

## Prophage Genomics

Carlos Canchaya, Caroline Proux, Ghislain Fournous, Anne Bruttin, and Harald Brüssow\*

*Nestlé Research Center, Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland*

|  |     |
|--|-----|
| <b>INTRODUCTION</b> .....  | 239 |
| <b>Technical Difficulties</b> .....                                  | 240 |
| <b>THEORETICAL FRAMEWORK</b> .....                                   | 241 |
| <b>The Phage Side</b> .....  | 241 |
| <b>The Bacterial Side</b> .....                                      | 241 |
| <b>PROPHAGES FROM LOW-G+C GRAM-POSITIVE BACTERIA</b> .....           | 242 |
| <i>Streptococcus pyogenes</i> .....                                  | 242 |
| <i>Streptococcus agalactiae</i> and <i>Streptococcus mitis</i> ..... | 247 |
| <i>Streptococcus pneumoniae</i> .....                                | 247 |
| <i>Streptococcus thermophilus</i> .....                              | 249 |
| <i>Lactococcus lactis</i> .....                                      | 250 |
| <i>Lactobacillus</i> .....   | 251 |
| <i>Listeria</i> .....  | 251 |
| <i>Staphylococcus aureus</i> .....                                   | 252 |
| <i>Bacillus</i> .....  | 254 |
| <i>Clostridium</i> .....   | 255 |
| <b>PROPHAGES FROM HIGH-G+C GRAM-POSITIVE BACTERIA</b> .....          | 257 |
| <i>Corynebacterium</i> .....   | 257 |
| <i>Streptomyces</i> .....  | 257 |
| <i>Bifidobacterium</i> .....   | 257 |
| <i>Mycobacterium</i> .....   | 257 |
| <b>γ-PROTEOBACTERIA</b> .....  | 258 |
| <b>Basic Phage Types</b> .....                                       | 258 |
| <b>Chimeric Prophages</b> .....                                      | 258 |
| <i>Escherichia coli</i> and <i>Shigella flexneri</i> .....           | 263 |
| <i>E. coli</i> O157.....   | 263 |
| <i>E. coli</i> K-12 .....  | 263 |
| Uropathogenic <i>E. coli</i> .....                                   | 263 |
| <i>S. flexneri</i> .....   | 263 |
| Location of virulence factors.....                                   | 264 |
| Expression of prophage genes.....                                    | 264 |
| <i>Salmonella</i> .....  | 264 |
| <i>S. enterica</i> serovar Typhimurium .....                         | 264 |
| <i>S. enterica</i> serovar Typhi.....                                | 265 |
| <i>Yersinia</i> and <i>Haemophilus</i> .....                         | 265 |
| <i>Vibrio</i> .....  | 265 |
| <b>Plant Pathogens</b> .....   | 266 |
| <i>X. fastidiosa</i> .....   | 266 |
| <i>Xanthomonas</i> .....   | 267 |
| <b>Opportunistic Pathogens and Free-Living Bacteria</b> .....        | 267 |
| <i>Pseudomonas</i> .....   | 267 |
| <i>Shewanella</i> .....  | 268 |
| <b>OUTLOOK</b> .....   | 268 |
| <b>Prophage Distribution</b> .....                                   | 268 |
| <b>Age of the Prophages</b> .....                                    | 269 |
| <b>Ephemeral Nature of Prophages?</b> .....                          | 269 |
| <b>Prophage Genome Diversity</b> .....                               | 269 |
| <b>Detecting Prophages</b> .....                                     | 270 |
| <b>Future Tasks</b> .....  | 270 |
| <b>ACKNOWLEDGMENTS</b> .....   | 270 |
| <b>REFERENCES</b> .....  | 271 |

\* Corresponding author. Mailing address: Nestlé Research Center, Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland. Phone: 41 21 785 8676. Fax: 41 21 785 8925. harald.brussow@rdls.nestle.com.

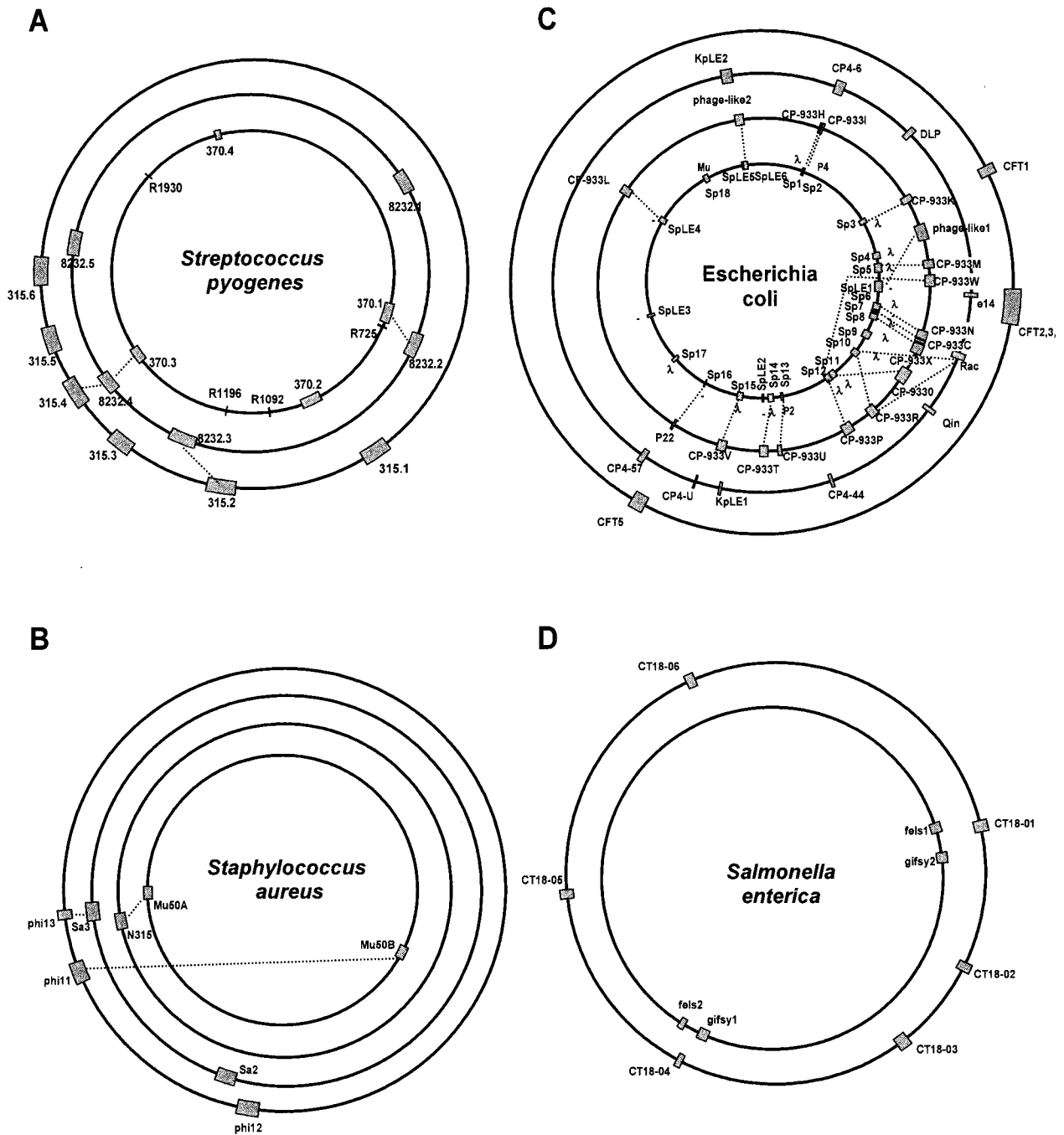


FIG. 1. Prophage content of four human bacterial pathogens. The prophages are indicated as shaded boxes on the bacterial genome maps. The lengths of the boxes correspond to the relative sizes of the prophage DNA with respect to the bacterial chromosome. Note that the circumference of the bacterial genomes does not correspond to their relative length. Prophages with extensive DNA sequence identity are linked by dotted lines. (A) *S. pyogenes* genomes of the sequenced M1, M18, and M3 strains (from center to periphery). (B) *S. aureus* genomes of the sequenced Mu50 (center), N315, MW2, and 8325 strains. (C) *E. coli* genomes from the O157:H7 Sakai (center) and O157:H7 EDL933 strains, the laboratory strain K-12, and the uropathogenic strain CFT073. (D) *S. enterica* serovar Typhimurium LT2 (center) and serovar Typhi CT18.

INTRODUCTION

Many bacterial genomes deposited in the public database contain phage DNA integrated into the bacterial chromosome. It is not rare for bacteria to contain multiple prophages in their chromosomes, which then constitute a sizable part of the total bacterial DNA (Fig. 1). The most extreme case is currently

represented by the food pathogen *Escherichia coli* O157:H7 strain Sakai. It contains 18 prophage genome elements, which amount to 16% of its total genome content (Fig. 1C). Less extreme but still impressive cases are represented by *Streptococcus pyogenes*, with four to six prophages, amounting to 12% of the bacterial DNA content (Fig. 1A). These prophages do

not represent exotic phage types: the *E. coli* O157 prophages resemble the well-known temperate *E. coli* phages  $\lambda$ , P2 (and its satellite phage P4), and Mu (160). The *S. pyogenes* prophages belong to the proposed Sfi11-, Sfi21-, and r1t-like *Siphoviridae*, which are also found in lactic acid bacteria (LAB) used in industrial milk fermentation. The taxonomy we use in this review for phages from low-G+C gram-positive bacteria is our own system based on comparative phage genomics (28, 171). Other authors have proposed partially overlapping and partially distinct phage taxonomy systems based on a phage proteomics tree (in these systems the Sfi11-like phages are called TP901-like phages after another type of phage from the same group) or a differentiation of phage genomes into a set of modi (modules) (114, 176).

Prophages are not only quantitatively important genetic elements of the bacterial chromosome. As mobile DNA elements, phage DNA is a vector for lateral gene transfer between bacteria (35). Indeed, numerous virulence factors from bacterial pathogens are phage encoded (22, 216, 215). It was postulated that this role of prophages is not limited to pathogenic bacteria but that some adaptations of nonpathogenic bacterial strains to their ecological niche might also be mediated by prophage genomes (30). Furthermore, prophages account for a substantial amount of interstrain genetic variability in several bacterial species (e.g., *Staphylococcus aureus* [7] and *S. pyogenes* [189]). When genomes from closely related bacteria were compared in a dot plot analysis, prophage sequences frequently accounted for a substantial, if not the major, proportion of the differences between the genomes (e.g., *Listeria monocytogenes* and *L. innocua* [79], *Salmonella enterica* serovars Typhi and Typhimurium [139, 164], and *E. coli* O157 and K-12 [166]). Microarray analysis (169) and PCR scanning (161) allowed researchers to explore the presence of specific prophages over a much larger set of related bacterial strains, and again prophages contributed a large part of strain-specific DNA, irrespective of whether pathogenic or nonpathogenic bacteria were investigated. Finally, when mRNA expression patterns were studied with microarrays in lysogenic bacteria that underwent physiologically relevant changes in growth conditions, prophage genes figured prominently in the mRNA species changing their expression pattern (190, 220). These data demonstrate that prophages are not a passive genetic cargo of the bacterial chromosome but are likely to be active players in cell physiology. Subtractive mRNA hybridization analysis demonstrated that prophage genes also make up prominent share of the *E. coli* genes upregulated when the bacteria invaded the lungs of infected birds (60). Apparently, prophage genomes are an important target for selection working on bacterial genomes. Indeed, in medical microbiology there are good indications that prophage acquisition actually shaped the epidemiology of some important bacterial pathogens (8). We summarize here some recently formulated ideas (22, 53, 115) on the coevolution of bacteria and phages, and we have screened the published bacterial genome sequences for prophage sequences. Specifically, we looked for the possible role of phage-encoded genes in the adaptation of the lysogenic bacterium to its specific environment, whether the bacterial host is an animal or plant pathogen or a commensal or a free-living bacterium. In addition, we asked where candidate lysogenic conversion genes (genes that could change the phenotype of

the lysogenic bacterial host) are integrated into the prophage genomes.

### Technical Difficulties

A review of prophage sequences in sequenced bacterial genomes has technical difficulties. On a very practical side, prophage sequences are currently not compiled in the National Center for Biotechnology Information (NCBI) phage database. Therefore, the interested scientist has to turn to the original publication and the annotations of the GenBank entry for the bacterial genomes to locate prophage sequences or has to reanalyze the genome sequence. However, no uniform criteria have been established for the diagnosis of prophages in bacterial genome sequences. Prophages can be present in many different forms ranging from inducible prophages to prophages showing deletions, insertions, and rearrangements to prophage remnants that have lost most of the phage genome. In addition, computer programs have difficulties in detecting prophage sequences. Only a few, if any, phage genes are sufficiently conserved and distinct from bacterial genes to serve as markers for prophage sequences. Computer programs efficiently detect integrase genes. However, it is not clear what qualifies an integrase as phage related. Several conjugative transposon-like elements contain lambda family integrases, as do integrons and pathogenicity islands. There are also chromosome-encoded integrases such as XerC/D. Given the presence of this gene family in several kinds of elements, it becomes problematic to use integrase as a prophage signature. In our own experience with one specific class of temperate phages (*Siphoviridae*), reasonably conserved phage proteins are the integrase (32), the portal protein, the terminase (52), and the tail tape measure protein. A further complication is that the current NCBI phage database is small (at the time of writing, it contained 136 complete phage genomes) and is dominated by a single phage group, *Siphoviridae* (1, 136) (contributing 53 complete genomes, followed by 18 *Inoviridae*, 17 *Podoviridae*, and 13 *Myoviridae* genomes). However, phages that are less well documented with respect to genome sequences can also integrate their DNA into the bacterial chromosome, e.g., P2- and Mu-like *Myoviridae* (147, 151), *Inoviridae* in *Vibrio* (217), *Xanthomonas* (48), *Xylella*, and *Pseudomonas* (194), and *Plasmaviridae* in *Acholeplasma* (137). In addition, psiM1-like *Siphoviridae*, lipothrixviruses, and fuselloviruses (167, 232) integrate their genomes into the chromosomes of *Archaea*. Still other forms of lysogeny exist that do not lead to the integration of phage DNA into the bacterial chromosome; e.g., prophages P1 and N15 are maintained as circular or linear plasmids (87, 174), respectively, and *Borrelia* prophages have a peculiar relationship to plasmids (65). We therefore anticipate an underreporting of prophage sequences in the published bacterial genomes.

On the other hand, *Bacillus subtilis* was reported to contain at least 10 prophage sequences (108). Two prophages corresponded to biologically well-defined incomplete prophages (197), and a third prophage was represented by the inducible prophage SPBc2. The diagnosis of the other prophage sequences was based on codon usage analysis (148). However, this analysis cannot easily differentiate prophages from other horizontally acquired DNA elements. For example, in *B. sub-*

*tilis* prophage 2 (numbered in order of appearance of the prophage in the genome) a typical lysogeny module was detected but no other phage links were detected. Prophage 6 showed only few isolated links to SPBc2, while the annotated prophage 7 lacked phage links altogether, casting doubt on their prophage nature. Overreporting of prophage-like elements might thus also be a problem.

## THEORETICAL FRAMEWORK

### The Phage Side

Before going into the analysis of prophage-containing bacterial genomes, it is appropriate to summarize the current ideas about phage-bacterial genome interaction from an evolutionary perspective. The peculiar life-style of temperate phages makes them model systems to address a number of fundamental questions in evolutionary biology. The viral DNA undergoes different selective pressures when replicated during lytic infection cycles compared to prophage DNA maintained in the bacterial genome during lysogeny. Darwinian considerations, along with the selfish-gene concept, lead to interesting conjectures (22, 30, 53, 115). One could anticipate that the prophage decreases the fitness of its lysogenic host by at least two processes: the metabolic burden to replicate extra DNA (Fig. 1) and the lysis of the host after prophage induction. To compensate for these disadvantages, one has to invoke the notion that temperate phages encode functions that increase the fitness of the lysogen. According to the selective value of these postulated phage genes, the lysogenic cell will be maintained or even be overrepresented in the bacterial population. An obvious selective advantage for the lysogenic host is the immunity (phage repressor) and superinfection exclusion genes of the prophage that protect the lysogen against phage infection. These genes also provide a direct advantage to the prophage since they exclude superinfecting phage DNA from competing with the resident prophage DNA for the same host. Where phages from the environment do not provide a sufficiently strong selection pressure, other phage genes have to increase the fitness of the lysogenic host, frequently in rather unanticipated ways (lysogenic conversion genes). Classic examples of such phage-located genes that increase host fitness include the nonessential phage  $\lambda$  genes *bor* and *lom*, which confer serum resistance and better survival in macrophages, respectively, to the *Escherichia coli* lysogen (9). In these cases, the reproductive success of the lysogenic bacterium carrying these new genes translates directly into an evolutionary success for the resident prophage. However, host-parasite relationships also constitute an arms race and therefore represent a highly dynamic genetic equilibrium. Gains from prophages carrying genes that increase host fitness are short-lived from a bacterial standpoint if the resident prophage ultimately destroys the bacterial lineage. In this way, prophages can be considered dangerous molecular time bombs that can kill the lysogenic cell on their eventual induction (115). One would therefore expect evolution to select lysogenic bacteria with mutations in the prophage DNA. Mutations that inactivate the prophage induction process avoid the loss of the lysogenic clone from the bacterial population. One would also expect that selection would lead to large-scale deletion of prophage

DNA in order to decrease the metabolic burden of extra DNA synthesis and a littering of the bacterial genomes by selfish DNA elements. One would predict, furthermore, that useful prophage genes (e.g., lysogenic conversion genes) are preferentially spared from this deletion process, since their loss would actually decrease the fitness of the cell (116). It was proposed that a high genomic deletion rate is instrumental in removing dangerous genetic parasites from the bacterial genome (115). These deletion processes could explain why the bacterial genomes (in general) have not increased in size despite a constant bombardment with parasitic DNA over evolutionary time. The streamlined bacterial chromosome containing few pseudogenes might be the consequence of this deletion process of parasitic DNA.

### The Bacterial Side

New data from comparative bacterial genomics highlight the importance of lateral DNA transfer in microbial evolution (35). Based on a variety of criteria such as sequence matches with other organisms, G+C-content, codon usage, association with mobile DNA elements, and proximity to tRNA genes, it has been estimated that some bacteria capture and fix DNA at a rate of at least 16 kb per  $10^6$  years (116). An interesting case is provided by *E. coli*. Genomic comparison between the pathogenic *E. coli* O157 EDL933 strain and the laboratory *E. coli* strain K-12 revealed 4.1 Mb of common chromosome backbone sequence, 1.3 Mb of O157-specific DNA, and 0.5 Mb of K-12-specific DNA (160). Approximately half of the O157-specific DNA was clear-cut mobile DNA, mostly prophage DNA (166). These observations led to the distinction of a conserved core genome sequence, which is shaped by the mechanisms of vertical evolution, and a variable part of the genome, which is dominated by processes of horizontal evolution. The replacement of tree-like with web-like phylogenies is the visual expression in our current understanding of evolution in the microbial world (58). Phage transduction and prophage integration are major mechanisms of lateral DNA transfer in prokaryotes. Bacteria are therefore confronted with a dilemma: phages are a threat to their survival, against which they must mount defensive countermeasures (surface changes, restriction modification, and a variety of abortive phage infection mechanisms), and at the same time phages are an important tool for the acquisition of genes which could help them to defend their ecological place or gain new ones. Apparently even closely related bacteria addressed this dilemma differently. For example, virulent phages against yogurt strains of *Lactobacillus delbrueckii* (subsp. *bulgaricus*) are a rarity in the food industry, and at the same time it is very hard to transform this organism with foreign DNA. Possibly this bacterium has opted for a strong barrier against intrusion of foreign DNA. In contrast, *Lactococcus lactis* and *Streptococcus thermophilus* phages are readily isolated from the dairy factory or raw milk, and many strains can be transformed in the laboratory, suggesting a greater permissiveness to DNA transfer. However, *L. lactis* has developed numerous antiphage strategies, and an intensive arms race exist between phage and host (44). In contrast, in *S. thermophilus* very few phage defense mechanisms have been identified so far. We even found molecular evidence for cooperation between the host and its phages; e.g.,

the bacterial *attB* site complements the 3' end of the phage integrase gene, which would otherwise lose its C terminus on integration (33). This observation suggests that the phage integrase is of selective value for this bacterial host.

### PROPHAGES FROM LOW-G+C GRAM-POSITIVE BACTERIA

#### *Streptococcus pyogenes*

In 1927, long before lysogeny was described, it was demonstrated that a filterable agent from scarlet fever isolates could convert nonscarlatinal *S. pyogenes* to toxigenic strains (22). We know now that this conversion is mediated by bacteriophages and that 90% of *S. pyogenes* isolates are lysogenic (Fig. 1A). For several reasons, *S. pyogenes* provides a neat test case for the role of prophages. Transformation and conjugation appear to play no or only minor roles in lateral DNA transfer in this species, giving phages a special role in this process (68, 189). *S. pyogenes* belongs to the lactic acid bacteria (LAB) branch of low-G+C gram-positive bacteria, and its phages are good examples of phages from LAB of medical and economical importance and will therefore be discussed in more detail. Phylogenetic relatives are used in the dairy industry as starter organisms, and their viruses have become a focus for comparative phage genomics studies (28). The only known habitat of *S. pyogenes* is the human; it is normally found on the skin and in the oral cavity in humans. *S. pyogenes* comes in many M serotypes and causes an astonishing range of diseases including pharyngitis, scarlet fever, pyodermitis, fasciitis, rheumatic fever, and toxic shock syndrome. With respect to the three sequenced strains, M1 strains were associated with wound infections, M3 strains were identified in patients with severe invasive infections, and M18 strains caused rheumatic fever outbreaks (8). Despite the protean character of this pathogen, the sequenced *S. pyogenes* isolates are genetically closely related. For example, the M18 strain shared 1,532 of 1,696 open reading frames (ORFs) identified in the M1 strain, and sequence identity ranged from 83 to 100% at the base pair level. In fact, a dot plot analysis revealed essentially a straight line between the two serotypes, with 1.7 of the 1.9 Mb of chromosomal DNA shared. There were only five larger regions of difference; all were prophage sequences (189). This observation leads to an interesting question: is the specific pathogenic potential of a given *S. pyogenes* strain influenced by its prophage content? Microarray analysis of 36 M18 strains demonstrated that prophages are not only a significant source of genomic divergence between strains of M1, M3, and M18 serotypes but also the predominant source of difference between M18 strains. The M18 strains differed by a maximum of 3% of the genes, and prophages were responsible for virtually all the variation in gene content. Variation in the prophages ranged from entire absence of the prophages from the test strain to small differences in the gene content of the prophages (189).

When the DNA sequences from the 15 available *S. pyogenes* prophages were compared against each other in a dot plot matrix, five clusters of phages consisting of two to four members could be distinguished, while two phages had only limited DNA sequence similarity to the rest of the phages (Fig. 2, green boxes). Next to *E. coli* prophages, this is the largest set

of prophage sequences from a single species for comparative analysis. Recently, a proposal was made to base the taxonomy of *Siphoviridae* on the genetic organization of the structural gene cluster (excluding the tail fiber genes) (171). All currently known LAB prophages showed a conserved overall gene order: left attachment site (*attL*)-lysogeny-DNA replication-transcriptional regulation-DNA packaging-head-joining-tail-tail fiber-lysis modules-right attachment site (*attR*) (Fig. 3 to 5). Particularly well conserved is the order of the structural genes, allowing the distinction of three major forms of head gene clusters exemplified by the *cos*-site *Streptococcus thermophilus* phage Sfi21, the *pac*-site *S. thermophilus* phage Sfi11, and the *cos*-site *Lactococcus lactis* phage r1t as prototypes. The corresponding phages were also observed in *S. pyogenes*.

The first group of *S. pyogenes* prophages are the r1t-like *Siphoviridae* (Fig. 2). The second group resembles Sfi11-like *Siphoviridae*. Sequence alignments allowed the distinction of two Sfi11 subgroups with *S. pneumoniae* phage MM1 and *S. thermophilus* phage O1205, respectively, as reference strains (Fig. 2). The third group of prophages are members of the Sfi21-like *Siphoviridae*. Database matches again permitted the differentiation of two subgroups: one had sequence similarity to *Lactobacillus gasseri* phage A2 (171), and the other had sequence similarity to staphylococcal phages (Fig. 2). Except for the tail fiber and lysis genes and part of the lysogeny genes, *S. pyogenes* prophage 315.5 lacks DNA sequence similarity to the other *S. pyogenes* prophages, while it has some protein sequence similarity to a *Bacillus halodurans* prophage (see below). Phage tail fibers and lysins have to interact with the cell surface and the cell wall of their bacterial host cell and should therefore be subjected to a strong adaptive selection pressure. Not surprisingly, all but one of the sequenced *S. pyogenes* prophages had a highly related tail fiber module (Fig. 2, red box). Central to this module is the phage hyaluronidase, an enzyme that splits the hyaluronic acid-containing capsule surrounding the bacterial cell. This lytic enzyme allows the phage to reach the cell surface, where it injects its DNA into the bacterial cell. Only prophage 315.6 lacks this tail fiber module typical for *S. pyogenes* prophages. Over these genes, prophage 315.6 had about 40% sequence identity at the amino acid level to *S. agalactiae* prophage SA1 (see below), suggesting a possible cross-species infection. In fact, these two phages had some DNA sequence similarity across their entire genomes.

One would expect competition and exclusion between prophages during the establishment of a polylysogenic cell. Exclusion could be mediated by three proteins encoded in the lysogeny module, namely, the phage integrase, the superinfection exclusion gene *sie* (140), and the phage repressor (immunity function). It is therefore not surprising that a substantial diversification was observed over this region between the *S. pyogenes* prophages (Fig. 2, yellow box; the largest group of phages sharing early genes is marked by blue circles). Eleven distinct prophage integration sites were identified in the three sequenced *S. pyogenes* strains. Three sites were occupied in more than one strain; in all cases, the corresponding phage integrases had at least 92% sequence identity at the amino acid level. Seven prophages used unique integration sites; however, not all possessed unique integrases. Prophages 8232.1 and 315.1 showed distinct chromosomal integration sites while sharing identical integrases. Seven- and 14-bp core sequences

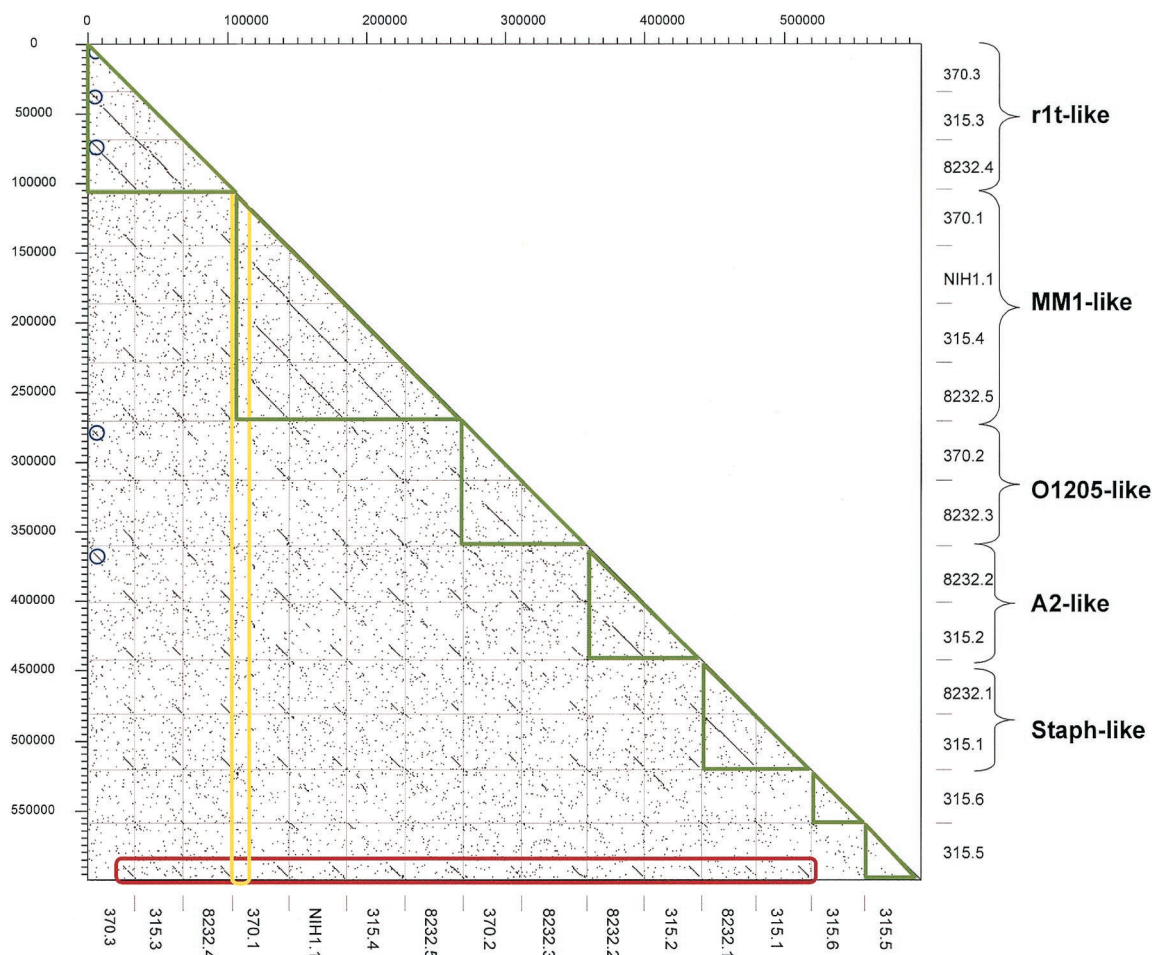


FIG. 2. Dot plot matrix for the currently available 15 *S. pyogenes* prophages identified by their prophage names on the *x* and *y* axes. According to their structural genes, the prophages were classified into distinct groups (green triangles) and annotated on the left ordinate. The prophage genomes were aligned with their integrase gene to the left (top). The extent of the conservation of the tail fiber and lysis genes is highlighted by the red box. The lack of conservation of the lysogeny and DNA replication genes is demonstrated by the yellow box. The largest group of prophages sharing early genes is indicated by the blue circles.

were deduced for the two integration sites, showing a 6-bp overlap. A similar observation was made for prophages 8232.5 and 315.5. Such low specificity of the phage integrase with respect to conserved nucleotides in the core sequence was also observed in other phages from LAB (*S. thermophilus* phage Sfi21 [33] and *Lactobacillus* phage mv4 [5]). It is therefore unlikely that the competition for integration sites is a limiting factor in the establishment of polylysogenic cells.

For the DNA-packaging, head, and tail genes, *S. pyogenes* prophage 370.3 had sequence similarity to *L. lactis* phage r1t (Fig. 3A). Differences between r1t-like *S. pyogenes* prophage were seen in a group of three nonstructural genes preceding the terminase gene (Fig. 3B, box C). An endonuclease gene and a point mutation interrupt and truncate the tail tape gene from prophage 8232.4 (boxes A and B, respectively). Prophage 315.3 showed distinct lysis and lysogenic conversion genes (box E). All three prophages demonstrated variations in the early genes (boxes D) and a disrupted replisome organizer gene (e.g., *orf7a* and *orf7b* in 370.3 [Fig. 3A]). *S. pyogenes* prophage 370.2 could be aligned with *S. thermophilus* phage O1205 in the DNA-packaging, head, and tail genes (Fig. 4A). Differences

between the two O1205-like *S. pyogenes* prophages 370.2 and 8232.3 were seen in the early-gene cluster and the lysis/lysogenic conversion genes. Prophage 8232.3 showed ORF-disrupting point mutations in the tail tape and a tail fiber gene, and a lysis gene is interrupted by an endonuclease, while prophage 370.2 contained a stop codon within the portal gene (Fig. 4A).

The four MM1-like *S. pyogenes* prophages had extensive DNA sequence similarity throughout the entire structural genes. Except for the tail fiber and two head genes, extensive protein sequence similarity to *S. pneumoniae* prophage MM1 was also detected (Fig. 4B). In the structural genes, differences between the MM1-like *S. pyogenes* prophages amounted for a few gene replacements (Fig. 4B and C, boxes B and C), the transfer of a holin gene to the opposite strand (box H), and a point mutation leading to an inactivating frameshift in a tail fiber gene (box A). In contrast, quite extensive differences were detected over the early and the lysogenic conversion genes.

The A2-related subgroup of Sfi21-like *S. pyogenes* prophages differed over the nonstructural early genes and the putative lysogenic conversion genes but showed closely related struc-



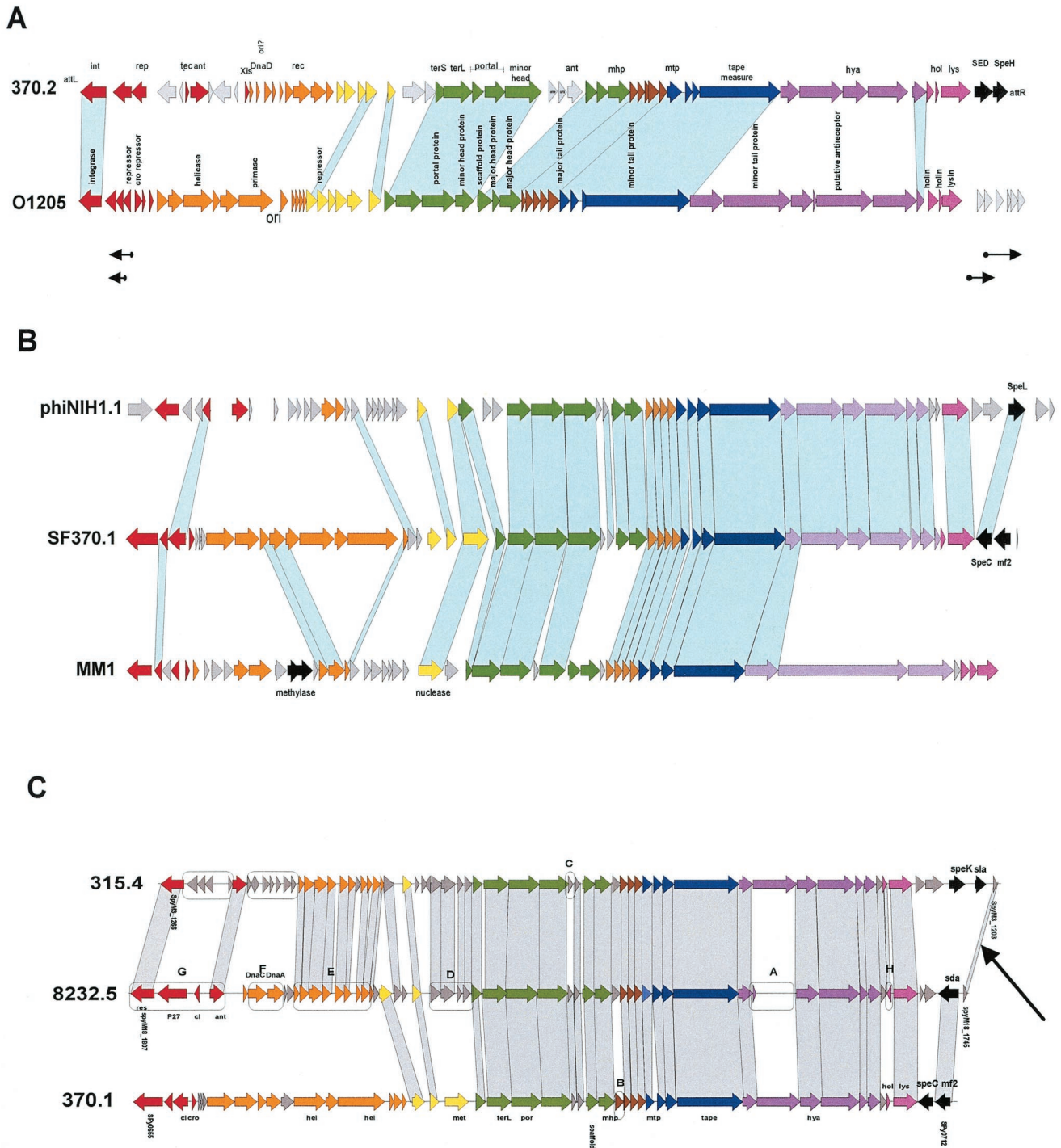


FIG. 4. Sfi11-like *S. pyogenes* prophages. (A) Alignment of *S. pyogenes* prophage 370.2 with *S. thermophilus* prophage O1205. The arrows under the map indicate phage O1205 transcripts detected in the lysogen. (B) Alignment of *S. pyogenes* prophage 370.1 with *S. pyogenes* prophage NIH1.1 and *S. pneumoniae* prophage MM1. (C) Alignment of Sfi11-like prophages from the three sequenced *S. pyogenes* genomes. For annotations, see Fig. 3.

tural genes (Fig. 5B); those of the *Staphylococcus*-related subgroup differed only for the genes near the *attR* site (Fig. 5C). A clear trend for prophage DNA loss was seen in the M1 *S. pyogenes* strain (Fig. 1A). A 13-kb prophage remnant, 370.4, encoded only lysogeny and DNA replication genes. A

closely related prophage remnant was identified at a corresponding position in the Manfredo strain. The prophages shared both flanking *att* sites but differed by internal insertions and deletions and gene replacements (37). In addition, three prophage remnants of only 2 kb were identified; they



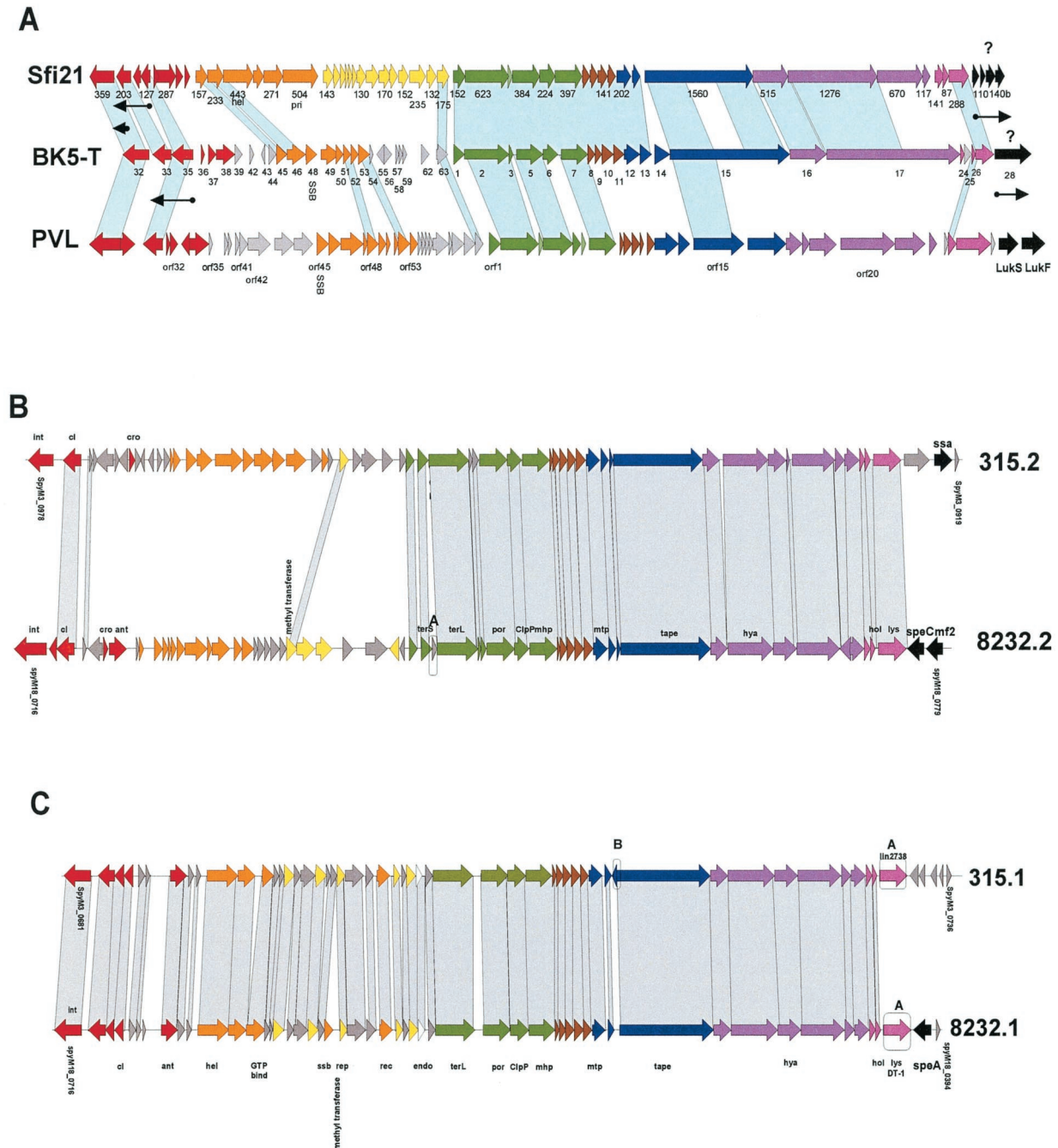


FIG. 5. Sfi21-like prophages. (A) Alignment of *S. thermophilus* prophage Sfi21, *L. lactis* prophage BK5-T, and *S. aureus* prophage PVL. Sfi21 and BK5-T genes transcribed in the lysogenic host are indicated by arrows under the gene map. Genes are annotated as in the original publications (103, 129, 134). (B and C) Alignment of the Sfi21-like *S. pyogenes* prophages of the A2-like (B) and *Staphylococcus*-like (C) subgroups.

consisted of the phage integrase accompanied by the phage repressor and a potential lysogenic conversion gene in the R-1092 remnant (37).

All prophages but one (315.1) encoded potential virulence factors between the lysin gene and the *attR* site (mitogenic factors, toxins/superantigens, enzymes) (8, 53, 68) (Fig. 3 to 5). The lysogenic conversion genes in the three prophages of the M1 strain differ in their G+C content from the surrounding prophage and bacterial DNA (68), suggest-

ing a faulty phage excision process in an unusual bacterial host with a lower G+C content as the origin of this DNA (230). The horizontal spread of these genes is also suggested by the presence of sequence-identical genes in the horse pathogen *Streptococcus equi* (37). Notably, there is a short stretch of sequence conservation adjacent to the right attachment site between different *S. pyogenes* prophages (Fig. 3B and 4C, arrows). This conserved segment and the highly conserved region around the hyaluronidase gene in the tail

fiber might allow an exchange of lysogenic conversion genes between different *S. pyogenes* prophages by homologous recombination.

The prophage-encoded hyaluronidase and DNase have been suspected of promoting bacterial spread through host tissue by their ability to hydrolyze glucosaminic bonds in hyaluronic acid, a major component of the extracellular matrix in the connective tissue, and the liquefaction of pus when degrading the DNA from decaying lymphocytes, respectively. Notably, antibodies against both phage enzymes are found in some post-streptococcal diseases (82). The virulence properties of the DNases are not entirely clear since they were also described in the literature as mitogenic factors (or streptodornases) (191). The prophage 315.4-encoded Sla protein showed phospholipase A<sub>2</sub> activity. Sla has sequence homology to a potent snake venom toxin and might contribute to inflammation and coagulopathy seen in streptococcal toxic shock syndrome (13).

Many *S. pyogenes* prophages encoded streptococcal pyrogenic exotoxins (Spe) in the lysogenic conversion region (8). However, the specific combination of toxins differed between the sequenced *S. pyogenes* strains: the M1 strain showed *speC*, *speH*, and *speI* genes; the M3 strain demonstrated *ssa*, *speK*, and *speA3*, genes; while the sequenced M18 contained the *speA1*, *speC*, *speL*, and *speM* genes (8). These are all distinct members of a large family of superantigens, and they include the scarlet fever toxin. These proteins bind the T-cell receptor and the major histocompatibility complex protein outside of the usual peptide binding site and lead to a pathological activation of the immune system, possibly allowing the escape of *S. pyogenes* from immune surveillance (170). It is conceivable that the variable combination of superantigens and mitogenic factors (*sda*, *sdn*, *mf2*, *mf3*, and *mf4*) provided by multiple prophages influences the pathogenic potential of the polylysogenic host. This is a theoretically interesting possibility to explain the strikingly distinct symptoms associated with pathogens whose bacterial genome sequences are so similar.

For example, prophage NIH1.1 was identified in an M3 *S. pyogenes* strain from a toxic shock syndrome patient. It resembled prophage SF370.1 over the entire structural gene cluster but encoded a distinct superantigen (SpeL instead of SpeC) (96) (Fig. 4B). Notably, possession of prophage NIH1.1 was a genetic marker for newly emerging M3 *S. pyogenes* strains in Japan (97), which had replaced an otherwise genetically identical strain that just lacked the prophage NIH1.1 (96). Prophage acquisition might thus be a major mechanism of short-term evolution in this epidemiologically highly dynamic and clinically variable bacterial species (8, 13). In an appealing model, the emergence of new, unusually virulent subclones of M3 strains is explained by the sequential acquisition of prophages 315.5, 315.2, and 315.4 in approximately 1920, 1940, and 1985 (13), suggesting bacterial pathogenicity evolution by prophage-mediated lateral gene transfer in the fast lane.

The possession of phage-located toxin genes does not automatically lead to the expression of these genes. Clinical isolates containing toxin genes showed a variable pattern of toxin expression when grown in broth culture (105). However, growth of these strains in mice or coculture with human pharyngeal cells led to the production of the toxins (30, 105). A small heat-stable factor released from the pharyngeal cells was identified as an inducer of the prophages (26). This is a fascinating

observation, since it means that streptococcal prophages respond via bacterial regulation systems to signals emitted from the eukaryotic host. Mobile DNA and prophages were also the most prominent group of genes that showed expression changes when mRNA from *S. pyogenes* cells grown at 29 or 37°C was assayed on an M1 strain-based microarray (190).

### *Streptococcus agalactiae* and *Streptococcus mitis*

*S. agalactiae* is a commensal organism that colonizes the gastrointestinal or genital tract of up to 40% of healthy women. However, it has substantial residual pathogenic potential for neonates and adults with underlying chronic illnesses. Two strains have been sequenced (80, 198). One strain contained two prophages, while the other showed only a large number of isolated prophage-related genes. Prophage SA1 showed a composite structure. Over the DNA-packaging, head, and tail genes, SA1 is closely related to *L. lactis* prophage r1t (differences are one indel and one gene replacement), while the tail fiber gene cluster show links to *S. pyogenes* prophage 315.6 and *S. thermophilus* prophage O1205. Near *attL*, SA1 has genes which match genes found at corresponding position in several prophages from gram-positive bacteria. Next to *attR*, a gene related to a candidate lysogenic conversion gene from a *Listeria* prophage is flanked by a mobile *S. thermophilus* DNA element.

A similar hybrid character was observed with *S. agalactiae* prophage SA2. The closest relative to the DNA-packaging, head, and tail genes from SA2 is *L. lactis* phage bIL170. This is a surprising affinity for a prophage since bIL170 belongs to the most frequently isolated group of obligate virulent lactococcal phages from the dairy environment (sk1-like or 936 species [47]) (Fig. 6), which is not known to contain temperate phages. The putative tail tape measure protein from SA1 and the putative phage adsorption protein from SA2 have low-level sequence similarity to two surface proteins, PbIB and PbIA, expressed from the *S. mitis* prophage SM1. Platelet binding is thought to be essential for the pathogenesis of infective endocarditis and was mediated in part by these prophage proteins (11). In prophage SA2, the structural genes from a virulent siphovirus apparently cooperated with nonstructural genes found only in temperate *Siphoviridae*. Interestingly, in the transition zone from nonstructural to structural SA2 genes, a XerC-like recombinase gene which might have mediated this remarkable modular exchange reaction was located.

### *Streptococcus pneumoniae*

*S. pneumoniae* is the major cause of acute bacterial pneumonia and otitis media. At the same time, it is also a transient commensal colonizing the throat and upper respiratory tract of 40% of humans. The great majority of clinical isolates carry prophages identified by hybridization experiments with phage lysin-specific probes or induction experiments (173, 181). Two types of prophages were induced: *Siphoviridae*, with a protein covalently linked to the 5' end of the 40-kb DNA genome represented by prophages HB-3 (177) and MM1 (78), and *Myoviridae*, with a 42-kb genome (55). Ground-laying work was performed with the lysins of *S. pneumoniae* phages and defined a two-domain structure consisting of a lytic domain and a binding domain to the choline part of the pneumococcal cell

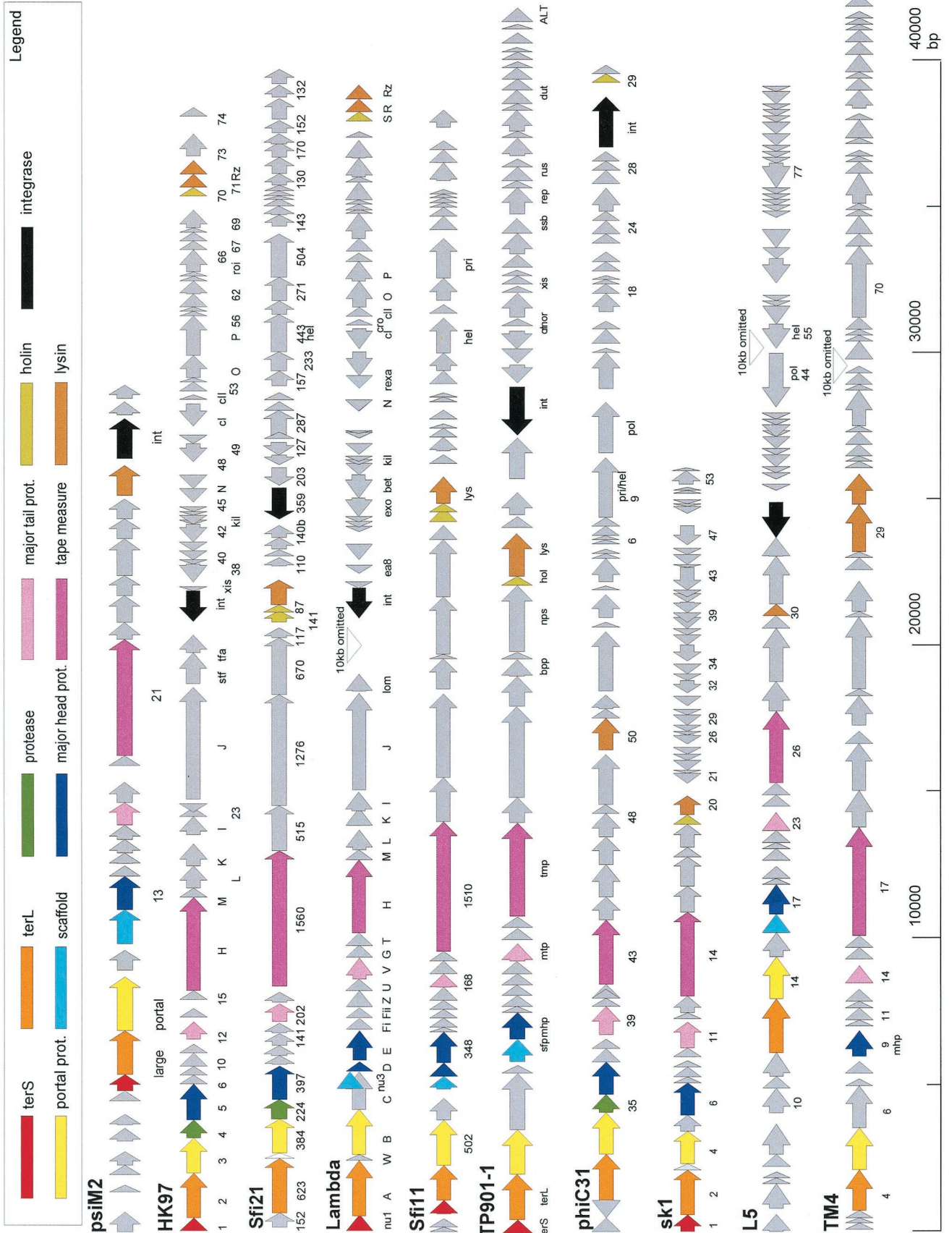


FIG. 6. Alignment of the gene maps from bacteriophages belonging to a postulated lambda-supergroup of *Siphoviridae*. The viruses represent prophages from *Archaea* (*Methanobacterium* virus psiM2),  $\gamma$ -proteobacteria (*E. coli* phages HK97 and lambda), low-G+C gram-positive bacteria (*S. thermophilus* phages Sfi21 and Sfi11; *L. lactis* phages TP901-1 and sk1) and high-G+C gram-positive bacteria (*Streptomyces* phage phiC31; *Mycobacterium* phages L5 and TM4). Virulent phages are underlined. To better visualize the similarity between the structural gene clusters from this diverse group of phages, the phage genome maps are aligned starting with the terminase genes at the left. Structural genes are identified by a color code, as indicated at the top. Selected ORFs are numbered to facilitate the orientation with the GenBank entry.

wall (74, 183). Prophage MM1 from the multiple-antibiotic-resistant epidemic *S. pneumoniae* strain Spain 23F is currently the only completely sequenced temperate phage of this species (accession number AJ302074). Its head and tail genes are closely related to those of *S. pyogenes* prophage 370.1 (Fig. 4B), while the tail fiber genes resemble *S. agalactiae* prophage SA1. MM1 is one of the few prophages from LAB lacking genes between the lysin gene and the *attR* site. A candidate lysogenic conversion region consists of two overlapping ORFs in the replication module of MM1, which encode the two subunits of a cytosine methyltransferase of a mobile DNA element. However, this DNA element is not specific to MM1, since similar genes were also found in *S. agalactiae* prophage SA2.

### *Streptococcus thermophilus*

*S. thermophilus* is naturally found in raw milk and represents a major starter bacterium in the dairy industry. Lysogeny is not widespread in this species (29). According to the mode of DNA packaging, two groups of temperate *S. thermophilus* phages have been characterized (122), represented by the *pac*-site Siphovirus O1205 (192) (Fig. 4A) and the *cos*-site Siphovirus Sfi21 (129) (Fig. 5A). Temperate and virulent *S. thermophilus* phages showed a peculiarly close genetic relationship. Virulent phages which are the predominant ecological isolates from both the factory and raw milk (31) are essentially the result of deletion, gene replacement and rearrangement events in the lysogeny module of temperate phages (127). In silico analysis demonstrated that prophages related to the two basic types of *S. thermophilus* prophages are found in many low-G+C gram-positive bacteria (Fig. 4A and 5A). Comparative genomics revealed distant relationships to lambdaoid phages from gram-negative bacteria and even prophages from *Archaea* (51) (Fig. 6). In fact, over the structural gene cluster, the Sfi21-like phages shared a gene map with *E. coli* phage HK097. Sfi11 phage even showed protein sequence similarity to phage lambda, suggesting distant phylogenetic relationships between these phages (28). No protein sequence similarity linked HK097 and Sfi21 prophages. However, some features characteristic for this group of phages were identified in both: the major head protein was proteolytically cleaved at amino acid 104 and 105, respectively, releasing an N-terminal protein fragment with strong coiled-coil structure (51). In both prophages, a protease gene precedes the major head gene. In phage Sfi21, this protease belongs to the ClpP protein family, as in many other Sfi21-like phages from LAB and even prophages from  $\gamma$ -proteobacteria (54). This specific gene constellation is a diagnostic criterion for Sfi21-like phages. Sfi11-like phages showed a distinct head gene constellation consisting of three phage head genes and one scaffold gene.

Sfi21 belongs to the few temperate phages from LAB for which a transcription map was established both in the lytic mode of infection (210) and in the prophage state (209) (Fig. 5A). In the lytic mode, essentially the entire Sfi21 genome was transcribed, allowing a distinction of early (transcription regulation module), middle (DNA replication), and late (structural and lysis genes) transcripts. In the lysogenic state, only two Sfi21 genome regions were transcribed from the otherwise transcriptionally silent prophage (209). One transcript com-

prised the DNA segment from the *cI*-like repressor (34) to the superinfection exclusion (*sie*) genes located directly upstream of the phage integrase (32). The cloned *cI* repressor protected a cell against superinfection with temperate phages (34), while the cloned *sie* gene conferred protection against many virulent phages (32). Another transcript covered four genes located between the lysin gene and the *attR* site (209). These genes lacked database matches, preventing speculations about their possible functions. *S. thermophilus* prophage O1205 carries a different set of genes near *attR*, and they also belong to the few genes transcribed from the prophage (209) (Fig. 4A). A lysogenic conversion phenotype was observed for a *S. thermophilus* strain lysogenic with the prophage TP-J34: it showed distinct growth properties (planktonic versus aggregated growth) when lysogenic or when prophage cured (H. Neve, personal communication). TP-J34 displayed a distinct set of genes between the lysin gene and the *attP* site (158). A database search revealed that many temperate phages from low-G+C gram-positive bacteria showed extra genes between the phage lysin and *attR* (209). With the exception of a *Bacillus halodurans* prophage (see below), these prophage genes from free-living bacteria showed an informative database match, precluding any speculation with respect to their function (209). In accordance with theoretical predictions, a prophage remnant consisting of the phage integrase and a few transcribed phage genes was described for *S. thermophilus* (209).

### *Lactococcus lactis*

*L. lactis* is the closest phylogenetic relative of the genus *Streptococcus* and is the major starter used in the cheese industry. Due to the economical impact of phage infections, lactococci and their phages became a focus of research in dairy microbiology. The completely sequenced *L. lactis* strain IL1403 contained six prophage elements (42). Two inducible and one noninducible prophages showed the genome organization of *cos*-site temperate *Siphoviridae* closely related to *S. thermophilus* phages Sfi21. Three 15-kb prophage remnants had maintained only lysogeny genes (integrase and repressor) and, in variable amounts, DNA replication and a few structural genes (42). In contrast to our interpretation, these authors viewed them as P4-like satellite phages.

The Sfi21-like lactococcal prophages are represented by prophage BK5-T (Fig. 5A). Over the structural gene cluster, BK5-T showed an interesting gradient of sequence similarity covering high and low DNA identity to prophages bIL286 (*Lactococcus*) and Sfi21 (*Streptococcus*) or moderate or low protein sequence identity to phages *adh* (*Lactobacillus*) and PVL (*Staphylococcus*). Since this gradient of prophage relatedness reflects the phylogenetic relationship of their host bacteria, a coevolution of prophages with their bacterial hosts was initially discussed (52). However, further analysis also demonstrated substantial sequence diversification within prophages from a single bacterial species (*L. lactis*), including DNA sequence (bIL286), protein sequence (bIL309), or only genome organization similarity to BK5-T (bIL285), creating a problem for phage taxonomy and models of phage evolution (171). Transcription in the BK5-T prophage was limited to two regions: three transcripts covered the phage integrase, the *sie* homologue and the *cI* repressor, and another transcript

was derived from an anonymous large gene located between the phage lysin and the *attR* site (21) (Fig. 5A).

The best-characterized Sfi11-like prophage in *L. lactis* is prophage TP901-1 (24, 25) (Fig. 6). The structural proteins from TP901-1 have been characterized by protein sequencing (100), immunoelectron microscopy (100), and mutational analysis (165). The results confirmed that the prediction of gene functions by comparative genomics, and specifically the alignment of the structural gene map with phage lambda, is quite reliable (28). For example, as in lambda, the length of the tail structure is determined in TP901-1 by the length of the tail tape measure protein (165). Also, the prediction of a transcriptional regulation module between the DNA replication module and the structural gene module in prophages from LAB was confirmed experimentally (24). In fact, many of the *in silico* predictions of gene assignments in phages from LAB were confirmed by experiments with one or the other phage from LAB, instilling some confidence in the power of comparative phage genomics. Indeed, in some cases the experiments were actually guided by comparative genomics. The genome analysis of *S. thermophilus* phages differing in host range provided keys to the location of the phage antireceptor on the genome map and suggested a mechanism of diversification by the exchange of highly variable gene segments flanked by conserved gene segments encoding collagen-like peptides (127, 200). The model was subsequently confirmed by the construction of chimeric phages with *S. thermophilus* phage DT1 (61). However, two genes occupied different genome positions in dairy phages from their positions in phage lambda. In prophages from low-G+C gram-positive bacteria, the lysis cassette is invariably located downstream of the tail fiber genes, in contrast to lambda, where they are found upstream of the DNA packaging genes. Second, the excisionase from TP901-1 was identified (23) within the early genes downstream of the *cro*-like repressor gene (131). In lambda, the *xis* gene is found directly upstream of the phage integrase gene. This position is occupied in lactococcal and streptococcal prophages by the *sie* gene (140).

The third class of lactococcal prophages is represented by phage r1t (206). Not only did its genetic switch region show a comparable structure to that of phage lambda (155), but also molecular modeling of its repressor on the basis of the lambda *cI*-repressor allowed the design of a thermolabile repressor mutant as a genetic tool (154). The functions of two predicted DNA replication genes were confirmed by biochemical experiments, including a replisome organizer (235) and a RusA protein. The latter is an endonuclease that resolves Holliday junction intermediates formed during DNA replication, recombination, and repair (182). Interestingly, the RusA protein of *E. coli* is also encoded by the defective prophage DLP12 and *rusA*-like sequences are associated with prophage sequences in several bacteria. Since phages from LAB dedicate more genes from their genome to DNA replication functions than the similarly organized phage lambda does, one might ask to what extent some of these genes are of potential use to the bacterial host. Such dual functions could also explain why prophage remnants in LAB demonstrated a trend to maintain genes from the lysogeny and DNA replication modules. As in *S. thermophilus* phages, closely related virulent derivatives of temperate r1t-like phages were described. Their lysogeny mod-

ule consisted only of the genetic switch structure, while the phage integrase has been eliminated (133).

Many lactococcal phages, including r1t, contain introns at various genome positions, demonstrating that selfish DNA elements such as prophages can also become the target for parasitic DNA elements. Intron homing is the process by which introns spread through a population of intronless alleles and is initiated by intron-encoded endonucleases. In dairy phages, these endonucleases are found relatively frequently (47, 71). The process of intron homing can be very efficient: an ecological survey in *S. thermophilus* phages revealed that all phage lysin genes possessing a 14-bp consensus sequence contained an intron. As with the prophage DNA, one would expect a selection pressure to remove the intron or to prevent its further spread. Indeed, large deletions within the homing endonuclease were detected in *S. thermophilus* phages (71).

### *Lactobacillus*

The use of *Lactobacillus*, another LAB, in various food fermentation processes and as probiotic (health-promoting bacteria) has motivated research into their phages and prophages. *L. delbrueckii* prophage mv4, for which closely related virulent phages were also described (142, 208), became the focus for research into the site specificity of the integration system (4, 5). *Lactobacillus gasseri* phage A2, in comparison, is the best-characterized LAB phage with respect to its DNA-packaging mechanism (75). Also, the genetic switch structure of A2 was studied in more detail than in any other phage from LAB: three operators located between divergently transcribed repressor genes were bound with different affinities by the two repressors, resulting in a repositioning of the RNA polymerase (76, 110, 111).

When corresponding genome segments were studied in different phages from LAB, substantial biological variability was frequently observed. The genetic switch region can serve as an example. *Lactobacillus casei* phage A2 still follows the phage lambda paradigm relatively closely. Substantial deviations from this theme were found in other phages from LAB; e.g., *Lactobacillus plantarum* phage phig1e showed seven 15-bp operators with dyad symmetry in this region, which were bound differentially by the repressors encoded by the flanking genes (102); in the *Lactococcus* phage BK5-T, the divergently transcribed repressor genes are separated by one ORF which is normally found further downstream of the early lytic transcript (134); and in the lactococcal phage TP901-1 and the streptococcal phage Sfi21, the lytic (Cro) repressor lacked binding activity for the DNA of the genetic switch region and inhibited the lysogeny (*cI*) repressor binding to the genetic switch region possibly by protein-protein interaction between the two repressors (34, 132).

The holin-lysin system provides another example. The similarity with the phage lambda holin S and lysin R gene constellation was demonstrated in experiments where *Lactobacillus* phage holin (162) and lysin (19) could complement lambda prophages containing mutations in both genes. However, *S. thermophilus* phages showed two holin genes with distinct biological properties, suggesting a holin-antiholin system in the control of the lytic process (184). Apparently, there are many different solutions to a given problem for phages with a com-

mon overall genome organization. This is not a peculiar situation in phages from LAB; similar observations were made with lambdoid phages (218).

Most sequenced *Lactobacillus* species contain prophage sequences. The 2-Mb chromosome of the gut commensal *Lactobacillus johnsonii* NCC 533 (Nestlé) contained two prophages showing the genome organization of Sfi11-like *pac*-site *Siphoviridae* (54). The lysogeny module of these prophages contained more genes than are commonly found in temperate phages of LAB (128). In one prophage, two of these extra genes showed links to a genomic island from *S. aureus*. Northern blot analysis revealed that these genes are transcribed in the lysogen. Microarray analysis demonstrated that the two prophages Lj928 and Lj965 represented quantitatively the majority of the strain-specific DNA of the sequenced *L. johnsonii* strain. Another *L. johnsonii* prophage, Lj771, had extensive DNA sequence identity to a prophage in the sequenced *L. gasseri* strain (Joint Genome Institute). Differences over the late genes were limited to few genome regions (lysin and anti-receptor) but were more extensive over the early genes.

The sequenced *L. plantarum* strain WCFS1 (106) contained two closely related Sfi11-like prophages that had a nearly identical structural gene cluster. One prophage contained a disruptive mutation in the terminase gene. Candidate lysogenic conversion genes were identified by database searches and transcription analysis near both the *attL* and *attR* sites. The extra genes shared similarity to a mitogenic factor encoded by an *S. pyogenes* prophage. This observation is notable since the sequenced *L. plantarum* strain was isolated from the oral cavity of a human, which is also the habitat of *S. pyogenes*. A prophage remnant consisted of truncated lysogeny, DNA replication, and a few structural genes typical for an Sfi21-like phage. It abutted directly one of the Sfi11-like prophages.

### *Listeria*

Although most if not all *Listeria* strains carry functional or cryptic prophages, the potential influence of lysogeny on the host phenotype is unknown. Only one *Listeria* prophage has been investigated in some molecular detail: A118 belongs to Sfi11-like *Siphoviridae*, but lacks a *pac*-site (125). The prophage integrates into *comK*, a putative transcriptional activator for various factors involved in competence for DNA uptake. However, *Listeria* is not easily transformable, and so a negative lysogenic conversion phenotype is not immediately obvious. A closely related prophage, EGDE, was identified in the sequenced *Listeria monocytogenes* strain (79). Over the structural gene cluster, differences from A118 were limited to the major head gene. In view of the intricate protein-protein interactions which occur during phage morphogenesis, it is surprising that a single protein can be exchanged without upsetting the other phage proteins participating in the head-building process. More substantial differences were detected over the nonstructural genes including the lysogeny module, which might explain why A118 can be propagated on a strain containing the EGDE prophage. From the *Listeria* strain ScottA, isolated during a large listeriosis epidemic in the United States, an Sfi21-like prophage, PSA, was induced and sequenced (accession number AJ312240). Like all sequenced *Listeria* prophages, PSA

contained a cluster of genes without database matches near the *attR* site. Parts of these genes were shared between different *Listeria* prophages.

*Listeria* is ubiquitous in nature; it can be found in soil and the gut, and it represents an opportunistic pathogen in animals and to a lesser extent in humans. *L. monocytogenes*, the etiological agent of listeriosis, a severe food-borne disease, and the nonpathogenic species *L. innocua* shared a closely related genome and an unexpected synteny with *B. subtilis* and *S. aureus* (79). Remarkably, all major gaps in the alignment of the two bacterial genomes were represented by the prophages integrated into *L. innocua*. Except for prophage genes, less than 10 and 5% of the genes were *L. monocytogenes* and *L. innocua* specific, respectively. *L. innocua* contains five prophages; only A118-like prophage 1 is shared with *L. monocytogenes*, but the two prophages are integrated into two different chromosome locations. Over the structural genes, prophages 2, 3, and 5 resembled *B. subtilis* prophage PBSX, *Xylella* prophage XfP3, and *Lactococcus* prophage bIL285 (171), respectively. The closest relative of prophage 4 was the *L. monocytogenes* prophage EGD, with which it had low to moderate sequence similarity in a patchwise fashion.

### *Staphylococcus aureus*

Staphylococcal enterotoxins cause an acute food-poisoning syndrome that is the second most frequent food-borne disease in the United States. Like botulism, the illness results from ingesting preformed bacterial toxins. The gene for enterotoxin A is carried by several staphylococcal prophages near their attachment sites (14). In addition, *S. aureus* causes a range of diseases from skin infections to life-threatening conditions such as sepsis. The organism produces many toxins and is highly efficient at overcoming antibiotics. A number of prophages have been found in clinical isolates. Their sequencing revealed the carriage of several toxin genes. Prophage PVL, a typical Sfi21-like siphovirus (Fig. 5A), encoded the clinically important bicomponent cytotoxin leukocidin S and F between the phage lysin and the *attR* site (103). Leukocidin is an established staphylococcal virulence factor, which causes leukocytolysis and tissue necrosis. The same toxin was found on prophage SLT, showing a distinct morphology (an elongated instead of icosahedral head as in PVL), suggesting horizontal transmission of toxin genes between temperate phages (153) (Fig. 7). Despite its distinct head morphology, SLT also showed the genome organization of an Sfi21-like siphovirus, with the characteristic gene constellation portal protein-ClpP protease-major head gene (identical to the prophage phi12 head protein, see below). The noninducible prophage PV83 shared with PVL the entire structural gene cluster (>86% amino acid identity) but showed a variant leukocidin (LukM/ LukF pore-forming complex) next to a distinct lysis cassette. The defective nature of this prophage might be linked to the incorporation of a transposase-containing insertion sequence into a head-to-tail joining gene of PV83. A second insertion element is found near the *attR* site of PV83 (234).

Exfoliative toxin is one of the extracellular staphylococcal proteins causing blistering skin disease. The exfoliative toxin A is encoded downstream of the lysin gene in prophage ETA (225) (Fig. 7). This prophage showed the genome structure of

an Sfi11-like siphovirus, with many protein sequence links to the phages described in the preceding sections. Comparison with prophage SLT identified a possibly inserted group of genes between the tail fiber and lysis genes (Fig. 7). These genes encode a cell hydrolase and a protein related to a collagen-like surface protein, a virulence factor in *S. pyogenes*, thus representing further candidate lysogenic conversion genes.

The genome sequence from the methicillin-resistant strain N315 and the vancomycin-resistant strain Mu50, isolated 15 years apart from Japanese patients, were closely related (99% at the nucleotide level); most of the differences were due to the insertion of Mu50-specific DNA elements (109). Both strains had related prophages phiN315 and phiMu50A integrated at the same chromosomal locus (beta-hemolysin) next to the pathogenicity island SaPI<sub>n</sub>1 (Fig. 1B). The two prophages belonged to the Sfi21-like *Siphoviridae* and had sequence similarity to prophage PVL over large parts of the genome (Fig. 7). However, the DNA-packaging, head, and tail genes belonged to different modules, showing some sequence similarity to *Listeria* prophage PSA. Differences between phiN315 and phiMu50A included several gene replacements in the lysogeny module (e.g., a sugar transferase), a larger replacement in the putative transcription regulation module, and a single-gene indel near the *attR* site, providing an additional truncated lysin gene in phiN315. Both prophages contained several candidate lysogenic conversion genes (encoding enterotoxin P, staphylokinase, and the M-like protein fragment) near but not in the direct vicinity of the *attR* site. The vancomycin-resistant strain contains an additional prophage phiMu50B that shares with prophage ETA the integrase and integration site and sequence similarity over part of the early, tail fiber, and lysis genes (Fig. 7). PhiMu50B is a close relative of prophage phi11 from *S. aureus* strain 8325, with which it could be aligned at the DNA level over nearly the entire genome length including the head gene cluster, defining a second allele of structural genes in Sfi11-like *S. aureus* prophages (Fig. 8). PhiMu50B contained candidate lysogenic conversion genes in the vicinity of the genetic switch region (two genes with links to the pathogenicity island SaPI<sub>n</sub>1) and genes downstream of the phage lysin (one showed a match with a *S. pyogenes* prophage gene located next to a superantigen or toxin) (Fig. 7).

In contrast to N315, which was isolated from a hospital infection, MW2 is a community-acquired methicillin-resistant *S. aureus* strain which is otherwise susceptible to many antibiotic classes. This strain had 95% identity to N315 and Mu50 at the nucleotide level. MW2 contains two prophages: phiSa3 and phiSa2. The first is found at a position occupied by prophages in four of the five currently sequenced *S. aureus* strains (Fig. 1B). Comparison of the corresponding prophage maps revealed patchwise relatedness. This mosaic structure was interpreted as evidence for multiple crossovers between the phages (7). PhiSa3 encodes two new enterotoxins in the vicinity of *attL* (enterotoxin G and K homologues, nearly identical to the corresponding genes in the SaPI<sub>n</sub>3 pathogenicity island [226]) and the *sea* toxin gene located between the tail fiber and lysis module. PhiSa3 differed from the prophage PVL essentially only in the associated virulence factors (Fig. 8). PhiSa2 has DNA sequence similarity to prophage phi12 essentially over the entire genome (Fig. 8). Differences included a few indels and some gene replacements. Notable was the possession of

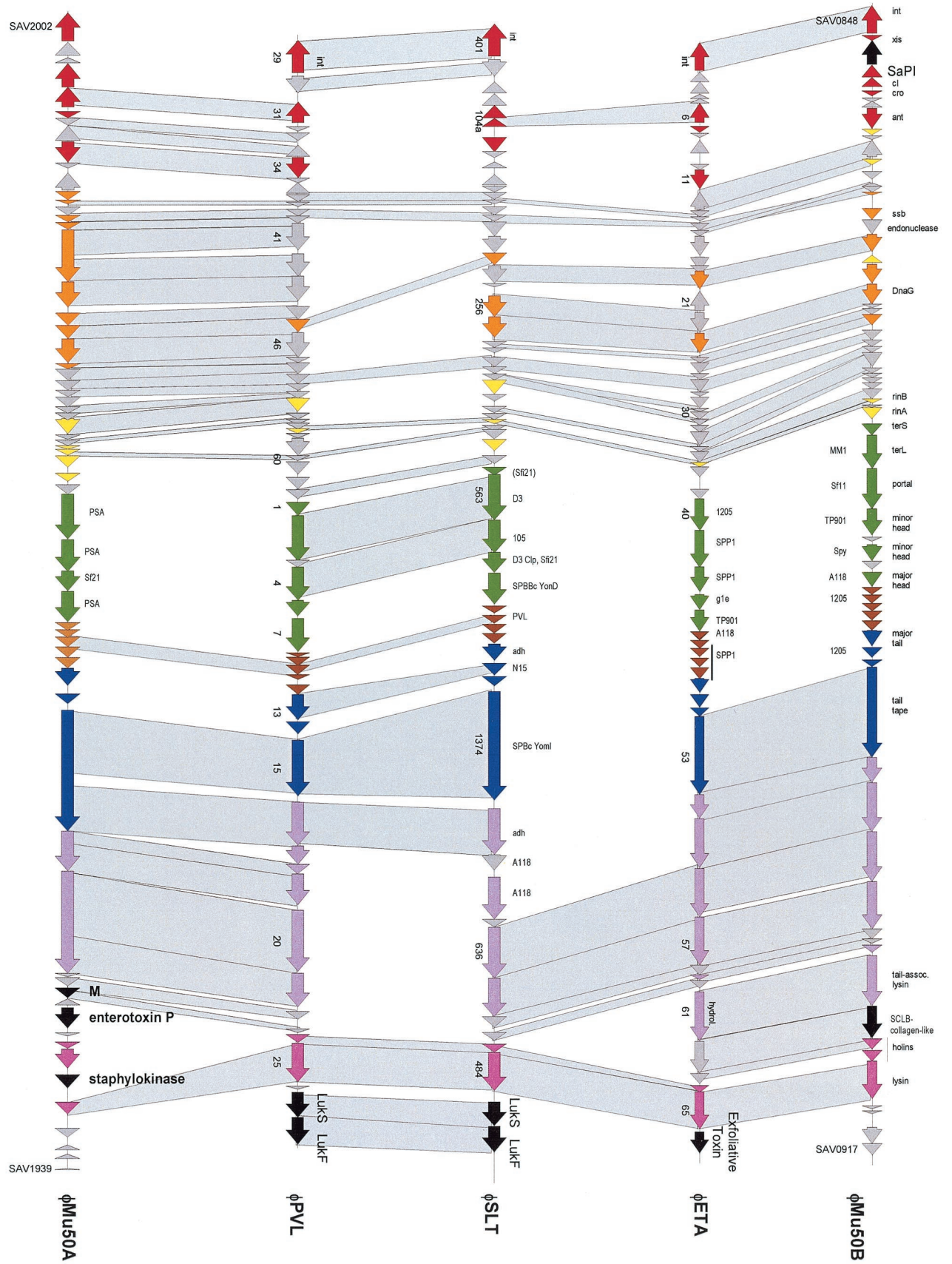


FIG. 7. Alignment of the five major types of *S. aureus* prophages, phiMu50A and phiMu50B are prophages extracted from the sequenced Mu50 *S. aureus* strain, while phages ETA, SLT, and PVL were induced from unsequenced *S. aureus* strains. The phage genes are color coded as in Fig. 3. Sequence-related genes between the prophages are linked by shading. Sequence matches of structural genes to those in other than staphylococcal phages are indicated (*Bacillus*: 105, SPP1, SPBc; *Listeria*, PSA, A118; *Lactobacillus*, adh, g1e; *Lactococcus*, TP901-1; *Streptococcus*, Sfi21, Sfi11, 1205, MM1 Spv; *E. coli*, N15; *Pseudomonas*, D3). Selected ORFs are numbered to facilitate the orientation with the GenBank entry.



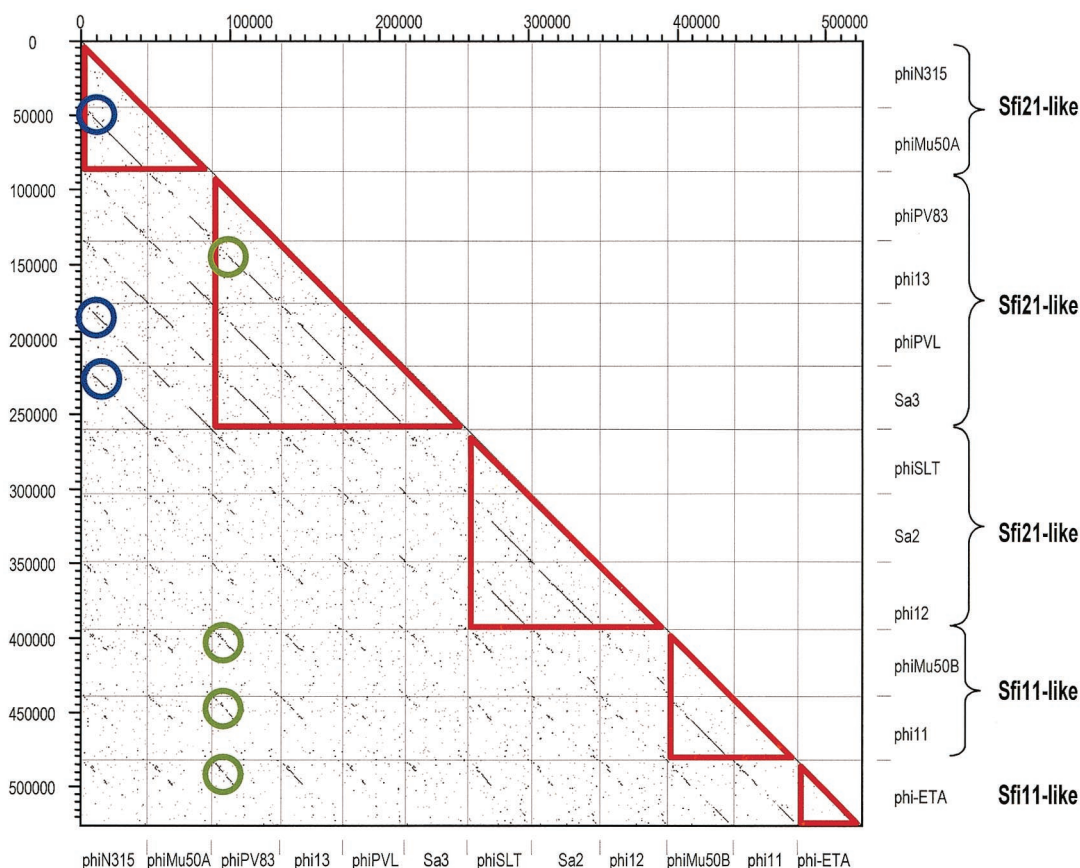


FIG. 8. Dot plot matrix for the currently available 12 *S. aureus* prophages. The prophages are identified by their names on the x and y axes. According to the structural genes, the prophages were classified into five distinct groups (red triangles) and annotated on the left ordinate. Two groups of phages sharing relatively highly conserved early genes are marked by blue and green circles.

the *lukF* and *lukS* genes between the lysin gene and *attR* in phiSa2, where phi12 lacked ORFs.

Strain 8325 was used for the construction of the first physical maps of *S. aureus*. It harbors three prophages, phi11, phi12, and phi13 (95) (Fig. 1B). phi11 and phi13 have been studied in some detail. phi11 DNA is 5% terminally redundant and 40% circularly permuted (126). phi11 is one of the few prophages from low-G+C gram-positive bacteria that showed the *attP-int-xis* gene constellation familiar from phage lambda (227, 228). This is, however, not the common situation even in staphylococcal phages (38). phi13 was the first staphylococcal phage associated with positive (staphylokinase) and negative (beta-toxin) phage conversion (222). The negative phage conversion occurred because phi13 integrated into the beta-toxin, leading to gene inactivation (46). This is not an isolated case. *S. aureus* phage L54a integration confers a lipase-negative phenotype due to insertional inactivation of a lipase gene (119). The positive phage conversion is conferred by the staphylokinase gene located between the phi13 lysin gene and *attR*.

A dot plot matrix of the available *S. aureus* prophages demonstrated five distinct groups of structural modules (Fig. 8). Three distinct groups of Sfi21-like *cos*-site *Siphoviridae* were identified: PVL-PV83-phi13-phiSa3 comprise the first group, the second is represented by SLT-phiSa2-phi12, while the third group is provided by phiMu50A-phiN315. In addition, two

different Sfi11-like *pac*-site *Siphoviridae* were revealed by the dot plot: prophages phiMu50B-phi11 on one side and phiETA on the other side. With respect to the early genes, the distinction of DNA homology groups was less obvious. Two loosely defined groups could be distinguished (Mu50B-PVL-Sa3 vs all the others), but an extensive mosaicism prevented a sharper distinction of modules (Fig. 8).

In contrast to the prophage-containing *S. aureus* strains, the sequenced *Staphylococcus epidermidis* strain ATCC 12228 (accession number AE015929) lacked prophage sequences.

### Bacillus

Comprehensive research was conducted on two virulent phages of the soil bacterium *Bacillus subtilis*: phi29, a podovirus, and SPP1, despite its life-style a typical Sfi11-like siphovirus (54) and by far the best-characterized phage of this proposed phage genus (see references 6 and 130 for recent examples and references therein). Much less is known about temperate *B. subtilis* phages, which have been classified into five groups (231). Only three groups are represented by sequenced prophages. The group I phage phi105 shows the genome organization of a typical Sfi21-like siphovirus (52) and was investigated mainly for repressor binding to the genetic switch region (205). The group III phage SPBc2 represents a

134-kb siphovirus consisting of 187 predicted ORFs, 70% of which lacked matches to the database (118). A mere 14 ORFs shared links with other phages. In contrast, about 30 ORFs had links with bacterial genes, mostly from *B. subtilis*. According to the orientation of the ORFs, three clusters could be distinguished. Cluster I contains the integrase/recombinase. Cluster II starts with the lysis cassette and continues with the structural gene module. The tail fiber genes had up to 50% amino acid sequence similarity to proteins from defective *B. subtilis* prophages. Cluster III contained genes involved in transcriptional regulation, DNA replication, and nucleotide metabolism.

Group V prophages are represented in the sequenced *B. subtilis* strain by the defective prophages PBSX and skin. Upon UV or mitomycin C induction, the cell releases phage-like particles consisting of small heads and large, complex tails that adsorb to and kill related bacilli acting like bacteriocins. The head contains randomly selected 13-kb fragments of the bacterial chromosome. In that respect, PBSX resembles a small bacteriophage-like particle discovered in the purple non-sulfur bacterium *Rhodobacter capsulatus*, which transfers random 4.5-kb segments of the genome of the producing cell to recipient cells, where allelic replacements occur. This particle was called a gene transfer agent, resulting in a genetic exchange process controlled by the bacterial cell (113). However, the DNA packaged into the PBSX head is not injected into the cell (223). The widespread occurrence of the PBSX-like defective phages throughout the *Bacillus* species and the failure to isolate strains cured of PBSX nevertheless suggested that their continued maintenance is advantageous, if not essential, for the host strain (223). The 28-kb PBSX prophage remnant consists of a shortened lysogeny and DNA replication module and a structural gene cluster whose organization resembles that of the Sfi11-like *pac*-site *Siphoviridae* (Fig. 9). In comparison with the standard genome map of Sfi11-like phages, PBSX lacks a large head protein gene normally located between the portal gene and the scaffold gene. In addition, there are fewer head-to-tail joining genes than usual, possibly explaining the small head morphology. The siphovirus-like tail fiber genes are followed by sequence links to putative tail genes from the myovirus prophage SPBc2 and end in a lysis cassette consisting of a holin and an amidase-type lysin gene (Fig. 9). During sporulation, the ca. 50-kb skin prophage element is excised from the *B. subtilis* chromosome by a DNA rearrangement event (197). The prophage remnant contains a seemingly complete set of structural genes characteristic of Sfi11-like *pac*-site *Siphoviridae*. Over these genes, it had sequence similarity to many structural genes from the PBSX prophage and *Listeria innocua* prophage 2 (79). The structural gene cluster is preceded by a DNA replication module. The lysogeny region is reduced to the genetic switch structure, while an integrase was not detected. An integrase was found downstream of the prophage lysis cassette, separated by a group of bacterial genes including an arsenic resistance operon. In contrast to PBSX, the skin element contains no genes essential for *B. subtilis* viability.

The sequencing of *B. halodurans*, an industrial source of enzymes used under alkaline pH, revealed 112 ORFs encoding transposases or recombinases, suggesting an important role of these enzymes in horizontal gene transfer in this species; however, no prophage was reported (196). A reanalysis of the

sequence revealed a complete prophage, showing the typical genome organization of an Sfi11-like siphovirus with sequence matches over the head and tail genes to *S. pyogenes* prophage 315.5. As in a number of other prophages, an isolated adenine methyltransferase gene was detected between the DNA replication module and the DNA-packaging module. More interestingly, however, was the presence of a type II restriction endonuclease and an associated cytosine-specific methyltransferase located between the phage lysin and the *attR* site. Possession of the prophage thus confers a potentially new restriction modification system to the lysogenic cell.

### *Clostridium*

The spore-forming clostridia are widely disseminated in soil and lakes but are also found in the intestinal flora. *Clostridium botulinum* is defined as any clostridial isolate that produces botulinum toxin, which causes an often fatal form of food poisoning. Biological experiments conducted 30 years ago established that lysogenization by some bacteriophages with contractile tails converted nontoxigenic into toxigenic isolates. Curing of the prophage leads to concomitant loss of the toxigenicity (66, 67). However, the first temperate *Clostridium* phage was sequenced only recently (233). Despite its relatively small genome size of 33.5 kb, the *Clostridium perfringens* phage phi3626 showed the typical genome organization of Sfi21-like *Siphoviridae* (Fig. 9). The presence of two genes related to sporulation-dependent transcription factors in the early-gene cluster suggests that this prophage also has a potential involvement with sporulation. *C. perfringens* is present in different pathotypes; strains producing the enterotoxin CPE are an important cause of food poisoning and recently also antibiotic associated diarrhea, while the histotoxic clostridia produce exotoxins that are implicated in gas gangrene. The histotoxic *C. perfringens* strain (185) contains a prophage with a variable extent of protein sequence identity, ranging from 25 to 80% to phi3626, essentially over the entire structural gene cluster. The structural module was flanked on one side by a lysis cassette. On the other side, the similarity to the genome organization of typical prophages from low-G+C gram-positive bacteria was less obvious: only a XerD/C-like recombinase and three potential transcriptional regulators could be identified. No phage-encoded candidate virulence factors could be identified by in silico analysis. The sequenced *Clostridium acetobutylicum* strain used in industrial fermentation (159) contained two prophages. Prophage 1 showed the structural gene cluster of an Sfi21-like phage. Sequence similarity to phi3626 was weak and was limited to five structural genes, while similarity to *S. aureus* prophages was more prominent. The structural gene cluster was preceded by 50 mostly very small ORFs that lacked links to phage genes except for three DNA replication and two repressor genes. This region ends with two closely spaced resolvase genes. The diagnosis of prophage 2 is less well backed by database matches. In fact, it showed an ORF organization reminiscent of the structural gene cluster from temperate dairy phages, but sequence similarities were limited to the tail tape measure protein and three ORFs sharing weak amino acid identity to *B. subtilis* phage SPBc2. Upstream of these genes, 60 ORFs were localized which lacked database matches except



for two DNA helicases, a phage lysin, and, again, three weak links to SPBc2.

*Clostridium tetani* strain E88 contains three prophages: phiCT3 is a hybrid combining PBSX-like tail and tail fiber genes with phage 3626-like DNA-packaging and Sfi21-like head genes (Fig. 9). The closest relatives of the DNA replication genes were found in *Listeria* phage A118.

PhiCT1 showed a gene map typical of Sfi11-like *Siphoviridae*, and some database matches backed this attribution. The structural genes were flanked by genes whose closest relatives were found in insect viruses (up to 39% amino acid identity with *Chilo iridescent* virus). Similar close relationships between phage and insect virus proteins were also reported for several dairy phages (128). SPBc2-like and A118-like integrase genes were detected upstream of the structural genes from phiCT2, suggesting that the putative early genes containing a ferritin-like gene resulted from a recombination event. Also, phiCT2 demonstrated a relatively typical structural gene cluster backed by numerous, but diverse phage links. The structural module is preceded by gene fragments of a resolvase and a transposase followed further upstream by another integrase gene. An intervening gene showed sequence relatedness to toxin A from *Clostridium difficile*.

#### PROPHAGES FROM HIGH-G+C GRAM-POSITIVE BACTERIA

The high-G+C gram-positive bacterial group, also called *Actinomycetales*, is characterized by having G+C contents above 60% and has traditionally been considered a sister group of the low-G+C gram-positive bacteria. It contains soil bacteria of substantial industrial interest (*Streptomyces*), gut commensals (*Bifidobacterium*), and important human pathogens (*Mycobacterium*), to name only the completely sequenced representatives.

##### *Corynebacterium*

The interest in prophages of corynebacteria stems from early work demonstrating that the diphtheria toxin is encoded by corynephage beta, morphologically a siphovirus with a 35-kb genome containing cohesive ends (17). Physical maps of the phage genome demonstrated that the toxin is encoded near the attachment site of the phage. The immunity genes of the phage are located on the opposite side of the attachment site (112). The two sequenced corynebacteria, namely, *Corynebacterium efficiens* and *Corynebacterium glutamicum*, lack prophages. Genomic information on corynephages must therefore await the completion of *Corynebacterium diphtheriae* genome sequencing.

##### *Streptomyces*

The sequenced *Streptomyces coelicolor* genome contains numerous mobile genetic elements; six of them share genes with the integrative plasmid pSAM2, while others contain transposases but no clear-cut prophage (12). However, tools for genetic engineering of *Streptomyces* have been developed from the temperate phage phiC31. This 41-kb siphovirus closely resembles Sfi21-like prophages with respect to the organiza-

tion of the structural genes (188). In contrast, the early genes, especially the lysogeny genes, showed a definitively distinct organization (221) (Fig. 6).

##### *Bifidobacterium*

The *Bifidobacterium longum* genome (180) contains a single approximately 15-kb prophage remnant identified by several phage structural (terminase, portal, head protease, minor tail, but no head gene), lysis, and integration genes. Another sequenced *B. longum* strain (Joint Genome Institute) showed a different prophage remnant of about 15 kb, consisting of a relatively complete structural module.

##### *Mycobacterium*

The two sequenced *Mycobacterium tuberculosis* strains, but not *Mycobacterium leprae*, contain two prophage remnants, phiRv1 and phiRV2 (89). Rv1 is a repetitive element which occurs in seven copies in the genome, but only one copy contains the prophage (the location differs between the two sequenced strains). The 10-kb Rv1 consists of three head genes (terminase, prohead protease, and capsid), a primase, and an integrase. A similarly organized second prophage remnant, phiRv2, which differs from phiRv1 mainly by the additional insertion of an IS element, was integrated into a tRNA gene. The phiRv1 element encodes an active-site-specific recombination system where an integrase of the serine recombinase family catalyzes integration and excision and an adjacent small ORF controls the directionality of recombination. Therefore, phiRv1 is probably a mobile DNA element (16). The avirulent strains of *Mycobacterium bovis* BCG have no prophage elements, while all clinical isolates of *M. tuberculosis* have at least one copy of either phiRv1 or phiRv2, suggesting that it could play a role in the physiology of *M. tuberculosis*, e.g., by conferring the ability to form virus-like particles capable of generalized transduction (113).

Much more is known about inducible prophages from non-sequenced mycobacteria strains. The molecular biology of these temperate phages has been developed to obtain genetic tools for the manipulation of pathogenic mycobacteria, e.g., integration-proficient vectors (120). The best-characterized *Mycobacterium* prophage is L5: the left arm of its 52-kb genome contains the structural gene cluster that shares its gene organization with many dairy and lambdoid coliphages (84) (Fig. 6). The right arm starts with the integrase gene and carries all the early genes on the opposite strand. Lysogeny is established by a phage repressor, which is encoded about 40 ORFs downstream of the integrase gene. The repressor needs to be stabilized by a protein encoded in an adjacent nonessential phage region to escape degradation (157). As in dairy phages, closely related virulent phages were described for phage L5. One of these isolates is mycobacteriophage D29: in theory it can be derived from L5 by a number of insertion and deletion events and the substitution of genes. Most notable is the deletion starting in the phage repressor gene and extending over the next 10 ORFs in the immunity region (72). The L5 repressor regulates transcription at an early lytic promoter, but it also affects gene expression throughout the prophage genome by binding to numerous "stopoperator" sites located within

short intergenic spaces in both the early and the late lytic operons (27).

The temperate mycobacteriophage Bxb1 is heteroimmune with L5 and shows a distinct host range. The overall genome organization of both prophages is quite similar, but sequence similarity to L5 or D29 proteins is patchwise and in general does not exceed 60% amino acid identity. Bxb1 encodes several enzymes that could degrade or modify the mycobacterial cell wall (141).

Further mycobacteriophages have been described, including TM4 with a 53-kb genome encoding proteins with similarity to haloperoxidase and glutaredoxins (73). In the structural genes, mycobacteriophages L5 and TM4 had some sequence relatedness to *Lactococcus* phage r1t (52). Still other sequenced mycobacteriophages showed very distinct morphology (G. Hatfull, personal communication).

## γ-PROTEOBACTERIA

### Basic Phage Types

Together with low-G+C gram-positive bacteria, the γ-proteobacteria represent the best-documented eubacteria with respect to complete genome sequences. The sequenced γ-proteobacteria comprise human- and plant-pathogenic bacteria, free-living bacteria, and intracellular symbionts. The symbionts lack any prophages, while some of the pathogens contain multiple prophages that confer important virulent factors to the lysogenic host. As shown above, in low-G+C gram-positive bacteria, a rather limited set of basic phage types represented the majority of prophages. This is also the case for γ-proteobacteria. The analysis of the prophages is facilitated by the fact that all basic prophage types have carefully characterized representatives in *E. coli*. Five types of prophages dominate in γ-proteobacteria.

The first group is represented by lambda-like *Siphoviridae*. Some of the lambda-like prophages have sequence similarity to phage lambda over one or more modules, as illustrated by the *Salmonella* phages Gifsy-1 and Gifsy-2 (Fig. 10A). However, many prophages from γ-proteobacteria maintain only the overall lambda-like genome map without demonstrating sequence similarity to lambda. It is not yet clear whether this represents gradients of relatedness or the existence of different lineages for the structural gene cluster in lambda-like phages from γ-proteobacteria.

The second group consists of prophages of the P2-like genus of *Myoviridae*. Not only is the basic genome organization of P2 conserved, but also many P2-like prophages from diverse γ-proteobacteria have substantial sequence identity to P2 (Fig. 10B).

The third group is composed of Mu-like prophages (Fig. 10C). This group of prophages is being found in a growing number of bacteria: the γ-proteobacterium *Shewanella* (86) the β-proteobacterium *Neisseria*, and the *Thermus* bacterium *Deinococcus* (147). There was a gradient of relatedness: *E. coli* phage Mu and the *Haemophilus influenzae* prophage FluMu had up to 50% amino acid identity over most late genes, while the sequence relatedness to the Mu-like prophage from *Neisseria* was restricted to a smaller number of late genes. The prophage from *Deinococcus* shared an overall genome organi-

zation with Mu but had only a few sequence relationships. The available Mu-like prophages from γ-proteobacteria are noted for their marked mosaicism in their genetic relationships (86).

The fourth group consists of filamentous phages that are related in their genome organization, but not in their sequence, to the single-stranded DNA phage M13 (Fig. 11). This affinity is somewhat surprising since M13 does not integrate its DNA into the *E. coli* chromosome. An infrequent observation is that of P4-like prophages, as demonstrated by the *E. coli* O157 prophage Sp2 (Fig. 10D). P4 is a satellite phage that depends for its propagation on the presence of the helper phage P2; this may explain its rare observation in sequenced genomes. Another rare finding is that of the P22-like *Podoviridae*. The *Salmonella* phage P22 is one of the model phages of molecular biology and is also notable for its O-antigen conversion mediated by two glycosyltransferases and a translocase encoded downstream of the phage integrase. The taxonomic status of P22-like phages is disputed, since P22 shows a chimeric genome with links to *Podoviridae*, *Siphoviridae*, *Myoviridae*, and even *Inoviridae* (202). Finally, *Pseudomonas putida* showed a T7-like prophage known so far only from obligate virulent phages (Fig. 12).

Several members of the major classes of prophages from γ-proteobacteria carry lysogenic conversion genes or virulence factors. The best-known example is that of the *Vibrio cholerae* prophage CTXphi, which encodes the cholera toxin responsible for the dramatic form of watery diarrhea associated with that bacterial infection (217).

### Chimeric Prophages

Prophage genomics showed that recombinant phages combining structural genes from distinct phage families, as seen in *Salmonella* phage P22, are not rare. Another example is *Shigella flexneri* prophage SfV (3). Like P22, it is an O-serotype-converting phage, which encodes enzymes involved in glycosylation of the O antigen, and as in P22, these genes are located between the integrase gene and *attL*. Phage SfV combines a lambda-like DNA-packaging/head morphogenesis module (with many sequence links to Sf21-like phages from gram-positive bacteria) with a Mu-like tail gene module (3). Morphologically, this results in a replication-competent Mu-like myovirus and another dilemma for phage taxonomy. A similar siphovirus-myovirus recombinant was found in *E. coli* phage P27 (175). Also, *Xylella fastidiosa* prophages XfP1 and XfP2 combine DNA-packaging genes from lambdaoid phages with tail genes from P2-like *Myoviridae* (187) (Fig. 13). *Salmonella* prophage CT18-02 is a Mu- and P2-like recombinant myovirus.

An even greater genome rearrangement occurred with a phage-like DNA element in *Pseudomonas aeruginosa*. It consists of a P2-like tail gene cluster separated by a lysis cassette from a lambda-like tail gene cassette (Fig. 11A). Here the prophage remnants have taken up new functions and have become bacteriocins (pyocin R2 and F2) (152) that present morphologically as phage tail structures. Superinfecting *Pseudomonas* phage PS17 showed phenotypic mixing with pyocin R2, resulting in an extended host range for PS17, but genetic recombination was not observed (186), unlike the situation for dairy phages confronted with natural or engineered phage resistance mechanisms. Some lactococcal phages escaped from

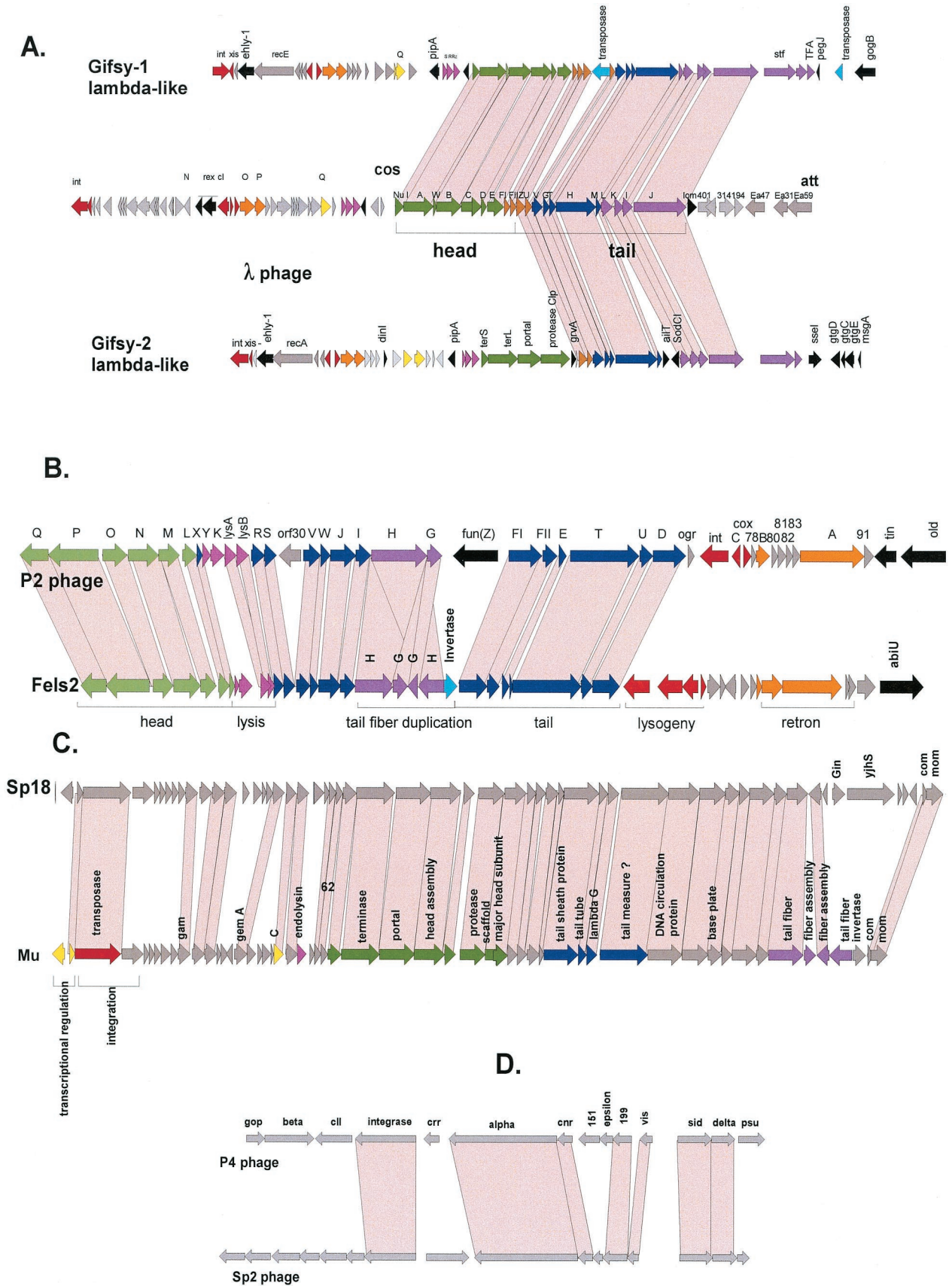


FIG. 10. Typical prophages found in the genomes of *E. coli* and *Salmonella*. (A) Alignment of the *Salmonella* prophages Gifsy-1 and Gifsy-2 with *E. coli* phage lambda. Genes with sequence similarity at the protein level are linked by shading. For the color coding of the genes, see Fig. 3. (B) Alignment of the P2-like *Salmonella* prophage Fels-2 with *E. coli* phage P2. P2 genes were annotated and color coded according to their attribution to phage modules. Genes linked by sequence similarity are connected by shading. (C) Alignment of the O157 prophage Sp18 with *E. coli* phage Mu. (D) Alignment of the O157 prophage Sp2 with *E. coli* satellite phage P4.

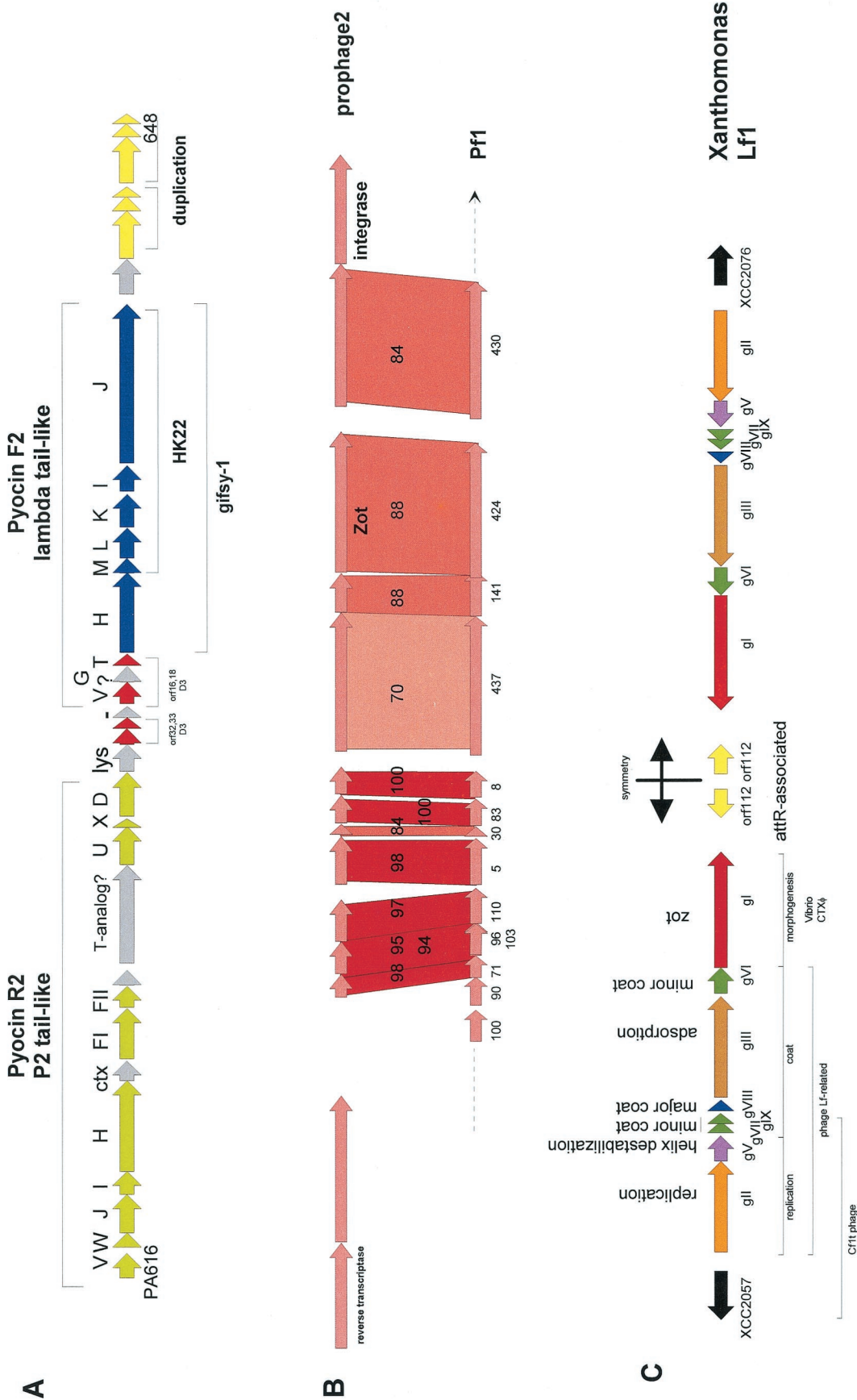
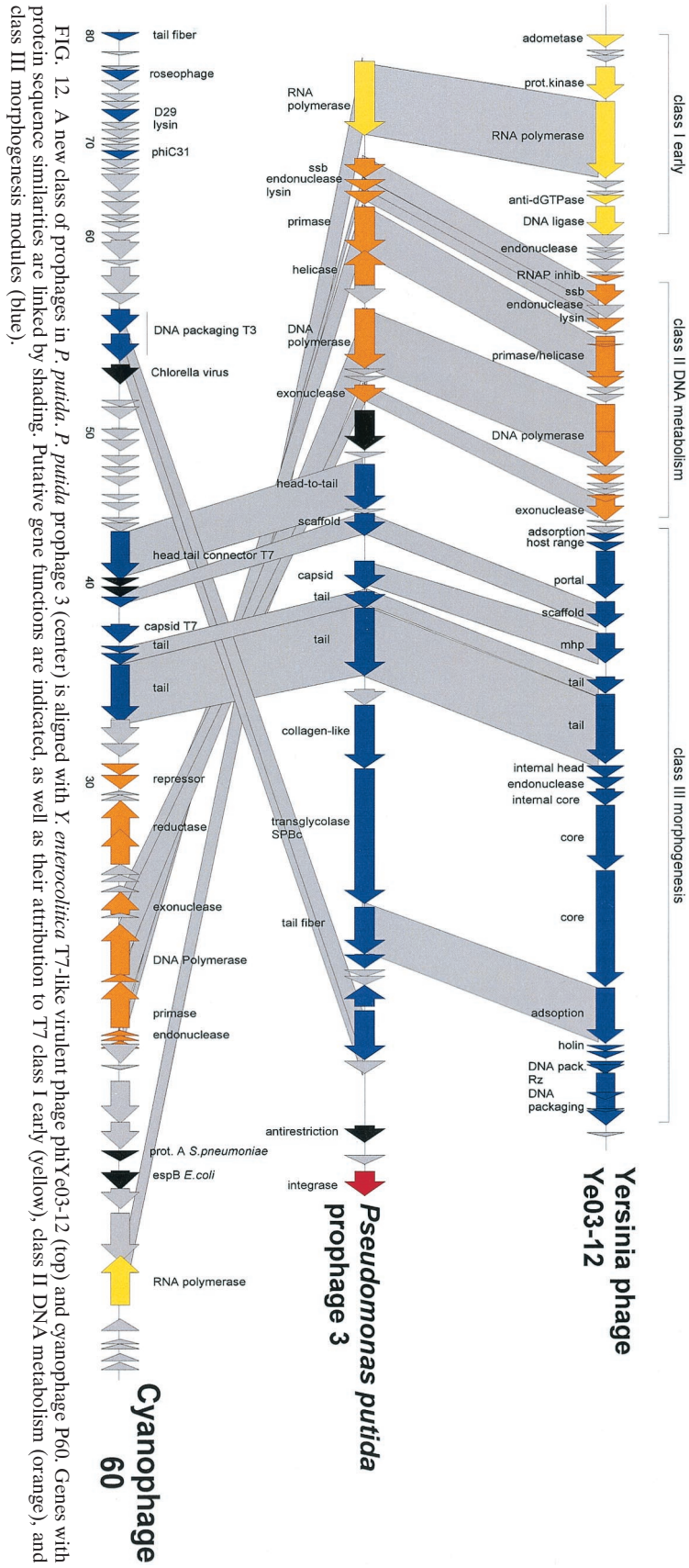


FIG. 11. *Pseudomonas* phage-derived bacteriocin and filamentous phages from *Pseudomonas* and *Xanthomonas*. (A) The gene map of a prophage-like DNA element in *P. aeruginosa* encoding the bacteriocins pyocin R2 and F2 resembling phage tail modules. Genes with sequence identity to phage tail genes from phage P2 are colored in green, and those with sequence identity to genes from lambdaoid phages are in red (*Pseudomonas* phage D3) or blue (*Salmonella* phage Gifsy-1 and *E. coli* phage HK22). The extent of lambda-like and P2-like tail genes is marked by brackets. Further unattributed likely phage tail genes are shown in yellow. For easier orientation, the gene numbers of the first and last genes of this likely recombinant prophage remnant are provided. (B) *P. aeruginosa* PAO1 prophage 2 aligned with filamentous *Pseudomonas* phage Pf1. The numbers below the Pf1 gene map identify the phage genes from the GenBank entry. The numbers between the maps indicate percent protein sequence identity. The symmetry axis of the gene map is indicated, as well as the tentative gene functions and sequence similarities to other filamentous phages.





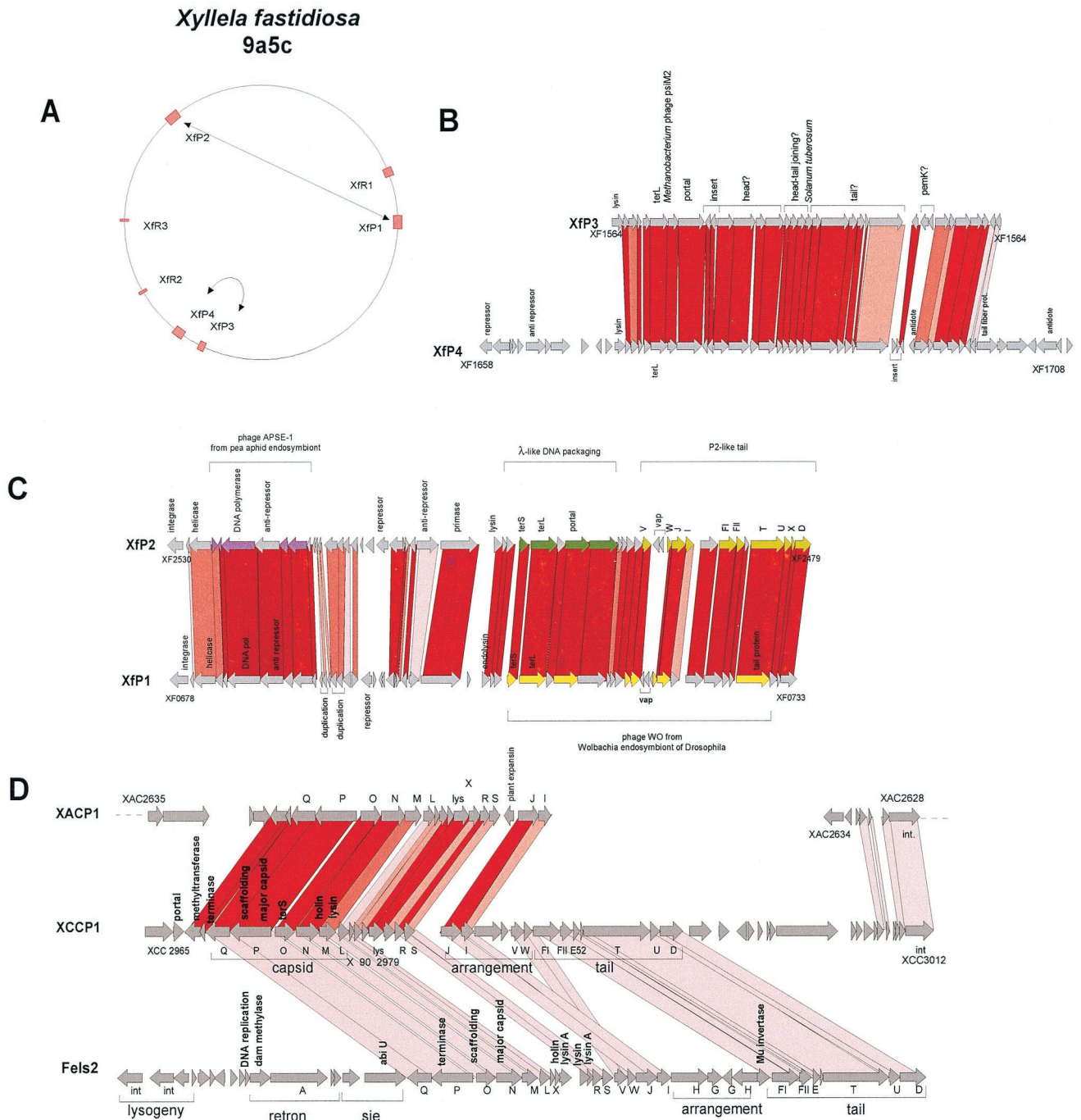


FIG. 13. Prophages from bacterial plant pathogens. (A) Prophage content of *X. fastidiosa* strain 9a5c. Prophages linked by high DNA sequence identity are connected by double-headed lines. (B) Alignment of the *X. fastidiosa* prophages XfP3 and XfP4. The degree of protein sequence identity is reflected by different intensities of red shading. Suspected gene functions are indicated. (C) Alignment of the *X. fastidiosa* prophages XfP1 and XfP2. Selected genes are annotated. Genes with protein sequence similarity to phage APSE-1 from an endosymbiont of a pea aphid are marked in violet, and those with similarity to phage WO from the *Drosophila* endosymbiont *Wolbachia* are marked in yellow. Genes corresponding to lambda-like DNA-packaging genes are in dark green, while tail genes resembling P2 phage proteins are in light green. The gene numbers for the first and last genes of the prophages are indicated to facilitate orientation on the *Xylella* genome. Candidates for virulence-associated genes are marked with vap. Prophage genes with protein sequence identity are linked by shading. The different shades of red indicate >90%, >80%, >70%, and any protein sequence identity. (D) Alignment of the *X. campestris* prophage XCCP1 (center) with *X. axonopodis* prophage XACP1 (top) and *Salmonella* prophage Fels-2 (bottom). Note that the leftmost XACP1 genes are virtually translocated and inverted to allow an optimal alignment with the XCCP1 gene map.

control by replacing part of their genome with DNA from prophages or prophage remnants which they encountered in the infected cell (62, 145). These observations demonstrate clearly that prophage DNA is the raw material for both phage and bacterial evolution.

### *Escherichia coli* and *Shigella flexneri*

***E. coli* O157.** The Shiga-toxin producing *E. coli* (STEC) strains represent serious emerging food-borne pathogens with an exceptional prophage content (Fig. 1C). The two sequenced O157 strains, EDL 933 and Sakai, are very similar with respect to their bacterial DNA but differ more in their prophage content (160, 166). Strain EDL933 lacks a Mu-like phage (Sp18) found in Sakai (Fig. 10C) and contains less lambda-like prophages (Fig. 1C). In EDL933, 12 prophage sequences were identified. Only the Shiga-toxin-converting phage 933W can produce infectious phage particles (168). There are nine defective lambda-like prophages which have large regions of high DNA sequence identity to phage lambda, to each other, and to prophages from the Sakai strain.

Among the 18 prophages of the Sakai chromosome, 13 are lambda-like phages (Fig. 1C). Prophage Sp18 closely resembled myovirus Mu (147); they had 43 to 74% amino acid identity over the structural genes. Differences were the replacement of Mu genes involved in host cell killing by similarly organized but sequence-unrelated genes (Fig. 10C). Mu showed extra genes involved in transcriptional regulation, while Sp18 showed extra genes adjacent to the tail fiber genes. Predicted proteins from prophages Sp13 and Sp2 had sequence similarity to myovirus P2 and its satellite phage P4, respectively. However, their genome map is rearranged with respect to P2 and P4 (Fig. 10D). All lambda-like Sakai prophages contained frameshift mutations and various types of deletions and insertions of IS elements, and none yielded an infectious phage (135, 229). A dot plot matrix demonstrated large regions of high DNA sequence identity between the different lambda-like prophages (22). It is not clear whether this intriguing similarity is the result of infection by closely related phages or propagation by a copy-and-paste mechanism within this cell lineage followed by some diversification via modular exchanges. For example, the Shiga toxin 2-producing prophages from the two STEC strains are integrated into the same locus, *wbrA*, and their DNA sequences are nearly identical over 85% of the prophage genome (135, 168). They differ, however, over the genetic switch and the DNA replication region. Numerous potential virulence factors are encoded by the prophages from both O157 strains (for gene maps of these prophages, see references 135, 160, 168, and 229). These factors also refer to toxins including Shiga-toxin 1 (168) or a necrotizing cytotoxin; the *bor* gene, implicated in serum resistance (9); the *lom* gene, encoding an outer membrane protein conferring the ability to survive in macrophages (172); an intestinal colonization factor; a type III secretion protein; extracellular enzymes like superoxide dismutase; tellurite resistance genes; a toxin-antitoxin system (*hok* genes) used by plasmids to assure their maintenance (77); and an avirulence-like gene from a bacterial plant pathogen. However, besides Shiga toxins, which are the principal virulence factors of STEC, there are no experimental data implicating other phage-encoded factors in STEC patho-

genicity. Unlike disease caused by *Streptococcus pyogenes*, the clinical manifestations of STEC (diarrhea, colitis, and hemolytic-uremia syndrome) are not very variable and are attributable exclusively to Shiga toxin.

***E. coli* K-12.** The sequenced laboratory *E. coli* strain also contained numerous prophage elements in addition to the inducible phage lambda (which was cured from the K-12 strain used for sequencing) (18). Several elements are clearly defective lambda-like prophage remnants (DLP-12, e14, Rac, and Qin), while the prophage nature of other elements is doubtful. No virulence genes were identified in the K-12 prophages DLP12, Rac, and e14; however, they encode proteins which might cooperate in the host physiology (36). K12 prophage *qsr'* contains the *bor* region, while lambda has *bor* and *lom* genes.

The *bor* gene product is an outer membrane lipoprotein that is constitutively expressed in the lysogen (10). This protein was detected in various *E. coli* lysogens including many clinical isolates, demonstrating its wide conservation and potential selective value. Expression of *bor* significantly and specifically increased the survival of *E. coli* in animal sera (9).

**Uropathogenic *E. coli*.** The extraintestinal *E. coli* strains differ from the diarrheal pathogens because they behave as either harmless human intestinal inhabitants or serious pathogens when they enter the urinary tract, bloodstream, or cerebrospinal fluid. The availability of the sequence of the uropathogenic *E. coli* strain CFT073 (219) allows comparisons with the laboratory *E. coli* strain and the enterohemorrhagic strains. Amazingly, the two pathogenic *E. coli* genomes are as different from each other as each is from the benign *E. coli* strain. The difference in disease potential is reflected in the absence of genes for the type III secretion system and of phage- and plasmid-encoded toxins found in diarrheagenic *E. coli*. CFT073 contains five prophages. Prophage 1 is a seemingly complete P2-like phage. Potential lysogenic conversion genes were detected upstream of the integrase, downstream of the replication genes, and at the position of the extra P2 genes *orf 30* and the *fun* gene. However, none showed links to known virulence factors. Only a few bacterial genes separate prophages 2, 3, and 4 from each other. Prophage 2 shows the structural gene cluster of a lambda-like siphovirus with a methyltransferase and the *lom* gene integrated into the tail fiber genes. Integration, recombination, transcriptional regulation, and lysis genes are present in the order characteristic for lambdoid phages. Prophage 3 is a short remnant consisting of an integrase gene, a few DNA replication genes, and three head genes. Prophage 4 demonstrates phage P21-like DNA packaging and head genes combined with lambda-like tail and tail fiber genes. The non-structural genes show a constellation reminiscent of lambdoid phages. Extra genes include *tonB* and *bor* downstream of the lysis genes, an outer membrane protein-encoding gene next to the Q gene, and six genes from O157 prophages flanked by two transposase genes between the integrase and exonuclease genes. Prophage 5 is a lambda-like phage containing cold shock genes downstream of the antiterminator and lysis genes. Between the tail fiber genes and *attR*, a DNA damage-inducible gene and gene resembling an abortive phage infection function from lactococcal phages were identified.

***S. flexneri*.** The sequencing of *Shigella flexneri* revealed its close genetic relationship to *E. coli*. In fact, *S. flexneri* serotype 2a is more closely related to *E. coli* strain K-12 than is the

pathogenic O157 strain (99), suggesting that *Shigella* evolved from *E. coli* with the appearance of early humans and does not merit its current genus taxonomic status. Important virulence factors include the seven *ipa* genes (invasion plasmid antigens). Five are located on *Shigella*-specific islands next to a gene with links to *E. coli* phage P27. Several genes of the islands have homologies to O157 prophages. The *sci* island carries genes related to phages P22 and HK620, suggesting another phage-transmitted pathogenicity island. The SfII island has been demonstrated to be a prophage in which two genes (*bgt* encoding a bactoprenol glycosyl transferase and *gtrII* encoding a glycosyltransferase) are required for expression of the type II O antigen (138). As in *S. flexneri* phage SfV and *Salmonella* phage P22, these genes are encoded adjacent to the phage integrase gene. Thus, phage-mediated horizontal DNA transfer appears to be one of the major routes by which *S. flexneri* gains virulence determinants.

**Location of virulence factors.** The potential virulence genes are located at a few preferred lambdoid prophage genome positions (Fig. 10A). These sites were less frequent downstream of the Q antiterminator, the tail fiber genes, the lysis cassette, and the N antiterminator (22, 215). Plugging genes downstream of the Q antiterminator gene puts them under a sophisticated prophage transcriptional control circuit (156). The location of virulence genes downstream of the tail fiber genes probably ensures their mobility. The tail fiber genes encode the phage antireceptor protein. This phage adhesin is often a multidomain protein with internal repeats and thus is the target of homologous recombination within the species, which allows the exchange of this and probably adjacent genes between morphologically distinct phage systems (81).

**Expression of prophage genes.** In *E. coli* H-19B lysogen, Shiga toxin 1 expression is low in uninduced lysogens but is induced via the iron-dependent Fur transcriptional repressor under low-iron conditions. It is also induced by phage induction from a promoter activated by the Q antiterminator (213). These coding regions plus transcription control sequences have been termed morons (more DNA). Moron accretion plays an essential role in the phage evolution model of the Pittsburgh phage group (88). Prophage release and toxin production are correlated. A physiological signal for prophage induction could be hydrogen peroxide released from human neutrophils attacking the lysogenic bacterium (212). Deletion of the antiterminator Q control region had no effect on the intestinal survival of the *E. coli* lysogen but prevented toxin production. Thus, toxin production is derived largely from the small fraction of bacteria lysing after prophage induction (214). It is not clear how toxin production helps the bacterial host; what is harmful to the human is not necessarily good for the bacteria. Isogenic lysogens of diverse *stx*-encoding phages produced markedly different amounts of Shiga toxin, pointing to distinct regulation mechanisms (211). In fact, genomic analysis revealed some heterogeneity around the *stx* genes in coliphages (149, 201).

### *Salmonella*

***S. enterica* serovar Typhimurium.** Biological experiments demonstrated that prophage-located virulence genes have played a decisive role in the emergence of *Salmonella enterica* serovar

Typhimurium, another important food-borne pathogen. The attenuated *Salmonella* LT2 strain contains four prophages and three prophage remnants (139). Gifsy-1 is a lambdoid phage with extensive sequence similarity to phage lambda over the structural genes (Fig. 10A). ORFs with links to putative virulence factors were demonstrated in three genome regions. One region is between the tail fiber genes and *attR* site (70). The *pagJ* gene is part of the PhoP-PhoQ-regulated transcripts that are activated after phagocytosis of *Salmonella* to increase bacterial survival within the phagosome environment. Adjacent genes have links to proteins involved in DNA transposition and rearrangement, suggesting that this region may have been or is a mobile element. The region also contains a gene with similarity to the type III-secreted protein of *Salmonella* (*gogB*) (Fig. 10A). Next to the lysis genes, Gifsy-1 carries the *pipA* gene, which is transcriptionally induced specifically when the lysogenic bacterium colonizes the small intestine of mice. Inactivation of *pipA* decreases the survival of *Salmonella* in Peyer's patches (193). In the vicinity of the integrase gene, prophage Gifsy-1 carries an enterohemolysin 1-associated gene (*ehly-1*) (Fig. 10A) that is also found in a corresponding region of prophage 933W from *E. coli* O157 (168).

Gifsy-2 is a more distant relative of lambda with rearranged capsid genes (a truncated portal protein followed by a ClpP proteinase fused to an unidentified/major capsid protein) (Fig. 10A). The genome position of the first head-to-tail connector gene is occupied by the "antivirulence" gene *grvA*: both its deletion and overexpression increase the virulence of the lysogen in an animal model (91). In the virulent *Salmonella* strain 14028, a Gifsy-2-like prophage is inducible, but it has a distinct genetic map around the *grvA* gene compared to Gifsy-2 (91). In the tail fiber region, two virulence factors are inserted: one is a homolog of the lambda *lom* gene, the other is a superoxide dismutase that protects the lysogen from phagocyte-derived reactive oxygen species (69). Between tail fiber genes and the *attR* site, a type III secretion protein (*sseI*) and a stress response gene (*dinI*) were identified as possible virulence factors (69) (Fig. 10A). The Fels-1 prophage is yet another lambdoid prophage in *Salmonella* strain LT2: a distinct *sodCIII* superoxide dismutase and a neuraminidase flank the tail fiber genes (70).

Prophage Fels-2 closely resembles the temperate myovirus P2, demonstrating a one-to-one gene correspondence over most of the gene map (Fig. 10B). Differences were observed at two regions where P2 contains lysogenic conversion genes. At the P2 *fun* gene position, Fels-2 encodes a Mu-like Gin DNA invertase. Notably, the adjacent tail genes H and G are duplicated and inverted in the Fels-2 prophage (Fig. 10B). At the position where P2 encodes superinfection exclusion genes (*tin* and *old*), Fels-2 encodes an *abiU*-like phage resistance protein known from *L. lactis*. P2-like prophages are also found in other *Salmonella* strains. An interesting case is prophage phiSopE, a P2-like prophage that encodes the type III effector protein SopE at the position of the P2 *fun* gene (143). This observation is interesting for two reasons. First, it suggests that there might be preferred places to insert virulence genes into the P2 genome (between the tail genes G and F1 and next to the *cos*-site). Second, a *sopE* gene was also detected in lambda-like *Salmonella* phages. Comparison of the sequences identified a conserved 1.2-kb cassette, suggesting that these virulence

genes were horizontally transferred between these two distinct phage families (143, 144). Gene transfers between phages could be of use when phages try to establish a polylysogenic state and are confronted with problems due to immunity systems, occupation of integration sites, and restriction enzymes. These observations led to the concept that a variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella* (70). Complexity is created not only by the possible permutation of distinct prophages associated with variable lysogenic conversion regions but also by recombination between different *Salmonella* prophages (144).

***S. enterica* serovar Typhi.** *S. enterica* includes several closely related serovars that cause distinct diseases: serovar Typhimurium causes in humans gastroenteritis, while serovar Typhi is the etiological agent of typhoid fever. Long before the advent of full genome sequencing, hybridization analysis indicated that the *Salmonella* serovars share >90% of their DNA content. One major source of sequence diversity are prophages (64, 139, 164). A dot plot analysis revealed substantial DNA sequence identity between the three lambda-like *S. enterica* serovar Typhimurium prophages Gifsy-1, Gifsy-2, and Fels-1 but only limited DNA sequence similarity to serovar Typhi prophages. The structural genes of serovar Typhi prophage CT18-01 form a typical gene map of lambdoid phages; its closest relative is *Xylella fastidiosa* prophage XfP3. A further block of early CT18-01 genes shows links to phage lambda. Notable is the sharing of a lipoprotein downstream of the Q antiterminator between CT18-01 and Gifsy-2.

Prophage CT18-02 is a chimeric prophage. In its nonstructural, DNA-packaging, and head modules, many genes resemble Mu-like phages from diverse bacterial species. The gene map is, however, rather different from that of Mu: the transcriptional orientation of the leftmost genes and the position of the transposase have been changed. In addition, two unusual genes are integrated into the early genome region: these encoding an enterohemolysin 2-like protein (56% amino acid identity to prophage 933W from *E. coli* O157) and a *Vibrio*-like type II secretion protein. The right genome half of CT18-02 consists of P2-like tail and tail fiber genes followed by four lambda-like side tail fiber genes and an invertase. CT18-03 is another prophage with a lambda-like genome organization. The structural gene module has extensive DNA sequence identity to CT18-01 and is followed by a cluster of recombination genes (encoding Mu-type invertase, lambda-type integrase, and RecT recombinase). The early genes show the characteristic organization of lambda-like phages into which a few candidate lysogenic conversion genes are inserted (enterohemolysin1-like gene upstream of the integration cassette, and *hok*-like gene with a transposase upstream of the Q anti-terminator). CT18-04 is a 7-kb prophage remnant corresponding to a contiguous cluster of P2-like tail genes flanked by an invertase and a reverse transcriptase-like gene. CT18-05 represents another P2-like prophage with respect to genome organization and 35 to 60% amino acid identity to P2. Where P2 carries the two lysogenic conversion genes *tin* and *old*, CT18-05 has two genes with links to *Xanthomonas*. As with many other prophages, it also encodes a DNA modification enzyme (Dam methylase).

### *Yersinia* and *Haemophilus*

*Y. pestis*, the causative agent of bubonic and pneumonic plague, has been sequenced for Mediaevalis (KIM) and Orientalis (CO92) biotypes. Phage genes but no complete prophages were detected in the KIM strain. The most complete are a P4-like prophage and a prophage remnant consisting of a tail and tail fiber genes with close sequence links to the *E. coli* phage N15 and further isolated phage genes (integrase, anti-repressor, antiterminator, *nin*, and lysis genes). Both observations are somewhat surprising: only a vague hint for a P2-like helper phage necessary for a P4-like prophage was detected in the KIM strain (50), and prophage N15 is maintained in *E. coli* as a linear plasmid but not as an integrated DNA. The CO92 strain contains a number of prophage-like DNA elements but no complete phage genome. For example, one of those elements consisted of genes X and V on one side and genes I and IV on the other side, with links to filamentous phages flanking the *puvA* virulence gene from *E. coli*. Another prophage element contains a long stretch of 10 nearly contiguous Mu-like tail genes interrupted only by a tRNA synthetase gene. With respect to the sequence, the closest relative is *Shigella* phage SfV. Seven genes upstream of the tail gene cluster, a *cI*-like repressor, and a *Q*-like antiterminator gene were identified bracketed by an *msgA*-type virulence gene and another tRNA synthetase gene. The most complete prophage showed a nearly complete lambda-type genome organization, lacking only the DNA replication genes. The prophage genome contained five transposases interrupting the phage integrase and terminase gene.

The sequenced *Haemophilus influenzae* genome harbors a nearly intact prophage with relatively close sequence relationships to *E. coli* phage Mu (147). In addition to a very small MiniFluMu prophage remnant, only one further highly deleted prophage remnant was seen in this bacterium (89).

### *Vibrio*

*Vibrio cholerae* is the etiological agent of cholera. *V. cholerae* as a species includes both pathogenic and nonpathogenic strains that vary in their virulence gene content. Virulence genes were acquired by horizontal gene transfer. The sequencing of the El T or strain N16961 (85), combined with microarray analysis (63), allows insights into the molecular ecology of this environmentally widely distributed species. The disease-causing cholera toxin (CT) is encoded by the filamentous prophage CTXphi, consisting of 10 genes (217). The genes of the so-called RS repeat encode repression, replication, and integration function (*rstR*, *rstA*, and *rstB*, respectively). The core genes *cep*, *orfU*, *ace*, and *zot* correspond to morphogenesis genes VIII, III, VI, and I of the reference *E. coli* filamentous phage M13. However, *Zot* has also been characterized as the zonula occludens toxin, *Cep* has been characterized as a core-encoded pilin, and *Ace* has been characterized as an accessory cholera enterotoxin. This gene cluster is followed by the *ctxAB* operon encoding CT.

*Zot* is an enterotoxin elaborated by *V. cholerae* that increases intestinal permeability by interacting with a mammalian cell receptor with subsequent activation of intracellular signaling, leading to the disassembly of the intercellular tight junctions

(56). Although there are several papers describing the toxic activity of the *V. cholerae* Zot and Ace proteins, it is very doubtful that these gene products function as toxins during *V. cholerae* infection. Zot and Ace are required for CTXphi phage morphogenesis; deletion of these genes from candidate vaccine strains did not reduce their virulence (195). When a pI (Zot) or pVI (Ace) orthologue is seen in a genome sequence, this should not be an indication of a lysogenic conversion gene. These are not dispensable genes (like Shiga, cholera, or diphtheria toxin genes) but essential phage genes.

The sequenced EI Tor strain carries only a single copy of the CT prophage; other *V. cholerae* strains carry several copies of this element, and strains of the classical *V. cholerae* biotype have a secondary copy of the prophage that is localized on the second and smaller *V. cholerae* chromosome. Thus, prophages are presumed to be subject to selective pressure, affecting their copy number and chromosomal location. On the other side of the CTXphi prophage is a region encoding an RTX toxin and its activator (RtxC) and transporters (RtxBD). Codon analysis suggested that the RTX region was horizontally acquired, along with the adjacent sensor histidine kinase and response regulator (124). The bacterial agent of epidemic cholera also carries a VPI element. VPI has many features of a bacterial pathogenicity island: it is 40 kb and contains genes associated with virulence, regulation, and mobility; it is inserted adjacent to a tRNA gene; and it has a different G+C content from the host chromosome. It was suggested that the VPI element represents an unusually long filamentous prophage (104). However, the data supporting this prophage claim have not been reproduced in several laboratories; the notion of a prophage link for this element should therefore be toned down. The VPI element encodes a toxin-corregulated pilus that functions as a colonization factor for the lysogenic bacterium and at the same time as a receptor for the CTXphi phage (104). The toxin-corregulated pilus is a fiber of polymerized pilin protein (TcpA) which was also reported as a coat protein of the postulated VPI prophage. The distribution pattern of the CTXphi and VPI elements can explain why only two of more than 100 *V. cholerae* serotypes, namely, O1 and O139, can cause cholera epidemics.

*Vibrio vulnificus* is an estuarine bacterium responsible for human gastroenteritis and wound infections. Its sequenced genome (accession number NC\_004459) contained a single prophage-like element containing a lambda-like lysogeny region next to a P2-like tail gene region flanked by genes encoding integrases and XerC/D-like recombinases.

### Plant Pathogens

It is thought that globally one-fifth of potential crop yields is lost each year because of disease, and a significant proportion of these diseases are of bacterial origin. The distinct anatomical structure and defense system of plants present other challenges to a would-be pathogenic bacterium than those encountered by animal pathogens. Only a few of the sequenced  $\gamma$ -proteobacteria are phytopathogens. Two observations are fascinating. First, *Xylella* needed just 10 years from its identification as phytopathogen to become one of the best-investigated bacteria with respect to comparative genomics. Second,

prophages turned out to play a decisive role in the genome evolution of this plant pathogen.

*X. fastidiosa*. The leafhopper-transmitted bacterium *X. fastidiosa* causes a serious disease of orange trees (187). Symptoms include a chlorosis of the leaves and hardening of the fruit. Anatomically, *X. fastidiosa* is xylem limited. The 2.7-Mb bacterial genome of strain 9a5c contains five prophages representing as much as 7% of the bacterial genome (187) (Fig. 13A) and several possible prophage remnants. Two of the prophages, XfP1 and XfP2, 41 and 43 kb, respectively, are closely related at the DNA sequence level and lie in opposite orientations in distinct genome regions (Fig. 13C). Notably, one of the few differences between these two prophage genomes is in the integrase genes, explaining the distinct integration sites. A second region of differences between the two prophages lies near a putative repressor gene. A third variant region is in the vicinity of tail genes. XfP1 carries a gene with similarity to *vapD*, a purported virulence-associated protein from the sheep pathogen *Dichelobacter* (40), but the two genes also showed similarity to a toxin-antidote maintenance system used by low-copy-number plasmids and some prophages (*hok*-like genes [77]). At exactly this position, XfP2 contains two ORFs from a transposon resembling a DeoR-family regulatory protein from *Yersinia pestis*. The leftmost 10 kb near the integrase gene consists of likely DNA replication functions with sequence matches to prophage APSE-1 (203) from an insect symbiont. In the center of the prophage, the endolysin gene separates nonstructural from structural genes. The DNA-packaging and head genes resemble in organization and sequence lambda-like phages. The tail and tail fiber genes, in contrast, are closely related to phage P2, suggesting a hybrid phage between a  $\lambda$ -proteobacteria-like siphovirus and a P2-like myovirus. Also, despite their length difference (21 and 35 kb respectively), prophages XfP3 and XfP4 are closely related at the DNA level (Fig. 13B). XfP3 is clearly a defective prophage; it consists mainly of a complete set of structural genes which recall in organization and by sequence similarity an Sfi11-like siphovirus known from prophage 4 of the gram-positive bacterium *Listeria innocua*. XfP3 is interrupted in the putative head-to-tail joining genes by a hypothetical gene from the potato plant (100% amino acid identity). A further interruption was observed in the putative XfP3 tail fiber genes, where a *pemK*-like plasmid antidote gene was inserted. The XfP4 prophage also contains two further tail fiber genes and a lysogeny module but no DNA replication genes, again suggesting a defective prophage.

Long regions of high DNA sequence identity between prophages residing in the same lysogen should allow homologous recombination across the prophages. Genome rearrangements are thus predicted in strain 9a5c. Indeed, the sequencing of the *X. fastidiosa* strain Temecula and its comparison with the 9a5c strain supports such a prediction (207). The two *Xylella* strains share 98% of their genes, and their average amino acid identity is 96%. Prophage genes figure prominently in the strain-specific genes of both bacterial isolates. The Temecula strain contains at least six prophages, none of which show sequence relatedness to prophages of strain 9a5c. For example, the Temecula strain contains a prophage copy from a filamentous phage; however, its closest relative is the *Xanthomonas* phage Lf and not Pf3 (90), as in a 9a5c prophage remnant. In addi-

tion, the Temecula strain harbors two *S. enterica* serovar Typhi-like prophages not detected in the 9a5c strain. An alignment of the two genomes revealed three chromosomal regions that were translocated and inverted. The three large rearrangements were all flanked at one border by a putative phage-related integrase gene, suggesting that they were phage mediated. The same is true for the few genomic islands identified in the two genomes.

**Xanthomonas.** Two different species of the genus *Xanthomonas* have been sequenced (48). *Xanthomonas axonopodis* is the cause of citrus canker, and *Xanthomonas campestris* induces black rot disease in cabbage. *X. campestris* harbors two prophages. Prophage XccP1 is located at about 3.5 Mb and represents a rearranged P2-like myovirus genome. A major rearrangement occurred around the attachment site, resulting in a displacement of the lysogeny and DNA replication modules from the left to the right side of the integrated prophage; a minor rearrangement occurred in the tail gene region, where the P2-like tail genes J and I and genes V and W exchanged places (Fig. 13 D). A closely related prophage, XacP1, was detected at about 3.1 Mb in *X. axonopodis*; however, it lacked most of the tail genes and is thus a defective prophage. In the middle of the remaining four tail genes, a protein is encoded that resembled a plant protein (49% amino acid identity to *Arabidopsis*) induced in citrus blight disease. This protein belongs to the widely distributed expansin family, a group of extracellular proteins that directly modify the mechanical properties of plant cell walls, leading to turgor-driven cell extension (121). Both prophages contain a methyltransferase which is not part of the P2 genome map.

*X. campestris* contains a second prophage at 2.4 Mb (near the terminus-associated genome inversion between the two *Xanthomonas* strains) that belongs to the *Inoviridae* family. *Inoviridae* are filamentous phages, which were extensively characterized as tools in molecular biology (*E. coli* phages M13 and fd). Over a major part of the genome, the prophage shows high-level sequence identity to the replication-competent *Xanthomonas* phage Lf; the exception was the *gl* gene (123) (Fig. 11C). The prophage showed the typical gene organization of filamentous phages, with about 10 genes covering DNA replication, coat, and morphogenesis genes and a putative phage export gene in the intergenic region (123). Interestingly, the gene at the position of the morphogenesis gene *gl* showed 30% amino acid identity to the Zot-like protein from *Vibrio* phage CTXphi. A further particularity of the Lf-like *Xanthomonas* prophage is its perfect side-by-side duplication. However, in contrast to *V. cholerae* prophage CTXphi, which shows in some strains a tandem repeat of the RS element which is necessary for phage production (49), the Lf-like *Xanthomonas* prophages show a symmetrical head-to-head constellation of two complete prophages (Fig. 11C).

### Opportunistic Pathogens and Free-Living Bacteria

**Pseudomonas.** *P. aeruginosa* grows in soil and coastal water, as well as on plants and animal tissues, and it represents an important opportunistic pathogen. With a 6.3-Mb genome, it is markedly larger than most other bacterial genomes, which explains its metabolic and ecological versatility. In the sequenced PAO1 strain, only two prophage elements were iden-

tified at a genome position of about 0.7 Mb (194). One element is the prophage tail-derived bacteriocin described above; the other is a filamentous prophage that shared the genome organization and high sequence similarity with *Pseudomonas* phage Pf1 (90). An interesting feature was the fact that the prophage was flanked upstream by a putative reverse transcriptase and directly downstream by an integrase (Fig. 11B). Filamentous phages normally do not encode an integrase and rely instead on bacterial enzymes for this process (XerCD in vibriophage CTXphi [94]).

PhiCTX is a temperate phage isolated from a *P. aeruginosa* strain that produces a pore-forming cytotoxin (note the subtle difference in terminology between phiCTX and vibriophage CTXphi). Not only does phiCTX share a genome organization with *E. coli* myovirus P2; over the structural genes, phiCTX shows marked homology to P2 genes, with 30 to 66% amino acid identity (151). Exceptions were only the tail fiber genes, involved in adsorption to the cellular receptor, and the lysis genes. In contrast, lysogeny, DNA replication, and lysogenic conversion genes differed between the two phages. In phiCTX, the integrase and attachment site were found to the left while the *ctx* toxin gene was found to the right of the *cos* site (cohesive ends). Interestingly, in phiR67 and phiR86, retron elements exist in the same location as the *ctx* element (57), suggesting that this position is a hot spot on the P2-like phage genome, where foreign genes can be picked up without disturbing the functions of other genes. Another interesting prophage of an unsequenced *P. aeruginosa* strain is prophage D3. It has been involved in serotype conversion of its lysogenic host by changing the O-antigenic polysaccharide side chain of the lipopolysaccharide. An O-acetylase responsible for part of the seroconversion was located on phage D3 upstream of the lysis cassette (107). D3 shows a genome organization characteristic of lambda-like coliphages, except that the lysis module has changed its position. In lambda, the lysis module is found between the *nin* region and the DNA-packaging genes, while in D3, the lysis cassette is located between the tail fiber and lysogeny genes. In this respect, D3 resembles the genome organization of Sfi21-like *Siphoviridae* from low-G+C gram-positive bacteria. Interestingly, the DNA-packaging and head-processing genes of D3 also had protein sequence similarity to Sfi21-like phages.

*Pseudomonas putida* is a soil bacterium that is associated in a nonpathogenic way with plant roots and has attracted the interest of biotechnologists for its versatile metabolic pathways. In fact, the recently sequenced strain KT2440 was shown to contain 33 putative genes encoding oxygenases (98). It also harbors four prophages. The first is a hybrid with an Sfi21-like siphovirus head gene cluster and a Mu-like myovirus-like tail gene module. The closest match for the head genes was found in the *Burkholderia* siphovirus phiE125. This is a surprising finding since phiE125 was isolated from a  $\beta$ -proteobacterium (224). Over its tail genes, the most closely related phage is *Shigella* phage SfV. The only difference from SfV was a group II intron maturase gene, which interrupts the otherwise conserved gene order after the tail sheath gene. The tail fiber genes were followed by genes encoding a lysin and an iron-regulated bacteriocin. Also prophage 2 is a chimeric phage: it combines an Sfi21-like siphovirus head gene module (again with sequence relatedness to phage from a  $\beta$ -Proteobacterium)

with P2-like myovirus tail and tail fiber modules (related to a pyocin-type *Pseudomonas* bacteriocin). The lysis genes have a unique constellation: the holin gene precedes the head gene cluster, and the lysin gene follows the tail fiber genes. Interestingly, this is an intermediate type of lysis gene constellation between temperate phages from low-G+C gram-positive bacteria and lambda-like phages. Prophage 2 showed repressor genes in a genetic switch configuration, but no integrase gene. In the vicinity of the genetic switch, a gene with links to a virulence-associated *vapE* gene from the sheep pathogen *Dichelobacter* was observed (40). *P. putida* prophage 3 demonstrated astonishing affinities. Its genome organization closely resembled that of T7-like *Podoviridae* (146) (Fig. 12). Class I early genes (encoding RNA polymerase and anti-restriction protein) were found at both ends of the *P. putida* prophage; in T7 they are located exclusively at the left side of the genome. As in phage T7, the class I genes are followed in prophage 3 by class II genes involved in DNA metabolism (endonuclease, primase-helicase, DNA polymerase, and exonuclease genes). In both phages, a lysis gene is placed between the endonuclease and primase genes. The right half of the genome is occupied by morphogenesis genes (portal, scaffold, capsid, tail fiber, and maturation and DNA-packaging genes). Two important differences were observed between this prophage and T7. The location of the T7 internal genes, which encode a hollow structure in the tail guiding DNA during genome packaging and ejection, was occupied by collagen-like and transglycosylase genes. T7 phages are known only as lytic phages; the observation of an integrase gene at the right end of prophage 3 and the integration of a T7-like prophage therefore comes as a surprise. T7-like phages are known from several  $\gamma$ -proteobacteria: an instructive case is the appearance of a close relative of T7 in *Y. enterocolitica* that could be aligned with T7 at the DNA level essentially over the entire genome (163). The observation that prophage 3 had 30 to 40% amino acid identity over numerous proteins to cyanophage P60 (41) is a further surprise, since a wide phylogenetic distance separates proteobacteria from cyanobacteria (Fig. 12). Interestingly, other cyanophages also had protein sequence similarity of up to 30% to coliphages, namely, the prototype virulent phage T4 (83). In contrast, neither T7-nor T4-like virulent phages were ever isolated from gram-positive bacteria.

*P. putida* prophage 4 shares the genome organization and up to 50% protein sequence identity with *Pseudomonas* phage D3 over the structural genes. The early genes are notable for the presence of two methyltransferase genes, a tellurite resistance gene, and a second integrase gene near the antiterminator gene and the lack of an identified lysis gene.

**Shewanella.** The biological activities of metal-reducing bacteria have considerable implications with regard to environmental pollutants. *Shewanella oneidensis* can directly reduce uranium and chromium from the dissolved liquid state to insoluble oxides. Its sequenced genome contains one lambda-like and two Mu-like prophages (86). Prophage lambda So shared sequence-related head and tail genes with *Pseudomonas* phage D3, while the tail fiber genes resemble those of coliphages HK97, HK22, and N15. The early genes shared a genome organization with phage lambda but contain a large insertion of four genes encoding methylating enzymes between the integrase gene and the genetic switch region. A second integrase

was found upstream of the lysis module. Prophage MuSo2 shared genome organization and sequence relatedness with *E. coli* phage Mu over the structural genes. However, neither duplication of tail fiber genes nor an invertase gene were found near the right genome end. Also, the early genes were organized as in Mu-like phages, but half of the genes lacked database matches. A potential secretion activator and a TraR-like protein are encoded upstream of the DNA-packaging genes. Also, prophage MuSo1 is a Mu-like phage. However, the syntax is punctuated by deletions of tail genes that were replaced by additional transposase and methylase genes and an insertion of four ORFs in the head gene cluster. Transposase genes were also inserted near the lysis cassette.

## OUTLOOK

We have restricted our analysis to prophages from sequenced genomes of three phylogenetic groups of eubacteria: low- and high-G+C gram-positive bacteria and  $\gamma$ -proteobacteria. This choice was motivated by two considerations. First, the majority of the sequenced phages are from these three groups, reflecting several decades of academic research with phages from *E. coli* and the medical and industrial interest in gram-positive bacteria. This allows instructive comparisons between the genomes of prophages and related replication-competent phages. Second, these three groups of bacteria also represent nearly half of the finished eubacterial genomes deposited in the NCBI database (at the time of the writing, they make up 44 of 91 genomes). Despite their importance for microbiological research, it should not be forgotten that these three groups of bacteria represent only a small spectrum of the phylogenetic diversity of bacteria. Much remains to be done in comparative prophage genomics, especially since prophages from bacteria outside of the surveyed groups will have substantial impact on our understanding of phage evolution.

What are the take-home lessons from a survey of prophage genomes in the three groups of eubacteria? We will try to provide some generalizations, which must be regarded as tentative in view of the recent nature of this research field.

### Prophage Distribution

Overall 40 (71%) of 56 sequenced genomes reviewed in this report contain prophage sequences exceeding 10 kb in length. Containing prophage genomes is thus more the rule than the exception in bacterial chromosomes. Is possession or nonpossession of prophages a significant property of the sequenced bacterial genome or just a question of statistical chance? There are some trends that can already be deciphered. One clear trend is that genomes from bacteria that experienced reductive evolution lacked prophage sequences. This applies to intracellular parasites in  $\gamma$ -proteobacteria (three genomes of *Buchnera* [204]), high-G+C gram-positive bacteria (*Mycobacterium leprae* [45]), low-G+C gram-positive bacteria (four genomes of *Mycoplasma* [179]) and endocellular symbionts (*Wigglesworthia glossinidia*, a  $\gamma$ -proteobacterium residing in tsetse flies [2]). All these genomes experienced a strong selection for genome reduction to only about 0.6 Mb, thereby drastically eliminating many genes and leaving no place for horizontally acquired DNA. In addition, within the confines of a eukaryotic cell, the

intracellular bacterium will experience few if any encounters with bacteriophages. Another trend is the lower prophage content in high-G+C gram-positive bacteria: four contained prophages, but only in the form of short prophage remnants, while three lacked prophages altogether (the numbers for prophage-containing versus prophage-lacking genomes were 17 and 9, respectively, for low-G+C gram-positive bacteria and 19 and 5, respectively, for  $\gamma$ -proteobacteria). However, the small number of sequenced high-G+C gram-positive bacteria indicates that this trend might still be affected by a sampling bias. A lively illustration of this statistical effect is *S. pneumoniae*: none of the three sequenced strains contained prophages or prophage remnants (59, 93, 199), while epidemiological surveys demonstrate that the majority of the clinical isolates contain prophages (173, 181).

Apart from endocellular bacteria, the only sequenced  $\gamma$ -proteobacterium lacking prophages is *Pasteurella*. In low-G+C gram-positive bacteria, the number of prophage-lacking sequenced bacterial genomes is greater but still small: one strain of *S. agalactiae*, one strain of *S. mutans*, three strains of *S. pneumoniae*, and the sequenced strain of *S. epidermidis* were devoid of prophages.

#### Age of the Prophages

Does comparative prophage genomics allow inferences on the residence time of prophage genomes? In many evolutionary analyses, one assumes tacitly that the very presence of a gene implies that it must confer a selectable function. That is, the gene could only have remained in the population if the encoded function is important. Evolutionary biologists have addressed the question of how important a function must be to ensure the maintenance of its gene (117). There is, however, a caveat to the argument. Genes without selective value might be encountered when they have just been introduced and have escaped selection because the process of gene elimination has not yet run its full course. Prophages seem to belong to this category. There is not much evidence that specific prophages are old parts of bacterial genomes. If they were, one would expect to encounter many examples of prophage remnants to be shared by most, if not all, strains within a given bacterial species. This was not observed. For example, *S. pyogenes* strains Manfredo and SF370 shared a prophage remnant (370.4) showing identical flanking sequences resulting from a common deletion event (they differed internally for indels and gene replacements) (37). In two other sequenced *S. pyogenes* strains, however, this particular prophage remnant was not detected (Fig. 11). Indeed, epidemiological data on *S. pyogenes* phages revealed that some prophage combinations were acquired over recent decades (8). A closer inspection of the prophages in the different *E. coli* genomes depicted in Fig. 1C demonstrated that close sequence relatedness was restricted mainly to the two closely related O157 strains, which still differed substantially in prophage content. Comparative prophage analysis in bacterial species that were sequenced in multiple strains (*Streptococcus*, *Staphylococcus*, and *Salmonella* strains) or in sister species (*Listeria*, *Xanthomonas*, and *Xylella*) all provided arguments against the maintenance of a given prophage profile over longer evolutionary times. Furthermore, microarray analysis and PCR scanning demonstrated that pro-

phages were generally strain specific within a given bacterial species (7, 13, 15, 39, 63, 150, 161, 169). In fact, there are examples from both pathogenic and commensal bacterial species where prophages represented the majority of the strain-specific DNA. All these observations speak against the antiquity of the currently encountered prophages in the sequenced bacterial genomes.

#### Ephemeral Nature of Prophages?

The various forms of prophage sequences seen in the survey (inducible prophages, defective prophages, prophage remnants, and isolated phage genes) argue for a dynamic model of phage-bacterial genome coexistence that fits some theoretical hypotheses relatively well (53, 115, 116). This relatively short-term association (in evolutionary terms) of a given phage DNA element with the host chromosome does not necessarily argue for an ephemeral role of prophage DNA for bacterial genome evolution. Clinical and epidemiological studies with many bacterial pathogens revealed the importance of prophage genes for the pathogenic properties of the given bacterial strain (e.g., *S. pyogenes*, *S. aureus*, *V. cholerae*, and *E. coli* O157). In other pathogens, the role of prophage genes for bacterial virulence was demonstrated in animal experiments (e.g., *S. enterica* serovar Typhimurium [for recent references, see reference 22]). Prophages apparently play an important part in the evolution of bacterial pathogenicity. Interestingly, comparative prophage genomics also showed "extra genes" in prophages from free-living bacteria or commensals, frequently at genome positions occupied by lysogenic conversion genes in pathogenic bacteria (209). However, only in exceptional cases (e.g., the *B. halodurans* prophage) could *in silico* analysis suggest a selective value for these extra genes. The majority of these extra genes in prophages from nonpathogenic bacteria lack database matches, and further experiments are needed to confirm or exclude a selective value associated with these extra genes.

#### Prophage Genome Diversity

Phage genomics confirmed the concepts of the modular theory of phage evolution. According to that theory, phage genomes are mosaics of modules (groups of functionally related genes) that are free to recombine in genetic exchanges between distinct phages infecting the same cell (20). One school of thought maintains that recombination essentially occurs everywhere and the apparent modular structure is the result of selection rejecting all genetic recombinations that do not lead to viable phage combinations (101). Selection also works against all recombinations that are less competitive than the existing phage types. Another group thinks that conserved linkers between individual modules are the sites of homologous recombination (43). Integration of the prophage into the bacterial chromosome liberates the phage DNA sequences from selection pressure on replicating phages. This change in selection pressure becomes evident from the survey of the prophage genomes. *In silico* analysis of the bacterial genomes suggested that the majority of the prophages are defective and apparently in a dynamic process of gradual decay. In the bacterial chromosomes, modular recombinations between different phage types are observed at a frequency not encountered in extracel-



lular phages. In  $\gamma$ -proteobacteria, prophages combining modules from lambda-, P2-, and Mu-like phages occurred in many different combinations. Even genomes only known from virulent phages, e.g., phages resembling *E. coli* phage T7, were observed in prophages. Similar observations were made for low-G+C gram-positive bacteria (bIL170-related streptococcal prophages). In addition, other mobile elements like transposons mixed with prophages, as demonstrated by the frequent observation of transposases in prophages. Apparently, the bacterial cell is a melting pot of diverse genetic elements. Genetic recombination between phages can lead to new chimeric phage types. In addition, phages can assume new functions and become phage-derived bacteriocins or gene transfer agents. Apparently, pathogenicity islands also have peculiar relationships with phages (178). These observations can be rationalized by the fact that the integrated prophage DNA becomes subject to the selective forces working on the bacterial chromosome. It is thus not surprising that prophages acquire new functions contributing to the fitness of the bacterial host into which they are integrated.

Even severely deleted prophage remnants can still provide interesting genetic material to a superinfecting phage and thus also contribute to the fitness of extracellular phages. For example, a phage integrase could lead the recombinant phage to new attachment sites, a distinct tail fiber cluster permits an altered host range to the recombinant phage, and a valuable lysogenic conversion gene increases the ecological spread of the recombinant temperate phage via the positive selective value it provides to its next lysogenic host. It is interesting that these genetic spare parts still come in small functional gene clusters, providing, for example, in the case of the *Pseudomonas* pyocin-prophage remnant, alternative solutions to the problem of new tail genes.

### Detecting Prophages

This genetic fluidity of prophages leads to the nontrivial practical problem of how to detect prophages in sequenced bacterial genomes. None of the phage genes is sufficiently specific or sufficiently highly conserved to serve as a single marker for prophages. Empirically, when screening a bacterial genome, we started with the search for integrases, which are easily detected and thus commonly annotated in the GenBank files. As a next step, we searched for a sequence approximately one phage genome distance upstream of the integrase that exactly matched a sequence directly downstream of the integrase. This process frequently defined the *attL* and *attR* sites of the putative prophage. Next, we inspected the predicted gene map for an unusually long ORF. This procedure often identified the tail tape measure gene (and sometimes also the anti-receptor gene). The genes preceding the tape gene showed, especially in low-G+C gram-positive bacteria, a characteristic length order, allowing a tentative diagnosis of prophage structural genes, which was then confirmed by BLAST matches to known phage or prophage genes. In this context, the large-subunit terminase and the portal protein turned out to be reasonably well-conserved phage proteins. Absence of matches to bacterial genes is a further criterion, but it should be used cautiously: Hypothetical bacterial genes might be unrecog-

nized prophages, and definitive bacterial genes might still be lysogenic conversion genes.

### Future Tasks

Prophage genomics has an impact on the field of phage-bacterium interaction at the genetic level. It can provide data on the genetic rules that underlie the arms race between the host bacterium and the infecting parasite and that underlie the postulated mutualistic interaction of prophages with the lysogen. We need more genome sequences from nonpathogenic bacteria inhabiting varied but defined environments to determine whether prophages also contribute factors with survival value to the lysogenic host. This will allow us to assess whether virulence factors are just the tip of an iceberg of prophage genes contributing a fitness increase to bacterial cells of any ecological affinity. If this is the case, prophage genomics could lead to important genes for the ecological adaptation of bacterial commensals and symbionts. Interestingly, bacteriologists increasingly see pathogenic, commensal, and symbiotic bacteria as steps in a continuum of bacterial adaptations to its hosts (92).

Another important area of research influenced by prophage genomics is the control of prophage gene expression. It is commonly anticipated that a prophage senses the physiological state of its host bacterium, which in turn influences its decision to induce a new replication cycle or to stay silent. Apparently, the expression of many lysogenic conversion genes from prophages is tightly controlled (214). We have here a mixture of phage-derived transcription control systems (e.g., Q antitermination) and systems where prophage genes respond directly or indirectly to molecular cues emitted by the mammalian host of the pathogenic bacterium (STEC and *S. pyogenes*) (for more references, see those in references 22 and 215). However, there are also more mundane tasks. One is the extension of the current phage database to include prophage sequences extracted from the different bacterial genome sequencing projects. This extended database will not only substantially increase the current phage database (probably more than double it) but also provide a greater phylogenetic breadth to the present database (30). Even more importantly, such a database will be a valuable tool for biologists annotating new phage, prophage, and bacterial genome sequences. We also need a terminology for the prophages, for which we might build on the example of the terminology for pathogenicity islands by combining an abbreviation for the bacterial species followed by PP for prophage and a number locating the prophage on the bacterial genome map starting at the origin of replication. Since the definition of a prophage is not clear-cut, this should be done by a subcommittee of the International Committee on Taxonomy of Viruses representing experts on phage genomics in different bacterial systems.

### ACKNOWLEDGMENTS

We thank the Swiss National Science Foundation for providing financial support for Carlos Canchaya (research grant 5002-057832). We also thank Nestlé for financial support of students in bioinformatics (C. Proux and G. Fournous) and for its encouragement in genome research.

We thank D. Pridmore for reading the manuscript.

## REFERENCES

- Ackermann, H. W. 1998. Tailed bacteriophages: the order caudovirales. *Adv. Virus Res.* **51**:135–201.
- Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat. Genet.* **32**:402–407.
- Allison, G. E., D. Angeles, N. Tran-Dinh, and N. K. Verma. 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. *J. Bacteriol.* **184**:1974–1987.
- Auvray, F., M. Coddeville, R. C. Ordenez, and P. Ritzenthaler. 1999. Unusual structure of the *attB* site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. *J. Bacteriol.* **181**:7385–7389.
- Auvray, F., M. Coddeville, P. Ritzenthaler, and L. Dupont. 1997. Plasmid integration in a wide range of bacteria mediated by the integrase of *Lactobacillus delbrueckii* bacteriophage mv4. *J. Bacteriol.* **179**:1837–1845.
- Ayora, S., R. Missich, P. Mesa, R. Lurz, S. Yang, E. H. Egelman, and J. C. Alonso. 2002. Homologous-pairing activity of the *Bacillus subtilis* bacteriophage SPP1 replication protein G35P. *J. Biol. Chem.* **277**:35969–35979.
- Baba, T., F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, H. Kuroda, L. Cui, K. Yamamoto, and K. Hiramatsu. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**:1819–1827.
- Banks, D. J., S. B. Beres, and J. M. Musser. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol.* **10**:515–521.
- Barondess, J. J., and J. Beckwith. 1990. A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* **346**:871–874.
- Barondess, J. J., and J. Beckwith. 1995. *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J. Bacteriol.* **177**:1247–1253.
- Bensing, B. A., I. R. Siboo, and P. M. Sullam. 2001. Proteins PblA and PblB of *Streptococcus mitis*, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. *Infect. Immun.* **69**:6186–6192.
- Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitz, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141–147.
- Beres, S. B., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarella, M. Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K. McCormick, D. Y. Leung, P. M. Schlievert, and J. M. Musser. 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci. USA* **99**:10078–10083.
- Bentley, M. J., and J. J. Mekalanos. 1985. Staphylococcal enterotoxin A is encoded by phage. *Science* **229**:185–187.
- Bhattacharyya, A., S. Stilwagner, N. Ivanova, M. D'Souza, A. Bernal, A. Lykidis, V. Kapatral, I. Anderson, N. Larsen, T. Los, G. Reznik, Selkov E Jr, T. L. Walunas, H. Feil, W. S. Feil, A. Purcell, J. L. Lassez, T. L. Hawkins, R. Haselkorn, R. Overbeek, P. F. Predki, and N. C. Kyriades. 2002. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc. Natl. Acad. Sci. USA* **99**:12403–12408.
- Bibb, L. A., and G. F. Hatfull. 2002. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. *Mol. Microbiol.* **45**:1515–1526.
- Bishai, W.R., and J. R. Murphy. 1988. Bacteriophage gene products that cause human disease, p. 683–724. *In* R. Calendar (ed.), *The bacteriophages*. Plenum Press, New York, N.Y.
- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1474.
- Boizet, B., Y. Lahbib-Mansais, L. Dupont, P. Ritzenthaler, and M. Mata. 1990. Cloning, expression and sequence analysis of an endolysin-encoding gene of *Lactobacillus bulgaricus* bacteriophage mv1. *Gene* **94**:61–67.
- Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann. N.Y. Acad. Sci.* **354**:484–490.
- Boyce, J. D., B. E. Davidson, and A. J. Hillier. 1995. Identification of prophage genes expressed in lysogens of the *Lactococcus lactis* bacteriophage BK5-T. *Appl. Environ. Microbiol.* **61**:4099–4104.
- Boyd, E. F., and H. Brüssow. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* **10**:521–529.
- Breuner, A., L. Brondsted, and K. Hammer. 1999. Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J. Bacteriol.* **181**:7291–7297.
- Brondsted, L., and K. Hammer. 1999. Use of the integration elements encoded by the temperate lactococcal bacteriophage TP901-1 to obtain chromosomal single-copy transcriptional fusions in *Lactococcus lactis*. *Appl. Environ. Microbiol.* **65**:752–758.
- Brondsted, L., S. Ostergaard, M. Pedersen, K. Hammer, and F. K. Vogensen. 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. *Virology* **283**:93–109.
- Broudy, T. B., V. Pancholi, and V. A. Fischetti. 2001. Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells. *Infect. Immun.* **69**:1440–1443.
- Brown, K. L., G. J. Sarkis, C. Wadsworth, and G. F. Hatfull. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. *EMBO J.* **16**:5914–5921.
- Brüssow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol. Microbiol.* **39**:213–222.
- Brüssow, H., M. Fremont, A. Bruttin, J. Sidoti, A. Constable, and V. Fryder. 1994. Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. *Appl. Environ. Microbiol.* **60**:4537–4543.
- Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. *Cell* **108**:13–16.
- Bruttin, A., F. Desiere, N. d'Amico, J. P. Guerin, J. Sidoti, B. Huni, S. Lucchini, and H. Brüssow. 1997. Molecular ecology of *Streptococcus thermophilus* bacteriophage infections in a cheese factory. *Appl. Environ. Microbiol.* **63**:3144–3150.
- Bruttin, A., F. Desiere, S. Lucchini, S. Foley, and H. Brüssow. 1997. Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage phi Sfi21. *Virology* **233**:136–148.
- Bruttin, A., S. Foley, and H. Brüssow. 1997. The site-specific integration system of the temperate *Streptococcus thermophilus* bacteriophage phiSfi21. *Virology* **237**:148–158.
- Bruttin, A., S. Foley, and H. Brüssow. 2002. DNA-binding activity of the *Streptococcus thermophilus* phage Sfi21 repressor. *Virology* **303**:100–109.
- Bushman, F. 2002. Lateral DNA transfer: mechanisms and consequences. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Campbell, A. M. 1996. Cryptic prophages, p. 2041–2046. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasarik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.) *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM, press, Washington, D.C.
- Canchaya, C., F. Desiere, W. M. McShan, J. J. Ferretti, J. Parkhill, and H. Brüssow. 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. *Virology* **302**:245–258.
- Carroll, D., M. A. Kehoe, D. Cavanagh, and D. C. Coleman. 1995. Novel organization of the site-specific integration and excision recombination functions of the *Staphylococcus aureus* serotype F virulence-converting phages phi 13 and phi 42. *Mol. Microbiol.* **16**:877–893.
- Chan, K., S. Baker, C. C. Kim, C. S. Detweiler, G. Dougan, and S. Falkow. 2003. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. *J. Bacteriol.* **185**:553–563.
- Cheetham, B. F., D. B. Tattersall, G. A. Bloomfield, J. I. Rood, and M. E. Katz. 1995. Identification of a gene encoding a bacteriophage-related integrase in a *vap* region of the *Dichelobacter nodosus* genome. *Gene* **162**:53–58.
- Chen, F., and J. Lu. 2002. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* **68**:2589–2594.
- Chopin, A., A. Bolotin, A. Sorokin, S. D. Ehrlich, and M. Chopin. 2001. Analysis of six prophages in *Lactococcus lactis* IL 1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29**:644–651.
- Clark, A. J., W. Inwood, T. Cloutier, and T. S. Dhillon. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdaoid phages. *J. Mol. Biol.* **311**:657–679.
- Coffey, A., and R. P. Ross. 2002. Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. *Antonie Leeuwenhoek* **82**:303–321.
- Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Fretwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. *Nature* **409**:1007–1011.
- Coleman, D. C., D. J. Sullivan, R. J. Russell, J. P. Arbutnot, B. F. Carey,

- and H. M. Pomeroy. 1989. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. *J. Gen. Microbiol.* **135**:1679–1697.
47. Crutz-Le Coq, A. M., B. Cesselin, J. Commissaire, and J. Anba. 2002. Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. *Microbiology* **148**:985–1001.
  48. da Silva, A. C., J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. Alves, A. M. do Amaral, M. C. Bertolini, L. E. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chambergo, L. P. Ciapina, R. M. Cicarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El Dorry, J. B. Faria, A. J. Ferreira, R. C. Ferreira, M. I. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. Lemos, M. V. Lemos, E. C. Locali, M. A. Machado, A. M. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. Sena, C. Silva, R. F. de Souza, L. A. Spinola, M. A. Takita, R. E. Tamura, E. C. Teixeira, R. I. Tezza, S. M. Trindade dos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal, and J. P. Kitajima. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**:459–463.
  49. Davis, B. M., and M. K. Waldor. 2000. CTXphi contains a hybrid genome derived from tandemly integrated elements. *Proc. Natl. Acad. Sci. USA* **97**:8572–8577.
  50. Deng, W., V. Burland, G. Plunkett, III, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner, and R. D. Perry. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**:4601–4611.
  51. Desiere, F., S. Lucchini, and H. Brüssow. 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate *Streptococcus thermophilus* bacteriophage Sfi21. *Virology* **260**:244–253.
  52. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brüssow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* **283**:240–252.
  53. Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brüssow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic *Streptococci*: evolutionary implications for prophage-host interactions. *Virology* **288**:325–341.
  54. Desiere, F., R. D. Pridmore, and H. Brüssow. 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. *Virology* **275**:294–305.
  55. Diaz, E., R. Lopez, and J. L. Garcia. 1992. EJ-1, a temperate bacteriophage of *Streptococcus pneumoniae* with a *Myoviridae* morphotype. *J. Bacteriol.* **174**:5516–5525.
  56. Di Pierro, M., R. Lu, S. Uzzau, W. Wang, K. Margaretten, C. Pazzani, F. Maimone, and A. Fasano. 2001. Zonula occludens toxin structure-function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. *J. Biol. Chem.* **276**:19160–19165.
  57. Dodd, I. B., and J. B. Egan. 1996. The *Escherichia coli* retrons Ec67 and Ec86 replace DNA between the *cos* site and a transcription terminator of a 186-related prophage. *Virology* **219**:115–124.
  58. Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* **284**:2124–2129.
  59. Dopazo, J., A. Mendoza, J. Herrero, F. Caldara, Y. Humbert, L. Friedli, M. Guerrier, E. Grand-Schenck, C. Gandin, M. de Francesco, A. Polissi, G. Buell, G. Feger, E. Garcia, M. Peitsch, and J. F. Garcia-Bustos. 2001. Annotated draft genomic sequence from a *Streptococcus pneumoniae* type 19F clinical isolate. *Microb. Drug Resist.* **7**:99–125.
  60. Dozois, C. M., F. Daigle, and R. Curtiss, III. 2003. Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. *Proc. Natl. Acad. Sci. USA*. **100**:247–252.
  61. Duplessis, M., and S. Moineau. 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Mol. Microbiol.* **41**:325–336.
  62. Durmaz, E., and T. R. Klaenhammer. 2000. Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. *Appl. Environ. Microbiol.* **66**:895–903.
  63. Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos. 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. USA* **99**:1556–1561.
  64. Edwards, R. A., G. J. Olsen, and S. R. Maloy. 2002. Comparative genomics of closely related *salmonellae*. *Trends Microbiol.* **10**:94–99.
  65. Eggers, C. H., B. J. Kimmel, J. L. Bono, A. F. Elias, P. Rosa, and D. S. Samuels. 2001. Transduction by phiBB-1, a bacteriophage of *Borrelia burgdorferi*. *J. Bacteriol.* **183**:4771–4778.
  66. Eklund, M. W., F. T. Poysky, and S. M. Reed. 1972. Bacteriophage and the toxigenicity of *Clostridium botulinum* type D. *Nat. New Biol.* **235**:16–17.
  67. Eklund, M. W., F. T. Poysky, S. M. Reed, and C. A. Smith. 1971. Bacteriophage and the toxigenicity of *Clostridium botulinum* type C. *Science* **172**:480–482.
  68. Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. USA* **98**:4658–4663.
  69. Figueroa-Bossi, N., and L. Bossi. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. *Mol. Microbiol.* **33**:167–176.
  70. Figueroa-Bossi, N., S. Uzzau, D. Maloroli, and L. Bossi. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol. Microbiol.* **39**:260–271.
  71. Foley, S., A. Bruttin, and H. Brüssow. 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *J. Virol.* **74**:611–618.
  72. Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. *J. Mol. Biol.* **279**:143–164.
  73. Ford, M. E., Stenstrom, C., R. W. Hendrix, and G. F. Hatfull. 1998. Mycobacteriophage TM4: genome structure and gene expression. *Tubercle Lung Dis.* **79**:63–73.
  74. Garcia, E., J. L. Garcia, P. Garcia, A. Arraras, J. M. Sanchez-Puelles, and R. Lopez. 1988. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Proc. Natl. Acad. Sci. USA* **85**:914–918.
  75. Garcia, P., J. C. Alonso, and J. E. Suarez. 1997. Molecular analysis of the *cos* region of the *Lactobacillus casei* bacteriophage A2. Gene product 3, gp3, specifically binds to its downstream *cos* region. *Mol. Microbiol.* **23**:505–514.
  76. Garcia, P., V. Ladero, J. C. Alonso, and J. E. Suarez. 1999. Cooperative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. *J. Virol.* **73**:3920–3929.
  77. Gerdes, K. 2000. Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bacteriol.* **182**:561–572.
  78. Gindreau, E., R. Lopez, and P. Garcia. 2000. MM1, a temperate bacteriophage of the type 23F Spanish/USA multiresistant epidemic clone of *Streptococcus pneumoniae*: structural analysis of the site-specific integration system. *J. Virol.* **74**:7803–7813.
  79. Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Cheoutani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fsihi, F. G. Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapatk, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart. 2001. Comparative genomics of *Listeria* species. *Science* **294**:849–852.
  80. Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot, and F. Kunst. 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol. Microbiol.* **45**:1499–1513.
  81. Haggard-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J. Bacteriol.* **174**:1462–1477.
  82. Halperin, S. A., P. Ferrieri, E. D. Gray, E. L. Kaplan, and L. W. Wannamaker. 1987. Antibody response to bacteriophage hyaluronidase in acute glomerulonephritis after group A streptococcal infection. *J. Infect. Dis.* **155**:253–261.
  83. Hambly, E., F. Tetart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann. 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. *Proc. Natl. Acad. Sci. USA* **98**:11411–11416.
  84. Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* **7**:395–405.
  85. Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleischmann, W. C. Nierman, and O. White. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477–483.
  86. Heidelberg, J. F., I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu,

- J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Neelson, and C. M. Fraser. 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat. Biotechnol.* **20**:1118–1123.
87. Heinrich, J., M. Velleman, and H. Schuster. 1995. The tripartite immunity system of phages P1 and P7. *FEMS Microbiol. Rev.* **17**:121–126.
88. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8**:504–508.
89. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* **96**:2192–2197.
90. Hill, D. F., N. J. Short, R. N. Perham, and G. B. Petersen. 1991. DNA sequence of the filamentous bacteriophage Pfl. *J. Mol. Biol.* **218**:349–364.
91. Ho, T. D., and J. M. Slouch. 2001. Characterization of *gvrA*, an antiviral gene on the gifsy-2 phage in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **183**:611–620.
92. Hooper, L. V., and J. I. Gordon. 2001. Commensal host-bacterial relationships in the gut. *Science* **292**:1115–1118.
93. Hoskins, J., W. E. Alborn, Jr., J. Arnold, L. C. Blaszczyk, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenry, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, Jr., P. L. Skatrud, and J. I. Glass. 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183**:5709–5717.
94. Huber, K. E., and M. K. Waldor. 2002. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature* **417**: 656–659.
95. Iandolo, J. J., V. Worrell, K. H. Groicher, Y. Qian, R. Tian, S. Kenton, A. Dorman, H. Ji, S. Lin, P. Loh, S. Qi, H. Zhu, and B. A. Roe. 2002. Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene* **289**:109–118.
96. Ikebe, T., A. Wada, Y. Inagaki, K. Sugama, R. Suzuki, D. Tanaka, A. Tamaru, Y. Fujinaga, Y. Abe, Y. Shimizu, and H. Watanabe. 2002. Dissemination of the phage-associated novel superantigen gene *speL* in recent invasive and noninvasive *Streptococcus pyogenes* M3/T3 isolates in Japan. *Infect. Immun.* **70**:3227–3233.
97. Inagaki, Y., F. Myouga, H. Kawabata, S. Yamai, and H. Watanabe. 2000. Genomic differences in *Streptococcus pyogenes* serotype M3 between recent isolates associated with toxic shock-like syndrome and past clinical isolates. *J. Infect. Dis.* **181**:975–983.
98. Jimenez, J. I., B. Minambres, J. L. Garcia, and E. Diaz. 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**:824–841.
99. Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou, and J. Yu. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.* **30**:4432–4441.
100. Johnsen, M. G., H. Neve, F. K. Vogensen, and K. Hammer. 1995. Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. *Virology* **212**:595–606.
101. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdaoid bacteriophages. *J. Mol. Biol.* **299**:27–51.
102. Kakikawa, M., S. Ohkubo, T. Sakate, M. Sayama, A. Taketo, and K. Kodaira. 2000. Purification and DNA-binding properties of the cro-type regulatory repressor protein *cng* encoded by the *Lactobacillus plantarum* phage phi g1e. *Gene* **249**:161–169.
103. Kaneko, J., T. Kimura, S. Narita, T. Tomita, and Y. Kamio. 1998. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Pantone-Valentine leukocidin genes. *Gene* **215**:57–67.
104. Karaolis, D. K., S. Somara, D. R. Maneval, Jr., J. A. Johnson, and J. B. Kaper. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**:375–379.
105. Kazmi, S. U., R. Kansal, R. K. Aziz, M. Hooshdaran, A. Norrby-Teglund, D. E. Low, A. B. Halim, and M. Kotb. 2001. Reciprocal, temporal expression of *SpeA* and *SpeB* by invasive MIT1 group streptococcal isolates in vivo. *Infect. Immun.* **69**:4988–4995.
106. Kleerebezem, M., J. Boekhorst, R. Van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. De Vries, B. Ursing, W. M. De Vos, and R. J. Siezen. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci. USA* **100**:1990–1995.
107. Kropinski, A. M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J. Bacteriol.* **182**: 6066–6074.
108. Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Conner-ton, A. Danchin, and . 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**:249–256.
109. Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu. 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**:1225–1240.
110. Ladero, V., P. Garcia, J. C. Alonso, and J. E. Suarez. 1999. A2 cro, the lysogenic cycle repressor, specifically binds to the genetic switch region of *Lactobacillus casei* bacteriophage A2. *Virology* **262**:220–229.
111. Ladero, V., P. Garcia, J. C. Alonso, and J. E. Suarez. 2002. Interaction of the Cro repressor with the lysis/lysogeny switch of the *Lactobacillus casei* temperate bacteriophage A2. *J. Gen. Virol.* **83**:2891–2895.
112. Laird, W., and N. Groman. 1976. Orientation of the tox gene in the prophage of corynebacteriophage beta. *J. Virol.* **19**:228–231.
113. Lang, A. S., and J. T. Beatty. 2000. Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc. Natl. Acad. Sci. USA* **97**:859–864.
114. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol.* **184**:4891–4905.
115. Lawrence, J. G., R. W. Hendrix, and S. Casjens. 2001. Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **9**:535–540.
116. Lawrence, J. G. and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95**:9413–9417.
117. Lawrence, J. G., and J. R. Roth. 1999. Genomic flux: genome evolution by gene loss and acquisition, p. 263–289. *In* R. L. Charlebois (ed.), *Organization of the prokaryotic genome*. ASM Press, Washington, D.C.
118. Lazarevic, V., A. Dusterhoff, B. Soldo, H. Hilbert, C. Mauel, and D. Karata. 1999. Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPbetac2. *Microbiology* **145**:1055–1067.
119. Lee, C. Y., and J. J. Iandolo. 1986. Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. *J. Bacteriol.* **166**:385–391.
120. Lee, M. H., L. Pascopella, W. R. Jacobs, Jr., and G. F. Hatfull. 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. *Proc. Natl. Acad. Sci. USA* **88**:3111–3115.
121. Lee, Y., D. Choi, and H. Kende. 2001. Expansins: ever-expanding numbers and functions. *Curr. Opin. Plant Biol.* **4**:527–532.
122. Le Marrec, C., D. van Sinderen, L. Walsh, E. Stanley, E. Vlegels, S. Moineau, P. Heinze, G. Fitzgerald, and B. Fayard. 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. *Appl. Environ. Microbiol.* **63**:3246–3253.
123. Lin, N. T., R. Y. Chang, S. J. Lee, and Y. H. Tseng. 2001. Plasmids carrying cloned fragments of RF DNA from the filamentous phage (phi)Lf can be integrated into the host chromosome via site-specific integration and homologous recombination. *Mol. Genet. Genomics* **266**:425–435.
124. Lin, W., K. J. Fullner, R. Clayton, J. A. Sexton, M. B. Rogers, K. E. Calia, S. B. Calderwood, C. Fraser, and J. J. Mekalanos. 1999. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc. Natl. Acad. Sci. USA* **96**:1071–1076.
125. Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar. 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol. Microbiol.* **35**:324–340.
126. Lofdahl, S., J. Zabielski, and L. Philipson. 1981. Structure and restriction enzyme maps of the circularly permuted DNA of staphylococcal bacteriophage phi 11. *J. Virol.* **37**:784–794.
127. Lucchini, S., F. Desiere, and H. Brüssow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* **73**:8647–8656.
128. Lucchini, S., F. Desiere, and H. Brüssow. 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content gram-positive bacteria. *Virology* **263**:427–435.
129. Lucchini, S., F. Desiere, and H. Brüssow. 1999. The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of *cos*-site phages Sfi19 and Sfi21. *Virology* **260**: 232–243.
130. Lurz, R., E. V. Orlova, D. Gunther, P. Dube, A. Droge, F. Weise, M. van

- Heel, and P. Tavares. 2001. Structural organization of the head-to-tail interface of a bacterial virus. *J. Mol. Biol.* **310**:1027–1037.
131. Madsen, P. L., and K. Hammer. 1998. Temporal transcription of the lactococcal temperate phage TP901–1 and DNA sequence of the early promoter region. *Microbiology* **144**:2203–2215.
132. Madsen, P. L., A. H. Johansen, K. Hammer, and L. Brondsted. 1999. The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901–1. *J. Bacteriol.* **181**:7430–7438.
133. Madsen, S. M., D. Mills, G. Djordjevic, H. Israelsen, and T. R. Klaenhammer. 2001. Analysis of the genetic switch and replication region of a P335-type bacteriophage with an obligate lytic lifestyle on *Lactococcus lactis*. *Appl. Environ. Microbiol.* **67**:1128–1139.
134. Mahanivong, C., J. D. Boyce, B. E. Davidson, and A. J. Hillier. 2001. Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. *Appl. Environ. Microbiol.* **67**:3564–3576.
135. Makino, K., K. Yokoyama, Y. Kubota, C. H. Yutsudo, S. Kimura, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, T. Iida, K. Yamamoto, M. Onishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 1999. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. *Genes Genet. Syst.* **74**:227–239.
136. Maniloff, J., and H. W. Ackermann. 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. *Arch. Virol.* **143**:2051–2063.
137. Maniloff, J., G. J. Kampo, and C. C. Dascher. 1994. Sequence analysis of a unique temperature phage: mycoplasma virus L2. *Gene* **141**:1–8.
138. Mavris, M., P. A. Manning, and R. Morona. 1997. Mechanism of bacteriophage SfiI-mediated serotype conversion in *Shigella flexneri*. *Mol. Microbiol.* **26**:939–950.
139. McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**:852–856.
140. McGrath, S., G. F. Fitzgerald, and D. van Sinderen. 2002. Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. *Mol. Microbiol.* **43**:509–520.
141. Mediavilla, J., S. Jain, J. Kriakov, M. E. Ford, R. L. Duda, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2000. Genome organization and characterization of mycobacteriophage Bxb1. *Mol. Microbiol.* **38**:955–970.
142. Mikkonen, M., L. Dupont, T. Alatosava, and P. Ritzenthaler. 1996. Defective site-specific integration elements are present in the genome of virulent bacteriophage LL-H of *Lactobacillus delbrueckii*. *Appl. Environ. Microbiol.* **62**:1847–1851.
143. Miroid, S., W. Rabsch, M. Rohde, S. Stender, H. Tschape, H. Russmann, E. Igwe, and W. D. Hardt. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl. Acad. Sci. USA* **96**:9845–9850.
144. Miroid, S., W. Rabsch, H. Tschape, and W. D. Hardt. 2001. Transfer of the *Salmonella* type III effector *sopE* between unrelated phage families. *J. Mol. Biol.* **312**:7–16.
145. Moineau, S., S. Pandian, and T. Klaenhammer. 2002. Evolution of a lytic bacteriophage via DNA acquisition from the *Lactococcus lactis* chromosome. *Appl. Environ. Microbiol.* **68**:1832–1841.
146. Molineux I. J. 1999. T7-like phages (*Podoviridae*), p. 1722–1729. In A. Granoff and R. Webber (ed.), *Encyclopedia of virology*. Academic Press, Ltd., London, United Kingdom.
147. Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J. Mol. Biol.* **317**:337–359.
148. Moszer, I., E. P. Rocha, and A. Danchin. 1999. Codon usage and lateral gene transfer in *Bacillus subtilis*. *Curr. Opin. Microbiol.* **2**:524–528.
149. Muniesa, M., J. Recktenwald, M. Bielaszewska, H. Karch, and H. Schmidt. 2000. Characterization of a shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. *Infect. Immun.* **68**:4850–4855.
150. Murray, A. E., D. Lies, G. Li, K. Nealson, J. Zhou, and J. M. Tiedje. 2001. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc. Natl. Acad. Sci. USA* **98**:9853–9858.
151. Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi. 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol. Microbiol.* **31**:399–419.
152. Nakayama, K., K. Takashima, H. Ishihara, T. Shinomiya, M. Kageyama, S. Kanaya, M. Ohnishi, T. Murata, H. Mori, and T. Hayashi. 2000. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* **38**:213–231.
153. Narita, S., J. Kaneko, J. Chiba, Y. Piemont, S. Jarraud, J. Etienne, and Y. Kamio. 2001. Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, phiSLT. *Gene* **268**:195–206.
154. Nauta, A., B. B. van den, H. Karsens, G. Venema, and J. Kok. 1997. Design of thermolabile bacteriophage repressor mutants by comparative molecular modeling. *Nat. Biotechnol.* **15**:980–983.
155. Nauta, A., D. van Sinderen, H. Karsens, E. Smit, G. Venema, and J. Kok. 1996. Inducible gene expression mediated by a repressor-operator system isolated from *Lactococcus lactis* bacteriophage r1t. *Mol. Microbiol.* **19**:1331–1341.
156. Neely, M. N., and D. I. Friedman. 1998. Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol. Microbiol.* **28**:1255–1267.
157. Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull. 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. *Mol. Microbiol.* **17**:1045–1056.
158. Neve, H., K. I. Zenz, F. Desiere, A. Koch, K. J. Heller, and H. Brüssow. 1998. Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: implications for the modular theory of phage evolution. *Virology* **241**:61–72.
159. Nolling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**:4823–4838.
160. Ohnishi, M., K. Kurokawa, and T. Hayashi. 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol.* **9**:481–485.
161. Ohnishi, M., J. Terajima, K. Kurokawa, K. Nakayama, T. Murata, K. Tamura, Y. Ogura, H. Watanabe, and T. Hayashi. 2002. Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc. Natl. Acad. Sci. USA* **99**:17043–17048.
162. Oki, M., M. Kakikawa, S. Nakamura, E. T. Yamamura, K. Watanabe, M. Sasamoto, A. Taketo, and K. Kodaira. 1997. Functional and structural features of the holin HOL protein of the *Lactobacillus plantarum* phage phi gl: analysis in *Escherichia coli* system. *Gene* **197**:137–145.
163. Pajunen, M. L., S. J. Kiljunen, M. E. Soderholm, and M. Skurnik. 2001. Complete genomic sequence of the lytic bacteriophage phiYeO3–12 of *Yersinia enterocolitica* serotype O:3. *J. Bacteriol.* **183**:1928–1937.
164. Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connor, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Fratwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848–852.
165. Pedersen, M., S. Ostergaard, J. Bresciani, and F. K. Vogensen. 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901–1 involved in tail length determination and baseplate assembly. *Virology* **276**:315–328.
166. Perna, N. T., G. Plunkett, III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grobeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**:529–533.
167. Pfister, P., A. Wasserfallen, R. Stettler, and T. Leisinger. 1998. Molecular analysis of *Methanobacterium* phage psiM2. *Mol. Microbiol.* **30**:233–244.
168. Plunkett, G., III, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* **181**:1767–1778.
169. Porwollik, S., R. M. Wong, and M. McClelland. 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **99**:8956–8961.
170. Proft, T., S. L. Moffatt, C. J. Berkahn, and J. D. Fraser. 1999. Identification and characterization of novel superantigens from *Streptococcus pyogenes*. *J. Exp. Med.* **189**:89–102.
171. Proux, C., D. van Sinderen, J. Suarez, P. Garcia, V. Ladero, G. F. Fitzgerald, F. Desiere, and H. Brüssow. 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. *J. Bacteriol.* **184**:6026–6036.
172. Pulkkinen, W. S., and S. I. Miller. 1991. A *Salmonella typhimurium* virulence protein is similar to a *Yersinia enterocolitica* invasion protein and a bacteriophage lambda outer membrane protein. *J. Bacteriol.* **173**:86–93.
173. Ramirez, M., E. Severina, and A. Tomasz. 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. *J. Bacteriol.* **181**:3618–3625.
174. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W.

- Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J. Mol. Biol.* **299**:53–73.
175. Recktenwald, J., and H. Schmidt. 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. *Infect. Immun.* **70**:1896–1908.
176. Rohwer, F. and R. Edwards. 2002. The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* **184**:4529–4535.
177. Romero, A., R. Lopez, R. Lurz, and P. Garcia. 1990. Temperate bacteriophages of *Streptococcus pneumoniae* that contain protein covalently linked to the 5' ends of their DNA. *J. Virol.* **64**:5149–5155.
178. Ruzin, A., J. Lindsay, and R. P. Novick. 2001. Molecular genetics of SaPI1—a mobile pathogenicity island in *Staphylococcus aureus*. *Mol. Microbiol.* **41**:365–377.
179. Sasaki, Y., J. Ishikawa, A. Yamashita, K. Oshima, T. Kenri, K. Furuya, C. Yoshino, A. Horino, T. Shiba, T. Sasaki, and M. Hattori. 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Res.* **30**:5293–5300.
180. Schell, M. A., M. Karmirantzou, B. Snel, D. Vilanova, B. Berger, G. Pessi, M. C. Zwahlen, F. Desiere, P. Bork, M. Delley, R. D. Priddy, and F. Arigoni. 2002. The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. USA* **99**:14422–14427.
181. Severina, E., M. Ramirez, and A. Tomasz. 1999. Prophage carriage as a molecular epidemiological marker in *Streptococcus pneumoniae*. *J. Clin. Microbiol.* **37**:3308–3315.
182. Sharples, G. J., E. L. Bolt, and R. G. Lloyd. 2002. RusA proteins from the extreme thermophile *Aquifex aeolicus* and lactococcal phage r1t resolve Holliday junctions. *Mol. Microbiol.* **44**:549–559.
183. Sheehan, M. M., J. L. Garcia, R. Lopez, and P. Garcia. 1997. The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Mol. Microbiol.* **25**:717–725.
184. Sheehan, M. M., E. Stanley, G. F. Fitzgerald, and D. van Sinderen. 1999. Identification and characterization of a lysis module present in a large proportion of bacteriophages infecting *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **65**:569–577.
185. Shimizu, T., K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, and H. Hayashi. 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. USA* **99**:996–1001.
186. Shinomiya, T. 1984. Phenotypic mixing of pyocin R2 and bacteriophage PS17 in *Pseudomonas aeruginosa* PAO. *J. Virol.* **49**:310–314.
187. Simpson, A. J., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. Briones, M. R. Bueno, A. A. Camargo, L. E. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El Dorry, A. P. Facincani, A. J. Ferreira, V. C. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. Leite, E. G. Lemos, M. V. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. Madeira, H. M. Madeira, C. L. Marino, M. V. Marques, E. A. Martins, E. M. Martins, A. Y. Matsukuma, C. F. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteiro-Vitorello, D. H. Moon, M. A. Nagai, A. L. Nascimento, L. E. Netto, A. Nhani, Jr., F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. Pereira, H. A. Pereira, Jr., J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J. de Rosa, V. E. de Rosa, Jr., R. G. de Sa, R. V. Santelli, H. E. Sawasaki, A. C. da Silva, A. M. da Silva, F. R. da Silva, W. A. Silva, J. F. da Silveira, M. L. Silvestri, W. J. Siqueira, A. A. de Souza, A. P. de Souza, M. F. Terenzi, D. Truffi, S. M. Tsai, M. H. Tsuchi, H. Vallada, M. A. Van Sluys, S. Verjovski-Almeida, A. L. Vettore, M. A. Zago, M. Zatz, J. Meidanis, and J. C. Setubal. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**:151–157.
188. Smith, M. C., R. N. Burns, S. E. Wilson, and M. A. Gregory. 1999. The complete genome sequence of the *Streptomyces* temperate phage straight phiC31: evolutionary relationships to other viruses. *Nucleic Acids Res.* **27**: 2145–2155.
189. Smoot, J. C., K. D. Barbian, J. J. Van Gompel, L. M. Smoot, M. S. Chaussee, G. L. Sylva, D. E. Sturdevant, S. M. Ricklefs, S. F. Porcella, L. D. Parkins, S. B. Beres, D. S. Campbell, T. M. Smith, Q. Zhang, V. Kapur, J. A. Daly, L. G. Veasy, and J. M. Musser. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proc. Natl. Acad. Sci. USA* **99**:4668–4673.
190. Smoot, L. M., J. C. Smoot, M. R. Graham, G. A. Somerville, D. E. Sturdevant, C. A. Migliaccio, G. L. Sylva, and J. M. Musser. 2001. Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc. Natl. Acad. Sci. USA* **98**:10416–10421.
191. Sriskandan, S., M. Unnikrishnan, T. Krausz, and J. Cohen. 2000. Mitogenic factor (MF) is the major DNase of serotype M89 *Streptococcus pyogenes*. *Microbiology* **146**:2785–2792.
192. Stanley, E., G. F. Fitzgerald, C. Le Marrec, B. Fayard, and D. van Sinderen. 1997. Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiolology* **143**:3417–3429.
193. Stanley, T. L., C. D. Ellermeier, and J. M. Schlauch. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar Typhimurium survival in Peyer's patches. *J. Bacteriol.* **182**:4406–4413.
194. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959–964.
195. Tacket, C. O., G. Losonsky, J. P. Nataro, S. J. Cryz, R. Edelman, A. Fasano, J. Michalski, J. B. Kaper, and M. M. Levine. 1993. Safety and immunogenicity of live oral cholera vaccine candidate CVD 110, a delta ctxA delta zot delta ace derivative of El Tor Ogawa *Vibrio cholerae*. *J. Infect. Dis.* **168**: 1536–1540.
196. Takami, H., K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hiram, Y. Nakamura, N. Ogasawara, S. Kuhara, and K. Horikoshi. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.* **28**:4317–4331.
197. Takemaru, K., M. Mizuno, T. Sato, M. Takeuchi, and Y. Kobayashi. 1995. Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology* **141**:323–327.
198. Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Dougherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc. Natl. Acad. Sci. USA* **99**:12391–12396.
199. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**:498–506.
200. Tremblay, D. M., and S. Moineau. 1999. Complete genome sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* **255**:63–76.
201. Unkmeir, A., and H. Schmidt. 2000. Structural analysis of phage-borne stx genes and their flanking sequences in Shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. *Infect. Immun.* **68**:4856–4864.
202. Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. *J. Bacteriol.* **182**:6472–6481.
203. van der Wilk, F., A. M. Dullemans, M. Verbeek, and J. F. van den Heuvel. 1999. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrtosiphon pisum*. *Virology* **262**:104–113.
204. van Ham, R. C., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, L. Jimenez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran, and A. Moya. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* **100**:581–586.
205. Van Kaer, L., Y. Gansemans, M. Van Montagu, and P. Dhaese. 1988. Interaction of the *Bacillus subtilis* phage phi 105 repressor DNA: a genetic analysis. *EMBO J.* **7**:859–866.
206. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Rutgers, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. *Mol. Microbiol.* **19**: 1343–1355.
207. Van Sluys, M. A., M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, L. E. Camargo, A. C. da Silva, D. H. Moon, M. A. Takita, E. G. Lemos, M. A. Machado, M. I. Ferro, F. R. da Silva, M. H. Goldman, G. H. Goldman, M. V. Lemos, H. El Dorry, S. M. Tsai, H. Carrer, D. M. Carraro, R. C. de Oliveira, L. R. Nunes, W. J. Siqueira, L. L. Coutinho, E. T. Kimura, E. S. Ferro, R. Harakava, E. E. Kuramae, C. L. Marino, E. Gigliotti, I. L. Abreu, L. M. Alves, A. M. do Amaral, G. S. Baia, S. R. Blanco,

- M. S. Brito, F. S. Cannavan, A. V. Celestino, A. F. da Cunha, R. C. Fenille, J. A. Ferro, E. F. Formighieri, L. T. Kishi, S. G. Leoni, A. R. Oliveira, V. E. Rosa, Jr., F. T. Sasaki, J. A. Sena, A. A. de Souza, D. Truffi, F. Tsukumo, G. M. Yanai, L. G. Zarus, E. L. Civerolo, A. J. Simpson, N. F. Almeida, Jr., J. C. Setubal, and J. P. Kitajima. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J. Bacteriol.* **185**:1018–1026.
208. Vasala, A., L. Dupont, M. Baumann, P. Ritzenthaler, and T. Alatossava. 1993. Molecular comparison of the structural proteins encoding gene clusters of two related *Lactobacillus delbrueckii* bacteriophages. *J. Virol.* **67**:3061–3068.
209. Ventura, M., A. Bruttin, C. Canchaya, and H. Brüssow. 2002. Transcription analysis of *Streptococcus thermophilus* phages in the lysogenic state. *Virology* **302**:21–32.
210. Ventura, M., S. Foley, A. Bruttin, S. C. Chennoufi, C. Canchaya, and H. Brüssow. 2002. Transcription mapping as a tool in phage genomics: the case of the temperate *Streptococcus thermophilus* phage Sfi21. *Virology* **296**:62–76.
211. Wagner, P. L., D. W. Acheson, and M. K. Waldor. 1999. Isogenic lysogens of diverse shiga toxin 2-encoding bacteriophages produce markedly different amounts of shiga toxin. *Infect. Immun.* **67**:6710–6714.
212. Wagner, P. L., D. W. Acheson, and M. K. Waldor. 2001. Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli*. *Infect. Immun.* **69**:1934–1937.
213. Wagner, P. L., J. Livny, M. N. Neely, D. W. Acheson, D. I. Friedman, and M. K. Waldor. 2002. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol. Microbiol.* **44**:957–970.
214. Wagner, P. L., M. N. Neely, X. Zhang, D. W. Acheson, M. K. Waldor, and D. I. Friedman. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. *J. Bacteriol.* **183**:2081–2085.
215. Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. *Infect. Immun.* **70**:3985–3993.
216. Waldor, M. K. 1998. Bacteriophage biology and bacterial virulence. *Trends Microbiol.* **6**:295–297.
217. Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**:1910–1914.
218. Weisberg, R. A., M. E. Gottesmann, R. W. Hendrix, and J. W. Little. 1999. Family values in the age of genomics: comparative analyses of temperate bacteriophage HK022. *Annu. Rev. Genet.* **33**:565–602.
219. Welch, R. A., V. Burland, G. Plunkett, III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Sonnenberg, and F. R. Blattner. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**:17020–17024.
220. Whiteley, M., M. G. Banger, R. E. Bumgarner, M. R. Parsek, G. M. Teitzel, S. Lory, and E. P. Greenberg. 2001. Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413**:860–864.
221. Wilson, S. E., C. J. Ingham, I. S. Hunter, and M. C. Smith. 1995. Control of lytic development in the *Streptomyces* temperate phage phi C31. *Mol. Microbiol.* **16**:131–143.
222. Winkler, K. C., J. de Waart, and C. Grooten. 1965. Lysogenic conversion of staphylococci to loss of beta-toxin. *J. Gen. Microbiol.* **39**:321–333.
223. Wood, H. E., M. T. Dawson, K. M. Devine, and D. J. McConnell. 1990. Characterization of PBSX, a defective prophage of *Bacillus subtilis*. *J. Bacteriol.* **172**:2667–2674.
224. Woods, D. E., J. A. Jeddelloh, D. L. Fritz, and D. DeShazer. 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. *J. Bacteriol.* **184**:4003–4017.
225. Yamaguchi, T., T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa, and M. Sugai. 2000. Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. *Mol. Microbiol.* **38**:694–705.
226. Yarwood, J. M., J. K. McCormick, M. L. Paustian, P. M. Orwin, V. Kapur, and P. M. Schlievert. 2002. Characterization and expression analysis of *Staphylococcus aureus* pathogenicity island 3. Implications for the evolution of staphylococcal pathogenicity islands. *J. Biol. Chem.* **277**:13138–13147.
227. Ye, Z. H., S. L. Buranen, and C. Y. Lee. 1990. Sequence analysis and comparison of int and xis genes from staphylococcal bacteriophages L54a and phi 11. *J. Bacteriol.* **172**:2568–2575.
228. Ye, Z. H., and C. Y. Lee. 1989. Nucleotide sequence and genetic characterization of staphylococcal bacteriophage L54a int and xis genes. *J. Bacteriol.* **171**:4146–4153.
229. Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, S. Kimura, C. H. Yutsudo, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, M. Yoh, T. Iida, M. Ohnishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 2000. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. *Gene* **258**:127–139.
230. Yu, C. E., and J. J. Ferretti. 1991. Molecular characterization of new group A streptococcal bacteriophages containing the gene for streptococcal erythrogenic toxin A (*speA*). *Mol. Gen. Genet.* **231**:161–168.
231. Zahler S. A. 1993. Temperate bacteriophages, p. 831–842. In A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
232. Zillig, W., D. Prangishvilli, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic, and D. Gotz. 1996. Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic *Archaea*. *FEMS Microbiol. Rev.* **18**:225–236.
233. Zimmer, M., S. Scherer, and M. J. Loessner. 2002. Genomic analysis of *Clostridium perfringens* bacteriophage phi3626, which integrates into *guaA* and possibly affects sporulation. *J. Bacteriol.* **184**:4359–4368.
234. Zou, D., J. Kaneko, S. Narita, and Y. Kamio. 2000. Prophage, phiPV83-pro, carrying panton-valentine leukocidin genes, on the *Staphylococcus aureus* P83 chromosome: comparative analysis of the genome structures of phiPV83-pro, phiPVL, phi11, and other phages. *Biosci. Biotechnol. Biochem.* **64**:2631–2643.
235. Zuniga, M., B. Franke-Fayard, G. Venema, J. Kok, and A. Nauta. 2002. Characterization of the putative replisome organizer of the lactococcal bacteriophage r1t. *J. Virol.* **76**:10234–10244.