# Evolutionary Genomics of an Ancient Prophage of the Order *Sphingomonadales*

Vandana Viswanathan[1,2], Anushree Narjala[1], Aravind Ravichandran[1], Suvratha Jayaprasad[1], and Shivakumara Siddaramappa[1,]*

[1]Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Electronic City, Bengaluru, Karnataka, India

[2]Manipal University, Manipal, Karnataka, India

*Corresponding author: E-mail: shivakumar@ibab.ac.in.

## Abstract

The order *Sphingomonadales*, containing the families *Erythrobacteraceae* and *Sphingomonadaceae*, is a relatively less well-studied phylogenetic branch within the class *Alphaproteobacteria*. Prophage elements are present in most bacterial genomes and are important determinants of adaptive evolution. An "intact" prophage was predicted within the genome of *Sphingomonas hengshuiensis* strain WHSC-8 and was designated Prophage I[WHSC-8]. Loci homologous to the region containing the first 22 open reading frames (ORFs) of Prophage I[WHSC-8] were discovered among the genomes of numerous *Sphingomonadales*. In 17 genomes, the homologous loci were co-located with an ORF encoding a putative superoxide dismutase. Several other lines of molecular evidence implied that these homologous loci represent an ancient temperate bacteriophage integration, and this horizontal transfer event pre-dated niche-based speciation within the order *Sphingomonadales*. The "stabilization" of prophages in the genomes of their hosts is an indicator of "fitness" conferred by these elements and natural selection. Among the various ORFs predicted within the conserved prophages, an ORF encoding a putative proline-rich outer membrane protein A was consistently present among the genomes of many *Sphingomonadales*. Furthermore, the conserved prophages in six *Sphingomonas* sp. contained an ORF encoding a putative spermidine synthase. It is possible that one or more of these ORFs bestow selective fitness, and thus the prophages continue to be vertically transferred within the host strains. Although conserved prophages have been identified previously among closely related genera and species, this is the first systematic and detailed description of orthologous prophages at the level of an order that contains two diverse families and many pigmented species.

**Key words:** *Sphingomonadales*, *Sphingomonadaceae*, *Erythrobacteraceae*, HGT, prophage.

## Introduction

The "purple bacteria" were originally classified into four groups based on the phylogenetic analyses of their 16S rRNA sequences (Woese 1987). These related groups of Gram-negative bacteria were assigned to a new class *Proteobacteria* because of their outstanding "diversity of shape and physiology" (Stackebrandt et al. 1988). A novel genus *Sphingomonas* was proposed around the same time to accommodate sphingoglycolipid-containing bacteria that were previously assigned to the genera *Flavobacterium* and *Pseudomonas* (Yabuuchi et al. 1990). Following the realization that the genus *Sphingomonas* is a member of the α-4 subclass of the *Proteobacteria* (Takeuchi et al. 1994), a novel family

*Sphingomonadaceae* was proposed using a polyphasic taxonomic approach (Kosako et al. 2000). The class *Proteobacteria* was later elevated to the level of a phylum (Phylum XIV) that contained five classes, including the novel class *Alphaproteobacteria*, within the hierarchical system of the second edition of the Bergey's manual of systematic bacteriology (Garrity et al. 2005a, 2005b). The novel order *Sphingomonadales*, containing *Sphingomonadaceae* as the sole family, was also proposed within this hierarchical system (Yabuuchi and Kosako 2005). Based on analyses of 16S rDNA sequences, the genera *Erythrobacter*, *Porphyrobacter*, and *Erythromicrobium* were proposed to be included in a novel family *Erythrobacteraceae* within the order *Sphingomonadales*. (Lee et al. 2005). Due to their clinical

relevance and applications in biotechnology and bioremediation, the genomes of several members of the order *Sphingomonadales* have been sequenced (Glaeser and Kämpfer 2014; Tonon et al. 2014). Comparative genomic studies have revealed that the coding potential of different species are highly variable, indicating their divergent evolution (Aylward et al. 2013; Zheng et al. 2016). These studies have also documented the occurrence of putative genes encoding oxygenases and glycoside hydrolases, which could facilitate survival in a variety of niches and the degradation of pollutants (Aylward et al. 2013; Verma et al. 2014).

The term "prophage" was originally proposed in the context of lysogeny and was defined as "the form in which lysogenic bacteria perpetuate the power to produce phage" (Lwoff 1953). It was later recounted that "in spite of its French origin, the Greek word was rapidly and unanimously adopted" and that "it seemed that the world eagerly awaited its coming" (Lwoff 1966). This definition could have been very appealing at a time when it was believed that "lysogeny is an attribute of every bacterial cell" (Joklik 1999). The fact that phage-related elements have been identified in almost all bacteria whose genomes have been sequenced bolsters this belief (Canchaya et al. 2003; Casjens 2003). Although several prophages have been identified among the genomes of *Sphingomonadales* (Aylward et al. 2013; Glaeser and Kämpfer 2014; Tonon et al. 2014; Zheng et al. 2014; García-Romero et al. 2016), comprehensive analyses of these horizontally transferred "selfish genetic elements" are conspicuously missing. Preliminary work at the authors' institution identified a prophage element that appeared to be partially conserved among several *Sphingomonas* spp., and was informally referred to as Prosphingophage (unpublished results). The objective of this study was to further characterize this element and its evolutionary significance.

## Tools and Methods

### Prophage Identification

Genome sequences (either complete or draft) of different species of *Sphingomonadales* were obtained from GenBank and annotated using the RAST server (http://rast. nmpdr.org/; last accessed January 11, 2017). A local BLAST was installed and a compatible genome database was created by following the instructions in the Sequencher® manual (http://www.genecodes.com/sites/default/files/documents/Tutorials/Local-BLAST.pdf; last accessed January 11, 2017). Prophages within these genomes were identified using PHASTER (Arndt et al. 2016). The presumptive prophages delineated using PHASTER were further subjected to intuitive curation. Manual curation relied on the criteria set forth by Casjens (2003) and involved using the "Feature Table" option within the SEED Viewer tool of the RAST server to locate (1) "cornerstone features", including open

reading frames (ORFs) encoding putative phage proteins and (2) a stretch of ORFs encoding hypothetical proteins, within the presumptive prophages. The locations of the curated prophages within the respective chromosomes were also assessed using the "Feature Table" option of the SEED Viewer. The GC% of the presumptive prophages (and that of their respective hosts) was calculated using the option available within the BioEdit tool. The order and orientation of the ORFs within the prophages (and those flanking the prophages) were checked using the sequence-based comparison tool available within the SEED Viewer. Among a pair of genomes, ORFs within the prophages were deemed orthologus if the putative proteins encoded by them had at least 30% identity and <20% difference in length during BLASTP analyses. This threshold was essential to identify (and exclude) ORFs that may be located outside the defined boundaries of the prophages. In addition, the synteny of ORFs (i.e., the occurrence of ORFs in the same order and orientation within a locus) was an important criterion to assign orthology. This parameter was essential to recognize (and include) ORFs that may show sequence divergence, or may be located on different contigs in draft genomes.

### Sequence Alignment and Analyses

Clustal omega (http://www.ebi.ac.uk/Tools/msa/clustalo/; last accessed January 11, 2017) was used to obtain an initial alignment of 16S rDNA and prophage protein sequences. This alignment was used as a guide to identify mismatches, and to manually trim the sequences at the ends when they were of different lengths. Trimming at the ends was performed to obtain similar number of positions (nucleotides or amino acids) for each aligned sequence and to ensure phylogenetic accuracy. ClustalW (http://embnet.vital-it.ch/software/ClustalW.html; last accessed January 11, 2017) was used to generate a "pir" output/alignment of the prophage protein sequences. The "pir" output from ClustalW was further aligned using the BOXSHADE server (http://www.ch.embnet.org/software/BOX_form.html; last accessed January 11, 2017) to highlight the conserved residues among prophage protein sequences from different species. Signal peptides were predicted using the PrediSi tool (http://www.predisi.de/; last accessed January 11, 2017). Secondary structures of proteins were predicted using the Chou and Fasman secondary structure prediction server (http://www.biogem.org/tool/chou-fasman/; last accessed January 11, 2017).

### Phylogenetic Analysis Using CVTree3

The web server CVTree3, which is an alignment- and parameter-free method that relies on the oligopeptide content (*K*-tuple length) of conserved proteins to deduce evolutionary relatedness (Zuo and Hao 2015), was used for phylogenetic analysis. The web server constructs phylogenetic trees using the Neighbor-Joining method based on a dissimilarity matrix.

Trees constructed by the web server are not subjected to statistical re-sampling (bootstrap or jackknife analyses) because the underlying method emphasizes on the "objective" correctness of phylogeny with respect to taxonomy. The method also does not provide a "scale" for the length of the branches because it emphasizes on tree topology and not evolutionary time (Zuo and Hao 2015).

Proteomes (excluding plasmid-encoded proteins) of *Sphingomonadales* were downloaded from UniProt (http://www.uniprot.org/proteomes/; last accessed January 11, 2017). Prophage protein sequences of *Sphingomonadales* were obtained from the respective genome annotations within the RAST server. The protein sequences were saved as multifasta files with the extension .faa. The multifasta files for each strain/prophage were uploaded on to the CVTree3 web server (http://tlife.fudan.edu.cn/archaea/cvtree/cvtree3/; last accessed January 11, 2017) and analyzed by selecting all available $K$-tuple length options (from 3 to 9). Because the best $K$-values for bacteria and viruses were shown to be 5–6 and 4–5, respectively (Zuo and Hao 2015), the proteome tree was visualized at $K = 6$ and the prophage protein tree was visualized at $K = 4$.

### Phylogenetic Analysis Using MEGA 6.0

Concatenated sequences of four predicted proteins were generated manually by joining them in the same order as their ORFs occurred within the orthologous prophages. Pairwise alignments of DNA/protein sequences were performed using ClustalW with default parameters. The pairwise distance matrix derived from these alignments was used to construct a guide tree by the Neighbor-Joining method. Subsequent progressive alignment was based on the guide tree. Phylogeny was reconstructed using the maximum likelihood method (with 1,000 bootstrap replicates) and the Tamura–Nei (for DNA sequences) or Jones–Taylor–Thornton (for protein sequences) substitution model in MEGA 6.0.

## Results and Discussion

### Several Genera of *Sphingomonadales* Have Stably Maintained a Specific Prophage

The genomes of at least hundred strains of various species of *Sphingomonas* have been sequenced. Although most of these genome sequences are incomplete, they could be used in analyses aimed at understanding the genomic diversity of the genus. *Sphingomonas hengshuiensis* strain WHSC-8 is an yellow pigmented bacterium isolated from "soil of Hengshui Lake Wetland Reserve in Hebei province, northern China" (Wei et al. 2015). The genome of strain WHSC-8 is among the few within *Sphingomonas* that have been completely sequenced and consists of a chromosome (5,191,536 bp; 66.7% GC) and a plasmid (36,853 bp; 62.6% GC). PHASTER predicted a single "intact" prophage

within the chromosome of strain WHSC-8 and assigned a "score" of 150, which indicated that it was unlikely to be a false positive result. Manual curation confirmed this prediction and the locus (3,774,962–3,832,206 bp) was designated Prophage I[WHSC-8] (57,245 bp; 68.5% GC). A total of 74 ORFs were annotated/identified within Prophage I[WHSC-8] and most of them encoded hypothetical proteins. Although 12 ORFs of Prophage I[WHSC-8] were predicted to encode putative phage-related proteins based on homology, ORFs encoding a phage integrase were absent among them. This was not surprising since, according to Casjens (2003), "although most temperate phages carry an integrase gene, its presence is neither necessary nor sufficient to prove the existence of a prophage".

Prophage I[WHSC-8] could be further delineated into three distinct regions. Region I (18,383 bp; 70.46% GC) contained 22 ORFs, including "cornerstone features" such as ORFs encoding putative phage capsid and portal proteins. The closest orthologs of 18 of these ORFs were found in *Sphingomonas* sp. Root 241 (see supplementary table S1, Supplementary Material online), which was isolated from the root of *Arabidopsis thaliana* (Bai et al. 2015). Notably, the orthologs of many of these ORFs were also identified among the genomes of various genera of the families *Sphingomonadaceae* and *Erythrobacteraceae*. Table 1 shows the features of 29 complete and three draft genomes of *Sphingomonadales* that contained a locus homologous to region I of Prophage I[WHSC-8]. From this table, it was apparent that the loci were of different sizes, and were the smallest in *Novosphingobium pentaromativorans* US6-1 and *Porphyrobacter neustonensis* DSM 9434 (each containing only three ORFs). It was also apparent that the GC% of the loci varied extensively, with the lowest (59.78%) in *Altererythrobacter epoxidivorans* CGMCC 1.7731, and the highest (73.20%) in *Sphingomonas taxi* ATCC 55669. Although there was no discernible relationship between the sizes of the loci and their GC%, in most cases the GC% was higher than that of the host genome. Furthermore, many of these homologous loci delineated by comparative genomics were also identified by PHASTER as prophages with a "score" similar to that of Prophage I[WHSC-8]. The locus in *Erythrobacter litoralis* HTCC2594 was previously identified as a prophage (Oh et al. 2009; Tonon et al. 2014).

Region II of Prophage I[WHSC-8] was the longest (22,178 bp; 67.07% GC) and contained 31 ORFs (see supplementary table S2, Supplementary Material online). Region III was the shortest (16,688 bp; 68.36% GC) and contained 21 ORFs (see supplementary table S3, Supplementary Material online). Whereas none of the ORFs from region III had orthologs among other *Sphingomonas* spp., at least five ORFs from region II had orthologs among members of the same genus. Furthermore, almost all ORFs from region III had orthologs in *Sphingobium ummariense* RL-3, which was a hexachlorocyclohexane-degrading bacterium isolated from soil in northern

**Table 1**
Bacteria Containing Loci Homologous to Region I of Prophage I^WHSC-8

| Species (Strain) | Chromosome Size (Genome Status) | GC% | GenBank/RefSeq Accession Number | Locus Homologous to Prophage I^WHSC-8 (Length; GC%) | Ortholog of ORF 1 (ompA) of Prophage I^WHSC-8 (Length; GC%) | Ortholog of ORF 3 of Prophage I^WHSC-8 |
|---|---|---|---|---|---|---|
| Sphingomonas sp. (Root 241) | 4,212,322 bp (draft)[a] | 66.00 | NZ_LMIV00000000 | 501,740–517,840 bp in contig 2 (16,101 bp; 70.19) | ASE13_13090; 377 aa | ASE13_13080; 715 aa |
| Sphingomonas sp. (MM-1) | 4,054,833 bp (complete)[b] | 67.20 | CP004036 | 3,619,742–3,637,287 bp (17,546 bp; 71.88) | G432_17145; 365 aa | G432_17135; 724 aa |
| Sphingomonas sp. (NIC1) | 3,408,545 bp (complete)[b] | 67.40 | CP015521 | 1,144,673–1,159,030 bp (14,358 bp; 71.33) | A7E77_05530; 373 aa | A7E77_05540; 710 aa |
| Sphingomonas taxi (ATCC 55669) | 3,859,099 bp (complete)[b] | 68.00 | CP009571 | 3,442,178–3,456,451 bp (14,274 bp; 73.20) | MC45_15835; 372 aa | MC45_15825; 721 aa |
| Sphingomonas melonis (TY) | 4,100,783 bp (draft)[b] | 67.10 | NZ_AQZT00000000 | 33,415–47,676 bp in contig 7 (14,262 bp; 71.42) | AVM11_15095; 373 aa | AVM11_15105; 710 aa |
| Sphingobium yanoikuyae (ATCC 51230) | 5,500,358 bp (draft)[b] | 64.40 | NZ_AGZU00000000 | 77,631–99,963 bp in contig 6 (22,333 bp; 68.24)[d] | HMPREF9718_00770; 358 aa | HMPREF9718_00772; 730 aa |
| Sphingobium japonicum (UT26S) | 3,514,822 bp (chromosome 1, complete)[c] | 64.80 | NC_014006 | 1,211,843–1,228,052 bp (16,210 bp; 69.36)[d] | SJA_C1-12400; 360 aa | SJA_C1-12380; 732 aa |
| Sphingobium chlorophenolicum (L-1) | 3,080,818 bp (chromosome 1, complete)[b] | 63.90 | CP002798 | 210,307–226,109 bp (15,803 bp; 68.03)[d] | Sphch_0217; 358 aa | Sphch_0215; 733 aa |
| Sphingobium sp. (TKS) | 4,249,857 bp (chromosome 1, complete)[b] | 63.40 | CP005083 | 1,471,870–1,487,480 bp (15,611 bp; 67.16)[d] | K426_07570; 358 aa | K426_07560; 732 aa |
| Sphingobium sp. (MI1205) | 3,351,250 bp (chromosome 1, complete)[b] | 62.30 | CP005188 | 2,337,681–2,353,409 bp (15,729 bp; 67.50)[d] | K663_11270; 362 aa | K663_11280; 731 aa |
| Sphingobium sp. (EP60837) | 2,669,660 bp (chromosome 1, complete)[b] | 62.40 | CP015986 | 2,323,770–2,339,688 bp (15,919 bp; 66.75)[d] | EP837_02291; 362 aa | EP837_02285; 776 aa |
| Sphingobium sp. (YBL2) | 4,766,421 bp (complete)[b] | 64.80 | CP010954 | 4,545,984–4,561,755 bp (15,772 bp; 70.00)[d] | TZ253_21010; 358 aa | TZ253_21020; 731 aa |
| Sphingobium sp. (SYK-6) | 4,199,332 bp (complete)[b] | 65.60 | NC_015976 | 3,576,052–3,591,409 bp (15,358 bp; 70.32) | SLG_33080; 354 aa | SLG_33100; 741 aa |
| Sphingobium baderi (DE-13) | 4,107,398 bp (complete)[b] | 62.40 | CP013264 | 3,523,708–3,542,006 bp (18,299 bp; 67.23)[d] | ATN00_17160; 349 aa | ATN00_17170; 735 aa |
| Novosphingobium sp. (PP1Y) | 3,911,486 bp (complete)[b] | 63.70 | FR856862 | 3,730,993–3,745,669 bp (14,677 bp; 67.55) | PP1Y_AT35659; 351 aa | PP1Y_AT35629; 731 aa |
| Novosphingobium pentaromativorans (US6-1) | 3,979,506 bp (complete)[b] | 63.50 | CP009291 | 2,157,279–2,159,382 bp (2,104 bp; 63.70) | JI59_10340; 351 aa | None |
| Novosphingobium aromaticivorans (DSM 12444) | 3,561,584 bp (complete)[b] | 65.20 | CP000248 | 3,331,159–3,345,964 bp (14,806 bp; 67.56)[d] | Saro_3134; 357 aa | Saro_3132; 734 aa |
| Sphingorhabdus sp. (M41) | 3,339,521 bp (complete)[b] | 56.70 | CP014545 | 2,894,894–2,910,876 bp (15,983 bp; 60.66) | AZE99_13750; 369 aa | AZE99_13760; 741 aa |

(continued)

**Table 1** Continued

| Species (Strain) | Chromosome Size (Genome Status) | GC% | GenBank/RegSeq Accession Number | Locus Homologous to Prophage I^WHSC-8 (Length; GC%) | Ortholog of ORF 1 (ompA) of Prophage I^WHSC-8 | Ortholog of ORF 3 of Prophage I^WHSC-8 |
|---|---|---|---|---|---|---|
| Sphingopyxis alaskensis (RB2256) | 3,345,170 bp (complete)[b] | 65.50 | CP000356 | 2,096,682–2,114,435 bp (17,754 bp; 69.66) | Sala_1988; 373 aa | Sala_1990; 742 aa |
| Sphingopyxis granuli (TFA) | 4,679,853 bp (complete)[c] | 66.20 | CP012199 | 924,316–941,647 bp (17,332 bp; 72.08) | SGRAN_0882; 372 aa | SGRAN_0880; 725 aa |
| Sphingopyxis macrogoltabida (EY-1) | 4,757,879 bp (complete)[b] | 64.90 | CP012700 | 201,720–219,343 bp (17,624 bp; 69.23) | AN936_01075; 369 aa | None |
| Sphingopyxis fribergensis (Kp5.2) | 4,993,584 bp (complete)[b] | 63.90 | CP009122 | 594,603–611,999 bp (17,397 bp; 67.56) | SKP52_03010; 372 aa | SKP52_03000; 725 aa |
| Citromicrobium sp. (JL477) | 3,258,499 bp (complete)[b] | 65.00 | CP011344 | 439,768–455,104 bp (15,337 bp; 69.50)[d] | WG74_02210; 371 aa | WG74_02200; 722 aa |
| Altererythrobacter marensis (KCTC 22370) | 2,885,033 bp (complete)[b] | 64.70 | CP011805 | 2,182,757–2,198,671 bp (15,915 bp; 70.54)[d] | AM2010_2068; 352 aa | AM2010_2066; 727 aa |
| Altererythrobacter dongtanensis (KCTC 22672) | 3,009,495 bp (complete)[b] | 65.80 | CP016591 | 280,293–296,157 bp (15,865 bp; 68.36)[d] | A6F68_00277; 384 aa | A6F68_00275; 730 aa |
| Altererythrobacter namhicola (JCM 16345) | 2,591,679 bp (complete)[b] | 65.00 | CP016545 | 266,462–280,590 bp (14,129 bp; 68.65)[d] | A6F65_01235; 318 aa | A6F65_00277; 729 aa |
| Altererythrobacter epoxidivorans (CGMCC 1.7731) | 2,786,256 bp (complete)[b] | 61.50 | CP012669 | 2,691,601–2,698,096 bp (6,496 bp; 59.78)[d] | AMC99_02724; 371 aa | AMC99_02725; 729 aa |
| Altererythrobacter atlanticus (26DY36) | 3,386,291 bp (complete)[c] | 61.90 | CP011452 | 2,535,141–2,542,708 bp (7,568 bp; 64.35)[d] | WYH_00683; 373 aa | None |
| Porphyrobacter neustonensis (DSM 9434) | 3,090,363 bp (complete)[b] | 65.30 | CP016033 | 789,090–791,267 bp (2,178 bp; 63.50) | A9D12_03720; 380 aa | None |
| Erythrobacter litoralis (HTCC2594) | 3,052,398 bp (complete)[b] | 63.10 | CP000157 | 2,811,166–2,826,784 bp (15,619 bp; 66.91)[d] | ELL_13950; 367 aa | ELL_13960; 725 aa |
| Erythrobacter atlanticus (s21-N3) | 3,012,400 bp (complete)[b] | 58.20 | CP011310 | 569,704–582,991 bp (13,288 bp; 59.85)[d] | CP97_02815; 357 aa | CP97_02860; 727 aa |
| Croceicoccus naphthovorans (PQ-2) | 3,543,806 bp (complete)[b] | 62.60 | CP011770 | 723,961–738,998 bp (15,038 bp; 62.73) | AB433_03655; 354 aa | AB433_03665; 733 aa |

[a]Genomes that contain an ortholog of TS85_17065 within the prophage.
[b]Genomes that contain an ortholog of TS85_17065 outside the prophage.
[c]Genomes that lack an ortholog of TS85_17065.
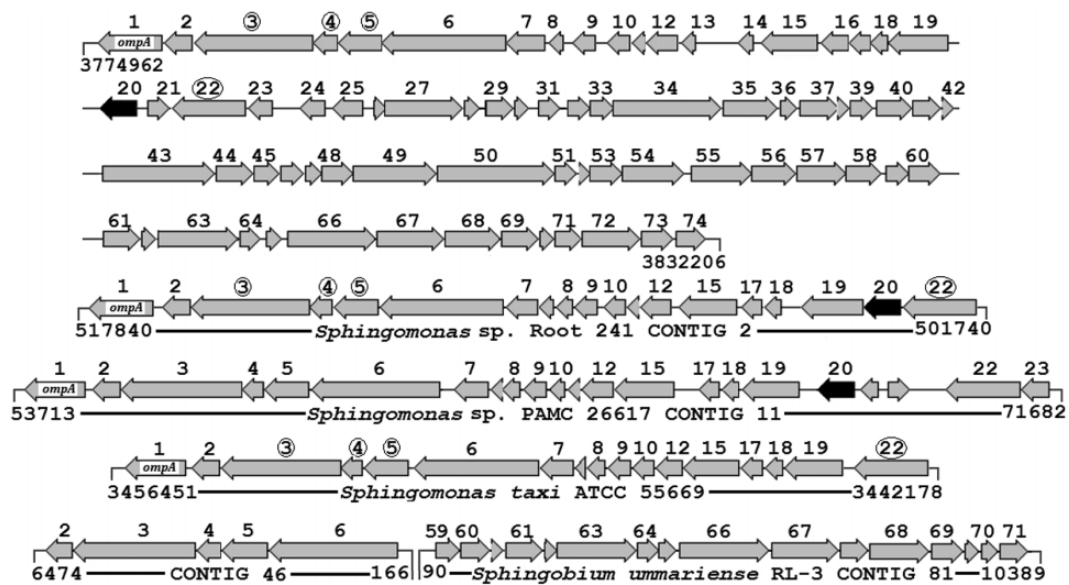[d]Prophage element was co-located with an ORF encoding a putative superoxide dismutase.

FIG. 1.—Comparison of prophages from five bacterial strains. A map of Prophage I^WHSC-8 is shown on the top (genome coordinates 3,774,962–3,832,206 bp) and contains 74 ORFs (represented by arrows). Region I consists of the first 22 ORFS, region II consists of ORFs 23-53, and region III consists of ORFs 54–74. The first ORF (ompA; TS85_16970) encodes a putative outer membrane protein A. The twentieth ORF (speE; TS85_17065) encodes a putative spermidine synthase (black arrow). The order and orientation of the ORFs in the orthologous prophages of Sphingomonas sp. Root 241 (genome coordinates 501,740–517,840 bp in contig 2), Sphingomonas sp. PAMC 26617 (genome coordinates 53,713–71,682 bp in contig 11), Sphingomonas taxi ATCC 55669 (genome coordinates 3,442,178–3,456,451 bp), and Sphingobium ummariense RL-3 (in contigs 46 and 81) are shown below the map of Prophage I^WHSC-8. Four ORFs (3, 4, 5, and 22) whose protein sequences were used for phylogenetic analyses in figures 3B and 4B are encircled.

India (Singh and Lal 2009; Verma et al. 2014). In contrast, only four ORFs from region II had orthologs in Sb. ummariense RL-3. Although orthologs of a few ORFs from region I were also found in Sb. ummariense RL-3, they were scattered among different contigs of the draft genome (fig. 1). Not unexpectedly, homologs of 12, 5, and 10 ORFs from regions I, II, and III, respectively, were found in the genomes of phylogenetically distant bacteria outside the Sphingomonadales. Whereas many of these ORFs encoded hypothetical proteins, some encoded putative phage-related proteins (see supplementary tables S1–S3, Supplementary Material online).

Prophages that are similar among different bacteria are either remnants of a lysogenic event in an ancestral genome/host, or recent independent integrations of the same bacteriophage (Bobay et al. 2013, 2014). Using four stringent criteria, Bobay et al. (2014) identified conserved prophage elements among fifteen strains of Salmonella enterica that had orthologs in two strains of Escherichia coli. These P2-like orthologous prophages of the family Myoviridae, although lacking an integrase, were found to be flanked by "homologous core genes". However, among distantly related bacteria, orthologous prophages are unlikely to be flanked by the same genes due to chromosomal rearrangements and gene shuffling. In such cases, the conserved order and orientation of ORFs could be a better indicator of common ancestry. Among 27 complete and three

draft genomes of Sphingomonadales, the order and orientation of the ORFs within the prophages were found to be conserved (data not shown). Furthermore, among 17 genomes, the prophage element was co-located with an ORF encoding a putative superoxide dismutase (table 1).

Because the orthologous prophages identified by Bobay et al. (2014) were from S. enterica and E. coli, they displayed a "high gene repertoire relatedness". Analyses of the putative proteins encoded by the 22 ORFs of region I of Prophage I^WHSC-8 using BLASTP indicated that the closest orthologs (in Sphingomonas sp. Root 241; see supplementary table S1, Supplementary Material online) had an identity range of 49–88% (average 71.5%). The closest orthologs of these proteins outside Sphingomonas had an identity range of 41–80% (average 60%). The closest orthologs of the putative proteins encoded by the ORFs of region III of Prophage I^WHSC-8 (in Sb. ummariense RL-3; see supplementary table S3, Supplementary Material online) had an identity range of 30–72% (average 51%). Taken together, these results indicate that the analyses of orthologous prophages at the level of genera and/or families may require the "threshold" of relatedness to be set to a value that is not >50%. While comparing the orthologous prophages of S. enterica and E. coli, Bobay et al. (2014) observed that such elements generally "display a gene diversity that does not greatly exceed the gene content of the ancestral prophage". Pairwise

comparisons revealed that the gene repertoire (42 ORFs) of the conserved prophages identified among the genomes of *Sphingomonadales* was much smaller than that of the gene content of Prophage I^WHSC-8^ (74 ORFs).

These analyses indicate that the conserved elements represent an ancient temperate bacteriophage integration, and this horizontal transfer event pre-dates natural selection-based speciation within the order *Sphingomonadales*. The possibility that these elements represent independent lysogenic conversions of different bacteria by the same broad host range bacteriophage is remote because they occur only among members of *Sphingomonadales*. This possibility is further ruled out by the fact that the host strains have a broad distribution in space and time. The extensive variation in the length of the orthologous prophages among different genomes suggests that they have been subjected to differential gene losses, and that some of them have "stabilized" in the respective genomes. The observation that the orthologous prophages differ in their GC% suggests that they have resided within their hosts for longer periods of time, and that most of them are evolving *in sync* with their respective host chromosomes. Because the GC% of the prophage was higher than that of the host chromosome in most cases, it is possible that the constituent genes are under similar selection, which may be required for maintaining their functions. Further analyses are required to determine whether these genes are under purifying selection, as demonstrated for the genes in the orthologous prophages of *S. enterica* and *E. coli* (Bobay et al. 2014).

The "stabilization" of prophages in the genomes of their hosts is an indicator of "fitness" conferred by the residual genes of these elements and adaptive evolution. Because the orthologous prophage elements among members of *Sphingomonadales* appear to be selectively maintained, it is possible that they bestow some fitness. The fact that such an element was absent in the complete genomes of *Sphingomonas wittichii* (strain RW1), *Sphingomonas sanxanigenens* (strain DSM 19645), *Sphingopyxis terrae* (strain NBRC 15098), *Sphingopyxis* sp. (strain 113P3), *Altererythrobacter ishigakiensis* (strain NBRC 107699), and *Zymomonas mobilis* (strains ATCC 10988, ATCC 29191, ATCC 29192, NCIMB 11163, NRRL B-12526, CP4, and ZM4) supports the differential gene loss hypothesis, and implies that the element is neither part of the "core genome" of the *Sphingomonadales*, nor is essential for bacterial function.

## The Orthologous Prophages Contain an ORF Encoding a Putative Proline-Enriched Protein

Among the 22 ORFs identified within region I of Prophage I^WHSC-8^, the first encoded a putative outer membrane protein A (TS85_16970, OmpA, 378 aa; fig. 1). Comparative genomics provided several interesting insights into this ORF. Of the 32 genomes shown in table 1, 27 contained a single copy of this ORF, which occurred as the first ORF of the orthologous prophage in each. The putative proteins encoded by these ORFs varied in their length (349–384 aa, average 365 aa) and identity (53–76%, average 58%, using TS85_16970 as the query sequence). In the genome of *Altererythrobacter namhicola* JCM 16345, there was a single copy of this ORF (A6F65_01235, 318 aa) that was not a part of the orthologous prophage. In the genomes of *P. neustonensis* DSM 9434 and *Croceicoccus naphthovorans* PQ-2, there were two copies of this ORF (A9D12_03720; A9D12_12925 and AB433_03655; AB433_05665, respectively) and only one of them occurred as the first ORF of the orthologous prophage in each (table 1). However, in the genomes of *Altererythrobacter atlanticus* 26DY36 and *Erythrobacter atlanticus* s21-N3, there were two copies of this ORF (WYH_00683;WYH_02434 and CP97_02815; CP97_06350, respectively) and neither of them was part of the orthologous prophage. The two copies identified in each of these four genomes could represent a gene duplication event or an independent acquisition. Furthermore, at least one copy of this ORF was also present in the genomes of several *Sphingomonadales* that were devoid of an ortholgous prophage (e.g., *Sm. sanxanigenens* DSM 19645, NX02_04620, 356 aa). The occurrence of this ORF in the absence of an ortholgous prophage indicates that it is either a relic of an ancient prophage or an independent acquisition that may have been selectively maintained. Outside of the *Sphingomonadales*, homologs (~35% identity) of TS85_16970 were found only in a few bacteria; for example, in the genera *Magnetospirillum* (A6A05_07025, MGR_1855, and H261_18647), *Brevundimonas* (ASC65_02295), and *Scytonema* (QH73_05930). Taken together, these results suggest that the putative OmpA-encoding ORF is as much an integral part of the ortholgous prophages in some *Sphingomonadales*, as it is of the pangenome of the order.

Each of the putative OmpA-orthologs from 29 *Sphingomonadales* contained a signal peptide (first 21 aa, fig. 2), which was similar to the signal peptide of OmpA of *E. coli* (Movva et al. 1980). This feature suggested that the putative proteins are secreted or membrane-bound. The most conspicuous feature of the putative OmpA orthologs listed in table 1 was a proline-rich region (fig. 2), which appeared to divide the protein into two unequal halves. The total number of prolines (and the number of contiguous prolines) differed among the orthologous proteins. The highest number of prolines (and the highest number of contiguous prolines, 27/27; fig. 2) were found in the OmpA ortholog from *Er. litoralis* HTCC2594 (ELI_13950, 367 aa). Although there appeared to be no relationship between the length of the putative proteins and the number of prolines, the OmpA orthologs from some *Sphingomonas* spp. contained few prolines (and fewer contiguous prolines, fig. 2). The OmpA proteins of Gram-negative bacteria typically contain a proline-rich region that acts as a linker/hinge between the N- and C-terminal domains
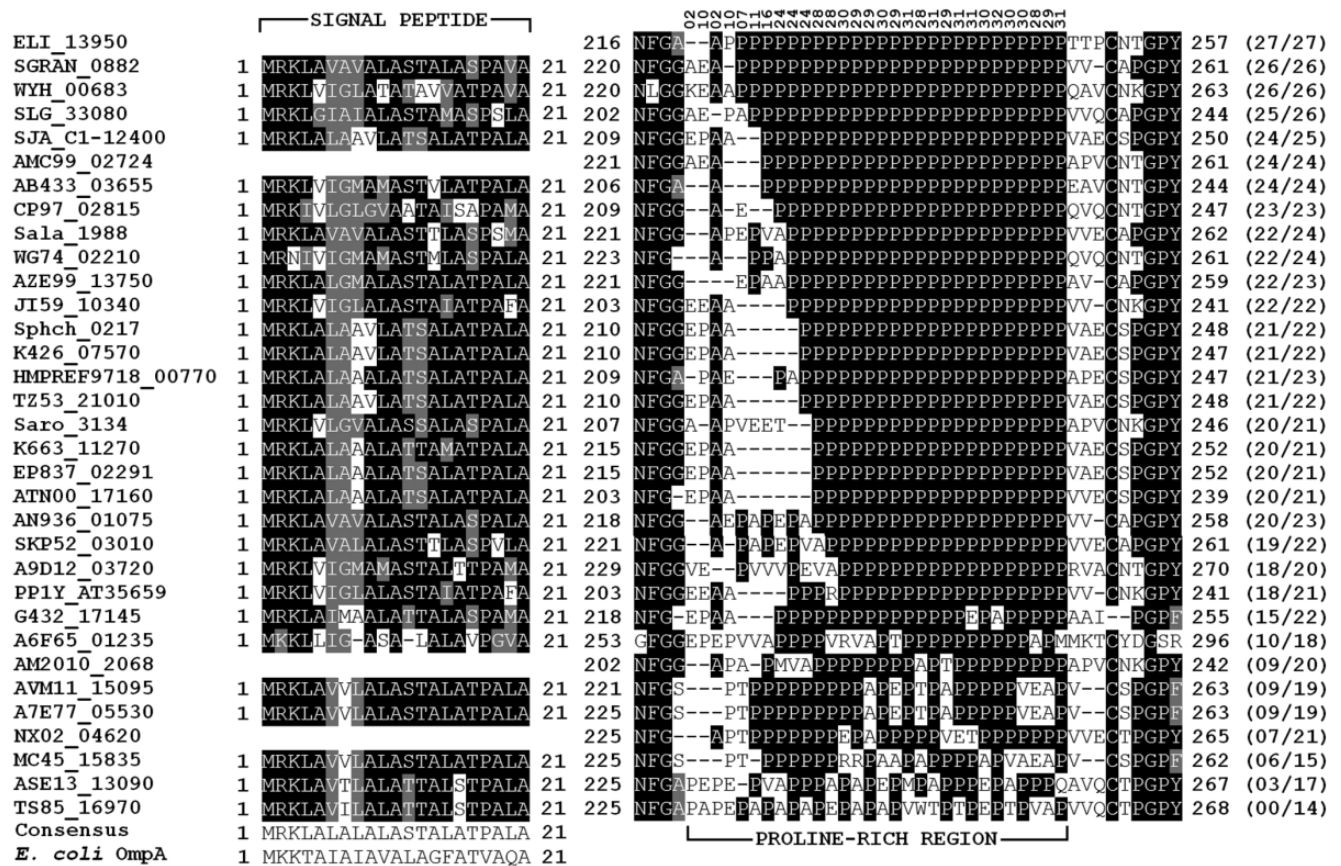
FIG. 2.—Comparison of putative OmpA orthologs from 33 bacterial strains (represented by locus tag numbers shown in table 1) by multiple sequence alignment. Signal peptide sequences (first 21 aa, shown on the left side) could not be identified in the OmpA orthologs of four bacterial strains (ELI_13950, AMC99_02724, AM2010_2068, and NX02_04620). Proline-rich linker/hinge regions are shown on the right side. Numbers on the top indicate the total number of prolines in each column. Numbers in parenthesis indicate the number of contiguous prolines versus the total number of prolines in each OmpA.

(Nikaido and Vaara 1985; El Hamel et al. 2001). These proteins are also referred to as β-barrel porins because they form transmembrane barrels that are predominantly composed of β-sheets (Delcour 2002; Tamm et al. 2004). The putative OmpA ortholog from *Er. litoralis* HTCC2594 appears to be a β-barrel porin based on its predicted secondary structure and model (see supplementary fig. S1, Supplementary Material online). It is possible that the other putative OmpA orthologs listed in table 1 are also β-barrel porins because their predicted secondary structures contained 8–10 β-sheets in the N-terminal domains (data not shown).

The occurrence of an ORF encoding a putative outer membrane protein among the orthologous prophages of *Sphingomonadales* was not surprising because many bacteriophages (and prophages) associated with Gram-negative bacteria have been shown to carry ORFs encoding porins (Highton et al. 1985; Zhao et al. 2011). It was speculated that these proteins may prevent superinfection by repressing the expression of other genes encoding outer membrane proteins that serve as phage receptors

(e.g., OmpC), and may also facilitate phage survival in "induced lysogens" (Highton et al. 1985). It was also proposed that these proteins "play a structural role in phage assembly" (Zhao et al. 2011). In addition to serving as receptors for antibiotics, bacteriophages, and colicins, β-barrel porins transport various molecules across the outer membrane (Petersen et al. 2007). Because the ORFs encoding putative OmpA proteins were consistently identified among the genomes of many *Sphingomonadales*, and they were present even in orthologous prophages that were highly truncated (e.g., in *N. pentaromativorans* US6-1 and *P. neustonensis* DSM 9434), it is likely that they confer some fitness and are thus selectively maintained. Their functional relevance within the respective hosts remains to be characterized.

## Six Orthologous Prophages Contain an ORF Encoding a Putative Spermidine Synthase

Among the 22 ORFs identified within region I of Prophage I[WHSC-8], the twentieth encoded a putative

spermidine synthase (TS85_17065, SpeE, 227 aa; fig. 1 and see supplementary table S1, Supplementary Material online). Comparative analyses indicated that the orthologous prophages in the draft genomes of *Sphingomonas* sp. Root 241 and *Sphingomonas* sp. PAMC 26617 contained a similar ORF (fig. 1). Furthermore, the orthologous prophages in the draft genomes of *Sphingomonas* sp. PAMC 26605 and 26621 as well as the complete genome of *Sphingomonas panacis* DCY99 also contained an ORF encoding a putative SpeE (data not shown). The draft/complete genomes of 28 other *Sphingomonadales* contained a similar ORF, but in each of these genomes the ORF was located outside the orthologous prophage (table 1). The proteins encoded by these ORFs contained ~222 aa and their identity was 55–82% (average 65%, using TS85_17065 as the query sequence). The above results point to two evolutionary possibilities for the *speE* ORF among *Sphingomonadales*; 1) that it was once part of the orthologous prophage, but has been subsequently translocated to another part of the chromosome in many species/strains; and 2) that it was part of the chromosome, but has been translocated into the orthologous prophage in the common ancestor of a few species/strains.

Neither the presence of polyamines within bacteriophages nor the occurrence of genes involved in polyamine biosynthesis in their genomes is unusual (Tabor and Tabor 1985; Shaw et al. 2010). However, the occurrence of such genes within prophages has not been reported hitherto. Therefore, the orthologous prophages from strains WHSC8, Root 241, PAMC 26605, PAMC 26617, PAMC 26621, and DCY99 are unusual in containing an ORF encoding a putative SpeE. Another feature of these and most other *Sphingomonas* strains is that their genomes either lacked an ORF (or contained a truncated/disrupted ORF) encoding a putative S-adenosylmethionine decarboxylase (SpeD). Consequently, these strains may not be able to produce spermidine and the *speE* ORFs within their genomes may be redundant. Indeed, chemotaxonomic studies of strains WHSC-8 and DCY99 have shown that they contain *sym*-homospermidine as the major polyamine (Singh et al. 2015; Wei et al. 2015). The fact that the *speE* ORFs are selectively maintained within their respective hosts indicates that they may have an as yet unknown function. In view of these observations, and the prediction that genes encoding homospermidine synthases have spread from *Alphaproteobacteria* to other bacteria and viruses through horizontal gene transfer (Shaw et al. 2010), it is possible that the ORFs encoding putative spermidine synthases were phage-borne.

## The Evolutionary Rates of Many Orthologous Prophages and Their Hosts Appear Similar

Because several bacterial strains (e.g., *Citromicrobium* sp. JL477, *Sphingorhabdus* sp. M41, *Sphingomonas* sp. Root 241, *Sphingomonas* sp. MM-1, *Sphingobium* sp. YBL2,

*Sphingobium* sp. MI1205, *Sphingobium* sp. EP60837, and *Sphingobium* sp. TKS) that were part of this study lacked proper taxonomic assignment, it was imperative that their phylogenetic position be assessed within the context of the order *Sphingomonadales*. Therefore, a phylogenetic tree based on the proteomes of 19 bacterial strains was constructed using the CVTree 3.0 tool. Five distinct clades could be recognized within this tree (fig. 3A). Clades I and II contained species and strains of the genera *Sphingomonas* and *Sphingobium*, respectively. Within clade I, strain Root 241 clustered close to strain WHSC8, further confirming the BLASTP results in supplementary table S1, Supplementary Material online. Within clade II, strains YBL2, EP60837, and TKS appeared distinct from each other. More importantly, the genera *Sphingomonas* and *Sphingobium* branched into sister lineages in the tree.

Clade III in the proteome-based tree (fig. 3A) contained *Sphingorhabdus* sp. M41, and clade IV contained two strains of the genus *Novosphingobium*. Interestingly, these two genera of *Sphingomonadaceae* branched away from clades I and II, and were placed on a main branch that contained members of the *Erythrobacteraceae* (clade V). A maximum likelihood phylogenetic tree (with 1,000 bootstrap replicates) constructed using 16S rDNA sequences also showed branching of *Sphingomonas* and *Sphingobium* into sister lineages, and placed *Sphingorhabdus* and *Novosphingobium* on a main branch that contained members of the *Erythrobacteraceae* (fig. 4A). Furthermore, a previous phylogenetic analysis based on concatenated sequences of 400 conserved proteins has also shown that some members of the genus *Novosphingobium* cluster near *Er. litoralis* HTCC2594 (Gan et al. 2015). Based on these results, the taxonomic assignment of some genera of *Sphingomonadaceae* needs to be reexamined.

While analyzing the prophages among *S. enterica* and *E. coli*, Bobay et al. (2014) proposed that the evolutionary rates of orthologous prophages would be similar to those of their hosts, and demonstrated that a phylogenetic tree based on phage-derived protein sequences mirrors that of their respective hosts. To test this hypothesis, a phylogenetic tree was constructed using the CVTree 3.0 tool by analyzing the protein sequences of four ORFs (TS85_16980, TS85_16985, TS85_16990, and TS85_17075; see supplementary table S1, Supplementary Material online) of Prophage I^WHSC-8 that were conserved in 18 other orthologous prophages. The topology of this tree (fig. 3B) was similar to that of the proteome-based tree (fig. 3A); the genera *Sphingomonas* and *Sphingobium* branched into sister lineages, and the genera *Novosphingobium* and *Sphingorhabdus* were placed closer to members of the *Erythrobacteraceae*. Furthermore, a maximum likelihood phylogenetic tree (with 1,000 bootstrap replicates) constructed using the concatenated protein sequences of the four ORFs was similar to the 16S rDNA-based tree and
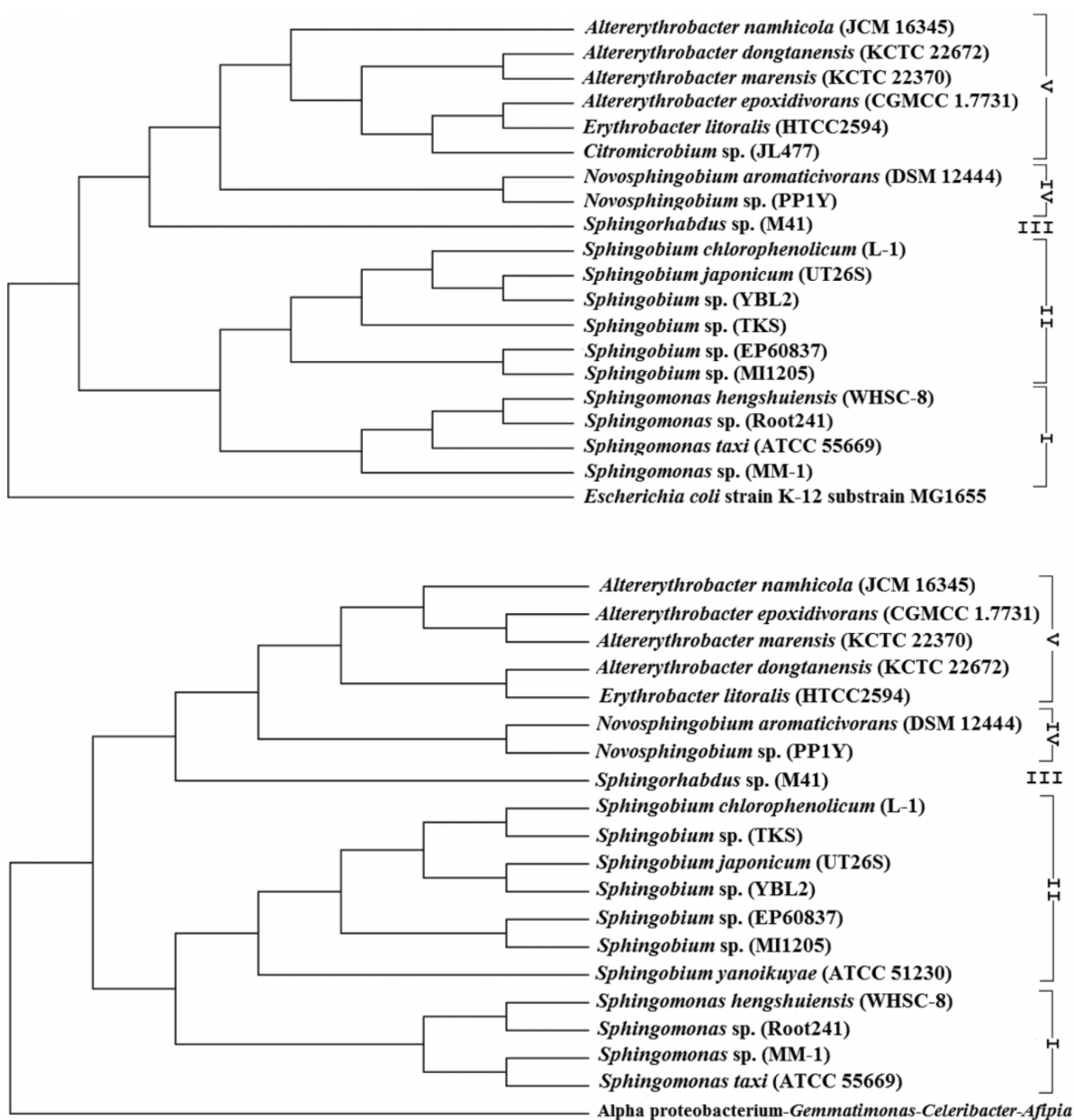
**FIG. 3.**—(A) (TOP) Phylogenetic tree based on the proteomes of 19 bacterial strains of the order *Sphingomonadales*. The proteome of *Escherichia coli* strain K-12 substrain MG1655 (UniProt Proteome ID: UP000000625) was used as an outgroup. Except the outgroup, all other strains contained an orthologous prophage (table 1). (B) (BOTTOM) Phylogenetic tree based on the protein sequences of four ORFs that were conserved in 19 orthologous prophages of the order *Sphingomonadales*. Homologous protein sequences from four bacteria (Alpha proteobacterium Q-1, GAK34242; *Gemmatimonas aurantiaca* T-27, BAH39687; *Celeribacter halophilus*, WP_066598903; *Afipia* sp. P52-10, ETR76025) were combined and used as an outgroup because a single species/strain that contained homologs of all four protein sequences could not be found. Both trees were constructed using the Neighbor-Joining method by the web server CVTree3. The top tree was visualized at $K = 6$ and the bottom tree was visualized at $K = 4$. In both trees, five distinct clades recognized based on species/strains represented within each are marked on the right side.

corroborated the inferences (fig. 4B). The discrepancies in the placement of species/strains within the main branches among the different trees could be due to the number and types of characters analyzed as well as the methods of analyses. Taken together, the results from the two types of phylogenetic analyses (Neighbor-Joining and maximum likelihood) indicate that the orthologous prophages have resided within their hosts for longer periods of time, and that most of them are evolving *in sync* with their respective host chromosomes.
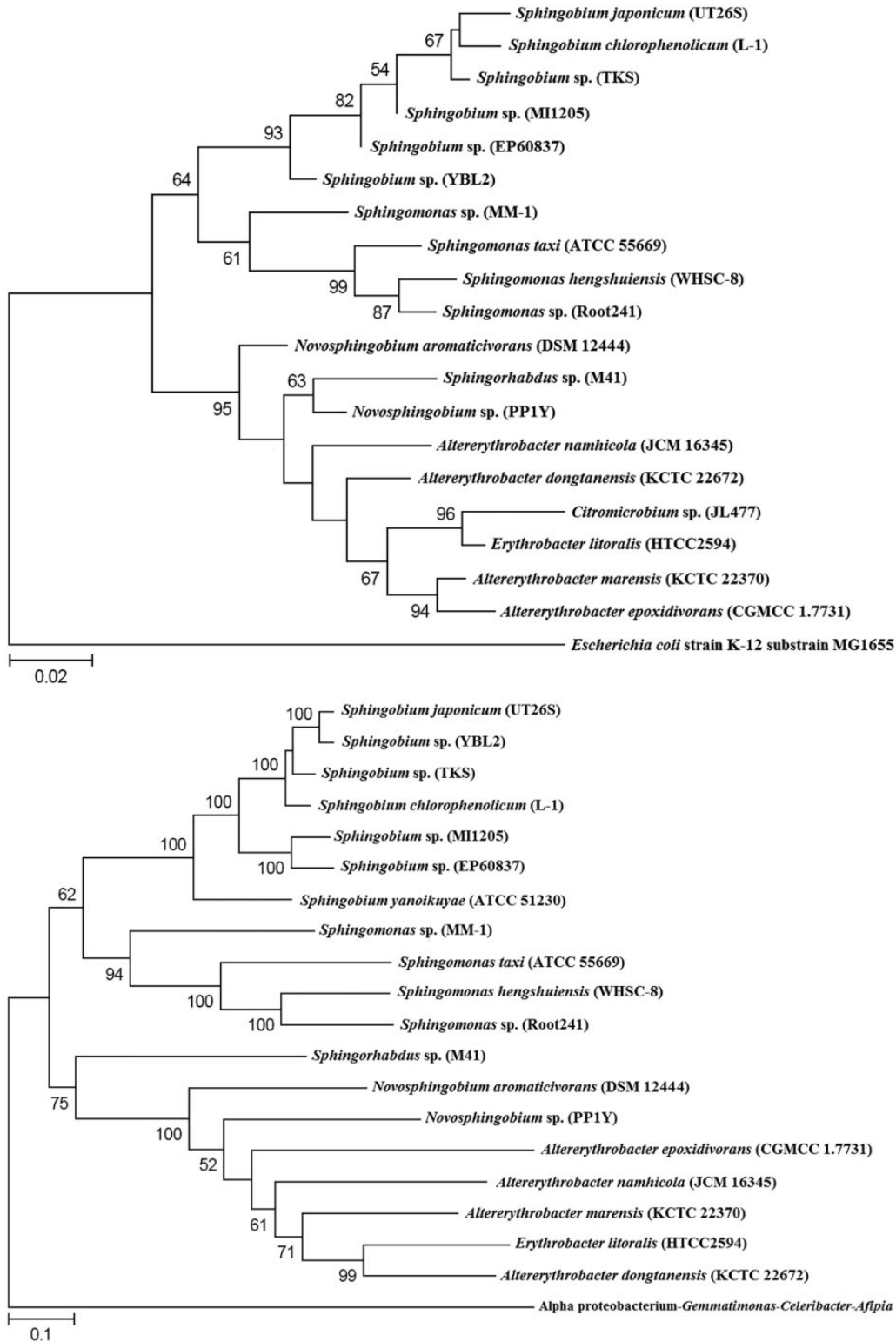
**Fig. 4.**—(A) (TOP) Phylogenetic tree based on the 16S rDNA sequences (~940 bp) of 19 bacterial strains of the order *Sphingomonadales*. The tree was rooted using the 16S rDNA sequence of *Escherichia coli* strain K-12 substrain MG1655 (GenBank locus tag AW869_04565) as the outgroup. Except the outgroup, all other strains contained an orthologous prophage (table 1). (B) (BOTTOM) Phylogenetic tree based on the concatenated protein sequences (~1,141 aa) of four ORFs that were conserved in 19 orthologous prophages of the order *Sphingomonadales*. Protein sequences were concatenated in the same order as their ORFs occurred in figure 1. The outgroup is similar to the one used in figure 3B. Both trees were constructed using the maximum likelihood method in MEGA 6.0. Bootstrap values of 1,000 replicates are indicated as numbers out of 100 at the nodes (only values >50 are shown). Scale bars show the number of nucleotide/aa substitutions per site.

## Conclusions

The results from the present study indicate that *Sm. hengshuiensis* strain WHSC-8 may contain an intact prophage. Whether this prophage is inducible remains to be explored. The results also suggest that large sections of this prophage are likely to be conserved among various species of *Sphingomonas* and *Sphingobium*, and that a small section (from region I) is likely to be conserved among various genera of *Sphingomonadaceae* and *Erythrobacteraceae*. These scenarios may become more apparent as the genomes of other species and strains of *Sphingomonadales* are sequenced and compared. The "stabilization" of a small section of this prophage is likely to be due to the selective fitness of one or more of the genes contained therein. Furthermore, it appears that the evolutionary rates of many of the orthologous prophages are similar to those of their hosts. However, it remains to be determined using a large number of complete genomes whether the genes within these prophages are under purifying selection. Finally, this group of bacteria has come a long way after being recognized as a novel genus (Yabuuchi et al. 1990), a novel family (Kosako et al. 2000), and a novel order (Yabuuchi and Kosako 2005) in terms of characterization of the genetic relationships and evolution.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Literature Cited

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:W16–W21.

Aylward FO, et al. 2013. Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. Appl Environ Microbiol. 79:3724–3733.

Bai Y, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528:364–369.

Bobay LM, Rocha EP, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol. 30:737–751.

Bobay LM, Touchon M, Rocha EP. 2014. Pervasive domestication of defective prophages by bacteria. Proc Natl Acad Sci U S A. 111:12127–12132.

Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage genomics. Microbiol Mol Biol Rev. 67:238–276.

Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? Mol Microbiol. 49:277–300.

Delcour AH. 2002. Structure and function of pore-forming β-barrels from bacteria. J Mol Microbiol Biotechnol. 4:1–10.

El Hamel C, Chevalier S, Dé E, Orange N, Molle G. 2001. Isolation and characterisation of the major outer membrane protein of *Erwinia carotovora*. Biochim Biophys Acta. 1515:12–22.

Gan HM, et al. 2015. Whole genome sequencing and analysis reveal insights into the genetic structure, diversity and evolutionary relatedness of *luxI* and *luxR* homologs in bacteria belonging to the *Sphingomonadaceae* family. Front Cell Infect Microbiol. 4:188.

García-Romero I, et al. 2016. Genomic analysis of the nitrate-respiring *Sphingopyxis granuli* (formerly *Sphingomonas macrogoltabida*) strain TFA. BMC Genomics. 17:93.

Garrity GM, Bell JA, Lilburn T. 2005a. Phylum XIV. *Proteobacteria* phyl. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey's manual of systematic bacteriology. 2nd ed., Vol. 2 (The Proteobacteria), Part B (The Gammaproteobacteria). New York: Springer. p. 1.

Garrity GM, Bell JA, Lilburn T. 2005b. Class I. *Alphaproteobacteria* class. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey's manual of systematic bacteriology. 2nd ed., Vol. 2 (The Proteobacteria), Part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria). New York: Springer. p. 1.

Glaeser S, Kämpfer P. 2014. The family *Sphingomonadaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes. 4th ed. (Alphaproteobacteria and Betaproteobacteria). Heidelberg (Germany): Springer. p. 641–707.

Highton PJ, Chang Y, Marcotte WR Jr, Schnaitman CA. 1985. Evidence that the outer membrane protein gene *nmpC* of *Escherichia coli* K-12 lies within the defective qsr' prophage. J Bacteriol. 162:256–262.

Joklik WK, editor. 1999. Microbiology: a centenary perspective. Washington (DC): ASM Press.

Kosako Y, Yabuuchi E, Naka T, Fujiwara N, Kobayashi K. 2000. Proposal of *Sphingomonadaceae* fam. nov., consisting of *Sphingomonas* Yabuuchi et al. 1990, *Erythrobacter* Shiba and Shimidu 1982, *Erythromicrobium* Yurkove et al. 1994, *Porphyrobacter* Fuerst al. 1993, *Zymomonas* Kluyver and van Niel 1936, and *Sandaracinobacter* Yurkove et al. 1997, with the type genus *Sphingomonas* Yabuuchi et al. 1990. Microbiol Immunol. 44:563–575.

Lee KB, et al. 2005. The hierarchical system of the '*Alphaproteobacteria*': description of *Hyphomonadaceae* fam. nov., *Xanthobacteraceae* fam. nov. and *Erythrobacteraceae* fam. nov. Int J Syst Evol Microbiol. 55:1907–1919.

Lwoff A. 1953. Lysogeny. Bacteriol Rev. 17:269–337.

Lwoff A. 1966. Interaction among virus, cell, and organism. Science 152:1216–1220.

Movva NR, Nakamura K, Inouye M. 1980. Amino acid sequence of the signal peptide of OmpA protein, a major outer membrane protein of *Escherichia coli*. J Biol Chem. 255:27–29.

Nikaido H, Vaara M. 1985. Molecular basis of bacterial outer membrane permeability. Microbiol Rev. 49:1–32.

Oh HM, Giovannoni SJ, Ferriera S, Johnson J, Cho JC. 2009. Complete genome sequence of *Erythrobacter litoralis* HTCC2594. J Bacteriol. 191:2419–2420.

Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007. Genes under positive selection in *Escherichia coli*. Genome Res. 17:1336–1343.

Shaw FL, et al. 2010. Evolution and multifarious horizontal transfer of an alternative biosynthetic pathway for the alternative polyamine sym-homospermidine. J Biol Chem. 285:14711–14723.

Singh A, Lal R. 2009. *Sphingobium ummariense* sp. nov., a hexachlorocyclohexane (HCH)-degrading bacterium, isolated from HCH-contaminated soil. Int J Syst Evol Microbiol. 59:162–166.

Singh P, Kim YJ, Hoang VA, Farh Mel A, Yang DC. 2015. *Sphingomonas panacis* sp. nov., isolated from rhizosphere of rusty ginseng. Antonie Van Leeuwenhoek 108:711–720.

Stackebrandt E, Murray RGE, Trüper HG. 1988. *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives". Int J Syst Bacteriol. 38:321–325.

Tabor CW, Tabor H. 1985. Polyamines in microorganisms. Microbiol Rev. 49:81–99.

Takeuchi M, Sawada H, Oyaizu H, Yokota A. 1994. Phylogenetic evidence for *Sphingomonas* and *Rhizomonas* as nonphotosynthetic members of the alpha-4 subclass of the *Proteobacteria*. Int J Syst Bacteriol. 44:308–314.

Tamm LK, Hong H, Liang B. 2004. Folding and assembly of β-barrel membrane proteins. Biochim Biophys Acta. 1666:250–263.

Tonon LAC, Moreira APB, Thompson F. 2014. The family *Erythrobacteraceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes. 4th ed. (Alphaproteobacteria and Betaproteobacteria). Heidelberg (Germany): Springer. p. 213–235.

Verma H, et al. 2014. Comparative genomic analysis of nine *Sphingobium* strains: insights into their evolution and hexachlorocyclohexane (HCH) degradation pathways. BMC Genomics 15:1014.

Yabuuchi E, et al. 1990. Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. Microbiol Immunol. 34:99–119.

Yabuuchi E, Kosako Y. 2005. Order IV. *Sphingomonadales* ord. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey's manual of systematic bacteriology. 2nd ed., Vol. 2 (The Proteobacteria), Part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria). New York: Springer. p. 230–286.

Wei S, et al. 2015. *Sphingomonas hengshuiensis* sp. nov., isolated from lake wetland. Int J Syst Evol Microbiol. 65:4644–4649.

Woese CR. 1987. Bacterial evolution. Microbiol Rev. 51:221–271.

Zhao H, Sequeira RD, Galeva NA, Tang L. 2011. The host outer membrane proteins OmpA and OmpC are associated with the *Shigella* phage Sf6 virion. Virology 409:319–327.

Zheng Q, et al. 2014. A marine inducible prophage vB_CibM-P1 isolated from the aerobic anoxygenic phototrophic bacterium *Citromicrobium bathyomarinum* JL354. Sci Rep. 4:7118.

Zheng Q, Lin W, Liu Y, Chen C, Jiao N. 2016. A comparison of 14 *Erythrobacter* genomes provides insights into the genomic divergence and scattered distribution of phototrophs. Front Microbiol. 7:984.

Zuo G, Hao B. 2015. CVTree3 Web Server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. Genomics Proteomics Bioinformatics 13:321–331.

**Associate editor:** Bill Martin