

Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs

Anne Lopes^{1,2}, Jihane Amarir-Bouhram³, Guilhem Faure^{1,2}, Marie-Agnès Petit^{3,*} and Raphaël Guerois^{1,2,*}

¹CEA, iBiTecS, F-91191 Gif sur Yvette, ²CNRS, URA 2096, Gif-sur-Yvette, F-91191 and ³INRA, UMR 1319, Micalis, 78350 Jouy en Josas, France

Received July 27, 2009; Revised February 2, 2010; Accepted February 3, 2010

ABSTRACT

Homologous recombination is a key in contributing to bacteriophages genome repair, circularization and replication. No less than six kinds of recombinase genes have been reported so far in bacteriophage genomes, two (UvsX and Gp2.5) from virulent, and four (Sak, Red β , Erf and Sak4) from temperate phages. Using profile–profile comparisons, structure-based modelling and gene-context analyses, we provide new views on the global landscape of recombinases in 465 bacteriophages. We show that Sak, Red β and Erf belong to a common large superfamily adopting a shortcut Rad52-like fold. Remote homologs of Sak4 are predicted to adopt a shortcut Rad51/RecA fold and are discovered widespread among phage genomes. Unexpectedly, within temperate phages, gene-context analyses also pinpointed the presence of distant Gp2.5 homologs, believed to be restricted to virulent phages. All in all, three major superfamilies of phage recombinases emerged either related to Rad52-like, Rad51-like or Gp2.5-like proteins. For two newly detected recombinases belonging to the Sak4 and Gp2.5 families, we provide experimental evidence of their recombination activity *in vivo*. Temperate versus virulent lifestyle together with the importance of genome mosaicism is discussed in the light of these novel recombinases. Screening for these recombinases in genomes can be performed at <http://biodev.extra.cea.fr/virfam>.

INTRODUCTION

The interest for bacteriophages, viruses infecting bacteria, and for viruses in general, is ever-growing with the discovery of their ubiquity in various natural ecosystems (1–4), their diversity (5–7), and their probable regulatory role in the equilibrium of bacterial populations (8–11). However, phage diversity also represents a challenge when their genomes need to be annotated. At present, each new genome brings in a majority of unknown predicted proteins (5–7). Large-scale analyses with ACLAME (12,13), a database dedicated to the comparative analysis of microbial mobile elements, has confirmed the abundance of orphan proteins in all completely sequenced phage genomes (14). This diversity may appear puzzling when one considers the simplicity of the bacteriophage life cycle, which consists mainly of reproducing its genome and its capsid in a given bacterial host. Phages ought to encode a few basic functions to this end, and one may expect to identify them easily by protein alignments. The fact that the number of unknown phage genes continues to grow suggests either that many phage functions do not derive from a common ancestor or that the current tools for protein alignments are not appropriate. It has been known for a long time that both mutations (15) and recombination rates (16) are much higher in phages than in bacteria, and this may blur the homology signal (17).

For this reason, we decided to apply sensitivity-enhanced techniques to investigate in detail one of these families of phage-encoded proteins. We focused on the phage proteins that are responsible for homologous recombination, which is an essential function in numerous phages: it contributes to genome repair (18), to the circularization of genomes with terminal repetitions (19), and to replication (20). When absent, the function

*To whom correspondence should be addressed. Tel: +33 1 34 65 28 67; Fax: +33 1 34 65 20 65; Email: marie-agnes.petit@jouy.inra.fr
Correspondence may also be addressed to Raphaël Guerois. Tel: +33 1 69 08 67 17; Fax: +33 1 69 08 92 75; Email: raphael.guerois@cea.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

may be complemented by the host-encoded RecA protein. However, the conservation of this phage-encoded function suggests that phages are better served by their own recombinases, as laboratory experiments may not reflect the variety of growth condition with which the phages and their hosts are faced in nature. Virulent and temperate phages are thought to contain different recombinases. So far, six families of recombinase genes have been reported in bacteriophage genomes, two (UvsX and Gp2.5) from virulent, and four (Sak, Red β , Erf and Sak4) from temperate phages, or from ex-temperate phages that have become virulent after the loss of a lysogeny module (designated collectively hereafter as temperate). Both virulent and temperate phages adopt radically different relationships with their hosts. Virulent phages can only kill them whereas temperate can integrate within their host genome. Different temperate genomes can therefore meet in the same host, and create mosaics by recombination, an event less likely occurring in virulent phages. Recombination processes play therefore a key role in these various lifestyles and the functions of these recombinases need to be further uncovered.

Among virulent phages, UvsX, a RecA-like enzyme present in phage T4, has orthologs in other phages with large genomes and strictly virulent lifestyles (25 orthologs in the ACLAME database version 0.3, March 2008, containing 465 phage genomes). The virulent phage T7 encodes the Gp2.5 recombinase, which shares functional and structural homology with SSB proteins (21), but in contrast to SSB it also provides a recombinase function through its single-strand annealing activity (22,23). It has only seven orthologs in the ACLAME database.

The recombinases identified in temperate phages seem more diverse although they share common properties. Sak, Red β and Erf, which were delineated as three distinct families on the basis of their primary sequences (24), harbour a single-strand annealing activity. All three proteins form similar ring-like quaternary structures visible by electron microscopy (25–27). Red β from phage λ and RecT from prophage *rac* are two well-known members of the Red β -like family that are used for genetic engineering (28,29). Three striking properties distinguish these recombinases from RecA-like enzymes: (i) their ATP-independence, (ii) their high efficiency in pairing short, 40–50 bp long sequences (30) and (iii) their ability to incorporate short single-strand DNA substrates into the bacterial chromosome (31). In addition, the Red β protein was recently reported to be 100-fold more efficient than was RecA for homologous recombination between 22% divergent sequences (16). All these characteristics may account for the remarkable plasticity and mosaicism of temperate bacteriophage genomes. Interestingly, the Sak protein shares homology with the N-terminal, globular domain of the eukaryotic Rad52 protein (24,26), and a distant homology between Red β and Rad52 has also been reported (32).

It is suspected that other families of recombinases are yet to be discovered. Indeed, most sequenced phage genomes do not encode homologs of any of the aforementioned proteins (78% of phage genomes in ACLAME). A genetic study by the group of Moineau has probably

uncovered one such new family called Sak4 (33). The common characteristic of all proteins called ‘Sak’ is their sensitivity to the plasmid-encoded phage-resistance system AbiK (for abortive infection system K). This plasmid is present in some *Lactococcus lactis* bacterial strains, and confers phage resistance. Sak and Sak2 belong to the same family, and Sak3 was identified as homolog to Erf (33), but Sak4 could not be related to any known recombinase. Up to now, Sak4 has not yet been formally proven to be a recombinase, but the fact that *sak4* bacteriophage mutants are as resistant to AbiK as are *erf* or *sak* recombinase mutants argues strongly for such an activity.

In summary, with the current protein alignment tools, at least six families of phage recombinases have been found or suspected, namely UvsX, Gp2.5, Sak, Red β , Erf and Sak4. The first two systems, UvsX and Gp2.5, seem quite specific to virulent bacteriophages lifestyle, whereas the four others appear to be more broadly distributed. Using in-depth bioinformatic methodologies, coupling profile–profile comparisons, structure-based modelling and gene-context analyses, we provide access to the global landscape of recombinases in bacteriophages: based on the global analysis of 465 bacteriophage genomes archived in the ACLAME database (12,13) we show that the Sak, Red β and Erf families belong to the same Rad52 superfamily. We unveil that Sak4 homologs are widespread and correspond to a shortcut version of Rad51. Coupled with gene-context analyses, these large-scale profile–profile analyses further reveal that Gp2.5-like recombinases are also widespread in bacteriophage genomes and are not limited to relatives of T7. For two newly detected recombinases belonging to the Sak4 and Gp2.5 families, we provide experimental evidence of their recombination activity *in vivo*. We therefore conclude that, previously unsuspected, a majority of bacteriophages encode a recombinase function that can be classified into one of only three superfamilies Rad52, Rad51 or Gp2.5. The discovery of so many recombination systems, which appear to segregate mainly according to the phage temperate or virulent lifestyle, raises fascinating questions as to their DNA repair strategies, their genome diversity and their evolution.

MATERIALS AND METHODS

Bioinformatics procedures

All the 465 genomes used for this study were taken from the ACLAME database version 0.3, release of March 2008 (A CLAssification of Mobile genetic Elements) (12,13). For each protein of the 465 phages, multiple sequence alignments were built using three iterations of PSI-Blast (34) against the nr90 database (non redundant sequence database filtered at 90% identity) and were transformed into HMM profiles using the HHmake algorithm (35) resulting in a total of 28 300 HMM profiles.

The HMM profiles of four recombinases [Sak from ul36 (NP_663647), Red β from 933W (NP_049474), Erf from D3 (NP_061548) and Sak4 from ϕ 31 (AAC48871)] were used as initial queries to screen the 28 300 profiles of the

465 genomes to identify homologs using the profile–profile comparison program HHsearch (v1.6.0), also called HHpred (35). Accordingly, a HHsearch hit was considered as significant when probabilities for being a true positive exceeded 90% with both global and local alignment modes (see Supplementary Figure S1 for justification of the threshold). The same thresholds of 90% were used in the next iterations, in which previously detected profiles were in turn compared to the whole-profile database until convergence is reached. Sak4 identification further required validation because it is composed of an ATPase domain which is a widely represented superfamily. Using Sak4 profile as a query matched not only Sak4 homologs but also other ATPase such as RecA proteins or DnaB helicases. Sak4 proteins are shorter than RecA or DnaB, which contain a DNA binding or a DnaB-like domain in their C-terminal region, respectively. To prevent the selection of any false positive in the assignment of Sak4 homologs, only short sequences (less than 290 residues), excluding the possibility of RecA or DnaB C-terminal domains, were assigned as Sak4.

Models of Sak, Red β and Erf undecamers (11-mer), were generated with Modeller 9v5 (36) using the human Rad52 N-terminal domain as template (PDB:1h2i). Initial alignments between one monomer of each recombinase and the template were obtained from profile–profile comparison (35). The sequence divergence between the three recombinases and the Rad52 template was such that several regions had to be carefully re-aligned in order to provide sequence alignments consistent with the structural constraints imposed by the Rad52-fold and optimizing the scores of Verify3D (37) and Prosa (38) methods assessing the 3D models likelihood (see scores in Supplementary Table S1 and Figure S2). The same procedure was followed for generating Sak4 and Gp2.5 models using RadB from *Thermococcus kodakarensis* (PDB:2cvh) and Gp2.5 from phage T7 (PDB:1je5) as templates, respectively. The multiple sequence alignments were processed with Jalview (39). Structural models were represented using PyMol (40). Conservation analyses were carried out using the Rate4Site algorithm (41). Profile–profile comparisons against the recombinase profiles database can be run and all profiles, models and alignments can be downloaded at <http://biodev.extra.cea.fr/virfam/downloads.html>.

Gene-context analyses were run considering the 20 genes upstream and downstream each of the 133 recombinases identified above. Profile–profile comparisons between the profiles of these genes and the 28 300 profiles database were performed as described above. Related genes were clustered together using full linkage hierarchical agglomerative clustering. This procedure was repeated for all the phage genomes in which no recombinase was detected focusing on the DNA replication modules.

Single-strand annealing assay

In the low-copy number and thermosensitive plasmid pKD46, *red β* of phage λ is cloned under a *ParaB* promoter, and its transcription is induced by AraC in

the presence of arabinose (28). The pKD46 backbone was used to replace the λ fragment encompassing *gam*, *red β* and *red α* by *uvrX* of phage T4, *gp2.5* of phage T7 or phage ϕ 12, or *sak4* of phage PA73, each preceded by its RBS, between the *EcoRI* and *NcoI* sites of pKD46. *EcoRI* and *NcoI* sites were artificially introduced at the 5' and 3' extremities of the recombinase genes, respectively, by PCR. Site-directed mutagenesis was realized by PCR, designing two 40-bp long complementary oligonucleotides centred on the mutation to be added to the gene. Left- and right-arm of the gene were then amplified separately, and both fragments were then combined in a third PCR to amplify the whole gene. All constructions were verified by sequencing. Plasmids, strains and bacteriophages used for the assay are reported in Supplementary Table S2.

Electro-competent cells were prepared essentially as described by the Biorad manufacturer, except that cells were grown at 30°C in SLB (5 g/l NaCl, 20 g/l yeast extract, 35 g/l bacto-tryptone and 1 ml/l NaOH 5 N). Ninety minutes prior harvesting, the recombinase was induced by addition of 0.2% arabinose. After 1 h expression of the electroporated cells in SOC medium supplemented with 0.2% glucose, the cultures were further incubated over-night without agitation at room temperature, before plating on LB supplemented with 50 μ g/ml rifampicin, and incubating at 37°C. All competent cells had similar levels of competence of around $2 (\pm 1) \times 10^7$ transformants per microgram, as estimated with the pACYC184 plasmid.

The 51-nt long oligonucleotide Maj32 is centred on the nucleotide 1576, starting from the ATG of the *rpoB* gene of *E. coli* MG1655, and corresponds to the non-coding strand. In this oligonucleotide, a G is changed to a C relative to the wild-type sequence at position 1576, so that the CAC codon on the coding strand becomes GAC, and the RpoB protein has a H526D mutation, which confers resistance to rifampicin (42). Maj32 was ordered from Eurogentec, as desalted and purified, and resuspended at a concentration of 1 mg/ml. Increasing amounts of the oligonucleotide mixture were tested, to determine the saturating concentration above which no more transformants could be obtained with the Red β positive control. Ten micrograms of Maj32 were found to be saturating.

RESULTS

Remote homology detection of recombinases among phage genomes

Three distinct families of recombinases, represented by Red β , Erf, Sak proteins, can be distinguished in bacteriophages, according to the classification scheme of Koonin's group (24) and a fourth one, Sak4, much less well-characterized, has emerged from recent experimental work (33). First, the presence of any of these recombinases among sequenced bacteriophages was assessed based on the 465 genomes archived in the ACLAME database (12,13). Profile–profile comparisons relying on algorithms such as HHsearch (35), are among the most powerful methods to search for homologs among rapidly diverging

sequences. Here, the efficiency of the approach was further enhanced by implementing an iterative and systematic profile–profile strategy. Every gene of every bacteriophage was used as a query to build a multiple sequence alignment so as to create a database of 28 300 profiles. This large ensemble of profiles was screened against each of the four Red β , Erf, Sak and Sak4 recombinase profiles using the HHsearch algorithm (35). This global analysis showed that 126 phages harboured a remote homolog of one of the recombinase superfamilies (Figure 1, thick coloured edges). Among them, 2 Red β -like, 5 Erf-like, 21 Sak-like and 40 Sak4-like were found, which had not been identified in the ACLAME database as recombinases (circles with thick outline in Figure 1). Using each of the 126 profiles as queries, a second iteration of profile–profile comparisons against the 28 300-profile database was sufficient to reach convergence. Seven additional bacteriophages with Sak or Red β recombinase were retrieved. Unexpectedly, this second iteration also

revealed that remote homologs of one recombinase family could match remote homologs of another family (shown as black connections between crowns in Figure 1).

As reported in Figure 1, three distinct recombinase families, namely Red β , Erf and Sak, were connected through remote homology relationships, whereas the Sak4 group of homologs remained isolated. These relationships strongly support the notion in that the three recombinases families belong in fact to a unique superfamily. This result is all the most surprising that none of the direct profile–profile comparisons between the initial Red β , Erf and Sak profiles revealed any significant remote homology signal. Hence, detection of remote homologies between drastically diverged sequences can be considerably enhanced through iterative all-against-all profile–profile comparisons and search for transitive relationships. In Supplementary Table S3, all retrieved recombinases are listed.

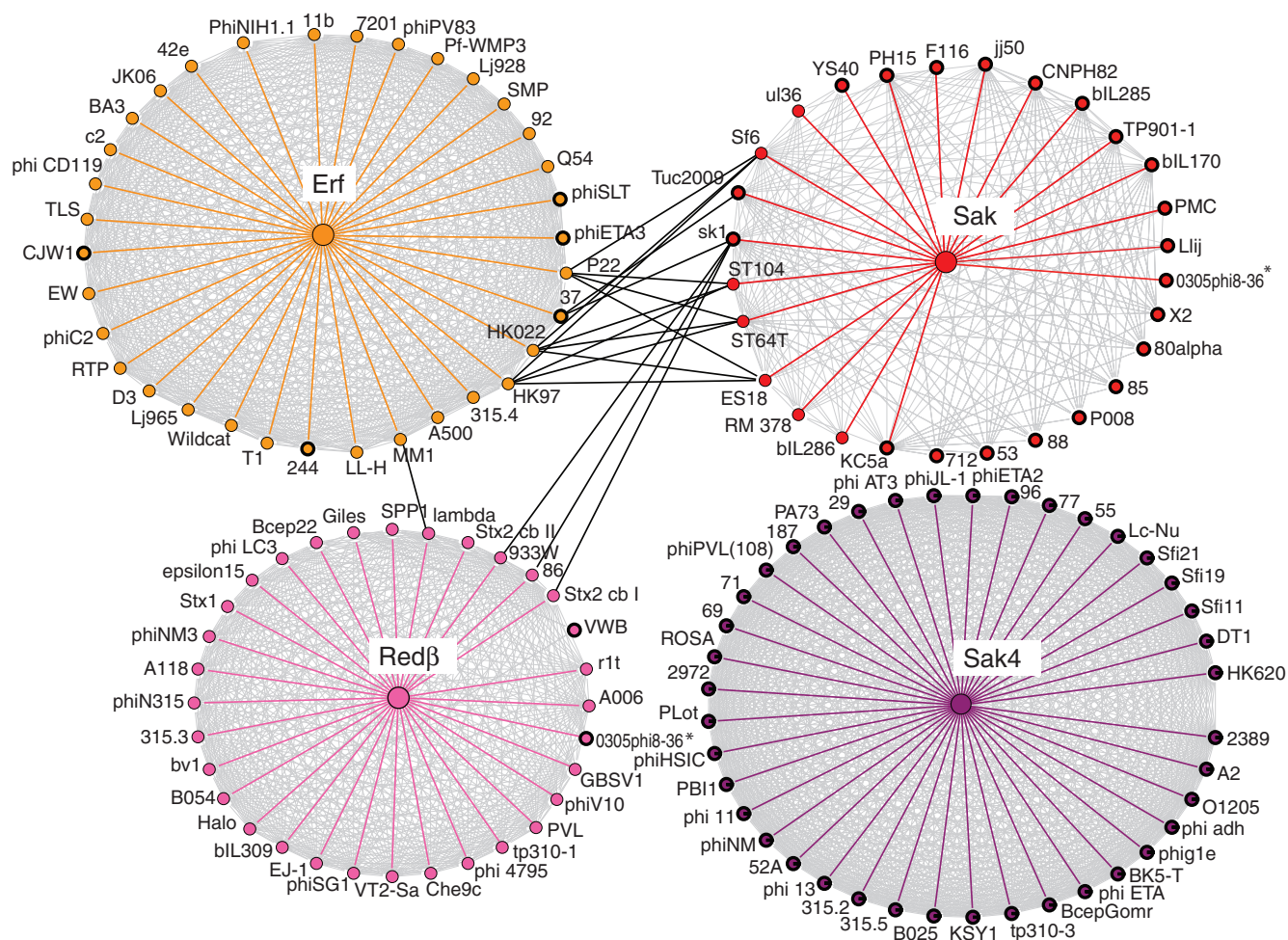


Figure 1. Homology relationships between recombinases revealed by iterative profile–profile comparisons. Network of homology relationships detected between the four recombinases superfamilies Red β /RecT (magenta), Erf (orange), Sak (red) and Sak4 (purple). Nodes represent the bacteriophage genomes for which a homologous recombinase was detected. Edges connect recombinases which were found homologous from the profile–profile comparison using HHsearch. Bold coloured edges correspond to significant hit obtained from the initial screening of the bacteriophage 28 300 profiles database with the four recombinase profiles. A second iteration using the detected recombinase profiles as queries revealed additional relationships among the same superfamily (light grey edges) and, most importantly, between different superfamilies (black edges). Nodes with a black outline pinpoint novel recombinases annotated as unknown function so far. Graph created using the Osprey program (53).

The structures for three families of recombinases are predicted to correspond to mini-Rad52

The remote homology between Sak, Erf and Red β proteins implies they adopt related folds and a structural model would provide critical insights into their common properties. Fortunately, a homology between Sak family and the human Rad52 was initially detected through sequence alignments (24) and further assessed recently through several experimental characterizations (26). The structure of the N-terminal domain of human Rad52 was solved and the domain was found to oligomerize into a ring-like shape bearing 11 subunits (43,44). In solution, full-length Rad52 forms homo-oligomers of seven or more subunits (45). Similarly, electronic microscopy visualization of Sak proteins revealed that they assemble into 11-mer stacked ring structures (26). A structural model of the monomeric Sak protein and of its oligomeric form could be generated using the structure of human Rad52 as a template. With respect to Red β and Erf, a structural model could also be generated based on the above assessed homology. Structural assessment of the models guided the optimization of the alignment (see 'Material and Methods' section) (Figure 2A) and the model obtained for Red β is compared to the Rad52

N-terminus structure in Figure 2B as a ribbon representation.

The major characteristic of the three recombinases with respect to Rad52 core structure is to match a shortcut Rad52 fold. Significant structural deletions are found in the main helix (α 3) and the three strands β 3– β 4– β 5 wrapping around α 3 (dashed regions in the secondary structure cartoon in Figure 2A). Such a deletion in the fold probably accounts for the difficulties encountered so far in detecting homologies among the Rad52-like superfamily. Interestingly, although the fold has been significantly rearranged through evolution, some specific features remained conserved in all these recombinases. Figure 2C maps the sequence conservation index at the surface of the recombinase models as calculated by the Rate4site algorithm (41). The clusters of red/orange positions in the groove testify to the specific conservation and functional importance of this region. In particular, one position, indicated by a red star in Figure 2A–C, is strictly maintained as a positively charged residue in all three recombinases. Mutation of the corresponding amino acid, K152, in human Rad52 was found to fully abrogate single-strand DNA binding (43). We mutated the corresponding position in Red β (R161) and reached similar

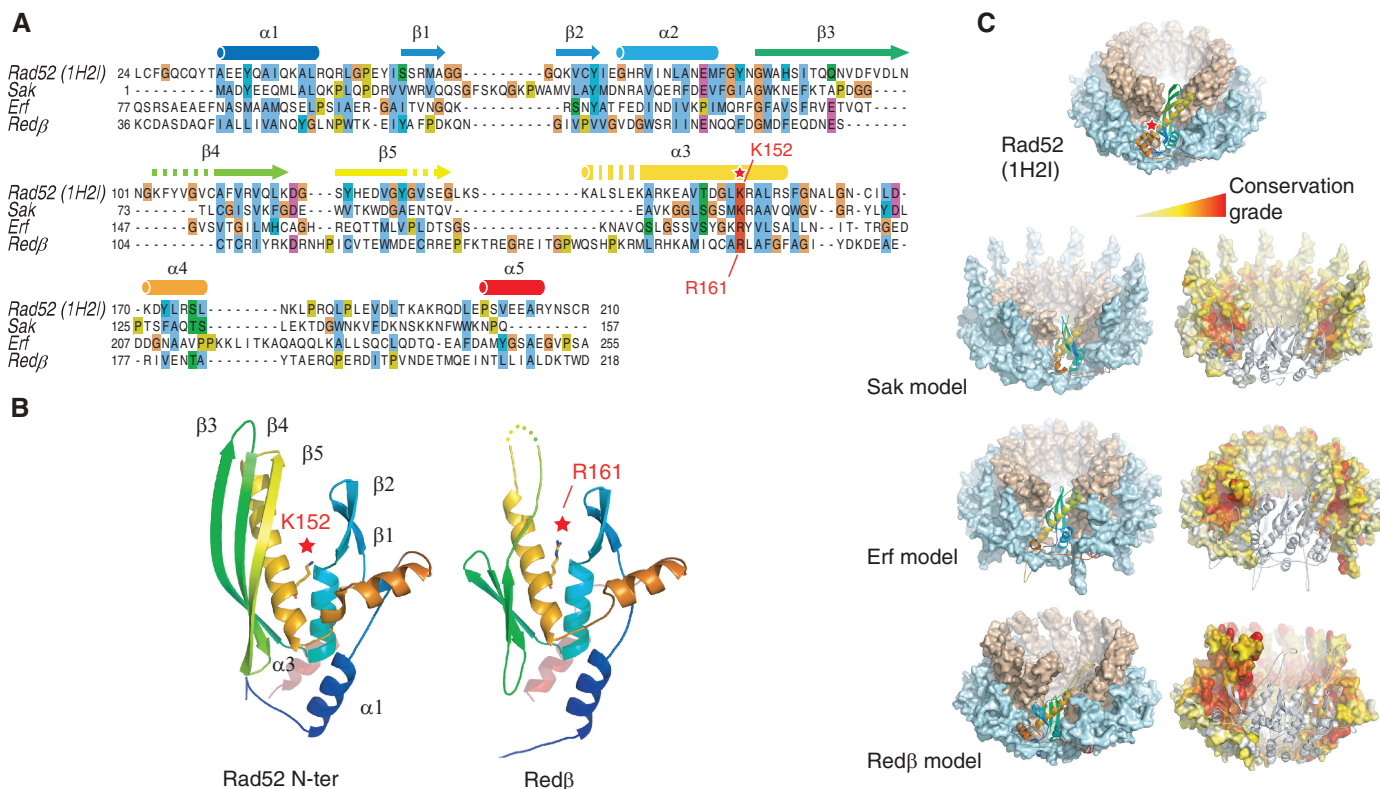


Figure 2. Structural models of the recombinases found evolutionarily related to the Rad52 superfamily. (A) Optimal sequence alignment resulting from the structural modelling procedure between the sequences of human Rad52 corresponding to the PDB (1H2I), Sak (from phage ϕ 36), Erf (D3), Red β (933W). Secondary structures are indicated on top of the alignment from blue to red colour as in (B and C). Truncated secondary structures between Rad52 and the phage recombinases are indicated by dashed secondary structures. The red star indicates the position of K152 in Rad52 whose mutation abrogated single-strand binding. (B) Cartoon representation of the Rad52 crystallized domain and of the Red β structural model with the red star pinpointing the side-chain of K152 and R161, respectively. (C) Surface representation of Rad52 11-mer oligomeric form as observed in the crystal structure 1H2I together with the 11-mer models of the three recombinases Sak, Erf and Red β . On the left, the surface is coloured in either light cyan or orange to help visualizing the external and internal ring, respectively. On the right the surface is coloured with respect to the conservation index calculated using Rate4Site program, conservation grade increasing from white to red.

conclusion (see ‘Experimental validation’ section). Hence, despite their sequence divergence, a remarkable selection pressure is apparent in the groove, strongly suggesting that bacteriophage recombinases function similarly to Rad52 in binding single-stranded DNA and promoting strand annealing.

Identification of Sak4 as a mini-Rad51

Sak4, the fourth recombinase, remained isolated in Figure 1 and could not be related to any feature of the Rad52 fold. In contrast, comparison of Sak4 profile against a database of PDB-derived profiles, revealed an unambiguous homology to the RecA-RadA-Rad51 family with the highest probability for the structure of the archeal RadB paralog. Sak4 only shares 15% identity with RecA and Rad51, but bears both Walker A and Walker B motifs involved in ATP binding, as shown in the alignment of Rad51 paralogs in Figure 3A. A striking feature of Sak4 is its small size, with 245 residues in bacteriophage ϕ 31 representing only two-thirds of the RecA or Rad51 proteins.

Our analysis brought to light 40 bacteriophages harbouring the Sak4 homolog and all of them shared this property of reduced size (Supplementary Table S3). How can the RecA or Rad51 fold tolerate such a drastic size reduction? First, only the region of the ATPase domain matched the Sak4 profile leaving no room for a secondary DNA binding domain as found in the Rad51 N-terminus and RecA C-terminus (Figure 3). Furthermore, specific deletions of helical turns need to be considered to permit a correct alignment (helices α 1, α 3 and α 4). A structural model accounting for these deletions could be reliably generated illustrating the architecture of Sak4 as a mini-RecA protein (Figure 3B). The absence of the secondary DNA-binding domain is also found in other Rad51 paralogs such as RadB in Archae, whose specific function with respect to RadA remains unclear (46), or XRCC2, which is probably a Rad51 cofactor (47). In the RecA protein, the secondary DNA binding domain is proposed to facilitate strand-exchange by retaining the displaced strand away from the heteroduplex (48). It may be, therefore, that the Sak4

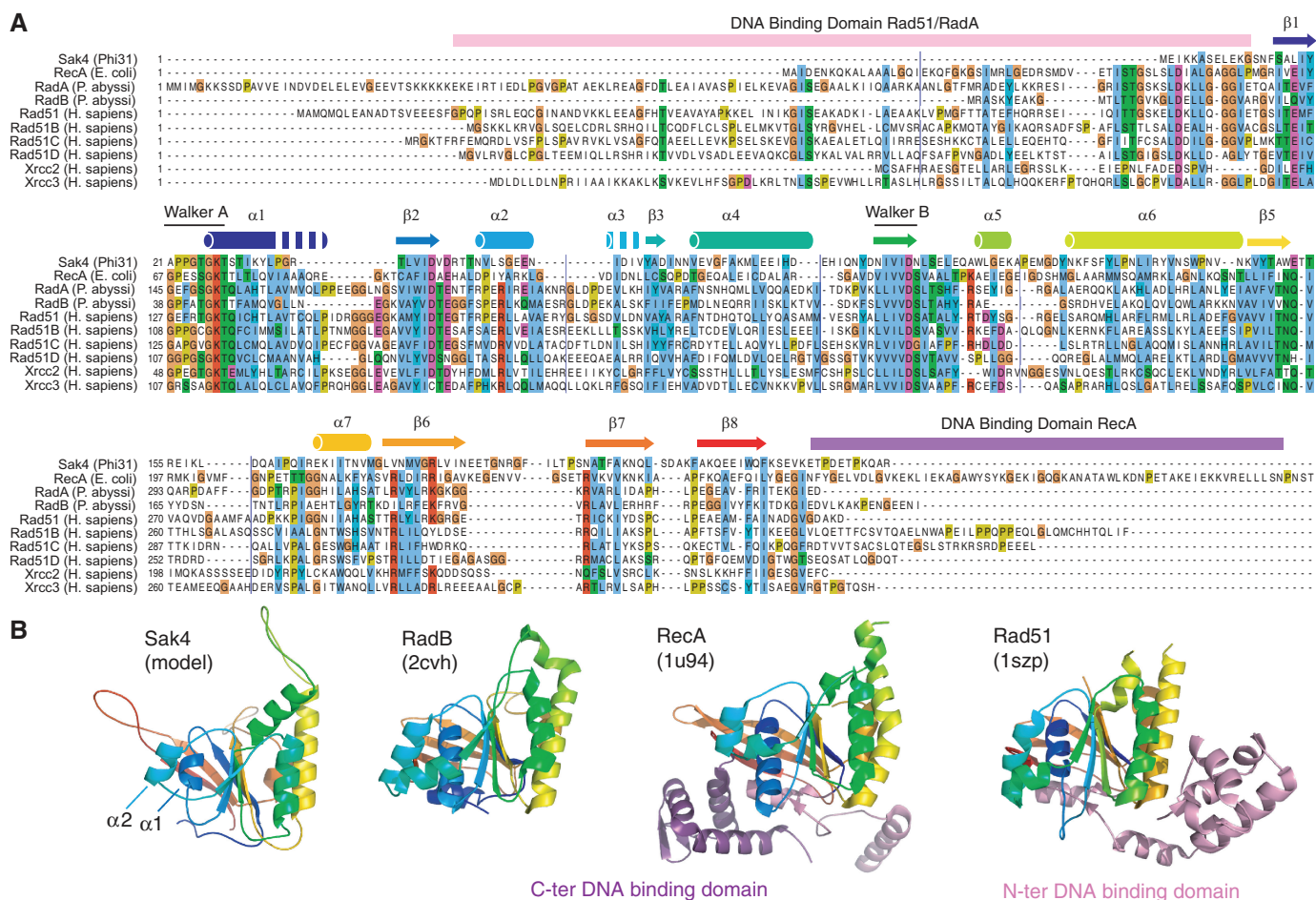


Figure 3. Structural Model of Sak4, a shortcut homolog of RecA/Rad51 superfamily. (A) Comparison between Sak4 from *Lactococcus* phage Φ 31 (gb:AA48871) and other members of the RecA/RadA/Rad51 superfamily, *E. coli* RecA (B7N6S9), *Pyrococcus abyssi* RadA (Q9V233) and RadB (Q9V2F6), human Rad51 (Q06609), Rad51B (O15315), Rad51C (O43502), Rad51D (O75771), XRCC2 (O43543) and XRCC3 (O43542). Secondary structures are indicated on top of the alignment from blue to red; pink and purple bars indicate the position of additional domains. (B) Ribbon representation of the structural model of Sak4 from *Lactococcus* phage Φ 31 compared to that of RadB (pdb 2cvh), RecA (pdb 1u94) and Rad51 (pdb 1szp) X-ray structures.

proteins have lost (or have never acquired) the capacity to promote strand exchange, and maintain only a single-strand annealing activity.

Alternative recombinase systems revealed by combining profile-profile and gene-context analyses

The architecture of the bacteriophage genomes often exhibits a conserved arrangement of gene functions. At a first level, genomes can typically be segmented into modules comprising either head or tail assembly factors, DNA replication machinery or cell lysis factors. Analysis of the gene neighbours for the 133 recombinases detected either as mini-Rad52 (Red β , Erf, Sak) or as mini-Rad51 (Sak4) analysis confirmed that the recombinase was located within a replication module (Supplementary Figure S4, Tables S4 and S5). Furthermore, we found examples of substitutions between the different recombinase families as illustrated in Figures 4A and Supplementary Figure S5. Can the pattern of genes surrounding these recombinases be used further to identify alternative recombinase systems in other phage genomes?

A key step in the gene-context analysis remains the identification of homologous proteins in the vicinity of the recombinases genes. Here again, high sequence divergence among the neighbouring sequences may limit the analysis. We pushed back the efficiency of the homology detection among the neighbours of the recombinases by applying the profile-profile comparisons protocol previously described, focusing on the genes in the DNA replication module. In particular, we looked for orthologs of RecE, a 5'-3'-exonuclease which prepares the single-strand substrate for the recombinase to perform the strand annealing. This protocol revealed 84 genomes containing a *recE*-like homolog, while only 26 could be detected through PSI-Blast analyses (yellow and red striped boxes in Figure 4A). However, among these 84 genomes, only 30 contained both a RecE-like exonuclease and a recombinase (Red β , Erf or Sak4) (shown as pink or dark pink in Figure 4A), suggesting that some other recombinase family may be hidden in the vicinity of the remaining genomes.

Indeed, in 19 genomes, a specific set of homologous but unknown proteins was detected, each of them clustering with a *recE*-like gene. One example of such a protein, found in bacteriophage ϕ 12, is indicated by a white box outlined in black in Figure 4A. Comparing the profiles of these 19 proteins against a database of PDB-derived profiles, revealed an unambiguous homology to the gene protein 2.5 (Gp2.5) of bacteriophage T7. Gp2.5 possesses a strand annealing activity which contributes to genetic recombination during growth of T7 phage (22,23). The structure of Gp2.5 from phage T7 showed that it contains an OB-fold (21) but differs from other SSBs by the presence of the long α 1 helix represented in Figure 4B. The R82C mutation (indicated by a red star) located just downstream of this helix specifically abrogated the single-strand annealing activity suggesting that this extra helix is important for this property (49). In the Gp2.5-like protein of phage ϕ 12, the long helix is also predicted and a conserved basic residue is found at the position

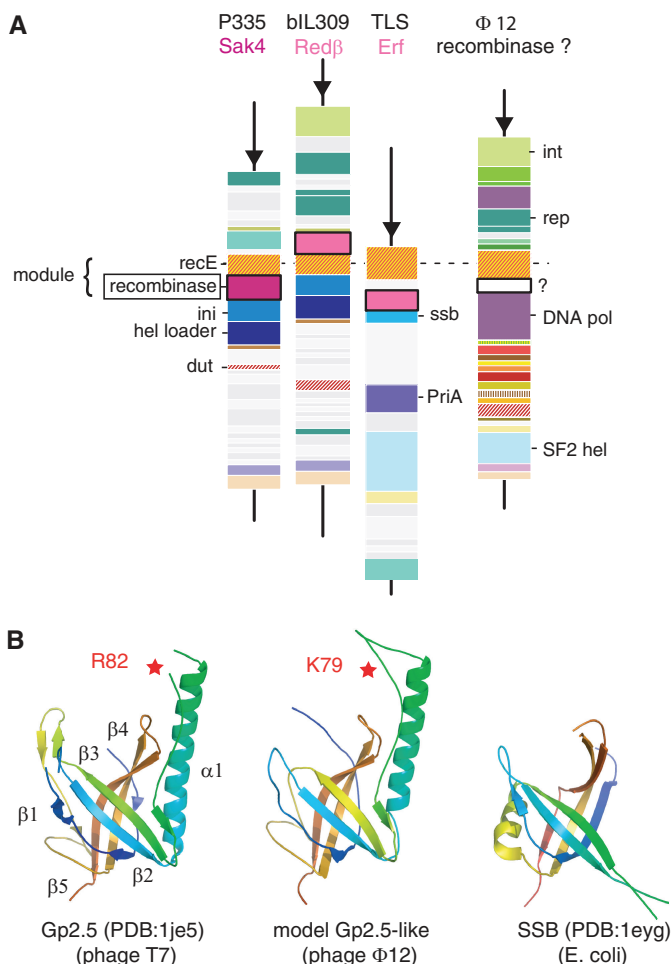


Figure 4. Analysis of the gene context in the DNA replication and recombination module of several bacteriophages. (see also Supplementary Figure S5) (A) Phage names are indicated on top with the type of detected recombinase annotated below. ORFs are represented by ordered boxes with size related to the ORF length. Light and medium grey boxes indicate genes not considered in the analysis. The different predicted recombinases are colored as pink. Other colours, when identical, pinpoint genes sharing a homology relationship (either remote or close). Genes surrounding the homolog of 5'-3' exonuclease RecE (yellow/red stripes) in P335 *sensu lato* (in fact, phage 4268), bIL309, TLS and ϕ 12 phages. A Rad52 (pink box) or Rad51-like (dark pink box) recombinase is often found in a close neighbourhood of RecE, suggesting that in ϕ 12 (the white gene labelled with a question mark) could itself be a recombinase. Some putative functions have been labelled by a short-term, *ini*, initiator; *hel loader*, DNA-C type helicase loader; *ssb*, single-strand binding; *RecE*, 5'-3' exonuclease *RecE*; *int*, phage integrase; *repr*, repressor; *DNA pol*, DNA polymerase (Pol I-type); *dut*, dUTPase; *endo*, endonuclease; *SF2 hel*, superfamily 2 helicase; *PriA*, DNA primase. (B) Structural model of the Gp2.5-like protein identified in phage ϕ 12. From left to right, ribbon representation of (i) the structure of Gp2.5 protein (phage T7) (PDB code: 1je5), (ii) the structural model of Gp2.5 distant homolog from phage ϕ 12 and (iii) the structure of the single-strand binding (SSB) protein of *E. coli* (PDB code: 1eyg). Red stars in Gp2.5 structures pinpoint residues whose mutation abrogated single strand annealing activity [in (49) and in this study].

corresponding to R82 in Gp2.5 of phage T7. Hence, not only the existence of the helix but also the conservation of important positions suggest that protein Phi12P11 (NP_803317) of phage ϕ 12 belongs to the specific Gp2.5 superfamily rather than to the larger SSB family.

As with Rad51- or Rad52-like recombinases, remote homologs of Gp2.5 are not always associated with RecE and the profile-profile comparison revealed three additional Gp2.5. All in all, 30 phages were found to contain a Gp2.5-like recombinase (Supplementary Table S3), underscoring the importance of this class of recombinase, besides the Rad52- and Rad51-like proteins.

Predicted Sak4 and Gp2.5 recombinases have single-strand annealing activities *in vivo*

It has been reported that various phage recombinases of the Rad52-like superfamily are able to promote single-strand annealing *in vivo* in *E. coli* (50). The assay is based on a 70-nt long oligonucleotide complementary to the *gal* gene that reverts cells to a Gal⁺ phenotype. Depending on the recombinase tested, yields of recombinants varied by a factor of 3000 with Red β giving the higher ratio of recombinants (2×10^{-1} Gal⁺ per viable cell), and GP35 of a *Bacillus subtilis* prophage the lower ratio (6.5×10^{-5}) (50). We used a similar assay to determine whether the predicted recombinases of the Sak4 and Gp2.5 families were able to recombine DNA *in vivo*. To do so, we used a 51-nt long single-strand oligonucleotide complementary to the *rpoB* gene of *E. coli*, except for a single point mutation that confers resistance to rifampicin. We then tested whether the presence of one of the predicted recombinases allowed the oligonucleotide integration into the *E. coli* chromosome, upon transformation by electroporation. The *E. coli* recipient strain was chosen to be defective for *recA*, as it is known that RecA is not needed for such a reaction, and proficient for mismatch repair, to minimize the background level of spontaneous rifampicin resistant (Rif^R) clones. The *rpoB* mutation was a C:C mismatch, which escapes from the MutLSHU repair system. We have tested that *in vivo* the five recombinases are produced at similar amounts, except for Gp2.5 from T7, which is produced at 5–10 times the level of the four other ones (Supplementary Figure S6).

With Red β recombinase (of the Rad52 superfamily), a yield of 2×10^{-3} Rif^R per viable cells was obtained at saturating amounts of oligonucleotide (Figure 5, see Supplementary Table S7 for exact numbers). Next, the full size RecA-like enzyme of phage T4, UvsX and the predicted minimal Rad51-like protein of the Sak4 family from phage PA73 (infecting *Pseudomonas aeruginosa*), were compared. Both recombinases produced $1\text{--}2 \times 10^{-6}$ Rif^R transformant per cells, 2000-fold less than Red β . Nonetheless, even this lower frequency was significantly higher than the background level of Rif^R recombinants produced with the control plasmid devoid of recombinase, which was on average at 4.9×10^{-7} per viable cell. Because the Gp2.5 family has various distant homologs, two members of the family were tested, those of phages T7 and ϕ 12 (a prophage of *Staphylococcus aureus*). A yield of 10^{-5} Rif^R clones per viable cells was obtained with the T7 protein, and 10-fold less with the ϕ 12 protein. The higher yield observed with the T7 protein compared to the ϕ 12 protein may be due to its higher expression level. We conclude therefore that the predicted

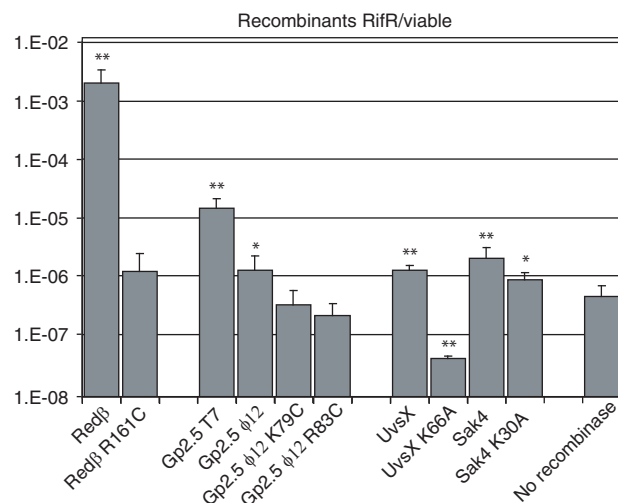


Figure 5. Single-strand annealing activities *in vivo* of the three superfamilies of recombinases. Efficiency of recombination is estimated by the ratio of rifampicin-resistant recombinants (generated by integration of the Maj32 single-strand oligonucleotide into *rpoB*), per viable cells. Each value corresponds to the average of at least three experiments, performed in an *E. coli* AB1157 derivative in which a given recombinase (*x*-axis) has been induced from a low-copy number pSC101 plasmid derivative. Spontaneous rifampicin-resistant mutants were obtained at a median frequency of 2×10^{-8} . To ascertain the effect of each recombinase, as compared to the empty plasmid vector ('No Recombinase'), a Student test was performed: double asterisks indicate significance at the 1% level, single asterisk indicate significance at the 5% level.

recombinases of phages PA73 and ϕ 12 promote single-strand annealing *in vivo*, albeit with a reduced efficiency with respect to Red β . The reduced efficiency may be related to the heterologous expression of these proteins in *E. coli*, and more complete studies are required to characterize fully these new recombinases.

Alignments of recombinases of the Rad52 family highlighted a conserved arginine/lysine residue across all sub-families (Figure 2, red star). A mutation converting this arginine into a cysteine (R161C) in the Red β gene led to a 1000-fold reduction of *in vivo* annealing efficiency (Figure 5 and Supplementary Table S7), confirming the critical role of this positively charged residue *in vivo*. The structure function analysis of Gp2.5 revealed the R82C substitution as affecting specifically the single strand annealing activity of the protein *in vitro* (49). A substitution affecting the lysine residue located at the equivalent position in the *gp2.5* gene of phage ϕ 12 (K79C) was therefore tested, as well as a neighbouring position, more conserved, R83C (Figure 4B and Supplementary Figure S3). Both protein variants produced a yield of recombinants undistinguishable from the background value. This confirmed the importance of both residues in the ϕ 12 protein, as well as the significance of the low activity detected for the wild-type ϕ 12 protein. Mutants in the Walker A motif of the ATPase of UvsX and Sak4 proteins were done, and interestingly the recombination activity of the Sak4 mutant was not affected at all, in contrast to the UvsX mutant (Figure 5 and Supplementary Table S7). The property of Sak4 is

reminiscent of the behaviour reported for XRCC2, a short version paralog of Rad51 which regulates the length of gene conversion tracts in vertebrate cells, irrespective of its ATPase activity (51).

DISCUSSION

Three major superfamilies of phage recombinases

Our systematic analysis of 465 bacteriophages combining sensitivity-enhanced homology searches, structural modelling and gene contextual analyses revealed many insights into the nature of the homologous recombination systems in bacteriophages (Supplementary Table S3 and Figure S7). First, we were able to draw a unified picture for most recombinases experimentally studied so far. We assessed with high confidence that the Rad52 fold has been substantially rearranged and shortcut throughout bacteriophage evolution, giving rise to a wide variety of recombinases that were previously thought to belong to different families (24). This feature has also been highlighted in two recent works for Sak (26) and Red β (32). In our work, careful structural modelling leads to optimized alignments. As a result, our Red β /Rad52 alignment (Figure 2A) predicts R161 as a hot functional residue in Red β , whereas this residue was not aligned with conserved positions in (32).

Strongly supported by gene-context analyses, we further discovered two unsuspected recombinase systems evolutionarily related either to Rad51 or to Gp2.5 superfamilies. Most members of the Rad51-like family were orthologs of Sak4, which is again a shortcut version of Rad51, and substituted extensively for Rad52 homologs in similar gene contexts. Experimentally, we validated that newly identified members of the Rad51-like and Gp2.5-like superfamilies do act as recombinase *in vivo*. Point mutations abrogating Red β (R161C) and Gp2.5 (R79C, R83C) activity were successfully designed and validate the alignments optimized with the help of the structural templates. The huge majority of genomes have at most one recombinase gene. This almost exclusive presence likely reflects their similar tasks in all these bacteriophages (Supplementary Table S3). The four exceptions (Phages CJW1 and 244 from *Mycobacterium smegmatis*, 0305phi8-36 from *Bacillus cereus*, and YS40 from *Thermus thermophilus*) include cases where a Rad51-like gene is combined with a Rad52-like gene, a situation reminiscent of the eukaryotic kingdom.

Recombinase distribution correlates with genome size

Are these recombinases found in certain classes of bacteriophages or are they randomly distributed? As regards genomes size, all genomes of <20 kb (140 genomes), including single-strand DNA genomes and RNA genomes, were devoid of recombinase. In the 20–80 kb range, 160 genomes out of 286 were identified with a recombinase from either the Rad52-like, Sak4-like or Gp2.5-like superfamily. Rad52-like and Sak4-like are almost exclusively integrated in temperate phages (P -value of $5.5 \cdot 10^{-5}$ and $1.2 \cdot 10^{-4}$, respectively), whereas Gp2.5-like are spread among both temperate and virulent

phages (Supplementary Figure S8B and Table S3). Within the 39 genomes of >80 kb, 24 have a UvsX/RecA-like recombinase, whereas only one have a Rad52-like one. In contrast to Sak4, UvsX/RecA recombinases are significantly associated with virulent phages (P -value of $2 \cdot 10^{-6}$). The presence of a recombinase is therefore strongly correlated with the phage virulence or temperate character (Supplementary Figure S8B and Table S3). Overall, 57% of the genomes >20 kb, have now a recombinase.

Whether large genomes with no detected recombinase (Supplementary Table S6) are really devoid of recombinase function, or encode an undetected one remains an open question. Phage genomes with terminal repeats, if being really orphan of a recombinase gene are not expected to grow in a *recA* host. A way to screen for putative new families of recombinase in the future will consist in testing whether orphan phages with terminal repeats, especially those encoding the *recE*-like accessory gene, grow in *recA* hosts.

Temperate phages and mosaicism

As mentioned in the 'Introduction' section, a major difference between UvsX/RecA and a recombinase such as Red β lies in the fidelity and the stringency of the recombination process. Red β was shown 100-fold more efficient than was RecA for homeologous recombination between 22% divergent sequences (16). We proposed earlier that a manifestation of this reduced stringency in Red β -containing phages may be the existence of genome mosaicism characterized by the presence of dispersed and almost identical segments recently exchanged between divergent phages infecting the same host (16). In contrast, comparative genomics of virulent T4-like genomes (which contain UvsX) gives no evidence for such mosaicism (52). The fidelity of Sak4-like recombinases is not known at present, but our previous study on mosaicism reveals that the nine *Staphylococcus aureus* phages (16) encompassing a Sak4-like recombinase (namely phages 187, 69, 77, 96, ROSA, 71, 55, 29 and 52A) have a high density of mosaics (an average of 10 per comparison), as compared to the genomes encompassing a Rad52-like protein (namely phages 53, 85, 42e, 37, EW, 88, 92, X2; 6.4 mosaics per comparison in average). Accordingly, Sak4 recombinases might behave similarly to Red β towards homeologous recombination. This is unexpected, given the high fidelity of RecA, which belongs to the same superfamily as Sak4.

Hence, two members of the Rad51 superfamily, Sak4 and UvsX/RecA may be associated with very distinct genomic structures related to the fidelity of the recombination process and existence of mosaicism. Comparing the capacity of Sak4 and UvsX proteins to promote homeologous recombination *in vivo* will tell whether these proteins effectively behave differently, and which domain or sequence motifs contribute to the distinct features. Along the same lines, it will be interesting to further probe whether the various Gp2.5, quite evenly distributed among virulent and temperate phages, also play distinct roles with respect to mosaicism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Raphael Leplae for providing phage classification data and to Johannes Söding for providing the latest version of the HHsearch algorithm. We thank Colin Tinsley and Xavier Veaute for critical reading of the article and the two anonymous referees for their helpful comments and suggestions.

Conflict of interest statement. None declared.

REFERENCES

- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L., Chau, B. *et al.* (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature*, **452**, 340–343.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L. *et al.* (2008) Functional metagenomic profiling of nine biomes. *Nature*, **452**, 629–632.
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S. *et al.* (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.*, **73**, 7059–7066.
- Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M. and Mead, D. (2008) Assembly of viral metagenomes from Yellowstone hot springs. *Appl. Environ. Microbiol.*, **74**, 4164–4174.
- Kwan, T., Liu, J., DuBow, M., Gros, P. and Pelletier, J. (2005) The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl Acad. Sci. USA*, **102**, 5174–5179.
- Kwan, T., Liu, J., Dubow, M., Gros, P. and Pelletier, J. (2006) Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J. Bacteriol.*, **188**, 1184–1187.
- Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R. *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell*, **113**, 171–182.
- Bossi, L., Fuentes, J.A., Mora, G. and Figueroa-Bossi, N. (2003) Prophage contribution to bacterial population dynamics. *J. Bacteriol.*, **185**, 6467–6471.
- Brown, S.P., Le Chat, L., De Paepe, M. and Taddei, F. (2006) Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. *Curr. Biol.*, **16**, 2048–2052.
- Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R. *et al.* (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, **449**, 83–86.
- Pal, C., Macia, M.D., Oliver, A., Schachar, I. and Buckling, A. (2007) Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, **450**, 1079–1081.
- Leplae, R., Hebrant, A., Wodak, S.J. and Toussaint, A. (2004) ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res.*, **32**, D45–D49.
- Leplae, R., Lima-Mendez, G. and Toussaint, A. (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.*, **38**, D57–D61.
- Lima-Mendez, G., Toussaint, A. and Leplae, R. (2007) Analysis of the phage sequence space: the benefit of structured information. *Virology*, **365**, 241–249.
- Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J.F. (1998) Rates of spontaneous mutation. *Genetics*, **148**, 1667–1686.
- Martinson, J.T., Radman, M. and Petit, M.A. (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet.*, **4**, e1000065.
- Hatfull, G.F. (2008) Bacteriophage genomics. *Curr. Opin. Microbiol.*, **11**, 447–453.
- Stahl, M.M., Thomason, L., Poteete, A.R., Tarkowski, T., Kuzminov, A. and Stahl, F.W. (1997) Annealing vs. invasion in phage lambda recombination. *Genetics*, **147**, 961–977.
- Fenton, A.C. and Poteete, A.R. (1984) Genetic analysis of the erf region of the bacteriophage P22 chromosome. *Virology*, **134**, 148–160.
- Mosig, G., Gewin, J., Luder, A., Colowick, N. and Vo, D. (2001) Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc. Natl Acad. Sci. USA*, **98**, 8306–8311.
- Hollis, T., Stattel, J.M., Walther, D.S., Richardson, C.C. and Ellenberger, T. (2001) Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7. *Proc. Natl Acad. Sci. USA*, **98**, 9557–9562.
- Araki, H. and Ogawa, H. (1981) The participation of T7 DNA-binding protein in T7 genetic recombination. *Virology*, **111**, 509–515.
- Kong, D. and Richardson, C.C. (1996) Single-stranded DNA binding protein and DNA helicase of bacteriophage T7 mediate homologous DNA strand exchange. *EMBO J.*, **15**, 2010–2019.
- Iyer, L.M., Koonin, E.V. and Aravind, L. (2002) Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics*, **3**, 8.
- Passy, S.I., Yu, X., Li, Z., Radding, C.M. and Egelman, E.H. (1999) Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proc. Natl Acad. Sci. USA*, **96**, 4279–4284.
- Ploquin, M., Bransi, A., Paquet, E.R., Stasiak, A.Z., Stasiak, A., Yu, X., Cieslinska, A.M., Egelman, E.H., Moineau, S. and Masson, J.Y. (2008) Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr. Biol.*, **18**, 1142–1146.
- Poteete, A.R., Sauer, R.T. and Hendrix, R.W. (1983) Domain structure and quaternary organization of the bacteriophage P22 Erf protein. *J. Mol. Biol.*, **171**, 401–418.
- Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
- Muyrers, J.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555–1557.
- Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **97**, 5978–5983.
- Ellis, H.M., Yu, D., DiTizio, T. and Court, D.L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA*, **98**, 6742–6746.
- Erler, A., Wegmann, S., Elie-Caille, C., Bradshaw, C.R., Maresca, M., Seidel, R., Habermann, B., Muller, D.J. and Stewart, A.F. (2009) Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.*, **391**, 586–598.
- Bouchard, J.D. and Moineau, S. (2004) Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J. Bacteriol.*, **186**, 3649–3652.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.
- Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355–362.

39. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
40. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. USA DeLano Scientific, San Carlos, CA.
41. Pupko,T., Bell,R.E., Mayrose,I., Glaser,F. and Ben-Tal,N. (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18(Suppl 1)**, S71–S77.
42. Wolff,E., Kim,M., Hu,K., Yang,H. and Miller,J.H. (2004) Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* rpoB to assess the role of polymerase IV in spontaneous mutation. *J. Bacteriol.*, **186**, 2900–2905.
43. Kagawa,W., Kurumizaka,H., Ishitani,R., Fukai,S., Nureki,O., Shibata,T. and Yokoyama,S. (2002) Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol. Cell*, **10**, 359–371.
44. Singleton,M.R., Wentzell,L.M., Liu,Y., West,S.C. and Wigley,D.B. (2002) Structure of the single-strand annealing domain of human RAD52 protein. *Proc. Natl Acad. Sci. USA*, **99**, 13492–13497.
45. Stasiak,A.Z., Larquet,E., Stasiak,A., Muller,S., Engel,A., Van Dyck,E., West,S.C. and Egelman,E.H. (2000) The human Rad52 protein exists as a heptameric ring. *Curr. Biol.*, **10**, 337–340.
46. Akiba,T., Ishii,N., Rashid,N., Morikawa,M., Imanaka,T. and Harata,K. (2005) Structure of RadB recombinase from a hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1: an implication for the formation of a near-7-fold helical assembly. *Nucleic Acids Res.*, **33**, 3412–3423.
47. Liu,N., Lamerdin,J.E., Tebbs,R.S., Schild,D., Tucker,J.D., Shen,M.R., Brookman,K.W., Siciliano,M.J., Walter,C.A., Fan,W. *et al.* (1998) XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol. Cell*, **1**, 783–793.
48. Mazin,A.V. and Kowalczykowski,S.C. (1998) The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *EMBO J.*, **17**, 1161–1168.
49. Rezende,L.F., Willcox,S., Griffith,J.D. and Richardson,C.C. (2003) A single-stranded DNA-binding protein of bacteriophage T7 defective in DNA annealing. *J. Biol. Chem.*, **278**, 29098–29105.
50. Datta,S., Costantino,N., Zhou,X. and Court,D.L. (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl Acad. Sci. USA*, **105**, 1626–1631.
51. Nagaraju,G., Hartlerode,A., Kwok,A., Chandramouly,G. and Scully,R. (2009) XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversion between sister chromatids. *Mol. Cell Biol*, **29**, 4283–4294.
52. Comeau,A.M., Bertrand,C., Letarov,A., Tetart,F. and Krisch,H.M. (2007) Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology*, **362**, 384–396.
53. Breikreutz,B.J., Stark,C. and Tyers,M. (2002) Osprey: a network visualization system. *Genome Biol.*, **3**, R22.