

# Composite Conserved Promoter–Terminator Motifs (PeSLs) that Mediate Modular Shuffling in the Diverse T4-Like Myoviruses

André M. Comeau<sup>1,7,\*</sup>, Christine Arbiol<sup>2,3,4,5</sup>, and Henry M. Krisch<sup>6,8</sup>

<sup>1</sup>Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, Canada

<sup>2</sup>CNRS, UMR7258, CRCM, Marseille, France

<sup>3</sup>Inserm, U1068, CRCM, Marseille, France

<sup>4</sup>Institut Paoli-Calmettes, Marseille, France

<sup>5</sup>Aix-Marseille Université, Marseille, France

<sup>6</sup>Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique-UMR5100, Université Paul Sabatier, Toulouse, France

<sup>7</sup>Present address: Department of Pharmacology, Dalhousie University, Halifax, Canada.

<sup>8</sup>Retirement address: Avenue du Général Guisan 38, Sierre, Switzerland.

\*Corresponding author: E-mail: andre.comeau@dal.ca.

Accepted: June 9, 2014

**Data deposition:** The complete genome sequences of PST and nt-1 have been deposited at GenBank under the accessions KF208315 and HQ317393, respectively.

## Abstract

The diverse T4-like phages (*Tqatrovirinae*) infect a wide array of gram-negative bacterial hosts. The genome architecture of these phages is generally well conserved, most of the phylogenetically variable genes being grouped together in a series hyperplastic regions (HPRs) that are interspersed among large blocks of conserved core genes. Recent evidence from a pair of closely related T4-like phages has suggested that small, composite terminator/promoter sequences (promoter<sub>early</sub> stem loop [PeSLs]) were implicated in mediating the high levels of genetic plasticity by indels occurring within the HPRs. Here, we present the genome sequence analysis of two T4-like phages, PST (168 kb, 272 open reading frames [ORFs]) and nt-1 (248 kb, 405 ORFs). These two phages were chosen for comparative sequence analysis because, although they are closely related to phages that have been previously sequenced (T4 and KVP40, respectively), they have different host ranges. In each case, one member of the pair infects a bacterial strain that is a human pathogen, whereas the other phage's host is a nonpathogen. Despite belonging to phylogenetically distant branches of the T4-likes, these pairs of phage have diverged from each other in part by a mechanism apparently involving PeSL-mediated recombination. This analysis confirms a role of PeSL sequences in the generation of genomic diversity by serving as a point of genetic exchange between otherwise unrelated sequences within the HPRs. Finally, the palette of divergent genes swapped by PeSL-mediated homologous recombination is discussed in the context of the PeSLs' potentially important role in facilitating phage adaption to new hosts and environments.

**Key words:** T4-like phages, *Yersinia*, *Vibrio*, genome evolution, PeSLs, genome plasticity.

## Introduction

The diverse T4-like phages (formally called the *Tqatrovirinae* subfamily by Lavigne et al. [2009]) are widespread in the environment (e.g., Jia et al. 2007; Comeau and Krisch 2008; López-Bueno et al. 2009; Butina et al. 2010). These morphologically complex phages have large genomes, varying in size from 160 to 240 kb, that can be adapted to infect a wide

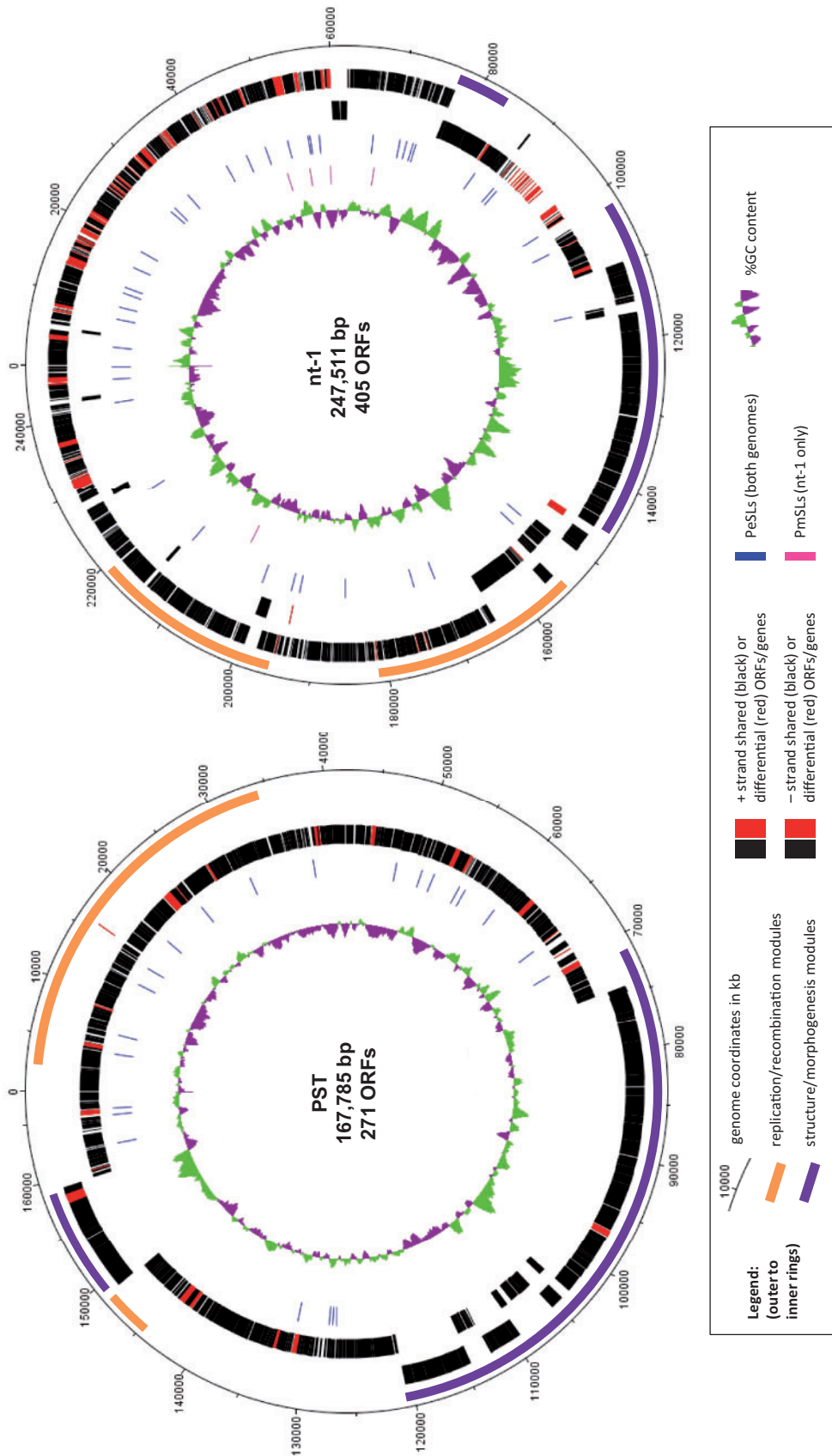
range of gram-negative bacteria (Ackermann and Krisch 1997). In the past decade, there has been considerable progress in sequencing the genomes of various representatives of the T4-like phages (reviewed in Krisch and Comeau 2008 and Petrov et al. 2010). A major focus of this effort was to understand the evolutionary diversity of this large and diverse yet coherent group of viruses. The picture that has emerged from

these studies is one of a common genome structure, with a well-conserved set of structural and replication genes grouped together in large contiguous blocks (modules). The evolution of this “core T4 genome” has been primarily vertical. However, interspersed between the conserved T4 core sequences are a series of hyperplastic regions (HPRs), which vary greatly in gene number and content. The HPRs are subject to much more horizontal transfer than the core genome (Filée et al. 2006; Comeau et al. 2007; Petrov et al. 2010). At least some, and perhaps most, of the genes in HPRs encode ancillary and adaptive functions that allow the core genome to propagate in a widely varied set of hosts. The analysis of a very closely related pair of T4 phages, the coliphages phi1 and RB49, revealed the presence of series of compound promoter–terminator elements (promoter<sub>early</sub> stem loops = PeSLs). Either homologous or site-specific recombination involving these well-conserved genetic regulatory sequences offered a simple, plausible explanation for some of the genetic hyperplasticity within the HPRs (Arbiol et al. 2010). As a consequence of these results, and our recent analysis of the mechanism that shuffles the major host-range determinants within the tail-fiber adhesin genes (Trojet et al. 2011), we have extended this analysis to additional closely related pairs of T4 phages that infect bacterial hosts that are phylogenetically distant. The phage PST, whose virion morphology is indistinguishable from T4, was originally isolated on the pathogen *Yersinia pseudotuberculosis* (Knapp and Zwillenberg 1964). However, a preliminary analysis of its genome (Krisch HM, unpublished data) indicated that it was phylogenetically very closely related to T4, a coliphage. Similarly, phage nt-1 was isolated on the halotolerant host *Vibrio natriegens* (Zachary 1974), but its genome sequence was clearly very closely related (Tétart et al. 2001) to the phage KVP40 that was isolated on the pathogen *V. parahaemolyticus* and also infects multiple *Vibrio* spp. (Matsuzaki et al. 1992; Miller et al. 2003a). Our notion was that an analysis of pairs of phages with differing host ranges could provide useful insights into the mechanism(s) and gene(s) responsible for their differences in host range. The additional pairs of phages analyzed were chosen to reflect the diversity with the T4-like group; one pair belonging to the T-even subgroup (the closest relatives of T4), whereas the other pair belongs to the much more evolutionarily distant Schizo-T-even subgroup that have a more elongated virion head and consequently a significantly larger genome size. This analysis suggests that, in spite of their phylogenetic differences, both pairs of T4-type phages have undergone similar genomic changes mediated in part by PeSLs sequences similar to those discovered in the Pseudo-T-even phages RB49 and phi1 (Arbiol et al. 2010). Hence, this study both confirms and extends our previous proposal about the genetic mechanism responsible for some of the early events in T4 phage genome differentiation.

High-throughput sequencing allowed us to quickly obtain single contiguous sequences for both the PST and nt-1

genomes. These contigs could be closed to circles (fig. 1), consistent with the circularly permuted, terminally redundant structure of the T4 group’s genomes (Casjens and Gilcrease 2009). The PST genome is only approximately 1 kb smaller than T4 and both sequences share many conserved features (table 1)—similar GC levels, tRNAs (supplementary table S1, Supplementary Material online), and gene/ORF contents. In contrast, the nt-1 and KVP40 genomes have modest differences in both their tRNA and gene/ORF contents. Both phage pairs have considerable overall nucleotide sequence conservation, with PST/T4 having 84% of their genomes with more than or equal to 95% similarity (fig. 2, pink shading), whereas the nt-1/KVP40 pair diverges more, 82% of their genomes have an identity level of  $\geq 66\%$  (essentially equivalent to ignoring the wobble position). In spite of their sequence divergence, the gene order and content of the paired genomes are remarkably well conserved (supplementary fig. S1, Supplementary Material online). Much of the divergence between the paired genomes (fig. 1) is located in the HPRs where the variable genes/ORFs are generally grouped together. For example, the HPR located in the first approximately 60 kb of the nt-1/KVP40 pair (figs. 1 and 2) contains many of the differences in their gene/ORF content. Less dramatically, the PST/T4 pair is differentiated by only 55 genes and ORFs. Although the majority of these encode unknown functions, a few have been shown to be involved in host-range determination. For example, in the tail-fiber adhesin locus (the major determinant of host specificity), PST has a typical gp38-type adhesin sequence (supplementary table S2, Supplementary Material online), whereas in T4, this adhesin function is encoded by an unrelated C-terminal domain of the adjacent gp37 long tail fiber (Trojet et al. 2011). There are also differences in this phage pair in their DNA modification systems (glycosylases/methylases), which could reflect differences in the specificities of *Escherichia* and *Yersinia* restriction-modification systems. Finally, there are five differences between PST/T4 in the internal protein (IP) sequences. The IPs are encapsidated in the virion head and injected upon infection along with the viral DNA. In at least some cases, these small proteins encode functions that negate host defensive mechanisms (Comeau and Krisch 2005; Bair et al. 2007). The nt-1/KVP40 pair differs in approximately 125 genes/ORFs, the majority of which encode small proteins of unknown function with no matches in the databases (ORFans). Only four differential genes have currently identifiable potential roles in host-range determination—a transcription factor and dCMP deaminase in KVP40, and two tRNA-modifying enzymes in nt-1 (supplementary table S3, Supplementary Material online).

As mentioned before, the examination of another closely related pair of T4-like genomes (RB49 and phi1) allowed us to identify the PeSL elements that were apparently closely associated with genome plasticity (Arbiol et al. 2010). The analysis of the genomes of the PST and nt-1 pairs revealed similar sets of PeSLs that differ slightly in sequence between them and



**Fig. 1.**—Genome maps and features of PST and nt-1.

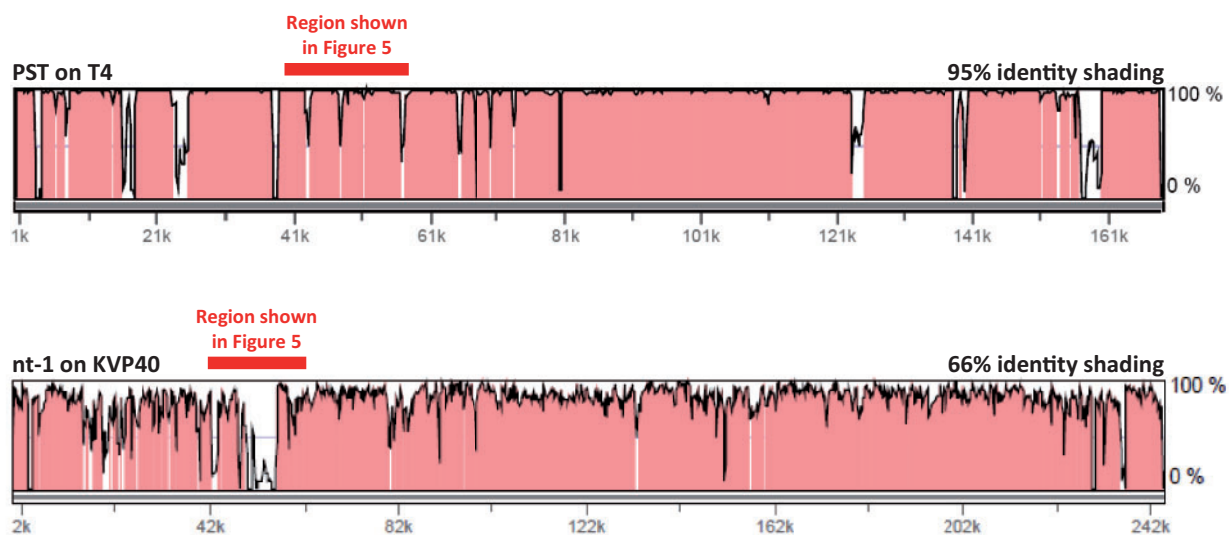
also from the RB49/phi1 pair (figs. 3 and 4). The stem-loop structures (SLSs) of the new PeSLs described here differ from the RB49/phi1  $G_N$ -loop- $C_N$  sequence: the nt-1 pair frequently have a consensus SLS sequence of  $A_{3-4}G_3$ -loop- $C_3T_{3-4}$ , whereas the PST pair often has a  $AG_NA$ -loop- $TC_NT$  sequence, but with greater sequence variability in their SLS sequences than the other pairs. The PeSLs motifs of nt-1 seem to be more

highly conserved than the others, suggesting that they may more efficiently promote recombinational shuffling than either the PST- or RB49-type PeSLs. The consensus promoter sequences in these PeSLs also differ from the near-perfect  $\sigma^{70}$  sequences (TTGACA... $N_{17}$ ...TATAAT) observed in the RB49/phi1 pair: TTTACW... $N_{17}$ ...TAYWAT for the PST/T4 pair and TTGYVH... $N_{17}$ ...TAWWAT for the nt-1/KVP40 pair. Interestingly, nt-1 also has some PeSL-like sequences based on the T4-like middle-mode (Mot) promoter consensus (therefore called promoter<sub>middle</sub> stem loop [PmSLs]), with a TGCTT Mot-box regulator sequence followed by, at the appropriate distance, a -10-like box with the consensus sequence TATTAT (fig. 4; Miller et al. 2003b). As in the original RB49/phi1 pair, the new PeSLs are preferentially located in the HPRs (fig. 1) and frequently associated with nearby ORF(an)s that are different between the pairs of phages. When located within the conserved core of the genome, they are often associated with particularly plastic nonessential adaptive genes. Figure 5 shows some segments of the genomes that are particularly variable between the pairs of phages. It is clear that PeSLs are frequently found in close proximity to chromosomal insertions, deletions, or sequence exchanges. For example, in the nt-1 genome between coordinates approximately 51–56 kb (fig. 5), there is an uninterrupted series of six ORFs with PeSL elements located in the intergenic spaces between them. Within this entire small PeSL-rich genomic segment, there have been numerous genetic rearrangements when compared with the KVP40 sequence: 5 insertions, 4 replacements, 15 deletions, and an ORF displacement. Figure 6 shows, in detail, several typical examples of ORF exchanges and deletions that were apparently mediated by PeSLs.

**Table 1**

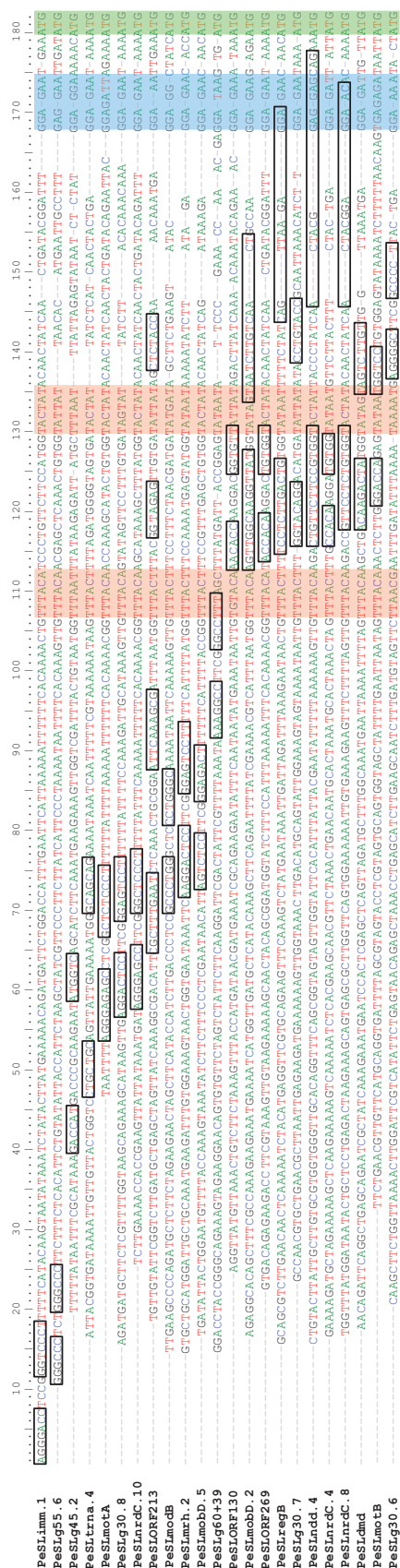
Summary of Genome Characteristics of PST versus T4 and nt-1 versus KVP40

| Characteristic                     | PST      | T4       | nt-1     | KVP40    |
|------------------------------------|----------|----------|----------|----------|
| Genome size (nt)                   | 167,785  | 168,903  | 247,511  | 244,834  |
| GC content (%)                     | 35.3     | 35.3     | 41.3     | 42.6     |
| # tRNAs                            | 9        | 8        | 29       | 30       |
| All ORFs/genes                     |          |          |          |          |
| Total number                       | 271      | 278      | 405      | 381      |
| Size range (aa)                    | 34–1,289 | 26–1,289 | 29–1,246 | 36–1,256 |
| Mean size (aa)                     | 193      | 197      | 187      | 194      |
| Median size (aa)                   | 131      | 135      | 126      | 133      |
| Differential ORFs/genes            |          |          |          |          |
| Total number                       | 27       | 28       | 70       | 54       |
| Potentially involved in host range | 5        | 7        | 2        | 2        |
| Other known virus functions        | —        | —        | 2        | —        |
| Mobile elements                    | —        | 7        | 1        | —        |
| Homologs of cellular functions     | —        | —        | 5        | 7        |
| Unidentified virus functions       | 22       | 14       | 8        | 2        |
| ORFans                             | —        | —        | 52       | 43       |



**FIG. 2.**—Whole-genome nucleotide level comparisons of PST to T4 (top) and nt-1 to KVP40 (bottom).





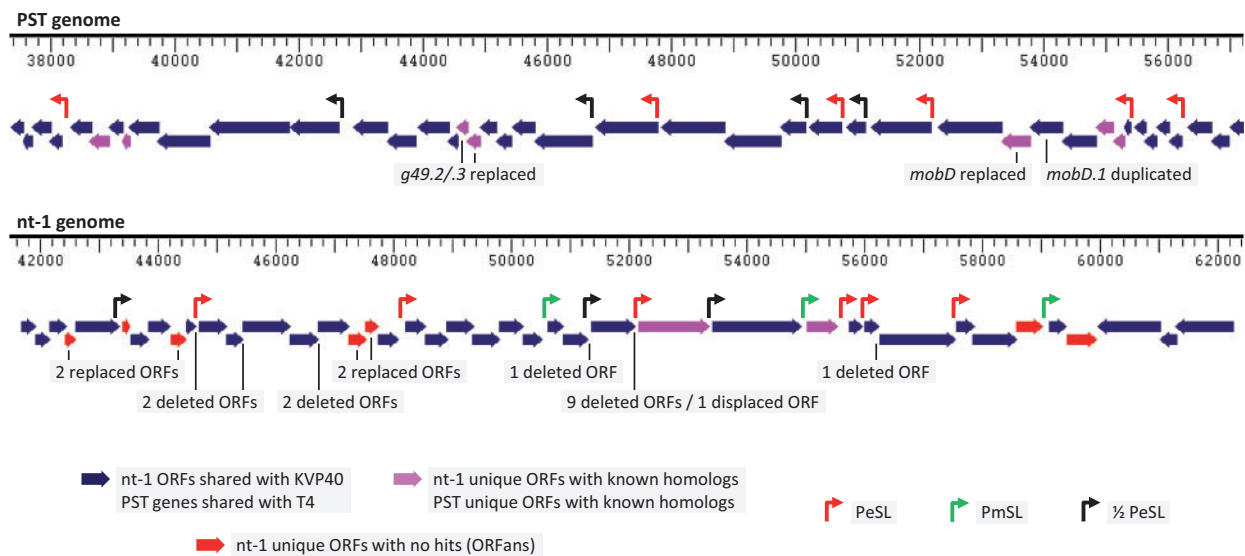
**Fig. 3.**—PeSL elements detected in PST. The PeSLs are composed of multiple motifs, starting with a generally AT-rich region located upstream of the –35 and –10 boxes (rose) of the  $\sigma^{70}$ -like promoters. The PeSLs have at their 3'-extremity a Shine–Dalgarno sequence (blue) and the start codon (green) of the downstream genes. The SLs (stems marked by black boxes) are located just upstream of the element's endogenous promoter. The PeSL sequences were aligned primarily on the basis of their constituent promoter and Shine–Dalgarno sequences.

## Conclusion

Despite being from two divergent groups within the T4-like viruses, both PST and nt-1 reveal a similar pattern of sequence variation when compared with their closest known relatives. PeSL elements, first identified in a yet another branch of the T4-like phages, appear to be responsible for a nontrivial portion of the initial genetic divergence that occurs within this diverse phage subfamily. Significantly, PeSL elements are located in close proximity to many of the divergent loci we have examined. For some, these PeSLs are close enough to each other (e.g., fig. 6A) to generate exchanges by a PeSL mini-circle formation mechanism, which has been previously demonstrated to delete the DNA interposed between two tandem PeSLs (Arbiol et al. 2010). For other divergent loci where there remains only a single PeSL on one side of the ORF(s), either the distal flanking PeSL sequence has been lost by a deletion event or by a rare homologous recombination event occurring between small randomly homologous sequences of non-PeSL origin (Albertini et al. 1982). Such non-PeSL-mediated mechanisms offer a plausible explanation for the *g49.2/3* exchange in PST (fig. 5). The frequency of PeSLs located in close proximity to each other is notably higher in the most variable genome segments. Such groupings of PeSLs are less frequent in genome regions containing the conserved virion structural genes, perhaps because in such regions the transcriptional units are generally longer, and there is strong coupling of expression of the different structural components that must be assembled together in the mature virion in precisely defined ratios. These observations, coupled with the previous evidence based on comparative genomics of exchange events occurring in or near the PeSL motifs (Arbiol et al. 2010), suggests that such elements play at least a contributory role to the creation of genetic diversity within the T4-like phages. The more “traditional” general homologous recombination and horizontal transfers of DNA adjacent to mobile elements (such as homing endonucleases common to these phages; Kadyrov et al. 1997; Brok-Volchanskaya et al. 2008) are clearly also involved, the latter (like the PeSL mechanism) is apparently more important in the less conserved, nonstructural regions of the genome. In such regions, we suggest that PeSL-mediated exchanges could make a nontrivial, but so far unrecognized, contribution to genome plasticity.

It should be noted that some of the sequence divergences between the PST/T4 and nt-1/KVP40 pairs could explain the differences in the host ranges of these closely related phages. As mentioned in our original observations on PeSLs (Arbiol et al. 2010), 20 years ago similar observations had been made within the T4 IP locus (Repoila et al. 1994), without detailing a mechanism, and it was hypothesized that a specific shuffling mechanism within the IP palette between related phages could allow major





**Fig. 5.**—Presence of multiple fully functional or degraded (“1/2”) PeSLs/PmSLs near indels in HPRs of the PST (top) and nt-1 (bottom) genomes. The degraded PeSLs either have 1) only intact  $-35$  regions remaining (for  $\sim 43$  K/51 K in nt-1) or 2) intact  $-35$  regions and stem loops (no  $-10$  s; for  $\sim 47$  K/51 K in PST); or 3) multiple weak/degraded copies of putative  $-35/-10$  regions inappropriately spaced (most probably caused by recombinational slippage) and degraded stem loops (for  $\sim 53$  K in nt-1) or with one or more still-intact stem loops (for  $\sim 43$  K/50 K in PST).

modifications in host range. It is clear that, due to the apparently wide distribution of PeSLs, we can expand this idea to yet larger segments of the T4 genomes (i.e., the HPRs but usually not structural regions) and probably to all of the T4-like phages. Further experimentation on the targets and the mechanism of PeSL-mediated genome divergence is clearly merited. Such an effort will also be essential for us to understand the numerous and largely unknown gene functions that allow these phages to so easily adapt to an ever-changing suite of environments and hosts.

## Materials and Methods

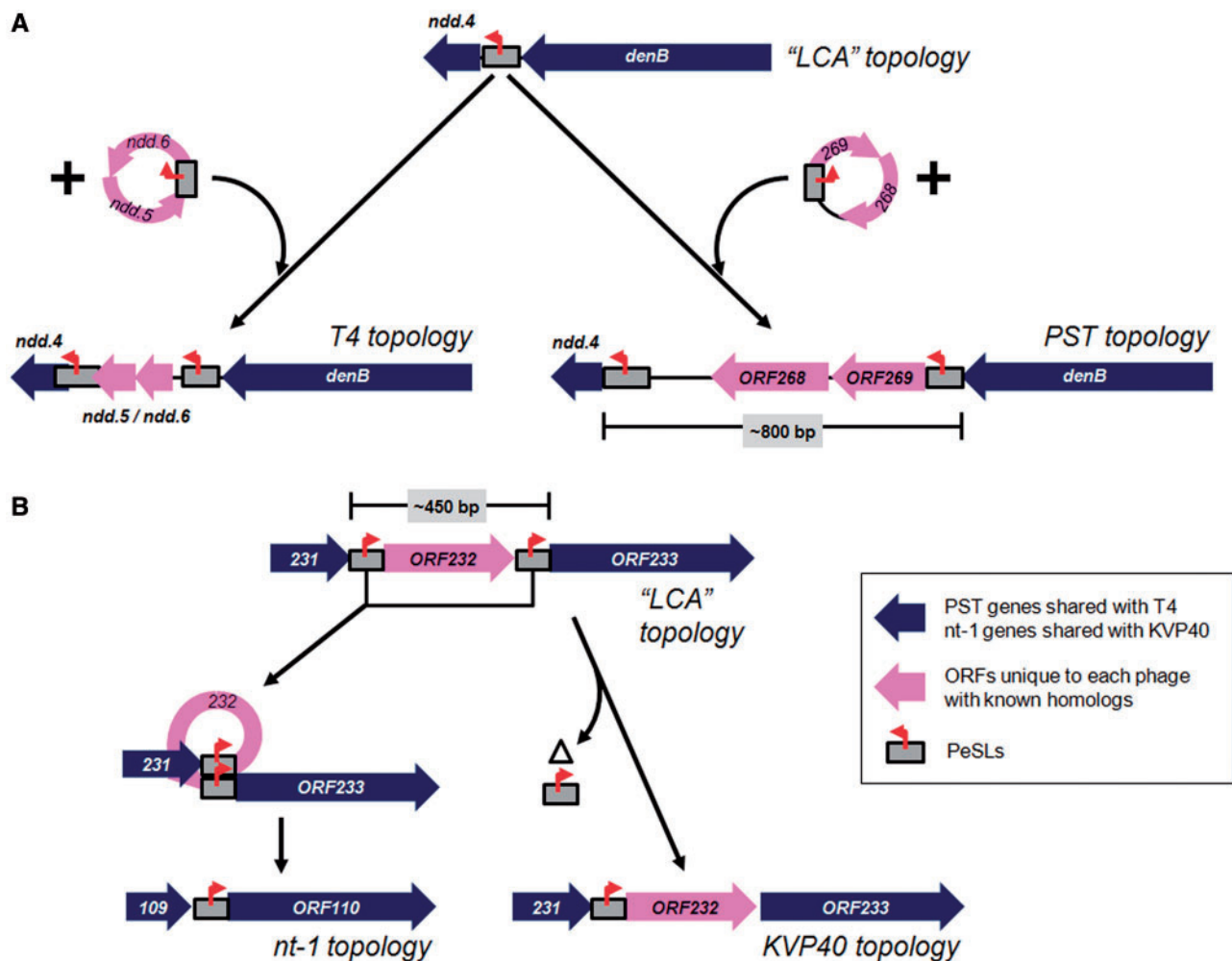
### Bacteriophage DNA Preparation and Pyrosequencing

The *Yersinia* phage PST and *Vibrio* phage nt-1 strains came from HMK’s Toulouse collection of myoviruses. PST was originally obtained from Dr Grimont of the Pasteur Institute in Paris, whereas nt-1 was obtained from the Félix d’Hérelle Reference Center for Bacterial Viruses (HER150). *Yersinia pseudotuberculosis* NCTC10275 and *V. natriegens* HER1138 were the host bacteria used to prepare stocks of the phages using standard techniques as described by Carlson and Miller (1994). All strains were grown at  $37^{\circ}\text{C}$  in Luria-Bertani medium. The DNA was extracted from high-titer stocks as detailed in Ackermann et al. (2011). The resulting pure DNAs were used for bar-coded library construction and 454 pyrosequencing that were performed according to the manufacturer’s instructions. PST was sequenced at the IBIS/Université Laval Plate-forme d’Analyses Génomiques

(Québec), and nt-1 was sequenced at the Broad Institute ([www.broadinstitute.org/annotation/viral/Phage](http://www.broadinstitute.org/annotation/viral/Phage), last accessed June 22, 2014) as part of a Gordon and Betty Moore Foundation (GBMF) Marine Microbiology Initiative ([www.moore.org/programs/science/marine-microbiology-initiative](http://www.moore.org/programs/science/marine-microbiology-initiative), last accessed June 22, 2014).

### Genome Assembly and Annotation

Raw reads were assembled using the de novo GS Assembler (Roche), resulting in one final contig each for PST and nt-1 with 54- and 30-fold coverage, respectively. Analyses of the genomes were done with the following programs: 1) GLIMMER ([www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer\\_3.cgi](http://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi), last accessed June 22, 2014;  $>100$  nt; bacterial genetic code) and GeneMark ([exon.gatech.edu/GeneMark](http://exon.gatech.edu/GeneMark), last accessed June 22, 2014; heuristic approach for prokaryotes and viruses;  $>90$  nt) for ORF determinations; 2) tRNA search using tRNAscan-SE ([lowelab.ucsc.edu/tRNAscan-SE](http://lowelab.ucsc.edu/tRNAscan-SE), last accessed June 22, 2014); 3) Java Word Frequencies ([athena.bioc.uvic.ca/virology-ca-tools/jfreq/](http://athena.bioc.uvic.ca/virology-ca-tools/jfreq/), last accessed June 22, 2014) and Dot Plot Alignments (MIPS Gepard; [www.helmholtz-muenchen.de/icb/gepard](http://www.helmholtz-muenchen.de/icb/gepard), last accessed June 22, 2014) for the exploration of DNA and protein “words”/patterns; 4) LAGAN ([lagan.stanford.edu/lagan\\_web/index.shtml](http://lagan.stanford.edu/lagan_web/index.shtml), last accessed June 22, 2014) for the visualization/calculation of whole genome-to-genome nucleic-acid-level identities; 5) the BLAST tools at NCBI ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov), last accessed June 22, 2014) for the characterization of genes/proteins and untranslated regions of the DNA; and 6) DNAPlotter for generating the circular genome visualizations



**FIG. 6.**—Examples of recombinational events causing the topology differences between the PST/T4 and KVP40/nt-1 phage pairs in which PeSLs are implicated. Deduced sites of recombinations are between the phages investigated here and a “last common ancestor” (“LCA”; may still be extant but not yet isolated and characterized). (A) The exchange of genes *ndd.4/1.5* in T4 (left) for *ORF268/269* in PST (right) could have come about through PeSL mini-circle integration (as outlined in Arbiol et al. 2010) of different DNA cassettes from different donors into the same (or similar) LCA. (B) The deletion of *ORF232* in nt-1 could have come about through PeSL mini-circle excision (Arbiol et al. 2010) from the LCA (left) and the resulting topology in KVP40 (right) achieved through loss/deletion of the downstream (potentially redundant) PeSL. Note here that *ORF109/110* in nt-1 are homologs of *ORF231/233* in KVP40 and that the reverse recombination is also possible (PeSL mini-circle integration of *ORF232* into nt-1 to generate KVP40/LCA).

([www.sanger.ac.uk/resources/software/dnaplotter](http://www.sanger.ac.uk/resources/software/dnaplotter), last accessed June 22, 2014). Specifically for the PeSL elements: Local BLASTn with previous PeSLs and self-on-self dot-plots easily initially detected the repeated elements; then all intergenic spaces of large-enough size were extracted and aligned to confirm the promoter elements and detect the SLSs.

#### Nucleotide Sequence Accession Numbers

The complete genome sequences of PST and nt-1 were deposited in GenBank under accession numbers KF208315 and HQ317393, respectively. According to the GBMF

guidelines, phage nt-1 was also deposited in CAMERA ([camera.calit2.net](http://camera.calit2.net), last accessed June 22, 2014).

#### Supplementary Material

Supplementary tables S1–S3 and figure S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

#### Acknowledgments

The authors thank their colleagues at the IBIS/Université Laval Plate-forme d’Analyses Génomiques (Québec, QC) for



sequencing and assembly advice for phage PST. This work was supported in part by the Gordon and Betty Moore Foundation through grant GBMF1799 to the Broad Institute, by the INSB of the French CNRS, and by the KriBurg Foundation to H.M.K. Sample G2350 (phage nt-1) was sequenced and assembled at the Broad Institute.

## Literature Cited

- Ackermann H-W, Krisch H. 1997. A catalogue of T4-type bacteriophages. *Arch Virol.* 142:2329–2345.
- Ackermann H-W, Krisch HM, Comeau AM. 2011. Morphology and genome sequence of phage  $\phi$ 1402. *Bacteriophage* 1:138–142.
- Albertini AM, Hofer M, Calos MP, Miller JH. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* 29:319–328.
- Arbiol C, Comeau AM, Kutateladze M, Adamia R, Krisch HM. 2010. Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. *Genome Biol Evol.* 2:140–152.
- Bair CL, Rifat D, Black LW. 2007. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI\*. *J Mol Biol.* 366:779–789.
- Brok-Volchanskaya VS, et al. 2008. Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage. *Nucleic Acids Res.* 36:2094–2105.
- Butina TV, Belykh OI, Maksimenko SY, Belikov SI. 2010. Phylogenetic diversity of T4-like bacteriophages in Lake Baikal, East Siberia. *FEMS Microbiol Lett.* 309:122–129.
- Carlson K, Miller ES. 1994. Experiments in T4 genetics. In: Karam JD, editor. *Molecular biology of bacteriophage T4*. Washington (DC): ASM Press. p. 421–483.
- Casjens SR, Gilcrease EB. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. In: Clokie MRJ, Kropinski AM, editors. *Bacteriophages: methods and protocols*. New York: Humana Press. p. 91–111.
- Comeau AM, Bertrand C, Letarov A, Tétart F, Krisch HM. 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* 362:384–396.
- Comeau AM, Krisch HM. 2005. War is peace—dispatches from the bacterial and phage killing fields. *Curr Opin Microbiol.* 8:488–494.
- Comeau AM, Krisch HM. 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol Biol Evol.* 25:1321–1332.
- Filée J, Bapteste E, Susko E, Krisch HM. 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. *Mol Biol Evol.* 23:1688–1696.
- Jia Z, Ishihara R, Nakajima Y, Asakawa S, Kimura M. 2007. Molecular characterization of T4-type bacteriophages in a rice field. *Environ Microbiol.* 9:1091–1096.
- Kadyrov FA, Shlyapnikov MG, Kryukov VM. 1997. A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. *FEBS Lett.* 415:75–80.
- Knapp W, Zwillenberg LO. 1964. Morphological differences between *Pasteurella* bacteriophages. *Arch Ges Virusforsch.* 14:563–566.
- Krisch HM, Comeau AM. 2008. The immense journey of bacteriophage T4—from d'Hérelle to Delbrück and then to Darwin and beyond. *Res Microbiol.* 159:314–324.
- Lavigne R, et al. 2009. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiol.* 9:224.
- López-Bueno A, et al. 2009. High diversity of the viral community from an Antarctic lake. *Science* 326:858–861.
- Matsuzaki S, Tanaka S, Koga T, Kawata T. 1992. A broad-host-range vibriophage, KVP40, isolated from sea water. *Microbiol Immunol.* 36:93–97.
- Miller ES, et al. 2003a. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol.* 185:5220–5233.
- Miller ES, et al. 2003b. Bacteriophage T4 genome. *Microbiol Mol Biol Rev.* 67:86–156.
- Petrov VM, Ratnayaka S, Nolan JM, Miller ES, Karam JD. 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology* 403:292–302.
- Repoila F, Tétart F, Bouet JY, Krisch HM. 1994. Genomic polymorphism in the T-even bacteriophages. *EMBO J.* 13:4181–4192.
- Tétart F, et al. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *J Bacteriol.* 183:358–366.
- Trojet SN, Caumont-Sarcos A, Perrody E, Comeau AM, Krisch HM. 2011. The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. *Genome Biol Evol.* 3:674–686.
- Zachary A. 1974. Isolation of bacteriophages of the marine bacterium *Beneckeia natriegens* from coastal salt marshes. *Appl Environ Microbiol.* 31:415–422.

Associate editor: Purificación López-García