Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere

Jonathan Filée*, Françoise Tétart*, Curtis A. Suttle†, and H. M. Krisch*‡

*Laboratoire de Microbiologie et Génétique Moléculaire, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5100, 118 Route de Narbonne, 31062, Toulouse, Cedex 9, France; and †Departments of Earth & Ocean Sciences, Microbiology & Immunology, and Botany, University of British Columbia, 1461-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4

Tailed bacteriophages are the most abundant biological entities in marine environments. However, most of these marine phages are uncharacterized because few of their hosts have been cultivated. To learn more about such phages, we designed a set of degenerate PCR primers for phage T4 g23, which encodes the major capsid protein in all of the T4-type phages, an important family of the tailed phage. These primers were used to amplify g23-related sequences from diverse marine environments (fjords and bays of British Columbia, the eastern Gulf of Mexico, and the western Arctic Ocean) revealing a remarkable level of molecular diversity, which in some cases was correlated with morphological variation of the virions. Phylogenetic analysis showed that although some of these sequences were closely related to well studied subgroups of the T4-type phage, such as the T-evens, the majority of them belong to five previously uncharacterized subgroups. These data indicate that the host range of T4-type phages is much broader than previously imagined and that the laboratory isolate T4 belongs to a phage family that is extraordinarily widespread and diverse in the biosphere.

diversity | T4-type phage | gene 23 | genomics | ecology

Bacteriophages are the dark matter of the biosphere. Numerous, ubiquitous, and inert, they become biologically active only when they infect an appropriate host among the innumerable bacteria in the environment. Phages are the most abundant entities in the biosphere: it has been estimated that the number of phage exceeds the number of bacteria by a factor of 5:25 (1, 2). Many lines of evidence indicate that their diversity is immense (3) and that only an infinitesimal fraction of the biosphere's phages have been inventoried.

Exerting enormous influence over the microbial world, phage play a critical role in the nutrient and energy cycle (2, 4, 5). Phages are also believed to be a major driving force in cellular evolution, serving as the primary vector for horizontal genetic exchange (6, 7). Recent evidence suggests that phage genes have been frequently subverted to provide cellular functions (8–10). The reverse is also true; host genes have been subverted by phages for their own purposes (11, 12). Such results argue that making a distinction between host and viral gene evolution may be artificial, because both systems seem to coevolve synergistically (12, 13).

More than 5,000 phages have been described, and 96% of these are tailed phages (14). Metagenomic analysis of uncultured marine viral samples indicate that sequences belonging to tailed phages dominate marine phage communities (15) and transmission electron microscopy has verified that the large majority of marine viruses are tailed (16, 17). It has been estimated that there are >10^30 tailed phage in the biosphere (18). Among the tailed phages, the myoviruses, those with contractile tails, are widespread and diverse. Quantitative transmission electron microscopy studies indicate that myoviruses constitute a major component of the marine phage communities of the North Sea and the North Atlantic (19). Furthermore, myovirus sequences represent 11–23% of all sequences obtained by metagenomic analysis of uncultured Pacific viral samples (15).

T4, the archetype of the myovirus family, can be shown by phylogenetic analysis of the head and tail genes to belong to the monophyletic group of T4-type bacteriophages (20, 21). These phages can be classed into subgroups (T-evens and PseudoT-evens, SchizoT-evens, and ExoT-evens) on the basis of this phylogeny. The most divergent group of these T4-type phages, the ExoT-evens, are marine cyanophages (22) that have only modest morphological resemblance to the T-evens. Recent sequencing of the genomes of such cyanophages indicated massive sequence divergence from the genomes of the T-evens (23, 24). However, despite this progress, our knowledge of the natural diversity of T4-type myoviruses is inadequate, because many environmental bacteria (and thus their phages) have not been cultivated. To overcome this barrier, culture-independent sequence analysis of PCR-amplified environmental DNA has been undertaken. By using PCR primers based on a structural gene (g20) found in T4-type phages infecting *Synechococcus* (25–27), these studies revealed substantial diversity. However, it is now clear that these primers amplified only a small subset of the g20 sequences of the T4-type phage present in aquatic environments.

With the objective of more accurately assessing the natural diversity of T4-type phage, we used degenerate PCR primers specific for each of two conserved sequence motifs in the major capsid gene (g23) of T4-type phages. These primers amplify g23 sequences for all of the T4-type phage so far tested, and we were able to use them to amplify DNA from a wide range of marine environments, including a 3,246-m-deep sample from the Arctic Ocean. Here, we characterize >100 g23 sequences from such samples.

Materials and Methods

Samples originated from 110 m deep in the Gulf of Mexico (sample 3B), the surface waters of the western Arctic Ocean (sample C200; 0.2-μm filtered, sample C106), 3,246 m deep in the Arctic Chuckchi Sea (sample CS), and several fjords and bays along the coast of British Columbia (Salmon Inlet surface, sample 375; Salmon Inlet at 3.6 m, sample 373; Jericho Pier surface, sample 485; Pendrell Sound at 6.2 m, sample 386; Teakerne Arm at 7–10 m, sample 440; and Malaspina Inlet surface, sample 444). For more details of the sites sampled see ref. 27. As described elsewhere (28), most organisms and particles larger than viruses were removed by filtration and virus-sized particles in the filtrate were concentrated by ultrafiltration. Subsamples (100 ml) from the Arctic Ocean were further concentrated to ≈500 μl by the use of Centricon Plus-20

EVOLUTION
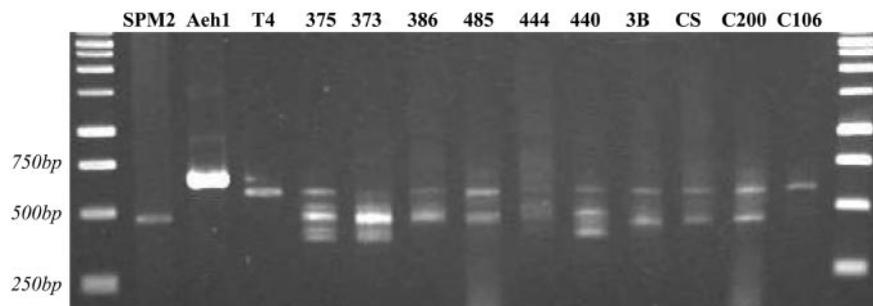
Fig. 1. Gel electrophoresis analysis of PCR-amplified g23 gene fragments. The first three lanes correspond to PCR amplification of g23 gene from the cyanophage S-PM2, the *Aeromonas* phage Aeh1, and the coliphage T4. The other lanes correspond to the g23 PCR products obtained with environmental concentrates of viruses from diverse marine locations (*Materials and Methods*).

centrifugal filters (Millipore) and then stored at −20°C. Other concentrates were stored at 4°C in the dark before analysis.

Degenerate gp23 primers MZIA1bis (5′-GATATTTGIG-GIGTTCAGCCIATGA-3′) and MZIA6 (5′-CGCGGTT-GATTTCCAGCATGATTTC-3′) were tested on phage T4 (a T-even), on *Aeromonas* phage Aeh1 (a SchizoT-even), and on the *Synechococcus* cyanophage S-PM2 (an ExoT-even).

Subsamples (35 ml) of viral concentrates were centrifuged at 85,000 × *g* for 3.5 h, and 2 ml of virus-free water was added to the virus pellets. A hot–cold technique (29) was used to extract nucleic acids from 100-μl subsamples of the viral concentrates. Two μl of each template were added to a 48-μl PCR mixture containing TaqDNA polymerase assay buffer (50 mM KCl/20 mM Tris·HCl, pH 8.4), 1.5 mM MgCl₂, a 0.20 mM concentration of each deoxyribonucleoside triphosphate, 40 pmol each of the primers and 1.0 unit of Taq polymerase (New England Biolabs). Negative controls contained all reagents and sterile water instead of a template. PCRs were performed with the following PCR cycle parameters: denaturation at 94°C for 90 s, 30 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 1 min, extension at 72°C for 45 s, and a final extension at 72°C for 5 min. PCR products were electrophoresed in 2% agarose in 0.5× Tris-acetate-EDTA, pH 8.0, at 100 V for 60 min. For five samples (3B, 373, 375, 485, and CS) a plug of agarose containing most of the amplified DNA was removed from each lane. PCR products were purified with the QIAquick Gel Extraction kit (Qiagen) and cloned with the pGEM-T Vector System I TA cloning kit (Promega). The resulting reactions were used to transform competent *Escherichia coli* JM109 (Promega). Twenty-five positive clones (white colonies) from each library were picked randomly and transferred by streaking onto agar plates. Plasmid DNAs were harvested from 3-ml, overnight cultures by using QIAprep spin miniprep kits (Qiagen). The clones were sequenced with a CEQ DTCS Quick Start kit (Beckman Coulter).

BIOEDIT (version 5.0.7) was used for sequence translation. Conserved T4-type genes of >100 aa in completely sequenced T4-type genomes were identified with the BLAST (30) program by using a cutoff value of 10^{-5}. Sequences were aligned with CLUSTALW (31) and refined manually with the help of the ED program in MUST 3.0 (32). Phylogenetic trees of the concatenated-conserved genes in genomes were constructed with maximum likelihood and distance-based methods without any distance correction (*p* distance) by using the PROML program in the PHYLIP package (33) and the program NJ in MUST, respectively. Finally, bootstrap proportions were calculated by analysis of 1,000 replicates with the SEQBOOT and CONSENSE programs in PHYLIP and the NJBOOT program in MUST.

Results and Discussion

Tailed bacteriophages are abundant in marine environments (15, 16, 34); yet the composition of these viral communities remains largely unknown. A unified molecular taxonomy is impossible because viruses contain no universally shared sequences. Consequently, studies on phage ecology and diversity require group-specific molecular markers. For this analysis, which is restricted to the T4-type phages, the customized marker must be conserved in all of these phages and give a phylogeny that is a reasonably accurate representation of that obtained for the entire genome. We have shown (20) that a phylogeny of the T4-type phage based on the conserved residues in the central portion of the major capsid protein (g23) was congruent with phylogenies based on several other essential phage functions (e.g., g18, g19, g32, and g43). To fully validate this result, we directly compared the phylogenies obtained with this g23 segment to those obtained from the T4-type phage genomes. There are 16 completely sequenced T4-type phage genomes available in the T4-Like Genome Database (http://phage.bioc.tulane.edu), and phylogenies obtained on the basis of the most conserved residues in our g23 segment or from a maximum likelihood phylogeny of a concatenation of 23 conserved T4-type genes (Fig. 4, which is published as supporting information on the PNAS web site) are topologically similar. Thus, our g23 segment appears to be a surprisingly good surrogate for the whole genome. A consequence of this result is that if a phage has a g23-like sequence, it has so far always had numerous homologs (40 to ≥100) of other T4 genes in its genome. Because no chimeric T4-type phages have been found that do not obey this rule, in the absence of a full genomic sequence, a phylogenic classification of T4-type phages based on g23 remains a useful shortcut. It is important to note that all phage groups do not have such a simple phylogenetic molecular marker. For example, the smaller and temperate lambdoid phages are subject to much greater modular genetic exchange in their essential genes than are the T4-type phages and no similarly valid phylogenetic marker has yet been identified.

Thus, in this study of phage diversity in the marine environment, we have used a PCR strategy based on the sequence of g23 to analyze the divergence of the widespread T4-type myoviruses (35). For all characterized T4-type phages, the major capsid protein is encoded by a homologue of T4 gene 23 (20). For this study of T4-type phages, we designed degenerate PCR primers that could amplify a homologous segment of the g23 sequence in all of the subgroups of T4-type phages. The primers yielded a 600-bp fragment from the coliphage T4 (a T-even), a 640-bp fragment from the aquatic *Aeromonas* phage Aeh1 (a SchizoT-even), a 480-bp fragment from the cyanophage S-PM2 (an ExoT-even), and similarly sized fragments from 10 of the 12 environmental samples examined (Fig. 1). These samples were collected from northeastern Gulf of Mexico, fjords and bays in northeastern Pacific, and the Arctic Ocean. Each positive sample generated PCR products that ranged
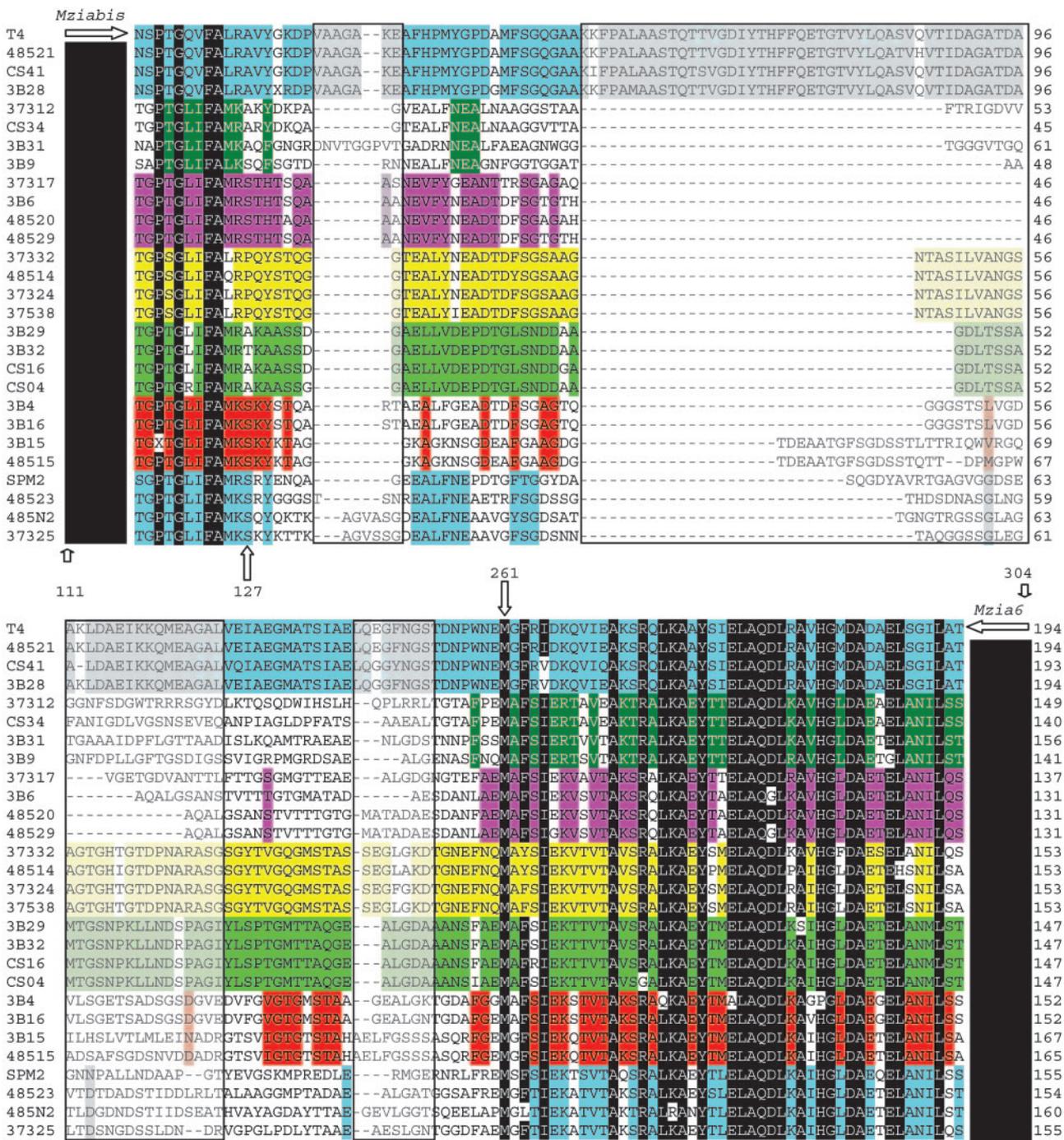
Fig. 2. Amino acid alignments of a representative subset of PCR-amplified g23 sequences. A black background indicates amino acid motifs conserved for >90% of the sequences. Amino acid motifs that are well conserved only within a subgroup of sequences are indicated by a color code. Sequences shown with a white background were not conserved among the sequences. A dash indicates that a space was inserted in the sequence to preserve the alignment. The two horizontal arrows represent the position of the two primers used for g23 amplification. The vertical arrows indicate the positions of the homologous residues in the amino acid sequence T4 g23. Positions excluded from the phylogenetic analysis are indicated by the semitransparent box.

from 380 to 600 bp. We randomly cloned and sequenced 125 of the amplification products from samples 373, 375, 485, 3B, and CS, and, with BLASTX (30), confirmed that 124 of 125 sequences were similar to T4 g23 ($10^{-97} < E$ values $< 10^{-20}$). Eighty-five of the sequences were unique, and these were used in the subsequent analyses. Although a dozen sequences were very similar to the g23 sequence of T4 (>90% amino acid identity), most sequences diverged considerably (<55% amino acid identity). Even by casual inspection, however, it was

apparent that the environmental sequences fell into subgroups (Fig. 2). Two highly conserved segments were located between amino acid residues 111–127 and 261–304 of the T4 sequence. These regions seem to have an important role in head morphogenesis (36), which could explain the strong constraints on their evolutionary divergence. Between these two conserved motifs genetic drift and insertion and deletion events have occurred. Nevertheless, even in this hypervariable segment, there are subsets of sequences that share common amino acid
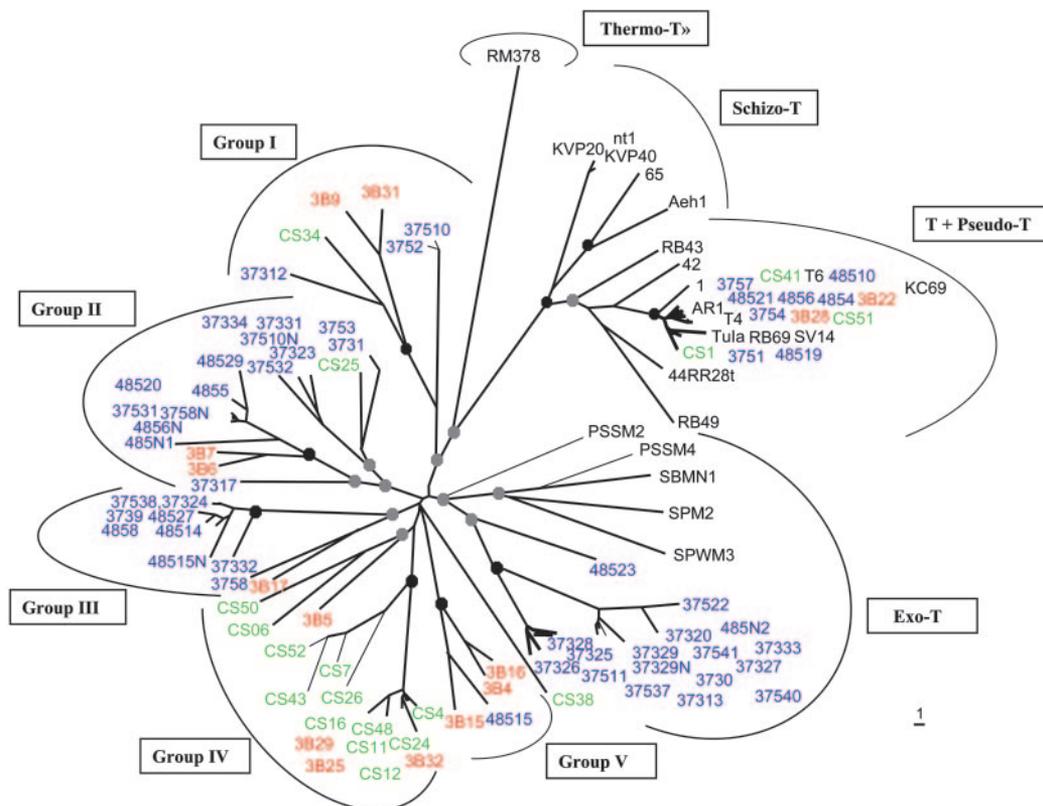
Fig. 3. Neighbor-joining tree of g23 sequences. The black circles and gray circles indicate, respectively, internal nodes with at least 95% and 50% bootstrap support. Sequences coming from the Pacific are in blue, from the Atlantic in red, and from the Arctic in green. Sequences from the database are in black. Groups I, II, III, IV, and V refer to newly identified subgroups of T4-type phages. The scale bar represents the number of amino acid substitutions per residue.

motifs, suggesting a close genetic relationship among groups of these phages. We constructed a database of the 85 environmental and 23 publicly available g23 sequences. A very recent structural analysis of the capsid protein of phage T4 reveals that the central region of gp23 from residue 164–233 contains a hypervariable loop structure (37). Because the inclusion of this hypervariable segment could lead to a misrepresentation of their phylogenetic relationships, we constructed a phylogenetic tree (Fig. 3) that uses only the most conserved residues and specifically excludes the hypervariable part of the loop structure (Fig. 2). This tree resolved the previously established subgroupings of the known T4-type phages as well as five previously uncharacterized clusters (clusters I–V). Some of the environmental sequences fell within the T-even/PseudoT-even group and the ExoT-even group (marine cyanophages), but the large majority fell within the newly resolved clusters. These data (Fig. 2) demonstrate that these sequences share a common distant ancestor and that numerous distinct and stable lineages have evolved from it.

Many T4 mutants with aberrant head morphology map within the two small highly conserved segments of g23 described above. For example, a T4 mutation (Ng191c) in the Asp-287 residue yields virions with significantly smaller ("petit") or larger ("giant") heads than the wild-type phage. In our environmental sequences, there are several sequence variations within the two "conserved" blocks, and some of these variations are perfectly conserved within subgroups. Such sequence variants could affect the size and morphology of the phage head. Because our environmental sequences are highly variable in length and sequence and because this segment of g23 determines the form of the head, it is probable that the head morphology of these phages differs from that of

the known T4-type phages. Indeed, microscopic observation of T4-type phage isolates has revealed variation in capsid size (38), which, together with the present data from environmental sequences, suggest that morphological diversity is probably underestimated. For example, the ExoT-even phage S-PM2 has a capsid of 85 × 85 × 85 nm and T4 of 78 × 78 × 111 nm; whereas, the SchizoT-even phage Aeh1 has a significantly bigger capsid of 78 × 78 × 137 nm. The smallest g23 PCR fragment found in our collection (samples 37317, 3B6, 48520, or 48529) is only 400 bp. It may be that such a sequence would produce a smaller phage when compared with T4. However, because the head size depends not only on the intrinsic size of the constituent gp23 subunits but also on the number of monomers and their packing in the phage head, the prediction of head size variation may not be so simple.

Some of the new T4-type phage subgroups have a wide geographical distribution. For example, subgroups I and II are present in all of the samples. Whereas subgroup IV is present in the Arctic and Gulf of Mexico, subgroup III is found mainly in the Pacific samples, and, finally, subgroup V is largely restricted to the Gulf of Mexico. The heterogeneous geographic distribution of the various T4-type phages implies that at least some of their host bacteria are probably also heterogeneously distributed in the marine environment. Interestingly, a few T4-type phages had been isolated from aquatic environments (38), but these infect only a fraction of marine bacterial species (*Vibrio* spp. and several cyanobacteria). Our results suggest that T4-like phages infect a broader repertory of hosts than previously appreciated, including the abundant and ubiquitous groups of bacteria in the sea, such as the SAR11 group of α-proteobacteria (39). Thus, T4-type phages could play a more important role in marine ecosystems than previously suspected.

As remarked above, some marine sequences are very similar to coliphage T4 (90–95% amino acid identities). Although several of these sequences were obtained from a deep and cold Arctic Ocean sample, they occurred most frequently in samples from Jericho Pier and Salmon Inlet (44% and 14% of the sequences analyzed per location, respectively). The host of all known T-even phage are enterobacteria that primarily inhabit the animal gut, but these bacteria are believed to be essentially absent from the open ocean (40, 41). Because Jericho Pier is adjacent to the city of Vancouver, in an area influenced by runoff from the Fraser River and subject to coliform contamination, the hosts for these phages could easily be enterobacteria, and a similar explanation is even possible for the phages from the more pristine Salmon Inlet. In contrast, it is unlikely that fecal coliforms are present 3 km deep in the Arctic Ocean, suggesting that T-even phage infect free-living marine bacterial hosts. It is conceivable that these phage may infect bacteria derived from the feces of marine animals, but this seems unlikely given the generally short half-life of phages, at least in surface seawater (42).

A number of our environmental sequences are closely related to the g23 sequences of T4-type cyanophage that were isolated on *Synechococcus* sp. (22). These sequences fall among the ExoT-even subgroup, a conclusion in accord with Short and Suttle (27). In their analysis of cyanophage diversity based on the vertex–portal–protein gene (g20), they concluded that half of the marine phage sequences belonged to the group of T4-type cyanophages that infect *Synechococcus* spp. They noted, however, that it is not clear what fraction of the remaining sequences comes from unclassified cyanophages or from phages infecting other bacteria. By using g23 primers, we obtained sequences related to the ExoT-even subgroup only in the surface samples, in which *Synechococcus* spp. are abundant (e.g., Salmon Inlet and Jericho Pier). In the deeper samples (Gulf of Mexico, 110 m), in which cyanophages are less abundant (20, 43–45) or probably absent (Arctic Ocean at 3,246 m), we did not obtain such sequences. These findings suggest that the g23 sequences within the ExoT-even group come from phages infecting *Synechococcus* or closely related bacterial species rather than phages infecting heterotrophic bacteria.

Although our primers efficiently amplified a PCR fragment from a freshwater SchizoT-even phage (Aeh1), we detected no marine environmental sequences that had a close relationship to any of the known phages in the SchizoT-even subgroup. Because one SchizoT-even phage [KVP40 (46)] has been isolated from a marine environment, our inability to detect members of this group may reflect some sequence bias inherent in our PCR primers.

Recently, the genome of a thermophilic myovirus, RM378, was sequenced (47). Our analysis of the genome of this *Rhodothermus marinus* phage indicates that it represents an example of a subgroup of the T4-type phages that we call "ThermoT-even."

Our results indicate that there are at least nine different subgroups of the T4-type phage, only four of which are represented by phages described in the literature. An interesting aspect of our results is that all of the environmental samples contained representatives of several different subgroups of T4-type phages. For example, the Gulf of Mexico samples contained sequences belonging to subgroups I, II, IV, and V and the T-evens, and the deep-Arctic sample had sequences belonging to subgroups I, II, and IV and the T-evens. These observations point out the remarkable diversity of the T4-type phage, even within a single seawater sample. Such results gain additional significance when one considers the peculiarities of phage evolution. Phage genomes appear to be a patchwork of DNA from disparate origins (48). In this

view, horizontal gene transfer would have occurred frequently between the diverse phages that share a common habitat. Many studies have documented the extent of mosaicism in the temperate *E. coli* phage (e.g., λ- and μ-like phages; for a review, see ref. 49), dairy phages (50), or Mycobacteriophages (51). However, among the large lytic phages, such as the T4-type, it appears that all of the related isolates share a common distant ancestor for a large fraction of their genomes, and modular swapping of the genes that encode essential functions is less common. However, some parts of these genomes are hyperplastic, but these parts seem to be primarily concerned with adaptation of the phage to a particular host or environmental niche (refs. 23, 46, and 52 and our unpublished results). The difference in the extent of modular swapping in the essential and nonessential segments of the T4-type genome may be explained by the fact that many of the critical genes have complex and highly specific interactions with other essential genes that are scattered about the core genome, which could severely restrict their modular shuffling. Such restraints are presumably much weaker for the facultative genes in the hyperplastic regions. Because of the efficient homologous T4-type recombination system, substantial horizontal gene transfer of such genes could occur between T4-type phages with overlapping host range. Because the genomes of T4-type phages are largely syntenous for essential genes, exchanges in any region of substantial homology will generate recombinants with intact genomes without loss of critical genes. There are numerous studies that indicate that gene transfer occurs most frequently between cellular organisms growing in the same niche (53–56). Similarly, T4-type recombinants are probably much more frequently created in niches that are densely populated by diverse T4-type phages (57). Clearly, the cohabitation of the diverse T4-type phages in various niches of the marine ecosystem and their very high levels of recombination (58) could result in substantial gene transfer between them.

Conclusion

The present study indicates that our knowledge of phage diversity in the biosphere is woefully inadequate and that what is known is biased because of sampling errors arising from the inability to propagate most of this biomass in the laboratory. Here, we show that T4-type phages are not only prevalent in a wide range of oceanic biotopes but are also extraordinarily diverse. These results suggest that the host repertory of T4-type phage is much more extended than previously suspected. For such reasons, we believe that the isolation and biological characterization of these newly identified groups of T4-type bacteriophages, coupled with the sequencing of their genomes, would be extremely informative. This view is based on the fact that the previous discovery of the T4-like cyanophages (11, 24) rapidly lead to the unexpected and exciting finding that such phages often encode and express photosynthesis genes. Among the five groups of T4-type phage described here, it seems very likely that at least a few startling insights will emerge on the role that phages play in the environment and in the evolution of their hosts and the biosphere.

EVOLUTION

1. Balter, M. (2000) *Science* **289,** 1866–1867.
2. Fuhrman, J. A. (1999) *Nature* **399,** 541–548.
3. Rohwer, F. (2003) *Cell* **113,** 141.
4. Wilhelm, S. W. & Suttle, C. A. (1999) *Bioscience* **49,** 781–788.
5. Weinbauer, M. G. (2004) *FEMS Microbiol. Rev.* **28,** 127–181.
6. Paul, J. H. (1999) *J. Mol. Microbiol. Biotechnol.* **1,** 45–50.
7. Waldor, M. K. & Mekalanos, J. J. (1996) *Science* **272,** 1910–1914.
8. Forterre, P. (1999) *Mol. Microbiol.* **33,** 457–465.
9. Filee, J., Forterre, P., Sen-Lin, T. & Laurent, J. (2002) *J. Mol. Evol.* **54,** 763–773.
10. Filee, J., Forterre, P. & Laurent, J. (2003) *Res. Microbiol.* **154,** 237–243.
11. Millard, A., Clokie, M. R., Shub, D. A. & Mann, N. H. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11007–11012.
12. Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F. & Chisholm, S. W. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11013–11018.
13. Krisch, H. M. (2003) *Res. Microbiol.* **154,** 227–229.
14. Ackermann, H. W. (2001) *Arch. Virol.* **146,** 843–857.
15. Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F. & Rohwer, F. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14250–14255.
16. Wommack, K. E. & Colwell, R. R. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 69–114.
17. Borsheim, K. Y. (1993) *FEMS Microbiol. Ecol.* **102,** 141–159.
18. Brussow, H. & Hendrix, R. W. (2002) *Cell* **108,** 13–16.
19. Frank, H. & Moebus, K. (1987) *Helgol. Meeresunters.* **41,** 385–414.
20. Tetart, F., Desplats, C., Kutateladze, M., Monod, C., Ackermann, H. W. & Krisch, H. M. (2001) *J. Bacteriol.* **183,** 358–366.
21. Desplats, C. & Krisch, H. M. (2003) *Res. Microbiol.* **154,** 259–267.
22. Hambly, E., Tetart, F., Desplats, C., Wilson, W. H., Krisch, H. M. & Mann, N. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 11411–11416.
23. Mann, N. H., Clokie, M. R., Millard, A., Cook, A., Wilson, W. H., Wheatley, P. J., Letarov, A. & Krisch, H. M. (2005) *J. Bacteriol.* **187,** 3188–3200.
24. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. (2005) *PLoS Biol.* **3,** e144.
25. Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. & Hodson, R. E. (2002) *Appl. Environ. Microbiol.* **68,** 1576–1584.
26. Dorigo, U., Jacquet, S. & Humbert, J. F. (2004) *Appl. Environ. Microbiol.* **70,** 1017–1022.
27. Short, C. M. & Suttle, C. A. (2005) *Appl. Environ. Microbiol.* **71,** 480–486.
28. Suttle, C. A., Chan, A. M. & Cottrell, M. T. (1991) *Appl. Environ. Microbiol.* **57,** 721–726.
29. Chen, F., Suttle, C. A. & Short, S. M. (1996) *Appl. Environ. Microbiol.* **62,** 2869–2874.
30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
31. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
32. Philippe, H. (1993) *Nucleic Acids Res.* **21,** 5264–5272.
33. Felsenstein, J. (1989) *Cladistics* **5,** 164–166.
34. Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P. & Rohwer, F. (2003) *J. Bacteriol.* **185,** 6220–6223.
35. Ackermann, H. W. (2003) *Res. Microbiol.* **154,** 245–251.
36. Black, L. W., Showe, M. K. & Steven, A. C. (1994) in *Molecular biology of Bacteriophage T4*, ed. Karam, J. D. (Assoc. Soc. Microbiol., Washington, DC), pp. 218–259.
37. Fokine, A., Leiman, P. G., Shneider, M. M., Ahvazi, B., Boeshans, K. M., Steven, A. C., Black, L. W., Mesyanzhinov, V. V. & Rossmann, M. G. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 7163–7168.
38. Ackermann, H. W. & Krisch, H. M. (1997) *Arch. Virol.* **142,** 2329–2345.
39. Morris, R. M., Rappe, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A. & Giovannoni, S. J. (2002) *Nature* **420,** 806–810.
40. Fuhrman, J. A., McCallum, K. & Davis, A. A. (1993) *Appl. Environ. Microbiol.* **59,** 1294–1302.
41. Bano, N., Ruffin, S., Ransom, B. & Hollibaugh, J. T. (2004) *Appl. Environ. Microbiol.* **70,** 781–789.
42. Suttle, C. A. & Chen, F. (1992) *Appl. Environ. Microbiol.* **58,** 3271–3729.
43. Ting, C. S., Rocap, G., King, J. & Chisholm, S. W. (2002) *Trends Microbiol.* **10,** 134–142.
44. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. (2003) *Nature* **424,** 1047–1051.
45. Suttle, C. A. & Chan, A. M. (1994) *Appl. Environ. Microbiol.* **60,** 3167–3174.
46. Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., Feldblyum, T. V., White, O., Paulsen, I. T., Nierman, W. C., *et al.* (2003) *J. Bacteriol.* **185,** 5220–5233.
47. Blondal, T., Hjorleifsdottir, S. H., Fridjonsson, O. F., Aevarsson, A., Skirnisdottir, S., Hermannsdottir, A. G., Hreggvidsson, G. O., Smith, A. V. & Kristjansson, J. K. (2003) *Nucleic Acids Res.* **31,** 7247–7254.
48. Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2192–2197.
49. Hendrix, R. W. (2003) *Curr. Opin. Microbiol.* **6,** 506–511.
50. Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G. F., Desiere, F. & Brussow, H. (2002) *J. Bacteriol.* **184,** 6026–6036.
51. Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., *et al.* (2003) *Cell* **113,** 171–182.
52. Desplats, C., Dez, C., Tetart, F., Eleaume, H. & Krisch, H. M. (2002) *J. Bacteriol.* **184,** 2789–2804.
53. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1998) *Trends Genet.* **14,** 442–444.
54. Wolf, Y. I., Aravind, L. & Koonin, E. V. (1999) *Trends Genet.* **15,** 173–175.
55. Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000) *Nature* **407,** 508–513.
56. Gallois, J. L., Achard, P., Green, G. & Mache, R. (2001) *Gene* **274,** 179–185.
57. Sandmeier, H. (1994) *Mol. Microbiol.* **12,** 343–350.
58. Mosig, G., Gewin, J., Luder, A., Colowick, N. & Vo, D. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 8306–8311.