

Phylogenetic structure and evolution of regulatory genes and integrases of P2-like phages

Hanna Nilsson, Carlos Cardoso-Palacios, Elisabeth Haggård-Ljungquist and Anders S. Nilsson*

Department of Genetics, Microbiology and Toxicology; Stockholm University; Stockholm, Sweden

Keywords: peduovirinae, P2-like bacteriophages, phylogenetic analysis, phage integration, lytic-lysogenic transcriptional switch, gamma-proteobacteria

The phylogenetic relationships and structural similarities of the proteins encoded within the regulatory region (containing the integrase gene and the lytic-lysogenic transcriptional switch genes) of P2-like phages were analyzed, and compared with the phylogenetic relationship of P2-like phages inferred from four structural genes. P2-like phages are thought to be one of the most genetically homogenous phage groups but the regulatory region nevertheless varies extensively between different phage genomes.

The analyses showed that there are many types of regulatory regions, but two types can be clearly distinguished; regions similar either to the phage P2 or to the phage 186 regulatory regions. These regions were also found to be most frequent among the sequenced P2-like phage or prophage genomes, and common in phages using *Escherichia coli* as a host. Both the phylogenetic and the structural analyses showed that these two regions are related. The integrases as well as the *cox/apl* genes show a common monophyletic origin but the immunity repressor genes, the type P2 C gene and the type 186 *cl* gene, are likely of different origin. There was no indication of recombination between the P2-186 types of regulatory genes but the comparison of the phylogenies of the regulatory region with the phylogeny based on four structural genes revealed recombinational events between the regulatory region and the structural genes.

Less common regulatory regions were phylogenetically heterogeneous and typically contained a fusion of genes from distantly related or unknown phages and P2-like genes.

DO NOT distribute.

Introduction

Although many phages are morphologically similar, their genomes may consist of genes with different evolutionary background, and it appears that many phages are composed of functionally interchangeable modules.^{1–3} Phages are characterized by a rapid evolution in which mutations, recombination between different phage genes, and horizontal gene transfer between phages result in low phylogenetic signals. This means that phylogenetic relationships can usually only be demonstrated for individual genes or modules, and not for entire genomes.

P2-like phages are temperate phages with 30–35 kb large genomes that infect γ -proteobacteria.^{4,5} They are perhaps different from other phage groups since their genome architecture, at least superficially considered, seems to be more conserved as compared with phages from other groups. T4-like viruses from *Escherichia coli* seem to have differentiated into several subgroups whereas P2-like phages from the same host are quite similar.⁶ As a consequence, the nucleotide sequences of many genes are also well-conserved among P2-like phages. The nucleotide sequence of five structural genes (the capsid scaffold gene, *O*, the major capsid precursor gene, *N*, the small terminase subunit gene, *M* and the capsid completion gene, *L*) from 18 P2-like phages isolated

from *E. coli* were shown to be 96% identical.⁷ There are however other coliphages that are less similar to P2. Phage 186 have merely 31 of its 43 genes in common with P2 and the nucleotide identity of the same five structural genes is only 76%. P2-like phages isolated from other γ -proteobacteria are less similar to P2 but have many homologous genes in common. For example, the two Salmonella phages Fels-2 and SopE Φ are 63–67% identical, and phage Φ CTX isolated from *Pseudomonas aeruginosa* is 53% identical, to P2 at the protein level.⁶ Even if phage 186 is similar to phage P2, and has been described as a coliphage, it has some genes in common with the Salmonella phages mentioned, particularly among the regulatory genes.

The similarities between P2-like phages have resulted in the taxonomic status of a subfamily, *Peduovirinae*, within the family *Myoviridae*.⁶ There are also data suggesting that the phylogenetic relationships of structural genes from different P2-like phages follows the phylogenetic relationship of their bacterial hosts, γ -proteobacteria.⁸ This could imply that P2-like phages are strictly host or strain specific and not prone to extensive recombination with phages with slightly different host range. However, we have previously shown that recombination between closely related P2-like phages occur, and that many P2-like phages contain unique horizontally acquired genes.^{7,9}

*Correspondence to: Anders S. Nilsson; Email: anders.nilsson@gmt.su.se
Submitted: 08/10/11; Revised: 10/18/11; Accepted: 10/19/11
<http://dx.doi.org/10.4161/bact.1.4.18470>

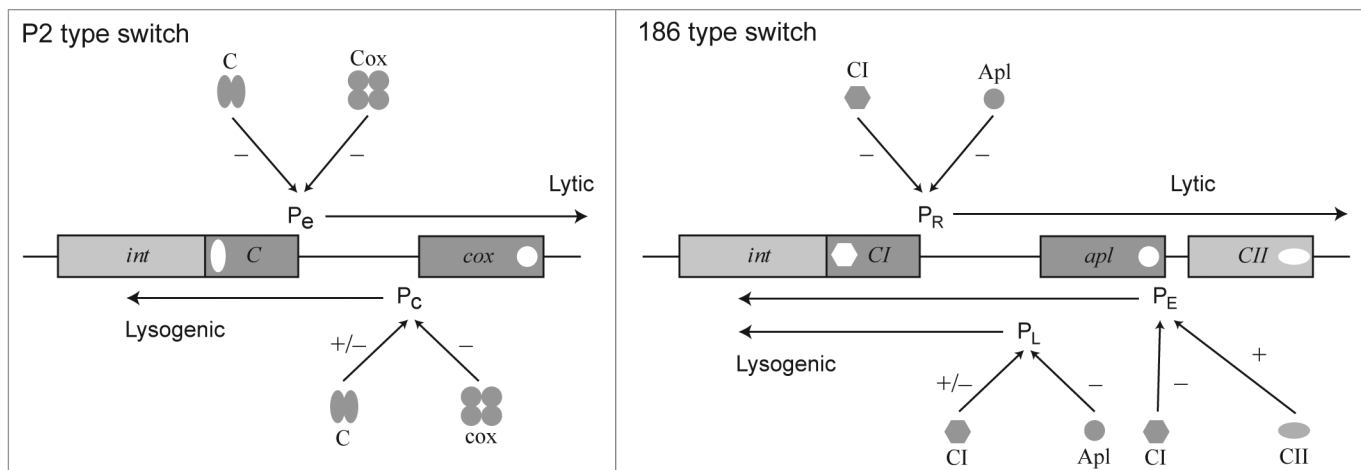


Figure 1. Schematic drawing of the transcriptional switch region of the two types of P2-like phages. The genes and patterns of repression and regulation of the P2 type (left) is mirrored by the more complex 186 type (right) which contain the additional gene *cII*. Both types have two converging promoters where the first gene of each operon encodes a repressor that controls the opposing promoter. In phage P2, *C* controls the expression of *Cox* and vice versa and these two repressors have their own operators located in the vicinity of, or overlapping with, the promoters they control. In addition, *C* also upregulates itself at low concentrations and downregulates at high. At high concentration, *Cox* also downregulates *C*. In phage 186, *CI* is expressed from different promoters during establishment and under maintenance. As in phage lambda, *CII* of phage 186 controls the establishment of lysogeny.

They are consequently not unaffected by mechanisms that cause mosaicism. The similarity of P2-like phages may indeed be due to a sampling bias since the first phages to be sequenced were originally isolated because of their phenotypical similarity to phage P2.

In this article we will investigate further the evolution of phages classified as belonging to *Peduvirinae* together with P2-like prophages identified in bacterial genomes. The aims are to assess the extent of modularity, i.e., check for recombination within and between two regions of the genome, analyze the evolution of functional differences of key proteins, and to evaluate possible host preferences of P2-like phages. Accordingly, the analyses consist of comparisons of structural features of three proteins encoded from functionally equivalent regulatory regions of P2-like phages, phylogenetic analyses of these proteins, and a comparison with the phylogenetic relationship of the phages' structural proteins.

The regulatory region in the phage P2 genome harbors the integrase gene (*int*), the immunity repressor gene (*C*), and the lytic repressor gene (*cox*) that also controls the directionality of the site-specific recombination. These three early genes are located next to each other, forming a co-adapted multifunctional regulatory region, since each of the three proteins bind to a DNA sequence that controls either transcription or function of at least one of the other two. The genes constitute the central mechanism for integration into, or excision out of, host genomes and are also central for directing the phage life cycle after infection, i.e., either to form lysogeny or to enter lytic growth. This is also the reason why the region can be said to contain a transcriptional switch, and there seems to be at least two different transcriptional switches among P2-like phages (Fig. 1). We wanted to investigate if there are more variants and also study the evolution of this region in more detail. Another aim was to assess

whether individual genes follow the same phylogenetic pattern or if there is significant recombination between them, something that potentially could have contributed to entirely new functional variation. We have also compared the phylogenies of the proteins of the early regulatory region with a phylogeny based on proteins inferred from four well-conserved late structural genes (corresponding to phage P2 genes *O*, *N*, *M*, *L* and *X*). Presumably, only small genetic changes are needed to make the regulatory genes to work together with another set of structural genes e.g., capsid genes. Recombination between structural and regulatory modules of P2-like phages should therefore appear as distinct phylogenetic branching patterns between the trees based on these two groups of genes.

Although structural genes from different P2-like phages are similar, recombination between these genes, or horizontal transfer of genes between genomes, cannot be ruled out. We have therefore analyzed the phylogenetic inter-relationship between the four structural genes.

We also investigated the hypothesis that the similarity between the phylogenetic relationship of P2-like phages and the relationship of their bacterial hosts could be explained by a close association between them for a long time. This would require that different P2-like phages are more or less host specific i.e., that some capsid- and regulatory types are found only in phages isolated from certain group of bacteria, or found in certain bacterial genomes. As mentioned above, there are two types of P2-like phages that can be discriminated by having different transcriptional switches. Representatives of both of these types have been isolated from *E. coli*, phage P2 itself and its relative phage 186. We have taken a closer look at the prevalence of these two prophages in *E. coli* and *Salmonella* using DNA-DNA dot-blot hybridizations with P2 and 186 genes as probes against bacteria from bacterial reference collections.

Table 1. Prophages and phages

| Prophage ^a | Accession number | Prophage host strain | Location in host genome | Integration site identity |
|--------------------------|------------------|---|--|---------------------------|
| ΦESP | NC_009436 | Enterobacter sp. 638 | 3783460-3815505 ^c | 186 |
| ΦKPN | NC_009648 | <i>Klebsiella pneumoniae subsp. pneumoniae</i> MGH 78578 | 3803150-3835929 ^c | 186 |
| ΦYFR | NZ_AALE00000000 | <i>Yersinia frederiksenii</i> ATCC 33641 | 11910-48192 ^d | |
| ΦECO1 | NC_009800 | <i>Escherichia coli</i> HS | 916812-948276 ^d | |
| ΦECO2 | NC_010468 | <i>Escherichia coli</i> C ATCC 8739 | 3023202-3058101 ^c | ΦECO1 |
| ΦECO3 | NC_004431 | <i>Escherichia coli</i> CFT073 | 908870-942351 ^d | ΦECO1 |
| ΦSEN1 | NC_003198 | <i>Salmonella enterica subsp. enterica serovar Typhi</i> CT18 | 3515381-3549053 ^d | ΦECO1 |
| ΦSEN2 | NC_003198 | <i>Salmonella enterica subsp. enterica serovar Typhi</i> CT18 | 4473835-4507389 ^c | |
| SopEΦ | AY319521 | <i>Salmonella typhimurium</i> DT204 | | Fels-2 |
| ΦECO4 | NZ_AAJV00000000 | <i>Escherichia coli</i> E22 | 125586-158737 ^c | P2 |
| ΦYPS | NC_006155 | <i>Yersinia pseudotuberculosis</i> IP 32953 | 3688279-3717980 ^d | |
| ΦYE98 | NC_008800 | <i>Yersinia enterocolitica subsp. enterocolitica</i> 8081 | 981219-1012614 ^d | |
| ΦECO5 | NC_007946 | <i>Escherichia coli</i> UTI89 | 911187-948481 ^c | |
| ΦECA29 | NC_004547 | <i>Erwinia carotovora subsp. atroseptica</i> SCRI1043 | 2935258-2966778 ^c | ΦD145 |
| ΦECO6 | NZ_AAMK00000000 | <i>Escherichia coli</i> 101-1 | 367299-403089 ^d | |
| ΦECO7 | NZ_AAKB00000000 | <i>Escherichia coli</i> 53638 | 63495-95171 ^c | WΦ |
| Phage^b | | | | |
| P2 | NC_001895 | <i>Escherichia coli</i> C | 2165238 ^e | |
| ΦD124 | AM159060 | <i>Escherichia coli</i> C | 2165238 ^e | |
| WΦ | NC_005056 | <i>Escherichia coli</i> C | 4104392 ^e | |
| L-413C | NC_004745 | <i>Yersinia pestis</i> , serovar 2 | 4104392 ^e | |
| ΦD160 | AM159075 | <i>Escherichia coli</i> C | 950293 ^e | |
| ΦD145 | AM158280 | <i>Escherichia coli</i> C | 950293 ^e | |
| | AM159077 | | | |
| | AJ298559 | | | |
| 186 | NC_001317 | <i>Escherichia coli</i> | 2783812 (IleY) 3213703 (IleX) (in <i>E. coli</i> MG1655) | |
| P _s P3 | NC_005340 | <i>Salmonella enterica subsp. enterica serovar Potsdam</i> | | |
| Fels-2 | NC_010463 | <i>Salmonella typhimurium</i> LT2 | 2844427-2879233 ^c | |
| F108 | NC_008193 | <i>Pasteurella multocida</i> | | |
| ΦO18P | NC_009542 | <i>Aeromonas media</i> | | |
| HP1 | NC_001697 | <i>Haemophilus influenzae</i> | | |
| HP2 | NC_003315 | <i>Haemophilus influenzae</i> | | |
| K139 | NC_003313 | <i>Vibrio cholerae</i> O139 | | |
| Kappa | NC_010275 | <i>Vibrio cholerae</i> O1 biovar eltor | | |

^aProphages that have not been shown to be viable, i.e., found in sequencing of host genomes. ^bPhages that have been shown to be viable. ^cCounter clockwise to the replication movement of the bacterial genome. ^dClockwise to the replication movement of the bacterial genome. ^eThe genome of *E. coli* C has not been sequenced. This is the corresponding integration site in *E. coli* MG1655.

Results

In total, 31 complete P2-like phage or prophage genomes were identified (Table 1). The phages left out from the following analyses typically had regulatory regions or regulatory genes of other types that were impossible to align with P2 or 186 type genes. There were for instance phages classified as *Peduvirinae* that did

not contain both P2 type of late genes (genes equivalent to P2 *O*, *N*, *M* or *L*) and a complete regulatory region (genes equivalent to P2 *int*, *C* or *cox*). The Burkholderia phages ΦE12-2, ΦE202 and Φ52237 are clear examples of mosaic genomes. The last two contain integrases of the lambda type, and none had genes identified as analogous to P2 *C* or *cox*, or to phage 186 *apl* or *cI*. The Mannheimia phage ΦMHaA1 also had a lambda-like integrase

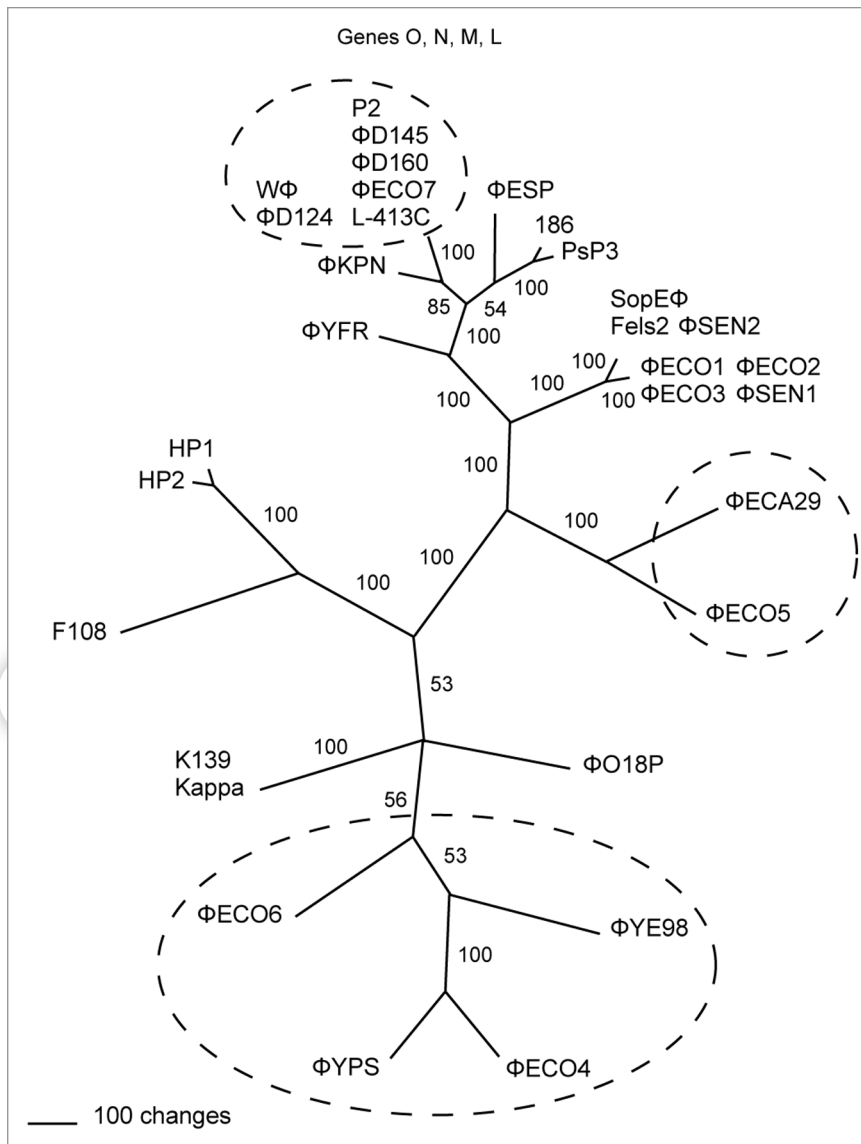


Figure 2. Unrooted phylogenetic tree of the structural genes *O*, *N*, *M* and *L* of P2-like bacteriophages. The tree is based on the concatenated inferred amino acid sequences of all four genes. Phages with the P2 type of regulatory region (*int-C-cox*) are encircled with dashed lines. The tree was generated under maximum phylogeny criteria in PAUP* and using branch-and-bound searches for finding the shortest tree. The robustness was tested in a 1,000 replicates bootstrap, resulting in the percentages on the branches, and with homogeneity partition tests of major clusters. Branches with less than a 50% bootstrap support are collapsed in the trees.

but contained an immunity repressor of the phage 186 *cl* type, and a gene similar to *cox* or *apl* was not found. The two phages ΦCTX and ΦRSA1 from *Pseudomonas* and *Ralstonia* had P2 type integrases but their immunity repressor genes or genes equivalent to *cox/apl* could not be identified.

The analyses of the relationship between the remaining 31 phages revealed a complex evolutionary history. Many genes seem to have coevolved for a very long time, but others show signs of recombination between closely related phages as well as between cohesive groups of phages. The alignments of the inferred amino acid sequences of the genes of the late gene region

were concatenated and analyzed phylogenetically. Bootstrap tests revealed that most major groups were stable, but also that the clustering within some groups could not be resolved (Fig. 2). Homogeneity partition tests (HPT) of all four genes resulted in a low probability that the capsid scaffold protein gene (equivalent to P2 *O*) in the Fels-2 group (Fels-2, SopEΦ, ΦSEN1, ΦSEN2, ΦECO1, ΦECO2 and ΦECO3) have had the same evolutionary history as the other three genes ($p < 0.002$). The major capsid precursor protein gene (P2: *N*) within the HP1 cluster (HP1, HP2, F108, Kappa, K139, ΦO18P, ΦECO6, ΦYE98, ΦYPS and ΦECO4), and the small terminase subunit (P2: *M*) and capsid completion (P2: *L*) genes in the phage 186 cluster (186, PsP3, ΦYFR, ΦESP and ΦKPN) also showed signs of different evolutionary histories than the other genes, but only weakly ($p = 0.027-0.047$). Recombination between closely related phages is the most likely explanation to the results of the HPT tests.

The analysis of the inferred amino acid sequences of the late genes also showed that the variation was greater than expected. Sequence analyses of the late genes in relatives to phage P2, that were able to grow on *E. coli* strains, have been shown to differ by only a few per cent.⁷ Among the prophages discovered in *E. coli* genomes, only one phage with a regulatory region closely related to P2 had late genes likewise related to phage P2. Others with an equally similar regulatory region (ΦECA29, ΦECO4, ΦECO5, ΦECO6, ΦYPS and ΦYE98) sometimes had more different late genes than the HP1/HP2 phages even though they were found in prophages of *E. coli* strains (Fig. 2).

The genetic variation of the regulatory region was too extensive to be subjected to a joint phylogenetic analysis. This could possibly be explained by the fact that some phages had genes that were unrelated to analogous genes from other phages (e.g., PsP3 and ΦCTX), but the difference between C and CI, which is twice the size of C, also contributed to the difficulties. These two proteins share no motifs at the sequence level and were impossible to align. Consequently, four separate phylogenetic analyses of the inferred amino acid sequences of the genes in the regulatory region were performed, i.e., *Cox/apl*, *int*, *C* and *cl*. The analyses resulted in an orderly distribution of genes from the two different types of regulatory region (Fig. 3). The variation of the integrase could be divided into two equally large groups, one containing all phages also having the C immunity repressor, and the other containing all with the CI immunity repressor. The variation of the *Cox/Apl*

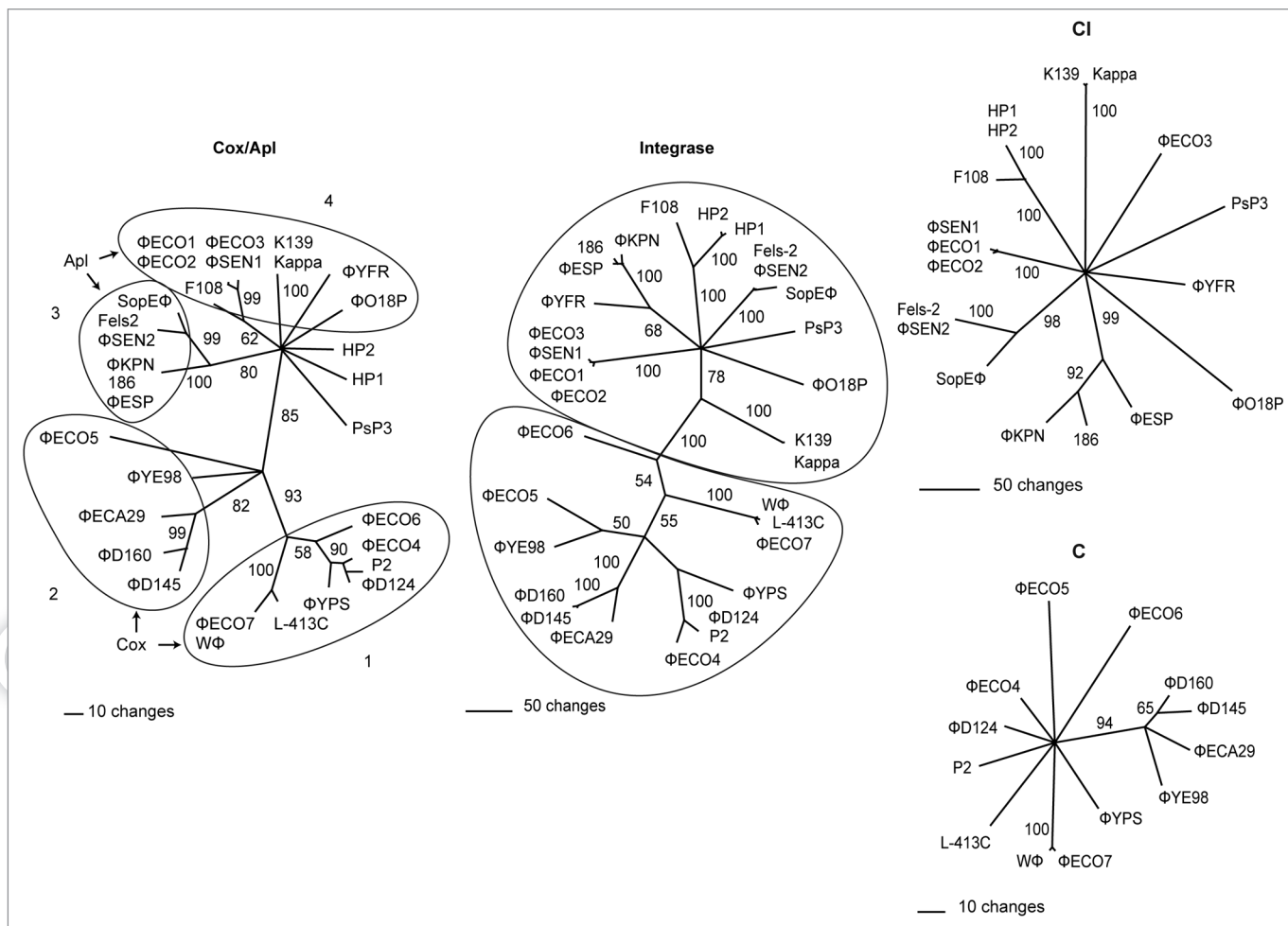


Figure 3. Unrooted phylogenetic trees depicting the relationship between the different Cox and Apl proteins (left), the Integrases (middle), the CI proteins (top right) and the C proteins (bottom right). The trees were generated with maximum phylogeny criteria in PAUP* using amino acid alignments and a heuristic search followed by a 1,000 replicates bootstrap resulting in the percentages on the branches. Branches with less than a 50% bootstrap support are collapsed in the trees. The trees show the same general division between P2 type phages and 186 type phages throughout the whole switch region.

proteins could be divided into several subgroups within these two major groups but there were also proteins that were completely different, i.e., HP1, HP2 and PsP3, and that may represent new types.

The most apparent result of recombinational events was however between the late genes and the regulatory region of many P2-like phages, where regulatory genes that cluster together in the Cox, Int and C trees (Fig. 3) have completely dissimilar late genes that, according to the phylogenetic analysis, were very distantly related (Fig. 2). Two ancient recombinational events are enough to explain the differences between the tree of the late genes and the trees of the genes in the regulatory region.

The immunity repressor proteins, C and CI. The C protein is part of the less complex transcriptional switch type represented by phage P2 (Fig. 1). It functions as a repressor of the lytic genes and confers immunity to infecting phages of the same immunity group. An immunity group is defined as a set of C proteins similar enough to recognize the same DNA binding site. Seven

immunity groups have earlier been found among the phages with a P2 type of switch, all sharing at least 95% identity within groups.⁹

Seven new C proteins were inferred from prophage nucleotide sequences within sequenced bacterial genomes in GenBank. Four of these were found in *E. coli* strains (ΦECO4, 5, 6 and 7), two in *Yersinia* sp (ΦYPS and ΦYE98), and one in *Erwinia* (ΦECA29) (Table 1 and Fig. S1). Six of the inferred proteins were not alike any of the C proteins of earlier identified immunity groups of phages with an *E. coli* host (I–VII; reviewed in ref. 9), and they also had spacer regions between genes C and *cox* that varied in size and sequence to such an extent that they could not be aligned. The spacer regions contain the important operator sites of the transcriptional switch. Consequently, these six C proteins were expected to represent new immunity classes, even though two were found in *Yersinia*, and one in *Erwinia*. The ΦECO7 prophage in the *E. coli* strain 53638 had a C protein and a spacer region that only differed by two residues compared

with the same sequences in phage WΦ, and most likely belongs to immunity class III. The three new immunity classes identified in *E. coli* are tentatively called class VIII (ΦECO4), class IX (ΦECO5) and class X (ΦECO6), since their function remain to be determined.

In the phage 186 type switch, the protein functionally analogous to P2 C is denoted CI. Apart from the ten CI proteins identified in phage genomes, we found eight new CI proteins in prophage genomes within five bacterial genera; *Klebsiella* (ΦKPN), *Enterobacter* (ΦESP), *Yersinia* (ΦYFR), *Salmonella* (ΦSEN1 and 2) and *E. coli* (ΦECO1, 2 and 3) (Table 1 and Fig. S2). The immunity classes of CI have not been as thoroughly studied as the classes of C proteins of the P2-like phages. However, the prophage ΦSEN2 from *Salmonella enterica* CT18-2 most likely belongs to the same immunity class as Fels-2. Their CI proteins were identical and their spacer sequences between *cI* and *apl/cox* were of equal length (116 nt), with only four mismatches. Another possible immunity class includes the prophages ΦECO1, ΦECO2 from *E. coli* strains and ΦSEN1 from *Salmonella enterica* since their CI proteins were over 96% identical. This was in addition supported by the 98–100% identity of the 125 nt long spacers. In a comparison, including all 18 CI repressor proteins, they were shown to be at least 40% identical and the length of the spacer between the *cI* and *apl/cox* genes varied between 88–125 nt.

The Cox/Apl proteins. The Cox/Apl protein of P2, WΦ or 186 functions both by regulating the transcriptional switch and as an architectural protein during site-specific recombination.^{10–13} The phylogenetic analyses revealed that the Cox/Apl proteins could be divided into four well-supported subgroups, two within Cox and two within Apl (Fig. 3). The same groups were to some extent present in the lesser resolved C and CI trees, and the JPRED predictions of their 2D structures separated them into the same four groups (Fig. S3). JPRED predicted a winged-HTH domain at the N-terminal half of the proteins for group 1 and 2, but the proteins in group 2 was larger compared with those in group 1, and there were only six conserved residues between the two groups. The proteins in group 3 were predicted to have a HTH motif at the N-terminus. This was not evident from the JPRED results but the helixturnhelix program in the EMBOSS package predicted a HTH motif followed by a β-α-β-α structure. In group 3, the Apl sequences of prophages ΦKPN and ΦESP were identical to that of phage 186, and the Apl of ΦSEN2 was identical to the Apl of Fels-2. The proteins of group 4 had a predicted β-sheet at the N-terminus that was followed by a HTH motif. There was no indication of two β-sheets forming a wing. Instead, the C-terminus contained a single α-helix.

The integrases. The integrase of the P2-like phages is a tyrosine site-specific recombinase that mediates integration and excision of the phage at the DNA attachment site. Dimers of the integrase bind to a core sequence and two arm sequences in the phage genome, the *attP* site, and to an *attB* site in the bacterial genome which is identical to the core of the *attP* site. The core of *attP* and *attB* is recognized by the core-binding domain of the integrase but the surrounding arm-binding sites in *attP* are recognized by the N-terminal domain. The C-terminal domain

contains the catalytic site. The amino acid sequence of both terminal domains of the integrase was similar in all P2 type and 186 type phages while the core-binding domain, where the integration specificity is determined, was much more variable and consequently harder to align than the N-terminal domains. The topology of the phylogenetic tree based on the complete residue sequence of the integrase proteins resulted in the same two major phage groups as in the C, CI and Cox/Apl trees (Fig. 3). In fact, when the sequence of the core-binding domain of Int was exempted, the HPT-tests supported the same evolutionary history for the complete regulatory region, except for the phages K139, HP1 and HP2. The arm-binding sites of P2 and phage 186 contain at least two direct repeats on either side of the core sequence. Direct repeats were identified in all phages or prophages, except ΦO18P, again forming two groups; one group with direct repeats similar to P2 and with a consensus sequence of tgTGGaCa, and another with direct repeats similar to phage 186 with a consensus sequence of tgccGCCActt (Fig. S4).

A comparison of the amino acid sequences of the two integrase types, those recognizing the P2 type arm-sites and those recognizing the 186 type arm sites, points at possible discriminators, i.e., Y vs. W at position 16, D vs. E at position 19, and R vs. Y at position 21 in the alignment (Fig. 4). Int defective point mutations in the N-terminal domain have previously been identified,¹⁴ and two of these are located at the potentially discriminating amino acids.

The integration sites of four phages (P2, WΦ, ΦD145 and phage 186) in *E. coli* have been identified previously in reference 5 and 15. When a phage is integrated, the attachment site core sequence can be deduced from the *attL* and *attR* junctions with the host DNA. Accordingly, we found many new integration sites by searching the direct repeats at the ends of the prophage genomes (Fig. 5). There was a tendency for phages with the 186 type of integrase to be integrated into tRNA genes, while the P2 type was found to be integrated between other genes, or more rarely into non-tRNA genes. For instance, phage 186 integrates into a tRNA^{Ile} gene and HP1 into a tRNA^{Leu} gene.^{5,16,17} We found that ΦESP and ΦKPN were similarly integrated into tRNA^{Met}. Just like phage 186 they both had long common core sequences to ensure an intact tRNA gene after insertion, but their lengths outside the tRNA gene differed extensively. ΦESP was found to have 50 nucleotides in common between *attL* and *attR* whereas ΦKPN had 110. ΦYFR was found to be integrated into tRNA^{Lys}, but in this case the core sequence was shorter. SopEΦ and Fels-2 have earlier been shown to be integrated into *ssrA* (tmRNA), which has a tRNA structure.^{18,19}

We identified two P2-like prophages in *Yersinia* sp, ΦYE98 and ΦYPS. They had relative short common core sequences when comparing the *attL* and *attR* junctions, 23 and 21 nt respectively, but in the latter case there were a few mismatches in the hypothesized core, so it might be even shorter (Fig. 5). The core sequence identified in ΦYPS is found in other *Yersinia* without the integrated prophage, while the core sequence identified in ΦYE98, including the host DNA on either side, can only be found in strain 8081 and thus does not seem to be common to other *Yersinia* species.

| | | → <i>int</i> | | | |
|----------|--------------------------|--|--|--------------|--|
| A | P2 ΦECO4 ΦD124 | AAAAAATAAGCCCGTGTAAAGGGAGATT | Within <i>ryeE</i> RNA = ■ | | |
| | WΦ ΦECO7 L-413C | GACACCATCCCTGTCTTCCCCACATGATGTGGGGGTTTTTTTTTATC | Between <i>cpxR</i> and <i>pfkA</i> | | |
| | ΦD145 ΦD160 ΦECA29 | AATGTCACCAATAACTGACTTGG | Within <i>pflA</i> | | |
| | ΦECO5 | AAAGCGCCTGCGGGCGCTTT | Between <i>serS</i> and <i>dmsA</i> | | |
| | ΦECO6 | TTTTTCATCAACAAGGATTTT | Within <i>pdxB</i> | | |
| | ΦYE98 | TCAATAAAAAACGCGCCCAAAGG | Between <i>pheA</i> and <i>yfiA</i> | | |
| | K139 | CAGAAAAAAGAAAAGCCCCTTTTC | Between <i>flaA</i> and <i>flaC</i> | | |
| | Kappa | GAAAAGGGGCTTTTCTTTTCTGTGCTGT | Between <i>flaA</i> and <i>flaC</i> | | |
| | ΦECO3 186 | ATAAATTTCAAGCAACAAAAAACCAT TGCTGGACTTGAACCAGCGACCAAGCGATTATGAGT | Within <i>rybB</i> RNA = ■ Within tRNA ^{Ile} | | |
| | HP1 HP2 | +169nt..AGGGATTTTAAATCCCTT | Within tRNA ^{Leu} | | |
| | F108 | +81nt..AGGGATTTTAAATCC | Within tRNA ^{Leu} | | |
| | SopEΦ Fels-2 | TTTGGTGGAGCTGGCGGGAGTTGAACCCGCGTCCGAAATTCCTACAT | Within <i>ssrA</i> RNA | | |
| | ΦKPN | +59nt..GTGGCCCTGTTGGGTTTGAACCAACGACCAAGCGATTATGAGT | Between an ORF and a DNA glycosylase | | |
| | ΦYFR | TTTGTGGTGGGTCGTGCAGGGTTCGAACCTGCGACCAATTGATTAAGAGT | Within tRNA ^{Lys} | | |
| | ΦCTX | TGAGTTCGAATCTCACCGCCTCCGCATAT | Within tRNA ^{Ser} | | |
| | B | ΦYPS | AAGGCATCGTGAGATGCCTTT AACGCATCGTGAGATGCCTTT | attL attR | Between YPTB3126 and tRNA ^{Gly} |
| | | ΦECO1 ΦECO2 | ATAAATTTTAGATAACAAAAAACCAT TTAAATTTCAAGCAACAAAAAACCAT | attL attR | Within <i>rybB</i> RNA = ■ |
| | | ΦSEN1 | ATCAATTTCCGGCTAAAAAAATAACCA TTAAATTTCAAGCAACAAAAAACCAT | attL attR | Within <i>yifB</i> |
| | | ΦESP | AAATTTGGTGGCCCTGCTGGACTTGAACCAGCGACCAAGCGATTATGAGT AAATTTGGTGGCCCGTCTGGACTTGAACCAGCGACCAAGCGATTATGAGT | attL attR | Within tRNA ^{Met} |
| | | ΦSEN2 | TTTGGTGGAGCTGGCGGGAGTTGAACCCGCGTCCGAAATTCCTGCATAC | attL attR | Within <i>samA</i> |
| ΦO18P | | N.D. | | | |
| PsP3 | | N.D. | | | |

Figure 5. Identification of host DNA attachment sites for P2-like phages. Attachment site core sequences resulting from a comparison of host-prophage *attL* and *attR* junction regions. All sequences are in the same orientation indicated by the location of the *int* gene. Sequences shaded in gray are within part of genes. The position of the sites relative to host genes are shown to the right. (A) Prophages where the suspected core region of the *attL* and *attR* regions were found to be identical, confirming the core nucleotide sequence of *attP* and *attB*. (B) Prophages with similar *attL* and *attR* regions, indicating the frame for the core sequence. N.D. = Not detected.

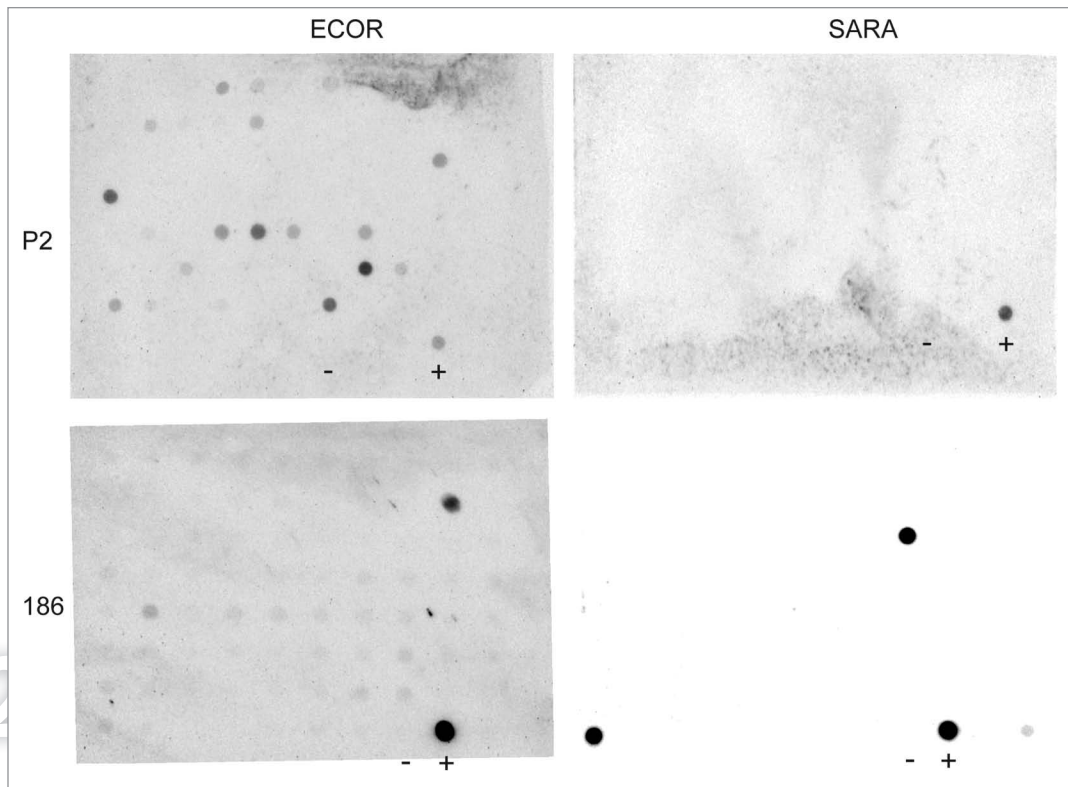


Figure 6. DNA-DNA dot blot hybridization of the P2 type and the 186 type of phage against the bacterial reference collections ECOR (72 strains of *Escherichia coli*) and SARA (72 strains of *Salmonella enterica* spp). The P2 DNA probe consisted of pooled amplifications of the genes *V-J*, gene *T* and *int-cox* and the phage 186 DNA probe of the pooled amplifications of the equivalent genes *32-L*, gene *G* and *int-apl*. Whole genomic bacterial DNA from individual strains were transferred onto a membrane and hybridized against the phage specific probes. In each picture, there are ten phage—bacterium hybridizations in each row. A dark spot indicates the presence of a closely related prophage in the particular strain. Controls are in the lower right corner and marked with a minus sign (-) for the negative control and a plus sign (+) for the positive control.

probe mix of unique P2 regions, including the regulatory region, resulted in 17 hits against the ECOR strains but none against the SARA strains. Hybridizing with the phage 186 mix, amplified from homologous genomic regions of the same size as the P2 probe, resulted only in a single hit against an ECOR bacterium, and only two hits against strains in the SARA collection (Fig. 6). The hybridizations against the SARB strains resulted in similar distributions of the two phage types; one hit when the P2 mix was used but five when the 186 mix was hybridized (data not shown). Taken together, it appears as both *E. coli* and *Salmonella* contain many P2-like prophages, but that the P2 type seems to be common in *E. coli* and almost absent in *Salmonella* whereas the 186 type seems to be rare in both bacteria. In addition, there must be either cryptic prophages or prophages with other regulatory regions than the P2 and 186 types in *Salmonella* since there are signals in the whole genome hybridization that are not present in the type specific hybridizations.

Discussion

The discrepancy between the phylogenetic trees of the late genes and the regulatory region of the 31 phages together with the many instances of regulatory region genes of other types than P2 or 186 types, e.g., the presence of lambda type integrases in phages

classified as *Peduvirinae*, being associated with P2-like late genes clearly demonstrates that the evolution of P2-like phages are no exception from the evolutionary processes found among other groups of phages. They indeed have mosaic genomes that consist of functional modules. In addition, GenBank contains hundreds of P2-like sequences coding for structural proteins found in bacterial genomes. Even though many of these are cryptic phages lacking the regulatory region, some ought to be functional P2-like prophages with other sets of regulatory regions than the P2 or 186 types.

The majority of phages with complete P2-like genomes (*sensu stricto*: alike phage P2) are of two distinct types, one with a P2 type of regulatory region and another with a phage 186 type, and it appears as only two recombination events have occurred between their regulatory regions and their late genes.

Recombination within the same type of regulatory region has been detected before,⁹ but it seems to be confined to closely related phages of the same regulatory region type. Although the phylogenetic inference is weak, no chimeric regulatory region was found among the 31 phages. The *int*—transcriptional switch regions of both types are strongly coadapted due to their multifunctionality, and the evolution of the regions is probably constrained by this complexity. The regions are simply too dissimilar to allow homologous recombination, and

illegitimate recombination would most likely result in dysfunctional regulatory genes and ruin the precise key and lock feature of the proteins. New regions that do arise of each type are more likely the result of many and smaller mutational changes, as well as recombination between similar regions. The two *int*—transcriptional switch regions have diverged into several subgroups accordingly, wherein the individual genes show a different evolutionary rate.

The crystal structure of P2 C has recently been determined.²² The N-terminal of the C protein is predicted to contain four α -helices where helix 3 has been hypothesized to be the DNA recognition helix in a helix-turn-helix (HTH) motif formed in association with helix 2. This is in accordance with the finding that this part of the 13 C proteins of the P2 type is very variable which corresponds to the variation of their target DNA sequences. There are only two conserved amino acids in helix 3, both at the C-terminal end. The structure of P2 C also revealed a fifth α -helix and a β -sheet not detected by the JPRED structure prediction. The sequences of these are quite conserved in all the 13 C proteins from phages with a P2 type transcriptional switch. The C-terminals of these C proteins also contain another conserved region, GQIAPALA, located after the structurally determined β sheet in P2 C but before the last 14 residues that form the C-terminal's flexible tail (Fig. S1).

The structure of the CI protein of phage 186 has also been determined, and the N-terminal domain contains five α -helices where helix 2 and 3 forms the HTH motif. The C-terminal domain consists of a highly twisted ten-stranded β -sheet that is involved in the assembly of a heptamer of dimers.²³ The alignment of the CI proteins showed that PsP3 and prophage Φ ECO3 have a longer C-terminal domain compared with phage 186, and also that SopE Φ , Fels-2 and Φ SEN2 have a longer coupler between the two domains (Fig. S2).

It is doubtful if these proteins, the C type and the CI type, have a common evolutionary ancestry. Both groups are however structurally conserved at the N-terminal end which interacts with DNA via a HTH motif, and it could be hypothesized that the C type of repressor protein is the result of an old truncation of CI. However, the CI proteins are all twice the size of the C proteins and the sequences of the two groups are impossible to align. A more plausible hypothesis is that the C gene is a horizontally transferred addition to the P2 type of regulatory region.

The Cox/Apl proteins have actually differentiated into more than four groups since there are analogous proteins that do not resemble any of the Cox/Apl proteins in the alignment. The proteins of HP1/HP2 and PsP3 are different from the rest and the Cox/Apl proteins in Φ CTX and Φ RSA1 have not been identified. The Cox/Apl proteins of HP1/HP2 have a secondary structure predicted to be similar to that of group 4 but the protein of PsP3 is a singleton without detectable conserved domains or relation to a protein with known function.

The identified Cox/Apl proteins in the analyses are also highly differentiated, especially between groups. Group 1 and 2, all belonging to the Cox type, share six residues but the other groups are only structurally similar over the HTH domains (Fig. S3). Thus the phylogenetic relationship of the groups cannot be

concluded, which suggests that the evolution of the Cox/Apl proteins to be further studied.

The integrases are more conserved than the transcriptional switch genes and obviously shares a common ancestry. They can clearly be divided into two groups, the P2 type and the 186 type of integrases, which share secondary structure and some residues. The same phylogenetical groups are present in the analyses of both transcriptional switch genes which, in combination with the analyses of the secondary structure of these proteins, points at an evolutionary monophyletic background of the entire regulatory region. Midpoint rooting of the Int and Cox/Apl phylogenetic trees places the Int-CI-Apl type closer to the root which suggests that it is the older type and that the less differentiated Int-C-Cox is a derived state that has evolved to become less complicated. The 186 type of transcriptional switch is similar to the intricate lambda switch as it is also dependent on the additional gene *cII*, whereas the P2-like switch not only lacks an equivalent to this gene, but also has a smaller C protein, half the size of CI. It cannot be ruled out that the 186 type of switch is distantly related to the transcriptional switch type of phage λ .

The distribution of these two types of transcriptional switches seems not to be strongly associated to phages utilizing certain hosts. Previous studies have concluded that the P2 type of switch is confined to phages with an *E. coli* host.^{8,9} Our analyses show that the P2 type can also be found in prophages in *Yersinia* and *Erwinia*, and that the 186 type is widely distributed among phages in many bacterial genera, e.g., *Salmonella*, *Klebsiella* and *Aeromonas*, as well as in *E. coli* and *Yersinia*. There are also prophages from *E. coli* and *Yersinia* with a genome containing a P2 type of regulatory region and structural genes more similar to phages like HP1, HP2 or K139. These results on the distribution of the two types of regulatory regions and their host preferences are not in accordance with earlier studies which showed a high occurrence of the P2 type of transcriptional switch in *E. coli* but no sign of phages with a 186 type. If the 186 type of transcriptional switch is actually common in all γ -proteobacteria it is surprising that not a single one was found among 38 sequenced *E. coli* phage switches.⁹ The results of the hybridizations may explain this discrepancy since they show that 186 like prophages are scarce within *E. coli* and *Salmonella*. In addition, phage 186 may have been isolated from an *E. coli* host but it grows poorly on many *E. coli* strains. Under laboratory conditions, phage 186 is propagated on *E. coli* B strains.²⁴ It could thus be questioned whether either *E. coli* or *Salmonella* is the preferred host for phage 186 or if it is another bacteria related to these or to *Klebsiella* or *Enterobacter*. In either case, the observations are in accordance with the view that the 186 type is older and spread over more bacterial genera than the P2 type.

From a taxonomic point of view, it could be motivated to let the subfamily *Peduovirinae* contain at least two genera; P2 types and 186 types. Since phage HP1 was the first phage with the 186 type of regulatory genes to be fully sequenced, we suggest naming the two groups "P2-like phages" and "HP1-like phages" within the *Peduovirinae* subfamily.

Previous studies have shown not only a preference but also a differentiation of P2-like phage genomes consistent with the

phylogeny of the hosts, indicating a host preference.⁸ Though we cannot see signs of host preference in this study it is undisputable that there can be several explanations for a lack of correlation between different sets of genes and host association. P2-like phages were initially isolated from commensal gut bacteria and sewage. Only phages that could be propagated on standard laboratory *E. coli* strains were isolated, and the sampling of phages was thus biased. Many of the prophages identified in bacterial genome sequencing projects might be old inactive prophages with mutationally deteriorated genomes showing a poor relationship to recent functional phages. Several of the bacteria that harbor the prophages identified in this study are not commensal but pathogens, due to a sequencing bias of such strains. Pathogenic *E. coli* may have as much as 20% larger genomes, and the extra genes are often organized into horizontally transferable pathogenicity islands which may contain prophages. There is also a possibility for conjugative transfer of genes between bacteria or by means of other vectors. Consequently, phage genes may hitch-hike along other genetic elements that are horizontally transferred and eventually be found in atypical genomes.

Materials and Methods

The protein sequences of all seven genes (genes equivalent to phage P2 structural genes *O*, *N*, *M* and *L*, and regulatory region genes *int*, *C* and *cox*) from all phages in this study were extracted from complete genomes as follows. First, from all phages classified by ICTV as *Peduvirinae* and second by blastp searches in the nr database, at the National Centre for Biotechnology Information website (www.ncbi.nlm.nih.gov), for potential complete P2-like prophage genomes within bacterial genomes (Table 1). In these searches, the amino acid sequences of the phage P2 and phage 186 *int* genes were used as probes, and the genomes of candidate phages were examined. The prophage genomes were considered to be complete and non-cryptic if they had a gene order similar to P2-like phages, and if they contained structural genes, a complete transcriptional switch region, an integrase, and if they were surrounded by identifiable *attL* and *attR* junctions.

The genomic regions *attP-int*-transcriptional switch genes and the structural genes of all phages or prophages were merged into one regulatory (including *int*) and one structural amino acid character matrix, which were aligned with ClustalX (version 1.83; IGBMC, University of Strasbourg, <ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/>) and the Jalview alignment editor (version 2.4; School of Life Sciences, University of Dundee, www.jalview.org). The phylogenetic analyses were executed using PAUP* (version 4.0b10; Sinauer Associates, Inc., <http://paup.csit.fsu.edu/about.html>).²⁵ The amino acid sequences of the highly differentiated regulatory region had to be analyzed one by one for maximum parsimony (MP) trees. The amino acid sequences of the genes of the structural region were however analyzed together. The searches for MP trees were performed under default settings except for the branch-and-bound (corresponding to phage P2 genes *O*, *N*, *M* and *L*) or heuristic (genes *int*, *C*, *cl* and *cox/apl*) search option and for random addition of starting trees. The support for each resulting shortest tree was assessed in

a 1,000 replicate bootstrap analysis, and branches with less than 50% support were collapsed. Congruence between trees based on different genes was tested in homogeneity partition tests (HPT) executed using the heuristic search setting in PAUP*, with random addition of starting trees and for 1,000 trees.

Secondary structure predictions of protein residue alignments were done with JPRED at www.compbio.dundee.ac.uk/~www-jpred/²⁶ and the program helixturnhelix in the EMBOSS program package at www.ch.embnet.org/EMBOSS/.²⁷

The difference in the natural distribution of the two phages P2 and phage 186 was investigated by DNA-DNA hybridizations where DNA from these two phages were used as probes and hybridized with bacterial DNA from collections of *Escherichia coli*, the ECOR collection,²⁸ and *Salmonella enterica*, the SARA and SARB collections.^{29,30} All bacterial strains, including the following strains used for construction of probe DNA, were grown overnight at 30°C in LB media and 1.5 ml of the cell suspension was used for DNA extraction with the QIAGEN DNeasy Blood and Tissue kit (69504). The probes were constructed either from whole genomic DNA (phage P2 only) or by PCR amplification of prophage genes from bacterial genomes (phage P2 and phage 186). The whole genomic probe from P2 was used to probe the bacterial collections for P2-like phages in general. The two other probes were made to be able to more specifically discriminate between the P2 type and the 186 type of phages and amplified from homologous genes or sets of genes from the two prophages. The regions amplified from P2 were genes *V-J*, gene *T* and *int-cox*, and from phage 186 the homologous regions genes *32-L*, gene *G* and *int-apl*. GE Healthcare Illustra PuRe Taq Ready-To-Go PCR beads (27-9557-01) were used for the amplification of the genes from the P2 lysogen C-117,³¹ and from the *E. coli* K12 strain C600, harboring a phage 186 lysogen,³² using primers from DNA Technology. The amplified probe DNA fragments were separated in a 1% agarose gel running in 1x TBE buffer, and extracted with the QIAEX II Gel extraction kit (20021). The probes were cut with Fermentas FastDigest Bsh1236I (ER0921) before DIG labeling. These two lysogenic strains were also used as positive controls in the hybridizations, and the non-lysogenic *E. coli* strain C-1757 was used as a negative control.³³

The concentrations of the DNA extracted from the three bacterial collections were spectrophotometrically quantified with a Thermo Scientific Nanodrop 8000, and 2 µg DNA from each strain spotted onto Bio-Rad Zeta-Probe GT membranes (162-0190), utilizing a Bio-Dot SF microfiltration device (170-6542/170-6543). The labeling of probes and the following hybridization process was performed using the Roche DIG High Prime DNA Labeling and Detection Starter Kit II (11585 614 910). The chemiluminescent detection of hybridization signals was made with a Fujifilm LAS-1000 Image analyzer equipped with Multi Gauge software. All the preparation and execution of hybridizations described above were performed according to protocols supplied by the manufacturers.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgements

H.N. performed the phylogenetic analyses, performed the DNA-DNA hybridizations of phage genes—bacterial genomes, and made the structural analyses of the C and CI proteins. C.C.P. made the structural analyses of the integrases, the host attachment sites, and arm binding DNA sequences. We would like to thank the anonymous reviewers for their helpful comments and suggestions. The work was partially supported by a grant from the Swedish Research Council to E.H.L.

References

- Hendrix RW, Hatfull GF, Smith MC. Bacteriophages with tails: chasing their origins and evolution. *Res Microbiol* 2003; 154:253-7; PMID:12798229; [http://dx.doi.org/10.1016/S0923-2508\(03\)00068-8](http://dx.doi.org/10.1016/S0923-2508(03)00068-8)
- Hendrix RW, Smith MCM, Burns RN, Ford ME. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc Natl Acad Sci USA* 1999; 96:2192-7; PMID:10051617; <http://dx.doi.org/10.1073/pnas.96.5.2192>
- Brüssow H, Desiere F. Evolution of tailed phages: Insights from comparative Phage genomics. In: Calendar R, Ed. *The Bacteriophages*. New York: Oxford University Press 2006; 26-36
- Bertani G. Studies on lysogenesis I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 1951; 62:293-300; PMID:14888646
- Nilsson AS, Haggård-Ljungquist E. The P2-like bacteriophages. In: Calendar R, Ed. *The Bacteriophages*. New York: Oxford University Press 2006; 365-90
- Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, et al. Classification of Myoviridae bacteriophages using protein sequence similarity. *BMC Microbiol* 2009; 9:224; PMID:19857251; <http://dx.doi.org/10.1186/1471-2180-9-224>
- Nilsson AS, Haggård-Ljungquist E. Detection of homologous recombination among bacteriophage P2 relatives. *Mol Phylogenet Evol* 2001; 21:259-69; PMID:11697920; <http://dx.doi.org/10.1006/mpev.2001.1020>
- Nilsson AS, Haggård-Ljungquist E. Evolution of P2-like phages and their impact on bacterial evolution. *Res Microbiol* 2007; 158:311-7; PMID:17490863; <http://dx.doi.org/10.1016/j.resmic.2007.02.004>
- Karlsson JL, Cardoso-Palacios C, Nilsson AS, Haggård-Ljungquist E. Evolution of immunity and host chromosome integration site of P2-like coliphages. *J Bacteriol* 2006; 188:3823-35; PMID:16707684; <http://dx.doi.org/10.1128/JB.01953-05>
- Lindahl G, Sunshine M. Excision-deficient mutants of bacteriophage P2. *Virology* 1972; 49:180-7; PMID:4556923; [http://dx.doi.org/10.1016/S0042-6822\(72\)80019-9](http://dx.doi.org/10.1016/S0042-6822(72)80019-9)
- Liu T, Haggård-Ljungquist E. The transcriptional switch of bacteriophage WΦ, a P2-related but heteroimmune coliphage. *J Virol* 1999; 73:9816-26; PMID:10559293
- Reed MR, Shearwin FE, Pell LM, Egan JB. The dual role of *Apl* in prophage induction of coliphage 186. *Mol Microbiol* 1997; 23:669-81; PMID:9157239; <http://dx.doi.org/10.1046/j.1365-2958.1997.2521620.x>
- Saha S, Haggård-Ljungquist E, Nordström K. The Cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J* 1987; 6:3191-9; PMID:2826134
- Eriksson JM. Structure-function studies of bacteriophage P2 integrase and cox protein. Department of Genetics. Stockholm: Stockholm University 2005; 82.
- Mandali S, Cardoso-Palacios C, Sylwan L, Haggård-Ljungquist E. Characterization of the site-specific recombination system of phage PhiD145, and its capacity to promote recombination in human cells. *Virology* 2010; 408:64-70; PMID:20875907; <http://dx.doi.org/10.1016/j.virol.2010.08.035>
- Woods WH, Egan JB. Integration site of noninducible coliphage 186. *J Bacteriol* 1972; 111:303-7; PMID:4559723
- Hauser MA, Scocca JJ. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1: Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J Biol Chem* 1992; 267:6859-64; PMID:1551893
- Pelludat C, Mirol S, Hardt WD. The SopEPhi Phage Integrates into the *ssrA* Gene of *Salmonella enterica* Serovar Typhimurium A36 and Is Closely Related to the Fels-2 Prophage. *J Bacteriol* 2003; 185:5182-91; PMID:12923091; <http://dx.doi.org/10.1128/JB.185.17.5182-91.2003>
- Dulebohn D, Choy J, Sundermeier T, Okan N, Karzai AW. Trans-translation: the tmRNA-mediated surveillance mechanism for ribosome rescue, directed protein degradation and nonstop mRNA decay. *Biochemistry* 2007; 46:4681-93; PMID:17397189; <http://dx.doi.org/10.1021/bi6026055>
- Garcia E, Chain P, Elliott JM, Bobrov AG, Motin VL, Kirillina O, et al. Molecular characterization of L-413C, a P2-related plague diagnostic bacteriophage. *Virology* 2008; 372:85-96; PMID:18045639; <http://dx.doi.org/10.1016/j.virol.2007.10.032>
- Nilsson AS, Karlsson JL, Haggård-Ljungquist E. Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. *Mol Biol Evol* 2004; 21:1-13; PMID:12949155; <http://dx.doi.org/10.1093/molbev/msg223>
- Massad T, Skaar K, Nilsson H, Damberg P, Henriksson-Peltola P, Haggård-Ljungquist E, et al. Crystal structure of the P2 C-repressor: a binder of non-palindromic direct DNA repeats. *Nucleic Acids Res* 2010; 38:7778-90; PMID:20639540; <http://dx.doi.org/10.1093/nar/gkq626>
- Pinkett HW, Shearwin KE, Stayrook S, Dodd IB, Burr T, Hochschild A, et al. The structural basis of cooperative regulation at an alternate genetic switch. *Mol Cell* 2006; 21:605-15; PMID:16507359; <http://dx.doi.org/10.1016/j.molcel.2006.01.019>
- Bertani LE, Bertani G. Genetics of P2 and related phages. *Adv Genet* 1971; 16:199-237; PMID:4947104; [http://dx.doi.org/10.1016/S0065-2660\(08\)60359-4](http://dx.doi.org/10.1016/S0065-2660(08)60359-4)
- Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts, USA: Sinauer Associates 1999.
- Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 2008; 36:197-201; PMID:18463136; <http://dx.doi.org/10.1093/nar/gkn238>
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000; 16:276-7; PMID:10827456; [http://dx.doi.org/10.1016/S0168-9525\(00\)02024-2](http://dx.doi.org/10.1016/S0168-9525(00)02024-2)
- Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 1984; 157:690-3; PMID:6363394
- Beltran P, Plock SA, Smith NH, Whittam TS, Old DC, Selander RK. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. *J Gen Microbiol* 1991; 137:601-6; PMID:2033380
- Boyd EF, Wang FS, Beltran P, Plock SA, Nelson K, Selander RK. *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. *J Gen Microbiol* 1993; 139:1125-32; PMID:8360609
- Sunshine MG, Thorn M, Gibbs W, Calendar R, Kelly B. P2 phage amber mutants: Characterization by use of a polarity suppressor. *Virology* 1971; 46:691-702; PMID:4944860; [http://dx.doi.org/10.1016/0042-6822\(71\)90071-7](http://dx.doi.org/10.1016/0042-6822(71)90071-7)
- Appleyard RK. Segregation of New Lysogenic Types during Growth of a Doubly Lysogenic Strain Derived from *Escherichia coli* K12. *Genetics* 1954; 39:440-52; PMID:17247495
- Bertani LE. Lysogenic conversion by bacteriophage P2 resulting in an increased sensitivity of *Escherichia coli* to 5-fluorodeoxyuridine. *Biochim Biophys Acta* 1964; 87:631-40; PMID:14220693

Note

Supplemental material can be found at:
www.landesbioscience.com/journals/bacteriophage/article/18470