JOURNAL OF BACTERIOLOGY, Jan. 2001, p. 358–366 Vol. 183, No. 1
0021-9193/01/$04.00+0 DOI: 10.1128/JB.183.1.358–366.2001
Copyright © 2001, American Society for Microbiology. All Rights Reserved.

Phylogeny of the Major Head and Tail Genes of the Wide-Ranging T4-Type Bacteriophages†

FRANÇOISE TÉTART,[1] CARINE DESPLATS,[1] MZIA KUTATELADZE,[2] CAROLINE MONOD,[1] HANS-WOLFGANG ACKERMANN,[3] AND H. M. KRISCH[1]*

Laboratoire de Microbiologie et Génétique Moléculaire du CNRS, Toulouse, France[1]; George Eliava Institute of Bacteriophage, Microbiology and Virology, Georgian Academy of Sciences, Tbilisi, Georgia[2]; and Félix d'Hérelle Reference Center for Bacterial Viruses, Laval University, Quebec, Quebec, Canada[3]

Received 16 June 2000 /Accepted 4 October 2000

We examined a number of bacteriophages with T4-type morphology that propagate in different genera of enterobacteria, *Aeromonas*, *Burkholderia,* and *Vibrio*. Most of these phages had a prolate icosahedral head, a contractile tail, and a genome size that was similar to that of T4. A few of them had more elongated heads and larger genomes. All these phages are phylogenetically related, since they each had sequences homologous to the capsid gene (gene *23*), tail sheath gene (gene *18*), and tail tube gene (gene *19*) of T4. On the basis of the sequence comparison of their virion genes, the T4-type phages can be classified into three subgroups with increasing divergence from T4: the T-evens, pseudoT-evens, and schizoT-evens. In general, the phages that infect closely related host species have virion genes that are phylogenetically closer to each other than those of phages that infect distantly related hosts. However, some of the phages appear to be chimeras, indicating that, at least occasionally, some genetic shuffling has occurred between the different T4-type subgroups. The compilation of a number of gene *23* sequences reveals a pattern of conserved motifs separated by sequences that differ in the T4-type subgroups. Such variable patches in the gene *23* sequences may determine the size of the virion head and consequently the viral genome length. This sequence analysis provides molecular evidence that phages related to T4 are widespread in the biosphere and diverged from a common ancestor in acquiring the ability to infect different host bacteria and to occupy new ecological niches.

Bacteriophages, the viruses that infect bacteria, are ubiquitous in the environment (5, 10, 11, 39). New techniques have made it possible to directly and rapidly estimate the total abundance of phages in environmental samples (10, 24). The high phage titers found by such methods in the soil and in aquatic habitats, such as the sea ($>10^6$/ml), indicate that phages constitute a large fraction of the total biomass. Phages are probably the most abundant biological entities on the planet (15), but their diversity makes it extremely difficult to assess their impact on the biosphere (10) because, in general, each phage is infectious for only a minute fraction of the bacteria in its surroundings. However, the aggregated phage population must significantly influence microbial ecology and consequently affect the entire ecosystem.

Until recently, remarkably little was known about phage diversity because most research was focused on only a few laboratory isolates. The T phages (7) isolated on *Escherichia coli* B were long used as the paradigms for all virulent phages, but these phages do not satisfactorily represent even virulent coliphages, let alone phages in general. For example, of the original seven T phages, the three even-numbered viruses (T2, T4, and T6) are identical in morphology and are very closely related (16).

T4, the archetype of the T-even phages, has been the subject of intensive study and is one of the best-characterized phages. Although it was demonstrated a number of years ago that urea-treated T4 could propagate in the spheroplasts of a broad spectrum of bacterial species (38), only recently has it become clear that T4, as well as the other laboratory phages such as λ or Mu, has numerous relatives in nature that infect a wide variety of bacterial species (2, 12, 15, 32). A recent survey of the literature revealed more than 140 descriptions of phages with a morphology that resembled that of T4 (2). Many of these T4-like phages had been isolated on enterobacterial species closely related to *E. coli* (*Klebsiella*, *Shigella*, and *Yersinia)* and less frequently on *Citrobacter, Proteus, Salmonella*, and *Serratia*. Others propagate on more distantly related bacteria (*Acinetobacter, Aeromonas, Burkholderia*, and vibrios). A previous PCR analysis of many of the T4-like phages had identified a subgroup of these phages, named the pseudoT-evens, whose genomes are substantially diverged from those of T-even phages (23); for example, the sequences of RB49, the pseudoT-even phage that is the best characterized, are invariably less than 70% identical to the corresponding T4 genes (23).

In this communication we examine the phylogenetic relationships of T4-like phages isolated on diverse bacterial species.

MATERIALS AND METHODS

Phages and bacteria. All of the T4-type phages are from the Toulouse collection (70 isolates). Phages T2, T4, and T6 were obtained from R. H. Epstein of the University of Geneva in Switzerland. Except for coliphage KC69 (obtained from K. Carlson, University of Uppsala, Uppsala, Sweden), the sources of the T4-type phages (coliphages Tu1a, SV14, RB69, RB49, RB42, and RB43; *Enterobacter cloacae* phage 1; *Aeromonas salmonicida* phages 44RR2.8t (synonym, 44RR) and

* Corresponding author. Mailing address: Laboratoire de Microbiologie et Génétique Moléculaire, CNRS UMR 5100, 118 Route de Narbonne, 31062 Toulouse Cedex, France. Phone: (33) 5 61 33 58 81. Fax: (33) 5 61 33 58 86. E-mail: krisch@ibcg.biotoul.fr.

† This paper is dedicated to the memory of Ulf Henning of Tübingen, Germany.

65; *Aeromonas hydrophila* phage Aeh1; *Burkholderia cepacia* phage 42; *Vibrio natriegens* phage nt-1; and *Vibrio parahaemolyticus* phages KVP20 and KVP40) are found in reference 2. For phages that grow on *E. coli* B[E], standard techniques were used for their propagation and the preparation of their DNA (23). The T4-type phages of non-*E. coli* hosts were grown as suggested in the information sheets provided with the phages by the Félix d'Hérelle Reference Center for Bacterial Viruses (Quebec, Quebec, Canada).

Electron microscopy. Cycles of differential centrifugation were used to purify some of the phages for electron microscopy (23). The non-*E. coli* phages Aeh1, KVP20, KVP40, nt-1, 42, 44RR, and 65 to be examined by electron microscopy were sedimented at 25,000 $\times$ *g* for 60 min in a JA-18.1 rotor using a Beckman J2–21 centrifuge. This was followed by two washes in 0.1 M ammonium acetate (pH 7.0). Particles were deposited on copper grids with carbon-coated Formvar films and were stained with 2% potassium phosphotungstate (pH 7.2) or 2% uranyl acetate (pH 4.0). Magnification was controlled with catalase crystals (19).

PCR and oligonucleotide primers. The consensus primers used to amplify the central portion of gene *23* of the various T4-type phages were Mzia1 (5′-TGT TATIGGTATGGTICGICGTGCTAT-3′) and CAP8 (5′-TGAAGTTACCTTC ACCACGACCGG-3′). These primers flank the sequences shown in the compilation (see Fig. 3), and they correspond respectively to the nucleotide sequences encoding T4 gene *23* amino acid (aa) residues 95 to 103 and 368 to 375 (26). The conditions used for the amplification reaction with these primers involved 28 cycles consisting of a 30-s denaturation at 96°C, a 2-min annealing at 62°C, and a 3-min extension at 72°C. With this protocol, an approximately 850-bp PCR product was produced from the DNA of phages T4, T6, Tu1a, KC69, SV14, RB69, RB49, RB42, RB43, 42, 44RR, 1, nt-1, 65, and Aeh1.

The primers initially used to amplify the gene *18* analogue of the T4-type phages were FT18-N2 (5′-GGTAAATTCCAATGGGGTCCAGCTT-3′) and FT18-C1 (5′-TATCAGCAGCCAACGGAACCCAA-3′) or FT18-C3 (5′-ATGT TAAACAGACGACGAACGTTAAT-3′). These oligonucleotides are based on the sequence motifs conserved in T4 and RB49 gene *18*. They respectively correspond to the T4 gene *18* sequences encoding aa residues 31 to 38, 467 to 473, and 551 to 558 (4). For the gene *18* amplification, the reactions involved 10 cycles consisting of a 10-s denaturation at 94°C, a 30-s annealing at 54°C, and a 10-s extension at 68°C, followed by an additional 15 cycles with a 10-s denaturation at 94°C, a 30-s annealing at 54°C, and a 15-s extension at 68°C. Hot Tub polymerase from Amersham International plc (Little Chalfont, United Kingdom) was used.

PFGE of phage genomes. For pulsed-field gel electrophoresis (PFGE) of phage genomes, blocks of 1% agarose containing 10^5 to 10^6 PFU were incubated overnight at 55°C in a lysis buffer containing 0.5 M EDTA, 10 mM Tris-HCl (pH 8), 1% sodium dodecyl sulfate, and proteinase K (0.5 mg/ml). These blocks were then dialyzed three times for 1 h against 10 ml of TE (Tris-EDTA). One-third of the block was analyzed by electrophoresis for 14 h in 1% agarose (0.5$\times$ TBE [1$\times$ TBE is 2 mM EDTA plus 90 mM Tris-borate]) at 275 V with a pulse time of 8 s on the Pulsaphor Plus system (Pharmacia & Upjohn AB, Uppsala, Sweden). The gel was then stained with ethidium bromide for 30 min, destained for 15 min, and photographed.

PCR sequencing. The PCR products were purified (6) and sequenced with an Amersham Life Science Thermo Sequenase kit. The gene *18, 19*, and *23* nucleotide sequences of the various T4-type phages were determined by a primer walking procedure (27).

Nucleotide sequence accession numbers. The nucleotide sequences of the central portion of gene *23* of T4-type phages have been deposited in the GenBank database under accession no. AF221994 to AF222003. The sequences of the entire gene *18* of T4-type phages RB49, 42, and nt-1 have been deposited in the GenBank database under accession no. Z78090, AF222058, and AF222059, respectively. The nucleic acid sequences of gene *19* of phages 42, RB49, and nt-1 have been deposited in the GenBank database under accession no. AF223001 to AF223003, respectively.

RESULTS

T-even and pseudoT-even subgroups of T4-type phages. The term "T4-type" refers to all of the 140 known phages with a virion morphology that generally resembles phage T4 (1, 2). Most of these phages have been isolated on various enterobacterial hosts, primarily *E . coli, Shigella*, and *Klebsiella*. The T4-type morphology is characterized by a moderately elongated icosahedral head (111 by 78 nm) connected by a collar to a contractile tail. The contractile tail (113 by 16 nm) is terminated by a base plate that carries six long, kinked tail fibers. These tail fibers are held in a folded configuration by "whisker fibers" extending from the collar (40). All of the phages initially in the Toulouse collection had T4-type morphology, and it was assumed, on this basis, that they all belonged to the T-even phages (T2, T4, and T6). Among the T-even phages the nucleotide sequences of homologous genes typically differ from each other by less than 5% (18, 22, 23, 29, 35). However, it was demonstrated (23) that some of the phages with T4-type morphology had genome sequences that were evolutionarily distant from the T-evens (namely, the coliphages RB42, RB43, and RB49 as well as the *Aeromonas* phage 44RR). In these phages, which we called the pseudoT-evens, only a small portion (about 10%) of the DNA hybridizes to the T4 genome under stringent conditions. In RB49 the nucleotide sequence encoding the major virion capsid protein was among those that diverged the least, about 30%, from the corresponding T4 sequence (23). Sequences of the RB49 genome containing the early and DNA replication genes had diverged even more from T4 (23; C. Desplats and H. M. Krisch, unpublished data). In addition, random genomic sequencing indicated that about a third of the RB49 DNA has no homology at all to T4.

Several of the T4-type phages (Tu1a, RB69, and SV14) that we previously analyzed seem to occupy an "intermediate" position between T-even and pseudoT-even phages (23). For example, phage RB69 appears to be a chimera composed of segments derived from both T-even and pseudoT-even genomes. Its gene *43* sequence differs from T4 at a level that would be expected for a pseudoT-even phage (42); yet, as will be shown, its gene *23* sequence is typical of that of a T-even phage. A similar genomic chimerism was indicated by DNA hybridization studies and sequence analysis of a few of the genes of Tu1a and SV14 (23; N. Vanzo and H. M. Krisch, unpublished data). Thus, there is good evidence that genetic exchanges occur between the T-even and pseudoT-even groups of phages, at least occasionally.

A more distant subgroup of T4-type phages, the schizoT-evens. We performed electron microscopy on a number of additional T4-type phages that were isolated on hosts other than *E. coli* (1, 2). Although some of them (e.g., *B. cepacia* phage 42) were indistinguishable from T4 (Fig. 1), others had slightly longer heads about 137 nm in length (phage Aeh1 of *A. hydrophila*, phage 65 of *A. salmonicida*, and phage nt-1 of *V. natriegens*). The recently described T4-like phages KVP20 and KVP40 of *V. parahaemolyticus* (20) have a morphology identical to that of *V. natriegens* phage nt-1 (data not shown). To distinguish phages with aberrant head morphology from the T4-type phages with the standard T4 morphology (the T-evens and pseudoT-evens), we propose to name them the schizoT-evens.

Genome size of T4-type phages. Since T4 packages its genome by a head-filling mechanism (30), any change in head size should simultaneously alter the size of the viral genome. The genome of T4 has been completely sequenced and is 169 kb in length (GenBank accession no. AF158101). Electron micrographs of other T-even genomes (T2 and T6) indicated a size close to that of T4 (16). We compared the sizes of the genomes of some of the pseudo- and schizoT-even phages to that of T4 by PFGE. As shown in Fig. 2, although the pseudoT-even phages had genomes with sizes comparable to that of T4,

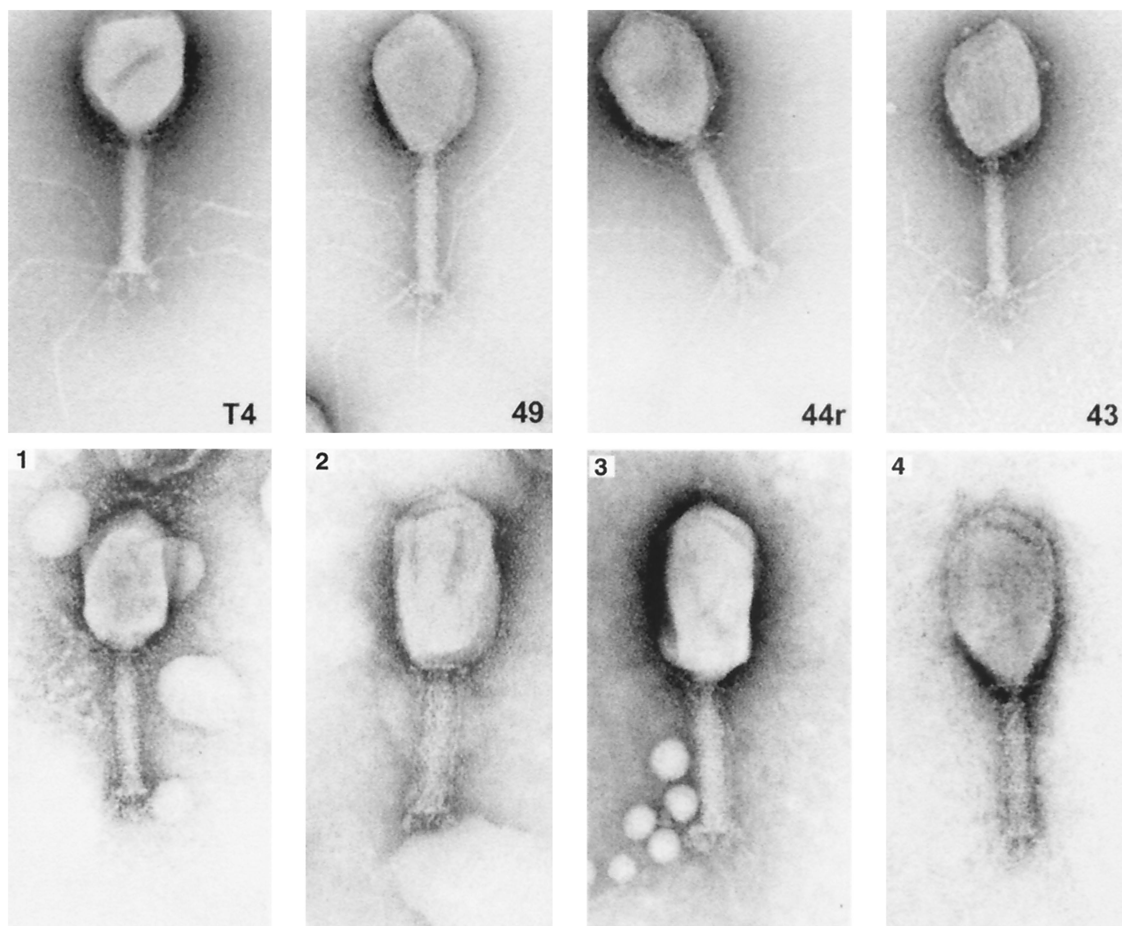

FIG. 1. Micrographs in the top row represent T4 and three previously characterized pseudoT-even phages (RB49, 44RR, and RB43) stained with phosphotungstate (23). Below are four additional T4-type phages that were isolated on nonenterobacterial hosts. (1) *B. cepacia* pseudoT-even phage 42; (2) *A. hydrophila* schizoT-even phage Aeh1; (3) *V. natriegens* schizoT-even phage nt-1; and (4) *A. salmonicida* schizoT-even phage 65. Magnification, about 150,000.

the schizoT-even phages, with their more elongated heads, had significantly larger genomes. The genome of phage Aeh1 from *Aeromonas* was estimated to be approximately 230 kb. Matsuzaki et al. (20) have recently reported that phage KVP40, which is closely related to nt-1, has a genome size of approximately 250 kb. The additional DNA sequence present in the schizoT-even phages is thus sufficient to encode more than 50 supplementary genes.

Sequence of the major capsid gene (23) in diverse T4-type phages. We had previously designed a pair of degenerate primers based on two regions in the gene *23* sequence of the phages T4, T6, SV14, and RB49 that are conserved (23). With these primers we could generate a gene *23*-specific PCR product with many of the T4-type phages, including members of T-even, pseudoT-even, and schizoT-even subgroups. The gene *23* sequences of some of the more distantly related T4-type phages were directly sequenced, including *V. natriegens* phage nt-1, *B. cepacia* phage 42, and *E. cloacae* phage 1. The inclusion of these gene *23* sequences in the alignment allowed us to further refine the consensus gene *23* primers and to amplify gene *23* from additional phages. The sequences of gene *23* from *A. hydrophila* phage Aeh1 and *A. salmonicida* phage 65 were determined from such PCR products. Both of these phages and nt-1 have elongated heads identical in length. Figure 3 shows the alignment of these sequences. Although this comparison reveals significant aa sequence differences between the phages, some blocks of aa sequence are universally conserved. The 70-amino-acid sequence between residues 230 and 300 of the T4 capsid gene is the most notable of these (Fig. 3). Many of the T4 mutants with aberrant head formation map within this conserved segment of gene *23*. An important role of this sequence in head morphogenesis would explain the unusually strong constraints on its divergence. Interestingly, a sequence closely related to these 70 aa is also found in T4 gene *24*, the head vertex protein. Gene *24* is believed to have arisen from an ancient duplication of gene *23.* The interactions between gp23 and gp24 (9, 21) are now thought to have only a minor role in the determination of virion shape and size (14). Some of the phylogenetically variable patches in gp23, however, could slightly modify the interactions between the gene *23* subunits in the virion, and this could result in differences in virion head size and shape. Other patches of variable sequence in gene *23* could be the sites that bind the various head accessory proteins on the surface of the capsid (41).

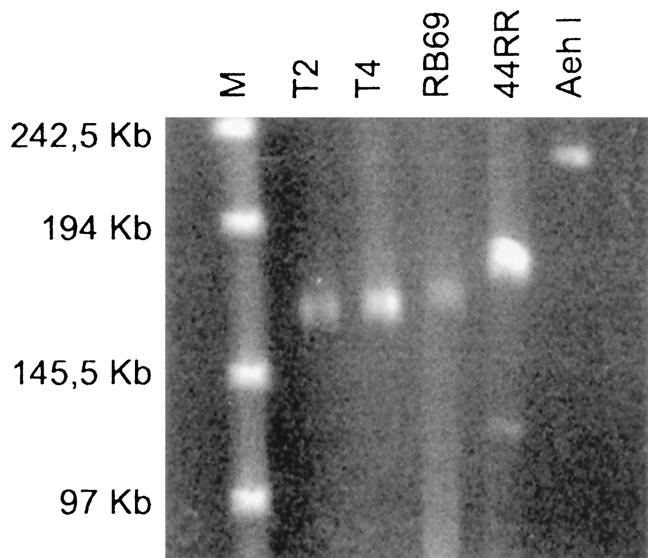

FIG. 2. PFGE of the genomes of phages T2, T4, RB69, 44RR, and Aeh1. There are differences in nucleotide modifications and glucosylation of the DNA in these phages (23); thus, besides their size, the migration of the genomes may be slightly influenced (5 to 10%) by such factors. The DNA minor band present in the 44RR lysate probably corresponds to a small fraction of heads that have an isometric morphology; a similar band was also detected in T4 lysates, where electron microscopy revealed some petite (isometric) variants. To the left is a lane containing size markers consisting of ligated concatemers of the phage λ genome.

Our alignment of 18 T4-type gene *23* sequences suggests that this phage gene has a modular or mosaic design. In a modular gene, recombination occurring in the flanking conserved motifs could replace a variable sequence by a different version. Further evidence for the modularity of gene *23* is illustrated by the sequence comparison in Fig. 4. For more than 90% of the gene *23* aa sequence, phage SV14 diverges from the T4 sequence by 5 to 10% (23), but in a 50-aa segment towards the center of the gene (Fig. 4A), it diverges by 50%. A similar, but less pronounced, increase in the divergence of the same sequence is also found (Fig. 4B) in phage AR1 (GenBank accession no. AAD01755). This is what would be expected if these phages had acquired this small segment of gene *23* by a genetic swap with a distantly related phage. Such events, even if extremely rare, could have major evolutionary consequences. For example, some nonconserved sequences in the gene *23* sequence must provide binding sites for the various head accessory proteins (41) that interact with exposed motifs of gp23 on the capsid surface. Altering the accessory capsid proteins would change the physical and antigenic properties of the virion head.

Sequences of major tail genes *18* and *19*. The contractile tail of the T-even phages has a complex structure that is perhaps the most reliable diagnostic feature of the T4-type phages. In T4, gene *18* encodes the tail sheath protein. The assembly of 144 molecules of gp18 into the sheath (24 stacked annuli each

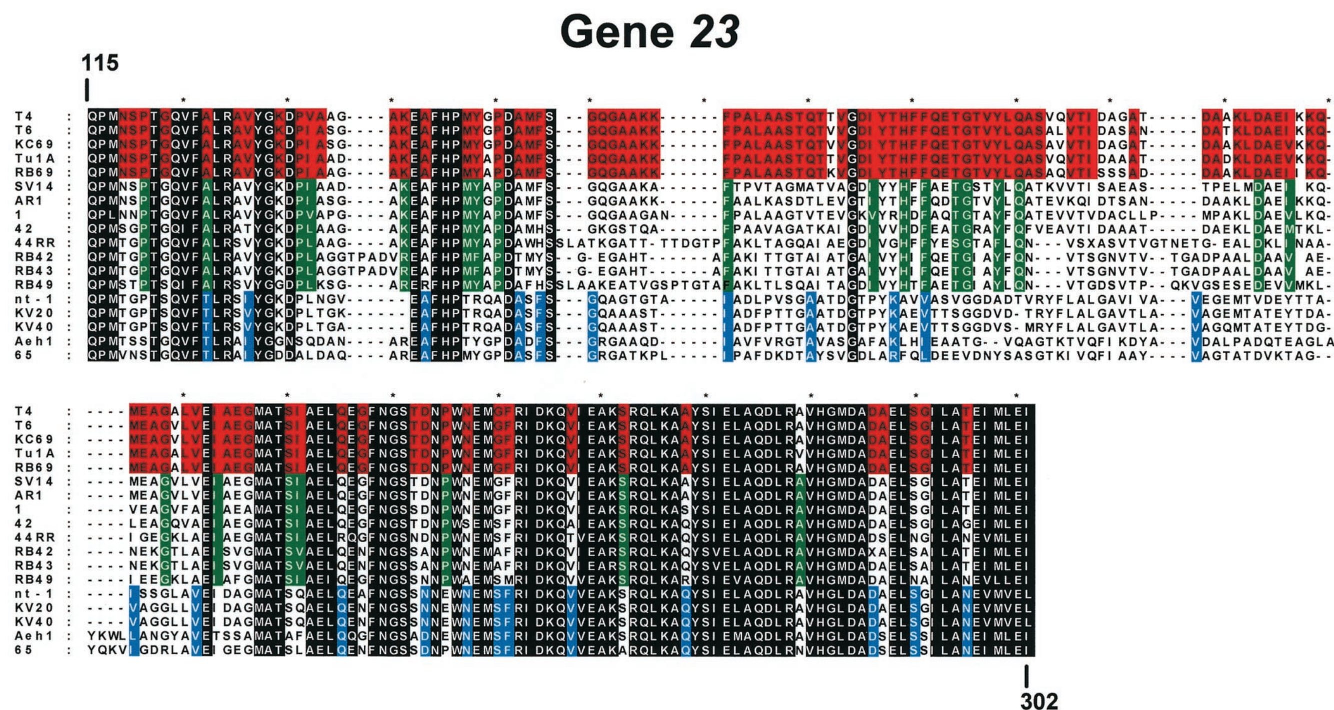

FIG. 3. Amino acid sequence alignments of the central portion of gene *23* of T4-type phages. For T4, the translation of the nucleotide sequence between codons 115 and 302 is shown. The protein sequences of the other 17 phages were aligned with the T4 sequence using the ClustalX program (36). Amino acid motifs common to all of the T4-type subgroups are indicated by a black background. Amino acid motifs that are well conserved within a T4-type subgroup phage are indicated by a color code (red, T-evens; green, pseudoT-evens; and blue, schizoT-evens). Sequences shown with a white background were not well conserved within their subgroup. The following groups of aa were considered equivalent in this presentation (D = N, E = Q, K = R, F = Y = W, and L = I = V = M). A dash indicates a space was inserted in the sequence to preserve the alignment. An ambiguity in the nucleotide sequence is indicated by an X for the corresponding codon. The coordinates of the T4 gene *23* aa sequence are shown at the extremities of the comparison in bold characters.

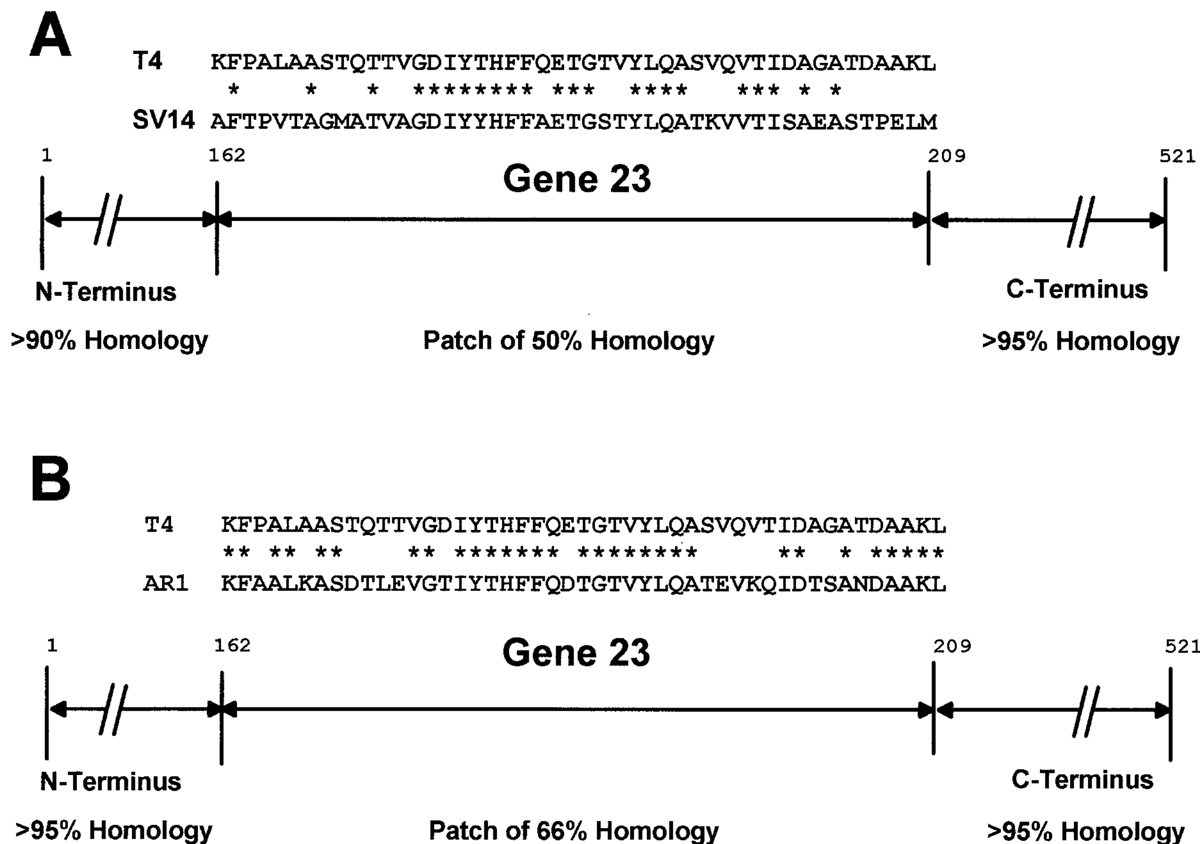

FIG. 4. (A) Comparison between the sequences of gene *23* in phages T4 and SV14. The level of homology is shown below the diagram for the N- and C-terminal segments as well as for a small patch of sequence near the center of the gene. The comparison of this divergent patch in each phage sequence is shown above the diagram. Asterisks between the alignments of the two sequences indicate identical residues in both genes. In the patch between aa 162 and aa 209, the SV14 DNA sequence differs by 47% from the T4 sequence, while in the remainder of the gene the divergence is 12%. (B) Comparison between sequences of gene *23* in phages T4 and AR1. The presentation is the same as in panel A, except that the sequence comparison is between T4 and AR1. In the patch between aa 162 and aa 209, the AR1 DNA sequence differs by 27% from the T4 sequence, while in the remainder of the gene the divergence is less than 2%.

composed of six subunits) depends on the underlying tail tube structure, containing a similar number of gp19 subunits (8, 17), but additional gene products such as gp3 appear to be involved in length determination (37). Relatively little is known about the structure of gp18 protein and how its conformational changes result in tail contraction, except that it hydrolyzes ATP to ADP during contraction and that it has two nucleotide binding motifs (3, 31). It is still not clear, however, that this ATP hydrolysis is directly involved in the conformational changes involved in tail sheath contraction.

The complete sequence of gene *18* of the pseudoT-even phage RB49 was determined by taking advantage of two small randomly determined sequences of this phage genome that, based on the gene order of the T4 genome, should flank gene *18* (23). PCR primers based on these sequences were used to amplify a fragment that was directly sequenced and shown to contain a gene *18* analogue. The RB49 gp18 aa sequence was 63% identical to the gp18 sequence of T4, but the differences were unevenly distributed. The 75 aa at the N terminus were 68% conserved. The next 325 aa of the gp18 protein diverged more (58% identity), but the last 250 aa (67% identity) of this 690-aa protein contained some blocks of sequence that were nearly identical to T4 (Fig. 5). These conserved sequences at the beginning and the end of gene *18* provided us with PCR primers that could amplify related sequences from many of the T4-type phages, including both pseudo- and schizoT-evens. We sequenced gene *18* fragments from the pseudo T-even phage 42 of *B. cepacia* and the schizoT-even phage nt-1 of *V. natriegens* by primer walking (27). The alignment of these sequences is shown in Fig. 5. The overall impression from comparing the gene *18* sequence is quite similar to that obtained from the gene *23* sequence comparison. The sequence of the schizoT-even phage nt-1 has clearly diverged more from T4 than have the pseudoT-evens. Gene *18* appears to be a patchwork composed of variable sequences interspersed between small conserved motifs and several large blocks of near identity. It seems likely that among the large conserved domains are those that mediate the structural changes in gp18 that contract the tail. In this regard, it should be noted that the nucleotide binding motif at residue 530 is located in a highly conserved domain. Some of the non-conserved sequences could provide binding sites for accessory structures that can interact with the contractile tail.

Using similar methods, we also analyzed the sequence of

Gene *18*

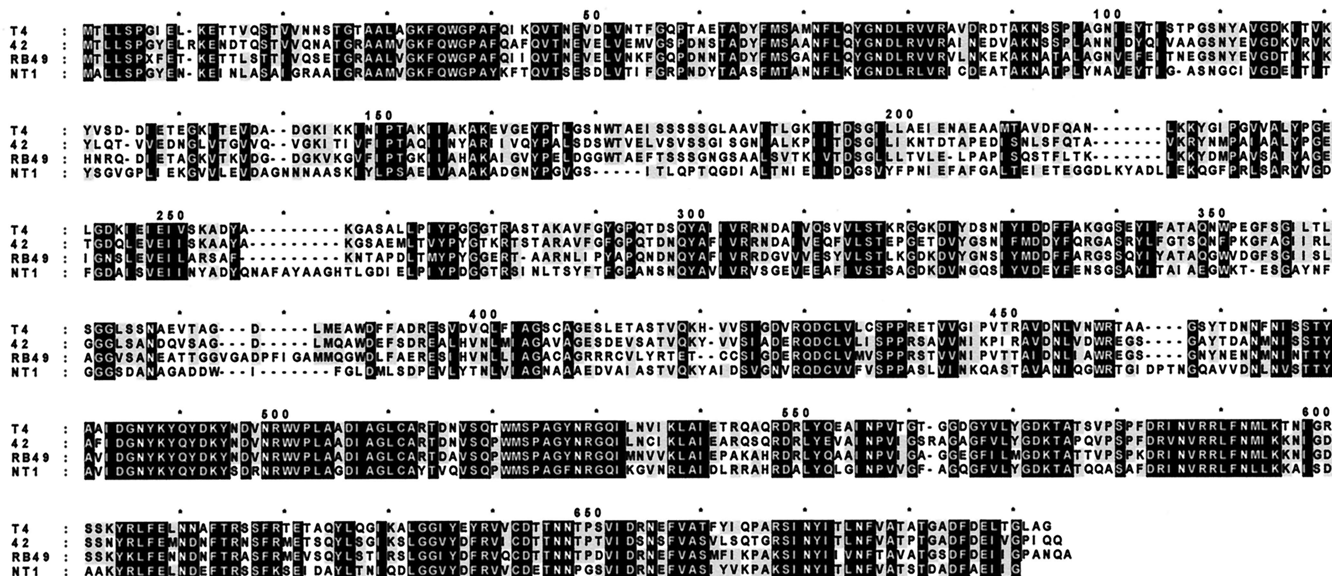

FIG. 5. Amino acid sequence comparison of the entire gene *18* of four T4-type phages (T4, RB49, 42, and nt-1). The homologous sequences of these phages were aligned with the T4 sequence as in Fig. 3. The conventions used in this presentation are the same as in Fig. 3, except that a black background indicates aa residues shared by all four of the phages while those that are present in only three phages are indicated by a gray background. The sequences shown with a white background were not conserved. The coordinates of the T4 gene *18* sequence are shown in the top line of the comparison.

gene *19* (31), which codes for the tail tube through which the phage DNA is injected into the cell. The strong constraints on the dimensions of this hollow tube might be expected to impose limits on gene *19* sequence divergence. Indeed, the alignment of gene *19* sequences from six T4-type phages shows a more uniform distribution of sequence conservation, but the overall level is similar to that in genes *18* and *23* (Fig. 6).

Phylogeny of the major virion genes of the T4-type phages. The phylogenetic trees (25) obtained for all of the viral structural genes are similar, and thus, only the data for gene *23* is shown in Fig. 7. The aa sequences of gene *23* differ by less than 10% within the T-even group, and in contrast, the eight pseudoT-evens differed from the T4 gene *23* sequence by 10 to 45%. Finally, five schizoT-evens differ in the gene *23* sequence from T4 by over 45%. Within each group the members were more closely related to each other than to any members of another group. This phylogenetic analysis classifies the T4-type phages as follows: T-evens (T4, T6, KC69 Tu1a, and RB69), pseudoT-evens (AR1, SV14, RB49, RB42, RB43, 42, 44RR, and 1) and schizoT-evens (nt-1, KVP20, KVP40, 65, and Aeh1).

Only the virion structural genes appear to be sufficiently

Gene *19*

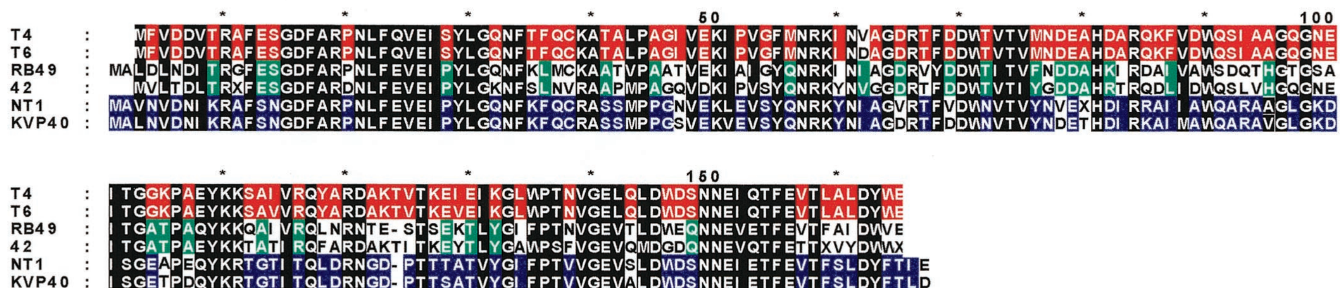

FIG. 6. Amino acid sequence alignments of gene *19* of T4-type phages. The protein sequences of the other five phages were aligned with the T4 sequence as in Fig. 3. The conventions used in this presentation are the same as in Fig. 3.

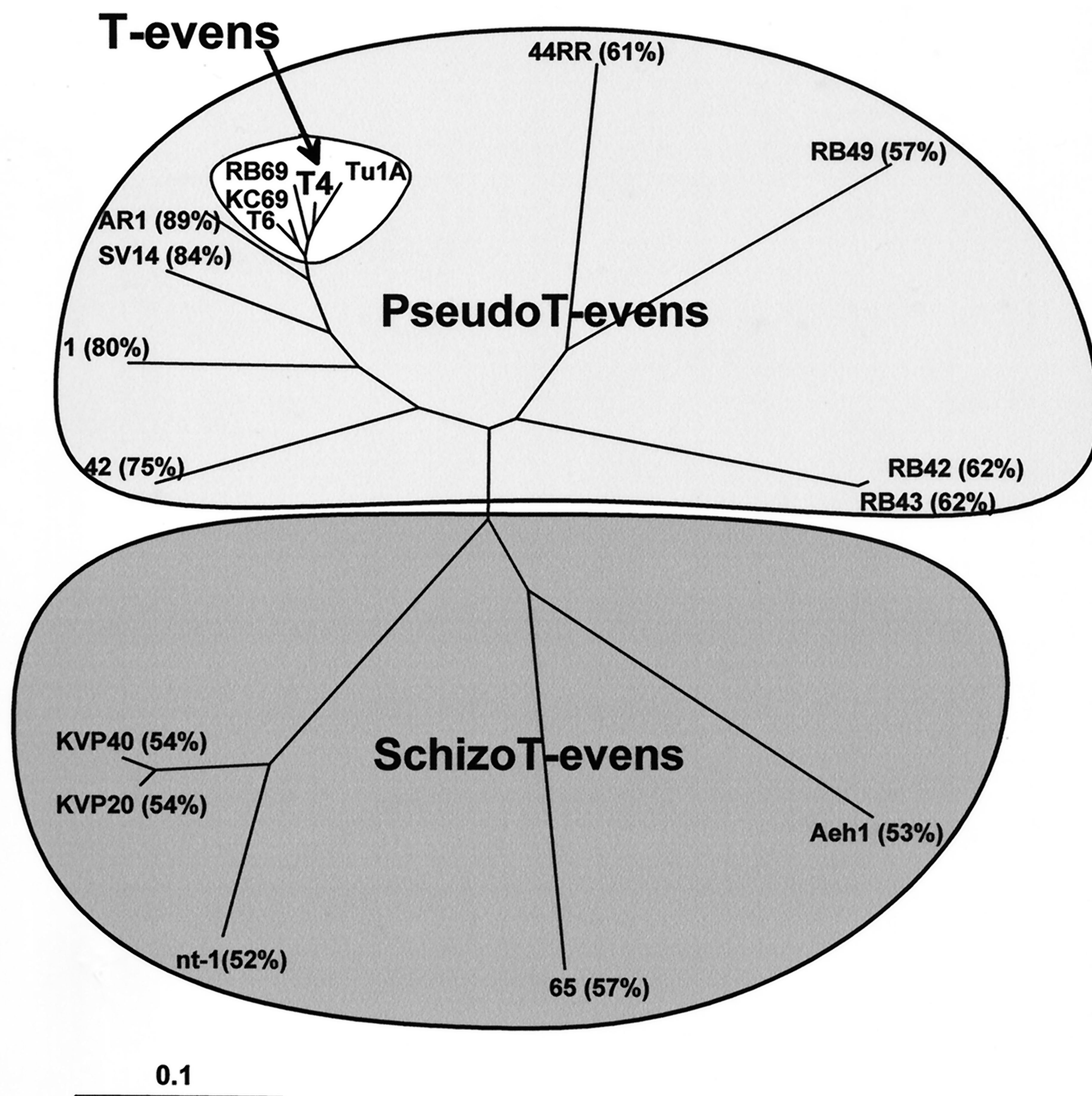

FIG. 7. Phylogenetic tree constructed by the TreeV32 program (25) using the sequence of gene *23* from 18 T4-type phages. The percentages of aa identities between the T4 gene *23* sequence and that of the other T4-type phages are shown beside each branch of the tree. The various subgroups are indicated by the shaded areas (white, T-evens; light gray, pseudoT-evens; and dark gray, schizoT-evens). In addition to the gene *23* sequence data, this classification reflects the sequences of other genes (genes *18* and *19*, etc.) and DNA hybridizations results (23).

conserved to construct a phylogeny by this method. Attempts to analyze nonstructural genes (e.g., the DNA polymerase gene *43*) were unsuccessful. The divergence of such genes in the various T4-type phages is too great to obtain PCR amplification with primers based on the T4 sequence (23). More laborious methods could be used to isolate and eventually sequence the gene *43* analogues of two of the pseudoT-even phages (RB49 and 44RR) (Desplats and Krisch, unpublished). However, primers based on the most conserved sequences in gene *43* of phages T4, RB49, and 44RR did not permit PCR amplification of gene *43* from any additional T4-type phages. Nevertheless, the gene *43* sequences of these pseudoT-evens (Desplats and Krisch, unpublished) are compatible with the phylogeny established for these phages on the basis of their virion structural genes. We have now sequenced more than 60% of the genome of the pseudoT-even phage RB49; this sequence analysis (Desplats and Krisch, unpublished) confirms the phylogenetic relationship of this phage to T4 predicted exclusively on the basis of the sequences of the major head and tail genes. Nevertheless, the RB49 genome does have one small sequence that it must have recently acquired from a genetic exchange with a T-even phage. Such events seem to be rare and do not alter a picture of a generally distant phylogenetic relationship between RB49 and T4. More limited

genomic sequencing of 44RR also confirms the position of this phage in the phylogenetic tree of the T4-type phages (Desplats and Krisch, unpublished).

DISCUSSION

Monod et al. (23) showed that primers based on the sequence of conserved segments of the capsid gene could be used to PCR amplify analogous DNA segments from distantly related T4-type phages. By sequencing such "PCR homologues," we have been able to compare diverse phages of the T4 type. All of the phage genomes analyzed had sequences homologous to the major structural genes of T4 (genes *18*, *19*, and *23*). Three subgroups of the T4-type phages can be distinguished: the T-evens, the pseudoT-evens, and the schizoT-evens. The vast majority of the known T4-type phages were isolated on enterobacteria, and most of them belong to the T-even subgroup. In contrast, the T4-type phages isolated on *Aeromonas*, *Burkholderia*, and *Vibrio* belong to either the pseudoT-even or schizoT-even subgroup and their genomes have diverged considerably.

The existence of the schizoT-even phage group demonstrates that the morphology of the virion head can vary within a family of phages. The similarity in morphology of the schizoT-evens to some of the aberrant phage produced by T4 "giant" point mutants (9) suggests that evolutionary transformations of head morphology may be accomplished without enormous difficulty. We expect that T4-type phages with smaller isometric heads or with a different length of the contractile tail can be isolated. It will be interesting to identify and analyze the sequences of the genes responsible for such morphological variations. The limits of the morphological variation of the T4-type phages have certainly not yet been circumscribed.

Our extensive analysis of a number of gene *23* sequences suggests a modular construction of its protein. Sequence comparisons of gene *18* (tail sheath protein) suggest that this gene has a similar modular organization. Gene mosaicism is most clearly evident in the T-even phage tail fibers (13, 28), where it facilitates swapping of the adhesin domains between phages (34). This may enable the phages to infect new hosts and thus would be an important evolutionary advantage. The patchwork structure of the tail fiber genes has been viewed as a consequence of the extreme pressures on their adhesin sequences to diversify (34). However, our sequence analysis of genes *23* and *18* in the T4-type phages now suggests that many more structural genes could have a similar, if less obvious, mosaic design.

As the fine structure of the T4-type virion becomes better understood, it will be interesting to determine the exact function of the various conserved motifs in the virion proteins. Many presumably have a role in determining the intrinsic structural features of the protein, but others could be engaged in protein-protein interactions between the different virion subunits (34). Another intriguing possibility is that some of these conserved sequences actually have a nonstructural role in promoting homologous recombination. As we have previously suggested, in a mosaic gene small conserved motifs could mediate the genetic swapping of the variable patches that they flank (33, 34). This could create increased diversity in specific domains of an otherwise conserved structural protein.

With the techniques reported here it is now easy to rapidly identify and characterize the T4-type phages in a feral phage population. For example, we have recently identified and characterized various T4-type phages isolated from pulmonary fluids of hospital patients with respiratory infections caused by *Pseudomonas aeruginosa* (M. Kutateladze, R. Adamia, F. Tétart, and H. M. Krisch, unpublished observations).

ACKNOWLEDGMENTS

We thank Rezo Adamia, Karin Carlson, Dick Epstein, Eric Miller, Jim Karam, and Lucien Caro for their diverse contributions. Our colleagues Marc Dreyfus, Jean-Pierre Claverys, and A. J. Carpousis tried to improve the manuscript. Germaine Burgener Krisch suggested the names of the T4-type phage subgroups, Yvette de Preval synthesized the oligonucleotides, and David Villa prepared the figures.

This research was financed by the CNRS and by grants from the Association pour la Recherche sur le Cancer (ARC), the Midi-Pyrénées Regional Council, the European Union INTAS-Georgia Initiative, and the NATO Collaborative Research Program.

REFERENCES

1. **Ackermann, H. W.** 1998. Tailed bacteriophages: the order Caudovirales. Adv. Virus Res. **51:**135–201.
2. **Ackermann, H.-W., and H. M. Krisch.** 1997. A catalogue of T4-type bacteriophages. Arch. Virol. **142:**2329–2345.
3. **Arisaka, F., L. Ishimoto, G. Kassavetis, T. Kumazaki, and S. Ishii.** 1988. Nucleotide sequence of the tail tube structural gene of bacteriophage T4. J. Virol. **62:**882–886.
4. **Arisaka, F., T. Nakako, H. Takahashi, and S. Ishii.** 1988. Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid sequence of its product. J. Virol. **62:**1186–1193.
5. **Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal.** 1989. High abundance of viruses found in aquatic environments. Nature **340:**467–468.
6. **Bouet, J. Y., J. Woszczyk, F. Repoila, V. François, J.-M. Louarn, and H. M. Krisch.** 1994. Direct PCR sequencing of the ndd gene of bacteriophage T4: identification of a product involved in bacterial nucleoid disruption. Gene **141:**9–16.
7. **Demerec, M., and U. Fano.** 1945. Bacteriophage-resistant mutants of *Escherichia coli.* Genetics **30:**119–136.
8. **DeRosier, D. J., and A. Klug.** 1972. Structure of the tubular variants of the head of bacteriophage T4 (polyheads). I. Arrangement of subunits in some classes of polyheads. J. Mol. Biol. **65:**469–488.
9. **Doermann, A. H., A. Pao, and P. Jackson.** 1987. Genetic control of capsid length in bacteriophage T4: clustering of *ptg* mutations in gene 23. J. Virol. **61:**2823–2827.
10. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. Nature **399:**541–548.
11. **Fuhrman, J. A. and C. A. Suttle.** 1993. Viruses in marine planktonic sytems. Oceanography **6:**51–63.
12. **Fuller, N. J., W. H. Wilson, I. R. Joint, and N. H. Mann.** 1998. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Appl. Environ. Microbiol. **64:**2051–2060.
13. **Haggård-Ljungquist, E., C. Halling, and R. Calendar.** 1992. DNA sequence of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. **174:**1462–1477.
14. **Haynes, J. A., and Eiserling F. A.** 1996. Modulation of bacteriophage T4 capsid size. Virology **221:**67–77.
15. **Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. Hatfull.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA **96:**2192–2197.
16. **Kim, J. S., and N. Davidson.** 1974. Electron microscope heteroduplex studies of sequence relations of T2, T4 and T6 bacteriophage DNAs. Virology **57:**93–111.
17. **King, J. and Mykolajewycz, N.** 1973. Bacteriophage T4 tail assembly: proteins of the sheath, core and baseplate. J. Mol. Biol. **75:**339–358.
18. **Loayza, D., A. J. Carpousis, and H. M. Krisch.** 1991. Gene 32 transcription and mRNA processing in T4-related bacteriophages. Mol. Microbiol. **5:**715–725.
19. **Luftig, R. B.** 1967. An accurate measurement of the catalase crystal period and its use as an internal marker for electron microscopy. J. Ultrastruct. Res. **20:**91–102.
20. **Matsuzaki, S., T. Inoue, M. Kuroda, S. Kimura, and S. Tanaka.** 1998. Cloning and sequencing of major capsid protein (mcp) gene of a vibriophage, KVP20, possibly related to T-even coliphages. Gene **222:**25–30.

21. **McNicol L. A., and L. E. Simon.** 1977. A mutation which bypasses the requirement for p24 in bacteriophage T4 capsid morphogenesis. J. Mol. Biol. **116:**261–283.
22. **McPheeters, D. S., G. Gosch, and L. Gold.** 1988. Nucleotide sequences of the bacteriophage T2 and T6 gene 32 mRNAs. Nucleic Acids Res. **16:**9341.
23. **Monod, C., F. Repoila, M. Kutateladze, F. Tétart, and H. M. Krisch.** 1997. The genome of the pseudoT-even bacteriophages, a diverse group that resembles the T-even phages. J. Mol. Biol. **267:**237–249.
24. **Noble, R. T., and J. A. Fuhrman.** 1998. Use of SYBR Green I for rapid epifluorescense counts of marine viruses and bacteria. Aquat. Microb. Ecol. **14:**113–118.
25. **Page, R. D. M.** 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:**357–358.
26. **Parker, M. L., A. C. Christensen, A. Boosman, J. Stockard, E. T. Young, and A. H. Doermann.** 1984. Nucleotide sequence of bacteriophage T4 gene 23 and the amino acid sequence of its product. J. Mol. Biol. **180:**399–416.
27. **Repoila, F., F. Tétart, J.-Y. Bouet, and H. M. Krisch.** 1994. Genomic polymorphism in the T-even bacteriophages. EMBO J. **13:**4181–4192.
28. **Sandmeier, H.** 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers. Mol. Microbiol. **12:**343–350.
29. **Selick, H. E., G. D. Stormo, R. L. Dyson, and B. M. Alberts.** 1993. Analysis of five presumptive protein-coding sequences clustered between the primosome genes, 41 and 61, of bacteriophages T4, T2, and T6. J. Virol. **67:**2305–2316.
30. **Streisinger, G., J. Emrich, and M. M. Stahl.** 1967. Chromosome structure in phage T4. III. Terminal redundancy and length determination. Proc. Natl. Acad. Sci. USA **57:**292–295.
31. **Takeda, S., F. Arisaka, S. Ishii, and Y. Kyogoku.** 1990. Structural studies of the contractile tail sheath protein of bacteriophage T4. 1. Conformational change of the tail sheath upon contraction as probed by differential chemical modification. Biochemistry **29:**5050–5056.
32. **Takeda, S., T. Sasaki, A. Ritani, M. M. Howe, and F. Arisaka.** 1998. Discovery of the tail tube gene of bacteriophage Mu and sequence analysis of the sheath and tube genes. Biochim. Biophys. Acta **1399:**88–92.
33. **Tétart, F., F. Repoila, C. Monod, and H. M. Krisch.** 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J. Mol. Biol. **258:**726–731.
34. **Tétart, F., C. Desplats, and H. M. Krisch.** 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J. Mol. Biol. **282:**543–556.
35. **Theimer, C. A., Y. Wang, D. W. Hoffman, H. M. Krisch, and D. P. Giedroc.** 1998. Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. J. Mol. Biol. **279:**545–564.
36. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
37. **Vianelli, A., G. R. Wang, M. Gingery, R. L. Duda, F. A. Eiserling, and E. B. Goldberg.** 2000. Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length. J. Bacteriol. **182:**680–688.
38. **Wais, A. C., and E. B. Goldberg.** 1969. Growth and transformation of phage T4 in E. coli B/4, Salmonella, Aerobacter, Proteus, and Serratia. Virology **39:**153–161.
39. **Wommack, K. E., R. T. Hill, M. Kessel, E. Russek-Cohen, and R. R. Colwell.** 1992. Distribution of viruses in the Chesapeake Bay. Appl. Environ. Microbiol. **58:**2965–2970.
40. **Wood, W. B., and M. P. Conley.** 1979. Attachment of tail fibers in bacteriophage T4 assembly: role of phage whiskers. J. Mol. Biol. **12:**715–729.
41. **Yanagida, M., Y. Suzuki, and T. Toda.** 1984. Molecular organization of the head of bacteriophage T-even: underlying design principles. Adv. Biophys. **17:**97–146.
42. **Yeh, L. S., T. Hsu, and J. D. Karam.** 1998. Divergence of a DNA replication gene cluster in the T4-related bacteriophage RB69. J. Bacteriol. **180:**2005–2013.