS1 and A-HIS2 respectively. By analysing these two sets of predicted ORFs, 95 and 104 putative ORFs were assigned to A-HIS1 and A-HIS2 respectively. All ORFs were subjected to BLASTP analysis against the nr database in July 2012 (summarized in Table 1A and 1B). This analysis allowed 16% of A-HIS1 ORFs and 13% of A-HIS2 ORFs to be assigned a putative function. Iterative psi-BLAST analysis

Table 1A. Predicted ORFs from the genome of A-HIS1 with identified putative function (identified by mass spectrometry* (MS) or BLASTP).

ORF	Start	End	Strand	Size (aa)	E-value	Conserved domains	Comment
1	1	1461	+	486	1E-19	Phage terminase large subunit, PBSX family, TIGR01547 (4E-25)	TerL, terminase large subunit Bacterial and phage hits
13	4916	5563	–	215	4E-09	AP2 DNA-binding domain, smart00380 (10E-06)	End, endonuclease, top hit bacterial pathogenesis-related transcriptional factor and ERF protein; also phage and eukaryote hits
14	5573	8554	–	993	6E-119	DNA polymerase family A, cd08641 (3E-62)	DNA pol γ, DNA polymerase γ similar to eukaryotic mitochondrial DNA polymerase γ Top hit fungal *Botryotinia fuckeliana* B05.10
16	8883	10 700	–	605	3E-53	Phage-associated DNA primase, COG3378 (2E-03)	DNA pri, putative DNA primase Top hit cyanophage S-TIM5 ORF159 hypothetical protein (AEZ65736)
17	10 700	11 248	–	182	6E-20	AP2 DNA-binding domain, cd00018 (1E-04)	End, top hit unnamed protein product (Escherichia phage vB_EcoP_G7C)
20	12 390	12 965	–	191	6E-14	DEDDh exonucleases, cd06127 (5E-20)	RNase T, putative bacterial ribonuclease T homologue, top hit gamma proteobacterium NOR51-B, also hits to DNA polymerase III
26	14 372	14 698	–	108	3E-05	AP2 domain, pfam00847 (2E-04)	End, similar to p42.1 Xanthomonas phage Xop411
27	14 703	15 512	–	269	3E-04	Methyltransferase domain, pfam13659 (1E-07)	DNA Mtase, putative DNA modification methylase similar to Pseudomonas sp. ND6 DNA methylase
28	15 590	16 543	–	317	1E-08	–	Top hit cyanophage S-TIM5 ORF139 hypothetical protein (AEZ65717)
30	16 964	18 061	–	365	2E-26	5′-3′ exonuclease, PRK14976 (1E-11)	DNA polI-like flap endonuclease (FEN), top hit cyanophage S-TIM5 ORF157 hypothetical protein (AEZ65734). Bacterial DNA polymerase I hits
33	19 735	20 526	–	263	5E-04	–	Similar to Desulfosporosinus acidophilus SJ4, AP2 domain-containing protein
36*	20 997	21 506	+	169	–	–	VSP, virion structural protein
37	21 542	23 608	+	688	–	–	VSP – based on similarity to AHIS-2 ORF 57 identified by MS; see Table S1
38	23 614	24 228	+	204	5E-64	–	Phage-related protein, top hit Acinetobacter baumannii AYE, similar to Burkholderia phage Bcep1 gp12 and cyanophage P-SSP7 protein PSSP7_001, also hit to Dcm coliphage TLS
39	24 225	24 383	+	52	1E-04	–	Similar to hypothetical protein SYNPCC7002_A1939, Synechococcus sp. PCC7002
55	30 097	31 716	+	539	1E-09	Phage Mu protein F like protein, pfam04233 (1E-11)	Min head, minor head protein, top hit phage phiJL001
57*	31 871	32 284	+	137	–	–	VSP
58	32 302	32 832	+	176	3E-14	AP2 domain, pfam00847 (1E-10)	End, top hit p42.1 Xanthomonas phage Xop411, other hits include HNH endonucleases, bacterial, phage and eukaryota
63	34 385	35 767	+	460	–	Hypothetical protein, PRK06202, S-adenosylmethionine-dependent methyltransferases superfamily (9E-03)	VSP – based on similarity to AHIS-2 ORF 73 identified by MS; see Table S1
73*	39 109	40 188	+	359	–	–	VSP, maj head, putative major capsid protein
77*	42 373	42 843	+	156	–	–	VSP
79*	43 158	45 371	–	737	2E-28	Phage-related minor tail protein, pfam10145 (2E-16)	TMP, phage tail tape measure protein, top hit Halomonas sp. TD01.
82	45 830	46 480	+	216	2E-15	dNK, deoxyribonucleoside kinase, cd01673 (9E-22)	dNK, deoxynucleoside/deoxyguanosine kinase, top hit Caldisericum exile AZM16c01.
88	48 034	49 149	–	371	7E-35	Peptidase C39 like family, pfam13529 (2E-14)	Carbpep, carboxypeptidase top hit Microcystis sp.T1-4. Cyanobacterial and proteobacterial hits
89*	49 193	49 975	+	260	–	Conserved bacterial chromosomal phage-associated hypothetical protein 2217, DUF2460, pfam09343 (3E-03)	VSP
93*	51 109	51 552	–	147	–	–	VSP
95	52 690	55 557	–	955	2E-05	Concanavilin A-like lectin/glucanases superfamily, pfam13385 (7E-09)	Putative tail fibre, top hit Immunoglobulin I-set domain protein, Pedosphaera parvula Ellin514

Table 1B. Predicted ORFs from the genome of A-HIS2 with identified putative function (identified by mass spectrometry* (MS) or BLASTP).

ORF	Start	End	Strand	Size (aa)	E-value	Conserved domains (e-value)	Comment
1	1	1437	+	478	2E-18	Phage terminase, large subunit, PBSX family, TIGR01547 (8E-22)	TerL, top hit Elusimicrobium minutum Pei191 phage terminase large subunit
20	6792	9671	–	959	2E-140	DNA polymerase family A, cd08641 (8E-83)	DNA pol γ, top hit Paracoccidioides brasiliensis Pb18
22	9997	11 829	–	610	–	Phage-associated DNA primase, COG3378 (5E-04)	DNA pri, putative DNA primase Top hit cyanophage S-TIM5 ORF159 hypothetical protein (AEZ65736)
27	13 373	13 972	–	199	9E-12	DEDDh exonucleases, cd06127 (2E-14), DnaQ-like exo superfamily	RNase T, putative bacterial ribonuclease T homologue, top hit Marinobacter algicola DG893, proteobacterial hits
31	15 153	15 578	–	141	3E-05	AP2 domain, pfam00847 (1E-06)	End, similar to p42.1 Xanthomonas phage Xop411
32	15 578	16 333	–	251	3E-04	Methyltransferase domain, pfam13659(9E-06)	DNA Mtase, similar to DNA N6-adenine methyltransferase, coliphage rv5
33	16 398	17 435	–	345	1E-11	–	Top hit cyanophage S-TIM5 ORF139 hypothetical protein (AEZ65717)
34	17 476	17 859	–	127	4E-04	–	RNA pol σ, RNA polymerase σ factor similar to Ruegeria sp. TM1040 ECF subfamily RNA polymerase σ -24 subunit
35	17 843	18 928	–	361	4E-32	53EXOc, 5′-3′ exonuclease, smart00475 (1E-14)	DNA poll-like FEN. top hit cyanophage S-TIM5 ORF157 hypothetical protein (AEZ65734) Bacterial DNA polymerase I hits
55*	24 908	25 414	+	168	–	–	VSP
57*	25 748	27 811	+	687	–	–	VSP
59	27 980	28 135	+	51	4E-08	–	Similar to *Acaryochloris* spp. hypothetical proteins MBIC11017 AM1_2888 and CCMEE5410 ACCM5_19468. Top hit Cyanothece sp. PCC 7425
68	31 483	32 319	+	278	3E-09	Phage Mu protein F like protein, pfam04233 (3E-09)	Min head, similar to SPP1 gp7 (Ruegeria sp. TM1040) and gp62 phage phiJL001
69*	32 321	32 707	+	128	–	–	VSP
73*	33 994	35 325	+	443	–	–	VSP
77	36 200	36 826	+	208	9E-04	–	Similar to VanW family protein, Oscillochloris trichoides DG6
82*	38 069	39 145	+	358	–	–	VSP, maj head, putative major capsid protein
83*	39 253	40 845	+	530	–	–	VSP
85*	41 700	42 140	+	146	–	–	VSP
87*	42 451	45 576	–	1041	1E-25	PhageMin tail, phage-related minor tail protein, pfam10145 (3E-11)	TMP, top hit Marinobacter adhaerens HP15, top phage hit Vibrio phage 1
90	46 099	46 698	+	199	2E-15	dNK, deoxyribonucleoside kinase, COG1428 (9E-17)	dNK, top hits Marinithermus hydrothermalis DSM 14884 deoxy-guanosine/adenosine kinase subunit
97	48 341	49 456	–	371	2E-34	Peptidase C39 like family, pfam13529 (3E-13)	Carbpep, top hits cyanobacterial Microcystis sp. T1-4 and proteobacterial Vibrio sp. EJY3 carboxypeptidase
99*	49 825	50 544	–	239	–	–	VSP
101	50 871	51 299	–	142	–	–	VSP – based on similarity to AHIS-1 ORF 93 identified by MS; see Table S1
103	52 448	56 743	–	1431	3E-05	Concanavalin A-like lectin/glucanases superfamily, pfam13385 (1E-05)	Putative tail fibre protein, based on similarity to A-HIS1 ORF 95; see Table S1
104	56 808	57 350	–	180	7E-18	–	End, top hit numod4 motif family protein, Paenibacillus elgii B69, associated with HNH (AP2) endonucleases

did not yield any further information. In addition, no pseudogenes or tRNAs were identified.

A-HIS1 and A-HIS2 are similar in their relative genomic architecture. This synteny is clearly visualized in a comparative genome plot in terms of the genes, which they both encode and the relative positions of those genes (Fig. 2). However, while there is synteny between the common genes of the *Acaryochloris* phages, the genes are considerably different at the nucleotide level, ranging from 45% to 67% identity (Table 2).

In addition, the genomes can roughly be divided into three regions based on the distribution of the predicted ORFs on the two DNA strands (Fig. 2); namely, we define regions 1, 2 and 3 as ~ 1.5–22 kb, ~ 22–42 kb and ~ 42–1.5 kb (i.e. including the terminase large subunit in region 3). Genes involved in DNA replication, metabolism and modification are encoded mainly in region 1 for both phages, with the exception of ORF (open reading frame) 38 in A-HIS1, which codes for a DNA methylase and is found in region 2. Structural proteins are encoded in regions 2 and 3, including those identified by mass spectrometry, which are referred to here as virion structural proteins (VSPs).

Proteobacterial ribonuclease T

Both phage genomes contain a putative bacterial DEDDh family RNase T, which shares 47% identity at the amino acid level (Table 2). RNase T has never previously been found in a phage genome. In particular, the occurrence of RNase T genes is clearly restricted to the proteobacteria, where it plays a role in the modification of tRNAs (Condon and Putzer, 2002), and therefore may potentially modify RNA species during phage infection. Top BLASTP hits of A-HIS1 and A-HIS2 RNase T were found against γ-Proteobacterial RNase T proteins (Table 1A and B). We performed a phylogenetic analysis to determine how the novel phage RNase T was related to other DEDD family proteins based on an analysis of exoribonuclease superfamilies by Zuo and Deutscher (2001). A phylogenetic network analysis of DEDD family proteins showed that the phage RNase T homologues fell closer to the other DEDD family proteins than to the RNase T clade (Fig. S3). However, as can be clearly seen, there are many possible trees highlighted by the netted region at the centre of the network. Interestingly, almost all the environmental sequences fell within the RNase T clade. Clearly, the phage RNase T and those of the α/γ-Proteobacteria share ancestry; however, the origins of the phage homologues remain unknown.

Phage structural proteins

Only two VSPs were identified by BLASTP analysis in each phage: a minor head protein and a tape measure protein. A further seven and nine ORFs were identified by mass spectrometry as VSPs for A-HIS1 and A-HIS2, respectively, which included each phage tape measure protein (Fig. 3 and Table S1). The observed protein sizes from SDS-PAGE correlated well with the predicted molecular weights. A further three ORFs (A-HIS1 ORFs 37 and 63 and A-HIS2 ORF 101) were then assigned as VSPs by virtue of their percent identity to homologous ORFs in the other genome, previously identified as VSPs by mass spectrometry (Table 2). In total, the combination of mass spectrometry, BLASTP analyses and ORF comparisons allowed 23 ORFs from each phage to be assigned a putative function (Table 1A and B).

SDS-PAGE analysis showed protein lanes to be always dominated by one particular protein band for each phage, around 38 kDa in size, which corresponded to ORFs 73 and 82 of A-HIS1 and A-HIS2, respectively (Fig. 3 and Table S1). ORFs 73 and 82 were also repeatedly identified by mass spectrometry in other protein bands on the gel. It has been well-documented that the major capsid protein in siphoviruses have a much higher copy number than the major tail protein. Specifically, coliphage T5 has 480 head copies to 190 tail copies, with the major capsid making up 57% of the total protein in a phage compared with the tail, which makes up 19% (Buchwald *et al*., 1970). For T5, Zweig and Cummings reported 65% for the head and 17% for the tail as a percentage of total protein in a T5 phage (Zweig and Cummings, 1973). This led us to conclude that these ORFs are the putative major capsid of phages A-HIS1 and A-HIS2 respectively.

A-HIS1 ORF 95 has been annotated as a putative tail fibre protein based on a BLASTP analysis, which revealed similarity to an immunoglobulin I-set domain protein. Recently, Fraser and colleagues (2006) found that Ig-like domains are present in many dsDNA phage structural proteins, and that these protein domains are likely to be on the outside of the phage (Barr *et al*., 2013). A-HIS2 ORF 103 was similarly annotated as a putative tail fibre protein based on its similarity to A-HIS1 ORF 95 (Table 1A and B).

Phage replication

We have previously reported that these phages encode a full-length family A DNA polymerase γ similar to mitochondrial DNA polymerase (Chan *et al*., 2011). Both phages also encode a putative flap endonuclease or FEN (Harrington and Lieber, 1994; Allen *et al*., 2009), which are found in viruses and all domains of life and are essential for DNA replication. The phage FENs are similar to eubacterial FENs characterized by the three aspartate residues required for metal-binding. Other ORFs encoded by both phages involved in nucleic acid modification are a putative DNA primase and DNA methylase.

Table 2. Summary of gene similarities between A-HIS1 and A-HIS2 based on standalone BLASTP (default parameters) analysis of each set of phage genes first as query against the other, then as the subject (i.e. database) for the other.

A-HIS1		A-HIS2			
	Size		Size		
ORF	aa[a]	ORF	aa[a]	% identity (nt/aa[a])	Putative function[b]
Shared gene pairs					
1	486	1	478	62/63	TerL
4	86	9	86	53/37	–
6	236	12	137	52/42	–
12	112	17	97	49/35	–
14	993	20	959	58/57	DNA pol γ
16	605	22	610	54/47	DNA pri
19	283	25	238	50/40	–
20	191	27	199	53/47	RNase T
22	57	28	63	53/37	–
24	209	30	210	59/49	–
25	59	95	58	52/36	–
26	108	31	141	60/50	End
27	269	32	251	51/36	rRNA Mtase
28	259	33	345	55/42	–
30	365	35	361	55/52	DNA pol I-like FEN
31	265	36	235	49/35	–
33	263	38	243	49/32	–
34	73	44	73	49/31	–
36	169	55	168	60/61	VSP
37	688	57	687	60/66	VSP
41	55	58	50	61/50	–
43	83	54	83	65/64	–
46	149	45	133	47/27	–
47	61	63	76	46/33	–
51	267	66	276	45/20	–
53	77	67	95	50/42	–
55	539	68	278	55/49	Min head
57	137	69	128	53/53	VSP
58[c]	176	104	180	59/56	End
63	460	73	443	58/57	VSP
64	190	74	201	49/36	–
65	61	76	59	51/30	–
66	214	77	208	51/43	–
70	78	79	78	52/31	–
71	54	81	51	45/31	–
73	359	82	358	64/60	Maj head
75	173	84	174	59/62	–
77	156	85	146	55/57	VSP
79	737	87	1041	55/52	TMP
82	216	90	199	58/56	dNK
84	89	91	95	50/30	–
86	160	94	158	59/50	–
88	371	97	371	67/70	Carbpep
89[c]	260	99	239	59/59	VSP
91	96	42	86	61/53	–
93	147	101	142	48/28	VSP
94	376	102	378	52/45	–
95	955	103	1431	52/43	Tail fibre
Similar gene pairs					
13	215	36 104	235 180	48, 56/12, 52	End (except for
17	182			46, 56/14, 48	A-HIS2 ORF36)
58	176	36	235	45/15	
41	55	4	60	48/29	–
43	83	42[c]	86	59/46	–
76	127	45	133	50/24	–
91	96	54[c]	83	58/44	–

Matches with e-values > 1e-03 were removed. Shared gene pairs were the top hits in both BLASTP searches which could mutually identify one another. Similar gene pairs listed are those that could also mutually identify one another, but were not top hits in the BLASTP searches. Per cent identity values are taken from per cent identity matrix (PIM) values calculated in CLUSTALX (1.83).

a. nt = nucleotides, aa = amino acids.

b. Abbreviations as in Table 1A and B.

c. ORFs with nucleotide sequences that are inverted with respect to each other.

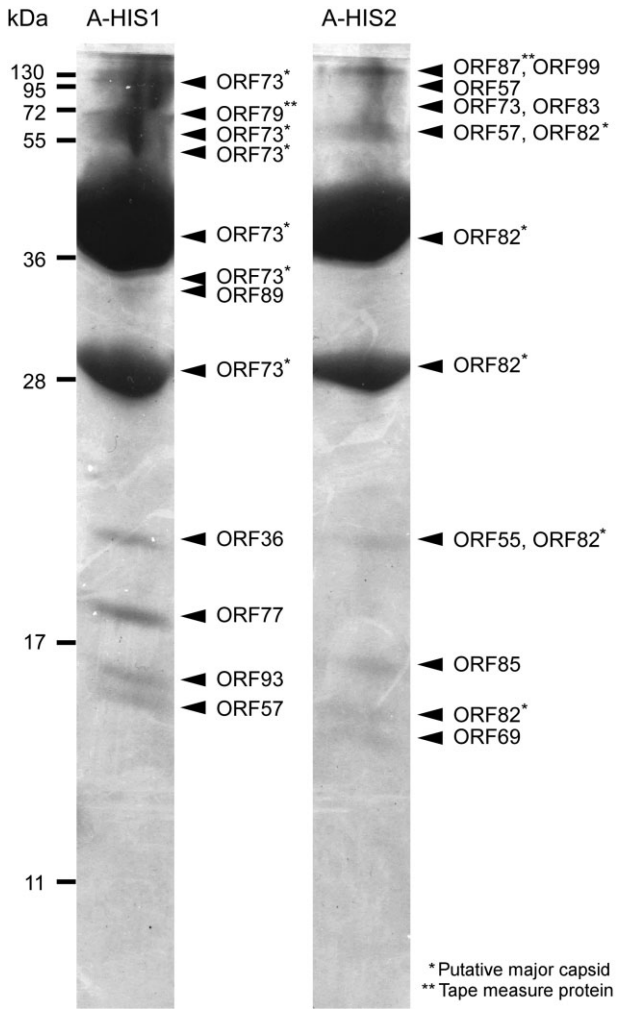

Fig. 3. Identification of phage structural proteins.
Coomassie-stained SDS-PAGE of purified proteins from A-HIS1 and A-HIS2 phage ghosts. Arrows indicate the position of bands removed from the gel and the corresponding ORFs, which were recognized after amino acid sequence identification. The putative major capsid and tape measure protein have been labelled in the figure. Other ORFs are annotated as virion structural proteins.

Genes associated with phage replication occur in regions 2 and 3 (Fig. 2). This includes a carboxypeptidase and a deoxynucleoside kinase (dNK), which are associated with protein maturation and DNA synthesis respectively (Alberts *et al.*, 2002; Eriksson *et al.*, 2002). The protease encodes two conserved domains: a C39 peptidase domain associated with the processing of bacteriocins and a family 19 chitinase domain. Each phage also contains a terminase, which is involved in the packaging of DNA into the virion heads (Catalano, 2000).

Putative regulatory elements

Using PHIRE (Lavigne *et al.*, 2004), a conserved sequence element was detected 10 times across the A-HIS2 genome, each between 19 bp and 27 bp upstream from the start codon of a predicted ORF (Fig. S4). The nucleotide sequence 5′-**TDACC**HN**A**K**GTTACA** NBDY**GAGGTATT**YHY-3′ was designated P^1 (consensus among the 10 putative sequences in bold). Notably, there are two highly conserved regions in this sequence (underlined), which are 4 bp apart in all versions of the sequence. The sequence motif extends a further 6 bp upstream of the P^1 region for ORFs 73, 79 and 83 at the 5′ end of P^1. The conservation among the P^1 elements and their location exclusively in intergenic spacers points to their possible role in transcriptional control, e.g. as a transcription factor binding site. The nucleotide sequence 5′-**GGGGGT**P^1-3′ was designated P^2. Both P^1 and P^2 were not found in any of the upstream regions of the predicted A-HIS1 ORFs. Six out of the 10 P^1 and P^2 elements occurred upstream of putative VSPs identified by mass spectrometry (ORFs 55, 57, 73, 83, 85 and 87, the tape measure protein). Interestingly, these motifs also occurred upstream of ORFs 97 and 103, which code for a carboxypeptidase and putative tail fibre protein respectively. Additionally, using ELPH one instance of a putative −10 and −35 promoter region in A-HIS1 was identified at positions complement(4891..4896) and complement(4914..4919) 18 bp upstream of ORF 11, with the optimal separation of 17 bases (as identified for the major class of *Escherichia coli* promoters; Alberts *et al.*, 2002). The predicted P^1 and P^2 elements are unique to A-HIS2, indicating that the corresponding genes are probably expressed together during the replication cycle.

CRISPRs

Small CRISPR RNAs have been shown to guide antiviral defence in bacteria (Brouns *et al.*, 2008). In the *Acaryochloris* chromosome, one CRISPR with three spacers (referred to here as SP1, SP2 and SP3) was found in the CRISPR database along with two putative CRISPRs, one in each of plasmids pREB7 and pREB8, with just one spacer each – SP4 and SP5 respectively (Grissa *et al.*, 2007a). BLAST analysis using the spacer sequences against the A-HIS1 and A-HIS2 phage genomes gave three and seven partial matches respectively (Table S2).

Genome comparisons between the phage and cyanobacterial genomes

BLASTP analyses indicated the *Acaryochloris* phages have four shared ORFs, which have homologues in the *Synechococcus* myovirus S-TIM5. In particular, A-HIS1/2 ORF 14/20 (mitochondrial DNA polymerase), ORF 16/22 (putative DNA primase), ORF 28/33 (hypothetical protein) and ORF 30/35 (putative DNA polI-like FEN) had 31/33%,

27/25%, 23/26% and 29/32% identity to S-TIM5 hypothetical proteins, ORFs 76, 159, 139 and 157, respectively.

A reciprocal BLASTP approach was used to compare the predicted ORFs from each *A. marina* phage to find the number of shared genes between the two phages. The phages A-HIS1 and A-HIS2 share 48 such genes (Table 2).

BLASTP against the *Acaryochloris* spp. genomes showed that only ORF 59 from A-HIS2 had a significant hit (e-value 4e-08) to *Acaryochloris* spp. hypothetical proteins MBIC11017 AM1_2888 and CCMEE5410 ACCM5_19468. This lone result suggests that lateral gene transfer between phages A-HIS1 and A-HIS2 and the host strain MBIC11017 has occurred infrequently. It is also worth noting that we were only able to classify five genes per phage as bacterial homologues (bacteria refers to bacteria in general) (Fig. 2, pink).

Discussion

A-HIS1 and A-HIS2 are the first *Acaryochloris* phages to be isolated and characterized. Transmission electron microscopy allowed us to classify these phages morphologically as siphoviruses. Notably, marine cyanobacterial siphoviral isolates are rare in the literature; siphoviruses infecting *Synechococcus* spp. include P1 (Lu *et al*., 2001), S-BBS1 (Suttle and Chan, 1993), S-CBS1, S-CBS2, S-CBS3, S-CBS4 (Wang and Chen, 2008) and KBS2A, and P-SS1 and P-SS2 infect *Prochlorococcus* sp. (Sullivan *et al*., 2003), of which S-CBS1, S-CBS2, S-CBS3, S-CBS4, P-SS2 and KBS2A have been sequenced establishing a cyanobacterial siphovirus genome database (Sullivan *et al*., 2009; Huang *et al*., 2012; Ponsero *et al*., 2013).

When the genomes of phages A-HIS1 and A-HIS2 were compared with each other, a high degree of synteny was observed. Fortuitously, the genome of the host strain MBIC11017 was also sequenced, and on examination it does encode phage-related genes (Swingley *et al*., 2008), notably a tail-related protein – a phage lysozyme and integrases which are commonly associated with temperate phages. However, none of these genes are related to genes in our lytic phages. Moreover, unlike the situation with some *Synechococcus* and *Prochlorococcus* phages, no photosynthesis-related genes were found in phages A-HIS1 and A-HIS2.

The area of phage taxonomy has aroused much discussion in recent years, and there is growing support for a comparative genomics led-approach to understand *Caudovirales* diversity (Nelson, 2004; Casjens, 2005). However, the *Acaryochloris* phage genomes have a unique composition and organization compared with other known siphoviruses. Recent work has shown that siphoviruses do have a conserved genome architecture (Brüssow and Desiere, 2001; Seguritan *et al*., 2003). However, these *Acaryochloris* phages do not conform to this idea. For example, a feature of the comparative genomics results presented by Brüssow and Desiere (2001) on the siphoviruses ψM2, HK97, Sfi21, λ, Sfi11, φC31, sk1, L5 and TM4 is that each siphovirus has a cluster of genes comprising most or all of the following genes: a large terminase, a small terminase, a portal protein, a protease and a major head protein. In particular, these genes occur very close or next to each other in more or less the same order. The equivalent genes in the *Acaryochloris* phages are spread out and interspersed by unknown genes. Moreover, in both *Acaryochloris* phages, the protease identified is a carboxypeptidase, and no portal proteins have been identified. A-HIS1 and A-HIS2 also contain the novel gene RNase T, which has not been found in other phages. In light of these observations A-HIS1 and A-HIS2 can be classically grouped with the *Siphoviridae* by morphology but form a unique subtype based on genomic data. Moreover, turning to the cyanobacterial siphovirus genomes specifically, genome comparison by TBLASTX analysis of the A-HIS1 and A-HIS2 genomes to the cyanobacterial siphoviruses S-CBS1, S-CBS2, S-CBS3, S-CBS4 and PSS2 indicated that the *Acaryochloris* phages are distinct from the other cyanobacterial siphoviruses as they have no genes in common. Compared with KBS2A, only A-HIS1 had two similar genes. These were KBS2A CPKG_00039 (minor head protein) and CPKG_00059 (hypothetical protein), which matched (e-value < 1e-03) A-HIS1 ORFs 55 (minor head protein) and 79 (tape measure protein).

In contrast, A-HIS1 and A-HIS2 have more in common with the myovirus S-TIM5 than they do with cyanobacterial siphoviruses as observed by the four shared ORFs, including the mitochondrial-like DNA polymerase they encode. This observation suggests that while more cyanobacterial siphovirus genomes may aid siphovirus classification, at the genomic level, the classification is not straightforward.

The *Acaryochloris* phages also uniquely encode a phage RNase T, which has been reported to have many functions, including playing a role in ribosomal RNA maturation (Li and Deutscher, 1995; Li *et al*., 1999). It was first discovered in 1984 by Deutscher and colleagues, who showed that RNase T removes adenosine monophosphate (AMP) from the 3' CCA terminus of specific tRNAs in the tRNA end-turnover process (Deutscher *et al*., 1984; Deutscher and Marlor, 1985). However, the physiological function of this process remains unknown. As observed by Deutscher (1973), RNase T acts as a regulator of protein synthesis by controlling which tRNAs have their AMP cleaved, as suggested by a model proposed by Stent (1964). Therefore, RNase T may specifically direct the production of proteins necessary for phage replication. However, RNase T may serve a more obvious purpose in that a pool of AMP would be released from idle

tRNAs, hence triggering tRNA degradation essentially recycling the raw material of the host as needed for the phages. Alternatively, RNase T activity may be part of an anti-CRISPR activity. It is well known that (eukaryotic) viruses have evolved strategies to avoid host defence mechanisms that are based on RNA interference. The prokaryotic CRISPR-dependent antiviral activity depends on short RNA molecules (crRNAs) annealing to phage DNA or RNA fragments. During the maturation process of the crRNAs, various single-stranded RNA species are present, which may constitute possible substrates for the phage RNase T (Brouns *et al*., 2008). However, this hypothesis lacks experimental evidence and hence requires further work. Noteworthy though is that after a number of weeks, phage-resistant *A. marina* MBIC11017 colonies readily appeared on confluently lysed plaque assay plates (Fig. 1E). The spontaneous occurrence of resistant colonies has been reported for other marine cyanophage-host systems (Avrani *et al*., 2011). Adsorption experiments indicated that the resistance is unlikely to be receptor-based, given that phages A-HIS1 and A-HIS2 were able to attach to clonal cultured resistant strains derived from infection with either phage (Fig. S2). This suggests that *A. marina* MBIC11017 may have evolved mechanisms to overcome the virulence of these lytic phages, perhaps by integrating further phage DNA fragments into their CRISPRs or using unknown mechanisms.

The *A. marina* MBIC11017 host genome lacks the RNase T gene, although this is not unsurprising given its absence in other cyanobacteria. On examination of the top BLASTP hits, the phage RNase T was found to be most similar to that of γ-Proteobacteria and with only a single α-Proteobacterial RNase T hit. The type of proteobacteria the *Acaryochloris* phage RNase T may have originated from remains inconclusive. While the current phylogenetic analysis may provide some insight, a more conclusive phylogeny could be obtained with more representatives of *Acaryochloris* phage RNase T-like sequences, preferably with a known origin. Other types of ribonucleases have also been unearthed in phages, including RNase H, found in a number of phages, including T4 and S-PM2, which remove the RNA primers within Okazaki fragments (Hollingsworth and Nossal, 1991) and RNase II found in lactococcal phage Q54 (Fortier *et al*., 2006).

Considering the genomes of A-HIS1 and A-HIS2 are the first phages that infect *Acaryochloris* to be sequenced, it is not surprising that only 24% (A-HIS1) and 22% (A-HIS2) of the predicted ORFs could be assigned putative function. Indeed, the ability to assign function to predicted ORFs in A-HIS1 and A-HIS2 solely by database searches is limited by the genomic data currently available, which is commonly encountered in other phage genome studies.

Our study further highlights the need for the isolation and characterization of phage infecting novel host organisms inhabiting unusual habitats. In particular, genomic analysis of our *A. marina* MBIC11017 phage isolates revealed many interesting features, including an intriguing evolutionary history. Specifically, these phages appear to provide evidence of a direct link among bacteriophages, mitochondria, proteobacteria and cyanobacteria.

Experimental procedures

Growth of bacterial host

Acaryochloris marina strain MBIC11017 was maintained in artificial sea water (ASW) for all experiments (Wyman *et al*., 1985; Clokie and Kropinski, 2009). Cells were grown in 1 l polycarbonate NALGENE® 4105 Fernbach culture flasks with aeration, stirring and under 30–50 μmol photons m^{-2} s^{-1} of continuous white fluorescent light (Osram, L65/80W/23) at 28°C. *Acaryochloris marina* MBIC11017 biofilms were prepared by pouring 30 ml of exponential phase culture into Petri dishes containing autoclaved circular glass coverslips and grown at room temperature under ambient light (10 μmol photons m^{-2} s^{-1}). After 3 days, the biofilms were transferred to Petri dishes containing freshly autoclaved ASW using sterile tweezers and left to grow.

Phage isolation

Seawater samples were collected from Heron Island, Great Barrier Reef, Australia (23° 25.800′ S and 51° 55.605′ E) as described previously (Chan *et al*., 2011). Both bacteriophages were found in one unfiltered sample that was procured by incubating an *Acaryochloris*-associated ascidian (*Lissoclinum patella*) in SM (100 mM NaCl, 8 mM MgSO4• 7H2O, 50 mM Tris-Cl (1 M, pH 7.5)) buffer. All seawater samples were stored in 50 ml falcon tubes, wrapped in aluminium foil and stored at 4°C before transportation.

Plaque assays were performed by adapting the method of isolating phages described for *Synechococcus* sp. WH7803, and all DIFCO Bacto Agar was cleaned as detailed in (Clokie and Kropinski, 2009; Millard, 2009). In short, *A. marina* MBIC11017 cells were grown to exponential phase from an initial optical density (OD) of 0.01–0.05 and harvested for plaque assays at an OD between 0.4 and 0.5. Typically, 1 l of cells was centrifuged, and the pellet was re-suspended in 20 ml ASW to allow for 0.5 ml of concentrated cells per plaque assay. Each 0.5 ml of cells was incubated for 1 h with 50 μl of seawater sample, mixed with 2.5 ml of 0.4% (w/v) agar and then plated onto 1% (w/v) agar plates. A negative control was performed by adding 50 μl of ASW instead of a seawater sample. Plaque assays were incubated at 23°C under 10–15 μmol photons m^{-2} s^{-1} continuous white light illumination.

Once clear plaques had appeared, plugs were removed from the Petri dish using a sterile Pasteur pipette. Each plaque plug was transferred to 1 ml ASW. Phage isolates were then made clonal by three rounds of purification using plaque assays from single plaques. ASW lysates of clonal

phages were stored at 4°C in the dark after bacterial debris was removed by centrifugation.

Isolation of phage resistant Acaryochloris

Spontaneous colonies that appeared on confluent plaque assays 1–2 months after infection were isolated and made clonal by repeated plating and culturing in ASW. All incubations were performed at 23°C under continuous white light illumination (10–20 μmol photons m^{-2} s^{-1}).

One-step growth curves

Before beginning the one-step growth experiment, phage lysates were titred using plaque assays. Of *A. marina* MBIC11017 cells, 1.25 l was grown to an OD (800 nm) between 0.4 and 0.5, and 150 ml of cells were transferred to a 500 ml conical flask per replicate and inoculated with phage to an MOI of 0.1.

Immediately after inoculation, two 0.5 ml samples were removed from each replicate. One sample was centrifuged at 13 000 *g* for 5 min at 4°C in the dark, after which the supernatant was transferred into another tube to be later used to assess free phage. The other 0.5 ml sample was incubated with four drops of chloroform to be later used to assess total phage. The cultures were then incubated at 28°C and shaken at 70 r.p.m. under continuous white light illumination (30 μmol photons m^{-2} s^{-1}) for 5 min to allow for phage adsorption.

In order to synchronize phage infection, infected cultures were subjected to centrifugation at 6693 *g* for 15 min at 28°C. Of each supernatant, 0.5 ml was kept to assess the number of unadsorbed free phage. Following this, the pellets of infected cells were re-suspended in 150 ml fresh ASW warmed to 28°C. Two 0.5 ml samples were then removed and processed as above every 2 h from inoculation (time zero). All phage samples were stored at 4°C in the dark.

Phages were counted by preparing dilution series for each sample collected during the time-course, and two appropriate dilutions were assessed per time point by plaque assay. To visualize the data, plaque forming units per infected cell (log scale) was plotted against time. The data presented are the average of three experimental replicates.

Adsorption assays

Exponentially growing axenic *Acaryochloris* MBIC11017 and phage-resistant strains A-HIS1R1 and A-HIS2R1 (OD800 between 0.4 and 0.5) were infected with phage A-HIS1 or A-HIS2 at an MOI = 1 in sterile conical flasks and shaken at 70 r.p.m. Samples were collected once an hour for 4 h, with the first sample immediately after phage inoculation. At each time point, 0.5 ml of cells were centrifuged, and the supernatant was transferred to a new tube and stored at 4 °C. The number of free phages in each sample was then counted by the plaque assay method as described in the phage isolation section using appropriate serial dilutions of the samples.

Transmission electron microscopy

Phages A-HIS1 and A-HIS2 were concentrated from lysates by precipitation with 2% NaCl (Fisher Scientific) and 10% polyethyleneglycol (PEG)-6000 (BDH) overnight at 4°C in the dark followed by centrifugation at 11 000 *g* for 10 min. Phage pellets were re-suspended in 2–3 ml ASW, and an equal volume of chloroform was added to remove the PEG. The sample was then shaken and centrifuged in a Hettich Rotina 46R centrifuge at 4754 *g* for 15 min at 4°C. The aqueous layer was removed from the chloroform layer and added to aqueous caesium chloride (CsCl, Fisher Scientific) to a final concentration of 0.75 g ml^{-1}. The solutions were then transferred to Beckman Ultra-Clear™ centrifuge tubes (14 × 95 mm) and subjected to ultracentrifugation at ~ 155 000 *g* in a SW40Ti rotor for 18 h at 4°C in an Optima L-80 XP centrifuge. Bands were removed using a syringe, and the concentrated phages were dialysed in dialysis tubing (size 3/MWCO 12-14 000 Da) for 1 h against 1 l ASW twice to remove the CsCl. Phage samples were negatively stained with a 1% uranyl acetate solution on glow-discharged (Emitech K100X Glow Discharger, EM Technologies) carbon film copper mesh grids (Agar Scientific). Phages were imaged using a JEOL 1200EX TEM. Images were processed using DigitalMicrograph™ (Gatan) and ImageJ.

Genome sequencing, annotation and characterization

DNA was extracted from CsCl-purified phages and stored in 70% (v/v) ethanol (Wilson *et al.*, 1993). The genomes of A-HIS1 and A-HIS2 were sequenced as described in (Chan *et al.*, 2011). A-HIS1 and A-HIS2 sequence data have been submitted to the EMBL database under accession numbers FN436268 and FN436269 respectively. The genomes were annotated using Artemis (Rutherford *et al.*, 2000). ORFs were predicted using both GeneMark.hmm 2.0 (Besemer and Borodovsky, 1999) and GLIMMER 3.02 (NCBI) (Salzberg *et al.*, 1998). To create the final set of predicted ORFs for each genome, the two sets of predicted ORFs from GeneMark and GLIMMER were first combined, then for ORFs predicted by both programmes where there was almost complete overlap the longest ORF prediction was kept. NCBI BLASTP and psi-BLAST were used to assign putative function to the predicted ORFs. A comparative genome plot was generated using GenoPlotR and a customized script. Comparison files were generated using standalone BLAST analysis, specifically TBLASTX, and compared using ACT (Artemis Comparison Tool) (Carver *et al.*, 2005). Gene similarity was assessed between the phage genomes by standalone BLASTP analysis using default parameters of each set of phage genes (amino acid sequences) first as query against the other, then as the subject (i.e. database) for the other. Matches with e-values > 1e-03 were removed. Shared gene pairs were the top hits in both BLASTP searches that could mutually identify one another. Similar gene pairs were those that could also mutually identify one another, but were not top hits in the BLASTP searches. Gene comparison was performed by comparing the percent identity matrix (PIM) values calculated from alignments in ClustalX 1.83. All PIM percentages were calculated with default parameters [gap opening : extension, 10:0.1 (pairwise) and 10:0.2 (multiple)]. Motifs and genetic regulatory elements were detected using tRNAscan-SE 1.21

(tRNAs) (Lowe and Eddy, 1997), PHIRE (Lavigne *et al*., 2004), ELPH (Gibbs sampler) (Pertea *et al*., 2007) and a custom Perl script for handling ELPH output files. The *Acaryochloris* genome chromosome and the two plasmids were submitted to CRISPRFinder to identify CRISPRs and accompanying spacers (Grissa *et al*., 2007b). The spacers were subjected to BLAST analysis against the phage genomes using standalone BLASTN with gap penalty settings: existence = 1 and extension = 2.

Proteomics

Phages were purified and concentrated as for TEM. After dialysis to remove the CsCl, the phage were re-suspended in 200 µl 10 M LiCl and heated at 46°C for 20 min to make phage ghosts. Samples were then diluted 10-fold in 50 mM Tris/HCl, 100 mM NaCl and 5 mM $MgCl_2$ (pH 8), and treated with 40 units DNase (Ambion) for 2 h at 37°C. Protein concentration was assessed with the BCA Protein Assay Kit (Sigma). Samples were subsequently concentrated further at 100 000 g for 30 min at 4°C and re-suspended in 10 mM HEPES pH 7.4 (Konopa and Taylor, 1979; Clokie *et al*., 2008). SDS-PAGE protein samples were prepared as described previously (Chan *et al*., 2007). Protein samples were resolved on a 15% SDS-PAGE gel. Molecular weight was assessed using Fermentas PageRuler™ Prestained Protein Ladder Plus. Proteins were sequenced by MALDI-ToF and LC-MS//MS, and protein bands were identified by comparing the sequence data to predicted protein sequences from the genome sequences.

Acknowledgements

Cyanophage genome sequencing was funded by NERC Grant NE/E01089X/1. Y-WC would like to acknowledge Svetla McPhie for electron microscopy training and Sam Mason for assistance with image formatting. Y-WC was supported by the MOAC Doctoral Training Program, via the EPSRC Life Science Initiative. WRH, RM and AWL thank the German Academic Exchange Service (DAAD) for supporting international cooperation, and WRH thanks the German Research Foundation (DFG) for support by Grant HE 2544/8-1.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) *Molecular Biology of the Cell*, 4th edn. New York, USA: Garland Science.

Allen, L.M., Hodskinson, M.R.G., and Sayers, J.R. (2009) Active site substitutions delineate distinct classes of eubacterial flap endonuclease. *Biochem J* **418:** 285–292.

Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. (2011) Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474:** 604–608.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., *et al*. (2013) Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci USA* **110:** 10771–10776.

Besemer, J., and Borodovsky, M. (1999) Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* **27:** 3911–3920.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., *et al*. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321:** 960–964.

Brüssow, H., and Desiere, F. (2001) Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. *Mol Microbiol* **39:** 213–222.

Buchwald, M., Murialdo, H., and Siminovitch, L. (1970) The morphogenesis of bacteriophage lambda. II. Identification of the principal structural proteins. *Virology* **42:** 390–400.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* **21:** 3422–3423.

Casjens, S.R. (2005) Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8:** 451–458.

Catalano, C.E. (2000) The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. *Cell Mol Life Sci* **57:** 128–148.

Chan, Y.W., Nenninger, A., Clokie, S.J., Mann, N.H., Scanlan, D.J., Whitworth, A.L., and Clokie, M.R.J. (2007) Pigment composition and adaptation in free-living and symbiotic strains of Acaryochloris marina. *FEMS Microbiol Ecol* **61** (1)**:** 65–73.

Chan, Y.W., Mohr, R., Millard, A.D., Holmes, A.B., Larkum, A.W., Whitworth, A.L., *et al*. (2011) Discovery of cyanophage genomes which contain mitochondrial DNA polymerase. *Mol Biol Evol* **28:** 2269–2274.

Clokie, M.R., and Kropinski, A.M. (2009) *Bacteriophages: Methods and Protocols*, *Methods in Molecular Biology*. New York, USA: Humana Press.

Clokie, M.R., and Mann, N.H. (2006) Marine cyanophages and light. *Environ Microbiol* **8:** 2074–2082.

Clokie, M.R., Thalassinos, K., Boulanger, P., Slade, S.E., Stoilova-McPhie, S., Cane, M., *et al*. (2008) A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. *Microbiology* **154:** 1775–1782.

Condon, C., and Putzer, H. (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res* **30:** 5339–5346.

Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W., and Frankenberg-Dinkel, N. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol* **18:** 442–448.

Deutscher, M.P. (1973) Synthesis and functions of the -C-C-A terminus of transfer RNA. *Prog Nucleic Acid Res Mol Biol* **13:** 51–92.

Deutscher, M.P., and Marlor, C.W. (1985) Purification and characterization of Escherichia coli RNase T. *J Biol Chem* **260:** 7067–7071.

Deutscher, M.P., Marlor, C.W., and Zaniewski, R. (1984) Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA. *Proc Natl Acad Sci USA* **81:** 4290–4293.

Eriksson, S., Munch-Petersen, B., Johansson, K., and Eklund, H. (2002) Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol Life Sci* **59:** 1327–1346.

Fortier, L.C., Bransi, A., and Moineau, S. (2006) Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of Lactococcus lactis. *J Bacteriol* **188:** 6101–6114.

Fraser, J.S., Yu, Z., Maxwell, K.L., and Davidson, A.R. (2006) Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol* **359:** 496–507.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007a) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8:** 172.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007b) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35:** W52–W57.

Harrington, J.J., and Lieber, M.R. (1994) The characterisation of a mammalian DNA structure-specific endonuclease. *EMBO J* **13:** 1235–1246.

Heidelberg, J.F., Nelson, W.C., Schoenfeld, T., and Bhaya, D. (2009) Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* **4:** e4169.

Hollingsworth, H.C., and Nossal, N.G. (1991) Bacteriophage T4 encodes an RNase H which removes RNA primers made by the T4 DNA replication system in vitro. *J Biol Chem* **266:** 1888–1897.

Huang, S., Wang, K., Jiao, N., and Chen, F. (2012) Genome sequences of siphoviruses infecting marine Synechococcus unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol* **14:** 540–558.

Kashiyama, Y., Miyashita, H., Ohkubo, S., Ogawa, N.O., Chikaraishi, Y., Takano, Y., *et al.* (2008) Evidence of global chlorophyll d. *Science* **321:** 658.

Konopa, G., and Taylor, K. (1979) Coliphage lambda ghosts obtained by osmotic shock or LiCl treatment are devoid of J- and H-gene products. *J Gen Virol* **43:** 729–733.

Larkum, A.W., and Kuhl, M. (2005) Chlorophyll d: the puzzle resolved. *Trends Plant Sci* **10:** 355–357.

Lavigne, R., Sun, W.D., and Volckaert, G. (2004) PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics* **20:** 629–635.

Li, Z., and Deutscher, M.P. (1995) The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. *Proc Natl Acad Sci USA* **92:** 6883–6886.

Li, Z., Pandit, S., and Deutscher, M.P. (1999) Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. *RNA* **5:** 139–146.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci USA* **101:** 11013–11018.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86–89.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., *et al.* (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449:** 83–86.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

Lu, J., Chen, F., and Hodson, R.E. (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine Synechococcus spp. in river estuaries. *Appl Environ Microbiol* **67:** 3285–3290.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

Millard, A., Clokie, M.R., Shub, D.A., and Mann, N.H. (2004) Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. *Proc Natl Acad Sci USA* **101:** 11007–11012.

Millard, A.D. (2009) Isolation of cyanophages from aquatic environments. *Methods Mol Biol* **501:** 33–42.

Miller, S.R., Wood, A.M., Blankenship, R.E., Kim, M., and Ferriera, S. (2011) Dynamics of gene duplication in the genomes of chlorophyll d-producing cyanobacteria: implications for the ecological niche. *Genome Biol Evol* **3:** 601–613.

Miyashita, H., Ikemoto, H., Kurano, N., Adachi, K., Chihara, M., and Miyachi, S. (1996) Chlorophyll d as a major pigment. *Nature* **383:** 402.

Nelson, D. (2004) Phage taxonomy: we agree to disagree. *J Bacteriol* **186:** 7029–7031.

Pertea, M., Mount, S.M., and Salzberg, S.L. (2007) A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. *BMC Bioinformatics* **8:** 159.

Ponsero, A.J., Chen, F., Lennon, J.T., and Wilhelm, S.W. (2013) Complete genome sequence of Cyanobacterial Siphovirus KBS2A. *Genome Announc* **1:** e00472-13.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

Sabehi, G., Shaulov, L., Silver, D.H., Yanai, I., Harel, A., and Lindell, D. (2012) A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proc Natl Acad Sci USA* **109:** 2037–2042.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26:** 544–548.

Seguritan, V., Feng, I.W., Rohwer, F., Swift, M., and Segall, A.M. (2003) Genome sequences of two closely related Vibrio parahaemolyticus phages, VP16T and VP16C. *J Bacteriol* **185:** 6434–6447.

Stent, G.S. (1964) The operon: on its third anniversary. Modulation of transfer RNA species can provide a workable model of an operator-less operon. *Science* **144:** 816–820.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature* **424:** 1047–1051.

Sullivan, M.B., Krastins, B., Hughes, J.L., Kelly, L., Chase, M., Sarracino, D., and Chisholm, S.W. (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ Microbiol* **11:** 2935–2951.

Suttle, C.A. (2005) Viruses in the sea. *Nature* **437:** 356–361.

Suttle, C.A. (2007) Marine viruses–major players in the global ecosystem. *Nat Rev Microbiol* **5:** 801–812.

Suttle, C.A., and Chan, A.M. (1993) Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser* **92:** 99–109.

Swingley, W.D., Chen, M., Cheung, P.C., Conrad, A.L., Dejesa, L.C., Hao, J., *et al*. (2008) Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium Acaryochloris marina. *Proc Natl Acad Sci USA* **105** (6)**:** 2005–2010.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., and Chisholm, S.W. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA* **108:** E757–E764.

Wang, K., and Chen, F. (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* **10:** 300–312.

Wilson, W.H., Joint, I.R., Carr, N.G., and Mann, N.H. (1993) Isolation and molecular characterization of five marine cyanophages propagated on Synechococcus sp. Strain WH7803. *Appl Environ Microbiol* **59:** 3736–3743.

Wyman, M., Gregory, R.P.F., and Carr, N.G. (1985) Novel role for phycoerythrin in a marine cyanobacterium, Synechococcus strain DC2. *Science* **230:** 818–820.

Zuo, Y., and Deutscher, M.P. (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29:** 1017–1026.

Zweig, M., and Cummings, D.J. (1973) Structural proteins of bacteriophage T5. *Virology* **51:** 443–453.

Supporting information

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

Fig. S1. One-step growth analysis. One-step growth experiments (MOI = 0.1) of phages (A) A-HIS1 and (B) A-HIS2 on *A. marina* MBIC11017. From time zero, the eclipse/latent period ends at i/ii where the total/free phage reaches one phage per infected cell respectively. Y-axis is on a log scale and $n = 3$.

Fig. S2. Phage adsorption to *Acaryochloris* strains. Percentage of free unadsorbed phages as determined by plaque assay. *Acaryochloris* strains (A) MBIC11017 (wild-type parent strain), (B) A-HIS1R1 (spontaneous strain from A-HIS1 infection) and (C) A-HIS2R1 (spontaneous strain from A-HIS2 infection) were inoculated at MOI = 1 ($n = 3$). Dotted lines join zero time-point (known titre value before experiment) to the first time-point at 5 min. Subsequent experimental time points are joined with solid lines.

Fig. S3. Phylogenetic network of DEDD family proteins. E, environmental (CAMERA), P, proteobacteria, A, *Alphaproteobacteria*, and Z indicates they are based on the sequences that were used in the alignment by Zuo and Deutscher (2001), which include sequences representing DNA exonucleases, DNA polymerase III, RNase T, oligoribonucleases, Pan2 proteins, DAN nucleases, RNase D and others (see Appendix S1 for sequence details and methods). In particular, Z4, Z16-Z20 and P00-P15 (except for P04) are RNase T sequences. Scale bar: substitutions per site.

Fig. S4. Conserved sequence elements possibly associated with A-HIS2 transcription and virion replication. Alignment of consensus sequences detected in the A-HIS2 genome by PHIRE, which occurred upstream of putative predicted ORFs (ORF numbers are shown on the left). The alignment was performed in CLUSTALX using default settings. The first adenosine (A) at the right end of each sequence is the start codon of the respective downstream ORF.

Table S1. ORFs identified by mass spectrometry.

Table S2. Alignment of CRISPR-spacer-associated sequences identified by standalone BLASTN. The same results were obtained from BLAST analyses using spacers 4 and 5 and so are also included in the alignment.

Appendix S1. Methods, sequence details and alignment used in the RNase T phylogenetic analysis.