Homologous recombination in Bacteriophages, less fidelity for more exchanges

Geoffrey Hutinet

UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 435 :
AGRICULTURE, ALIMENTATION, BIOLOGIE, ENVIRONNEMENT ET SANTÉ

Laboratoire : MICALIS

THÈSE DE DOCTORAT SUR TRAVAUX

SCIENCES AGRONOMIQUES, BIOTECHNOLOGIES AGRO-ALIMENTAIRES

par

Geoffrey HUTINET

Homologous recombination in Bacteriophages : less fidelity for more exchanges

Date de soutenance : 31/10/2014

Composition du jury :

Directeur de thèse :	Marie-Agnès PETIT	Directeur de Recherches (INRA)
Président du jury :	Cécile FAIRHEAD	Professeur (Université Paris Sud XI)
Rapporteurs :	Mart KRUPOVIC	Chargé de Recherches (Institut Pasteur)
	Stéphanie MARSIN	Directeur de Recherches (CEA)
Examinateurs :	François ENAULT	Maitre de Conférences (Université Blaise Pascale)
	Mauro MODESTI	Directeur de Recherches (CNRS)

To the truth

Remerciements

Ce que m'a appris ce doctorat sur la recherche peut se résumer par le célèbre adage d'Edward A. Murphy Jr., plus connu sous le nom de loi de Murphy, « *Anything that can go wrong, will go wrong* » (« *Tout ce qui peut mal tourner va mal tourner* »). Effectivement, la plupart des expériences menées en laboratoire se soldent par des échecs. Et des échecs j'en ai essuyé de nombreux avant d'arriver au rendu final que vous allez lire. Fort heureusement, on apprend de ses échecs, on peut alors se relever et recommencer grâces aux conseils, aux remontrances ou au réconfort des gens qui nous entourent, à la fois dans notre vie professionnelle et privée, afin de réussir à mener un projet jusqu'à son terme. C'est pourquoi je souhaite remercier les personnes qui m'ont permis de mener à bien ce projet de thèse, et de rentrer dans cet étrange, mais merveilleux, monde qu'est la recherche.

Tout d'abord, je souhaite remercier les membres de mon jury de thèse, chargés de m'évaluer pour l'obtention du grade de docteur. Merci à Mart Krupovic et Stéphanie Marsin qui ont accepté de relire et juger mon manuscrit en tant que rapporteurs, et surtout merci pour leurs commentaires plus qu'encourageants. Merci à François Enault et à Mauro Modesti d'avoir accepté d'être examinateurs. Enfin merci à Cécile Fairhead, la présidente du jury.

Je souhaite ensuite remercier les membres de mon comité de suivi de thèse, qui m'ont aidé chaque année avec leurs remarques pertinentes et pour les discussions passionnantes qui m'ont permis d'avancer. Merci donc à Xavier Veaute, Bénédicte Michel et Michael Dubow.

Merci aux membres de « l'ex-UBLO » pour m'avoir diverti quand le besoin s'en faisait sentir ! Je remercie plus particulièrement Aurélie Bobillot, Mélanie Guillemet, Renata Matos, Laureen Crouzet et Olivier Telle. Les « potins » et autres discussions, scientifiques ou non, en français ou en anglais, m'ont permis de faire des pauses aux moments nécessaires.

Merci à mon équipe d'accueil, l'équipe « phages ». Marianne De Paepe et Colin Tinsley qui ont su me conseiller lors de mes expériences grâce à leurs compétences personnelles, mais aussi pour les blagues et discussion sur tout et rien. Merci à Oussema Souai, Théo Mozzanino, Hadrien Delattre, Jeffrey Cornault et Arnaud Briet d'avoir partagé des moments de rigolade avec moi. Merci à Olivier Son et Elisabeth Moncault de m'avoir aidé à réaliser les expériences en fin de thèse. Merci à Jihane Amarir-Bouhram pour avoir ouvert le champ de recherche grâce à sa thèse, précédant la mienne. Enfin, merci à mon encadrante, Marie-Agnès Petit, avec qui il n'a pas été toujours évident de se confronter mais qui, au final, m'a aidé et aiguillé au bon moment pour arriver aux résultats que vous allez lire.

Merci à nos collaborateurs du CEA de Gif-sur-Yvette, l'équipe « Molecular assemblies and signaling », dirigée par Raphaël Guerois et Françoise Ochsenbein. Ils m'ont accueilli et appris à purifier des protéines. Merci à Nicolas Richet de m'avoir aidé à mettre au point ces purifications. Merci à Arthur Belse qui a fini de les mettre au point et qui purifiait les protéines pour moi. Merci au reste de l'équipe et au reste du bâtiment pour la bonne entente et la bonne ambiance.

Je souhaite aussi remercier toute la troupe de théâtre du centre de Jouy-en-Josas : Abdel Boukadiri, Aline Viel, Annie Mathieu, Emmanelle Bourneuf, Deborah Jardet, Elisabetta Giuffra, Evelyne Lhoste, Olivier Telle, Setha Ketavong, Pierre-Yves Holtzmann, Elodie Poumerol, Guillaume Piton, François

Deniau et surtout notre metteur en scène Jean-Marc Mollines. Les séances du jeudi soir me permettaient vraiment de sortir de la thèse, et les pièces montées et jouées étaient vraiment divertissantes et enrichissantes.

Merci à tous mes amis, qui m'ont littéralement supporté à plusieurs moments. Merci à Marc Descrimes, Cecile Pereira, Drago Haas, Kevin Verstraete pour les plus présents et merci à tous les autres, présents ponctuellement. Un grand merci surtout à Aurélien Lancien qui à su, coûte que coûte, me faire sourire à chaque instant de ma dernière année de thèse.

Enfin, merci à mes parents, Johny et Véronique Hutinet, pour m'avoir mis au monde, aidé, tant sentimentalement que financièrement, et conseillé tout au long de ma vie. Je sais que je pourrai toujours compter sur eux si besoin. Merci au reste de ma famille qui croit en moi et me soutient en permanence pour que je puisse réaliser mes rêves et mes projets personnels.

On a tous beaucoup d'attentes, dans la thèse comme dans la vie, mais finalement la thèse c'est comme la vie, on rame beaucoup et, à la fin, on te dit que tu as été génial.

Table of contents

Table of figure contents

Chapter I:

Chapter II:

Chapter III:

Abbreviation list

µM	microMolar
ADP	Adenosine DiPhosphate
ATP	Adenosine TriPhosphate
bp	base pair
Chi	Crossover hotspot instigator
Ct	carboxy-terminus
DNA	DeoxyriboNucleic Acid
DSB	Double Strand Break(s)
dsDNA	double strand DNA
E. coli	*Escherishia coli*
FRET	Fluorescence Resonance Energy Transfert
GFP	Green Fluorescent Protein
HJ	Holliday Junction
HR	Homologous Recombination
HSV	Herpes Simplex Virus
ICTV	International Comittee on Taxonomy of Viruses
kb	kilo-bp
kDa	kiloDalton
Mg	Magnesium
MMEJ	Microhomologous Mediated End Joining
MMR	MisMatch Repair

Chapter I: Introduction

The study of bacteriophage genome evolution is a challenge, because these organisms are probably the most diversed on earth. No convinced species definition has been proposed up to now for them, classification remains an open field of research. Even the prediction of gene functions by comparative means is hindered by this diversity. Most phage genes of completely sequenced genomes remain of unknown function. This diversity is probably the consequence of the high mutation and recombination rates observed among Bacteriophages. My thesis has been centered on the study of homologous recombination in phages.

This introduction will first give an overview of phage genome structures, compared to bacterial genomes. In particular the concept of phage genome mosaïcism will be explained. The hypothesis guiding my doctoral research has been that the great facility of genetic exchanges among phage genomes may be due in part to the particular propriety of phage proteins that catalyse homologous recombination.

The key proteins for homologous recombination are the recombinases. A second part of this introduction will therefore summarize what was known on bacterial and phage recombinases at the onset of my work. The particularity of the bacterial machinery is the extreme fidelity of the Homologous recombination system. Indeed, this system permits to maintain the genome integrity when some stresses induce DNA damages. In phages, the homologous recombination machinery is not as faithful as that of Bacteria. Indeed, a 2008 paper has shown that exchange of genetic material catalyzed by λ phage recombinase occurs even between two 22% diverged sequences. However, whether this unfaithful property was restricted to the λ phage or common among phage recombinase was no known.

This will lead us to the specific questions that I have addressed during my doctorate: what is the fidelity of different phage-encoded homologous recombination machinery? Can homologous recombination explain part of the temperate phage mosaicism? Does the unexplored Sak4 family of phage recombinase behave similarly to the bacterial recombinase RecA, or to other phage-encoded recombinases, and what is its fidelity?

I) Bacterial genomes

The first observation of a bacterium is attributed to Anthony Leewenhoeck in a letter sent in 1683 to the Royal Society of London (Leewenhoeck 1684). He constructed the first microscopes and observed teeth scurf (dental biofilms) samples resuspended into rainwater. He observed what he designated as “little animals”. The name “bacterium” came long after; Christian Gottfried Ehrenberg named them because of their rod shapes (from the greek βακτιον, bacterion) in 1838.

After the discovery of the electron microscope, further investigations on bacteria showed they had no nucleus, contrary to Eukaryotes. They were then named Prokaryotes (Stanier 1962). With the appearance of the sequencing techniques, the taxonomy for Prokaryotes changed and started to be based on molecular sequences rather than phenotypic traits: In 1977, Woese et al. sequenced 16S-rDNA from Prokaryotes and 18S-rDNA from Eukaryotes (Woese, Fox 1977). This led him to distinguish Archaebacteria from Eubacteria, and to demonstrate that Archaebacteria are closer to the Eukaryotes than Eubacteria. He concluded that the tree of life is divided into three branches: the Eukaryotes, the Bacteria and the Archaea ((Woese, Kandler, Wheelis 1990), figure1)

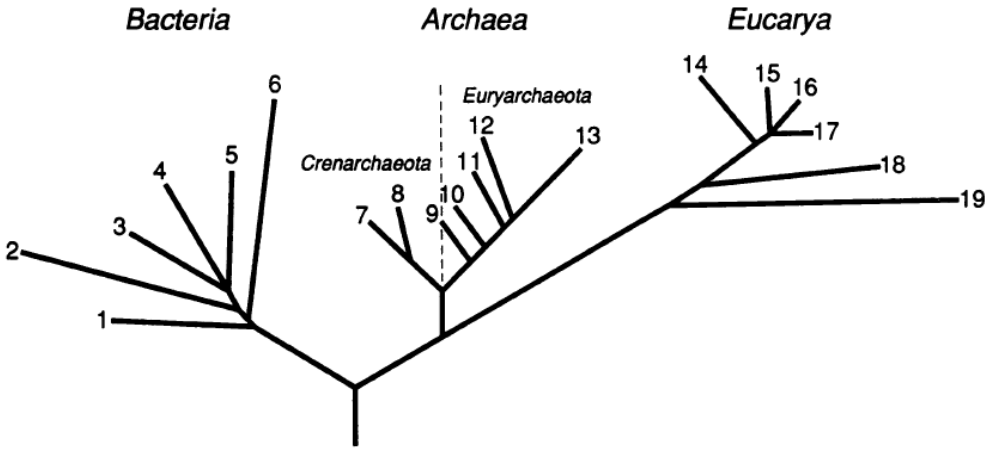


Figure 1 : Universal phylogenetic tree in rooted form, showing the three domains. Branching order and branch lengths are based upon rRNA sequence comparisons. The position of the root was determined by comparing sequences of pairs of paralogous genes that diverged from each other before the three primary lineages emerged from their common ancestral condition. The numbers on the branch tips correspond to the following groups of organisms. Bacteria: 1, the Thermotogales; 2, the flavobacteria and relatives; 3, the cyanobacteria; 4, the purple bacteria; 5, the Gram-positive bacteria; and 6, the green nonsulfur bacteria. Archae: the kingdom Crenarchaeota: 7, the genus Pyrodictium; and 8, the genus Thermoproteus; and the kingdom Euryarchaeota: 9, the Thermococcales; 10, the Methanococcales; 11, the Methanobacteriales; 12, the Methanomicrobiales; and 13, the extreme halophiles. Eucarya: 14, the animals; 15, the ciliates; 16, the green plants; 17, the fungi; 18, the flagellates; and 19, the microsporidia. (Woese, Kandler, Wheelis 1990)

Bacterial genomes have begun to be completely sequenced and compared at the eve of the 21st century. However, the task of classifying them, as has been done for the Eukaryotes, turned out to be more complicated than expected. Classification has been successively based on their phenotypes, their capacity to cross hybridize their chromosomal DNA, their 16S-rDNA sequences, their gene content proximity, or their average nucleotide identity (ANI). However, bacterial species boundaries are still unclear and discussed (Konstantinidis, Ramette, Tiedje 2006). The reason for this perplexity is the incredible plasticity of bacterial genomes due to horizontal gene transfers. Indeed at the species level, we can distinguish genes of the core-genome, comprising genes shared by all (or almost all) strains within a species, and those from the pan-genome, comprising the constant and the variable gene pool, made of genes found in only some of the strains of a given species. There seems to be no correlation between the gene function and its belonging or not to the core genome (Huang, Hsiang, Trevors 2013). It is nevertheless recognized that prophage genes constitute a large fraction of a species pan-genome (15% in *Salmonella enterica*, 30% in *Escherichia coli*, (Bobay, Rocha, Touchon 2013)). Whole genomes can be aligned at the species level thanks to their core-genome, and partitioned into a conserved backbone, interrupted by a collection of variable segments (Figure 2, (Chiapello et al. 2005)). Variable segments have various lengths: the large ones correspond to prophages and other genomic islands, the small ones to intra-genic microdiversity (Touzain et al. 2010).Extraction of the species backbone is useful for the detection of small DNA motifs used by bacteria for the repair of the chromosome (Chi sites), or its segregation (KOPS sites, MatS sites, (Touzain et al. 2011)). It seems therefore that the last 15 years of intense sequencing efforts have permitted to reach a good overview of bacterial genomes, where for most species a common core-genome is observable, whose genes can be used for deriving robust phylogenies (Touchon et al. 2009). Strain-specific variable segments that correspond mainly to the gene flux inside a species interrupt this backbone. These segments are exchanged among strains by horizontal transfer, or loosed by deletion.

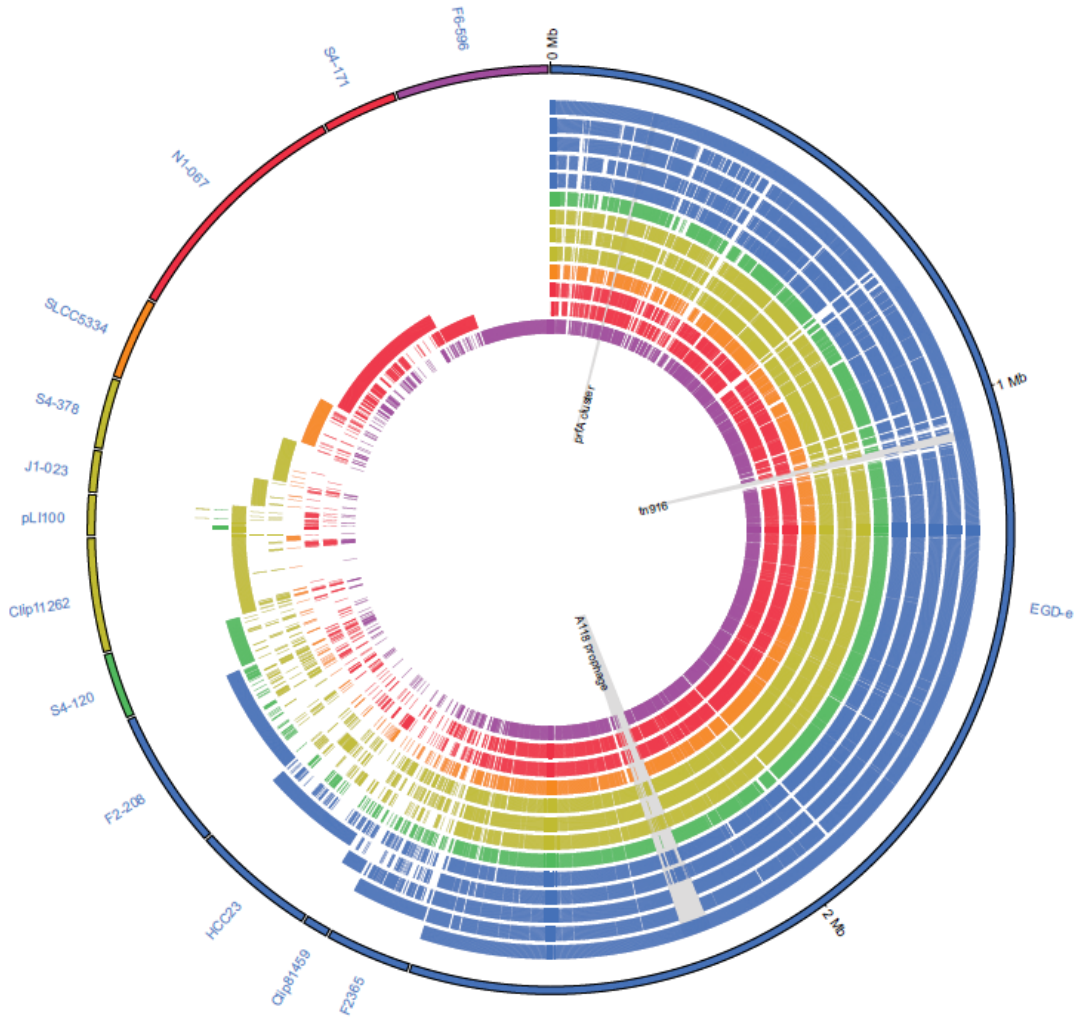


Figure 2: Comparative genome content of 13 *Listeria* chromosomes and *L. innocua* plasmid pLI100. The outermost circle indicates the source of each gene in the pan-genome with each gene represented by a constant width wedge. Starting at the top of the figure (0 Mb) and moving clockwise, all EGD-e genes are arranged in chromosomal order. Continuing clockwise, all genes not present in EGD-e are grouped by strain (as indicated by segment labels). Genes in the F2365 segment are present in F2365, but absent from EGD-e, and genes in the Clip81459 segment are present in Clip81459, but absent from F2365 and EGD-e, and so on. In this way, each gene is represented only once in the diagram. Gene order in all segments except EGD-e is monotonically increasing, but discontinuous, since shared genes may be represented in other segments. Internal circles indicate gene presence (solid color) or absence (unfilled) of each gene in each of the 13 strains examined. Circles from outer to inner are in the same order as strains on the outer circle, starting with EGD-e, followed by F2365, etc. *L. monocytogenes* strains are in blue; *L. marthii* is in green; *L. innocua* strains are in gold; *L. welshimeri* is in orange; *L. seeligeri* strains are in red; *L. ivanovii* subsp. *londoniensis* is in purple. The location, in the EGD-e genome, of the prfA virulence cluster, conjugative transposon tn916 and prophage A118 are specifically indicated. (den Bakker et al. 2010)

II) Bacteriophage genomes

Bacteriophages are viruses infecting Bacteria. They are generally specific to a species, but some infect bacteria belonging to the same genus, or others are restricted to some strains into the same species.

1) History and definition

The first observation of bacteriophages was reported by Frederick Twort in 1915 (Twort 1915). He described a "transparent material" that grew on a lawn of a *Micrococcus* strain. He called it ultra-microscopic (*i.e.* beyond detection under the microscope) viruses. He observed that these viruses could not grow alone, and could not grow on other bacteria, animals, neither human.

Felix d'Herelle isolated an anti-Shiga microbe (named Shigella today) in 1917, noted that it lyses liquid Shiga culture, and made on Petri dishes "a certain number of circles of around 1 mm in diameter on which the culture is non-existent". From this, he concluded that the anti-Shiga is not a chemical compound but a life form. He could not isolate this invisible microbe alone in a culture medium, thus he concluded this anti-Shiga was an "obligate bacteriophage". "Bacteriophage" (phage in short) was then coined as the name used to define viruses that grow on bacteria (D'Herelle 1917).

There are two categories of phages: virulent and temperates (Figure 3). Both infect Bacteria and inject their genome from the virion into the bacterium. Both kinds can undergo a lytic cycle, whereby genomes are rapidly replicated and encapsidated. The bacterium is lysed to release the virions formed, or phages are secreted from the bacterium in the case of filamentous phages such as M13 (Beaudoin, Pratt 1974). In addition, temperate phages can insert their genome into the bacterial chromosome instead of replicating themselves. Almost all genes are repressed during this cycle. The phage is then named a prophage or lysogenous phage. The dormant prophage is replicated passively during bacterial multiplication, and can excise itself, if induced by an endogeneous or exogenous stress, to start again a lytic phase.

Phages are very difficult to classify. The first manner to classify them is based on the type of genetic material present in the virion. It can be DNA or RNA, single strand or double strand. However, most bacteriophages known today are dsDNA, and their virion has a tail (Caudoviridae). A sub-classification is then made among Caudoviridae, based on the tail type: short for Podoviridae, long and non-contractile for Siphoviridae, and long and contractile for Myoviridae (ICTV classification, last update in 2013). However, this sub-classification is not monophyletic for some groups, such as Mu-like phages, that are closely related but can be either Sipho- or Myoviridae. Genomics analyses are

now on the front stage for virus classification. Rohwer and Edwards were the first in 2002 to try to construct a proteomic tree from 105 completely sequenced phage genomes (Rohwer, Edwards 2002). They used a distance based on the inverse of the number of genes shared by two phages. However, their branch history did not fit with the capsid classification. Lima-Mendez et al. had a similar approach on 306 completely sequenced phage genomes, except that they eliminated the hierarchical tree-like approach (Lima-Mendez et al. 2008). They looked for pairwise similarities and drew a network to cluster phages. They obtained separate graphs for ssDNA and RNA phages, and most Caudoviridae were connected into a single network, with temperate phages occupying the center of the network, and virulent phages (such as T4, T7) forming more distant 'segregating' nodes. However, the difficulty with network representations is their fluidity: the network is expected to change when more genomes are added, so it may not be usefull for classification matters. Even if no universal gene such as the 16S rDNA gene is present among phages, single gene approaches have been tried on sub-categories of phages, such as Caudoviridae, that share a related major capsid protein, but also portal and terminase genes that are well conserved (Smith et al. 2013). Very recently, a classification based on the retrieval of a collection of 3 to 9 proteins involved in the connection between the virion head and tail has been proposed, that permits to classify >90% of the completely sequenced Caudoviridae (Lopes et al, submited). This classification identifies several

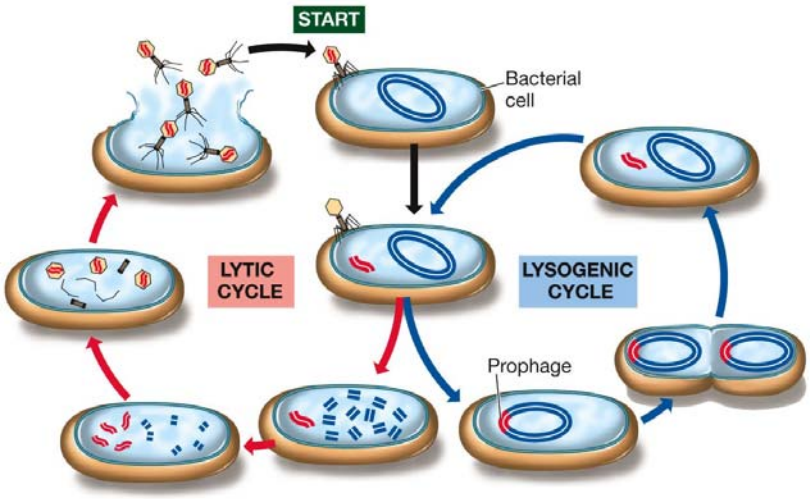


Figure 3: Phage life cycles. When a phage binds its receptor on a bacterium, it injects its genome. Virulent phages enter in the lytic cycle: the host genome is sometimes degraded (as represented here), and the phage multiplies itself. Capsid proteins are expressed and assembled. The phage genetic material is then encapsidated and the lysis of the cell releases virions. Temperate phages have another option after genome injection. They can enter into the lysogenic cycle instead of the lytic cycle in certain conditions. The phage genome is integrated into the bacterial genome and replicated at the same time as the bacterial chromosome. The phage is then said to be in a prophage state. After a stress, the prophage is induced and the phage enters in the lytic cycle.

groups that are closely related phages, despite having Myo or Siphoviridae virions, as observed for Mu-like phages. It may be that comparative genomics will finally help reaching a satisfying method of classification for phages.

2) Genome organization

A) Virulent phages

Virulent DNA phages possess either very large genomes, in the 80-200 kb range, or very small ones (below 20 kb). Those of the T4 family belong to the large genome category; they have been largely sequenced, and compared to one another. They have a core genome, and the shared genes can be used to build evolutionary trees (Sarker et al. 2012). The rest of the genome of T4-like phages is hyper-plastic and encodes for genes probably implied in the adaptation to the host (Figure 4, regions that are not grey shaded) (Comeau et al. 2007). In fact, many of these genes have unknown functions. Somehow, the structure of T4-like genomes is reminiscent of that of bacteria, with a common backbone and phage-specific variable segments.

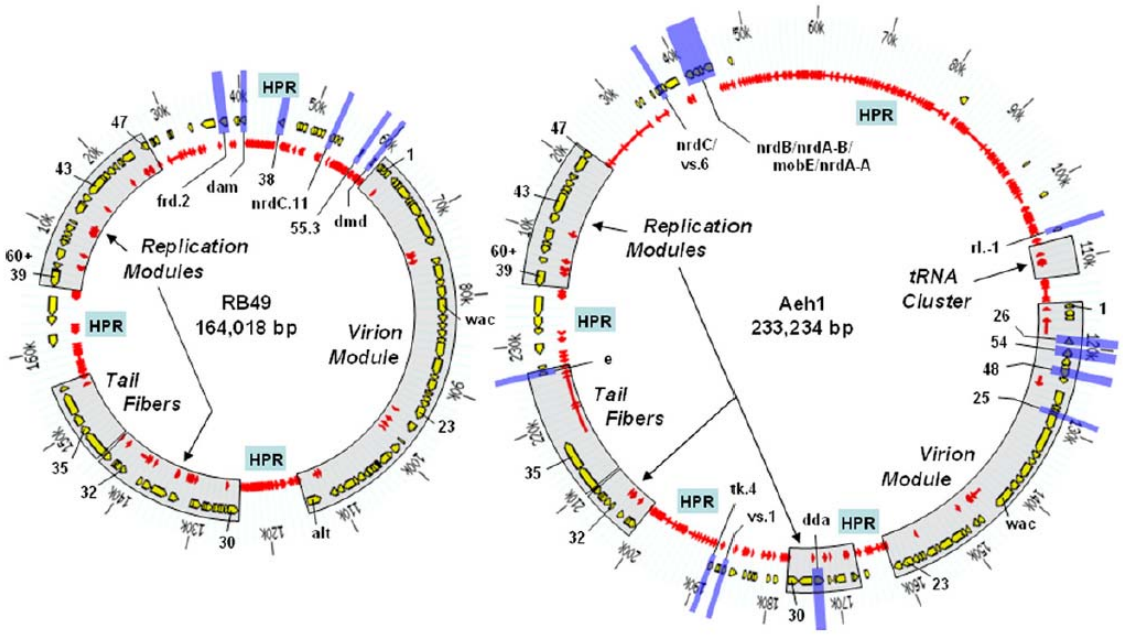


Figure 4: Genome organization of two virulent phages: RB49 and Aeh1 (T4 family). Genome maps of the phages are drawn to scale and, in each, the outer rings show the scales (in kb), the middle rings (yellow arrows) the T4 homologues, and the inner rings (red arrows) the ORFs with no homology to T4. T4 genes that have been rearranged (radiating blue bars) are indicated in their new locations. Various T4 genes are given as landmarks on the outside of the rings and the hyperplastic regions are indicated. The virion modules, tail fiber genes, replication modules and tRNA cluster are indicated by gray boxes. (Comeau et al. 2007)

A recent study showed that T4-like *Cyanobacteria* phage genomes are prone to high levels of horizontal transfer, probably mediated by homologous recombination (Ignacio-Espinoza, Sullivan 2012). Indeed, the T4 group of phages possesses genes encoding for a homologous recombination system. This system is similar but independent from that of the host (see below for further explanations).

B) Temperate phages

Temperate DNA phages fall into two genome categories: small sized genomes (4-10 kb range), usually encapsidated as ssDNA (Roux et al. 2012), and medium sized genomes (20-80 kb range), usually forming dsDNA virions. This last category has a genomic structure completely different from that of T4-like phage genomes. It is impossible to delineate a shared backbone even among related genomes: rather, each genome seems to be composed by an assemblage of parts coming from recent horizontal transfers. The phage genomes are said to be “mosaic” ((Brussow, Canchaya, Hardt 2004) for review, Figure 5)

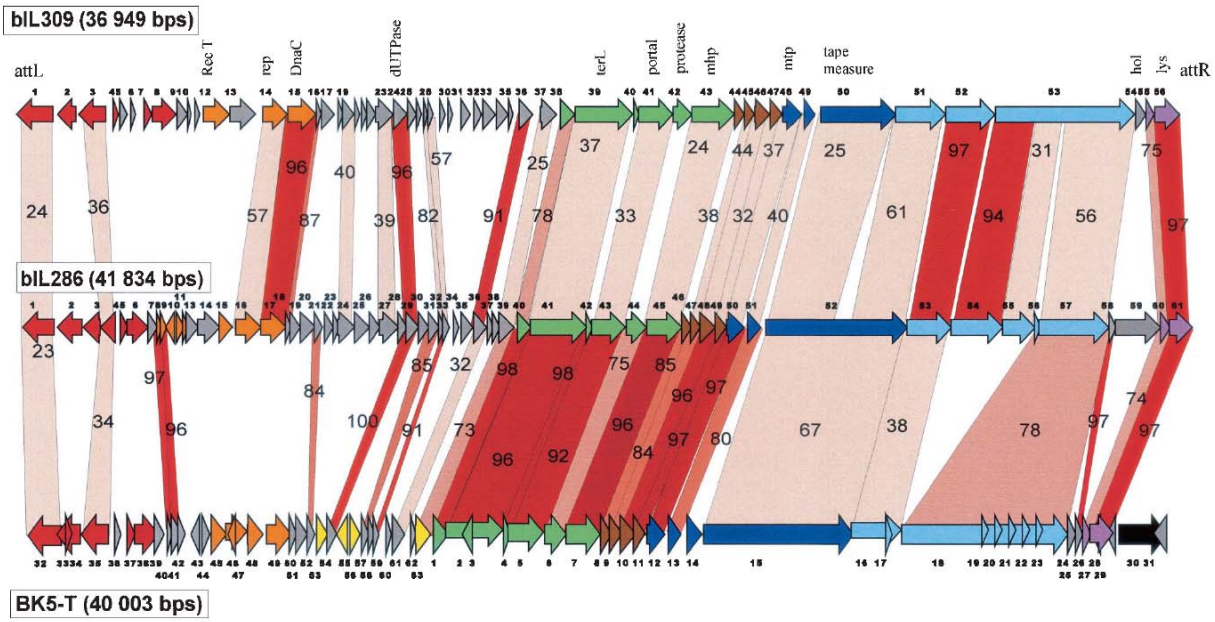


Figure 5: Alignment of the genetic maps from the *L. lactis* prophage bIL309 with the *L. lactis* prophage bIL286 and temperate *L. lactis* phage BK5-T. The open reading frames are color coded according to their predicted function (Green, predicted DNA packaging and head morphogenesis genes; brown, predicted head-to-tail joining genes; dark blue, tail morphogenesis genes; light blue, tail fiber genes; mauve, lysis genes; black, predicted lysogeny conversion genes; red, lysogeny genes; orange, DNA replication genes; yellow, transcription regulation genes). Selected genes or genomic features are denoted. Genes encoding proteins that showed significant amino acid sequence are linked by red shading, and the percentage of amino acid identity is indicated. The degree of amino acid identity (>90, >80, >70, or <70%) is reflected in the color intensity of the red shading. Inspired from (Proux et al. 2002).

The Hatfull team, helped with students, has been isolating numerous mycobacteriophages and sequencing their genomes. In 2011, 62 genomes were sequenced and compared. Each phage can be classified into one of 11 super groups depending on its genome structure and sequence homologies. Interestingly, although groups can clearly be delineated, genetic exchanges are observed between groups ((Pope et al. 2011). Recently, an upgrade with 471 phages led to the addition of 8 new super groups (Hatfull 2014)).

These authors think that genetic exchanges among phages are mainly due to illegitimate recombination: two phages in the same bacterium break and ligate their genomes to form a new viable phage. Other argued on the possibility that phage recombinases (see below) may play a role in these exchanges, because of their unique propriety to recombine diverged sequences, as described below (Martinsohn, Radman, Petit 2008).

Mosaicism among temperate dsDNA phages of *Staphylococcus* is remarkably extant (Kwan et al. 2005). A recent study tried to recreate the evolutionary history of 31 *Staphylococcus aureus* phages with the hypothesis of homologous recombination (Swenson et al. 2013). To do so, they truncated these phage genomes into segments. Each sequence of a segment is classified as an allele, called here variant, based on its identity (at least 94% of identity to belong to a group of variants, and at most 76% identity with any other variant). The basic postulate is that between two identical variants of the same segment, homologous recombination can occur and form a new mosaic phage from the two ancestral phages. They were able to observe these events and to recreate three ancestral phage genomes from the isolated phages, which exchanged genetic material with other phages.

3) Ways to exchanges genomes fragments

Different molecular processes will be described here, that can explain the mosaic formation among temperate phages. They will be presented starting from the mechanisms requiring little or no homology between the exchanged fragments (the broad “illegitimate recombination” category), up to those involving more extant homology stretches. Interestingly, in the phage world, the frontier between these two classical categories seems to be somewhat blurred.

A) Polymerase slippage

DNA polymerase can switch from a matrix to another if some homology (superior to 14nt) is present between the two ends of the recombination event. A model proposes that the polymerase stops due to the presence of a block, such as a hairpin structure. The replicated strand unwinds from its matrix,

and if complementarity to the tip of the replicated strand is present downstream of the pausing site, annealing occurs. DNA polymerase binds again to the replicated strand extremity and resumes polymerisation ((Viguera, Canceill, Ehrlich 2001), Figure 6A). Poteete has proposed that phage recombinases play a role in conjunction with polymerase slippage: by catalyzing the annealing of a phage 3'-ssDNA extremity with the lagging strand of the replication fork. The leading strand polymerase slips from the initial matrix to the phage strand, thanks to its close vicinity ((Poteete 2008), Figure 6B).

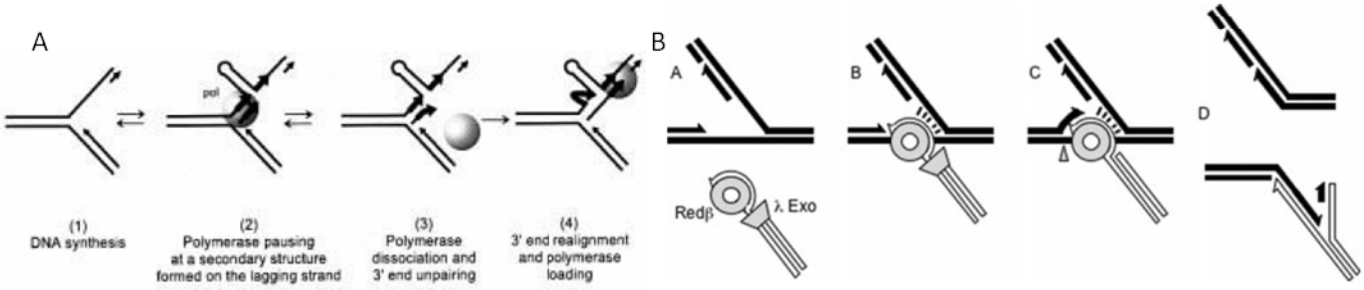


Figure 6: Polymerase slippage mecanism A. (Viguera, Canceill, Ehrlich 2001) Schematic representation of the slippage process at replication fork. During DNA synthesis of a repeated sequence (step 1), the polymerase reaches a barrier on the lagging strand (a hairpin structure here) and pauses (step 2). Polymerase dissociation occurs (step 3). If the polymerase is not able to disrupt the barrier (with its strand displacement activity), then the tip of the newly synthetized strand can unpair from its template and anneal to the second repeat beyond the barrier, allowing polymerase re-loading and resumption of the synthesis (step 4). B. (Poteete 2008) Red-mediated replisome invasion/template switch. The drawing on top shows a replication fork moving through λ DNA, left to right. The drawing on bottom shows a dsDNA end processed by λ exonuclease, with the overhanging 3' end wrapped around a Redβ ring (Passy et al. 1999) (step A). When the replisome exposes complementary bases on the strand serving as the template for lagging strand synthesis, Redβ anneals its captive strand to them (step B). Redβ positions the invading strand to serve as the new template for leading strand synthesis. The resulting joint molecule is resolved by an endonucleolytic cleavage (triangle). λ exonuclease falls off or is removed (step C). The detached lagging strand arm and unreplicated remainder of the right side of the parent chromosome is shown on top of the figure (step D). The incoming DNA now serving as the template for both leading and lagging strand synthesis, with the transferred replication fork continuing to move in the same genetic direction on its new template, is shown on bottom of the figure (step D).

B) NHEJ (Non-Homologous End Joining)

NHEJ (Figure 7) is a mechanism of double strand break DNA repair well known in Eukaryota. A protein, called Ku, binds to the two extremities of a double-strand break. A machinery is then recruited to get the two broken parts closer. Finally, a ligase ligates the two DNA (see (Ochi, Wu, Blundell 2014) for review). No homology between the two broken ends is needed to perform NHEJ, and this process is often associated with mutagenesis or large rearrangements. Ku orthologs were found in Bacteria (Bowater, Doherty 2006). These genes are most widely distributed in Proteobacteria (α, β, γ, and δ families), Firmicutes, and Actinobacteria. There is no evidence whether

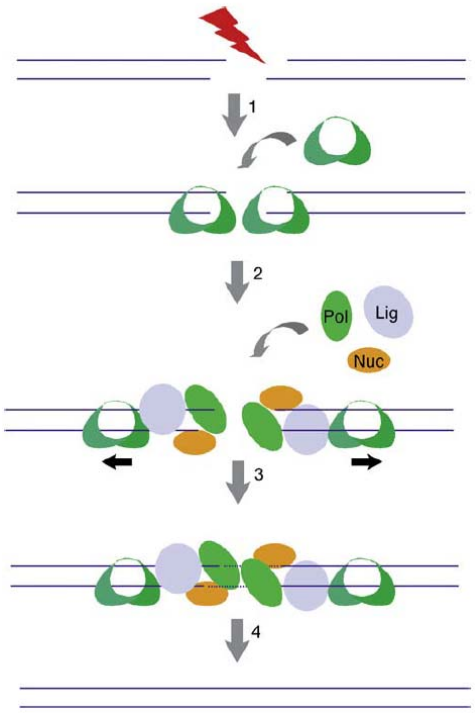


Figure 7: DSB Repair by the NHEJ Complex The Ku complex locates to the break site (step 1), where it may serve as an end-bridging and alignment factor. Following binding to the broken DNA ends, additional processing enzymes are recruited by Ku to the break site (step 2). Ku may translocate away from the ends, allowing access by other factors to the break termini. When DNA ends are non- complementary and/or are damaged, the DNA end-processing, gap- filling, and nucleolytic activities generate DNA termini, capable of being ligated, prior to ligation (step 3). Subsequently, the broken ends are joined by an NHEJ-specific DNA ligase (step 4) and the NHEJ complex dissociates. (Bowater, Doherty 2006)

these genes are acquired by horizontal transfer or if they were present in the common ancestor and disappeared from some bacterial branches. This bacterial Ku forms homodimers that bind to dsDNA extremities. The complete NHEJ activity was not proved in bacteria and deletion of the Ku-like protein is not deleterious for UV resistance (a classical test for repair of DNA break systems). In *Bacillus subtilis*, in addition to Ku, some essential homologs to Eukaryotic NHEJ machinery were identified: a ligase and a polymerase (de Vega 2013).

Ku orthologs were also found in phages (d'Adda di Fagagna et al. 2003). They share more identity with bacterial Ku than with Eukaryotic Ku. In phages, this protein is Gam and was first identified in the Mu phage. For the sake of clarity, as a different Gam protein is present in λ, the Mu type of Gam will be designated as Mu-Gam hereafter. This protein interacts with DNA as Eukaryotic Ku. There is no evidence of NHEJ function in phages carrying Mu-Gam, but it is still a possibility than could explain the mosaisism.

C) MMEJ (Microhomology-Mediated End Joining)

MHEJ (Figure 8) is a another way to repair double strand breaks involving an annealing step between the two ends, between small repeats of less than 10 pb. *E. coli* can promote this type of events at a low frequency of 10^{-5} (Chayot et al. 2010). The authors transformed *E. coli* with a digested plasmid, mimicking a double strand break. They showed that the RecBCD exonuclease is essential for the reaction. RecBCD resects DNA at the DSB and exposes microhomologies that anneal each other, independently of the host recombinase RecA. Then the annealing product is ligated. An alternative

MHEJ pathway has been proposed, which involves RecQ, a DNA helicase. This pathway can only be seen in an *E. coli* strain with a *recB* mutation, that leads to the loss of the RecBCD exonuclease activity but maintains the helicase activity (Ivankovic, Dermic 2012). RecQ would participate to the annealing of microhomologies. Due to it slow efficiency, MHEJ is a minor pathway in *E. coli*, that may take place when RecBCD exonuclease activity is inactivated and no sister chromatid, or no other homologous sequence, is available to perform a RecA-dependent repair.

Figure 8: Proposed model for MMEJ. It relies on the degradation of DNA ends by the RecBCD nuclease/helicase complex (blue incomplete circle). Regions of microhomology (one to nine nucleotides, red strips) are available to promote the synapse between DNA ends. End-ligation likely relies on the essential LigA. Adapted from (Chayot et al. 2010).

In an assay based on the selection of lambda prophage imprecise excision events (a process occurring in the absence of the prophage integrase gene), RecOR was also found to facilitate MHEJ, by annealing microhomologies (Shiraishi et al. 2002). In the same study, the authors also reported that the RecET pathway stimulates MHEJ by a factor of seven. These proteins are encoded by *rac*, a defective prophage in *E. coli.* The *recET* genes are normally silent, but become activated by a *sbcA* mutation, that brings a promoter upstream of *recET*. RecT is a recombinase, that performs single strand annealing (SSA) of complementary DNA sequences (see below), and RecE is a 5'-3' dsDNA exonuclease. They also saw that RecJ, a 3' to 5' ssDNA exonuclease, is essential in this reaction. Later, the same team found that the RecET pathway was inhibited by an *uvrD* mutation. They proposed that the UvrD helicase prevents the RecA loading and so enhances the illegitimate recombination by the RecET pathway (Shiraishi et al. 2006). These studies suggest that phage recombinases such as RecT may promote illegitimate recombination by MHEJ.

D) HR (Homologous recombination)

Homologous recombination should not be confounded with site-specific recombination. In this last process, a specific DNA site is recognized and recombined by a resolvase or an integrase. Most temperate phages encode an integrase, which permits integration of phage in the host genome at an *attB* sequence (see the review in Annex). Homologous recombination is not sequence-dependent, but relies on the homology between the DNA partners that recombine.

In 2010, phages genomes were analyzed to identify gene encoding protein catalyzing the homologous recombination events, named recombinases. Among 365 dsDNA phage genomes of a size superior to 20Kbp, 60% contained a recombinase (Lopes et al. 2010). By detecting distant homology between amino acid sequences, all recombinases were classified into three super families, Rad52, Rad51 and Gp2.5-like proteins (these super-families are described in details later in this chapter). The crystallographic structures for members of Rad51-like and Gp2.5-like recombinases have been solved (Hollis et al. 2001; Gajewski et al. 2011). None of any phage Rad52-like recombinase has been crystallized to sufficient resolution up to now to permit 3D reconstitutions. The Redβ 3D model below is taken from Lopes et al. (Lopes et al. 2010). It was model on the human Rad52 Nt domain, which has been cristalized (Kagawa et al. 2002)(Figure 9). Several phage recombinases, belonging to the Rad52 and Gp2.5 families, are known to promote the single strand annealing (SSA) reaction ((Kmiec, Holloman 1981; Rezende et al. 2003; Datta et al. 2008; Ploquin et al. 2008) see paragraph III of this chapter for further explaination).

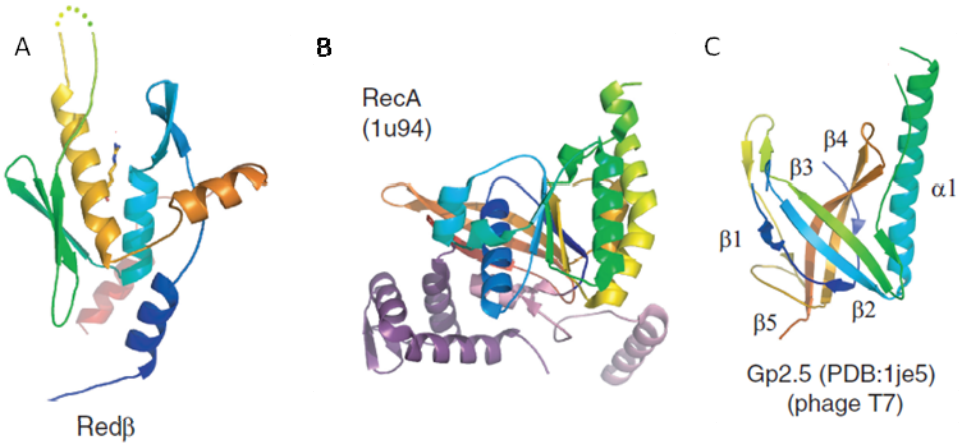


Figure 9: Structure of the three phage recombinases families. Model of Redβ (A) from phage λ as a representative of the Rad52-like family. Model of RecA (B) of *E. coli* represent the Rad51-like family. Structure of Gp2.5 (C) from the phage T7 represent the Gp2.5-like family. Adapted from (Lopes et al. 2010).

Some phages infecting Enterobacteria also encode for proteins that inhibit the main host exonuclease, RecBCD. This is a way to protect the phage rolling-circle intermediates of replication from degradation (Figure 10). The first one to be reported was Gam, from phage λ, that infects *E. coli*. The Gam protein was crystallized and looks like a globule with two mobile α-helixes that have the same diameter as ssDNA. The hypothesis is that at least one of the two helixes takes the place of ssDNA in RecBCD complex, inhibiting its DNA degradation activity (Figure 10, (Court et al. 2007)). Phage Phi80 encodes GamL, which performs the same function as Gam, while sharing no homology with the Gam protein (Rotman et al. 2012). Phage P22 encodes a third type of RecBCD modulator,

Abc2, (see part III of this Manuscript) that only partially inhibits the ds exonuclease, and is proposed to convert it into a ssDNA 5'-3' exonuclease (Murphy 2000). Finally, the temperate phage Mu encodes a fourth form of RecBCD 'protector', Mu-Gam, (d'Adda di Fagagna et al. 2003), which binds to the extremities of the phage genome, and thereby prevents their degradation by RecBCD. This protein has already been mentioned in paragraph 2B, as it shares homology with Ku proteins. Whether phages infecting Gram positive bacteria also encode Gam like functions is unknown at present. Given the panel of diverse proteins performing similar functions, the possibility exists that this activity is also present among temperate phages infecting Gram+ bacteria, but remains unveiled until now.

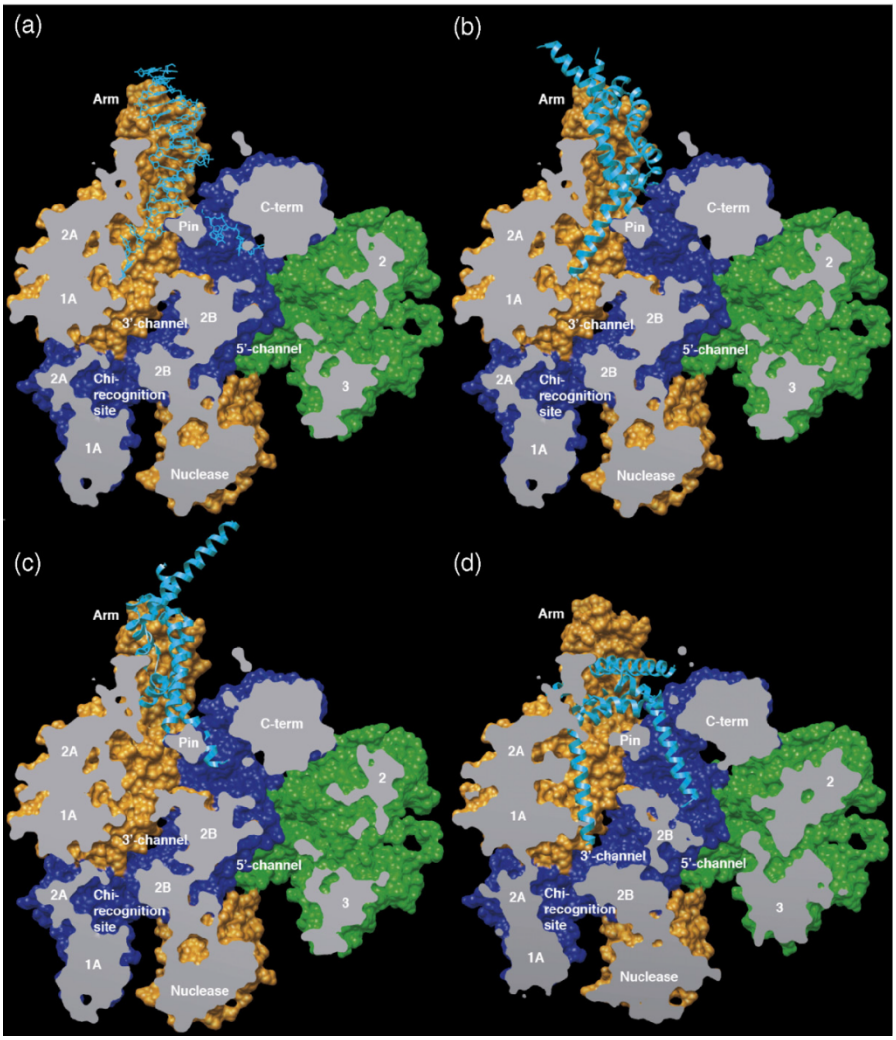


Figure 10: Surface cutaways showing proposed models for RecBCD inhibition by the Gam protein. (A) Co-crystal structure of RecBCD and DNA. Models of Gam inhibition show (b) the 3' channel occupied by helix H1 of Gam, (C) the 5'-channel occupied by helix H1 Gam and (B) the re-orientated H1 helices occupying both the 3' and 5' channels of RecBCD. RecB is shown in orange, RecC in blue, RecD in green.With DNA and Gam both shown in cyan. Key domains and features of the RecBCD complex are labelled. For clarity, (D) shows a different cutaway region but the same orientation. (Court et al. 2007)

We can ask why phages encode their own recombinase, since almost all bacterial hosts already encode one. This question has been addressed in 2013 for 275 prophages extracted from genomes of Enterobacteria (Bobay, Rocha, Touchon 2013). They compared the occurrence of (1) phage recombinase gene, (2) a phage gene coding for Gam- or Abc2-like inhibitors of RecBCD, and (3) the presence on the phage genome of Chi sites (GCTGGTGG, a site recognized by RecBCD, which converts the potent double strand exonuclease into an ssDNA exonuclease and RecA loader (Smith 2012). Interestingly, and contrary to what was generally suposed, they found that a majority of the prophages had Chi sites. They also reported evidence for a clear counter selection for Chi-sites on prophages encoding a Gam or Abc2-like inhibitor of RecBCD. They proposed that phages had to choose between the host recombination system (i.e. accumulate Chi sites) and their own system (i.e. encode both a RecBCD inhibitor and a recombinase). Nevertheless, their analysis did not permit to conclude that. First, some of the prophages used in this analysis are defective, and no exact assessment of the proportion of active phages among their set was performed. Furthermore, some RecBCD inhibitors might have been missed, as they looked for only two of them: Gam and Abc2. Finally, their analysis uncovered a significant category of prophages (20% of the set) encoding both a recombinase and Chi sites, suggesting that phages are not in an "either-or" choice between the two recombination systems. Indeed, we have shown (see Chapter II) that phage λ recombines with prophages using independently either the host RecA and its own Redβ system (redundancy), with equal efficiency when the homology region is long enough (2 kb). Interestingly, in λ, no Chi sites are present. This suggests that the RecA-dependent recombination occurs independently of RecBCD, probably by the RecFOR pathway. An alternative explanation to the observations of Bobay et al. is that defective phages, in their decay process, acquired Chi sites and lost some part of their genome, potentially containing a recombinase and a RecBCD inhibitor.

Their analysis shows notwithstanding that phages carrying a recombinase have genomes bearing traces of more recent exchanges than recombinase-less phages, suggesting that phage recombinases facilitate DNA exchanges. Our own analysis of phage mosaics has also uncovered traces of homologous recombination at the vicinity of 40% of all recent exchanges (see Chapter II).

Other clues on the particularities of phage-encoded recombinases can be suggested by interaction maps. The interaction map of the λ proteins has been explored (Rajagopala, Casjens, Uetz 2011) and interactions between its recombinase Redβ and P, N, Int and A were found. P is the replicative helicase loader, and such an interaction may help to load a replication fork outside of the origin, in a process linked to recombination. However, a compared analysis of replication yield/products of λ in the presence and absence of Redβ indicated no loss of products in the mutant, and only marginal changes in the proportion of circular and sigma-shaped structures (Better, Freifelder 1983). N is an

anti-termination protein, which permits transcription past early terminator region. How Redβ may facilitate N action is not immediately evident, and the interaction may be an artifact. Int is the site-specific integrase, it is curious Redβ interacts with it. It suggests that site-specific recombination and homologous recombination may somehow be linked. A is the DNA large sub-unit of the Terminase of λ. It fills capsids until recognition of a cos site on λ genome, which it cuts. However, the packaging cannot continue if the cos site is not complete, for instance if the DNA molecule is a monomer of the genome. It is possible that Redβ, through its interaction with the terminase, is brought to the capsid in such cases, and reanneals two cos sites to continue the packaging.

III) Homologous recombination

My research during this PhD has focused on homologous recombination, especially that mediated by phage recombinases. I will therefore summarize here what is known on the bacterial host homologous recombination system before those of phages.

1) History and definition

The comprehension of homologous recombination began long before the understanding of DNA structure and gene concept. Mendel proposed a model in 1865 for the segregation of characters from a generation to another, where each character segregates independently (Mendel 1866). Later, Morgan understood that the characters encoded by genes are carried by chromosomes and so can be genetically linked (Morgan 1911). Muller showed a synaptic attraction between homologous chromosomes and saw chromosome rearrangements after X-ray induced breaks (Muller, Altenburg 1930).

Since the discovery of DNA structure (Watson, Crick 1953), some molecular models for recombination were published (all models are presented in Figure 11)**.** The following summary is inspired by the historical review of Haber (Haber 2007). The first model to be published sought to explain recombination between a broken and an intact λ phage molecule (Figure 11A, (Meselson, Weigle 1961)). They thought that each strand of the broken DNA anneals to an intact complementary DNA, which serves as a template for DNA polymerization. Holliday conceived another model in 1964 (Figure 11B). He proposed that recombination begins by a coordinated pair of single-strand nicks on compatible strand pairs of homologous chromosomes. The nicked strands unwind and move to the other strand to form a single strand exchange. This leads to a four-stranded structure called now Holliday junction. This junction is symmetrical and can be then resolved by two cleavages that yield

different products: (1) one cleavage gives an exchange between the two chromosomes, the so-called crossing-over. (2) The symmetrical cleavage leads to a small exchange of the single strands annealed previously, called non-cross over, or patch repair (Holliday 1964).

In 1966, Thomas proposed another mechanism, whereby 4 nicks are formed at different places, DNA strands are unwound so that ssDNA is produced. If this ssDNA finds a complementary ssDNA, an heteroduplex DNA can be formed (Figure 11C, (Thomas 1966)). This mechanism is now known as single strand annealing (SSA). Later on, it was proposed that only one of the two DNA molecules receive one nick (Figure 11D, (Meselson, Radding 1975)). The 3' end could serve as a primer for DNA polymerization and displace the complementary strand on the intact matrix. In this intermediate, called a D-loop, the displaced strand would invade the homologous sequence of the other chromatid and the DNA would isomerize into a Holliday junction.

Based on the activity of Redα from the λ phage, which is a 5' to 3' exonuclease, Resnick proposed that the 3' ended tail produced was displaced to a homolog intact matrix, and served as primer for DNA polymerization, until the repair was long enough to be brought back to the initial break. The repair ended by a last DNA polymerization step. This is the first model of what is now called synthesis-dependent strand annealing (SDSA, Figure 11Eα, (Resnick 1976)). However, this mechanism did not explain crossovers. So Resnick imagined further that after the 3' end invasion of an intact matrix, the endonuclease could nick the displaced DNA and promote crossover (Figure 11Eβ, (Resnick 1976)). Following a different lane of thought, Szostak et al. proposed a model for crossing-overs based on the formation of double Holliday junction (Figure 11F, (Szostak et al. 1983)): both 3' ends obtained after resection of a DSB invade a homologous sister chromatid and pair on it. After DNA synthesis and ligation, a double Holliday junction structure is obtained and can be resolved.

After that, the models have been inspired by previous model and did not evolve much. However, proteins capable of promoting these events were studied. Clark and Margulies showed in 1965 that one gene is essential for recombination and repair, based on UV irradiation of *E. coli* K12 strain. This gene is *recA* (Clark, Margulies 1965)

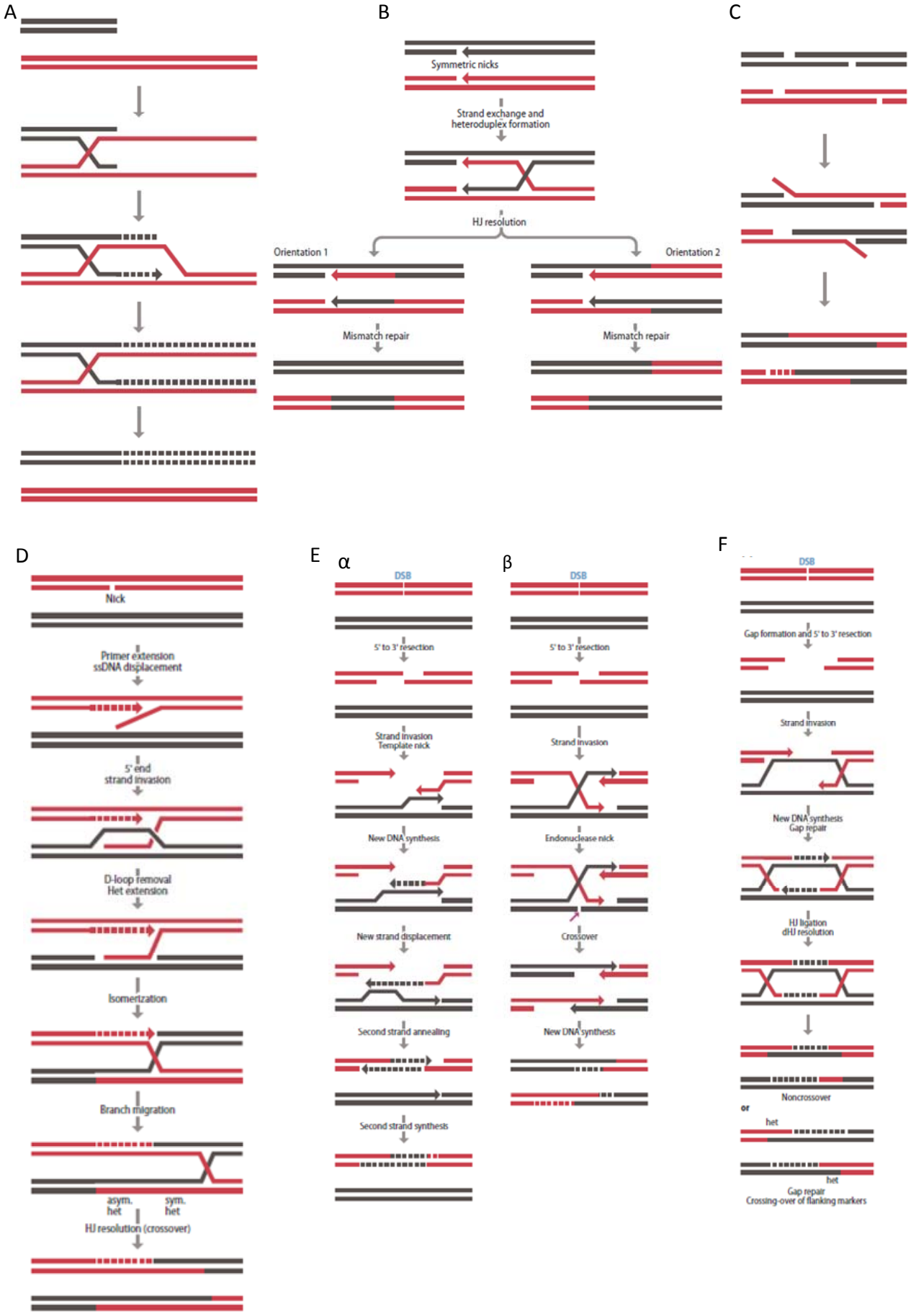

A
B
Symmetric nicks
Strand exchange and heteroduplex formation
HJ resolution
Orientation 1
Orientation 2
Mismatch repair
Mismatch repair
C
D
Nick
Primer extension ssDNA displacement
5' end strand invasion
D-loop removal Het extension
Isomerization
Branch migration
asym. het
sym. het
HJ resolution (crossover)
E
α
DSB
5' to 3' resection
Strand invasion Template nick
New DNA synthesis
New strand displacement
Second strand annealing
Second strand synthesis
β
DSB
5' to 3' resection
Strand invasion
Endonuclease nick
Crossover
New DNA synthesis
F
DSB
Gap formation and 5' to 3' resection
Strand invasion
New DNA synthesis Gap repair
HJ ligation dHJ resolution
Noncrossover
or
het
het
Gap repair Crossing-over of flanking markers

Figure 11: Recombination Models. A Meselson and Weigle's break–copy recombination mechanism. The two strands of a broken chromosome fragment can form base pairs with an intact template and promote copying to the end of the template, thus producing a recombined, full-length product (Meselson, Weigle 1961). B. Holliday's 1964 model. A pair of nonsister chromatids after meiotic DNA replication are shown; the two other chromatids, uninvolved in recombination, are not shown. A pair of same-strand nicks leads to a reciprocal exchange and formation of symmetric heteroduplex connected by a four-stranded symmetric structure now known as a Holliday junction (HJ). The HJ can be cleaved by cutting either of two pairs of strands (orientations 1 and 2). Crossovers occur when the HJ is cleaved so that only the crossing-strands connect the two homologous chromosomes. In the example shown, mismatch corrections lead to a 6:2 gene conversion (Holliday 1964). C. Charles Thomas's SSA model to obtain reciprocal recombination by annealing overlapping single strands of DNA from two chromosomes with offset nicks on both strands (Thomas 1966). D. Meselson and Radding's 1975 model. A single nicked strand is displaced by new DNA synthesis primed from the 3' end of the nick. The displaced strand can form a region of heteroduplex DNA by strand invasion and the formation of a displacement loop (D-loop). Cleavage of the D-loop leaves a single region of heteroduplex DNA adjacent to an HJ that is always distal from the initiating nick. Isomerization of the HJ and subsequent branch migration leads to the formation of a symmetric (sym. het) region of heteroduplex adjacent to an asymmetric (asym. het) segment (with only one heteroduplex region), still with the crossover point far from the initiating lesion (Meselson, Radding 1975). E. The DSB repair model of Resnick (1976). (α) A DSB is resected and one of the 3' ends invades a template. The template strand is nicked. The paired DSB end initiates new DNA synthesis, which is displaced, allowing the original sequences to reanneal with the template. When the newly-copied strand is long enough, it can anneal with the second end, priming a second round of new synthesis and the repair of the DSB. This leads to a noncrossover repair of the DSB. (β) Crossovers can be generated by DSB repair. Here, the nicked template strand itself can anneal with the single-stranded sequences created by 5' to 3' resection. Strand invasion of the original DSB end creates a Holliday junction. An endonuclease nick at the base of the HJ results in connections between the molecules that result in a crossover. DNA synthesis fills in the ssDNA gaps and leads to a crossover (Resnick 1976). F. The double Holliday junction (dHJ) model of Szostak, Orr-Weaver, Rothstein and Stahl (1983). A DSB was enlarged into a double- stranded gapped region, which was subsequently resected to have 3'-ended single-stranded tails that could engage in strand invasion. The first strand invasion would produce a D-loop to which the end of the second resected end could anneal. The initial structure has one complete and one half HJ, but branch migration of the half HJ allows the formation of a second complete HJ. New DNA synthesis completes the formation of a fully ligated structure that can be resolved into crossovers if the two HJs are cleaved in different orientations. Resolution of both HJs as crossovers should leave an apparent double crossover in the middle; such outcomes are rare (Szostak et al. 1983). Adapted from (Haber 2007).

2) RecA

RecA is the ubiquitous recombinase found in the bacterial reign. It participates to homologous recombination *in vivo* and repairs DNA.

A) RecA activity

RecA is a 38 kDa protein that catalyzes ATP hydrolysis, homologous pairing and recombination. Renaturation of complementary strand was the first activity discovered for RecA. In 1985, Bryant and

Lehman explored this activity. They tried to renaturate the NaOH-denatured 5386 bp long ϕX174 genome (circular duplex linearised by PstI digestion) in the presence of RecA and they demonstrated that when the RecA concentration is low, the protein needs ATP and Mg^{2+} to promote the reaction. They separated the products of renaturation on gel and observed them under an electronic microscope. A high molecular weight complex was separated on gels, and a network of DNA was seen under the microscope (Bryant, Lehman 1985). This suggests either that RecA promotes multiple single strand pairing reactions within the same ssDNA molecule, either it promotes another form of recombination: the strand invasion reaction. Indeed RecA is able to search homology by binding to ssDNA and searching of a complementary a dsDNA.

This activity has been observed more precisely by single molecule assays (van der Heijden et al. 2008). A dsDNA is bound chemically to a surface at one end, and to a magnetic bead at the other end. A magnetic field pulls up the magnetic bead with a chosen force to elongate the DNA. The space position of the bead is measurable and represents the length of the DNA molecule. Homologous ssDNA covered of RecA is added thanks to a flow. Here the ssDNA can invade the dsDNA and pair to its complementary sequence. To verify the invasion the product is digested with a restriction enzyme. Without the strand invasion reaction, the digestion cuts the DNA and the bead is lost. But after the strand invasion, the digestion cut the dsDNA resulting of the annealing but the bead is retained by the displaced ssDNA strand. Furthermore, the author saw the structure after invasion is torsionally constrained by trying to turn the bead in both directions. This constraint is lost after the digestion. For the authors, this result indicates the displaced strand is wrapped around the heteroduplex dsDNA formed by the invasion.

Single molecule experiments have also permitted to ask questions about the search of homology by RecA. A first research looked for the sliding of RecA on DNA (Ragunathan, Liu, Ha 2012). To do so, a ssDNA labeled with an acceptor for fluorescence resonance energy transfer (FRET) is bound to a glass surface. RecA next covers the ssDNA. The target dsDNA is labeled with an acceptor for FRET. With this, the authors measured the distance between the acceptor and the donor as a function of time. They found that RecA slides on dsDNA with a diffusion coefficient of 7700bp^2/s until it finds homology with a minimum of 6 complementary nucleotides. This rapid 1D search is to be confronted with another discovery where the authors observed that the ssDNA-RecA complex can be transfered from a dsDNA segment to another (Forget, Kowalczykowski 2012). A dsDNA is bound between two beads, and the distance between them can be controlled, thereby changing the containment of the DNA. They showed that the less the DNA is constrained and the longer the homology, the quicker is the reaction. They also showed that ssDNA-RecA complexes can create a bridge inside the dsDNA to search homology at two different positions at the same time with two different part of the ssDNA.

The question now is how RecA finds this homology. For this, the resolution of the structure of RecA has been very suggestive. Story et al. resolved its structure in 1992. This protein possesses a core N-terminus (Nt) domain and a smaller C-terminus (Ct) domain. The core-domain has two activities: it binds DNA and binds and hydrolyzes ATP. The Ct domain is a secondary DNA binding domain. RecA polymerizes on DNA in a helix formation, with 6 monomers of RecA per turn. The core DNA binding site is at the center of the structure, but the Ct one is turned to the outside (Story, Weber, Steitz 1992). With this structure, De Vlaminch et al. (2012) proposed a mechanism for homology recognition by a shift of DNA between the two binding sites. With a single molecule set-up of bead and dsDNA as described previously, they tried to make the DNA move with a DNA-RecA complex (Figure 12). They found that the secondary DNA binding site of RecA can bind to ssDNA but not to dsDNA, unlike the primary DNA binding site, which can bind both. However the secondary DNA binding site can bind dsDNA in negative supercoiled structures, suggesting it can bind to an unwinding DNA. This suggests the homology search begins in a breathing DNA. One of the two strands of the breathing DNA is caught by the secondary DNA binding site of a RecA-ssDNA complex, and then the homology is found on the primary DNA binding site (De Vlaminck et al. 2012).

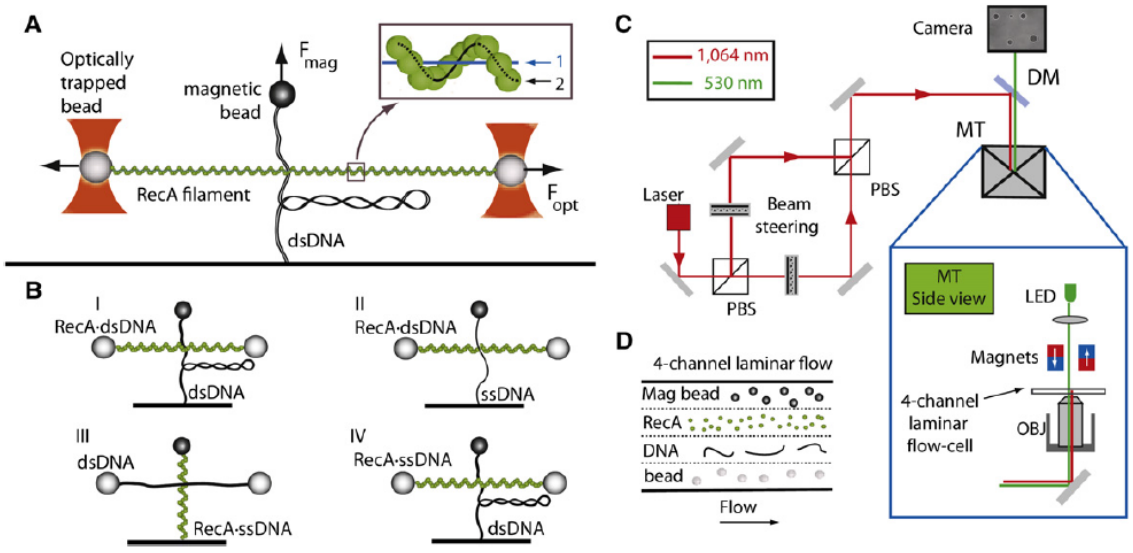


Figure 12: Experimental Approach of De Vlaminch et al. (A) Side view of dual-molecule assay where a RecA filament that is held in the optical tweezers interacts with a coiled dsDNA that is tethered in a magnetic- tweezers configuration. Inset: schematic of the RecA filament represented with two DNA-binding sites, a primary site (1) and secondary DNA-binding site (2). (B) The four different experimental configurations I–IV used in their work. (C) Schematic outline of the setup indicating the main components. Two independently steerable optical traps are generated using a 1064nmlaser system. The beams are split and recombined using a polarizing beam splitter and focused using a high numerical aperture objective in the volume of a four-channel laminar-flow cell. Positioning and rotation of an external magnet pair allows stretching and coiling of a molecule tethered in the magnetic tweezers configuration. (D) Top view of the four-channel laminar-flow system used for the assembly of the molecular constructs. (De Vlaminck et al. 2012)

The ATPase site of RecA permits the dissociation of the RecA-DNA complex. Once the homology is found, the RecA filament continues to polymerize to continue the annealing (Rossi et al. 2011). By this mechanism, there is the creation a Holliday junction. This junction is next resolved by dissolution of the annealing or by cleavage into the junction.

In vivo the question is a little different. Indeed RecA is used to repair breaks into the chromosome of bacteria. Nevertheless, how can RecA find homology between two DNA segregated at the two poles of the cell in preparation for the division? DNA damage can occur long times after the DNA polymerization, so we need to imagine how the two DNA are caught and brought together. This question has been addressed in the laboratory of Pr. Sherratt (Lesterlin et al. 2014). In a system creating double strand breaks that are detectable under the optic microscope, they used a GFP tagged RecA to observe how the RecA-DNA complex acts at the breaking point. They observed that RecA is able to displace the break cross the cell to find homology on the sister DNA. After the repair, the two sister DNAs are separated across the cell in prevision of the cellular division.

B) Regulation of the RecA filament

Regulation for the RecA filament occurs at two distinct steps. The first one is the regulation before filament formation, which is inhibited by SSB and promoted by RecBCD or RecFOR. The second is the regulation after filament formation, stabilized by RecX and disassembled by DinI, UvrD and maybe by RdgC (Figure 13).

It is usually admitted that SSB inhibits the RecA filament formation by binding the ssDNA before RecA. Indeed *in vitro* when both proteins are mixed together, SSB catches the ssDNA before RecA (Reddy et al. 2000). Nevertheless, if RecA is already nucleated on the ssDNA, SSB facilitates the polymerization of RecA by unwinding the secondary structures of ssDNA. A recent single molecule study (Bell et al. 2012) showed that RecA is able to nucleate on an ssDNA covered by SSB but with a high lag time (3 minutes) and a low nucleation rate. However, the RecA filament polymerization is not disturbed.

RecA has manners to overcome the SSB barrier. The first one relies on RecBCD activity. This complex is composed of three different proteins: RecB, that has helicase and nuclease activities, RecC, which recognizes a specific site on DNA called Chi, and RecD, which is an helicase. This heterotrimer detects DSB on DNA and degrades both strands until RecC recognizes a Chi site. RecBCD pauses on Chi sites

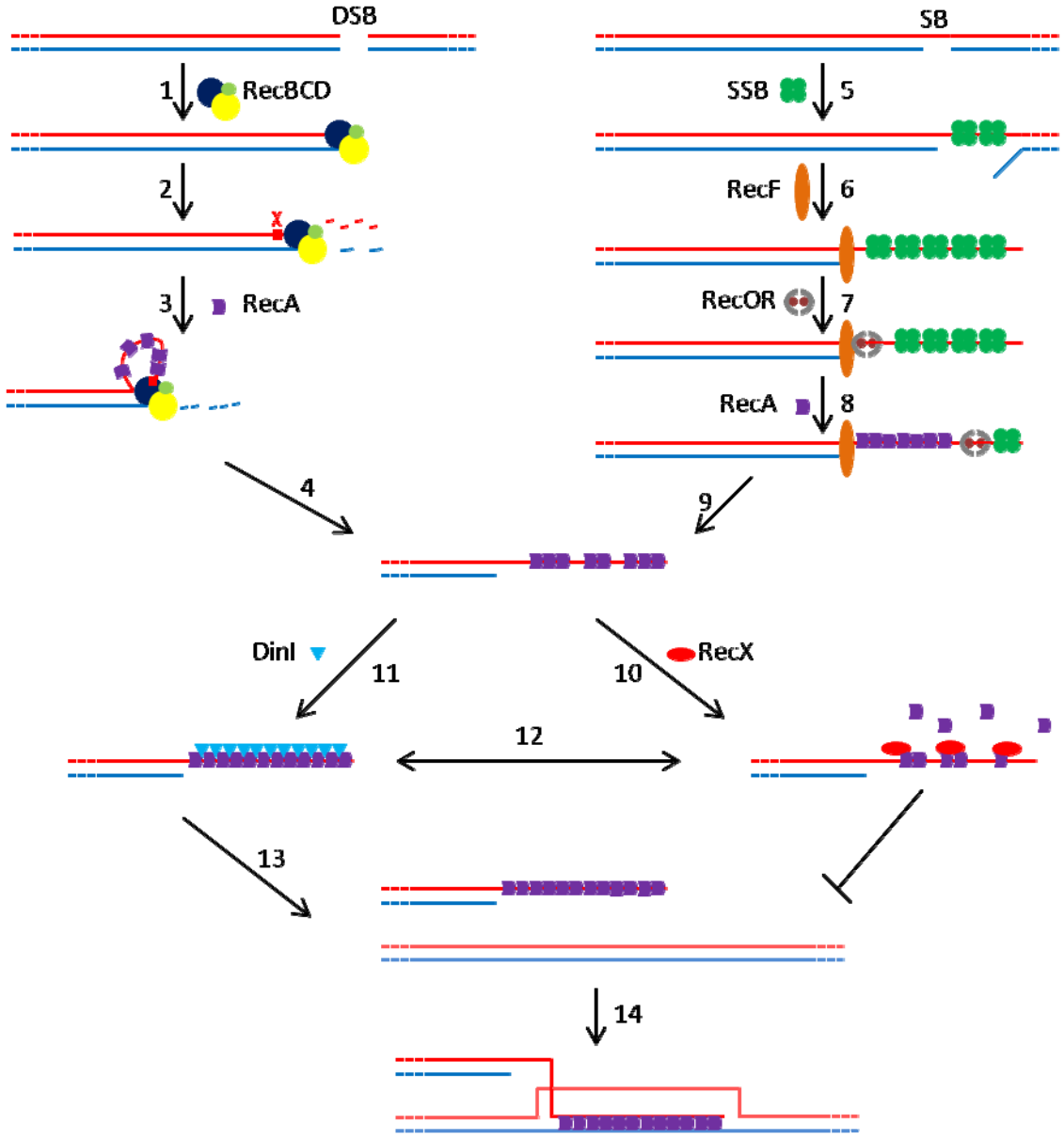


Figure 13: Regulation of RecA activity. There are two RecA loading pathways depending on the DNA damage. In the case of double strand break (DSB), RecBCD recognize blunt DNA (1, RecB in blue, RecC in Yellow, RecD in green) and degrade both strand but not at the same speed (2). When RecB recognize a χ site, the 3' to 5' exonuclease activity stop but not the 5' to 3' and RecC load RecA on the 3' overhang created (3). The ssDNA-RecA filament complex is ready to recombine (4). In the other case of single strand break (SB), the DNA is digested and SSB load into the ssDNA produced (5). RecF recognise a ssDNA break (6) and load RecOR on the ssDNA (7, RecO in brown and RecR in grey). RecOR moves SSB away from the ssDNA permiting the loading of RecA (8). The ssDNA-RecA filament complex is ready to recombine (9). RecX can dismantle the RecA filament (10) and DinI stabilise it (11). Each one are interchangeable (12). If the RecA filament remain stable (13), it searches for homology into an intact matrix by strand invasion (14).

and its activity changes to degrade only one of both strands. RecB loads RecA on the 3' overhang DNA created before SSB come (Spies, Kowalczykowski 2006).

The second one is the RecFOR pathway. RecF recognizes a single strand break on dsDNA. RecO can bind to ssDNA, RecF, RecR and SSB. A study showed it also anneals two complementary ssDNA, in the presence and in the absence of SSB (Kantake et al. 2002). RecR binds a dimer RecO. The idea here is that the RecOR complex either binds to SSB and moves it away to permit the nucleation of RecA, or binds to RecF which finds a ssDNA break on dsDNA and polymerizes RecA from the break (Morimatsu, Kowalczykowski 2003; Handa et al. 2009). Recent single molecule analyses showed that RecOR and RecFOR complexes decrease the lag of RecA nucleation and increase the polymeration rate (Bell et al. 2012). Last years, the structure of RecOR was resolved (Radzimanowski et al. 2013) and revealed a ring structure composed by two RecO and four RecR. This ring rounds the ssDNA that permits to push SSB and nucleate RecA.

RecX and DinI have antagonist activities. RecX is expressed in operon with RecA and has an inhibition activity. RecX binds to RecA and inhibits its polymerization, leaving it to its depolymerization, whereas DinI binds to RecA, and stabilizes the RecA filament preventing the depolymerization and promoting the polymerization (Lusetti et al. 2004; Renzette, Gumlaw, Sandler 2007).

UvrD is a helicase which dismantles the RecA filament (Veaute et al. 2005). UvrD prevents the polymerization of RecA at the replication fork *in vivo*. This may prevent the reannealing of the fork by RecA (Lestini, Michel 2007).

RdgC is the most ambiguous inhibitor protein. Some activities were shown *in vitro*, but no phenotypes were observed *in vivo*. Indeed, when RecA is competed by RdgC *in vitro* in the classic strand exchange reaction, RecA loses its activity proportionally to RdgC addition (Drees et al. 2006). The structure was resolved in 2007 (Briggs et al. 2007). It dimerizes to form a ring that rounds the DNA with a preference for the ssDNA. The idea is that RdgC slides on DNA and moves away RecA from DNA to inhibit its action.

C) Fidelity of this system

It is admitted that homologous recombination is a very faithful mechanism. However, it is currently discussed. Malkov et al. studied this property using a strand invasion assay, whereby a supercoiled plasmid is incubated with a labeled short oligonucleotide having more or less homology with the plasmid. If RecA pairs the oligonucleotide to the plasmid, a D-loop is formed and easily separated on gel. Using a 27-mer oligonucleotides, with 6 nt of perfect identity on both extremities with the

plasmid, and 15 nt of random sequence in the center, they showed that RecA can recombine until diverged sequences 4 mismatches on 15bp (~25% of divergence) randomly distributed but in their model, mismatches were surrounded by short perfect homologies (6bp, (Malkov, Sastry, Camerini-Otero 1997)). Sagi et al. have proposed a molecular model for RecA fidelity: RecA may have a low sensibility to mismatches in the absence of ATP hydrolysis. This *in vitro* situation is nevertheless quite unlikely *in vivo*, unless the bacterium is running out of ATP. It succeeds to anneal strands more rapidly if the first mismatch is distant from the invading 3' end. The sensitivity to mismatches and the directionality in the strand exchange process support a mechanism for homology recognition that can be modeled as a kinetic proofreading cascade: RecA pairs the DNA monomer by monomer until the loss of similarity, thus RecA is stopped by consecutive mismatches (Sagi, Tlusty, Stavans 2006).

Beyond the intrinsic sensitivity of RecA to mismatches, a second mechanism for proofreading relies on the mismatch repair of the cell. MutS is the sensor of this system by detecting mismatches, with different affinity, depending on the mismatch (Brown, Brown, Fox 2001). MutL binds to MutS and polymerises on DNA. MutS recognizes a mismatch during the recombination exchange and together with MutL blocks the recombination intermediate by preventing rotation of DNA strands during the strand exchange. UvrD comes next to disrupt the RecA filament. Consequently, the invading strand is removed (Tham et al. 2013).

D) Links with Eukaryotes and Archaea

Proteins sharing homology with the Bacterial RecA are found in the two other branches of life. In Eukaryotes, Rad51 has the same function as RecA, although the homology spans only the core domain, that is an ATPase and a DNA binding domain. As already explained, RecA has a Ct domain that binds ssDNA. In Rad51, this is a Nt domain that has this function, sharing no homology with the Ct of RecA. Eukaryotes have several other Rad51 paralogs, which will not be discussed here. In Archaea, RadA is the ortholog of Rad51, having the same Nt domain. Furthermore, Archaea have a paralog of RadA called RadB. Lin et al. tried to reconstitute the evolution history of these genes (Figure 14 (Lin et al. 2006)). They compared sequences of Rad51 and its paralogs from different Eukaryota, RadA and RadB from Archaea, and RecA from Bacteria and plants. They think that the RecA ancestor was conserved in the reign of Bacteria. In the common ancestor of Eukaryotes and Archaea, there was a gene, or genome, duplication, leading to two copies of the *recA* gene, that they designated as *radα* and *radβ*. In Archaea, Radα became RadA and Radβ became RadB. The ancestor of Eukaryotes acquired a bacterial copy of RecA by endosymbiosis of an *α-Proteobacteria*, the mitochondria. Then Radα duplicated into Rad51 and DMC1, its paralog in meiosis, and Radβ into

different paralogs named differently in Mammals and Fungi. Plants conserved the unduplicated, bacterial RecA copy, contrary to Mammals and Fungi, and acquired a second one from the *Cyanobacteria* endosymbiont, the chloroplast.

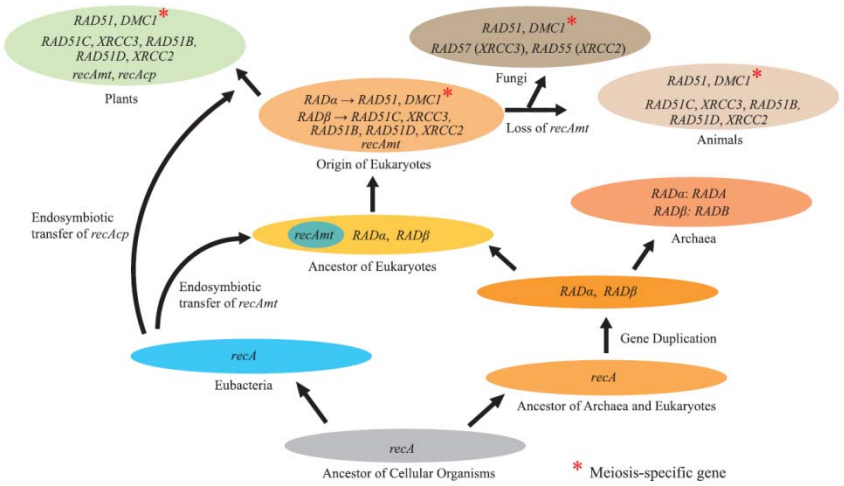


Figure 14: A model of the evolutionary history of recA/RAD51 gene family. The gene duplication that occurred before the divergence of archaea and eukaryotes gave rise to two lineages, RADα and RADβ, and, in eubacteria, recA has remained as a single-copy gene. In eukaryotes, both RADα and RADβ genes experienced several duplication events, but, in archaea, they remained as single-copy genes. Eukaryotic recA genes originated from proteobacteria (recAmt) and cyanobacteria (recAcp) recA genes after two separate endosymbiotic events. recAmts were subsequently lost in the ancestors of animals and fungi (Lin et al. 2006).

RecA and Rad51 were compared for their *in vitro* activities. Both promote DNA pairing and strand exchange. They both need ATP and magnesium to do so (Rice et al. 2001). RadA and RecA were also compared. In the same manner RadA promotes ATP hydrolysis, DNA pairing and strand exchange (Seitz et al. 1998). Recently, Rad51 was spoted by chromatin immunoprecipitation *in vivo* on yeast genome with a system of inducible DSB. Rad51 binds first to the DSB as expected, and on the homologous region of the DSB (Renkawitz et al. 2013). In yeasts and mammals, paralogs are used to stabilize the Rad51 filament (see (Karpenshif, Bernstein 2012) for review). In Archaea, RadB also enhances the activity of RadA strand invasion (Graham, Rolfsmeier, Haseltine 2013).

3) Phage recombinases

As said earlier, three families of recombinases are encoded by phages (Lopes et al. 2010). Each family has its particularities and its model of study. Here we propose to present recent researches on these different models to illustrate the capacities of each family.

A) Rad52-like recombinases

The most studied Rad52-like recombinase is Redβ from λ. This protein is known to promote SSA but not strand invasion (SI, Figure 11). *In vivo*, SSA is the observed way of action for Redβ (Stahl et al. 1997). However, when Redβ has annealed a part of a ssDNA on its complementary ssDNA, it can promote branch migration along the dsDNA until the end of the complementarity region (Li et al. 1998). In vitro, Redβ can promote SI under conditions of low Mg^{2+} concentration (0 to 2 mM) (Rybalchenko et al. 2004). Another well studied Rad52-like recombinase, the RecT encoded by the *rac E. coli* prophage, has the same capacity at low magnesium concentration (Noirot, Kolodner 1998). In fact, under such conditions, the dsDNA is less stable, because the negative charges of phosphates are not compensated by the positive charges of the magnesium. The two strands of the dsDNA repel themselves and open the double strand. The complementary ssDNA covered by Redβ can anneal into this breathing dsDNA. *In vivo*, the supercoiled dsDNA breathes too. However, for strand invasion to proceed, the breathing needs to be at the exact position where the ssDNA caught by Redβ can anneal, and it is a rare event. After that, Redβ can promote the branch migration and complete the reaction for homologous recombination.

Redβ has not been crystallized yet, but electron microscopy observations have permitted to observe its structure. It forms ring structures. In the absence of DNA, the ring is composed of 12 subunits, whereas with ssDNA, it is made of 15 to 18 subunits. The structure changes in the presence of dsDNA: Redβ becomes a helical filament (Figure 15A, (Passy et al. 1999)). The hypothesis is that in the ring structure form, Redβ exposes ssDNA at the outside of the ring. After annealing, Redβ protects the dsDNA inside the helical filament. The biochemical behavior of Redβ was reminiscent of that of yeast Rad52, which forms also such rings under the microscope, but no homology was found, even by the Koonin group many years later (Iyer, Koonin, Aravind 2002). Erler et al. proposed an alignment in 2009 of Redβ with the eukaryotic Rad52 protein, and observed that the ring structure of Redβ without DNA had only 11 monomers and was not closed. The authors hypothesized this ring is a right helix. In the presence of ssDNA, they observed Redβ is not organized, contrary to previous observations. As previously, in the presence of dsDNA, Redβ formed a helix filament (Figure 15B, (Erler et al. 2009)). In 2010, our group also found distant homology between Rad52 and various phage recombinase: Erf, Redβ/RecT and Sak. Last year, Matsubara et al. realigned different Redβ, Rad52 and Sak recombinases (Matsubara et al. 2013). They experimentally distinguished amino acids involved in the quaternary structure (K148 and R149 in Redβ), in ssDNA binding (V77, G78, V79, W82, R152 and probably R161) and in SSA activity (K69, G78 and K132)

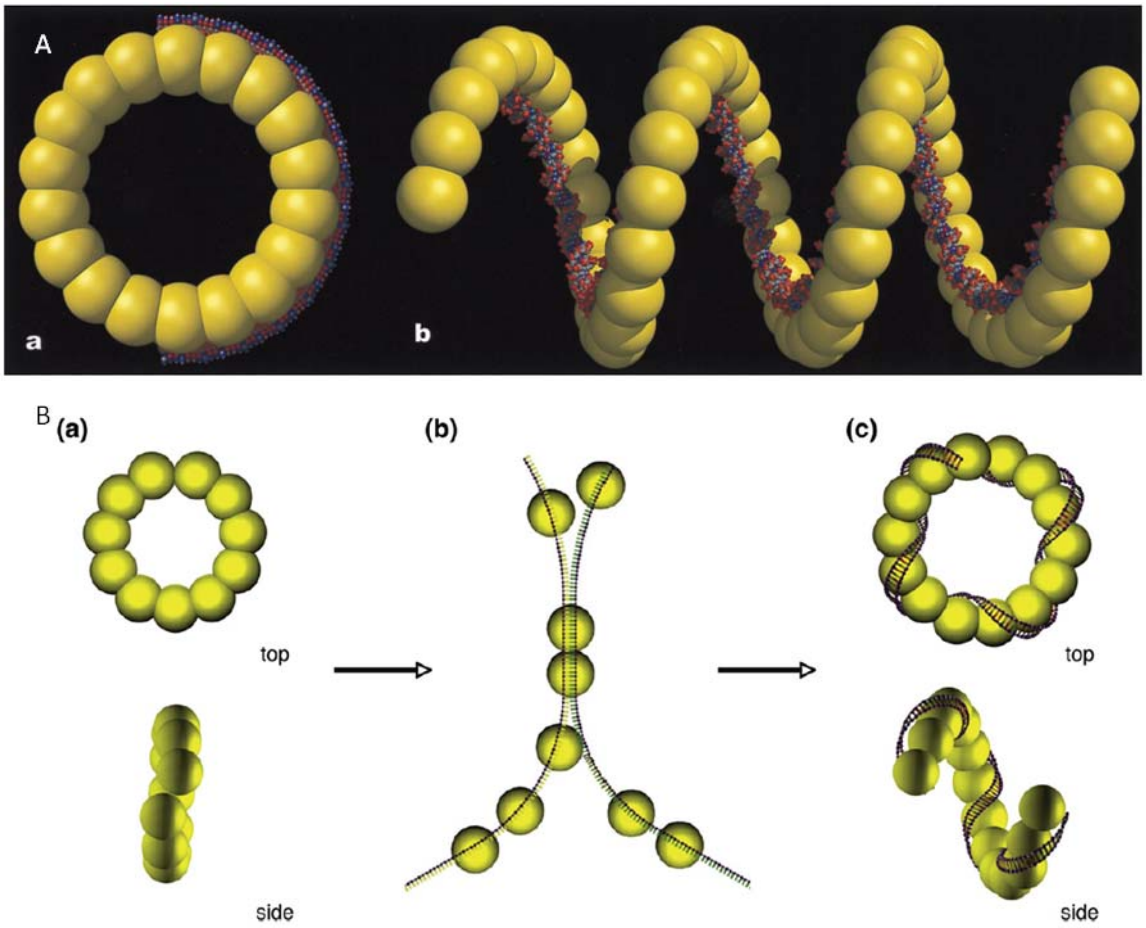


Figure 15: Model of Redβ quaternary structure. (A) Models for the complex of the large Redβ ring and ssDNA (a) and the Redβ helix formed on dsDNA (b). The ssDNA is believed to wrap around the outside of the 18-subunit rings. The exposure of the bases would provide the catalytic role played by these rings in DNA annealing. The helical filaments are formed by Redβ protein on preformed dsDNA, or on dsDNA products of the Redβ-catalyzed annealing reaction. The Redβ filament induces a left-handed supercoil in the DNA, with about one supercoil turn for every 100 bp. (Passy et al. 1999) (B) Models of the Redβ quaternary structure in the absence of DNA, the disordered Redβ–ssDNA complex, and the Redβ–dsDNA complex after annealing. (a) Model of the right-handed helix of a Redβ polymer in the absence of DNA. (b) Model of the heterogeneous and disordered binding of Redβ to ssDNA and annealing initiation. (c) Model of the stable helical Redβ–dsDNA annealing intermediate, assuming that the bound DNA is stretched by a factor of 1.53 and takes a right-handed superhelical path around the left-handed Redβ helix. The two DNA strands do not cross each other with respect to the Redβ protein surface. (Erler et al. 2009)

How exactly functions Redβ *in vivo* is not completely understood. Two groups have shown that recombination is replication dependent. Poteete first reported in 2008 in vivo experiments with a non-replicating λ and a plasmid with conditional replication. The recombination was inhibited when the plasmid replication was prevented. He showed also that SSA was more efficient when the ssDNA to anneal was complementary to the plasmid lagging strand, compared to the leading strand. His model proposed that Redβ anneals a λ 3'-overhang ssDNA onto the lagging strand template and that the replication machinery of the plasmid switches on to the invading DNA (presented in Figure 6B (Poteete 2008)). In 2011, Maresca et al. also observed replication-dependent recombination, but

proposed a different mechanism of SSA, where the λ exonuclease, Redα, digests completely one strand of the dsDNA. Redβ anneals the extremities of this new ssDNA at the replication fork (Figure 16, (Maresca et al. 2010)). This annealed DNA is then used as a primer for an Okazaki fragment. Furthermore, if the DnaG primase is mutated so that less primers are formed on the lagging strand template, the recombination is enhanced (Lajoie et al. 2012). A recent study showed that Pol III, the major DNA polymerase at the replication fork, had a strong effect on dsDNA integration, but a low effect on ssDNA integration at the replication fork (Poteete 2013). In summary, all present models for recombination by Redβ include an annealing step at the replication fork.

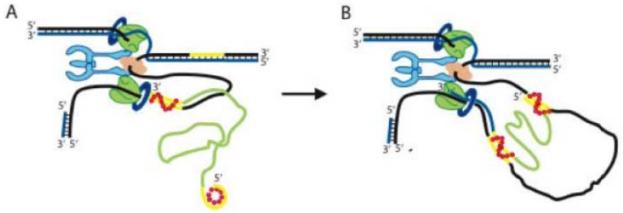


Figure 16: Model for recombination at the replication fork. Annealing of an ssDNA molecule to complementary regions on the lagging strand template at the replication fork is depicted. The ssDNA molecule comprises two flanking homology arms (~50 nt; yellow), interspaced by a heterologous sequence (light green). The Redβ annealing intermediate is shown as a curved line of red dots. The leading strand is shown in blue, lagging strand in black, DnaB helicase in light orange, the two Pol III holoenzymes are green, which are tethered to the γ/τ clamp loader (light blue), and the β sliding clamps are dark blue rings. (A) The Redβ-ssDNA protein complex anneals the 3' end first, which then primes DNA synthesis for an Okazaki fragment. (B) After replication fork progression, the second homology region becomes exposed and annealing of the 5' homology arm creates the ssDNA heteroduplex intermediate. (Maresca et al. 2010)

The fidelity of Redβ is much lower than that of RecA. In 2008, Martinsohn et al. found that Redβ recombined inverted 800 bp long sequences, up to 22% of divergence, with an efficiency 100 fold higher than RecA (Martinsohn, Radman, Petit 2008). A recent study (Li et al. 2013) also showed that Redβ recombined diverged sequences up to 17% using a different assay, the annealing of an oligonucleotide at the replication fork (also named single-strand recombineering). In addition, they found that the position of mismatches had impact on the recombination frequency: the most important for insertion of an oligonucleotide is the complementarity at the two extremities of the oligonucleotide, the 5' extremity permits the annealing before the mutation point and the 3' extremity permits the loading of the polymerase to complete the insertion. The authors also showed that the polymerases Pol I and III have major impact on the recombination frequency in the recombineering reaction.

The Eukaryotic homologue of Redβ is named Rad52. Its main function in yeast is to mediate Rad51 binding to DNA, in a way analogous to RecBCD or RecFOR in bacteria. This function of Rad52 is

essentially carried out by its Ct domain, which contains Rad51-interacting sequences. This domain is absent in phage homologs. In addition, yeast Rad52 performs Rad51-independent recombination *in vivo*, by an SSA mechanism (Mortensen et al. 1996). This activity lies in its Nt domain, which is homologous to the phage proteins, and has been crystallized. This SSA activity has also been demonstrated *in vitro* (Van Dyck et al. 2001). In addition, Rad52 promotes strand invasion in a supercoiled dsDNA, and D-loops can be observed on gel (Kagawa et al. 2001). Inspired by the fact that Rad52 is an accessory protein for Rad51 homologous recombination, *in vitro* experiments have shown that the RecA strand exchange activity is stimulated by the phage-encoded Sak recombinase (Ploquin et al. 2008), which shares a detectable homology to Rad52, (Iyer, Koonin, Aravind 2002). However, no *in vivo* evidence strengthens this set of data. On the contrary, for Redβ, we have shown that its recombination activity *in vivo* is independent of RecA (see chapter II). Interestingly, it was recently recognized that yeast encode a short Nt version of the Rad52 protein, called Mgm101, exported into the mitochondria that promotes SSA and has the same ring structure under electron microscope. This protein is essential for DNA repair in mitochondria (Mbantenkhu et al. 2011).

B) Gp2.5-like recombinases

(a) On phages

Gp2.5 is an SSB-like protein encoded by the phage T7 infecting *E. coli*. It was first shown to bind ssDNA and to be needed for T7 DNA replication (Kim et al. 1992). Later, it was observed that it anneals two ssDNA without hydrolyzing ATP. Finally, with a classic test of strand exchange, it was demonstrated that Gp2.5 promotes the formation of joint molecules (Kong, Richardson 1996), provided that the T7 helicase Gp4 is added to the strand exchange reaction. It may be that Gp2.5 is not able to promote branch migration alone, but helped with an helicase it can do so.

Hollis et al. crystallized Gp2.5 in 2001. It has the classical OB-fold found in SSB protein with a supplementary α-helix, positioned in the sequence between the β-sheets forming the OB-barrel. This structure suggests that Gp2.5 binds to ssDNA as a monomer, unlike SSB that binds ssDNA as dimers. It is possible that the α-helix plays a crucial role in the annealing activity (Hollis et al. 2001).

(b) The ICP8 orthologs in Herpes Simplex Virus

ICP8 from HSV (Herpes Simplex Virus) and Gp2.5 from T7 share a common ancestor (Kazlauskas, Venclovas 2012). Indeed, different ICP8-liks and Gp2.5-like were aligned by profile comparison. The core of ICP8 revealed the same structure as Gp2.5. These two proteins are known to have a function related to replication and recombination.

A ring structure in the presence of ssDNA was observed by electron microscopy for ICP8 (Tolun et al. 2013). Two superposed homo-nonamer rings compose this structure. The authors fitted the crystal structure into their reconstruction map. They hypothesize that the ssDNA is inside the ring structure and each two sub-rings present ssDNA to the other (Figure 17). From here, they made a novel SSA model where the two rings slide one on another until they matched two complementary ssDNA.

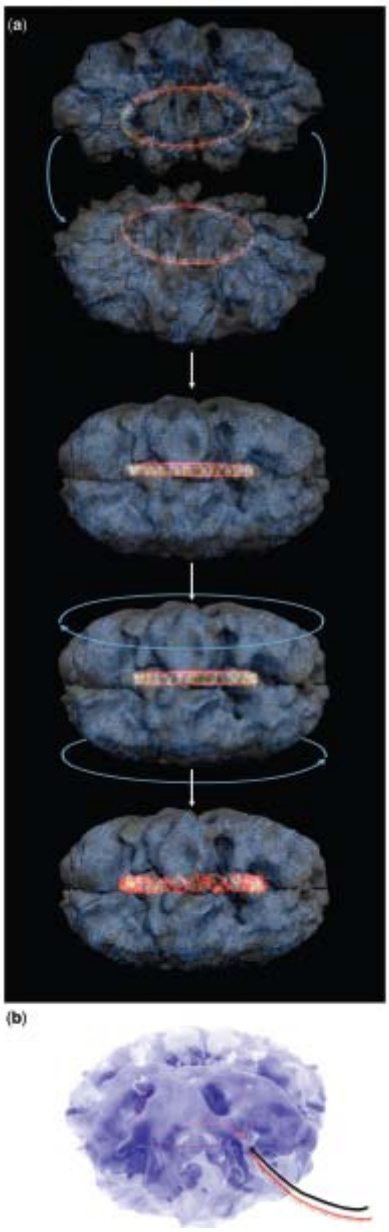


Figure 17: Artistic renderings depicting the models for two proposed mechanisms for annealing. (a) ssDNA molecules are first bound by ICP8, and then annealing follows catalyzed by the docking and sliding of these two rings. (b) It shows the ssDNA entering a pre-formed oligomeric ICP8 ring. (Tolun et al. 2013)

C) Rad51-like recombinases

Rad51-like recombinases can be split into two sub-families: RecA orthologs, named UvsX, and core-only RecA, named Sak4.

(a) UvsX

UvsX is a protein encoded by the phage T4. The UvsX structure was resolved in 2011. It is strikingly similar to RecA (Gajewski et al. 2011). UvsX has also the same activity as RecA: it promotes strand invasion and branch migration, in an ATP-dependent manner. UvsX works better and quicker to promote strand exchange in the presence of the phage helicase UvsW. A recent review compared UvsX, RecA and their homologues in yeast and human, RAD51. It appears that UvsX acts quicker than the others do, in the ATPase activity and strand exchange activity (Liu et al. 2011). However, RAD51

and RecA/UvsX proteins have not the same structure. They share a domain: the core domain, which hydrolyzes ATP and binds DNA. RecA and UvsX have the same Ct, which binds ssDNA, while RAD51 has a Nt that shares the very same activity. When the structures of recombinase filaments for RecA, UvsX and Rad51 are compared (Yang et al. 2001), RecA and UvsX share the same while Rad51 is different, due to the position of the secondary domain. Rad51 seems therefore to have evolved independently from UvsX and RecA, and to have converged phenotypically.

T4 phages have the same strategy for UvsX loading as Bacteria, Eukaryotes and Archaea. Two proteins, Gp46 and Gp47, encode for an exonuclease activity that resects dsDNA to a 3'-overhang ssDNA. Then a SSB protein catches this ssDNA, this is Gp32. After that, a mediator exchanges Gp32 by UvsX. This mediator is UvsY. It can bind both Gp32 and UvsX to exchange them on the ssDNA ((Liu, Morrical 2010), Figure 18).

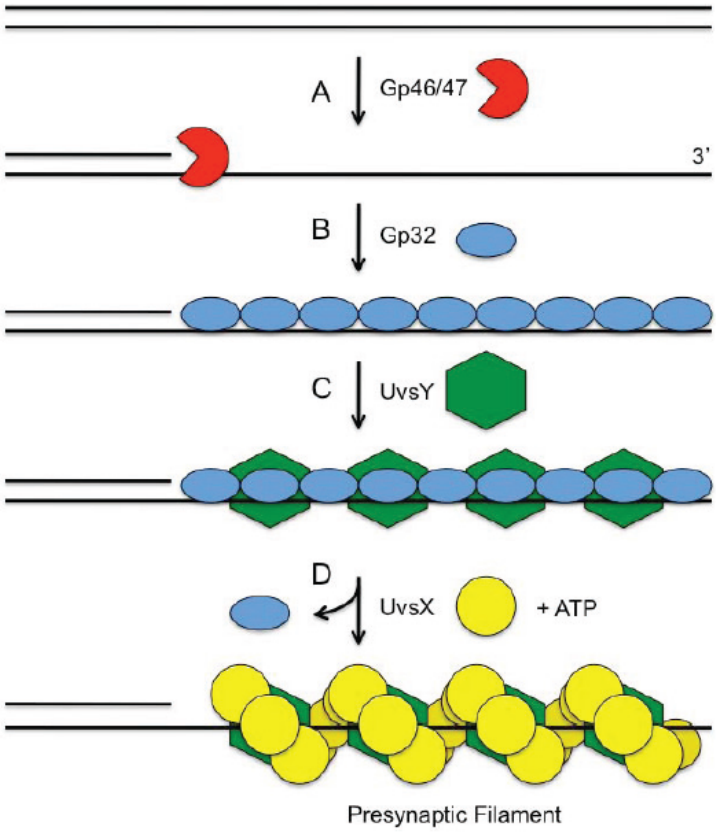


Figure 18: Presynapsis pathway in bacteriophage T4 homologous recombination. (A) A dsDNA end may be nucleolytically resected to expose a 3' ssDNA tail. The Gp46 and Gp47 proteins are thought to be the major enzymes involved in the resection step. (B) The exposed ssDNA is sequestered by the Gp32 ssDNA-binding protein, which denatures secondary structure in ssDNA and keeps it in an extended conformation. (C) The UvsY recombination mediator protein forms a tripartite complex with Gp32 and ssDNA and "primes" the complex for recruitment of UvsX recombinase. (D) UvsY recruits ATP-bound UvsX protein and nucleates presynaptic filament formation. Gp32 is displaced in the process. (Liu et al. 2011)

Maher and Morrical studied the DNA binding activity of UvsX. They observed the dissociation constants of UvsX for ssDNA and dsDNA in the presence or absence of ATP or ATPγS (a non-hydrolysable form of ATP). It seems that without ATP, UvsX has a low affinity for dsDNA, and cannot bind the ssDNA at all. It has a better affinity for the dsDNA in presence of ATPγS than in presence of ATP, which means UvsX probably releases dsDNA after ATP hydrolysis. The affinity for the ssDNA is the same in presence of ATP and ATPγS. The affinity is the highest when UvsX is mixed with dsDNA

and homologous ssDNA. This correlates with the present model for recombination: UvsX binds to ssDNA and searches for homology into dsDNA. After the resolution of the recombination intermediate, UvsX releases the dsDNA by hydrolyzing ATP (Maher, Morrical 2013).

(b) Sak4

The Sak4 family was discovered with three other protein families called Sak, Sak2 and Sak3 in an *in vivo* assay (Bouchard, Moineau 2004). *Lactoccoccus lactis* possesses systems that stop the phage infection. They are called Abi for "abortive infection". The authors searched for mutations in phages that escape to the AbiK system. They classified these mutations into four families, named above. Sak, Sak2 and Sak3 were identified to be recombinases of the Rad52 family. But Sak4 remained different from the three first, and was supposed to be a recombinase only by analogy. In 2010, Lopes et al. showed that Sak4 is a core-only RecA which promotes SSA *in vivo* (recombineering assay, (Lopes et al. 2010)). Sak4 was not further studied at that time, but some other core-only RecA had been studied in various living organisms, which are briefly presented below.

In Eukaryotes, Rad51 paralogs help Rad51 to promote homologous recombination. First in *S. cerevisiae*, two core-only RecA, Csm2 and Psy3, and two other proteins, Shu1 and Shu2, compose the SHU-complex. This complex permits the polymerization of the Rad51 filament and stabilizes it (She et al. 2012). In human, the complex BCDX2 has a core-only RecA designated Xrcc2. The other proteins of this complex, Rad51B, Rad51C and Rad51D are Rad51 paralogs with the same core-domain and with different Nt domain. This complex promotes SSA, and binds to branched-DNA (Yokoyama et al. 2004). *In vivo*, this complex participates to the recruitment of Rad51 on ssDNA (Chun, Buechelmaier, Powell 2013).

In Archaea, two RadA paralogs are core-only RecA: RadB and RadC. RadB is only found in *Euryarchaea*. It binds ssDNA and dsDNA with the same efficiency as RadA. RadB binds and hydrolyzes ATP but does not release ADP. It does not promote strand exchange either and inhibits the strand exchange activity of RadA when both are mixed together. However, when RadB is added 15 minutes after RadA in a strand exchange test, the RadA activity is improved (Komori et al. 2000). In this study, RadB seems to regulate the Holliday junction clivage by an inhibition in the absence of ATP.

RadC was first annotated as KaiC because of its high level of homology with KaiC, the circadian clock protein found in *Cyanobacteria*. The only difference is that KaiC had two RecA core domains while RadC has one only (Haldenby et al. 2009). Deletions in the *radC* gene provoke sensitivity to DNA damage agents. So these genes participate to the repair of DNA damages, contrary to *kaiC* genes (Liang et al. 2013).

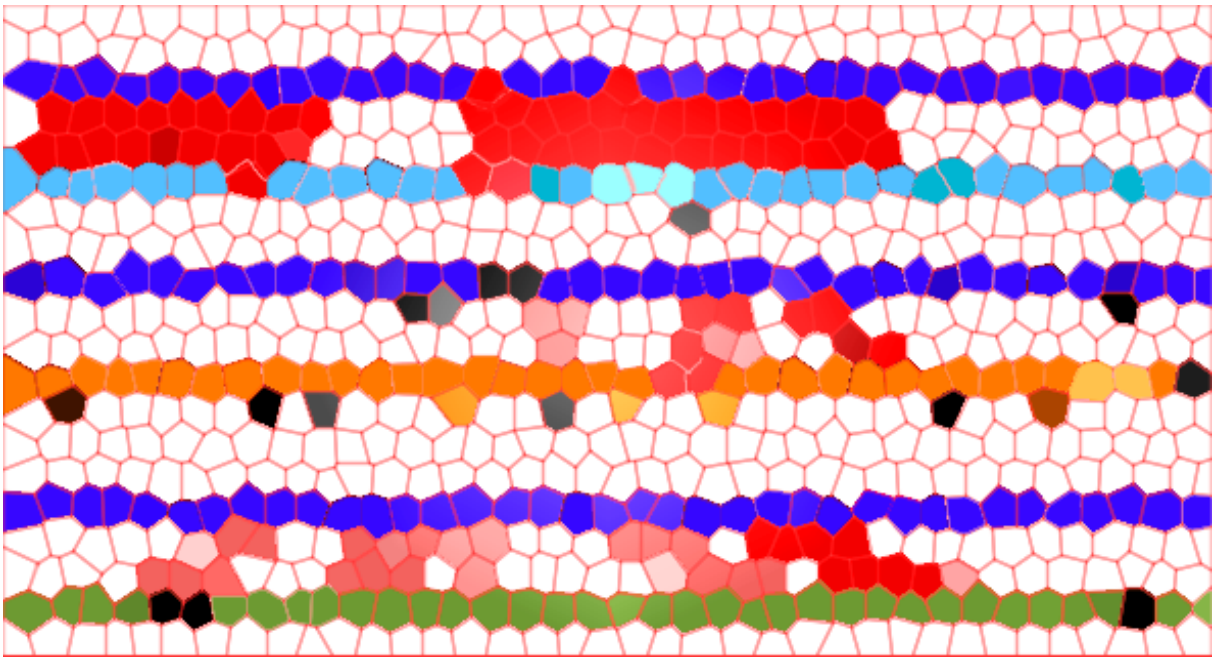
IV) All the questionings

The mechanisms permitting the remarkable acquisition of mosaic structures for temperate genomes are unknown at present. In our team, we think the unique activities of phage recombinases participate to this acquisition. We are particularly interested by their capacity to recombine diverged sequences, unlike RecA. The Martinsohn et al. study had clearly suggested a higher performance of Redβ on 22% diverged sequences, compared to RecA (Martinsohn, Radman, Petit 2008). However, the complete demonstration that this recombination was Redβ dependent could not be obtained, because the recombination assay itself relied on Redβ expression for the read-out. This led us to set up a different assay, where the acquisition of a chromosomal gene marker by the phage is followed up. Depending on the presence of homologous regions flanking this marker, we found its efficient acquisition by the phage. The question was to know if these exchanges of genetical material are mediated only by phage recombinases, only by the host recombinase RecA or both. The result is to be confronted against the hypothesis of Bobay et al., where the phage needs to choose between a phage homologous recombination system and this of the host (Bobay, Rocha, Touchon 2013). This first work is reported in chapter II. My particular contribution to this work has been to address the generality of the observation made with Redβ. Have other Rad52-like recombinases relaxed fidelity as well? I also investigated phage genomes with a bio-informatics approach, to systematically detect recent exchanges and scan for the presence of traces of homologous recombination at their vicinity.

In a second period of my research, I decided to turn to the study of Sak4 recombinases, as these were newly uncovered, and poorly characterized. I used a combined genetics and biochemical approach, to characterize this core-only RecA, and find out whether it could recombine DNA by itself, or would rather behave as accessory factor, as found in Eukaryotes and Archaea. Would this recombinase promote SSA like phage recombinases of the Rad52 family, or SI like UvsX and RecA? And would its fidelity be more similar to RecA-like recombinases, or to the Rad52-like recombinases I had studied first? This is reported in chapter III of this Manuscript.

Chapter II:

Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of rad52-like recombinases.



OPEN ACCESS Freely available online

PLOS | GENETICS

Temperate Phages Acquire DNA from Defective Prophages by Relaxed Homologous Recombination: The Role of Rad52-Like Recombinases

Marianne De Paepe[1,2]*, Geoffrey Hutinet[1,2], Olivier Son[1,2], Jihane Amarir-Bouhram[1,2], Sophie Schbath[3], Marie-Agnès Petit[1,2]*

1 INRA, UMR1319, Micalis, domaine de Vilvert, Jouy en Josas, France, **2** AgroParisTech, UMR1319, Micalis, domaine de Vilvert, Jouy en Josas, France, **3** INRA, UR1077, MIG, domaine de Vilvert, Jouy en Josas, France

Abstract

Bacteriophages (or phages) dominate the biosphere both numerically and in terms of genetic diversity. In particular, genomic comparisons suggest a remarkable level of horizontal gene transfer among temperate phages, favoring a high evolution rate. Molecular mechanisms of this pervasive mosaicism are mostly unknown. One hypothesis is that phage encoded recombinases are key players in these horizontal transfers, thanks to their high efficiency and low fidelity. Here, we associate two complementary *in vivo* assays and a bioinformatics analysis to address the role of phage encoded recombinases in genomic mosaicism. The first assay allowed determining the genetic determinants of mosaic formation between lambdoid phages and *Escherichia coli* prophage remnants. In the second assay, recombination was monitored between sequences on phage λ, and allowed to compare the performance of three different Rad52-like recombinases on the same substrate. We also addressed the importance of homologous recombination in phage evolution by a genomic comparison of 84 *E. coli* virulent and temperate phages or prophages. We demonstrate that mosaics are mainly generated by homology-driven mechanisms that tolerate high substrate divergence. We show that phage encoded Rad52-like recombinases act independently of RecA, and that they are relatively more efficient when the exchanged fragments are divergent. We also show that accessory phage genes *orf* and *rap* contribute to mosaicism. A bioinformatics analysis strengthens our experimental results by showing that homologous recombination left traces in temperate phage genomes at the borders of recently exchanged fragments. We found no evidence of exchanges between virulent and temperate phages of *E. coli*. Altogether, our results demonstrate that Rad52-like recombinases promote gene shuffling among temperate phages, accelerating their evolution. This mechanism may prove to be more general, as other mobile genetic elements such as ICE encode Rad52-like functions, and play an important role in bacterial evolution itself.

Citation: De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, et al. (2014) Temperate Phages Acquire DNA from Defective Prophages by Relaxed Homologous Recombination: The Role of Rad52-Like Recombinases. PLoS Genet 10(3): e1004181. doi:10.1371/journal.pgen.1004181

Editor: Josep Casadesús, Universidad de Sevilla, Spain

Received August 2, 2013; **Accepted** January 4, 2014; **Published** March 6, 2014

Funding: This work was performed with the financial support of Fondation pour la Recherche Médicale (MDP) and Agence Nationale de la Recherche (Dynamophage project). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: marianne.depaepe@jouy.inra.fr (MDP); marie-agnes.petit@jouy.inra.fr (MAP)

Introduction

Bacteriophages, or phages, the viruses that attack bacteria, have gained a renewed interest in the last decade with the emergence of antibiotic resistant bacteria. Despite their early discovery [1,2] and the in-depth molecular and genetic characterization of few model phages, overall phage biology is still globally poorly understood comparatively to their ecological importance. Indeed, phages and related elements dominate the biosphere both numerically and in terms of genetic diversity. Despite more than 10^5 genes already described in phage genomes, recent studies suggest that the majority of phage genes remains to be discovered [3]. The great genetic diversity of these viruses is due to their very ancient origin, their large population size and their high evolvability. Understanding evolvability of bacterial viruses will likely become a major issue with the prospective massive use of phages as alternatives to antibiotics. Notably, the mutation rate of viruses is much higher than that of cellular organisms [4,5]. Horizontal gene transfer also plays a major role in virus evolution by creating new combinations of genetic material through the pairing and shuffling of related DNA sequences [6–8].

Initial observations of hybridizing segments between phage genomes by electron microscopy, and more recent genomic analyses, have revealed the pervasive mosaicism of temperate phage genomes [9–11]. Mosaicism refers to the patchwork character of phage genomes, which can be considered as unique combinations of exchangeable genomic segments [11–15]. Temperate phages, as opposed to lytic phages, have the ability to enter a prophage dormant state upon infection, in which they stably replicate with the bacterial genome. Nearly all bacterial genomes contain multiple active or defective prophages, the latter being unable to produce phage particles. In *Escherichia coli*, prophage genes can constitute up to 14% of the genome [16], and represent 41% of a 20 species pangenome [16–18]. Intergenomic rearrangements are thus facilitated for temperate phages by frequent encounters of different viruses inside the same bacterial

Author Summary

Temperate bacteriophages (or phages) are bacterial viruses that, unlike virulent phages, have the ability to enter a prophage dormant state upon infection, in which they stably replicate with the bacterial genome. A majority of bacterial genomes contain multiple active or defective prophages, and numerous bacterial phenotypes are modified by these prophages, such as increased virulence. These mobile genetic elements are subject to high levels of genetic exchanges, through which new genes are constantly imported into bacterial genomes. Phage-encoded homologous recombination enzymes, or recombinases, are potentially key actors in phage genome shuffling. In this work, we show that gene acquisition in temperate phages is strongly dependent on the presence of sequence homology, but is highly tolerant to divergence. We report that gene exchanges are mainly catalyzed by recombinases found on temperate phages, and show that four different Rad52-like recombinases have a relaxed fidelity *in vivo*, compared to RecA. This high capacity of exchange speeds up evolution of phages, and indirectly also the evolution of bacteria.

host, for example between an invasive virus and a resident prophage.

While genome shuffling appears as a key driver of phage evolution, a quantitative description of mosaicism and analysis of its underlying molecular mechanisms are lacking. In particular, it is still debated whether phage genetic mosaicism is the product of recombination at sites of limited homology between genomes [19,20], or the result of random, cut and paste, illegitimate recombination [11,21]. In the latter hypothesis, the conservation of synteny would result from the counterselection of deleterious non-ordered gene combinations. Functions involved in homologous recombination (HR) have been extensively studied in *E. coli* and phage λ [22,23]. In *E. coli*, the RecA recombinase is essential for catalyzing DNA exchanges between homologous molecules. However, the rate of successful RecA-dependent exchanges rapidly decreases with increasing sequence divergence [24,25], suggesting that homologous exchanges should not happen between very divergent phage genomes. Phage genomes also encode recombinases catalyzing HR reactions, that have been classified into 3 super-families known as Rad51-, Rad52- and Gp2.5-like [26]. These recombinases are also found sporadically on non-bacterial viruses: archaeal proviruses encode Rad52-like genes [27] and Mimi and Herpes viruses encode a recombinase sharing homology with Gp2.5 [28]. Among the phage recombinases, the Rad52-like family is the largest and most diversified, and is itself subdivided into the Redβ, Erf and Sak groups. Among them, the λ recombinase Redβ is known to be efficient in the recombination of diverged sequences [19,29], leading us to suggest that phage recombinases could be key actors in genomic shuffling of related phages.

The λ HR system, known as Red, is expressed during phage lytic development. Red consists of two functions encoded by the *redβ* and *redα* genes. The main activity of Redβ is to mediate single-strand DNA annealing between a Redβ-bound single-stranded region and a complementary sequence. λ Redα is a double-strand-specific exonuclease that generates single-strand DNA for Redβ annealing. Redβ can also promote recombination by a RecA-like strand-invasion mechanism, especially on short DNA sequences [30]. Two other genes in the λ nin genomic region, *orf* (former name *ninB*) and *rap* (former name *ninG*), were shown to facilitate RecA-dependent gene exchanges *in vivo* between strictly identical sequences [31]. The *orf* gene product is a mediator protein that participates in the loading of the bacterial RecA recombinase on SSB-coated DNA in the absence of the three bacterial proteins RecFOR [32–36]. The *rap* gene codes for a Holliday junction resolvase [35]. Orf and Rap have numerous homologs among temperate phages, and form ~500 members families in Pfam. Interestingly, their distribution pattern among phage genomes is contrasted: among the 465 completely sequenced phage genomes collected in the ACLAME database, 191 encode a recombinase [26]. The presence of Orf is tightly associated with Recombinase+ genomes, as among the 55 phages encoding Orf, 49 also encode a recombinase, of the Rad52- or Rad51-like family. On the contrary, among the 180 genomes encoding Rap, only 100 are Recombinase+, as if Rap and phage recombinase occurrences were independent (MAP et al., to be published elsewhere). These distributions are suggestive of a more important role of Orf on phage recombinase activity. However, whether Orf or Rap stimulates Redβ mediated recombination is unknown at present.

Here, we study quantitatively the generation of mosaics between functional (i.e., infectious) temperate phages and defective prophages, and identify the genetic determinants of these exchanges. These assays reveal the preponderant role of Rad52-like phage recombination genes in exchanges involving short and diverged sequences. Moreover, a global analysis of mosaics between active temperate phages and defective prophages further reveals that these exchanges are commonplace in phage genome evolution.

Results

A recombination assay to measure formation of mosaics in phage λ

The aim of this first assay was to determine the extent and the mechanisms of genomic exchanges between an invasive infectious phage and defective prophages residing in the host chromosome. Homologous genomic regions between the MG1655 *E. coli* strain and λ phage were identified by a Blast search using relaxed parameters (same parameters as for the ANI analysis, see Fig. S6 legend). Twelve regions sharing over 70% mean identity on a stretch of at least 100 bp were found, all inside defective prophages. 3 such regions, differing in size and in the extent of identity, were selected for the experiments: i) a region in Dlp12 defective prophage, sharing 98% mean identity over 3353 bp with λ, ii) a region in Qin defective prophage sharing 96% mean identity over 657 bp, and iii) a region in Rac defective prophage sharing 88% mean identity over 703 bp, followed by another homology region nearby of 95% mean identity (Fig. 1 A, B & C). The length distribution of segments of perfect identity (segment without mismatch) is accordingly very different in the 3 regions (Fig. S1). Notably, the longest stretches of perfect identity are 298, 115 and 49 bp for Dlp12, Qin and Rac respectively. An antibiotic resistance gene, conferring resistance to either chloramphenicol (*cat*) or kanamycin (*kanR*), was introduced in the middle of each selected region (Fig. 1 and Materials and Methods). As a control, we used a strain with the *cat* gene into *ilvD*, which has no homology with λ and whose deletion does not impact phage growth.

Defective prophages can reportedly excise and even replicate in different strain backgrounds under certain conditions, notably during phage infection [31,37,38]. We thus verified by PCR that the 3 studied defective prophages do not excise during infection by λ. Moreover, as Rac has an active but normally repressed replication origin, we checked by semi-quantitative PCR whether

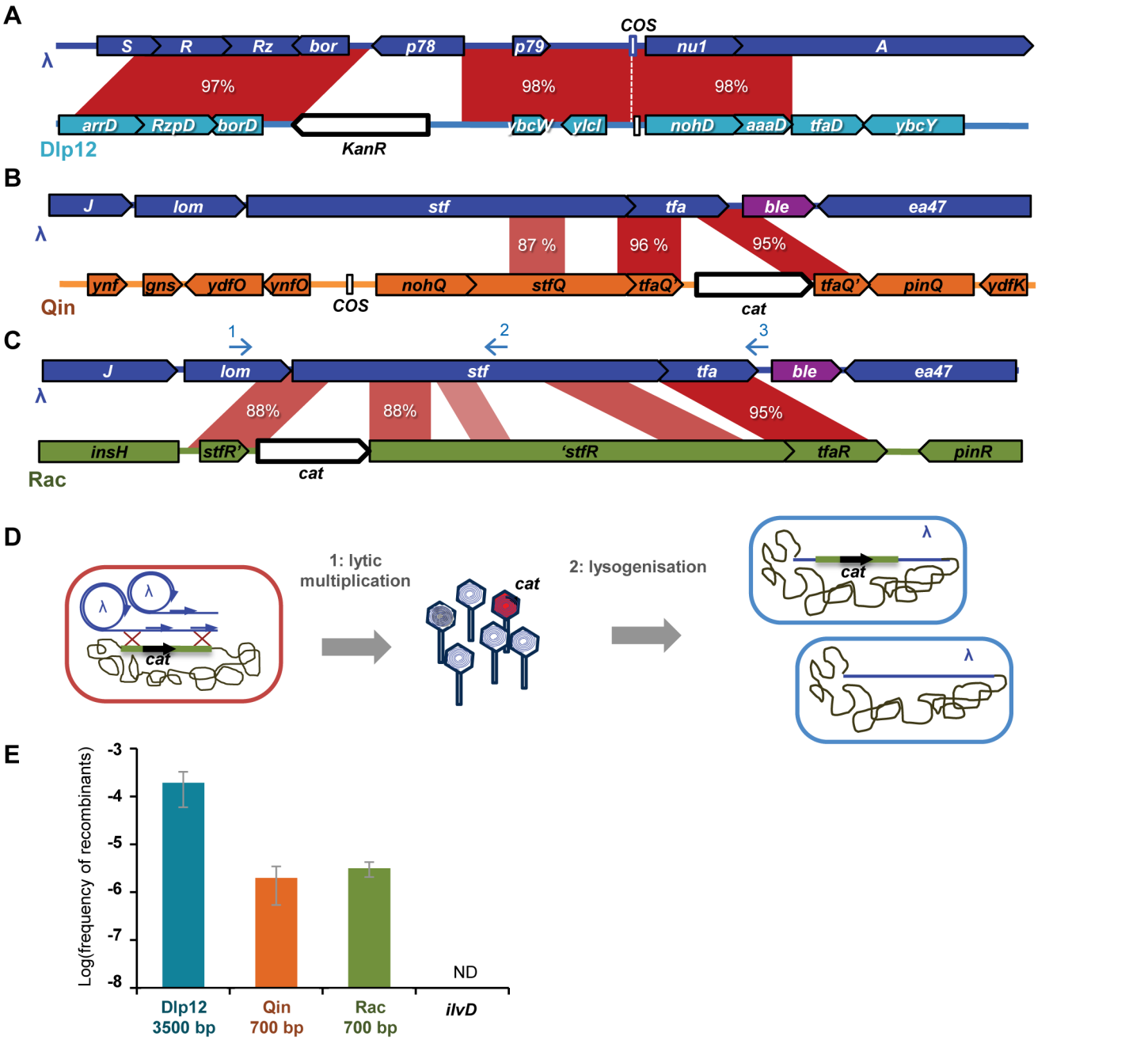


Figure 1. Recombinants between λ and defective prophages are formed during lytic cycle. A, B and C: Maps of the three regions of similarity between λ and MG1655 used in this study. Antibiotic resistance genes (white arrows) were inserted in the defective prophages, in the middle of the identity regions. *KanR* stands for the gene conferring resistance to kanamycine, the *cat* gene confers resistance to chloramphenicol. Identity region between λ and Dlp12 (A) spans across essential lysis genes (R and Rz) and terminase genes (A and nu1), separated by the *cos* site. Identity regions between λ and both Qin (B) and Rac (C) span across side tail fibers genes (*stf* and *tfa*). The *ble* gene, that confers resistance to phleomycine, was inserted between *tfa* and *ea47*, under the constitutive promoter P*sacB*. Blue arrows in C indicate the position of the 3 oligonucleotides used to sequence the recombinants. D: Schematic representation of the recombination assay: (1) λ phage is multiplied on a strain in which an antibiotic cassette has been inserted in a region of homology. (2) The phage produced is used to lysogenize a new strain. The total number of lysogenized bacteria is determined by their phleomycine resistance, while bacteria lyzogenised by a recombinant phage are also resistant to either chloramphenicol or kanamycin. E: Frequency of λ recombinants with Dlp12, Qin and Rac. Bp numbers indicate the size of the homology regions. Mean ± standard deviation of at least 3 independent recombination assays is indicated.
doi:10.1371/journal.pgen.1004181.g001

its replication was induced upon λ infection. We found no evidence of such a phenomenon (Fig. S3).

Recombination between λ and the marked defective prophages can result in the integration of *cat* or *kanR* antibiotic resistance gene into λ (Fig. 1E). As this does not produce a directly detectable phage phenotype, *E. coli* cells were further lysogenized with the resulting phage, and the proportion of cells lysogenized by WT or antibiotic-resistant recombinant phages was determined by plating

on selective antibiotic media (Fig. 1D and Materials and Methods). To be able to detect all lysogenized bacteria, we used a λ strain marked with the phleomycine resistance gene *ble* (Urλ*ble* strain, Table S1 and Materials and Methods). This recombination assay enables the estimation of the horizontal gene transfer rate into the λ genome when it infects its host.

Mosaicism is produced by homologous recombination

λ recombinants with Rac and Qin were observed at frequencies around 2×10^{-6}, and were a hundred fold more numerous with Dlp12 (Fig. 1E). Interestingly, the frequencies of recombinants were similar for Rac and Qin despite different extent of identity on comparable lengths (Fig. 1 and S1). In contrast, the recombinant frequency with the control chromosomal gene *ilvD*, presenting no homology with λ genome, was below our detection threshold of 5×10^{-9}. PCR analysis on 20 λ recombinants within each of the 3 loci showed that all had incorporated the resistance gene at the expected position for an homologous exchange. 12 to 19 were sequenced to confirm that homologous recombination occurred in the targeted region and to identify the exact junctions (Fig. 2). Among recombinants with Rac, the majority resulted from recombination events within the adjacent 88% identity regions, but one third had recombined on the right in the nearby 95% homology region in *tfa*.

HR with Rac and Qin leads to the inactivation of λ side tail fibers genes (*stf* and *tfa*), which reportedly improves phage growth compared to the parent in a soft-agar overlay [39]. To evaluate if this increases the recombinant frequency by favoring their growth, we performed the same recombination assay with the sequenced λPaPa strain of λ, mutated in its *stf* gene [40]. The frequency of λPaPa recombinants with Rac was not significantly different from that with Urλ ($p = 0.18$, Student T-test), ruling out the possibility that the advantage of side tail fiber inactivation distorts our conclusions. This can be explained by the low recombination rate: most recombinants statistically arise only during the last (and third) cycle of phage growth, as phage engaging in the third lytic cycle are 10,000 more numerous than those involved in the first one. Sequencing of 12 λ-Rac recombinants showed that they all correspond to different hybrids (Fig. 2A), confirming the absence of amplification and also indicating the absence of a recombination hotspot.

The high recombination frequency with Dlp12 is expected for two reasons. First, Dlp12 shares the largest region of homology with λ (3353 bp), and secondly, the region of homology contains the λ *cos* site (Fig. 1A). Double-strand DNA breaks at *cos* sites, created for genome encapsidation, stimulate recombination on its left side, the right side bound by the terminase being protected [41]. Double-strand breaks in both λ and Dlp12 might thus stimulate the recombination in this region. Sequencing of 19 λ-Dlp12 recombinants indeed revealed a high proportion of junctions in the immediate proximity to *cos* (10 out of 19 clones, Fig. 2B). 3 out of 19 junctions were nevertheless found on the right side of *cos*, indicating that double-strand breaks at *cos* improve but are not necessary for recombination events. Interestingly, all tested λ-Dlp12 recombinant phages were active, including those that had replaced the essential λ genes R and Rz (lysis) or *nu1* (terminase) by

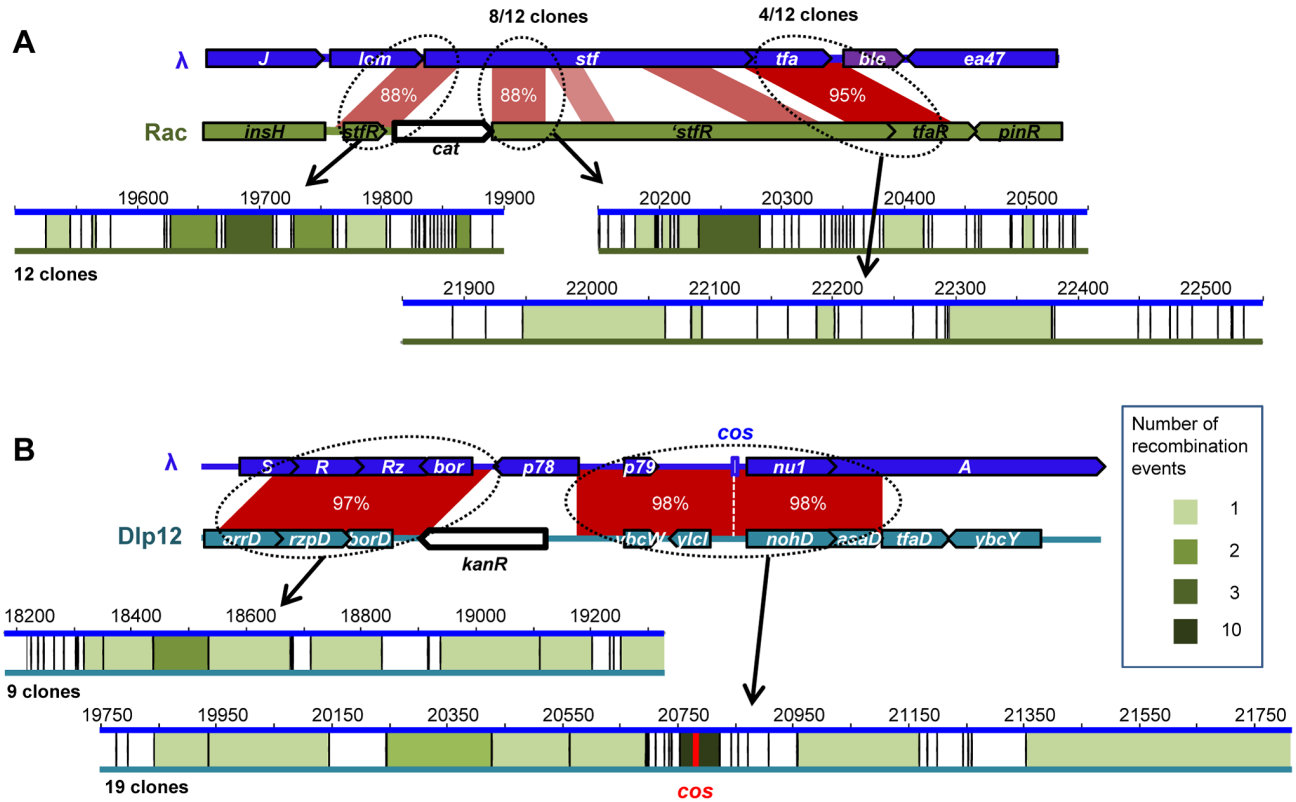


Figure 2. Position of recombination events in homology regions between λ and defective prophages. A: Positions of recombination events between Rac and λ for 12 sequenced clones. Mismatches in similarity regions between rac and λ are represented by vertical black bars. Intervals between two mismatches in which sequencing revealed that recombination occurred are colored in green, intensity representing the number of recombination events (see legend). B: Positions of recombination events between Dlp12 and λ. Nine and 19 clones were sequenced on the left and right sides of *cat* gene, respectively. Half of the recombination events scored occurred at direct proximity of the *cos* site.
doi:10.1371/journal.pgen.1004181.g002

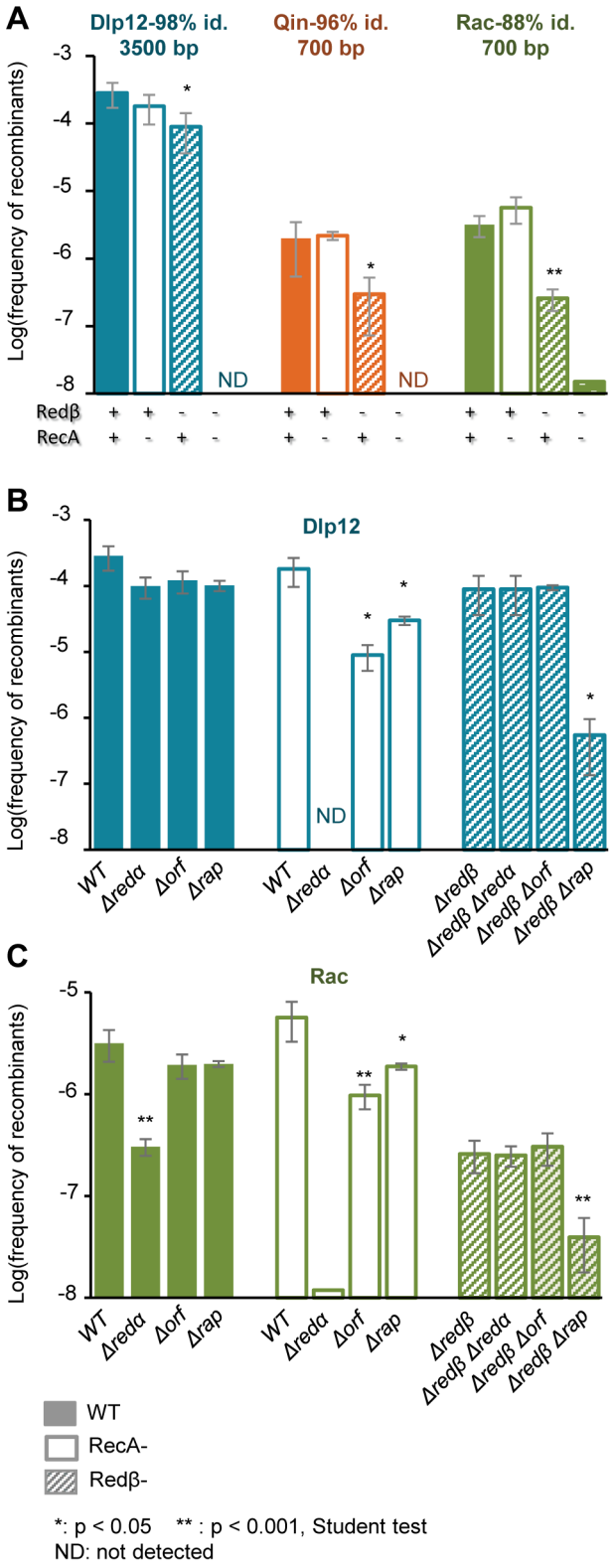


Figure 3. Recombinants between λ and defective prophages are formed preferentially by the Red-pathway, especially when sequences are short or diverged. A: Frequency of recombinants with Dlp12, Qin and Rac, as a function of Redβ and RecA presence (indicated below the bars). Both recombinases are able to catalyze the exchanges, but Redβ participation is more pronounced, especially on short and diverged sequences. When both recombinases are absent, almost no recombinants are obtained. B: Role of *redα*, *orf* and *rap* genes on the exchanges in the large region of high homology between λ and Dlp12. Analysis is performed in the presence of both RecA and Redβ (full bars), and also with only Redβ (empty bars), or RecA (stripped bars). Phage genotypes are shown below the bars. C: Same genetic analysis on the short (700 bp) and more diverged region of homology between λ and Rac. Mean ± standard deviation of at least 3 independent recombination assays are indicated.
doi:10.1371/journal.pgen.1004181.g003

those of Dlp12, indicating that these Dlp12 genes are functional in λ.

Recombination genes involved in mosaicism

We have shown that genetic exchanges with λ are driven by the presence of homologous regions. We next questioned the respective roles on these exchanges of the bacterial and phage HR genes, *recA* and *redβ*, and also of two other λ recombination genes, *orf* and *rap*. As *redβ* and *redα* λ mutants have a reduced burst size (Fig. S2), conditions of phage growth on the marked strains were adapted to ensure that the same number of phage generations was realized (see Materials and Methods).

The frequency of recombination with the large, quasi-identical segment of Dlp12 (3.3 kb, 98% identity) was not affected by a single *recA* deletion, and only 3-fold by a *redβ* deletion (Fig. 3A, blue bars). However, when both recombinase genes were deleted, exchanges were completely abolished. This shows that both RecA and Redβ produce recombinants independently of one another, which is indicative of redundant pathways. Within Qin and especially Rac regions, showing more divergence with λ, single *recA* deletion had similarly no effect on recombination frequencies. The single *redβ* deletion however had a pronounced effect on the recombination frequencies (6- and 13- fold respectively, Fig. 3A, orange and green bars), indicating that most recombination events are formed by Redβ when homology is reduced, reflecting a lowest activity of RecA on these substrates. The residual recombinants formed in the absence of Redβ almost completely disappeared with the double deletion of *recA* and *redβ*. Only 2 recombinant clones were obtained in the *recA redβ* double mutant, with Rac, and were found by PCR to result from homologous recombination. They might have originated from the expression of the *recET* recombination genes of Rac. Normally the *recET* genes are completely repressed in *E. coli* MG1655, even upon λ infection [42], and thus cannot promote recombination, but rare *sbcA* mutations [43] or incorporation in λ (the so-called λ-rev genotype, [44]) can activate them.

Role of other phage encoded recombination promoting genes

The activities of recombinases are stimulated by numerous host or phage-encoded cofactors, which either prepare the substrate for recombination or act at latter stages to resolve the DNA heteroduplex structure. The main partner of Redβ is the double-strand-specific exonuclease Redα that transforms double-strand DNA into single-strand DNA, the substrate for Redβ. Redα is dispensable for Redβ recombination if the DNA substrate is initially in a single-strand form [45].. Here, deletion of *redα* gene had the same effect as the *redβ* deletion, and basically disabled Red-mediated recombination in Dlp12 and Rac regions (Fig. 3B&C). This finding reveals a need for single-strand DNA formation for Red-mediated recombination in our assay.

We then tested the effects of other recombinase helper proteins on recombinant frequencies. λ encodes two such proteins, Orf and Rap, which belong to protein families highly prevalent in phage genomes (respectively 55 and 180 homologues in 465 phage

genomes, our unpublished work). Orf and Rap participate in RecA-dependent recombination by supplying a function equivalent to bacterial encoded recombination cofactors, respectively RecFOR and Ruv resolvase [36,46]. Here we asked whether they could also participate in Red recombination, and whether their role was more pronounced on diverged sequences than on easy to recombine long sequences. Single deletion of *orf* or *rap* resulted in a not significant 3-fold reduction of λ recombination with Dlp12 (Fig. 3B, first set of data). However, in a *recA* strain, in which all gene exchanges are Red-mediated, recombination was reduced by 10-fold in the Δ*orf* mutant, and by significant 3-fold in the Δ*rap* mutant. This result reveals that Orf and to some extent Rap participate in Red-mediated recombination. Finally, in the genetic context where RecA mediates all gene exchanges, the *orf* deletion had no effect, while the *rap* deletion decreased recombinant frequency by 300-fold, as expected from an earlier report [47]. The same genetic analysis in the Rac region assay revealed the same dependencies (Fig. 3C). In conclusion, Orf and to a lesser extent Rap facilitate homologous recombination with λ when Redβ is the only recombinase, with no indication of an increased role when DNA sequences are more divergent.

Homologous recombination-dependent gene shuffling also occurs with phage Φ80

In order to determine the generality of our observations with λ, we performed similar recombination assays with the lambdoid phage Φ80. Its genome homology with λ is mostly clustered in the capsid region, while the rest of the genome is highly divergent with only few segments of homology [48]. Φ80 encodes a putative recombinase from the Rad52 family, hereafter named $RecT_{\Phi 80}$, that shares 32% identity at the amino acid level with Redβ.

As previously described with λ, we detected by Blast search 9 regions of homology between Φ80 and *E. coli* MG1655 genomes, and selected 3 of them for the recombination assay: i) the 3′ part of the bacterial core gene *yecD*, of unknown function, presenting 96% mean identity over 300 bp, ii) in the *nohD* region of the defective prophage Dlp12, sharing 79% identity over 980 bp with the terminase genes *nu1* and *A*, and iii) within the *nin* region of Dlp12, sharing 67% identity over 500 bp (Fig. 4 A, B, C). For each region, the longest stretch of perfect identity is 116, 41 and 11 bp for *yecD*, *nohD* and *nin*, respectively (Fig. S1B). Homology regions were labeled with the *cat* gene, and Φ80 genome was marked with the *ble* gene. Φ80 phage was then propagated on the modified *E. coli* strains and the number of recombinant phages was scored (Fig. 4D).

Recombinants were produced at a frequency of 6×10^{-7} per phage with the *yecD* locus and 8×10^{-7} with the *nohD* locus of Dlp12 (Fig. 4D). In the last case, recombination disrupts the essential terminase genes *nu1* and *A*, but this does not prevent encapsidation of recombinant Φ80 genomes, due to the terminases encoded by the other copies of Φ80. However, the recombinants formed will not give lytic progeny upon infection of a new host, ensuring that the recombinants counted in the assay are exclusively those formed during the last lytic cycle. As discussed above, the presence of the *cos* site within the homology region probably increases the production of recombinants. Within the last region, sharing only 67% identity over 500 bp with λ, the recombination frequency was below 1×10^{-8}, our detection threshold (*ybcN* locus, Fig. 3C). PCR analysis revealed that the 12 recombinants scored had incorporated the resistance gene at the locus expected for homologous exchange.

As with λ, we then determined the respective role of homologous recombination enzymes RecA and $RecT_{\Phi 80}$ in these genetic exchanges. *recA* deletion slightly diminishes recombinant frequency by 2-fold on both loci, whereas $recT_{\Phi 80}$ deletion has a pronounced effect on the more diverged sequences (Fig. 3C).

These results show that homology-dependent mosaic formation driven by phage encoded recombinases is not restricted to λ and may be a general event among temperate phages encoding their own homologous recombination functions. Exchanges occurred even when the resulting recombinant phages were no longer viable as a lytic phage, and appeared to be only limited by the degree and length of sequence homology.

Like Redβ and $RecT_{\Phi 80}$, Erf_{D3} and $RecT_{Rac}$ recombinases are efficient on diverged sequences

The capacities of Rad52-like phage encoded recombinases and cofactors were further explored with an intra-bacteriophage recombination assay, which enables the comparison of recombinase activities on the same DNA substrate. We monitored homologous recombination within the λ genome, between two inverted 800 bp sequences introduced on each side of the P_L promoter (described in Fig. 5A and [19]). Briefly, inversion of the P_L promoter leads to a detectable phenotypic switch, as it prevents transcription of the *red* and *gam* genes, enabling growth on a P2-lysogen, contrary to the non-inverted phage. Vice versa, inverted P_L Red^- Gam^- phage cannot grow on a *recA* strain, in contrast to the non-inverted wild-type phage. Two different recombination cassettes were used, the inverted sequences being either strictly identical, or 78% identical. The switch was measured in the two directions, and in all cases, the recombination assay was performed by growing phages on a restrictive host for the multiplication of the recombinants, ensuring that the recombinants are produced only during the last generation.

In this study, λ *redβ* and *redα* genes were replaced by other pairs of recombinase genes and their respective associated exonuclease. The first pair is *recT* and *recE* (fragment coding for the Cter part of the protein), from the defective prophage Rac. The second pair is the predicted Erf family recombinase gene *erf*, with its associated predicted exonuclease gene *exo*, from D3 phage infecting *Pseudomonas aeruginosa* [49]. Resulting phages were grown to confluence on a bacterial lawn, and the frequency of recombinants was measured by differential platings. On both the 100% and 78% identical substrates, efficiency of recombination mediated by the $RecT/RecE_{Rac}$ and Erf/Exo_{D3} pairs was similar to that mediated by λ Red proteins (Fig. 5B). First, this demonstrates that the D3 *erf* gene product is indeed a recombinase, and as efficient as Redβ and $RecT_{Rac}$. For all 3 recombinase/exonuclease pairs, recombination between identical or diverged sequences was respectively 20- and 100-fold higher than in the RecA pathway.

This result extends our previous observation with λ and Φ80, and suggests that during the lytic cycle, the high efficiency and low fidelity of phage-dependent recombination might be a general phenomenon that promotes horizontal gene transfer into phage genomes.

Genomic analysis reveals abundant mosaics among and between temperate and defective phage genomes, but none between virulent and temperate genomes

We thus attempted to quantify the traces that such exchanges might have left on phage genomes on an evolutionary time scale by analyzing a collection of *E. coli* phages (Table S4). In particular, we examined whether traces of HR events could be detected at the boundaries of recently exchanged DNA fragments. In a previous study of mosaics formed between temperate lambdoid phages [19], we showed that half of the mosaics, defined by Blast hits with more than 90% identity, were flanked by at least one region

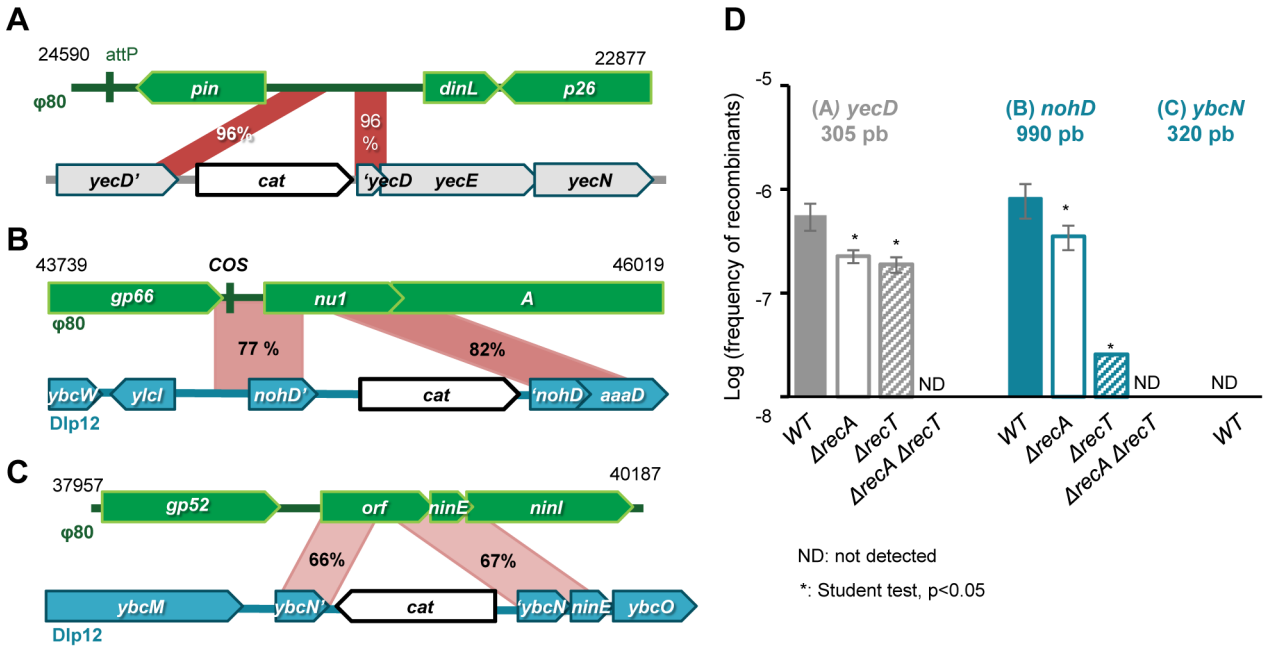


Figure 4. Formation of Φ80 hybrids also depend on phage recombinase. A: map of the 300 bp region of 96% identity between the Φ80 region near attP site and *E. coli* gene *yecD*. B: Map of the 980 bp homology region between Dlp12 and Φ80, spanning through the essential terminase genes *A* and *nu1*. C: Highly divergent 500 bp homology region between the non-essential nin region of Φ80 and Dlp12. D: Frequency of recombinants for each homology region, as a function of Redβ and RecA (genotypes are indicated below the bars). Mean ± standard deviation of at least 3 independent recombination assays are indicated.
doi:10.1371/journal.pgen.1004181.g004

sufficiently similar to be an indication of a region of homology preexistent to the recent exchange identified by the Blast hit. These above background homology regions were hereafter named "HR traces". This study was extended and refined here, so as to include more genomes from temperate (24 genomes), defective (34 genomes) or virulent (26 genomes) *E. coli* phages.

Results are summarized in Table 1. We detected a large number of recently exchanged genomic segments among temperate and defective phages (between 0.8 and 2.4 mosaics per genome pair). No mosaic was found between virulent and temperate or defective phages. We also found a few mosaics between virulent phages (0.06 per genome pair, Table 1). Interestingly, among temperate and defective phages, the density of mosaics, *i.e.* the numbers of mosaics per 10 kb, was similar (Fig. 6, lower part of the diagonal), suggesting that exchanges do occur at similar frequencies among defective and functional temperate phages. Among temperate phages, two sub-groups not sharing mosaics were found: the larger one contained lambdoid phages (i.e. λ-like, ε15-like and P22-like), the smaller group contained P2-like, P1 and Mu-like phages. Among pairs involving defective phages, exchanges between these two groups are found, probably because non functional gene combinations are not counterselected in defective prophages. Interestingly, the size distribution of mosaics indicates that small gene fragments are much more often exchanged than complete functional modules (Fig. S5).

We next investigated whether the mosaics had HR traces, i.e. flanking regions that have a lower mean identity than the mosaic but a higher than expected identity (see Figure 7 and Text S1 for the principle of analysis). Briefly, HR traces were detected by complete realignment of the mosaic flanking regions, and identification of above-background levels of nucleotide identity. We found regions suggestive of HR for 32% of the mosaics occurring among temperate phages, 35% of those formed between defective and temperate phages, and 31% for the inter-defective mosaics (Fig. 6, upper side of the diagonal, and Table 1). In most cases, only one, rather than two HR traces were found flanking the mosaics. This may be explained by other mechanisms than HR to generate exchanges but also by successive rounds of exchanges, involving different phage pairs, as illustrated in Fig. S7.

Discussion

It has been known for a long time that related phages can form recombinant hybrids when infecting simultaneously the same bacterial cell, or by recombination with integrated prophages [6,50–52]. λ defective mutants in particular can be rescued by genes present on prophages [31]. Here we quantified for the first time these exchanges between phage genomes, and helped to precise the rules that dictate phage mosaicism, demonstrating that they are dependent on homologous recombination enzymes, with a preponderant role for the phage-encoded recombinases when sequences are diverged. Our results suggest that illegitimate recombination is probably much less frequent during phage genome evolution. This high level of exchanges among phage sequences is to be contrasted with the barrier observed against similar exchanges during bacterial recombination: on *E. coli* plasmids, recombination involving 4% diverged sequences is reduced by 10,000-fold compared to identical sequences [53], whereas during the lytic cycle, recombination between 12% diverged sequences is reduced only by a 100-fold compared to identical sequences (Fig. 5B).

Mosaic formation is driven by Rad52-like phage recombinases

We demonstrate that phage encoded Rad52-like recombinases play a primordial role in recombining regions sharing only short

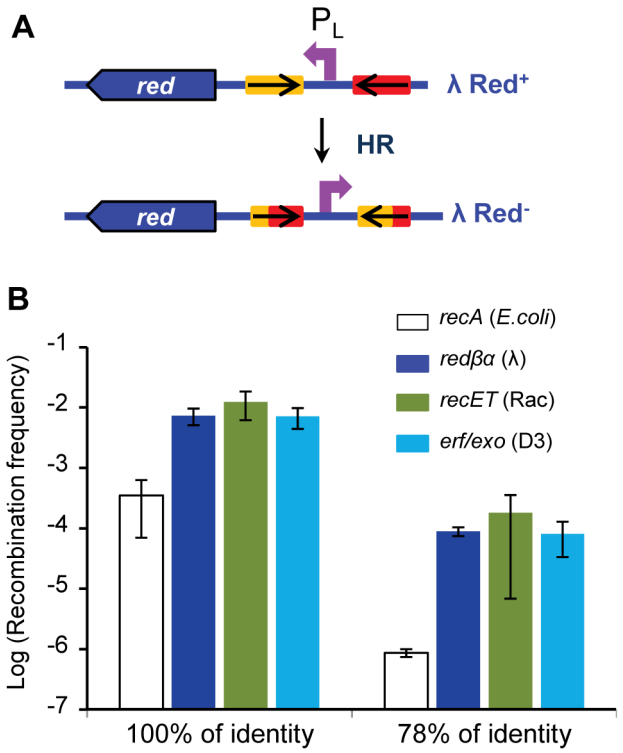


Figure 5. High efficiency of recombination of two other Rad52 recombinases. A: Principle of the assay: two 800 bp homologous sequences (*oxa* genes, as in [19]), inversely repeated, represented by the red and yellow rectangles, are inserted so as to flank the P_L promoter in the λ genome. Inverted sequences are either 100% or 78% identical. When homologous recombination occurs between the repeated sequences, the P_L promoter is inverted, which leads to a phenotypic switch, because the *red* and *gam* genes are no longer expressed (see Materials and Methods). B: Recombination frequencies scored with three different pairs of phage recombinase/exonuclease, and compared to RecA pathway. The *recET* (from Rac prophage) and *erf/exo* (from phage PA73) genes were substituted in place of the λ *red* genes. Values shown for *redαβ* and *recA* pathways are those reported in [19]. Mean ± standard deviation of at least 3 independent recombination assays are indicated.
doi:10.1371/journal.pgen.1004181.g005

stretches of homology. Interestingly, on these substrates, all four Rad52-like recombinases that we tested (Redβ, $RecT_{Rac}$, Erf_{D3} and $RecT_{\Phi 80}$) had a higher activity than RecA during λ replication. These results strongly support the view that phage Rad52-like recombinases, predominant among *E. coli* lambdoid phages [54], play a crucial role in genomic shuffling. Interestingly, a recent bioinformatic study showed that the level of mosaicism is higher for phages encoding a recombinase than for others [54], supporting this hypothesis. Whether the same holds true for the two other large families of recombinases encountered in phages (Gp2.5 and Rad51, [26]) remains to be investigated experimentally. *Staphylococcus aureus* temperate phages exhibit a large level of mosaicism [19], and encode indifferently all three types of recombinases, which suggests that the two other families might also have the same property of relaxed fidelity. Further work aiming at comparing these recombinases side by side will allow addressing this point in detail.

The high efficiencies of Rad52-like recombinases on diverged sequences could be due to their higher concentration or activity, compared to RecA, during phage infection. However, on highly homologous sequences, RecA or phage recombination pathways result in a similar yield of recombinants (Fig. 3A and 4D). The 10–20 fold higher efficiency of Redβ or $RecT_{\Phi 80}$ compared to RecA pathway is unveiled only on more divergent sequences. It could result from activity on shorter perfect sequence identity segments to recombine as compared to RecA. The Minimal Efficient Pairing Segment (MEPS), the minimal size of exact pairing required for efficient exchange *in vivo*, has indeed been found to be 31–34 bp for RecA [24] and only 23–27 for Redβ [24,55]. Our *in vivo* assays involve sequences that do not have regularly spaced mismatches, so that slight changes in the number of segments above MEPS size may have drastic effects. Interestingly, Li and collaborators have recently reported that recombineering with regularly spaced, diverged oligonucleotides is effective up to 1 mismatch every 5 nucleotide, which corresponds to a mean divergence of 17% [29], and fits nicely our observations. The low sensibility to divergence of phage recombinases could also result from different sensitivities to methyl-directed mismatch repair (MMR), comprising MutSLH proteins. Indeed, the high fidelity of homologous recombination is caused not only by the intrinsic properties of the enzymes, but also by MMR inhibition of exchange if mismatches are present in recombination intermediates [56]. Classically, defects in MMR provoke a 100-fold increase of HR, either in RecA [57] or Redβ pathways [58]. However HR is less affected by the MMR during the λ lytic cycle: RecA-dependent recombination is enhanced by only 2 to 8 fold in a *mutS* background, depending on the level of divergence, and Redβ-dependent recombination is not affected at all [19]. Whether mismatch repair is titrated or inhibited during the phage lytic cycle is an open question that deserves further inquiry.

In λ, *red* mutants have a 5-fold reduced burst size (Fig. S2), which at present remains unexplained (see [59] for a review discussing this point). Our observation that RecA substitutes completely for Red for HR on identical sequences seems to exclude that the loss of viability of *red* mutants is due to a strict HR defect. RecBCD pathway is inactive in λ infected cells as λ lacks Chi sites and moreover express a RecBCD inhibitor. Further, we could not detect any synergy of the *red* and *recA* mutations on plaque sizes (Fig. S4). Whether Red impacts replication or any other stage of the phage cycle remains to be investigated, as well as why Φ80 recombinase does not impact Φ80 plaque size.

Biochemical evidence suggests that Sak and Sak3, two Rad52-like recombinases of lactococcal phages, act as cofactors of RecA [60,61]. Some *in vivo* work also pointed to such a role for Redβ on non-replicating λ genomes [62]. In the present work, the phage and host recombinases appear rather to work independently from each other, and redundantly on highly similar sequences.

The RecA cofactor function of phage-encoded Rad52-like proteins may therefore be a minor activity *in vivo*. This situation is contrasted with the yeast-encoded Rad52 protein, which is essentially known for its Rad51 cofactor activity. However, Rad52 also performs some repair reactions in a Rad51 independent way [63]. Whether these Rad52 activities are tolerant to diverged DNA is unknown at present. Interestingly, another kind of mobile genetic elements that present genomic mosaicism, the integrative conjugative elements of the SXT/R391 family, has been reported to encode a recombinase of the Rad52 family (named s065) [64]. It was found to act in the formation of hybrid ICEs, independently of RecA. The assay involved 95–97% identical sequences, and RecA was the dominant pathway. Whether s065 becomes dominant when recombination involves more diverged sequences, remains to be investigated.

Table 1. Data collected on mosaics, by categories of genome pairs compared.

Phages pairs	T-T*	T-D	D-D	V-V	V- others
N genome pairs analysed (p)	266	807	549	289	1508
N of Blast matches	711	895	670	16	0
% of IS matches	9.7	26.7	33.6	0.0	0
N of mosaics analyzed for HR traces (m)	635	631	430	16	0
N mosaics/genome pair (m/p)	2.39	0.78	0.78	0.06	0
Median mosaic length (bp)	404	494	458	131	0
Median mosaic id%	96.1	94.8	95.1	91.5	0
N of mosaics without HR traces	432	411	298	5	0
N with 1 trace of HR	169	171	106	6	0
N with 2 traces of HR	34	49	26	5	0
P (≥1 trace of HR at random)	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$1.8\ 10^{-14}$	

*T, temperate, D, defective, V, virulent.
doi:10.1371/journal.pgen.1004181.t001

Contribution of accessory proteins Orf and Rap to recombination

Interestingly, both Orf and Rap proteins have numerous homologs among temperate phages, and the presence of Orf is strongly associated with the presence of a recombinase. Notably, Φ80 possesses homologs of both *orf* (*gp53*) and *rap*. Rap (recombination adept with plasmid) increases RecA-dependent recombination between λ and a plasmid sharing perfect identity by a 100-fold [47]. In line with this result, we found that in the RecA pathway, *rap* deletion results in a 500-fold decrease in recombination with Dlp12. With the Rac substrate, where the activity of the RecA pathway is minor, *rap* deletion results in a further 10-fold decrease, again underlining the importance of *rap* in this pathway. For Redβ-dependent recombination, the decreases due to *rap* deletion were only 3 and 6-fold for Dlp12 and Rac substrates, respectively. This is probably related to the fact that the Rap substrates, Holliday junctions, are not necessarily formed by Redβ Indeed, Redβ principally recombines by a strand assimilation mechanism, which consists in the single-strand annealing between Redβ-bound single-strand DNA and the lagging strand template of a replication fork, a reaction that does not need a resolution step [65–68]. On the contrary, RecA catalyses mainly strand invasion reactions, which generate Holliday junctions [69,70]. The Rap protein acts essentially independently from Redβ, a conclusion which agrees nicely with the independent genomic distribution of these two genes.

Orf is involved in displacing of SSB from single-stranded DNA, which facilitates RecA binding. We found that *orf* deletion decreases Redβ dependent recombination by 10-fold, and does not affect the RecA pathway. This is in line with the strong association observed between *orf* and recombinase genes in phage genomes. This Orf effect might reflect the proportion of cases where Redβ enters in competition with SSB. As Redβ loads onto single-strand DNA immediately after Redα, the two proteins being supposedly in interaction, the Orf activity to remove SSB should not be essential, as SSB might not be present on most of the substrates recombined by Redβ.

Quantification of mosaics and traces of homologous recombination in phage genomes

The presence of all these recombination genes in phage genomes influences their long term evolution. Indeed, here we add evidences that HR leaves frequent traces among phage genomes. We found recently formed mosaics - defined as sequences sharing more than 90% identity - between temperate and defective genomes, but also a few ones among virulent phages. Interestingly, recent gene exchanges were identified only inside the lambda-like group or the P1-P2-Mu group, but not between each group. Traces of homology flanking the mosaic were detected in ~30% of the mosaics between temperate phages and/or defective prophages. Whether the remaining 70% mosaics were formed by HR but have shorter or masked traces (Fig. S7), or were created by non-homologous recombination remains an open question.

Remarkably, no recent mosaics were found between temperate and virulent phages infecting *E. coli*. In fact, a well documented case of ancient horizontal gene transfer among a large group of unrelated temperate and virulent phages exists, and concerns the side tail fiber (*stf*) genes [71–73]. However, this case is very specific, as these genetic exchanges are most probably driven by *stf* associated phage invertases that catalyze site-specific recombination [74]. The very low potentiality of virulent phages to acquire genes from temperate phages, which often carry virulence genes, is reassuring for the development of phage therapy as a means to combat bacterial infections. It should be noted however that some virulent phages are in fact former temperate phages that lost their lysogeny module. The case is well documented among dairy phages [75]. Such "virulent" phages, that should rather be named "ex-temperate", do exchange sequences with temperate and defective prophages [75] and should therefore be avoided for phage therapy, as already indicated [76]. To ascertain the choice of virulent phages for therapeutic use in the future, it may be relevant to conduct an analysis similar to the one presented here, by looking for mosaics (*i.e.* Blast matches above 90% identity) between the selected (and sequenced) phage, and all possible sequenced phages and prophages infecting the targeted bacterial species.

Consequences for evolution

The relaxed exchange of genetic information among phages might result from different selective pressures acting on viruses as compared to bacteria. The level of HR is indeed the result of contradictory demands. Recombination between homologous sequences must be an efficient process to repair rapidly DNA lesions. On the other hand, recombination between different homologous loci, by creating genomic rearrangement, can destroy

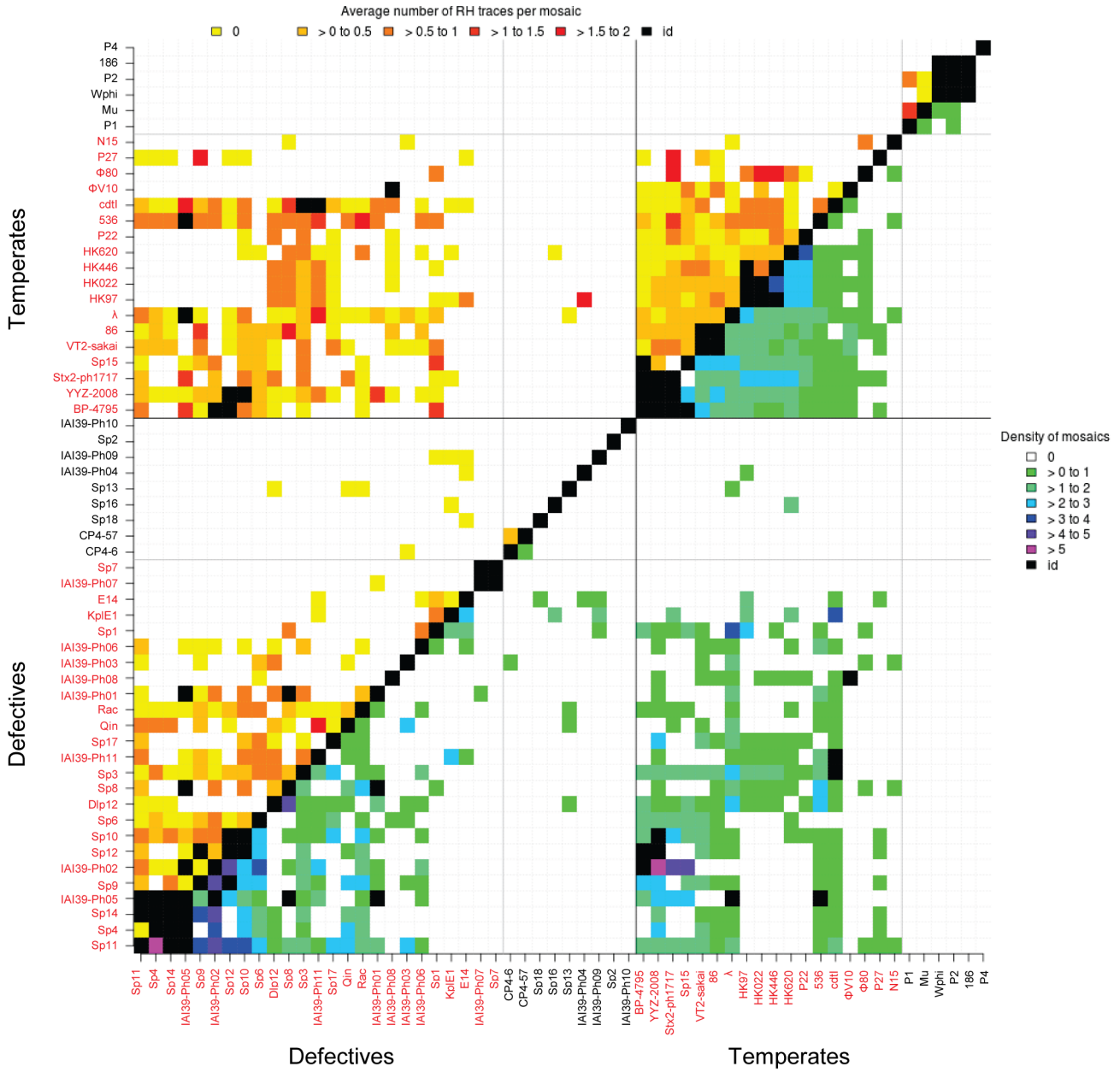


Figure 6. Heat map of mosaic characteristics in temperate and defective phages. Density of mosaics found among pairs of phages (lower part of the diagonal, purple to green colours), and average number of traces of recombination detected at the boundary of mosaics (higher part, yellow to red colours). The density is the number of mosaics per 10 kb of phage genome. Names of λ-like phages are in red.
doi:10.1371/journal.pgen.1004181.g006

chromosomal integrity and functional gene associations and must be avoided. This problem is especially pronounced in eukaryotes, which contain repeated sequences with slight polymorphism, that are prone to recombine together, but bacterial chromosomes also possess repeated elements and related genes. Phage genomes, however, do not possess repeated sequences usually, and can thus better tolerate recombination between diverged sequences without risking chromosomal rearrangements. This might explain the phage tolerance to low fidelity recombination, which then accidentally contribute to rapid phage evolution. Alternatively, as recombination of diverged sequences is sometime advantageous for phages, it is tempting to speculate that it is under positive selection. First, high levels of shuffling inside genes can generate new functions (reviewed in [77]). We found that small sequences of 100 bp within genes are frequently exchanged. This mechanism to generate new genes could explain the much larger viral gene repertoire as compared to bacterial gene repertoire. Secondly, creation of new gene combinations is likely valuable in numerous functions to allow greater phage propagation, e.g., escape from CRISPR-mediated bacterial immunity systems [78] or expansion of bacterial host range [72,79].

Temperate phage genes are alternatively submitted to very different selective pressures depending on the nature of their replicating cycle. Once integrated, prophage DNA becomes subject to the selective forces working on the bacterial chromosome. This explains that temperate phages constitute a reservoir of genes that improve bacterial phenotypes by lysogenisation [76,80]. Likewise, defective prophages, that were long considered to be

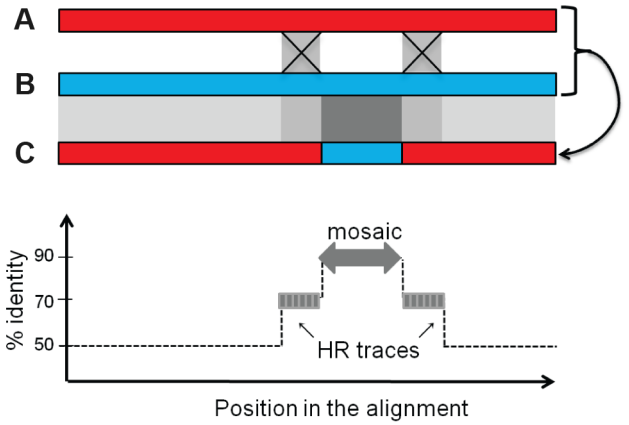


Figure 7. Rationale for the bio-informatics detection of mosaics. Upper panel: two ancestral phages A and B recombine across two regions of partial identity (light grey squares). As a consequence, the new piece of DNA in the recombined phage C, when compared to its parent B providing the mosaic, exhibits a 100% identity region (dark grey square), flanked by the two partially identical sequences (light grey), above the background level *b* of low identity shared by B and C. Upon alignment of genome C with B around the mosaic (lower panel), if the regions flanking the mosaic have a percentage of identity above the background level of identity ($\geq b$+10%), they will be counted as traces of homologous recombination (HR trace).
doi:10.1371/journal.pgen.1004181.g007

mere genetic junk, are now known to confer a broad range of beneficial phenotypes to the bacterial hosts, with respect to virulence, stress resistance, or even mutation rate [80–83]. The present study illustrates that inversely, prophage remnants can be a reservoir of functional lytic cycle genes for temperate phages. Dlp12 lysis module, helpful in certain strains of *E. coli* for biofilm development [43], is also active in λ for lysis. The constant exchange of genes between prophage remnants entrapped in host chromosome and active phages, *via* homologous recombination, blurs even more the distinction between evolutionary pressures acting on temperate phages and their bacterial host, tightly links their evolution, and indirectly accelerates bacterial evolution itself.

Materials and Methods

Bacterial and phage strains

All phage and bacterial strains are listed in Table S1. Unless specified, all gene replacements, deletions or insertions were done by recombineering as already described [84], with primers specified in Table S2.

The MG1655 *stfR*::*cat* (in Rac prophage) and *tfaQ*::*cat* (in Qin prophage) were constructed by inserting the *cat* cassette from the pKD3 plasmid into the respective genes with the *stfR*::*cat* and *tfaQ*::*cat* oligonucleotide pairs listed in Table S2, respectively. The MG1655 *ybcV*::*Kan* (in Dlp12 prophage) was constructed by P1 transduction from the Keio collection strain ECK0550 [85].

The Urλ*ble* phage was constructed from Urλ isolated from the *E. coli* K12 ancestral strain. The phleomycine/bleomycine resistance gene *ble*, cloned under the *Bacillus subtilis* promoter p*sacB*, taken from plasmid pUCphleo (gift from E. Dervyn), was introduced between the *tfa* and *ea47* genes of λ, and oriented as *tfa*. The presence of *ble* does not modify phage growth (Fig. S2), nor the frequency of lysogenization. Complete sequencing of Urλ*ble* showed that no important mutation other than the presence of the *ble* gene and the expected frameshift in the *stf* gene [40] differentiate our strain from the sequenced λPaPa strain (Table S2).

Deletions of *redβ*, *redα*, *orf* and *rap* genes were done by recombineering in *E. coli* K12 Urλ*ble*. In the case of the deletion of *redβ*, a RBS was introduced to maintain the expression of *redα*, as the two genes are overlapping and *redα* RBS is inside *redβ*.

Phage Φ80 was isolated in our laboratory from a strain contamination. Phage Φ80*ble* was constructed by introducing the p*sacB-ble* construct after *gp63*, and in the same orientation.

The λNec9 and 10 were constructed by replacing *redα* and *redβ* by the Rac-encoded *recE* and *recT* genes, in λNec4 and λNec6, respectively (phage strains listed in Table S1). To do this, the fragment corresponding to the C-terminal (from a.a. 588) of RecE and the full *recT* gene was added into pKD4 at the *Bmg*BI site, to give pJA100. The PCR fragment generated from pJA100, using the pair of oligonucleotides pKD4 (Table S3) was then introduced between the *gam* and *orf60a* genes of λNec4 and λNec6 by recombineering [84]. To construct λNec11 and λNec12, *redα* and *redβ* were replaced by *erf* and *exo* genes from D3 (a *P. aeruginosa* phage) in λNec4 and λNec6, respectively. First, these two genes were cloned into pKD4 (pJA82, constructed with erf/exo oligonucleotides described in Table S3) at the *Bmg*BI site, then the PCR fragment generated from this plasmid (using the oligonucleotides pKD4 described in Table S3) was introduced between the *gam* and *orf60a* genes of λNec4 and λNec6, following the same steps as for the *recET* constructions. Final constructs were verified by sequencing.

Measure of phage growth

Burst size was determined on *E. coli* MG1655 bacteria growing exponentially in LB broth supplemented with maltose (0.2% w/v) and magnesium (MgSO4 at a concentration of 10 mM). When OD_{600} reached 0.2, 10 ml of the culture were concentrated 10 times, and the phage added at a multiplicity of infection of 0.002. The mix was incubated for 7 minutes at 37°C and then diluted 100-fold and 10,000-fold in LB at 37°C. The number of non-adsorbed phage at the start of the phage growth was evaluated by plaque assay after bacterial centrifugation. Samples of the two dilution mixes were taken repeatedly throughout time and assayed immediately for plaque-forming units. The burst size is the factor between final phage number and initial phage number, subtracting the number of non-adsorbed phage at time of dilution.

Plaque size was determined after overnight growth on a layer of *E. coli* MG1655 bacteria embedded into top agarose supplemented with maltose and magnesium (2 g/L agarose, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 10 mM MgSO4, 2 g/L maltose). Plates imaging and analysis was realized with the Colony Doc-it imaging station (UVP, Upland, Canada) with the same settings for all plates.

Phage-prophage recombination assay

The phage to be tested was multiplied on the appropriate *E. coli* strains on plates. Briefly, 10^4 (for Red+ λ strains and Φ80 strains) or 10^6 (for Δ*redβ* λ strains) PFU were mixed to 100 µl of overnight bacterial culture grown on LB maltose. After 5 minutes of incubation, 4 ml of top agar containing MgSO4 10 mM were added and then poured on a fresh LBA plate containing 0.2% of glucose. After 6 to 8 hours of incubation at 37°C, when lysis was confluent, 5 ml of water were poured on the top of the plates and incubated at 4°C for two hours. The phage supernatant was then recovered and filtrated at 0.2 µm. Stocks were around 10^{10} PFU/ml for Red positive strains and 10^9 PFU/ml for Δ*redβ* or Δ*exo* strains. A theoretical calculation based on the measured burst sizes in liquid medium allowed to estimate that under such conditions

Redβ$^+$ and Redβ$^-$ phages performed similarly 2.8 generations during the recombination assay. Φ80 experiments were performed similarly but with few differences: no maltose was added to the medium, and the same amount of RecT+ and RecT− phages (10^4) were used for inoculation as no difference in plaque sizes was observed between the two genotypes.

Read-out of the phage-prophage recombination assay by lysogenization

1 ml of *E. coli* MG1655 culture growing in LB+0.2% maltose+ MgSO4 10 mM was added to the phage stock in which recombinants were produced (m.o.i = 1) when the OD_{600} of the culture reached 1. The mix was then incubated for 1.5 hours at 37°C in a closed 2 ml tube. The total number of lysogenized bacteria was measured by plating the bacteria at the appropriate dilution on phleomycine (5 mg/L) plates. Recombinant phage concentration was estimated by plating on either chloramphenicol (20 mg/L) or kanamycine (50 ml/L) plates. In these conditions, 5 to 10% of the bacteria were lysogenized, as indicated by the proportion of bacteria that acquired phleomycine resistance. Recombinant frequencies are the ratio of chloramphenicol or kanamycine resistant bacteria on phleomycine resistant bacteria. Figures given are an average of at least three independent recombination assays, followed for each of them by at least 3 read outs by lysogenisation.

Analyses of recombinant clones

PCR amplification of the λ genomic region susceptible of containing the resistance gene acquired from the host chromosome, if recombination was guided by homology, were realized on more than 20 colonies per genotype (oligonucleotide sequences are given in Table S3). For each colony tested, a fragment at the expected size was obtained. About 35% of the colonies gave an additional band corresponding to the native size of the λ region tested. A PCR using divergent oligonucleotides at *int* and *ea59* (that give a product if λ is excised or integrated in multiple copies) allowed confirming that these colonies were polylysogens. As the same proportion of polylysogens was found for each mutant, this phenomenon was neglected for the calculation of the frequency of recombinants.

P_L inversion assay on phage λ

2×10^6 λNec phages (100 µL) were incubated 5 min with 2×10^7 (500 µL) of exponentially growing C600 *recA* cells. Then 5 mL of top agar containing 10 mM MgSO4 was poured onto the mix and plated on LB plates, which were incubated 6 h at 37°C, until confluent lysis. 5 mL of a 10 mM MgSO4 solution were poured on the lysed plates, and the whole top agar layer was recovered, vortexed, and centrifuged for 10 min at 4°C. The supernatant, containing the phage, was then recovered and filtered at 0.2 µm. To estimate the amount of recombinants produced during lytic growth on plates, the stock was titrated on C600 *recA* strain to count parental phage, and on a C600 P2 lysogen strain, where only recombinant phage can grow, to count the recombinants. The frequency of recombination was then calculated by the number of recombinant PFU divided by the number of PFU corresponding to parental phages, counted on the *recA* strain.

Bioinformatics analysis of mosaicism

A set of 50 non-redundant genomes (less than 92% overall identity or coverage <80%) of phages infecting *E. coli*, and having either temperate (n = 24) or virulent (n = 26) lifestyles were chosen (see Table S4). The 34 non-redundant defective prophages were those of strains MG1655 and 0157:H7 Sakai [37] as well as the prophages identified in IAI39 *E. coli* strain (all contained Insertion Sequences (IS) in genes important for the phage life cycle, and were therefore classified as defective). The analysis was focused on recent exchanges only (mosaics with more than 90% identity), because it aimed at detecting traces of recombination in mosaic flanking sequences. Among the selected phages, some pairs were too closely related to contribute to the analysis: when more than 50% of the smaller genome of the pair shared more than 70% average nucleotide identity with its partner, the pair was excluded (black squares in Fig. 6, the strategy used to examine phage relatedness is shown Fig. S7). To detect recent exchanges, a Blastn was run on each genome pair, and all hits of a minimal size of 100 bp and sharing more than 90% identity were selected. Among these hits, some corresponded to IS sequences: they contributed for 9% of the total hits found among temperate phages, but as much as 33% among the defective phage pairs. The rest of the hits were named mosaics.

Under the hypothesis of mosaics formed by homologous recombination between diverged sequences, the scenario depicted in Fig. 7 is supposed to take place. Two ancestral phages A and B sharing in average 60% identity except for some more conserved regions at 80% identity, recombine across these 80% identical sequences (light grey). As a consequence, the new piece of DNA in the recombined phage C, when compared to its parent B providing the mosaic, exhibits a 100% identity region (dark grey), flanked by the two 80% identical sequences, above the background level of 60% identity shared by the two genomes at the time of exchange, and somewhat less at the time where genomes are analyzed. Following this scheme, to detect traces of homologous recombination at the vicinity of the mosaics, pairs of 2 kb-long sequences flanking the mosaics were realigned by dynamic programming, and the identity level of successive 50 bp-long windows along the alignment was measured. The 50 bp length was chosen as a close value to the minimum required for Redβ recombination (30 bp). The probability to encounter such HR traces at random (last line of Table 1) was calculated as described in Text S1.

Supporting Information

Figure S1 Distribution of perfect identity segment sizes in the different homology regions. Lengths of homology regions with λ (A) and Φ80 (B) are shown in blue and red for the regions placed left and right from the antibiotic cassette, respectively. Green bars indicate the segments in which a recombination event was detected. (C) Purple: Length distribution of homology regions for the *oxa5/oxa7* sequences (P_L inversion assay). Green: segments in which a recombination event was detected. (D) Numbers of MEPS (Minimal Efficiency Pairing Segment), defined as the minimal length of strict identity to initiate pairing, for Redβ and RecA in the different homology regions.
(DOCX)

Figure S2 Single burst growths of Urλ, Urλ*ble* and Urλ*ble* Δ*redβ* (λSOC27). The insertion of the phleomycin resistance gene (*ble*) does not modify either the burst size or the latency period of the phage. Deletion of the *redβ* gene reduces the burst size by 5 fold, but the latency period is unchanged. The experiment has been repeated four times independently, and a representative result is shown.
(DOCX)

Figure S3 Examination of Rac prophage replication. As the Rac prophage possesses an active origin of replication, *oriJ*, we investigated by semi-quantitative PCR if λ infection activates Rac replication. Cells infected or not with λ at an m.o.i. of 1, were

harvested 30 min after infection for total DNA purification. DNA was purified after phenol-chloroform extraction and ethanol precipitation steps. PCR was realized on 50 pg of total DNA, with 22 cycles of amplification. 1 to 2: PCR with primers framing *oriJ* does not indicate Rac replication following λ infection (line 2, MG+λ) compared to the non-infected MG1655 strain (line 1, MG). 3 to 4: internal control with primers within the *gyrB* gene. 5 & 6: PCR on the λ gene Q shows λ replication in the MG+ λ sample (lane 6).
(DOCX)

Figure S4 Plaque size distribution of λ mutants in the presence or absence of RecA in the host. To determine the effect of phage mutations on growth, we used the plaque size method. As all phage parameters like virion size, adsorption rate and latency period are equal in other respects, a reduction in burst size should directly translate into a smaller plaque of lysis on a lawn of bacteria. The same phage mutants were thus grown on MG1655 or isogenic *recA* lawns. *redβ and redα* mutants give a 2.5-fold smaller plaque area, but no difference in plaque size for *orf* and *rap* mutants is observed. On a *recA* lawn, a uniform and modest 15% decrease of plaque size was found for all mutants, probably reflecting the lower growth rate of the *recA* mutant. Interestingly, on a *recA* host, the *redβ* deleted phage makes plaques of a similar size compared to the WT host, whereas homologous recombination on phage DNA is abolished. This suggests that RecA, while being essential for HR when a λ *red* phage is used, is dispensable for growth of such λ *red* phage, and therefore that the Red function in growth is distinct from its role in homologous recombination.
(DOCX)

Figure S5 Distribution of mosaic sizes in the 4 categories of comparisons.
(DOCX)

Figure S6 Strategy to determine the phage genome pairs to be excluded from the mosaic analysis. Mosaic analysis was focused on recent exchanges (mosaics with more than 90% identity), because it aimed at detecting the relatedness of the mosaic flanking sequences. The genome pairs too closely related to permit detection of such flanking sequences needed to be eliminated. We reasoned that if the average nucleotide identity (ANI) of the genome pair was above 70%, no clear signal above the ANI value, and different from the mosaic signal would be detected. We also specified that at least 50% of the genomes needed to be included for the measurement of the ANI value, to be relevant. To measure the ANI of each genome pair, the method of Goris et al. was adapted [86]. Instead of using standard Blast parameters, with which ANI below 80% are seldom generated, we used the following values: cost to open a gap: 5, cost to extend a gap: 2, penalty for nucleotide mismatch : - 4, reward for nucleotide match : 3, word size : 10. This allowed reaching ANI values down to 60%. ANI values as a function of genome coverage for temperate and defective pairs (**A**), and for virulent pairs (**B**) are shown. The virulent pairs of genomes are dominated by T4-like phages, which group nicely into discrete sets of ANI values. In order to avoid cutting within a group, virulent pairs with ANI value above 68% and coverage above 48% were removed from the mosaic analysis (red frame). In a marked contrast, temperate/defective pairs form a cloud of points, among which the subset with ANI value above 70% and coverage value above 50% are the pairs removed from the analysis (red frame). ANI analysis was performed for pairs of genomes listed in TableS4 having mosaics. In the black frames of Figure S6, (>92% id, >80% cov), genomes are so similar that they were completely removed from the analysis (indicated with * as redundant in the last column of Table S4).
(DOCX)

Figure S7 Scheme for the masking of homologous recombination traces. Genome A/B is the product of a recombination between A and B in regions of homology. As described in Fig. 7, alignment of the two genomes should reveal traces of homologous recombination on both sides of the mosaic. However, recombination of the A/B hybrid with a third genome, C, can mask one of the two HR traces. Alignment of A/B/C hybrid with the ancestral B genome will reveal only one trace of homologous recombination.
(DOCX)

Table S1 Bacterial and phage strains used in this study.
(DOCX)

Table S2 Differences between the sequence of λ (λPapa) and the Urλ-*ble* phage used in this study.
(DOCX)

Table S3 Primers used in this study.
(DOCX)

Table S4 List of phage genomes used.
(DOCX)

Text S1 Probability to encounter traces of HR at random around mosaics.
(DOCX)

Acknowledgments

We are grateful to Drs Plunkett and Kuzminov for communicating the sequence of phage Φ80 prior to its publication. We also thank Michael Dubow and Magali Leroy for the sequencing of the phages used in this study. We are also indebted to Ludovic Lechat for the gift of Urλ*ble*, and to David Halpern for the constructions of the *orf* and *rap* deletions. We thank A. Gruss, P. Serror and M. Ansaldi for critical reading of the manuscript prior to submission. Our special thanks go to Colin Tinsley for his constant support along this work, and his analysis of the IAI39 defective prophages.

Author Contributions

Conceived and designed the experiments: MDP GH JAB MAP. Performed the experiments: MDP GH OS JAB MAP. Analyzed the data: MDP GH OS JAB SS MAP. Wrote the paper: MDP GH SS MAP.

References

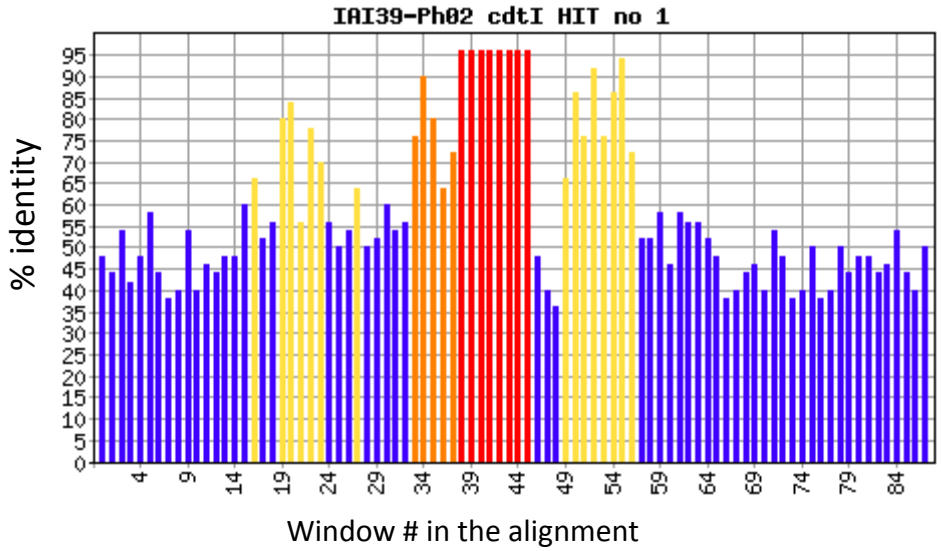
1. D'Herelle F (1917) On an invisible microbe antagonistic toward dysenteric bacilli. Compte Rendu de l'Académie des Sciences 165: 373–375.
2. Twort FW (1915) An Investigation on the Nature of ultra-microscopic Viruses. The Lancet 2: 1241–1243.
3. Kristensen DM, Waller AS, Yamada T, Bork P, Mushegian AR, et al. (2013) Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. J Bacteriol 195(5):941–50.
4. Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9: 267–276.
5. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84: 9733–9748.
6. Baker J, Limberger R, Schneider SJ, Campbell A (1991) Recombination and modular exchange in the genesis of new lambdoid phages. New Biol 3: 297–308.
7. van der Walt E, Rybicki EP, Varsani A, Polston JE, Billharz R, et al. (2009) Rapid host adaptation by extensive recombination. J Gen Virol 90: 734–746.
8. Muylkens B, Farnir F, Meurens F, Schynts F, Vanderplasschen A, et al. (2009) Coinfection with two closely related alphaherpesviruses results in a highly diversified recombination mosaic displaying negative genetic interference. J Virol 83: 3127–3137.
9. Hatfull GF (2008) Bacteriophage genomics. Curr Opin Microbiol 11: 447–453.
10. Niwa O, Yamagishi H, Ozeki H (1978) Sequence homology in DNA molecules of temperate phages phi81, phi80 and lambda. Mol Gen Genet 159: 259–268.

11. Hendrix RW (2003) Bacteriophage genomics. Curr Opin Microbiol 6: 506–511.
12. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H (2003) Prophage genomics. Microbiol Mol Biol Rev 67: 238–276, table of contents.
13. Botstein D (1980) A theory of modular evolution for bacteriophages. Ann N Y Acad Sci 354: 484–490.
14. Lawrence JG, Hatfull GF, Hendrix RW (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J Bacteriol 184: 4891–4905.
15. Brussow H, Desiere F (2001) Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol 39: 213–222.
16. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, et al. (2001) Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 8: 11–22.
17. Bobay LM, Rocha EP, Touchon M (2013) The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol 30: 737–751.
18. Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal Escherichia coli. Nat Rev Microbiol 8: 207–217.
19. Martinsohn JT, Radman M, Petit MA (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. PLoS Genet 4: e1000065.
20. Clark AJ, Inwood W, Cloutier T, Dhillon TS (2001) Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J Mol Biol 311: 657–679.
21. Hatfull GF, Hendrix RW (2011) Bacteriophages and their genomes. Curr Opin Virol 1: 298–303.
22. Murphy KC (2012) Phage recombinases and their applications. Adv Virus Res 83: 367–414.
23. Kuzminov A (1999) Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol Mol Biol Rev 63: 751–813, table of contents.
24. Shen P, Huang HV (1986) Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics 112: 441–457.
25. Majewski J, Cohan FM (1999) DNA sequence similarity requirements for interspecific recombination in Bacillus. Genetics 153: 1525–1533.
26. Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R (2010) Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res 38: 3952–3962.
27. Duboc H, Rajca S, Rainteau D, Benarous D, Maubert MA, et al. (2013) Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. Gut 62: 531–539.
28. Turpin W, Humblot C, Noordine ML, Wrzosek L, Tomas J, et al. Behavior of lactobacilli isolated from fermented slurry (ben-saalga) in gnotobiotic rats. PLoS One 8: e57711.
29. Li XT, Thomason LC, Sawitzke JA, Costantino N, Court DL (2013) Bacterial DNA polymerases participate in oligonucleotide recombination. Mol Microbiol 88: 906–920.
30. Muyrers JP, Zhang Y, Buchholz F, Stewart AF (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. Genes Dev 14: 1971–1982.
31. Hayes S, Asai K, Chu AM, Hayes C (2005) NinR- and red-mediated phage-prophage marker rescue recombination in Escherichia coli: recovery of a nonhomologous immlambda DNA segment by infecting lambdaimm434 phages. Genetics 170: 1485–1499.
32. Curtis FA, Reed P, Wilson LA, Bowers LY, Yeo RP, et al. (2011) The C-terminus of the phage lambda Orf recombinase is involved in DNA binding. J Mol Recognit 24: 333–340.
33. Sawitzke JA, Stahl FW (1997) Roles for lambda Orf and Escherichia coli RecO, RecR and RecF in lambda recombination. Genetics 147: 357–369.
34. Tarkowski TA, Mooney D, Thomason LC, Stahl FW (2002) Gene products encoded in the ninR region of phage lambda participate in Red-mediated recombination. Genes Cells 7: 351–363.
35. Sharples GJ, Curtis FA, McGlynn P, Bolt EL (2004) Holliday junction binding and resolution by the Rap structure-specific endonuclease of phage lambda. J Mol Biol 340: 739–751.
36. Poteete AR (2004) Modulation of DNA repair and recombination by the bacteriophage lambda Orf function in Escherichia coli K-12. J Bacteriol 186: 2699–2707.
37. Asadulghani M, Ogura Y, Ooka T, Itoh T, Sawaguchi A, et al. (2009) The defective prophage pool of Escherichia coli O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog 5: e1000408.
38. Ruzin A, Lindsay J, Novick RP (2001) Molecular genetics of SaPI1–a mobile pathogenicity island in Staphylococcus aureus. Mol Microbiol 41: 365–377.
39. Gallet R, Shao Y, Wang IN (2009) High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment. BMC Evol Biol 9: 241.
40. Hendrix RW, Duda RL (1992) Bacteriophage lambda PaPa: not the mother of all lambda phages. Science 258: 1145–1148.
41. Thaler DS, Stahl MM, Stahl FW (1987) Double-chain-cut sites are recombination hotspots in the Red pathway of phage lambda. J Mol Biol 195: 75–87.
42. Liu X, Jiang H, Gu Z, Roberts JW (2013) High-resolution view of bacteriophage lambda gene expression by ribosome profiling. Proc Natl Acad Sci U S A 110: 11928–11933.
43. Kolodner R, Hall SD, Luisi-DeLuca C (1994) Homologous pairing proteins encoded by the Escherichia coli recE and recT genes. Mol Microbiol 11: 23–30.
44. Mills S, Shanahan F, Stanton C, Hill C, Coffey A, et al. (2013) Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes 4: 4–16.
45. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98: 6742–6746.
46. Poteete AR, Fenton AC, Wang HR (2002) Recombination-promoting activity of the bacteriophage lambda Rap protein in Escherichia coli K-12. J Bacteriol 184: 4626–4629.
47. Hollifield WC, Kaplan EN, Huang HV (1987) Efficient RecABC-dependent, homologous recombination between coliphage lambda and plasmids requires a phage ninR region gene. Mol Gen Genet 210: 248–255.
48. Rotman E, Kouzminova E, Plunkett G, 3rd, Kuzminov A (2012) Genome of Enterobacteriophage Lula/phi80 and insights into its ability to spread in the laboratory environment. J Bacteriol 194: 6802–6817.
49. Iyer LM, Koonin EV, Aravind L (2002) Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics 3: 8.
50. Campbell A (1994) Comparative molecular biology of lambdoid phages. Annu Rev Microbiol 48: 193–222.
51. Bouchard JD, Moineau S (2000) Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. Virology 270: 65–75.
52. Durmaz E, Klaenhammer TR (2000) Genetic analysis of chromosomal regions of Lactococcus lactis acquired by recombinant lytic phages. Appl Environ Microbiol 66: 895–903.
53. Hall SD, Kolodner RD (1994) Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. Proc Natl Acad Sci U S A 91: 3205–3209.
54. Bobay LM, Touchon M, Rocha EP (2013) Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet 9: e1003825.
55. Sawitzke JA, Costantino N, Li XT, Thomason LC, Bubunenko M, et al. (2011) Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. J Mol Biol 407: 45–59.
56. Radman M, Wagner R (1993) DNA mismatch repair systems: mechanisms and applications in biotechnology. Biotechnol Genet Eng Rev 11: 357–366.
57. Elez M, Radman M, Matic I (2007) The frequency and structure of recombinant products is determined by the cellular level of MutL. Proc Natl Acad Sci U S A 104: 8935–8940.
58. Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci U S A 100: 15748–15753.
59. Poteete AR (2001) What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol Lett 201: 9–14.
60. Ploquin M, Bransi A, Paquet ER, Stasiak AZ, Stasiak A, et al. (2008) Functional and structural basis for a bacteriophage homolog of human RAD52. Curr Biol 18: 1142–1146.
61. Sokol H, Seksik P (2010) The intestinal microbiota in inflammatory bowel diseases: time to connect with the host. Curr Opin Gastroenterol 26: 327–331.
62. Sokol H, Vasquez N, Hoyeau-Idrissi N, Seksik P, Beaugerie L, et al. (2010) Crypt abscess-associated microbiota in inflammatory bowel disease and acute self-limited colitis. World J Gastroenterol 16: 583–587.
63. Mott C, Symington LS (2011) RAD51-independent inverted-repeat recombination by a strand-annealing mechanism. DNA Repair (Amst) 10: 408–415.
64. Wrzosek L, Miquel S, Noordine ML, Bouet S, Chevalier-Curt MJ, et al. (2013) Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. BMC Biol 11: 61.
65. Court DL, Sawitzke JA, Thomason LC (2002) Genetic engineering using homologous recombination. Annu Rev Genet 36: 361–388.
66. Poteete AR (2008) Involvement of DNA replication in phage lambda Red-mediated homologous recombination. Mol Microbiol 68: 66–74.
67. Maresca M, Erler A, Fu J, Friedrich A, Zhang Y, et al. (2010) Single-stranded heteroduplex intermediates in lambda Red homologous recombination. BMC Mol Biol 11: 54.
68. Mosberg JA, Lajoie MJ, Church GM (2010) Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics 186: 791–799.
69. Noirot P, Kolodner RD (1998) DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem 273: 12274–12280.
70. Rybalchenko N, Golub EI, Bi B, Radding CM (2004) Strand invasion promoted by recombination protein beta of coliphage lambda. Proc Natl Acad Sci U S A 101: 17056–17060.
71. Montag D, Schwarz H, Henning U (1989) A component of the side tail fiber of Escherichia coli bacteriophage lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. J Bacteriol 171: 4378–4384.

72. Haggard-Ljungquist E, Halling C, Calendar R (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J Bacteriol 174: 1462–1477.
73. Sandmeier H, Iida S, Arber W (1992) DNA inversion regions Min of plasmid p15B and Cin of bacteriophage P1: evolution of bacteriophage tail fiber genes. J Bacteriol 174: 3936–3944.
74. Sandmeier H (1994) Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibres. Mol Microbiol 12: 343–350.
75. Chopin A, Bolotin A, Sorokin A, Ehrlich SD, Chopin M (2001) Analysis of six prophages in Lactococcus lactis IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res 29: 644–651.
76. Miquel S, Martin R, Rossi O, Bermudez-Humaran LG, Chatel JM, et al. (2013) Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol 16: 255–261.
77. Long M, Betran E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4: 865–875.
78. Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated Adaptive Immune Systems in Bacteria and Archaea. Annu Rev Biochem 82:237–66.
79. Lwoff A (1953) Lysogeny. Bacteriol Rev 17: 269–337.
80. Brussow H, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68: 560–602, table of contents.
81. Panis G, Franche N, Mejean V, Ansaldi M (2012) Insights into the functions of a prophage recombination directionality factor. Viruses 4: 2417–2431.
82. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, et al. (2010) Cryptic prophages help bacteria cope with adverse environments. Nat Commun 1: 147.
83. Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA (2012) Prophage Excision Activates Listeria Competence Genes that Promote Phagosomal Escape and Virulence. Cell 150: 792–802.
84. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640–6645.
85. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2: 2006 0008.
86. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, et al. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57: 81–91.

Text S1: Probability to encounter traces of HR at random around mosaics :

At the vicinity of the mosaics, pairs of 2 kb-long sequences flanking the mosaics were aligned, and the identity level of successive 50 bp-long windows along the alignment was measured. An example of the identity profile flanking a mosaic is shown below. In red, the mosaic, in orange, the flanking region that could be a trace of HR, and in yellow some other region of higher-than-background similarity, that does not flank a mosaic. In blue, the background level of identity of the two regions compared.



From such identity profiles, ***b***, the background level of identity shared by the two genomes (taken from the third median value) is derived. Next, regions having a higher-than-background level of identity, namely above ***b***+10%, and placed next to the mosaic are classified as "trace of HR", whereas the other such regions that are distant from the mosaic are classified as "bumps". In the above example, the mosaic flanking region has 1 trace of HR and 4 bumps. For a given category of phages being compared, the total number of mosaics found is ***Nm***, the total number of traces of HR is ***Nth***, and the total number of bumps is ***Nb***. To estimate whether ***Nth*** is significantly above ***Nr***, the number of bumps expected at random just next to the mosaic, the following calculation is made: we model the occurrences of bumps along the concatenated realigned sequences by a Poisson process with intensity ***i*** estimated by the number of bumps ***Nb*** divided by the cumulative length of all sequences realigned. The number ***Nr*** of bumps expected at random in the 100 bp next to all mosaics of the group is then distributed according to a Poisson distribution with mean ***Nm*** x 100 x ***i***. It implies that the probability to observe as much as ***Nth*** traces of HR is given by the tail of this

Poisson distribution and found to be in the 10^{-14}- 10^{-16} range for all groups of phages tested (values reported in Table 1 of the main text). We conclude therefore that the number of traces of HR found next to the mosaics is highly significant. As a control, the same study was done on the hits corresponding to IS. 71 among 533 hits analysed contained a trace of homologous recombination at their boundary, for a random number expected ***Nr*** of 46. In this case, a p-value of 0.025, at the limit of significativity, was obtained, as expected for IS that should not exchange by HR but by transposition (unless degenerated).

Table S1: bacterial and phage strains used in this study.

Strain	Genotype	Commentary
MAC1117	MG1655 *ilvD:cat*	gift from S. Delmas (INSERM U1001, Paris, France)
MAC1400	MG1655 *stfR::cat*	this study
MD7	MG1655 *stfR::cat* recA306 *srl::Tn10*	this study
MD19	MG1655 *tfaQ::cat*	this study
MD63	MG1655 *tfaQ::cat recA306 srl::Tn10*	this study
MD36	MG1655 *ybcV::Kan*	this study. Allele from Keio collection
MD61	ybcV::Kan *recA306 srl::Tn10*	this study
MD22	MG1655 *yecD::cat*	this study
MD58	MG1655 *nohD::cat*	this study
MD68	MG1655 *ybcN::cat*	this study
MAC	C600 P2 lysogen	[19]
MAC	C600 recA306 *srl::Tn10*	[19]
	Phage strains	
λMD 50	λPaPa	reference laboratory strain of phage λ . Gift from D. Refardt (Refardt & Rainey, 2010)[e]
λMD51	λPaPa *ble*	this study
Urλ	Urλ	Original isolates of λ. From the *E. coli* K12 strain of the *E. Coli* Genetic Stock Center.
Urλ *ble*	Urλ *ble*	Urλ strain carrying the *ble* gene that confers resistance to phleomycine. gift from L. Le Chat
λSOC 27	Urλ *ble redβ::FRT*	this study
λSOC 28	Urλ *ble exo::FRT*	this study
λSOC 29	Urλ *ble orf::FRT*	this study
λSOC 30	Urλ *ble rap::FRT*	this study
Φ80 MD2	Φ80 *ble*	this study
λNec4	[800bp IR, 100% id][a] cI857ts bΔ(22297-27728)	[19]
λNec1	[invPL, 800bp IR, 100% id][b]cI857ts bΔ(22297-27728)	[19]
λNec6	[800 bp IR, 78%id][c] cI857ts bΔ(22050-26766)	[19]
λNec3	[invPL, 800 bp IR, 78%id][d] cI857ts Δ(22050-26766)	[19]
λNec9	[800bp IR, 100% id] cI857ts Δ(22297-27728) Δ*red::recET*	this study
λNec10	[800 bp IR, 78%id] cI857ts Δ(22050-26766) Δ*red:: recET*	this study
λNec11	[800bp IR, 100% id] cI857ts Δ(22297-27728) Δ*red*::*erf-exo*	this study
λNec12	[800 bp IR, 78%id] cI857ts Δ(22050-26766) Δ*red:: erf-exo*	this study

a. [800bp IR, 100% id] : recombination cassette composed of 800 bp-long inverted and 100% identical repeats flanking the pL promoter in its native orientation. The genotype of this region is ea10:oxa7 rexA:(oxa7-cat-χ+) Δorf28-ral

b. [invPL, 800bp IR, 100% id]: same as a., but inverted orientation of the sequences inside the IR, including PL.

c. [800bp IR, 78% id] : recombination cassette composed of 800 bp-long inverted and 78% identical repeats flanking the pL promoter in its native orientation. The genotype of this region is ea10:oxa7-5 rexA:(oxa5-7-cat-χ+) Δorf28-ral

d. [invPL, 800bp IR, 78% id] : same as c, but inverted orientation of the sequences inside the IR, including PL. The genotype of this region is ea10:oxa7 rexA:(oxa5-cat-χ+) Δorf28-ral

e. Refardt D, Rainey PB (2010) Tuning a genetic switch: experimental evolution and natural variation of prophage induction. *Evolution* **64:** 1086-1097

Table S2: Differences between the published sequence of λ and the Urλ-*ble* phage used in this study.

λPaPa NC_001416	**Ur-λ*ble***	**orf**	**Change***
138	- G	IR *cos-nu1*	- 1
542	C -> T	*nu1*	
after 14267	+ G	IR *L-K*	+ 1
17183	A -> G	*J*	E_{560} -> G
18835	T -> G	*J*	Y_{1111}-> D
20661	A -> G	*stf*	
after 20835	+ C	*stf*	+ 1
21714	G -> A	*stf*	S_{689} -> N
22445	A -> G	*tfa*	K_{158} -> E
22597-22633	***deletion -> psacB-phleoR***		
25662	T -> G	*ea59*	M_{438} -> L
31016	T -> C	*orf61*	
34934	A -> G	*sieB*	
45618	T -> C	*R*	

IR: intergenic region

*: Absence of change indicates a synonymous mutation. In addition to the expected C addition in *stf*, that restores the reading frame and therefore intact side tail fibers in Urλ, one deletion and two insertions are observed in intergenic regions (IR). Among the 8 substitutions, 3 are neutral, and 4 of the 5 non-synonymous changes are in tail component genes.

Table S3: primers used in this study [a]

Gene deletions or insertions		
stfR::cat	CGCCGTCACATGCCGGGACCATTACCGTGTATGAAGATTCCAACTTTTGGCGAAAATGAG	
	GTTGATGCCGTTCCTGCGCTGGAAGACGCTGACTGAGCCGACGGGAACTTCATTTAAATGGCG	
tfaQ::cat	GTTATTTATTTCTGAACTCGGTCCGTTACCGGAAAATGTTACCTGTCCTCCTTAGTTCCTATTCC	
	TATCCTTCACCCAGGCTGTGCCGTTCCACTTCTGAAACTCCCCTTCTGCCACTCATCGCAGTAC	
redβ::FRT	GATCCCGGTACGCTGCAGGATAATGTCCGGTGTCATGCTGTCCTCCTTAGTTCCTATTCC	
	CATTGCTCACCACCAGGTTGATATTGATTCAGAGGTATAAAACGAGTGTAGGCTGGAGCTGCTTC	
exo::FRT	CTTGGATTCCTGAAACAGAAAGCCGCAGAGCAGAAGGTGGCAGCATGAACACAGGAACACTTAACGGC	
	CCTACCCGGATATTATCGTGAGGATGCGTCATCGCCATTGCTCCCCTGTAGGCTGGAGCTGCTTCG	
orf::FRT	GGGAGAGGGAAGTCATGAAAAAACTAACCTTTGAAATTCGGCAGCATTACACGTCTTGAG	
	CATGCAGCCCTGTCTCCCCATCTCGCTTTCCACTCCAGAGCCCATGGTCCATATGAATATCCTCC	
rap::FRT	GATGATGGCTAAACCAGCGCGAAGACGATGTAAAAACGATGGCAGCATTACACGTCTTGAG	
	GTCATGCGGCCTCACTTCTGCTATTTCGCAGGTCTTTGAGCCATGGTCCATATGAATATCCTCC	
Φ80-*ble*	CAACAAAGAGATAAATCCAAGCCCGTTTTGTACGGGCTGTATGACCATGATTACGCCAAG	
	ATGCAGGTAACGCAGCCAGAATTGTTGATTGGCTGATCGTTAAGTTGGGTAACGCCAGG	
yecD::cat	CTTCCACGCTACGCACACGGGCGATGCGCGGGTAGATATGGTTGACAACTTTTGGCGAAAATGAG	
	AAGACGCCTGTAGTGCCGCCAGCGCCGAGCAGCACAATAACAGTAGGAACTTCATTTAAATGGCG	
nohD::cat	ATTGAGTACGAACGCCATCGACTTACGCGTGCGCAGGCCGATGCACAGGACAACTTTTGGCGAAAATGAG	
	CACCTGCGATCCGCGACAGCACGAAAGTACAGAATGCGGTTTCCACCACTGGAACTTCATTTAAATGGCG	
ybcN::cat	CTCCAGCTTGGGTAAAAGATGCTCTCAAACACACATATCTCGGTTATGAAGGAACTTCATTTAAATGGCG	
	TCGGAGGTATGGCGTAACGACTGGATAGTGGTGATATCACCGGTTACGACACAACTTTTGGCGAAAATGAG	
Detection of excision of defective prophages		
Rac	CTCCAGCATGGTATAGCTGTCTTTAC	described in [82]
	CAGATTTCTTATGCTGGGCGTTCCG	
Qin	CGACAATACGCGCCACATAA	
	AACGGCGAGTAAGTAGTACGC	
Dlp12	CAAAAGCCATTGACTCAGCAAGG	
	ACGGATAAGACGGGCATAAATGA	
Activity of oriJ (rac), semi-qPCR		
gyrB	GTCGAAGTGGCGTTGCAGTG	described in (Fogg et al, 2010) [b]
	AGCCTGCCAGGTGAGTACCG	
λ *Q*	GAGTGCGGAAGATGCAAAGG	
	TTAACAGTGCGTGACCAGG	
oriJ	GTTGAGTCAGGCGAAATCC	
	CCACGGTAACCACAATCAC	
Identification of recombination junctions in λ hybrids		
Qin	GCTCATGCCCATTCTTTC	
	ACGTTTCTGCGGCATAT	
Rac	ACTGGCGTACTGACGGATTC (rac 1)	
	ACGTTTCTGCGGCATAT (rac 2)	
	ACGTTTCTGCGGCATATC (rac 3)	
Dlp12	TTAGCATCCGCCCATTCAAC	
	TGGTCAGTTCGAGCATAAGG	
Construction of *recET* and *exo/erf* λ		
recE	CGAGCTCCCGCTCGAGCTTTTGCGATTGGTGGTTGC	
	GGAGGATGAAACTGGCGAAGTCGCA	
recT	CGAGCTCCATTACACCGCCAGGCTGAA	
	CCGCTCGAGGGAGGCTGGGCTAAGGAATATGCAA	
exo/erf	GTTTGCCGCATGTGGCTTTG	
	GGAGGCCATCAACAACGAAGCCC TG	
pKD4	TGATTGCGCCTACCCGGATATTATCGTGAGGATGCGTCATTCCTCCTTAGTTCCTATTCC	
	CATTGCTCACCACCAGGTTGATATTGATTCAGAGGTATAATGCAGATTGCAGCAT TACAC	

a. Unless specified, primers were designed for this study.

b. Fogg PC, Allison HE, Saunders JR, McCarthy AJ (2010). *J Virol* **84:** 6876-6879

Table S4: List of phage genomes used.

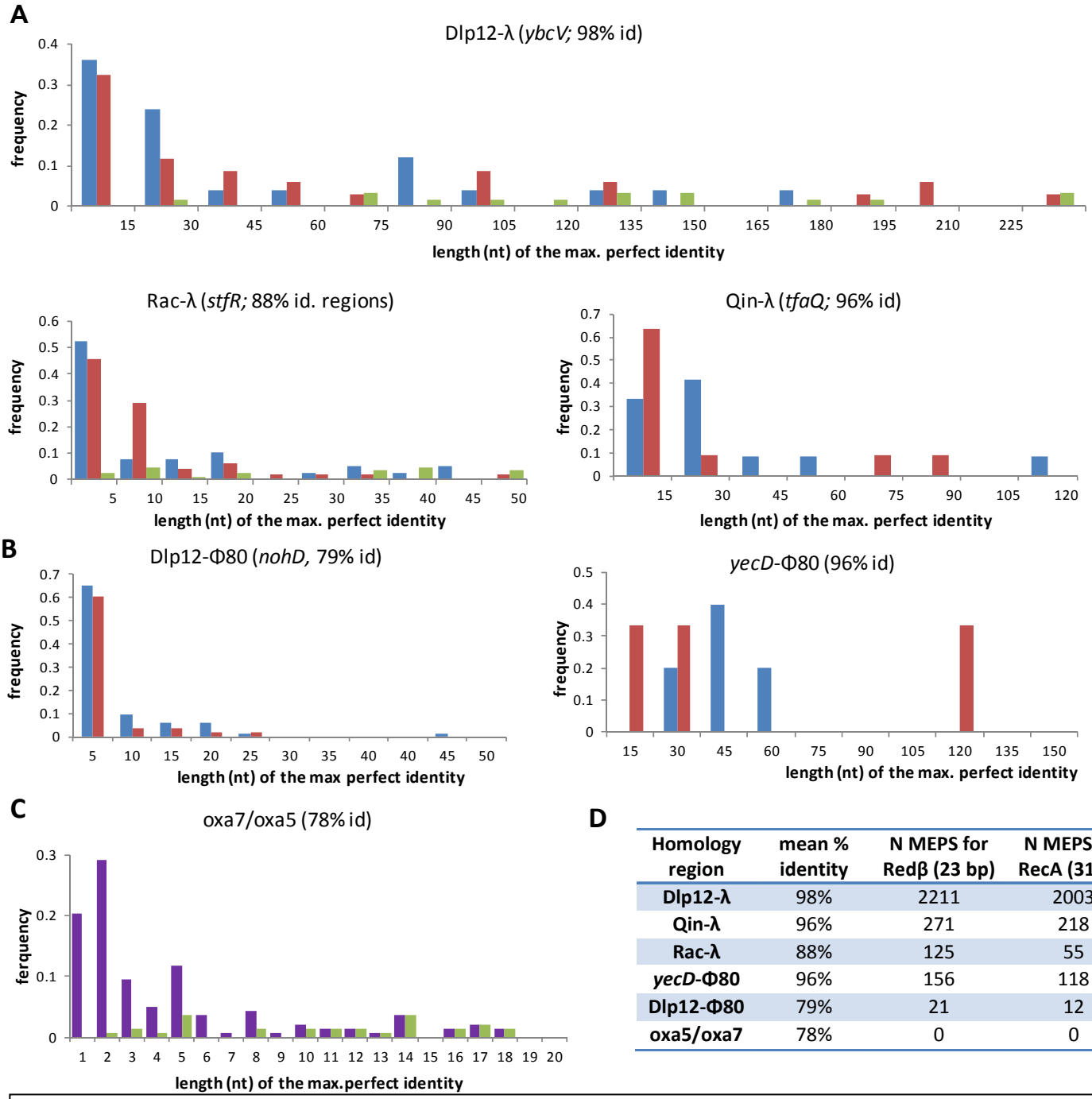
Phage	length (bp)	life style[a]	family	accession number, *redundancy with
186	30624	T	P2	NC_001317
536	46023	T	λ	NC_008253(1189882..1228488)
86	60238	T	λ	NC_008464
933W	61670	T	λ	NC_000924 , *VT2Sakai
BP-4795	57930	T	λ	NC_004813
Bp7	168066	V	T4	NC_019500
CC31	165540	V	T4	NC_014662
cdtI	47021	T	λ	NC_009514
CP4-57	22030	D	P2	NC_000913.2(2753978..2776007)
CP4-6	34306	D	fragments	NC_000913.2(262184..296489)
dlp12	21302	D	λ	NC_000913.2(564025..585326)
e14	15193	D	λ	NC_000913.2(1195443..1210646)
HK022	40751	T	λ	NC_002166
HK446	39026	T	λ	NC_019714
HK620	38297	T	λ	NC_002730
HK97	39732	T	λ	NC_002167
HX01	169158	V	T4	NC_018855, *RB69
IAI39-Ph01	32853	D	λ	NC_011750.1(532595..565447)
IAI39-Ph02	12130	D	λ	NC_011750.1(923080..935209)
IAI39-Ph03	28321	D	λ	NC_011750.1(1555796..1584116)
IAI39-Ph04	18544	D	P2	NC_011750.1(1882552..1901095)
IAI39-Ph05	23629	D	λ	NC_011750.1(1932603..1956231)
IAI39-Ph06	55581	D	unknown	NC_011750.1(1992456..2048036)
IAI39-Ph07	13835	D	λ	NC_011750.1(2060694..2074528)
IAI39-Ph08	41562	D	ε15	NC_011750.1(2751408..2792969)
IAI39-Ph09	11389	D	unknown	NC_011750.1(2934628..2946016)
IAI39-Ph10	14803	D	Mu	NC_011750.1(4625205..4640007)
IAI39-Ph11	44948	D	λ	NC_011750.1(5025091..5097038)
IME08	172253	V	T4	NC_014260
jk06	46072	V	T1	NC_007291
JS10	171451	V	T4	NC_012741, *JS98
JS98	170523	V	T4	NC_010105
JSE	166418	V	T4	NC_012740, *RB49
K1-5	44385	V	Sp6	NC_008152
K1E	45251	V	Sp6	NC_007637
K1F	39704	V	T7	NC_007456
KP15	174436	V	T4	NC_014036
KplE1	9687	D	fragments	NC_000913.2(2464404..2474619)
lambda	48502	T	λ	NC_001416
Mu	36717	T	Mu	NC_000929
N15	46375	T	N15	NC_001901
N4	70153	V	N4	NC_008720
NJ01	77448	V	C3	NC_018835
P1	94800	T	P1	NC_005856
P2	33593	T	P2	NC_001895
P22	41724	T	P22	NC_002371
P27	42575	T	unknown	NC_003356
P4	11624	T	P2	NC_001609

Phi1	164270	V	T4	NC_009821
phi80	46150	T	λ	JX871397
phiV10	39104	T	ε15	NC_007804
Qin	20458	D	λ	NC_000913.2(1630450..1646830)
rac	23060	D	λ	NC_000913.2(1409966..1433025)
RB14	165429	V	T4	NC_012638, *T4
RB15	166877	V	T4	phage.ggc.edu/gbrowse/RB15/RB15.fa, *T4
RB16	176788	V	T4	NC_014467
RB18	166677	V	T4	phage.ggc.edu/gbrowse/RB18/RB18.fa, *T4
RB26	163036	V	T4	phage.ggc.edu/gbrowse/RB26/RB26.fa, *T4
RB32	165890	V	T4	NC_008515, *T4
RB43	180500	V	T4	NC_007023
RB49	164018	V	T4	NC_005066
RB51	168394	V	T4	NC_012635, *T4
RB69	167560	V	T4	NC_004928
RTP	46219	V	T1	NC_007603
rv5	137947	V	RV5	NC_011041
sp1	10586	D	unknown	NC_002695.1(300041..310626)
sp10	51112	D	λ	NC_002695.1(1921414..1972525)
sp11	45778	D	λ	NC_002695.1(2158174..2203951)
sp12	46142	D	λ	NC_002695.1(2203952..2250093)
sp13	21120	D	P2	NC_002695.1(2592901..2614020)
sp14	44029	D	λ	NC_002695.1(2668007..2712035)
sp15	47879	T	λ	NC_002695.1(2895926..2943804)
sp16	8551	D	unknown	NC_002695.1(3192983..3201533)
sp17	24199	D	unknown	NC_002695.1(3475965..3500163)
sp18	38759	D	Mu	NC_002695.1(5040843..5079601)
sp2	12887	D	P2	NC_002695.1(310627..323513)
sp3	38586	D	λ	NC_002695.1(891123..929708)
sp4	49650	D	λ	NC_002695.1(1161091..1210740)
sp5	62708	T	λ	NC_002695.1(1246012..1308719), *VT2Sakai
sp6	48423	D	λ	NC_002695.1(1541470..1589892)
sp7	15463	D	unknown	NC_002695.1(1594570..1610032)
sp8	46897	D	λ	NC_002695.1(1618153..1665049)
sp9	58175	D	λ	NC_002695.1(1757506..1815680)
SPN3US	240413	V	unknown	JN641803
stx2-phage1717	62147	T	λ	NC_011357
stx2-phageI	61765	T	λ	NC_003525, *VT2Sakai
T1	48836	V	T1	NC_005833
T2	163833	V	T4	http://phage.ggc.edu/gbrowse/T2/T2.fa, *T4
T3	38208	V	T7	NC_003298
T4	168903	V	T4	NC_000866
T4T	168920	V	T4	http://phage.ggc.edu/gbrowse/T4T/T4T.fa
T5	121750	V	T5	NC_005859, *T4
T6	168705	V	T4	http://phage.ggc.edu/gbrowse/T6/T6.fa, *T4
T7	39937	V	T7	NC_001604
TLS	49902	V	T1	NC_009540
vB-EcoM-ACG-C40	167396	V	T4	NC_019399, *T4
vB-EcoM-VR7	169285	V	T4	NC_014792
VT2-Sakai	60942	T	λ	NC_000902
Wphi	32684	T	P2	NC_005056

wV7	166452	V	T4	NC_019505, *T4
YYZ-2008	54896	T	λ	NC_011356

a. D, defective; T, temperate; V, virulent.

FIGURE S1



Homology region	mean % identity	N MEPS for Redβ (23 bp)	N MEPS for RecA (31 bp)
Dlp12-λ	98%	2211	2003
Qin-λ	96%	271	218
Rac-λ	88%	125	55
***yecD*-Φ80**	96%	156	118
Dlp12-Φ80	79%	21	12
oxa5/oxa7	78%	0	0

Figure S1 Distribution of perfect identity segment sizes in the different homology regions. Lengths of homology regions with l (A) and W80 (B) are shown in blue and red for the regions placed left and right from the antibiotic cassette, respectively. Green bars indicate the segments in which a recombination event was detected. (C) Purple: Length distribution of homology regions for the oxa5/oxa7 sequences (PL inversion assay). Green: segments in which a recombination event was detected. (D) Numbers of MEPS (Minimal Efficiency Pairing Segment), defined as the minimal length of strict identity to initiate pairing, for Redb and RecA in the different homology regions.

FIGURE S2

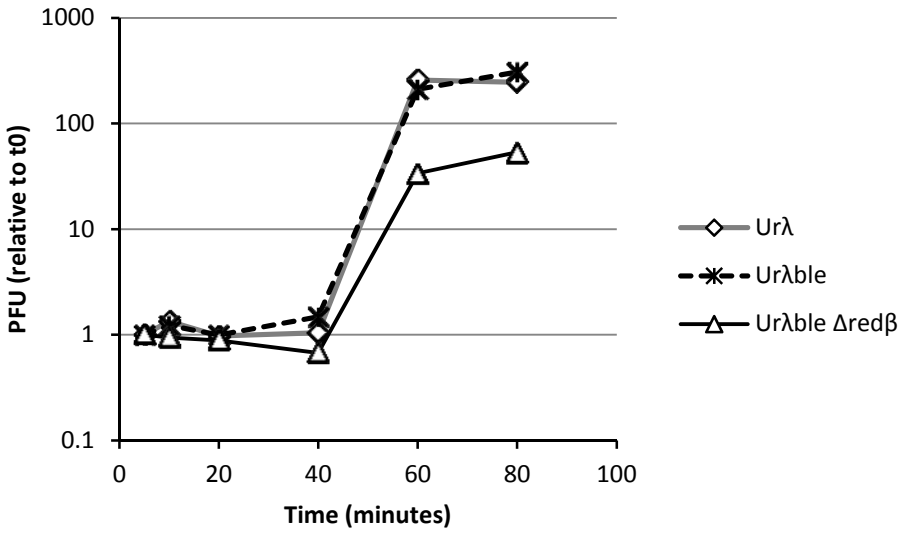
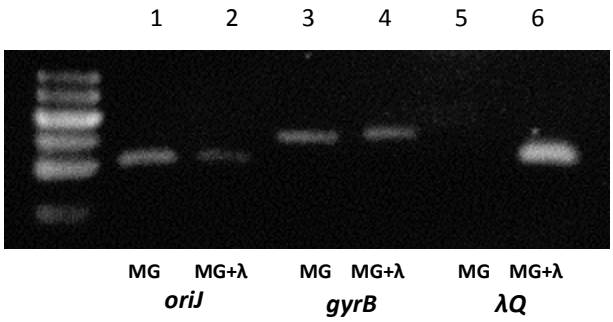


Figure S2 Single burst growths of Url,Urlble and Urlble Dredb (lSOC27). The insertion of the phleomycin resistance gene (ble) does not modify either the burst size or the latency period of the phage. Deletion of the redb gene reduces the burst size by 5 fold, but the latency period is unchanged. The experiment has been repeated four times independently, and a representative result is shown.

FIGURE S3



MG: MG1655 *E. coli* strain

MG+λ : MG1655 infected by λ

Figure S3 Examination of Rac prophage replication. As the Rac prophage possesses an active origin of replication, oriJ,we investigated by semi-quantitative PCR if l infection activates Rac replication. Cells infected or not with l at an m.o.i. of 1, were harvested 30 min after infection for total DNA purification. DNA was purified after phenol-chloroform extraction and ethanol precipitation steps. PCR was realized on 50 pg of total DNA, with 22 cycles of amplification. 1 to 2: PCR with primers framing oriJ does not indicate Rac replication following l infection (line 2, MG+l) compared to the non-infected MG1655 strain (line 1,MG). 3 to 4: internal controlwith primerswithin the gyrB gene.5&6:PCR on the l gene Q shows l replication in the MG+ l sample (lane 6).

FIGURE S4

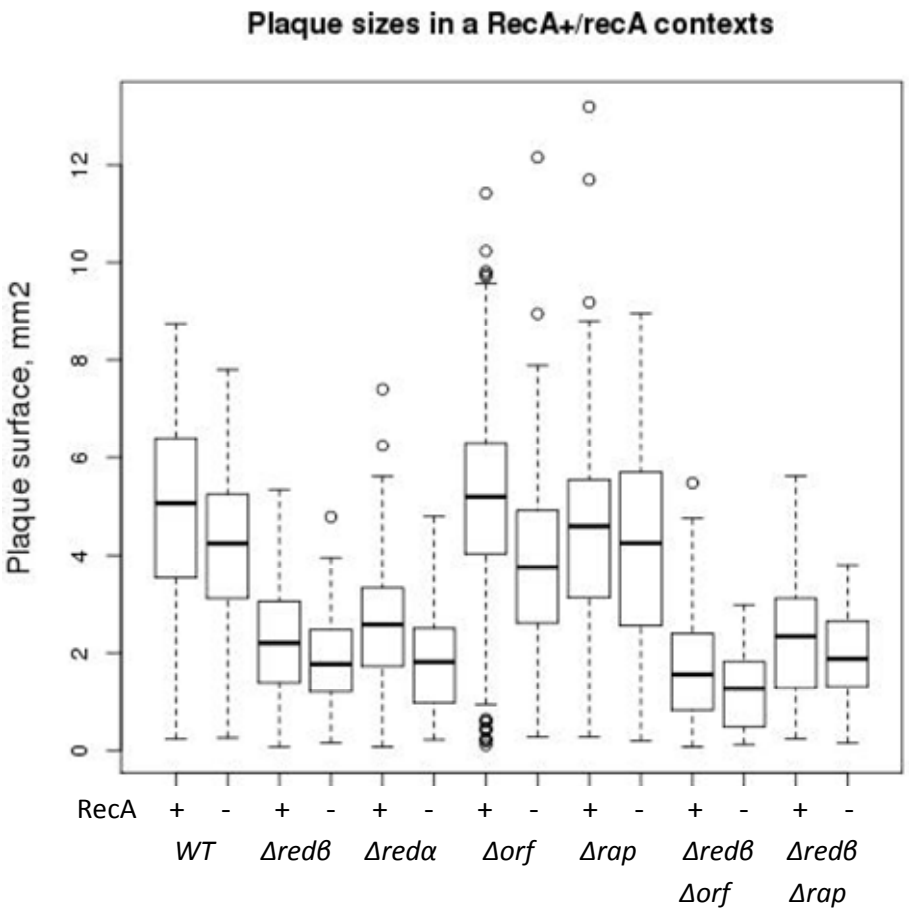
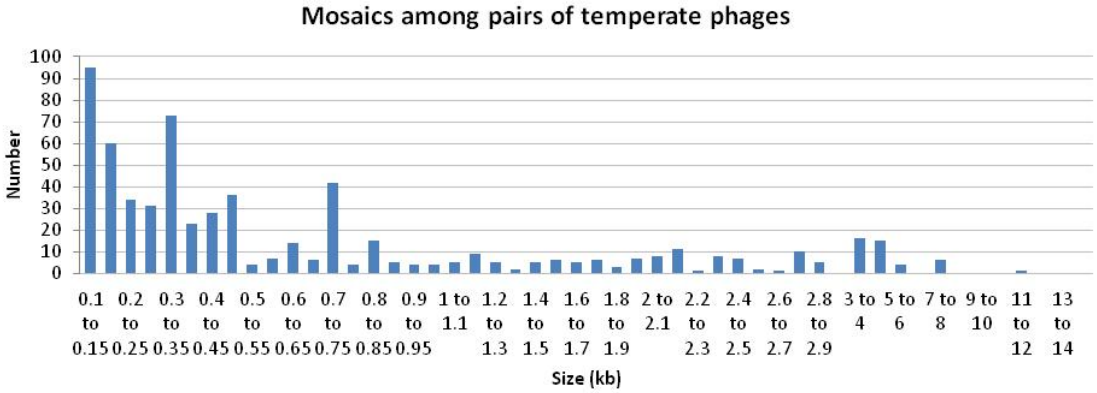
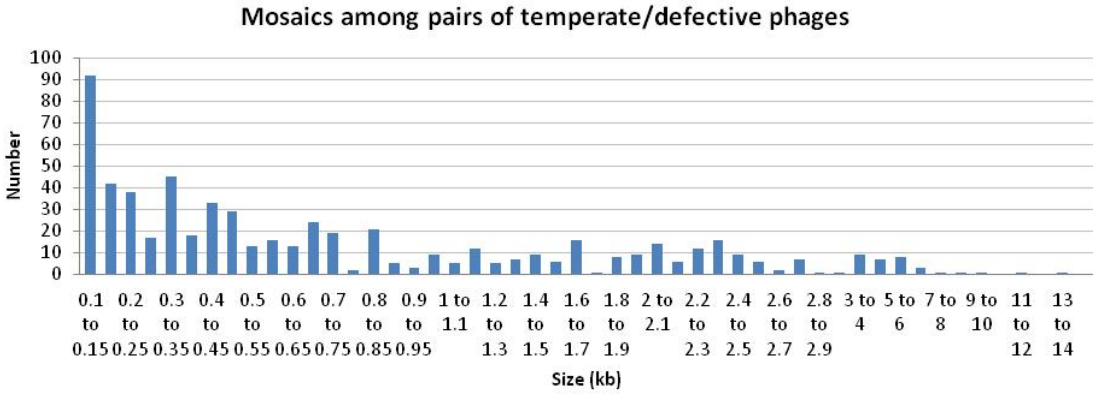
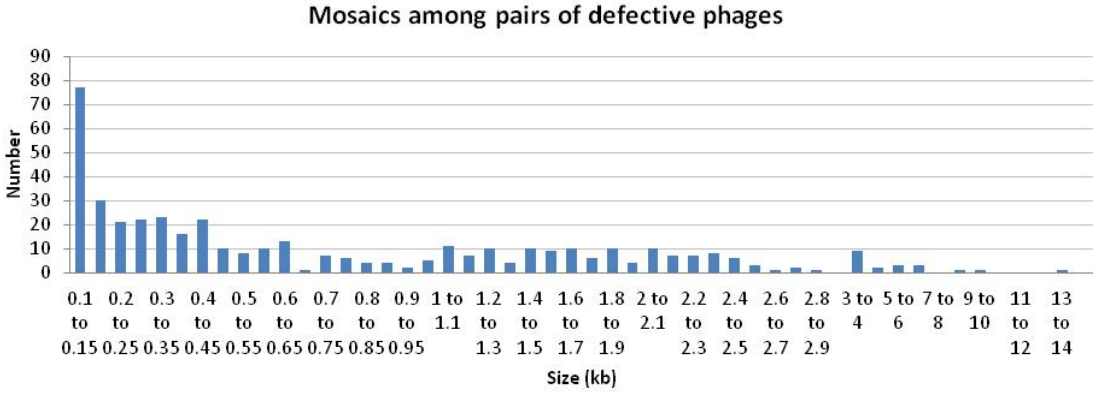


Figure S4 Plaque size distribution of l mutants in the presence or absence of RecA in the host. To determine the effect of phage mutations on growth, we used the plaque size method. As all phage parameters like virion size, adsorption rate and latency period are equal in other respects, a reduction in burst size should directly translate into a smaller plaque of lysis on a lawn of bacteria. The same phage mutants were thus grown on MG1655 or isogenic recA lawns. redb and reda mutants give a 2.5-fold smaller plaque area, but no difference in plaque size for orf and rap mutants is observed. On a recA lawn, a uniform and modest 15% decrease of plaque size was found for all mutants, probably reflecting the lower growth rate of the recA mutant. Interestingly, on a recA host, the redb deleted phage makes plaques of a similar size compared to the WT host, whereas homologous recombina- tion on phage DNA is abolished. This suggests that RecA, while being essential for HR when a l red phage is used, is dispensable for growth of such l red phage, and therefore that the Red function in growth is distinct from its role in homologous recombination.







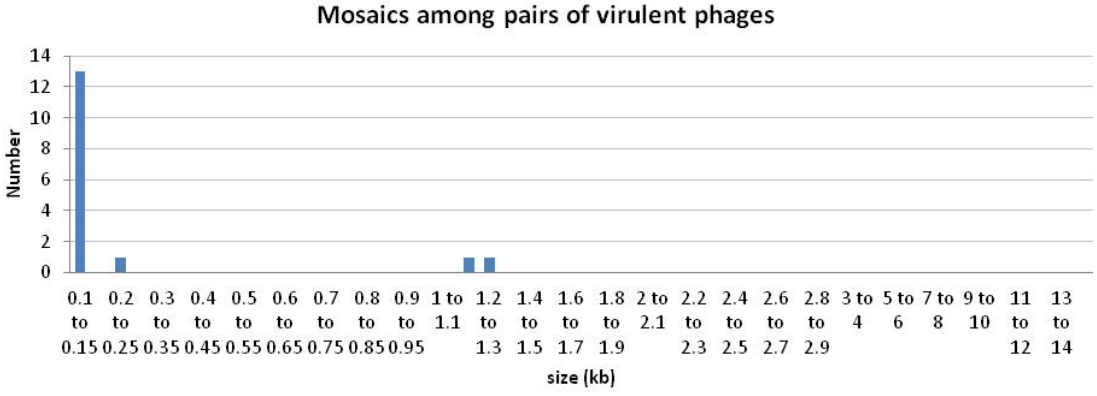


Figure S5: Distribution of mosaic sizes in the 4 categories of comparisons.

FIGURE S6

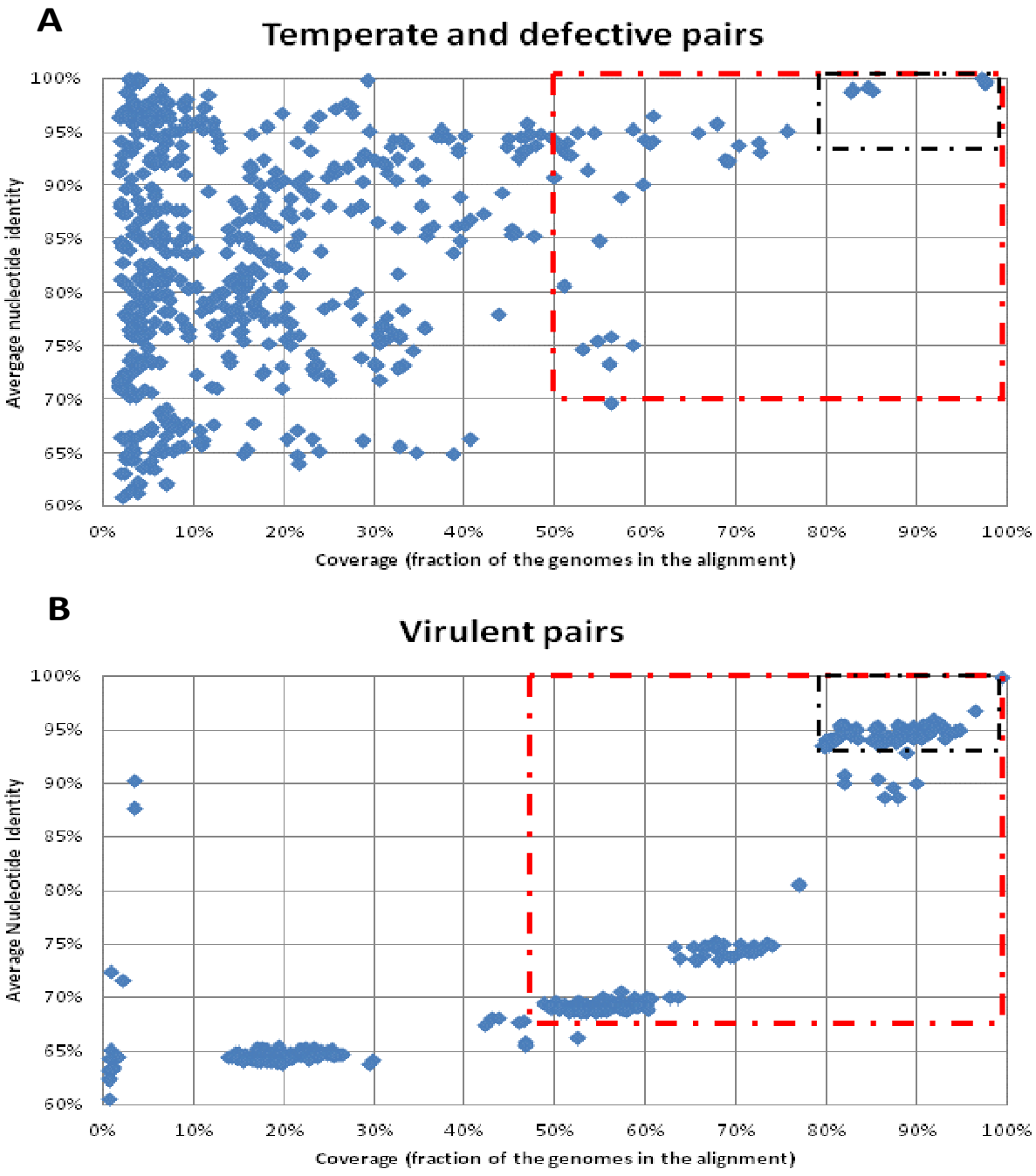

A
Temperate and defective pairs
Avergage nucleotide identity
100%
95%
90%
85%
80%
75%
70%
65%
60%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Coverage (fraction of the genomes in the alignment)
B
Virulent pairs
Average Nucleotide Identity
100%
95%
90%
85%
80%
75%
70%
65%
60%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Coverage (fraction of the genomes in the alignment)

Figure S6 Strategy to determine the phage genome pairs to be excluded from the mosaic analysis. Mosaic analysis was focused on recent exchanges (mosaics with more than 90% identity), because it aimed at detecting the relatedness of the mosaic flanking sequences. The genome pairs too closely related to permit detection of such flanking sequences needed to be eliminated. We reasoned that if the average nucleotide identity (ANI) of the genome pair was above 70%, no clear signal above the ANI value, and different from the mosaic signal would be detected. We also specified that at least 50% of the genomes needed to be included for the measurement of the ANI value, to be relevant. To measure the ANI of each genome pair, the method of Goris et al. was adapted [86]. Instead of using standard Blast parameters, with which ANI below 80% are seldom generated, we used the following values: cost to open a gap: 5, cost to extend a gap: 2, penalty for nucleotide mismatch : - 4, reward for nucleotide match : 3, word size : 10. This allowed reaching ANI values down to 60%. ANI values as a function of genome coverage for temperate and defective pairs (A), and for virulent pairs (B) are shown. The virulent pairs of genomes are dominated by T4-like phages, which group nicely into discrete sets of ANI values. In order to avoid cutting within a group, virulent pairs with ANI value above 68% and coverage above 48% were removed from the mosaic analysis (red frame). In a marked contrast, temperate/defective pairs form a cloud of points, among which the subset with ANI value above 70% and coverage value above 50% are the pairs removed from the analysis (red frame). ANI analysis was performed for pairs of genomes listed in TableS4 having mosaics. In the black frames of Figure S6, (.92% id,.80% cov), genomes are so similar that they were completely removed from the analysis (indicated with * as redundant in the last column of Table S4).

FIGURE S7

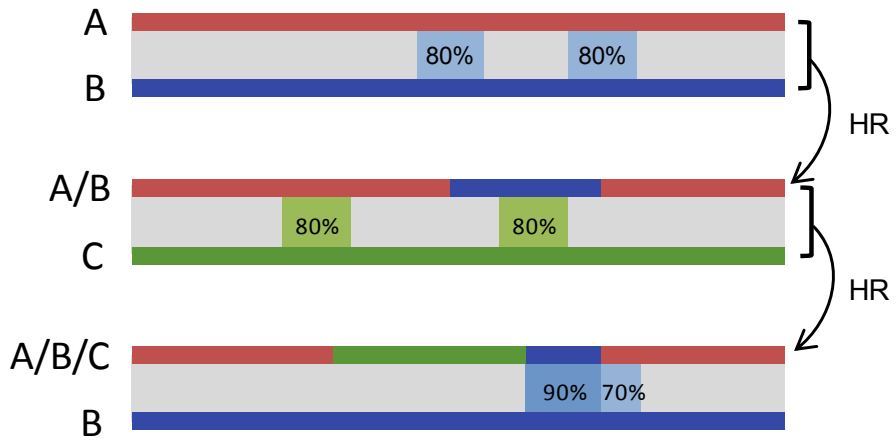


Figure S7 Scheme for the masking of homologous recombination traces. Genome A/B is the product of a recombination between A and B in regions of homology. As described in Fig. 7, alignment of the two genomes should reveal traces of homologous recombination on both sides of the mosaic. However, recombination of the A/B hybrid with a third genome, C, can mask one of the twoHR traces. Alignment of A/B/Chybrid with the ancestral B genome will reveal only one trace of homologous recombination.

Chapter III:

Sak4, a phage encoded core-only RecA, needs its cognate SSB to perform single strand annealing *in vivo*.

Sak4, a phage encoded core-only RecA, needs its cognate SSB to perform single strand annealing *in vivo*.

Hutinet Geoffrey[1], Besle Arthur[2], Son Olivier[1], Moncaut Elisabeth[1], Guerois Raphael[2], Ochsenbein Françoise[2*], and Petit Marie-Agnès[1*]

[1] INRA, UMR1319 Micalis, 78350 Jouy en Josas, France
[2] Laboratoire de Biologie Structurale et Radiobiologie, CNRS UMR 8221 and CEA iBiTec-S, Gif-sur-Yvette, France

* To whom correspondence should be addressed. Tel: 33 1 34 65 20 77; Fax: 33 1 34 65 20 65; Email: marie-agnes.petit@jouy.inra.fr . Correspondence may also be addressed to francoise.ochsenbein@cea.fr

ABSTRACT

Bacteriophages, the viruses infecting bacteria, are remarkable under many aspects, one of them being the incredible diversity of proteins they encode to perform DNA replication and homologous recombination processes. Looking back at these extremely simplified forms of life may help to untangle how similar proteins work in more sophisticated organisms. In a previous study apprehending phage recombinase diversity, we uncovered the Sak4 family, which are core-only RecA-like proteins. Such proteins are not found in Bacteria, but are present in Archaea (RadB), and in Eukaryota (CSM2 and PSY3 of the Shu-complex in yeast, XRCC2 in human), where they participate in recombination events by helping the nucleation/polymerization of Rad51/RadA filaments. But none of these core-only RecA seems to directly promote some sort of recombination by itself. Here, we report a biochemical and genetic characterization of two Sak4, encoded by phages HK620 of *Escherichia coli* and PA73 of *Pseudomonas aeruginosa*. We show that a distant homolog of SSB is present next to the *sak4* gene in most phages. Using a ssDNA recombineering assay, we found that $Sak4_{HK620}$ promotes SSA *in vivo*, provided its cognate SSB is co-expressed with it. Similarly to $Red\beta_{\lambda}$, $Sak4_{HK620}$ is active on diverged DNA substrates, up to 12% *in vivo*. *In vitro*, Sak4 proteins were insoluble, except for the GST-$Sak4_{PA73}$ fusion, which was further studied. It was unable to perform the typical, RecA-dependent strand exchange reaction, but performed efficient SSA, pairing up to 20% diverged oligonucleotides. Its cognate SSB had affinity to ssDNA, but did not enhance any of its biochemical properties.

INTRODUCTION

Homologous recombination is a key mechanism to the survival of all living organisms and a major force to their evolution. The recombinase is key to the process and conserved in all the three living kingdoms. In Bacteria, this recombinase is named RecA, and its mechanism of action is well described, particularly in *Escherichia coli* (see (Kuzminov, 1999;Cox, 2007) for reviews). In Eukaryota, its homologue Rad51 is also well studied in many organisms, from yeast to humans (Krogh and Symington, 2004;Krejci et al., 2012). In Archaea, the RadA protein shares homology with RecA and Rad51 (in the 30% identity range) and shares the same activity (Seitz et al., 1998;Haldenby et al., 2009). Briefly, these proteins bind ssDNA, form a filament on it and invade dsDNA to search homology and anneal the two DNA, by a so-called strand invasion mechanism. Recently, for RecA and Rad51, these activities were dissected using single molecule assays (van der Heijden et al., 2007;van der Heijden et al., 2008;Forget and Kowalczykowski, 2010; 2012).

Redβ of the phage λ, infecting *E. coli*, promotes homologous recombination by a mechanism distinct from RecA, namely single strand annealing (SSA). *In vitro*, it binds to ssDNA and facilitates annealing of two complementary ssDNA ((Kuzminov, 1999) for review). It is proposed to act *in vivo* by first loading on an ssDNA fragment (prepared by its cognate exonuclease Redα), and then promoting the annealing of this ssDNA fragment onto the lagging strand matrix at the replication fork, taking the place of an Okazaki fragment (Maresca et al., 2010). Furthermore, Redβ also promotes some level of strand invasion *in vitro*, under conditions of low of magnesium and AT-rich DNA template (Komori et al., 2000). Some *in vivo* studies are also suggesting that Redβ and the related RecT recombinase perform strand invasion, albeit less efficiently than SSA (Muyrers et al., 2000;Martinsohn et al., 2008).The efficiency of Redβ *in vivo* is high, and we showed recently that homologous recombination was as efficient for Redβ and RecT as for RecA on perfectly identical sequences, and even 10 to 100 fold more efficient compared to RecA, on diverged sequences (up to 22% divergence) (Martinsohn et al., 2008;De Paepe et al., 2014). This suggests that phage-encoded recombinases of this family have a relaxed fidelity, which may be at the root of the remarkable mosaicism of some temperate phage genomes.

Redβ was recently recognized to share distant similarity with the N-terminal domain of the eukaryotic Rad52 protein, together with other phage recombinases of the Erf and Sak families (Ploquin et al., 2008;Erler et al., 2009;Lopes et al., 2010). Altogether these three families of Rad52-like proteins constitute the most populated superfamily of phage recombinases. Whereas the main activity of Rad52 in *S. cerevisiae* is that of a cofactor of Rad51, it should be noted that in phages, the protein is fully active as a recombinase by itself *in vivo*: it acts equally well in wild type strains and in

strains deleted for *recA*. Some evidence has nevertheless been gathered, suggesting that the Sak protein of the *Lactococcus lactis* phage ul36 also facilitates RecA-dependent strand exchange *in vitro* (Ploquin et al., 2008), and that the Redβ protein of phage λ is accessory to RecA-dependent recombination *in vivo*, under conditions where the phage is not replicating (Stahl et al., 1997).

Besides this important and particularly well studied family of phage recombinases, comparative genomics on phage genomes permitted us to detect two other important superfamilies of recombinases involved in homologous recombination (Lopes et al., 2010). One is composed of proteins similar to Gp2.5 of phage T7, infecting *E. coli*. This recombinase has an OB-fold, as the single strand binding (SSB) proteins, with a supplementary α-helix (Hollis et al., 2001;Rezende et al., 2003): it functions both as an SSB and a recombinase *in vitro* and *in vivo*, and specific mutations can alter independently each of these two activities (Rezende et al., 2003). The other superfamily includes recombinases similar to RecA, and is subdivided into two groups: (i) the full-size RecA-like proteins, such as UvsX of phage T4 infecting *E. coli* ((Liu and Morrical, 2010)for review), that are mostly found in some virulent phages with large genomes. These recombinases function in a way similar to RecA or Rad51 (Liu et al., 2011). (ii) The other family (named hereafter 'core-only RecA') shares with RecA only the core domain, which in RecA is responsible for two key functions: ATP hydrolysis and ssDNA binding. They are present on medium-sized genomes of temperate phages, mostly from Gram positive bacteria. A few of them (4 among the 44 Sak4 of the Lopes et al. study (Lopes et al., 2010)) are also found in phages from Gram negative bacteria, such as HK620 of *E. coli* and PA73 of *Pseudomonas aeruginosa*. The first one to be described was named Sak4, and should not be confounded with Sak and Sak3 proteins that are both Rad52-like. It was discovered after isolation of a mutant of the *Lactococcus lactis* phage 31 that could infect the phage-resistant strain of *L. lactis* strain expressing the AbiK protein. The mutation was found in a gene of phage 31 of previously uncharacterized function, and named *sak4* (sak stands for suppressor of AbiK). As AbiK is a plasmid-encoded protein permitting abortive phage infection by targeting the phage-encoded homologous recombination gene, it was supposed that Sak4 was a recombinase (Bouchard and Moineau, 2004). Indeed, using a single strand recombineering assay, we reported that the Sak4 protein encoded by phage PA73 of *P. aeruginosa* promoted a low level of SSA (Lopes et al., 2010).

Some other core-only RecA are present in Archaea, where they are designated as RadB, and in Eukaryota, where the most studied are the Csm2 and Psy3 proteins of the Shu-complex in *Saccharomyces cerevisiae*, and the Xrcc2 protein in human. These proteins participate in recombination events by stabilizing or helping the nucleation/polymerization of Rad51 or RadA filaments (Komori et al., 2000;Godin et al., 2013). But none of these core-only RecA seems to be a recombinase by itself, able to directly promote strand invasion or SSA. However, the BCDX2 complex,

composed by 3 full-size paralogs of Rad51 (Rad51B, C and D) and a core-only Rad51 (Xrcc2) can promote SSA in human (Yokoyama et al., 2004).

Here, we report a genetic characterization of *sak4* encoded by phage HK620 (infecting *E. coli*) and a biochemical characterization of Sak4 of phage PA73 of *P. aeruginosa*. We show that a distant homolog of SSB is present next to the *sak4* gene in most phages, and using a ssDNA recombineering assay, we found that $Sak4_{HK620}$ promotes SSA *in vivo*, provided its cognate SSB is co-expressed with it. Similarly to $Red\beta_{\lambda}$, $Sak4_{HK620}$ is active on diverged DNA substrates, up to 12% *in vivo*. *In vitro*, Sak4 proteins turned out to be insoluble, except for the GST-$Sak4_{PA73}$ fusion, which was further studied. It was unable to perform the typical, RecA-dependent strand exchange reaction, contrary to RecA, but performed efficient SSA, pairing up to 20% diverged, 81nt long oligonucleotides *in vitro*. Its cognate SSB had affinity to ssDNA, but did not enhance any of its biochemical properties.

MATERIAL AND METHODS

Strains & oligonucleotides and proteins

Strains of *E. coli* used in this study are described in Table 1. Oligonucleotides used in recombineering and *in vitro* SSA tests are described in Table 2. The purified RecA and SSB proteins of *E. coli* were purchased from Tebu.

PSI-Blast search

HkaK protein sequence from HK620 (annotated later in the text by SSB_{HK620}) and Orf31 protein sequence from PA73 (annotated later in the text by SSB_{PA73}) were compared by PSI-BLAST against the NRDB (non redundant data base) with standard parameters. These comparisons permitted to link the *orf31* and *hkaK* genes to atypical two OB-fold SSB proteins from some phages infecting *Staphylococcus aureus*.

Recombineering in the *cat* gene

To select recombinants, an assay based on the acquisition of chloramphenicol resistance upon recombination was set up. For this, a cat_{SS} gene with two stop codons separated by 6 nucleotides was constructed: insertion of T528G and C535T stop mutations into the *cat* gene of pACYC184 was

done by site-directed mutagenesis. The *E. coli thiF* and *thiE* genes were replaced by this mutated gene in an AB1157 *recA mutS* background (see Suppl. Text for details for the recombineering selection choice, and for strain constructions). To test easily various phage recombinase genes, a plasmid expression system derived from the thermosensitive, low copy number plasmid pKD46 was used. The plasmids carrying recombinases under the control of an arabinose inducible promoter, with or without accessory factors, were transformed into this strain.

An over-night culture grown at 30°C was diluted at the hundredth in 50mL of SLB+ampicillin (35 g/L Bactotryptone, 20 g/L yeast extract, 80 mM NaCl and 5 mM NaOH, 100µg/mL ampicillin), supplemented with 0.1 M of MOPS to maintain the pH at 7 all along the cell growth, and grown at 30°C until an OD_{600nm} of 0.2. The recombinases and their partner were induced by addition of arabinose to a final concentration of 0.2% and the culture was shifted to 37°C. After 45 minutes, the growth was stopped by chilling in an ice-cold bath (OD_{600nm} between 0.6 and 1). The cells were centrifuged in a pre-cooled centrifuge for 7 minutes at 4°C. Cells were resuspended in 50 mL of cooled glycerol 10% and centrifuged again. They were washed in 1 mL of cold glycerol 10% in a pre-cold 1.5 mL tube twice, and then resuspended in 300 µL of cold glycerol 10%.

Electroporation was performed with 1µg of 81 nt-long oligonucleotide gently mixed with 100 µL of prepared cells and electroporated at 2.5 V, with a resistance of 200 Ω and a capacitance of 25 µFd. 900 µL of SOC media was added and electroporated cells were grown for 1 h at 37°C. Dilutions (done in SOC media) were then mixed to 5 mL of heated Top Agar 7 (10 g/L Bactotryptone, 2.5 g/L NaCl and 7 g/L Agar) and plated on LB agar to count total viable cells (titers of ~10^9 cells/ml are obtained, with this “top layer” method, which allows the best recovery of electroporated bacteria). To count CmR recombinants, the same procedure was used, except that after 2 h of growth in the top-agar layer at 37°C, a second top-layer of a mix of 5 mL of heated Top Agar 7 supplemented with 60 µg/mL of chloramphenicol was added. Viable cells were numerated after 24h of growth at 37°C, and recombinant cells were numerated after 48 h. The frequency of recombination was established by the ratio recombinant cells divided by viable cells.

Protein purification

$UvsX_{T4}$, $Red\beta_{\lambda}$, $Sak4_{PA73}$, SSB_{PA73} and SSB_{HK620} were purified from Lemo21-DE3 *Escherichia coli* transformed by pETM30-derived plasmids (see supplementary methods for plasmid construction details). Cells were cultured into 800mL of LB+kanamycin (30µg/mL) at 37°C until an OD_{600nm} of 1. The culture was then transferred at 20°C and the recombinases tagged by GST were induced by IPTG at a

final concentration of 100µM. Cells were then cultured overnight at 20°C and harvested by centrifugation at 3500 rpm for 30 min at 4°C. They were resuspended at 4°C in solution A (100 mM Tris HCl pH7.5, 100mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10% Glycerol, triton 1X, 1 mM PMSF, 8 µg/mL aprotinine) for $UvsX_{T4}$, $RedB_{\lambda}$, SSB_{PA73} and SSB_{HK620} constructions or solution B (100 mM Tris HCl pH8.5, 500 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10% Glycerol, triton 1X, 1 mM PMSF, 8µg/mL aprotinine) for $Sak4_{PA73}$ construction. These solutions were supplemented by 2.5 mg/mL of lysosyme and stirred on ice for 15 min. Cells were then lysed by sonication using a vibracell 75042 sonicator at 25, 35, 45, and 55 % intensity (1 min at each intensity). The lysate was then clarified by centrifugation (30 min at 20000 rpm at 4°C) and 5mL of GSH-agarose beads resuspended in buffer C (10 mM tris-HCl pH7.5, 5 mM $MgCl_2$, 100mM NaCl, 1 mM DTT) was added. The mixture was incubated at 4°C for 1 h under agitation. The beads were then transferred on a gravity flow column (Biorad) and washed with 25 mL buffer D (10 mM tris-HCl pH7.5, 5 mM $MgCl_2$, 1 M NaCl, 1 mM DTT). A second wash was done with 25 mL of buffer C. The fusion proteins were eluted with 15mL of a buffer E (10 mM tris-HCl pH7.5, 5 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, 10 mM GSH) for $UvsX_{T4}$ and $RedB_{\lambda}$ constructions, or buffer F (10 mM tris-HCl pH8.5, 5 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, 10 mM GSH) for $Sak4_{PA73}$, SSB_{PA73} and SSB_{HK620} constructions. Except for $Sak4_{PA73}$, the his-GST-tag was removed by the addition of TEV protease (his-tagged) at a concentration of 1/100 (w/w) overnight at 4°C. The his-GST-tag and his-TEV protease where then removed using a His-trap fast flow 5mL column. The flow through containing the proteins was dialysed overnight at 4°C against buffer C. The elution containing the GST-$Sak4_{PA73}$ construction was dialysed overnight at 4°C against buffer D. SDS-PAGE was used to control each step of the purification process.

The protein concentrations were determined by the A_{380nm}, using the ϵ of each protein: 1.13 cm^{-1} M^{-1} for $UvsX_{T4}$, 1.45 cm^{-1} M^{-1} for $RedB_{\lambda}$, 1.25 cm^{-1} M^{-1} for GST-$Sak4_{PA73}$, 1.51 cm^{-1} M^{-1} for SSB_{PA73} and 1.17 cm^{-1} M^{-1} for SSB_{HK620}.

SSB gel shift assays

SSB proteins were incubated with the Cy5-labelled oligonucleotide GO47 (10 nM) for 15 minutes at 37°C in 10 µL of buffer G (10 mM tris-HCl pH7.5, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/mL BSA). Saccharose to 8% final concentration was added prior to gel loading. Electrophoresis was performed with 10% acrylamide/bis-acrylamide (19:1) gels and 1X TBE migration buffer, at 120V for 2 hours, and the gel was revealed by the BioRad ChemiDoc™ MP imaging system.

ATPase assay

ATPase activity was assayed by linking ATP hydrolysis to NADH oxidation: in the presence of lactate deshydrogenase (LDH) and pyruvate kinase (PK), for each ATP molecule hydrolyzed, a NADH molecule is oxidized and the solution absorbance at 340nm decreases (Pullman et al., 1960). The reactions were carried out, with or without 10 nM of GO34, at 37°C in a 70 µL final volume of Buffer H (10 mM Tris-HCl pH7.5, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/mL BSA, 7.5 mM phosphoenol pyruvate, 0.8mM NADH, 1mM ATP) with 1 µL PK/LDH mix from Sigma-Aldrich®. The recombinases were used at 0.5 µM for $RecA_{E.coli}$, 2 µM for $Red\beta_{\lambda}$, 0.5 µM for $UvsX_{T4}$, 0.5 µM for GST-$Sak4_{PA73}$. Aliquots of 10µL were taken at different times and mixed in 40µL of 25mM EDTA to stop the reaction. The A_{340nm} was measured in 96 wells plates by Tecan Infinite® M200 reader. The concentration of NADH was determined as a function of absorbance, using a standard NADH curve.

Single strand annealing assay

The Cy5-labelled oligonucleotide GO47 (10 nM) was preincubated with one recombinase (0.5 µM for $RecA_{E.coli}$, 2 µM for $Red\beta_{\lambda}$, 0.5 µM for $UvsX_{T4}$, 0.5 µM for GST-$Sak4_{PA73}$) for 2 minutes at 30°C in 20µL of final volume of the buffer G (10 mM tris-HCl pH7.5, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/mL BSA). The reaction was started by the addition of one more or less diverged oligonucleotide (10 nM, see Table2). After 7 minutes, 4.5µL of the reaction was mixed with 3µL of stop buffer (by 10 mM of tris-HCl pH7.5, 3 µM of GO35 (untagged GO47), 0.5% of SDS and 0.2 mg/ml of proteinase K), and incubated 5 minutes at 30°C. The sample was loaded on a 10% acrylamide/bis-acrylamide (19:1) gel with a 1X TBE migration buffer. Electropheresis was performed at 120V for 2 hours, and the gel was revealed by the BioRad ChemiDoc™ MP imaging system. For kinetics, the same protocol is repeated in 50µL of reaction volume.

Strand exchange assay

Circular ssDNA from ΦX174 (10 µM in nucleotide) was incubated with 5 µM for $RecA_{E.coli}$, or 1.16 µM for GST-$Sak4_{PA73}$, or both, for 10 min at 37°C in 8 µL of final volume of the buffer H (10 mM Tris-HCl pH7.5, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/mL BSA, 1 mM ATP, 20 mM phosphocreatin and 10 units of creatin kinase). A second step of incubation with $RecA_{E.coli}$ or GST-$Sak4_{PA73}$ was added for 1 minute at 37°C in certain reactions, indicated in the figure legend. SSB was added (1 µL for a final concentration of 0.5 µM for $SSB_{E.coli}$ or 1 µM for SSB_{PA73}; milliQ water was added when there was no SSB) to the

solution. The reaction was kept at 37°C for 10 min. Strand exchange reaction was started by addition of 1µL containing *Pst*I-linearized double strand ΦX174 was added for a final concentration of 10 µM (in nucleotides). After 30 minutes incubation at 37°C, the reaction was stopped by adding 2.5 µL of SI stop buffer (150 mM EDTA, 6% of SDS and 0.8 mg/ml of proteinase K). After at least 15 minutes at 37°C, the sample was loaded on a 0.8% GTG-agarose gel with a 1X TAE migration buffer. Electrophoresis was run at 100 V for 5 hours at 4°C. The gel was colored by Ethidium Bromide and revealed by the BioRad ChemiDoc^{TM} MP imaging system.

RESULTS

Temperate phages encoding Sak4 often encode an atypical SSB nearby

HK620 is the only sequenced phage infecting *E. coli* which encodes a Sak4. In the original annotation of its genome (Clark et al., 2001), four predicted genes were suspected to participate to homologous recombination by comparison with the *Salmonella typhimurium* phage P22 genome organization, as phage genomes are often organized into functional modules. These genes were therefore initially annotated as *arf*, *erf*, *abc1* and *abc*2 respectively, similarly to P22. A map of this region in the two phages is shown Fig. 1. Upon closer inspection, only one of these four genes shares homology with a P22 gene: *abc2*. The arf_{HK620} gene is orfan, and has no homology to *arf* of P22, we therefore name it $hkaM_{HK620}$. The erf_{HK620} gene is not homologous erf_{P22}, but similar to the new family of recombinases exemplified by *sak4* of phage Φ31, a core-only RecA-like recombinase (Lopes et al., 2010). Below, it will therefore be designated as $sak4_{HK620}$. Finally, the $abc1_{HK620}$ gene has no similarity to $abc1_{P22}$, despite occupying a similar genetic position in the recombination module. A PSI-Blast analysis indicated rather homology with a family of SSB-like proteins, which have the particularity to contain two OB-folds. This family encompasses members from various origins (phages, *Deinococcus radiodurans*, Archaea) (see methods). Some of the SSB detected by PSI-Blast were nevertheless shorter in length, and contained only one OB-fold. Interestingly, all these SSB orthologs of phage origin happen to be placed next to a *sak4* gene (Suppl. Fig. 1). However, the reverse is not true, only 80% of the *sak4* genes were found next to a gene encoding an SSB (suppl. Table 3). We renamed therefore $abc1_{HK620}$ as ssb_{HK620}. Finally, the $abc2_{HK620}$ gene shares 94% identical amino-acids with $abc2_{P22}$ (Clark et al., 2001), which functions as a hijacker of $RecBCD_{E.coli}$, converting the potent ds exonuclease into a 5'-3' exonuclease which prepares DNA extremities for recombination (Murphy, 2000).

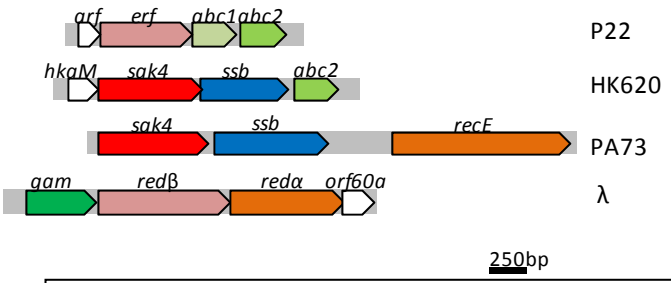


Figure 1: Genetic organization of homologous recombination partners in four phages. P22, infecting *Salmonella typhimurium*, has four genes in operon involved in homologous recombination: arf_{P22} of unknown function, erf_{P22} encoding a Rad52-like recombinase (Lopes et al., 2010), $abc1_{P22}$ and $abc2_{P22}$, producing $RecBCD_{E.coli}$ hijackers. HK620, infecting *E. coli*, also has four genes potentially involved in homologous recombination (Clark et al., 2001): $hkaM_{HK620}$ of unknown function, $sak4_{HK620}$, encoding a core-only Rad51-like recombinase (Lopes et al., 2010), ssb_{HK620}, encoding a single strand binding protein, and $abc2_{HK620}$, homologous to $abc2_{P22}$. In the *Pseudomonas aeruginosa* phage PA73, the two genes flanking $sak4_{PA73}$ are ssb_{PA73}, homologous to ssb_{HK620}, and $recE_{PA73}$, encoding an exonuclease. In the *E. coli* phage λ the recombination module contains four genes: gam_{λ}, encoding an inhibitor of $RecBCD_{E.coli}$, $red\beta_{\lambda}$, homologous to erf_{P22} (Lopes et al., 2010), $red\alpha_{\lambda}$, encoding an exonuclease, and $orf60a_{\lambda}$ of unknown function.

We searched for other 'local co-occurrences' among the 44 predicted *sak4* genes: phage HK620 was the only one to possess an *abc2* gene (2%), but 25% had a *recE*-like exonuclease (an accessory factor), which prepares single-stranded substrates for the RecT recombinase ((Kuzminov, 1999) for review), and the remaining 73% had no detectable equivalent function.

We therefore undertook the study of two Sak4 in parallel: the one encoded by HK620, and one present on PA73, a phage infecting *Pseudomonas aeruginosa* (*recE* accessory factor, see Fig. 1). In both phages, a gene encoding a putative single strand binding protein was present next to the *sak4* gene (fig. 1), and given the preferred association of these two genes detected above, the study involved these SSB as well. In bacteria, the SSB protein is ubiquitous and essential: the gene cannot be deleted. Its main function is to facilitate the replication process, but it also has a role in homologous recombination: $SSB_{E.coli}$ protects the ssDNA from $RecA_{E.coli}$ nucleation, but once $RecA_{E.coli}$ is nucleated on DNA, $SSB_{E.coli}$ facilitates its polymerization by suppressing DNA secondary structures. In addition, it is supposed that $SSB_{E.coli}$ catches the displaced strand during the D-loop formation, to stabilize the recombination intermediate (Cox, 2007;Shereda et al., 2008). The reason why phages encode their own SSB proteins is not clear at present. Recently, it was shown for phage SPP1 that its SSB intoxicates the bacterial replication fork (Lo Piano et al., 2011). Here, we investigated whether phage SSB also plays a role in phage recombination.

The SSB_{HK620} is essential for $Sak4_{HK620}$ SSA activity *in vivo*

To test the capacity of Sak4 recombinases to anneal ssDNA *in vivo*, alone or in combination with potential cofactors, a recombineering assay was performed: the Maj98 oligonucleotide, an 81-mer complementary to the mutated cat_{SS} gene placed into the *E. coli* bacterial chromosome, was designed such that it could pair with the lagging strand, relative to the replication fork. Upon successful annealing, the chloramphenicol resistance is reestablished (Fig.2A,(Li et al., 2013)). Because we also planned to test the tolerance of the recombineering reaction to mismatches, the mismatch repair gene *mutS* of the strain to be tested was invalidated. To decrease the background level of spontaneous mutants, that is high in such *mutS* strains, the cat_{SS} gene had two mutations

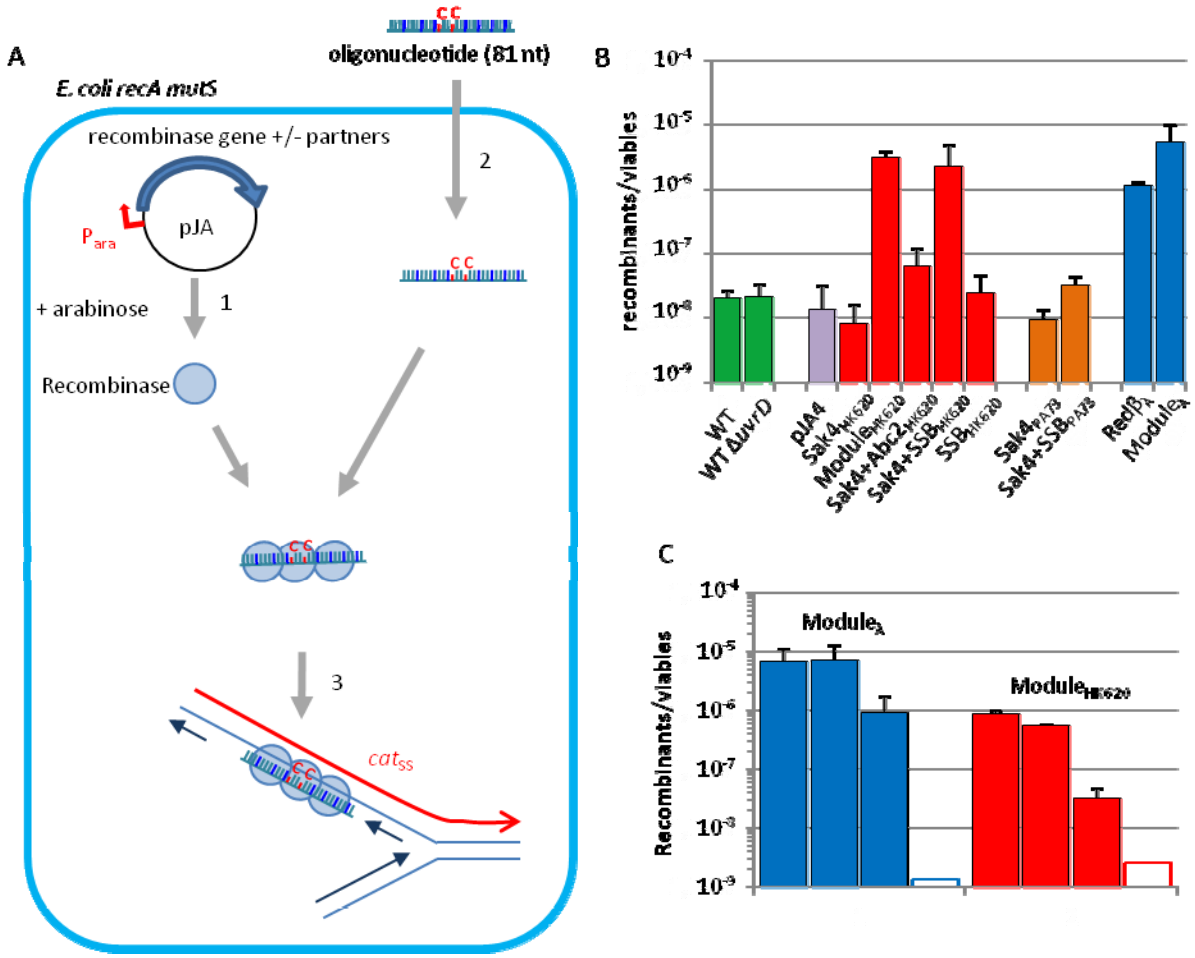


Figure 2: Recombineering activity. A. Assay: a *recA mutS E. coli* [Cm^S], carrying a plasmid coding for a recombinase under the control of an arabinose inducible promotor (1), is electroporated with an 81-mer oligonucleotide (2) complementary to a mutated *cat* gene (cat_{SS}, see Suppl. Text). Integration of the oligonucleotdide at the replication fork (3), confers a Cm^R phenotype to the bacterium. **B.** recombineering efficiency for four phage recombinases, with or without their partners described Fig.1 : $RecA_{E.coli}$, $Sak4_{HK620}$, $Sak4_{PA73}$ and $Red\beta_{\lambda}$. **C.** Effect of the oligonucleotide divergence relative to the target catSS gene, on the efficiency of recombineering. The recombinases were tested with all their partners. White bars show the detection limit.

placed next to each other to be restored for chloramphenicol resistance (CmR) reestablishment. The resulting strain was transformed with different plasmids coding for recombinases with or without their accessory factors (see methods). In this strain the spontaneous occurrence of CmR revertants was below the detection level (10^{-9}). However, a mock experiment of transformation with the Maj98 oligonucleotide in the strain carrying the vector plasmid devoid of recombinase, pJA4, indicated that spontaneous SSA occurs *in vivo*, at a frequency of $1.2x10^{-8}$.

In a RecA^{+} *E. coli* strain, the SSA remains at the same level as the spontaneous annealing. Indeed, it is known that RecA is not able to promote such annealing in vivo (Lopes et al., 2010). Among the many mutants that have been tested in an attempt to overcome the barrier to recombineering in *E. coli*, *uvrD* has not been tested, to our knowledge. As the deletion of this helicase is known to increase the recombination rate of RecA in classical conjugation assay (Veaute et al., 2005), we performed the recombineering experiment in an *uvrD* mutant, and observed no increase in recombination efficiency (Fig. 2B, 1st panel).

We performed all following experiments in a *recA* background, to exclude any potential interference of RecA in the assay. In the presence of a pJA4 derivative expressing Redβ_{λ}, SSA was increased by a factor of 80, relative to the spontaneous annealing. next, the pKD46 plasmid encoding the *redβ_{λ}* module was tested (Datsenko and Wanner, 2000), Fig. 1, last map): in addition to the gene encoding Redβ_{λ}, it contains the genes for Redα_{λ}, a dsDNA-dependent 3'-5' exonuclease (Subramanian et al., 2003), Gam$_{\lambda}$, an inhibitor of the exonuclease RecBCD$_{E.coli}$ (Court et al., 2007;Murphy, 2007) and Orf60a$_{\lambda}$, of unknown function, but reported to enhance dsDNA-recombeening (Datsenko and Wanner, 2000). SSA was increased by a factor of 5 over the Redβ_{λ} annealing (Fig. 2B, last panel).

In contrast to Redβ_{λ}, Sak4$_{HK620}$ did not promote SSA by itself (Fig. 2B, 2nd panel). To test whether Sak4 required an accessory protein, a plasmid encoding all three surrounding genes (*ssb*$_{HK620}$, *abc2*$_{HK620}$ and *hkaM*$_{HK620}$, see Fig.1) was tested. This time, a frequency of recombinants 240-fold over the spontaneous annealing was obtained. We next investigated which gene was implied in this effect. The *abc2*$_{HK620}$ and ssbHK620 genes were tested separately. Abc2$_{HK620}$ had a very small effect on Sak4$_{HK620}$-dependent SSA, but SSB$_{HK620}$ restored almost completely the phenotype observed with the all three surrounding genes, whereas this protein alone did not increase the spontaneous SSA level. We conclude therefore that *in vivo*, Sak4-promoted SSA annealing depends completely on the presence of its companion SSB.

Whereas Sak4$_{PA73}$ was found to promote SSA three fold above spontaneous annealing with a different recombineering assay (Lopes et al., 2010), it did not promote SSA in the present assay (Fig. 2B, 3rd panel). A construction with two *sak4*$_{PA73}$neighboring genes, *ssb* $_{PA73}$ and *recE* $_{PA73}$ (Fig 1) could

not be performed, as it was toxic to *E. coli*, but a plasmid containing *sak4*$_{PA73}$ and *ssb*$_{PA73}$ was obtained. No SSA was observed with this construction. We cannot exclude that these proteins of a *Pseudomonas* phage are inactive in the heterologous host *E. coli*.

Both Sak4$_{HK620}$ and Redβ$_{\lambda}$ anneal up to 12% diverged sequences *in vivo*

The capacity of the Sak4$_{HK620}$ recombinase to anneal diverged sequences was tested *in vivo*, using several 81 nt-long oligonucleotides with equally spaced mismatches (divergence from 2% to 28%), as presented in Table 2. Redβ$_{\lambda}$ was tested in parallel, as this protein is already known to tolerate up to 12-16% of divergence during recombination *in vivo* (Martinsohn et al., 2008;Li et al., 2013;De Paepe et al., 2014). Both Redβ$_{\lambda}$ and Sak4$_{HK620}$ were tested together with all their accessory factors, given their high frequency of SSA (Fig. 2B). Annealing efficiency was high for 2% and 6% diverged pairs, and decreased sharply for 12% of divergence (Fig. 2C). No recombinants were detected for 16% diverged sequences. Interestingly, the two recombinases were similarly sensitive to divergence *in vivo*, suggesting that some common factor limits the integration of highly diverged pairs, such as those formed *in vitro* between 20% diverged oligonucleotides. This limiting factor may be exonuclease degradation of the oligonucleotides prior to their annealing, so that final duplexes are shorter and therefore less stable. Indeed, sequencing of 8 and 9 recombinants for the Redβ$_{\lambda}$ and Sak4$_{HK620}$ assays respectively, among those obtained with 12% diverged sequences, showed that all but one had integrated less than the total length of the input oligonucleotide (Suppl. Fig. 2). A slight tendency of a preferred degradation of the 3' end of the oligonucleotide was observed: 80% of integrated oligonucleotides were degraded in 3' relative to the incoming oligonucleotide, and 65% in 5' (Suppl. Fig. 2). A previous study from the Court laboratory has shown that the degradation in 5' is due to the exonuclease activity of Pol I and degradation in 3' to ssDNA exonucleases (Li et al., 2013).

Sak4 proteins are hardly soluble

To further study the Sak4 activity *in vitro*, Sak4$_{HK620}$ and Sak4$_{PA73}$ were expressed in *E. coli* with a GST tag (see Methods). The GST-Sak4$_{HK620}$ fusion formed inclusion bodies and was not studied further *in vitro*. The GST-Sak4$_{PA73}$ fusion was more soluble, but precipitated upon GST tag removal. It was therefore characterized as a fusion protein below (Fig. 3, lane GST-Sak4$_{PA73}$, the two lower migrating bands are heat shock proteins, and the small MW band is GST). We investigated the biochemical properties of the purified fusion, hereafter designated as GST-Sak4$_{PA73}$. In addition to Sak4, two well characterized phage recombinases were included in this study, as a comparison: UvsX$_{T4}$ and Redβ$_{\lambda}$.

They were also expressed as GST tagged fusion, remained soluble upon tag removal, and were therefore studied in their untagged form (Fig. 3). Finally, a commercial preparation of $RecA_{E.coli}$ was also included in the assays (Fig. 3).

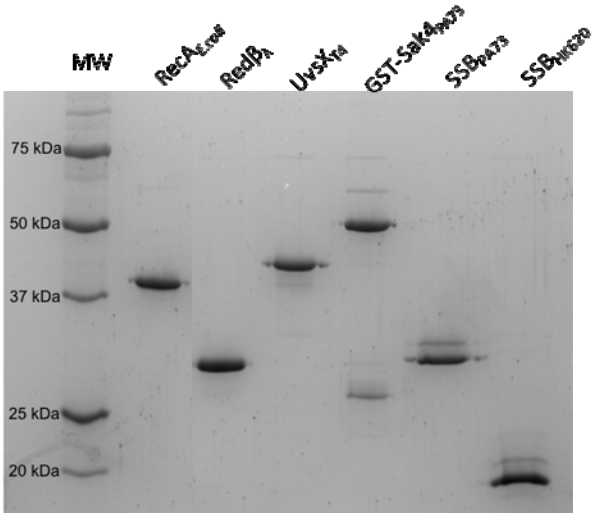


Figure 3: Purified proteins on gel. $Red\beta_{\lambda}$, $UvsX_{T4}$, $Sak4_{PA73}$, SSB_{PA73} and SSB_{HK620} were purified as described in the methods section, $RecA_{E.coli}$ was from Tebu. The molecular weights (MW) of the ladder are indicated.

The phage-encoded SSB-like proteins have affinity for ssDNA

As the activity of Sak4 is dependent on its cognate SSB *in vivo*, SSB_{HK620} and SSB_{PA73} were purified as well, and compared to $SSB_{E.coli}$. The two phage SSB proteins were produced as GST-tagged fusion, and remained soluble upon GST removal (Fig. 3, two last lanes, SSB_{HK620} is shorter in size and corresponds to a single-OB fold version of phage SSB). In order to validate the predicted single strand binding function, increasing concentrations of each SSB were incubated in the presence of Cy5 labelled, 81-nt long oligonucleotide (GO47), and the products were separated by electrophoresis. Each phage SSB shifted the ssDNA on gel. Based on the intensity of the non-displaced band, corresponding to the free oligonucleotide, an 8-fold reduced affinity was observed, compared to $SSB_{E.coli}$ (Fig. 4). The shifted bands were of a much lower mobility in the case of SSB_{PA73}, probably proportional to the protein size. We next tested whether GST-$Sak4_{PA73}$ and its cognate SSB interacted, by GST pull-down, and did not find any evidence of a direct interaction. No indication of any interaction between GST-$Sak4_{PA73}$ and $SSB_{E.coli}$ was found either.

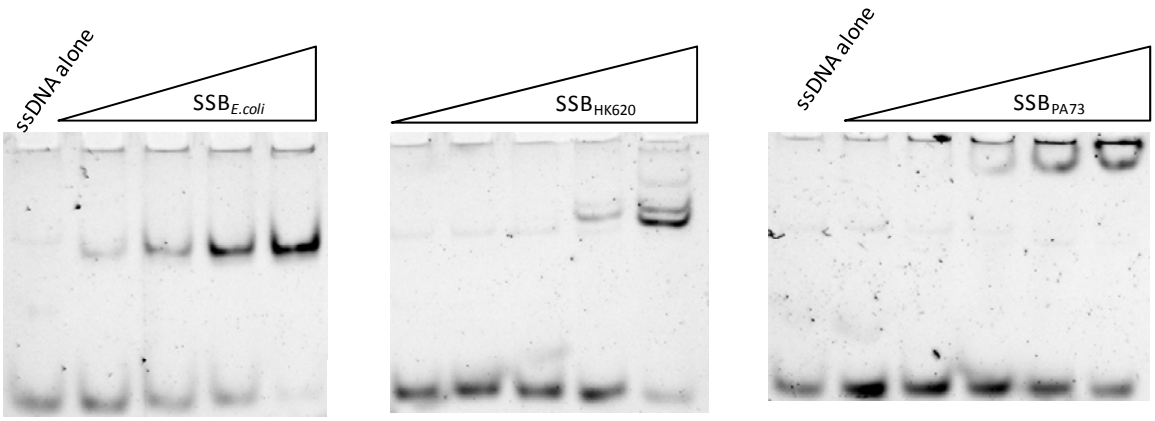


Figure 4: Gel shift of ssDNA by SSB proteins. 10nM of GO47 was mixed with SSB proteins for 15 minutes and then loaded on a 10% acrylamide/bis-acrylamide (19:1) gel. SSB$_{E.coli}$ was used at 0.25µM, 0.5µM, 1µM and 2µM. SSB$_{HK620}$ and SSB$_{PA73}$ were used at 0.5µm, 1µM, 2µM, 4µM and 8µM.

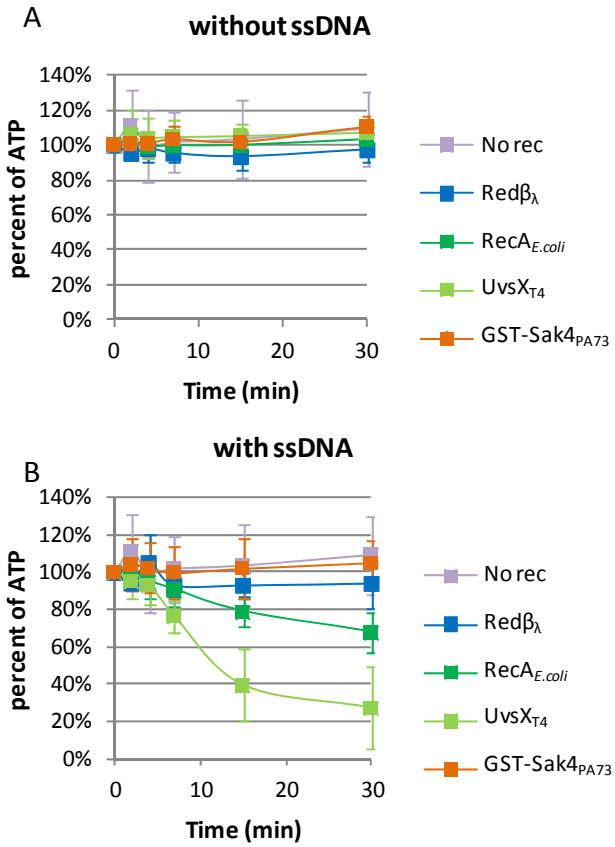


Figure 5: GST-Sak4$_{PA73}$ does not degrade ATP as UvsX$_{T4}$ and RecA$_{E.coli}$. A. Redβ$_{\lambda}$ (2µM), UvsX$_{T4}$ (0.5µM), RecA$_{E.coli}$ (0.5µM) or GST-Sak4$_{PA73}$ (0.5µM) were mixed with ATP, NADH and LDH/PK (see methods). The A$_{340nm}$ was measured and different times and the mean percentage (3 measures) of the initial absorbance is reported. Error bars indicate the lower and the higher observed values. **B.** Same experiment in the presence of 10nM of the GO34 oligonucleotide.

GST-Sak4$_{PA73}$ does not degrade ATP

As Sak4$_{PA73}$ has Walker A and B ATP-binding sites, we investigated its capacity to hydrolyze ATP. ATP hydrolysis was monitored using a coupled spectrophotometric assay (see Methods). RecA$_{E.coli}$ and UvsX$_{T4}$ hydrolyzed ATP in the presence of the 81nt long GO34 ssDNA, whereas Redβ$_{\lambda}$ did not, as expected. Interestingly, GST-Sak4$_{PA73}$ did not hydrolyze ATP either (Fig. 5).

GST-Sak4$_{PA73}$ does not promote strand exchange, nor stimulates the RecA reaction

GST-Sak4$_{PA73}$ being similar to RecA-like protein, its strand exchange activity (SE), which is typical of RecA, was investigated, using the classical assay which follows the exchange of one strand of a linear ΦX174 dsDNA (dsl) on a homologous circular ssDNA

(ssc, Fig. 6A). For $RecA_{E.coli}$, the reaction is ATP-dependent (Rehrauer and Kowalczykowski, 1993;Rice et al., 2001), and stimulated by SSB (Kowalczykowski and Krupp, 1987). We first tested $RecA_{E.coli}$ and GST-$Sak4_{PA73}$ in the presence and absence of $SSB_{E.coli}$ (Fig. 6B and 6C). $RecA_{E.coli}$ at low concentration (1µM, Fig. 6B) did not promote SE unless $SSB_{E.coli}$, was added, in which case joint molecules (JM) were formed. At a higher concentration (2µM, Fig.6C), $RecA_{E.coli}$ was able to promote JM alone, and the presence of $SSB_{E.coli}$ permitted completion of the SE reaction up to the formation of nicked circular dsDNA (nc). GST-$Sak4_{PA73}$ was tested at 2µM, and no indication of any strand exchange activity was observed (no JM nor nc products in the gels), even in the presence of $SSB_{E.coli}$.

The core-only RecA of Archaea, RadB, interacts with the RadA filament, and even slightly stimulates SE, when added late in the reaction (Komori et al., 2000). Similarly, the core-only RecA Psy3/Csm2 in Yeast and Xrcc2 in Human stabilize the Rad51 filament and stimulate SE (Krejci et al.,

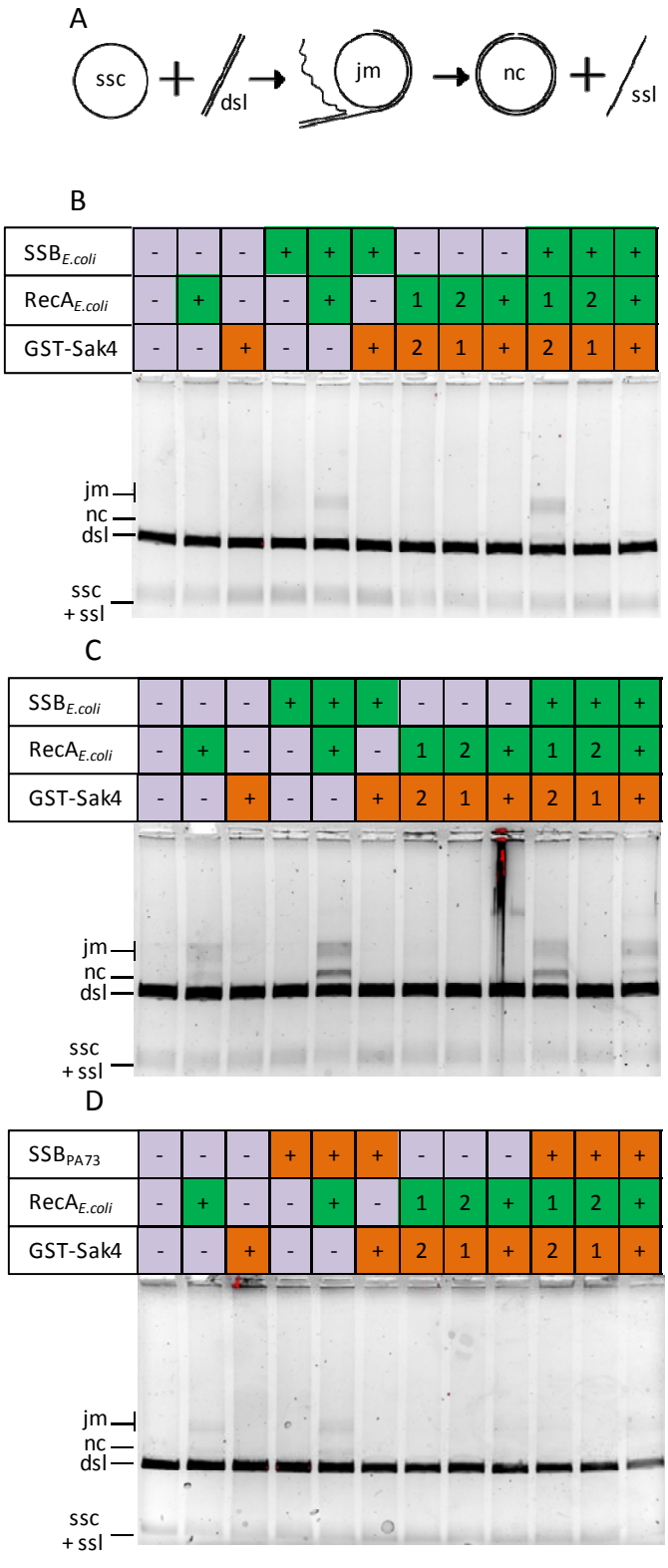


Figure 6: GST-$Sak4_{PA73}$ does not promote strand exchange. A. Scheme of the strand exchange reaction: circular ssDNA (ssc) bound to the recombinase reacts with its homologous linear dsDNA (dsl) to form joint molecules (jm) between ssc and dsl. The final products of the strand exchange are a nicked circular dsDNA (nc) and linear ssDNA (ssl). **B.** Different order of addition of proteins (1µM of $RecA_{E.coli}$, 2µM of GST-$Sak4_{PA73}$ and 0.5µM of $SSB_{E.coli}$) before addition of the dsl. The order of addition of $RecA_{E.coli}$ and GST-$Sak4_{PA73}$ is specified by 1 and 2, or added at the same time. $SSB_{E.coli}$ is always added last (see methods). **C.** Same experiment with 2µM of $RecA_{E.coli}$. **D.** Same as C but 1µM of SSB_{PA73} was used instead of $SSB_{E.coli}$.

2012). So, we tested whether GST-Sak4$_{PA73}$ could stimulate RecA$_{E.coli}$, in the 1-2µm range, where RecA$_{E.coli}$ is stimulated by SSB$_{E.coli}$. No Sak4-dependent stimulation of the SE was observed, regardless of the order of addition of the two proteins (Fig. 6B and 6C, lanes 7-9). In fact, the presence of GST-Sak4$_{PA73}$ inhibited SE, at 2µM of RecA$_{E.coli}$ (Fig. 6C, compare lane 2 and 7-8). The experiment was also performed in the presence of SSB$_{E.coli}$, and again no stimulatory effect of GST-Sak4$_{PA73}$ was found. Its inhibitory effect was lessened by the presence of SSB$_{E.coli}$, and observed only when GST-Sak4$_{PA73}$ was added prior to RecA$_{E.coli}$ (Fig. 6C, compare lanes 10 and 11, a simultaneous addition, lane12, has intermediate effect). Finally, the whole experiment was repeated with SSB$_{PA73}$ instead of SSB$_{E.coli}$, for the 2µM RecA reactions. SSB$_{PA73}$ did not stimulate RecA$_{E.coli}$, nor did it permit to unveil any SE activity for GST-Sak4$_{PA73}$ (Fig. 6D).

GST-Sak4$_{PA73}$ performs SSA *in vitro*, and anneals up to 20% diverged single stranded DNAs

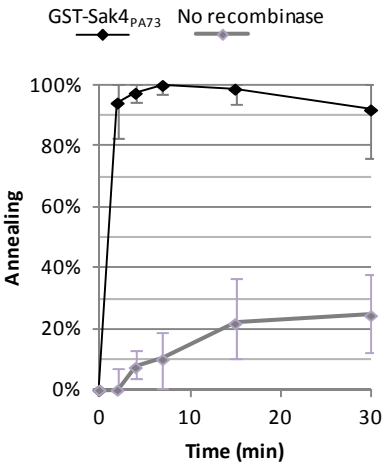


Figure 7: Annealing kinetics of GST-Sak4$_{PA73}$. Annealing kinetic without ATP, the same curves are obtained with 1 mM ATP. The error bars represent the lower and higher values obtained.

Finally, the SSA activity of GST-Sak4 was monitored, by incubating it in the presence of two complementary 81 nt long oligonucleotides, followed by protein removal and analysis of the DNA products on acrylamide gels. Kinetics of recombinase-promoted SSA is shown Fig. 7, compared to spontaneous annealing. GST-Sak4 performed efficient and rapid annealing *in vitro,* as ~100% of the dsDNA product was formed within the first minute of incubation, whereas spontaneous annealing had not occurred yet at this time point, and appeared slowly over time, with 20% ds products being made at the 30 minutes point.

To test the effect of divergence on the efficiency of annealing, increasingly diverged oligonucleotides were next used (0 to 28% divergence, the sequence of the various 81 nt-long nucleotides are shown in Table 2). GST-Sak4 was compared to RecA and, as a negative control, the GST protein was tested. The results of Fig. 8A and 8B are shown at a single time point of 7 min. No activity above the level of spontaneous annealing (10%) was observed with GST. Both GST-Sak4$_{PA73}$ and RecA$_{E.coli}$ promoted SSA, up to 20% diverged sequences. These two recombinases have therefore an efficiency of annealing comparable to that observed for Redβ$_{\lambda}$ up to 20% diverged sequences (see Suppl. Fig. 3). Interestingly, Redβ$_{\lambda}$ also annealed 28% diverged sequences, but the dsDNA product had a mobility in gel lower than expected (Suppl. Fig. 3). We have no hint at the nature of this shifted

band at present, it may be due to partial melting of this highly diverged duplex DNA. We noted a slight, but reproducible increase of the Redβ$_{\lambda}$ SSA efficiency when divergence was increased, up to 20% (Suppl. Fig. 3). We have no explanation for this tendency. Contrary to GST-Sak4$_{PA73}$ and RecA$_{E.coli}$, UvsX$_{T4}$ did not have any SSA activity in the absence of ATP (Suppl. Fig. 3). This fits with its dependence on ATP to bind ssDNA (Maher and Morrical, 2013).

The same SSA assay was repeated in the presence of 1mM ATP (Fig. 8C and 8D, Suppl Fig. 4). UvsX$_{T4}$ performed SSA with oligonucleotides up to 28% divergence this time (Suppl. Fig. 4). In contrast, the efficiency of SSA was unchanged by the presence of ATP for GST-Sak4$_{PA73}$, RecA$_{E.coli}$ and Redβ$_{\lambda}$, except that RecA$_{E.coli}$ was active also on 28% diverged sequences.

We next asked whether the host, and/or the phage encoded SSB proteins would inhibit the various recombinase's SSA activity. We first verified that none of the three SSB performed SSA *in vitro* (Suppl. Fig. 5). Increased concentrations of SSB$_{E.coli}$ or SSB$_{PA73}$ were incubated with recombinases and

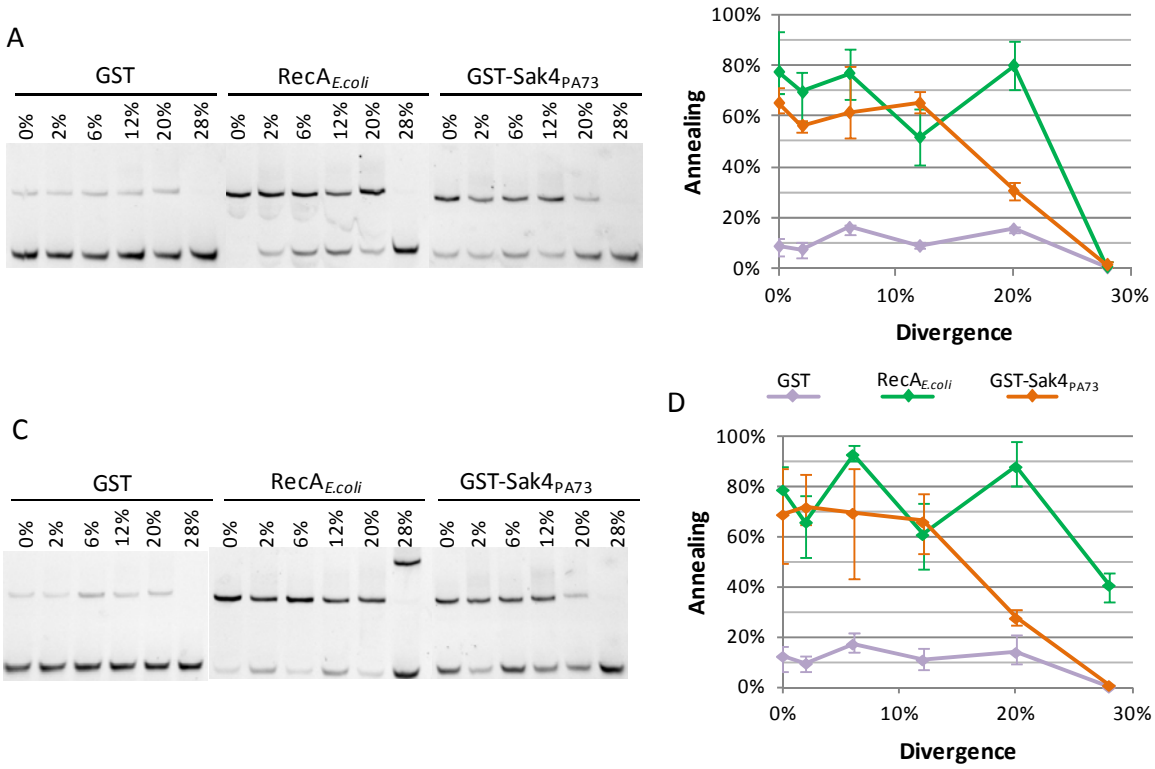


Figure 8: *in vitro* SSA activity of recombinases. A and C. Representative gels of SSA assay without (A) or with (C) 1mM ATP respectively: the GST control, RecA$_{E.coli}$ and Sak4$_{PA73}$-GST. The lower band represents the Cy5-tagged GO47. The higher band is the dsDNA corresponding to the annealing of GO47 and the more or less diverged complementary oligonucleotide (described in Table2). **B and D.** Quantification of SSA activity without (B) or with (D) 1mM ATP respectively for the GST control, RecA$_{E.coli}$, Redβ$_{\lambda}$, UvsX$_{T4}$ and Sak4$_{PA73}$-GST in function of divergence (3 measures at each point, see supplementary Figure 4 for details).

oligonucleotide GO47 for 2 minutes before the addition of the complementary oligonucleotide GO34 (Suppl. Fig. 6). Remarkably, the two SSB had completely different effects on the SSA reactions: SSB_{PA73} did not inhibit the SSA of any recombinase, whereas $SSB_{E.coli}$ inhibited the SSA activity of $RecA_{E.coli}$, GST-$Sak4_{PA73}$ and $UvsX_{T4}$ in a dose dependent manner, with complete inhibition achieved at 0.5µM of $SSB_{E.coli}$, corresponding to SSB:recombinase molar ratios of 1:1 (as monomers). The $Red\beta_{\lambda}$ activity was only partially inhibited at the highest SSB concentrations tested, which corresponded to $SSB_{E.coli}$/$Red\beta_{\lambda}$ ratio of 1:4 (as monomers). *In vivo*, both SSB should co-occur in the host during a phage cycle, we therefore tested the effect of a mix of both SSB proteins on the SSA activity: using a 250µM concentration of $SSB_{E.coli}$ and 8-fold more of SSB_{PA73}, the SSA activity of both $RecA_{E.coli}$ and GST-$Sak4_{PA73}$, were inhibited. This suggests a dominance of $SSB_{E.coli}$ over SSB_{PA73}, which may be related to its higher affinity for ssDNA (Fig. 4).

DISCUSSION

Sak4 has SSA activity

The Sak4 protein, despite sharing homology with RecA, behaves more like a phage recombinase of the Rad52 superfamily, both *in vitro* and *in vivo*. Technical reasons prevented the parallel *in vivo* and *in vitro* study of a single Sak4, but the results reported here for $Sak4_{HK620}$ *in vivo* and $Sak4_{PA73}$ *in vitro* are consistent. $Sak4_{HK620}$ promoted SSA *in vivo*, in a recombineering assay, to levels comparable to $Red\beta_{\lambda}$, provided its companion SSB was co-expressed. The only form of Sak4 that was soluble in vitro was the GST-$Sak4_{PA73}$ fusion, so it was investigated further. It exhibited an SSA activity, but no strand exchange activity, like $Red\beta_{\lambda}$. We cannot exclude that the GST fusion prevented the observation of the strand exchange activity. Furthermore, despite the clear presence of Walker A and B motifs in both Sak4, no ATPase activity was detected for GST-$Sak4_{PA73}$. This may suggest that the main Sak4 activity is not strand-invasion mediated recombination, like RecA, but SSA. This SSA would occur without degrading ATP, as is also observed for Rad52-like proteins.

Interestingly $Sak4_{HK620}$ needs the presence of its cognate SSB protein *in vivo* to carry out recombineering. This SSB dependence is not observed *in vitro* in SSA assays with GST-$Sak4_{PA73}$. But no inhibition is observed either, contrary to what is seen when the host SSB protein is added into SSA reactions. One may imagine a scenario where phage-encoded SSB are present at the replication fork,

and do not shield ssDNA from the phage recombinase. Otherwise, it is possible that SSB_{HK620} prevents the inhibition by an unknown factor that would inhibit $Sak4_{HK620}$-mediated SSA (see below). Contrary to GST-$Sak4_{PA73}$, we noticed that the SSA mediated by $Red\beta_{\lambda}$ was not completely inhibited by $SSB_{E.coli}$ on linear substrates (Suppl. Fig. 6). It was previously shown that $Red\beta_{\lambda}$ anneals ssDNA in the presence of $SSB_{E.coli}$ at low concentration of DNA (Muniyappa and Radding, 1986). It may be that $Red\beta_{\lambda}$ is loaded first, before $SSB_{E.coli}$, as is supposedly the case if it interacts with its cognate exonuclease $Red\alpha_{\lambda}$.

In humans, the core-only RecA XRCC2 protein is thought to act essentially as a recombination accessory factor. Indeed, when present in a complex together with RAD51BCD, it promotes SSA *in vitro* (Yokoyama et al., 2004). We found no evidence of any interaction between Sak4 and RecA *in vivo*, as SSA occurred at similar levels in $RecA^{+}$ or *recA*-deleted strains (not shown). The function of the archaeal RadB protein is not well understood at present, but it presents some activities similar to Sak4. It inhibits the SE activity of RadA, as we found for GST-$Sak4_{PA73}$, when added onto ssDNA prior to $RecA_{E.coli}$ in the reaction. Furthermore, RadB does not degrade ATP, but it binds to it, and hydrolyses without releasing it (Komori et al., 2000). It is possible that Sak4, in the same manner, binds to ATP but does not release it.

Sak4 being essentially truncated for its C-terminal domain, relative to RecA, it may behave as partially deleted RecA mutants. The removal of amino acids from the C-terminal end of $RecA_{E.coli}$ leads to improved SE reactions at alkaline pH (Lusetti et al., 2003). But at pH7, the most C-terminus deleted $RecA_{E.coli}$ no longer promotes SE (Lusetti et al., 2003). This may explain the SE incapacity of Sak4. Indeed, this domain of RecA is thought to be important for stabilizing the three-strands intermediate (De Vlaminck et al., 2012).

The *in vitro* SSA activity of $RecA_{E.coli}$ is inhibited *in vivo*

We report here that $RecA_{E.coli}$ promotes SSA *in vitro*. This was the first demonstrated recombination activity of this recombinase, deduced from nitrocellulose binding assays (Bryant and Lehman, 1985). No trace of SSA activity for $RecA_{E.coli}$ is nevertheless found *in vivo,* in a recombineering assay ((Lopes et al., 2010) and this work). It is usually considered that the high affinity of $SSB_{E.coli}$ for ssDNA shields the substrate for $RecA_{E.coli}$, as we saw effectively *in vitro.* Still, one may have expected that $RecO_{E.coli}$, through its mediator role, would remove $SSB_{E.coli}$ to insert $RecA_{E.coli}$ (Morimatsu and Kowalczykowski, 2003;Handa et al., 2009). Moreover, a recent work in single molecules (Bell et al., 2012) showed that $RecA_{E.coli}$ can nucleate on SSB covered 48 nt ssDNA and drive away $SSB_{E.coli}$ by polymerization. Another possible inhibitor of $RecA_{E.coli}$ annealing *in vivo* is $UvrD_{E.coli}$, a helicase known to disrupt the $RecA_{E.coli}$ filament (Veaute et al., 2005). But even in an *uvrD* mutant, $RecA_{E.coli}$ is prevented from annealing

oligonucleotides at the replication fork (Fig. 2B). At present, we have no clue therefore on the reason why $RecA_{E.coli}$ does not perform recombineering *in vivo*.

We found in addition a remarkable tolerance of $RecA_{E.coli}$ to mismatches *in vitro*, as it annealed oligonucleotides differing by up to 28% of divergence. The tolerance of $RecA_{E.coli}$ to mismatches has already been observed both in strand invasion assays on 13% diverged substrates, *i.e.* 2 mismatches over a 15 bp heteroduplex (Malkov et al., 1997), and a strand exchange assay between M13 and fd molecules (Worth et al., 1994;Tham et al., 2013), but strand invasion and strand exchanges reactions are different, and probably more complex than SSA.

$Red\beta_{\lambda}$ and GST-$Sak4_{PA73}$ tolerate the presence of mismatches during SSA, both *in vitro* and *in vivo*

We reported earlier that several Rad52-like phage recombinases recombine diverged sequences *in vivo*, both during DNA inversion on a phage substrate (Martinsohn et al., 2008), and during exchanges between chromosomal and phage sequences (De Paepe et al., 2014). In addition, $Red\beta_{\lambda}$-dependent recombineering was reported to proceed with diverged oligonucleotides up to 17% (Li et al., 2013). Here, we found *in vivo* annealing for $Red\beta_{\lambda}$ up to 12% diverged sequences, but no longer at 16% (the difference may lie in the different locus used for recombineering, compared to the Li et al. study (Li et al., 2013)). Interestingly, $Sak4_{HK620}$ behaved in a way similar to $Red\beta_{\lambda}$, with respect to divergence *in vivo*. *In vitro*, however, it could not pair 28% diverged oligonucleotides, contrary to $Red\beta_{\lambda}$. Overall, the tolerance to mismatches *in vivo* seems therefore a property shared by two distinct families of phage-encoded recombinases.

As mentioned in the introduction, phages encode three recombinase superfamilies (Lopes et al., 2010). The first and most populated one is the Rad52-like, comprising the Redβ/RecT, Erf and Sak families, in which some members have recognized SSA activity (Kmiec and Holloman, 1981;Datta et al., 2008;Ploquin et al., 2008). The same holds true for the second, Gp2.5 family (Lopes et al., 2010). Here we report that members of the third family of Rad51-like recombinases, $UvsX_{T4}$, $Sak4_{PA73}$, but also $RecA_{E.coli}$ itself, are capable of SSA. Furthermore, it is known that some Rad51 orthologues in Eucaryota and Archaea promote SSA and not SI: this is the case of the BCDX2 complex (Yokoyama et al., 2004), composed by Rad51BCD and the core-only XRCC2, and for RadB of *P. furiousous* (Komori et al., 2000). It may be that the very first activity of recombination of the ancestral RecA was to simply anneal ssDNA, and this activity evolved to a strand invasion activity leading to the homologous recombination as we know it now.

ACKNOWLEDGEMENT

We thank Marianne De Paepe and Xavier Veaute for their insights all along this work and help with the manuscript, and Hadrien Delattre for his help with gene context analysis.

FUNDING

This work was supported by the Agence Nationale de la Recherche [Dynamophage contract].

REFERENCES

Bell, J.C., Plank, J.L., Dombrowski, C.C., and Kowalczykowski, S.C. (2012). Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* 491**,** 274-278. doi: 10.1038/nature11598.

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. doi: 10.1128/JB.186.11.3649-3652.2004.

Bryant, F.R., and Lehman, I.R. (1985). On the mechanism of renaturation of complementary DNA strands by the recA protein of Escherichia coli. *Proc Natl Acad Sci U S A* 82**,** 297-301.

Clark, A.J., Inwood, W., Cloutier, T., and Dhillon, T.S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Mol Biol* 311**,** 657-679. doi: 10.1006/jmbi.2001.4868.

Court, R., Cook, N., Saikrishnan, K., and Wigley, D. (2007). The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. *J Mol Biol* 371**,** 25-33. doi: 10.1016/j.jmb.2007.05.037.

Cox, M.M. (2007). Motoring along with the bacterial RecA protein. *Nat Rev Mol Cell Biol* 8**,** 127-138. doi: 10.1038/nrm2099.

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 U S A* 97**,** 6640-6645.

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 U S A* 105**,** 1626-1631. doi: 10.1073/pnas.0709089105.

De Paepe, M., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S., and Petit, M.A. (2014). Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* 10**,** e1004181. doi: 10.1371/journal.pgen.1004181.

De Vlaminck, I., Van Loenhout, M.T., Zweifel, L., Den Blanken, J., Hooning, K., Hage, S., Kerssemakers, J., and Dekker, C. (2012). Mechanism of homology recognition in DNA recombination from dual-molecule experiments. *Mol Cell* 46**,** 616-624. doi: 10.1016/j.molcel.2012.03.029.

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. doi: 10.1016/j.jmb.2009.06.030.

Forget, A.L., and Kowalczykowski, S.C. (2010). Single-molecule imaging brings Rad51 nucleoprotein filaments into focus. *Trends Cell Biol* 20**,** 269-276. doi: 10.1016/j.tcb.2010.02.004.

Forget, A.L., and Kowalczykowski, S.C. (2012). Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. *Nature* 482**,** 423-427. doi: 10.1038/nature10782.

Godin, S., Wier, A., Kabbinavar, F., Bratton-Palmer, D.S., Ghodke, H., Van Houten, B., Vandemark, A.P., and Bernstein, K.A. (2013). The Shu complex interacts with Rad51 through the Rad51 paralogues Rad55-Rad57 to mediate error-free recombination. *Nucleic Acids Res* 41**,** 4525-4534. doi: 10.1093/nar/gkt138.

Haldenby, S., White, M.F., and Allers, T. (2009). RecA family proteins in archaea: RadA and its cousins. *Biochem Soc Trans* 37**,** 102-107. doi: 10.1042/BST0370102.

Handa, N., Morimatsu, K., Lovett, S.T., and Kowalczykowski, S.C. (2009). Reconstitution of initial steps of dsDNA break repair by the RecF pathway of E. coli. *Genes Dev* 23**,** 1234-1245. doi: 10.1101/gad.1780709.

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 U S A* 98**,** 9557-9562. doi: 10.1073/pnas.171317698.

Kmiec, E., and Holloman, W.K. (1981). Beta protein of bacteriophage lambda promotes renaturation of DNA. *J Biol Chem* 256**,** 12636-12639.

Komori, K., Miyata, T., Diruggiero, J., Holley-Shanks, R., Hayashi, I., Cann, I.K., Mayanagi, K., Shinagawa, H., and Ishino, Y. (2000). Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. *J Biol Chem* 275**,** 33782-33790. doi: 10.1074/jbc.M004557200.

Kowalczykowski, S.C., and Krupp, R.A. (1987). Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J Mol Biol* 193**,** 97-113.

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Res* 40**,** 5795-5818. doi: 10.1093/nar/gks270.

Krogh, B.O., and Symington, L.S. (2004). Recombination proteins in yeast. *Annu Rev Genet* 38**,** 233-271. doi: 10.1146/annurev.genet.38.072902.091500.

Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* 63**,** 751-813, table of contents.

Li, X.T., Thomason, L.C., Sawitzke, J.A., Costantino, N., and Court, D.L. (2013). Bacterial DNA polymerases participate in oligonucleotide recombination. *Mol Microbiol* 88**,** 906-920. doi: 10.1111/mmi.12231.

Liu, J., Ehmsen, K.T., Heyer, W.D., and Morrical, S.W. (2011). Presynaptic filament dynamics in homologous recombination and DNA repair. *Crit Rev Biochem Mol Biol* 46**,** 240-270. doi: 10.3109/10409238.2011.576007.

Liu, J., and Morrical, S.W. (2010). Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol J* 7**,** 357. doi: 10.1186/1743-422X-7-357.

Lo Piano, A., Martinez-Jimenez, M.I., Zecchi, L., and Ayora, S. (2011). Recombination-dependent concatemeric viral DNA replication. *Virus Res* 160**,** 1-14. doi: 10.1016/j.virusres.2011.06.009.

Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M.A., and Guerois, R. (2010). Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* 38**,** 3952-3962.

Lusetti, S.L., Wood, E.A., Fleming, C.D., Modica, M.J., Korth, J., Abbott, L., Dwyer, D.W., Roca, A.I., Inman, R.B., and Cox, M.M. (2003). C-terminal deletions of the Escherichia coli RecA protein. Characterization of in vivo and in vitro effects. *J Biol Chem* 278**,** 16372-16380. doi: 10.1074/jbc.M212917200.

Maher, R.L., and Morrical, S.W. (2013). Coordinated Binding of Single-Stranded and Double-Stranded DNA by UvsX Recombinase. *PLoS One* 8**,** e66654. doi: 10.1371/journal.pone.0066654.

Malkov, V.A., Sastry, L., and Camerini-Otero, R.D. (1997). RecA protein assisted selection reveals a low fidelity of recognition of homology in a duplex DNA by an oligonucleotide. *J Mol Biol* 271**,** 168-177. doi: 10.1006/jmbi.1997.1164.

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A.F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* 11**,** 54. doi: 1471-2199-11-54 [pii]

Martinsohn, 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.

Morimatsu, K., and Kowalczykowski, S.C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* 11**,** 1337-1347.

Muniyappa, K., and Radding, C.M. (1986). The homologous recombination system of phage lambda. Pairing activities of beta protein. *J Biol Chem* 261**,** 7472-7478.

Murphy, K.C. (2000). Bacteriophage P22 Abc2 protein binds to RecC increases the 5' strand nicking activity of RecBCD and together with lambda bet, promotes Chi-independent recombination. *J Mol Biol* 296**,** 385-401. doi: 10.1006/jmbi.1999.3486.

Murphy, K.C. (2007). The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J Mol Biol* 371**,** 19-24. doi: 10.1016/j.jmb.2007.05.085.

Muyrers, J.P., Zhang, Y., Buchholz, F., and Stewart, A.F. (2000). RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* 14**,** 1971-1982.

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. doi: 10.1016/j.cub.2008.06.071.

Pullman, M.E., Penefsky, H.S., Datta, A., and Racker, E. (1960). Partial resolution of the enzymes catalyzing oxidative phosphorylation. I. Purification and properties of soluble dinitrophenol-stimulated adenosine triphosphatase. *J Biol Chem* 235**,** 3322-3329.

Rehrauer, W.M., and Kowalczykowski, S.C. (1993). Alteration of the nucleoside triphosphate (NTP) catalytic domain within Escherichia coli recA protein attenuates NTP hydrolysis but not joint molecule formation. *J Biol Chem* 268**,** 1292-1297.

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. doi: 10.1074/jbc.M303374200.

Rice, K.P., Eggler, A.L., Sung, P., and Cox, M.M. (2001). DNA pairing and strand exchange by the Escherichia coli RecA and yeast Rad51 proteins without ATP hydrolysis: on the importance of not getting stuck. *J Biol Chem* 276**,** 38570-38581. doi: 10.1074/jbc.M105678200.

Seitz, E.M., Brockman, J.P., Sandler, S.J., Clark, A.J., and Kowalczykowski, S.C. (1998). RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev* 12**,** 1248-1253.

Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43**,** 289-318. doi: 10.1080/10409230802341296.

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.

Subramanian, K., Rutvisuttinunt, W., Scott, W., and Myers, R.S. (2003). The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Res* 31**,** 1585-1596.

Tham, K.C., Hermans, N., Winterwerp, H.H., Cox, M.M., Wyman, C., Kanaar, R., and Lebbink, J.H. (2013). Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures. *Mol Cell* 51**,** 326-337. doi: 10.1016/j.molcel.2013.07.008.

Van Der Heijden, T., Modesti, M., Hage, S., Kanaar, R., Wyman, C., and Dekker, C. (2008). Homologous recombination in real time: DNA strand exchange by RecA. *Mol Cell* 30**,** 530-538. doi: 10.1016/j.molcel.2008.03.010.

Van Der Heijden, T., Seidel, R., Modesti, M., Kanaar, R., Wyman, C., and Dekker, C. (2007). Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Res* 35**,** 5646-5657. doi: 10.1093/nar/gkm629.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. *EMBO J* 24**,** 180-189. doi: 10.1038/sj.emboj.7600485.

Worth, L., Jr., Clark, S., Radman, M., and Modrich, P. (1994). Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci U S A* 91**,** 3238-3241.

Yokoyama, H., Sarai, N., Kagawa, W., Enomoto, R., Shibata, T., Kurumizaka, H., and Yokoyama, S. (2004). Preferential binding to branched DNA strands and strand-annealing activity of the human Rad51B, Rad51C, Rad51D and Xrcc2 protein complex. *Nucleic Acids Res* 32**,** 2556-2565. doi: 10.1093/nar/gkh578.

TABLE AND FIGURES LEGENDS

Table 1 : Strains

strain	genetic background	plasmid	recombinational genes on plasmid	Origin
Map1798	Map1798	pJA17	*hkaM*$_{HK620}$ *sak4*$_{HK620}$ ssb$_{HK620}$ *abc2*$_{HK620}$	This study
Map1801	Map1798	pKD46	*orf60a*$_{\lambda}$ *redα*$_{\lambda}$ *redβ*$_{\lambda}$ *gam*$_{\lambda}$	(34)
Map1802	AB1157 *recA mutL* *cat*$_{ss}$	none		This study
Map1879	Map1798	pJA4	none	(Lopes et al., 2010)
Map1881	Map1798	pGH3	*sak4*$_{HK620}$	This study
Map1883	Map1798	pJA12	*sak4*$_{PA73}$	(Lopes et al., 2010)
Map1894	Map1798	pJA12	*redβ*$_{\lambda}$	This study
Map1895	Map1798	pGH19	*sak4*$_{HK620}$ ssb$_{HK620}$	This study
Map1897	Map1798	pGH20	*sak4*$_{HK620\ 0}$ *abc2*$_{HK620}$	This study
Map1922	Map1798	pGH21	ssb$_{HK620}$	This study
Map1931	Map1798	pGH24	*sak4*$_{PA73}$ ssb$_{PA73}$	This study

Table 2 : Oligonucleotides for *in vivo* recombineering and *in vitro* SSA assays. Red letters highlight the mismatches relative to the cat_{SS} gene

Name	N of mismatches	Divergence relative to *catss*	Sequence
GO34	0	0 %	CAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATAGTATACGTAAGGCG ACAAGGTGCTGATGCCGCTGGCGATTCAGGT
Map98	2	2 %	CAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCG ACAAGGTGCTGATGCCGCTGGCGATTCAGGT
Map99	5	6 %	CAACTTCTTCGCGCCCGTTTTCACTATGGGCAAATATTATACGCAAGGCG ACAAGGTGCTGATGCCCCTGGCGATTCAGGT
Map100	10	12 %	CAACTTTTTCGCGCCCGTCTTCACTATGGGCAAATATTATACGCAAGGCG ATAAGGTGCTCATGCCCCTGGCCATTCAGGT
Map101	16	20 %	CAACTTTTTTGCGCCCGTCTTTACTATGGGGAAATATTATACGCAAGGGG ATAAGGTCCTCATGCCCCTCGCCATTCAGGT
Map102	23	28 %	CAATTTTTTTGCGCCGGTCTTTACTATGGGGAAGTATTACACGCAAGGGG ATAAAGTCCTCATGCCCCTCGCCATCCAAGT

Supplementary material

Choice of the recombineering assay

The recombineering assay that we conducted previously (1) was based on the selection of Rifampicin resistant (RifR) colonies produced by the recombineering of a 51nt-long oligonucleotide complementary to the rpoB gene. Several improvements were made in the present study. In the former protocol, competent cells were prepared in large batches, frozen at -80°C in aliquots, and used repeatedly. In addition, once electroporated and let grown 1hour for expression, cells were further incubated overnight on the bench, prior to titrations on plates. These two factors proved on the long range to be responsible for variations in cell viability, and hence estimations of recombineering efficiencies were not always reproducible. Following the protocol recommended by Dr Court, competent cells were prepared freshly, and the overnight incubation on the bench was suppressed. However, this last change reduced drastically the yield of RifR colonies, suggesting that 1h expression time was not sufficient to allow the plain recovery of recombinants. To increase the expression time without letting grow cells in liquid (as expected if the expression time is extended), cells were plated after the 1h expression period embedded into top agar on non-selective LB plates, and let grown 2 more hours at 37°C. At this stage only, the rifampicin selection was applied, by a top-agar overlay. This (old) technique proved to increase by a factor of 100 the yield of RifR colonies. The top-agar technique also permitted a better recovery of viable cells (that were in the range of 10^9-/ml, rather than 10^8 otherwise). However, this new protocol led to an unexpectedly high "background" of recombineering with the recombinase-less, control plasmid pJA4: this level, which was estimated at 5 10^{-7} spontaneous recombinants per viable cell in Lopes et al., reached now 2 10^{-6}. This high background hindered the detection the recombineering activity of the $Sak4_{PA73}$ recombinase (1.5 10^{-6} in Lopes et al., 2 10^{-6} here).

To reduce this background level, and also in perspective of the experiments we wanted to conduct in a *mutS* background (to use degenerated oligonucleotides), a new selection system was constructed. Based on the cat gene, two points were mutated to introducing stop codons, next to each other in *cat* gene (strain MAC1802, see construction details below). The *cat* cassette (together with a KanR marker gene) was introduced at the vicinity of *rpoB*, within the *thi* operon, so as to delete *thiE* and *thiF* (the *thiE* gene is already mutated in AB1157). Spontaneous recombineering at *cat* was low (5 10^{-9}). However, recombineering with pKD46 was reduced by a factor of 200 at this locus, compared to the *rpoB* locus: 4.8 10^{-6} CmR/viable cell, for 10^{-3} RifR/viable cell. Comparison of recombineering efficiencies obtained at these two loci is shown in the Table below. Another set of recombineering assays were performed into strain G205, a derivative of HME57 (a strain used in the Court laboratory, which is *galK*) rendered *recA*, so as to compare recombineering efficiencies at the *rpoB* and *galK* loci, in a MutS+ background (both oligonucleotides generate single C:C mispairs, that are not recognized by MutS, (3)). For the Gal^+ selection, no top agar overlay is used (to follow Court's lab protocol), and viable cells are counted on the same M9 agar medium, supplemented with glucose 0.2%.

	Gal+	RifR	CmR
G205 pJA4	1.1 (+/- 0.9) 10^{-5}	2 10^{-6}	
G205 pKD46	2.7 (+/- 1) 10^{-3}	1 10^{-3}	
G205 pKDsak4ssb	3.8 (+/- 3.3) 10^{-4}	2 10^{-4}	
MAC1802 pJA4		1,3 (+/- 0.1) 10^{-5}	4.4 (+/-2.8) 10^{-9}
MAC1802 pKD46		9,1 (+/- 6) 10^{-4}	5.1 (+/-4) 10^{-6}
MAC1802 pKDsak4ssb		1,0 (+/- 0.9) 10^{-4}	1.8 (+/- 1.3) 10^{-6}

Construction of MAC1802 strain

MAC1802 is a strain containing in its chromosome a mutated *cat* gene. Site directed mutagenesis, to introduce two stop codons separated by 6nt into the *cat* gene of pACYC184 (cat_{SS} allele), was performed with the help of oligonucleotides J85 and J86 (supplementary table 1), that are complementary. The two stop codons generate G:A and T:G mispairs during the recombineering experiment using oligonucleotide Maj115, which has the polarity of a lagging strand. Oligonucleotides J84 and J87 flank the cat gene of pACYC184 (4) and overlap two BmgBI sites of this plasmid. The two halves of the mutated *cat* gene were amplified with the J84-85 and J86-87 pairs (652bp and 243bp fragments, respectively), and the full cat_{SS} mutated gene was then produced using these two PCR fragments as initial primers, as well J84 and J87. The full gene fragment was then cloned into the BmgBI site of pKD4 (2). This cat_{SS} mutated gene, together with the KanR marker of pKD4 were next introduced by recombineering into the *thi* operon (10kbp away from *rpoB*), in place of *thiE* and *thiF*, and in the same orientation, of strain JJC40 (AB1157 *hsdR*), using oligonucleotides Maj80 and Maj81. The resulting cassette was next transduced by P1 into AB1157 (strain MAC1716). The *mutS*Ω (specR) allele (5) was next introduced by P1 transduction in this strain (MAC1774), and finally the Δ*recA*306 *srl*:*Tn*10 allele (from strain GY5902, a gift of Dr. S. Sommer), to give strain MAC1802.

Plasmid constructions

Constructions of all plasmid are described in supplementary Table 2. Unless otherwise stated, plasmids (column 1) were constructed by the cloning of a PCR product (columns 2 to 5) into a plasmid (column 6 and 7). The gene amplified is indicated in the second column. PCR were performed with the high fidelity polymerase Phusion, using the DNA matrix indicated in third column, and the oligonucleotides indicated in columns 4 and 5. The vector (column 6) and the PCR product are digested by the enzyme stipulated in column 7. Brief plasmid description in column 8.

Supplementary references

1. Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M.A. and Guerois, R. (2010) Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res*, **38**, 3952-3962.
2. 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 U S A*, **97**, 6640-6645.
3. Brown, J., Brown, T. and Fox, K.R. (2001) Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. *The Biochemical journal*, **354**, 627-633.
4. Chang, A.C. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol*, **134**, 1141-1156.
5. Brégeon, D., Matic, I., Radman, M. and Taddei, F. (1999) Inefficient mismatch repair: genetic defects and down regulation. *Journal of Genetics*, **78**, 211-228.

Supplementary tables

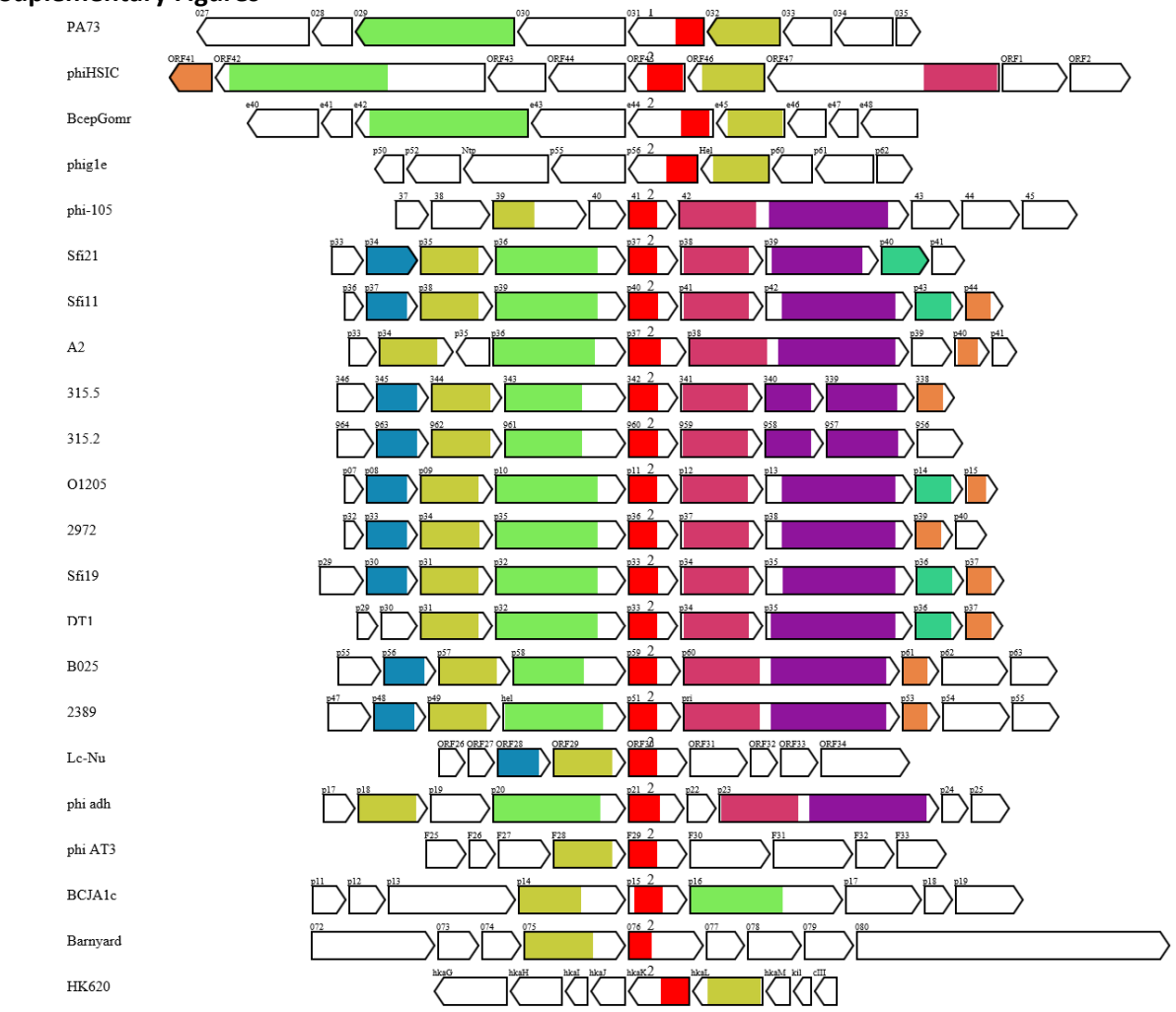
Supplementary Table 1: Oligonucleotides for plasmid and strain constructions.

name	sequence
GO18	ACCACTCCATGGGAATGAGTACTGCACTCGCAAC
GO20	ACCACTCCATGGGAATGTCTGATTTAAAATCTCGTTTGATTAAAGCTTCTAC
GO22	ACCACTCCATGGGAATGGGAACTGCGACATTAAT
GO24	ACCACTCCATGGGAATGCAAATGTCGCAATTAAA
GO26	CAAATTGGATCCTATGCTGCCACCTTCTGCTCTG
GO27	CAAATTGGATCCTATTCATTAAATTCTTCCATATCACTTAGCTGTTCGAGGTC
GO28	CAAATTGGATCCTATGCGGCGTTTTCCTTAATTT
GO29	CAAATTGGATCCTATTGCATCGCTTTCGCGAACA
GO32	GCAAGCTTGTCGACGGAGCT
GO33	CGGATGGATGAGCGAGAATC
GO48	ACCACTCCATGGGAATGAGCAACGTGATTTTTAC
GO49	CAAATTGGATCCTAGAAAGGAGGGTAATCGTTCT
GO50	ACCACTCCATGGGAATGCAACTTATCCAAGCGTT
GO51	CAAATTGGATCCTAACGCTGGCCCCAAGGTGCCG
GO56	TTAACTTTAAGAAGGAGATA
GO58	TTCGGATCCTATTGCATCGC
J55	GGAATTCCGGAGGTATAAAACGAATGAGT
J56	CATGCCATGGCATGTCATGCTGCCACCTTCTGCT
J67	CATGCCATGGCATGTCAGTCATTACTGATAGCGC
J68	CCGGAATTCGGAGGCAAAATGATTCCGGTAGAACTGGC
J80	CCGGAATTCGGAGGATGGGAACTGCGACATTAAT
J81	ATGCCATGGTCATGCGGCGTTTTCCTTAA
J84	GACGTGCACGTAAGAGGTTCCAACTTT
J85	GCACCTTGTCGCCTTACGTATACTATTTGCCCATGGT
J86	ACCATGGGCAAATAGTATACGTAAGGCGACAAGGTGC
J87	CACGTCCTTTCGAATTTCTGCCATTCAT
Maj40	TTGCACGGCGTCACACTTTG
Maj41	CAGTGAAGCATCAAGACTAAC
Maj80	ATGTATCAGCCTGATTTTCCTCCTGTACCTTTTCGTTCAGGACTGTACGCTTGCATGCAGATTGCAGC
Maj81	TTAAACAGGATCTGCATTGCTTCCTCCGCATACCGGGCAACCACTGGCCCATATGAATATCCTCCTTA
Maj110	CCATTACCATGGTCAGAAAGGAGGGTAATCGTT
Maj111	CCATTACCATGGATTGTCATTACTGATAGCGCCATA
Maj112	ATGGCACCATGGTACTGAAAATAAGGCTCCCATT
Maj113	CCGGAATTCGGAGGCAAATTATGGGAACTG
Maj115	AAGTCGCGGTCGGAACCGTATTGCAGCAGCTTTACCATCTGCCGCTGGACGGCGCACAAATCGCGCTTAA
Maj116	CCGGAATTCGGAGGACGCCGCATGAGCAACGTG

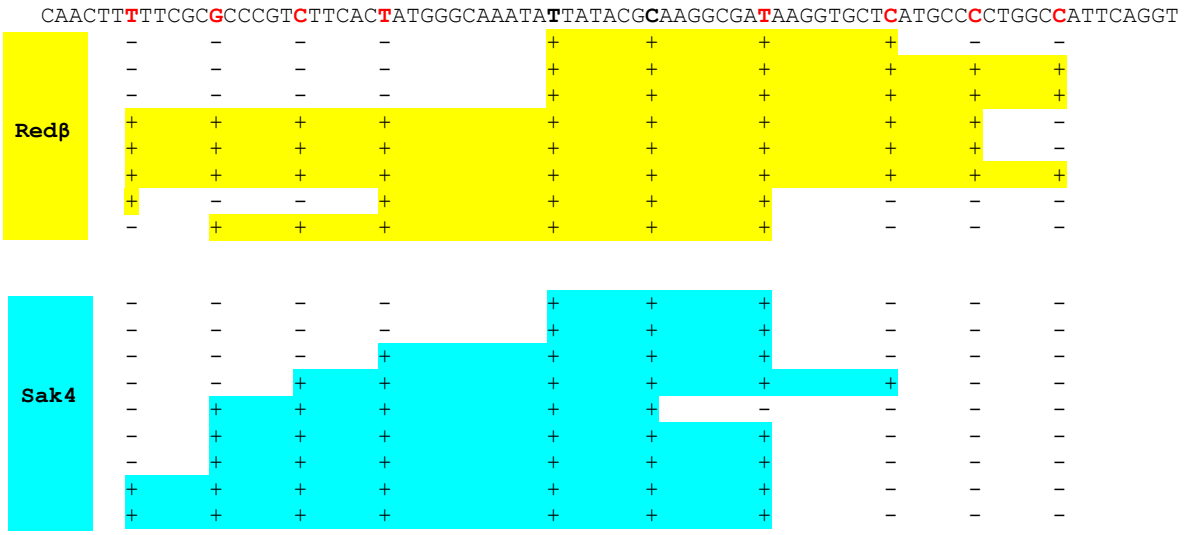
Supplementary Table 2: Plasmid constructions

	PCR for insert				vector			
plasmid name	gene amplified by PCR	DNA matrix	oligo 1	oligo 2	original plasmid	digested by	origin	plasmid description
pGH3	*sak4*$_{HK620}$	HK620	J80	J81	pKD46	EcoRI and NcoI	this study	Para: *sak4*$_{HK620}$
pGH5	*uvsX*$_{T4}$	T4	GO20	GO27	pETM30	NcoI and BamHI	this study	Plac: GST-*uvsX*$_{T4}$
pGH7	*sak4*$_{HK620}$	HK620	GO22	GO28	pETM30	NcoI and BamHI	this study	Plac:GST-*sak4*$_{HK620}$
pGH8	*sak4*$_{PA73}$	PA73	GO24	GO29	pETM30	NcoI and BamHI	this study	Plac:GST-*sak4*$_{PA73}$
pGH11	*redβ*$_{λ}$	pKD46	GO18	GO26	pETM30	NcoI and BamHI	this study	Plac:GST-*redβ*$_{λ}$
pGH15	*ssb*$_{HK620}$	HK620	GO48	GO49	pETM30	NcoI and BamHI	this study	Plac:GST-*ssb*$_{HK620}$
pGH16	*ssb*$_{PA73}$	PA73	GO50	GO51	pETM30	NcoI and BamHI	this study	Plac:GST-*ssb*$_{PA73}$
pGH19	*sak4+ssb*$_{HK620}$	pJA17	Maj110	Maj113	pKD46	NcoI and BamHI	this study	Para:*sak4+ssb*$_{HK620}$
pGH20	*abc2*$_{HK620}$	pJA17	Maj111	Maj112	pGH7	NcoI	this study	Para:*sak4+abc2*$_{HK620}$
pGH21	*ssb*$_{HK620}$	HK620	Maj41	Maj116	pKD46	EcoRI and NcoI	this study	Para:*ssb*$_{HK620}$
pGH22	*sak4-GST*$_{HK620}$	pGH8	GO56	GO58	pKD46	EcoRI and NcoI	this study	Para:*sak4-GST*$_{HK620}$
pGH23					pGH19 pGH22	AlwNI AlwNI	this study	Para:*sak4-GST+ssb*$_{HK620}$
pJA12							(1)	Para:*sak4*$_{PA73}$
pJA16	*redβ*$_{λ}$	pKD46	J55	J56	pKD46	EcoRI and NcoI	this study	Para:*redβ*$_{λ}$
pJA17	*module*$_{HK620}$	HK620	J67	J68	pKD46	EcoRI and NcoI	this study	Para:*module*$_{HK620}$
pJA4							(1)	None
pKD46							(2)	Para:*module*$_{λ}$

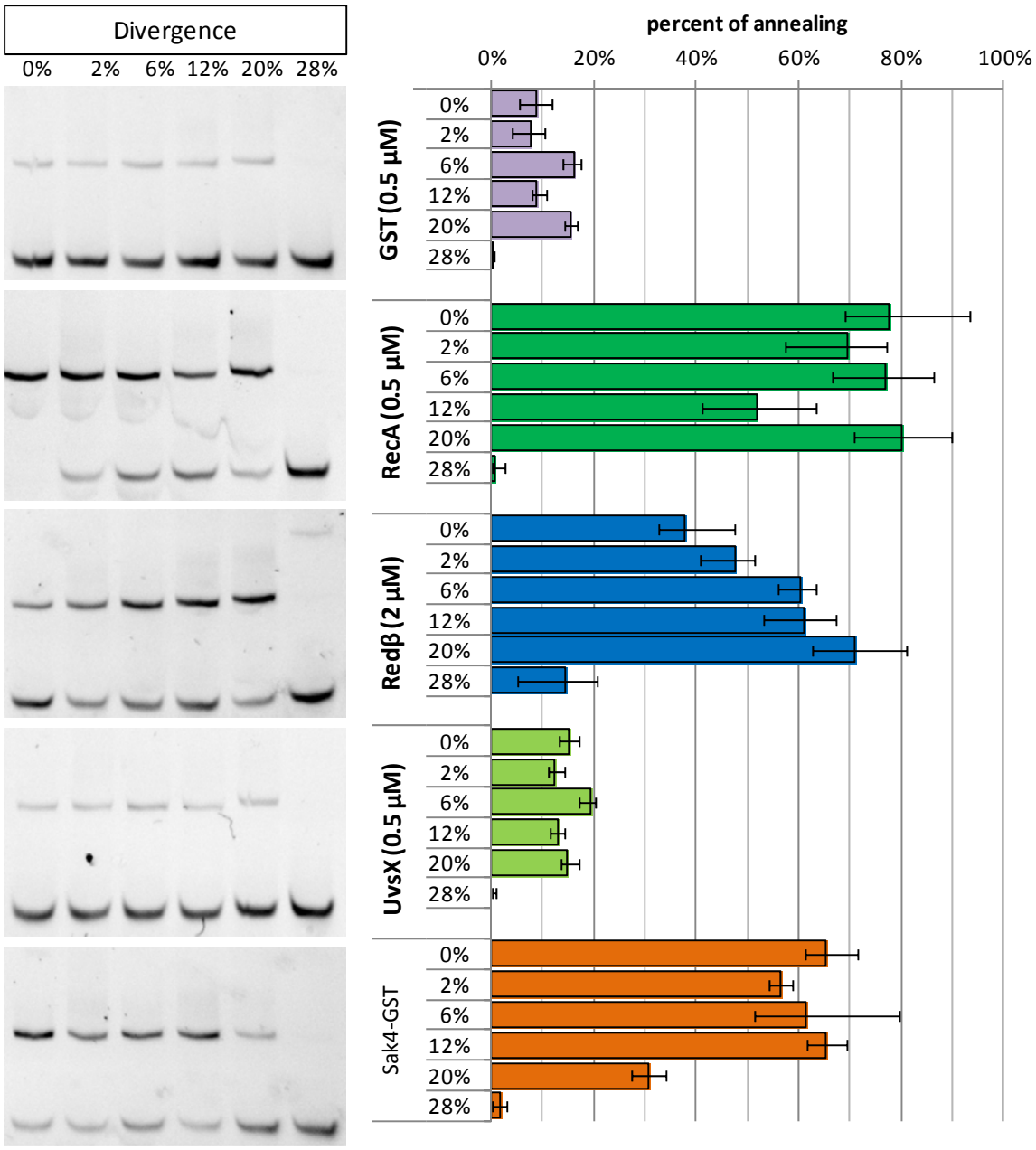
Suplementary Figures



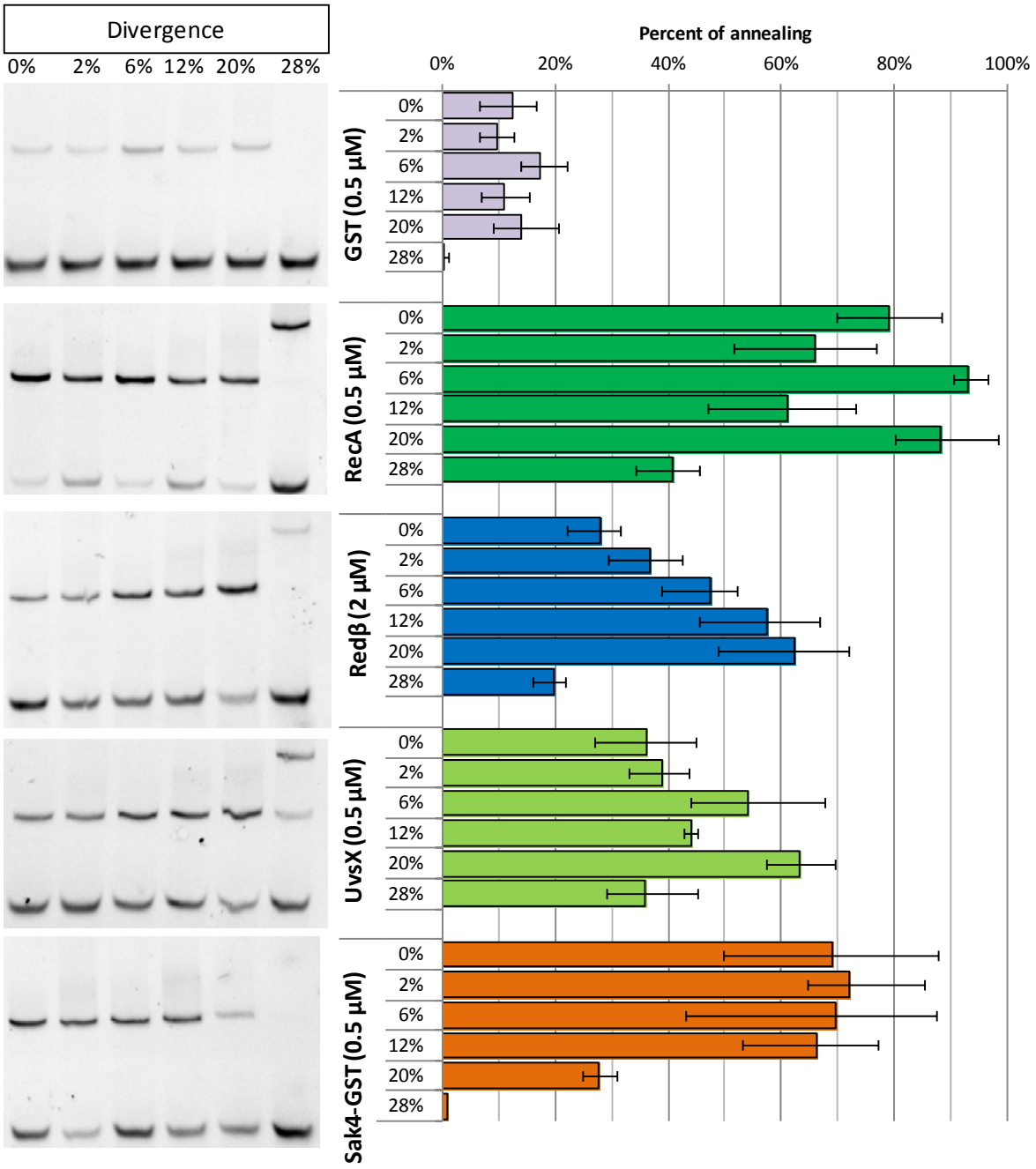
Supplementary Figure 1: Genome context for the SSB_{HK620} protein: Phage homologs of SSB_{HK620} (in red) were searched using HHsearch, and those retrieved with a 95% probability cutoff are shown. The same search was performed for the 4 genes left and right off all SSB homologs that had been retrieved, to inspect local co-occurences. Sak4 homologs are coloured in yellow. Under the figure are the annotation taken from PFAM, ACLAME or GENBANK.



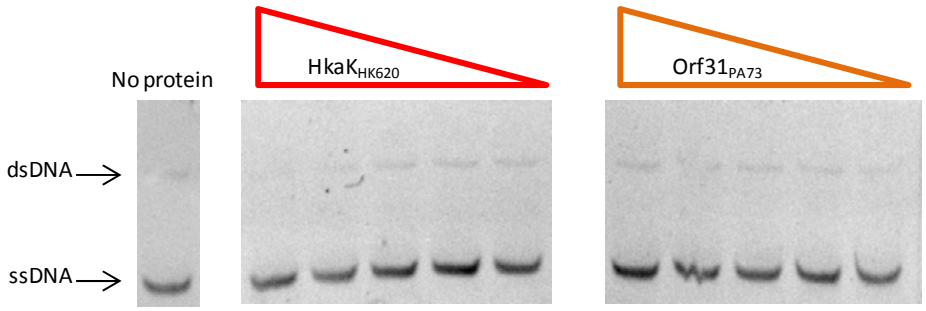
Supplementary Figure 2: Sequence analysis of ~10 recombinants obtained by ssDNA recombineering *in vivo*, with Redβ (upper, blue sequences) and Sak4$_{HK620}$ (lower, yellow sequences) using oligonucleotide map100 (12% diverged, relative to the catss sequence). On top of the figure is the sequence of map100 with the central T and C in bold letters showing the two mismatches permitting correction of the 2 stop codons in catss. The red bases correspond to silent mutations introduced into map100 to create a 12% diverged sequence, relative to catss. + indicates mutations present in map100 that have been incorporated into the CmR recombinants obtained, - : mutations that were not incorporated.



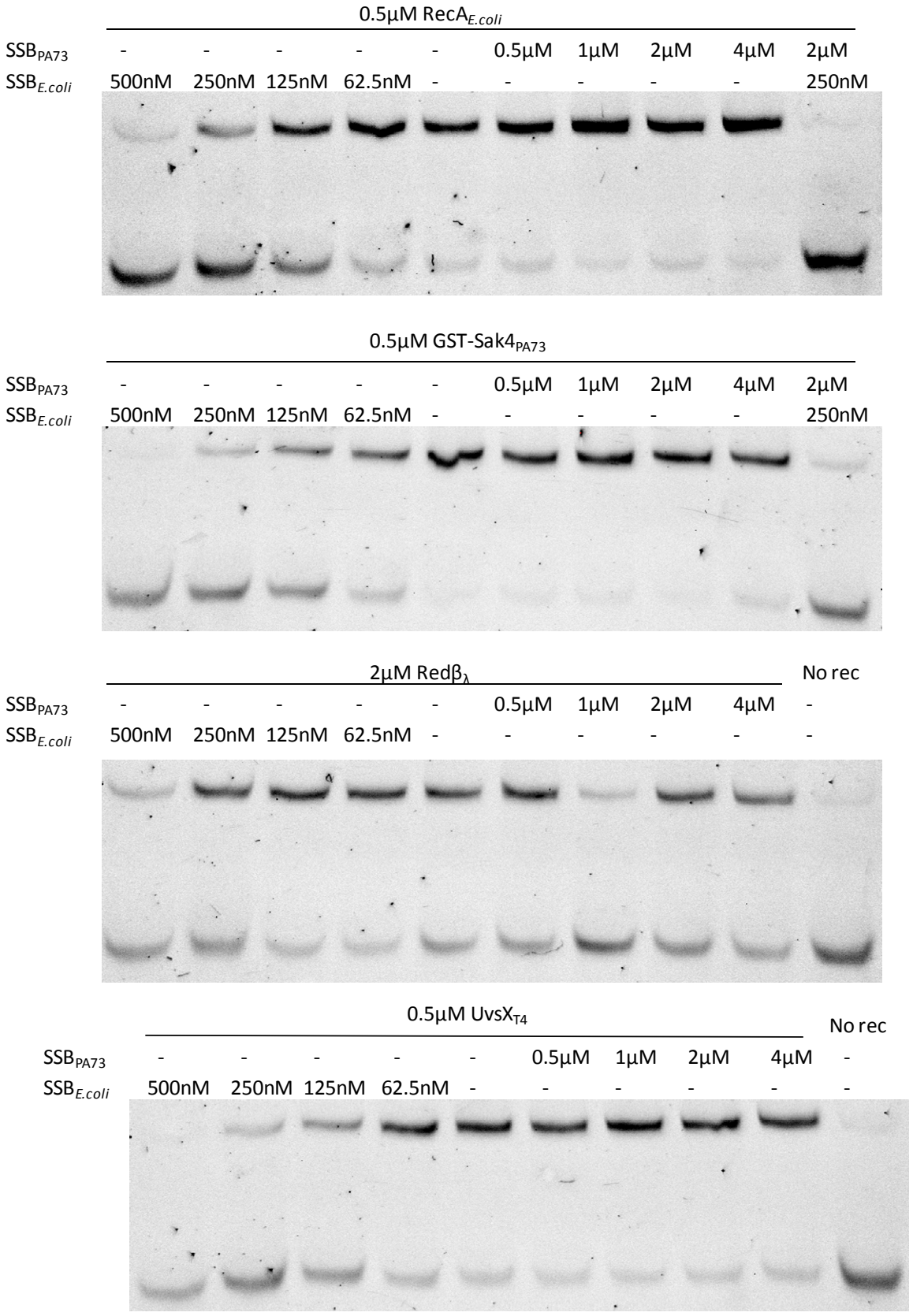
Supplementary Figure 3: SSA of all recombinases used in this study, and of GST, in the absence of ATP. A representative gel is shown for each recombinase on the left of the figure (time point 7 min). On the right part, the mean (3 replicates) percents of annealing were quantified for each reaction. The error bars represent the minimum and maximum annealing measured.



Supplementary Figure 4: SSA of all recombinases used in this study, and of GST, with 1mM ATP. A representative gel is shown for each recombinase on the left side of the figure (time point 7 min). On the right, the mean (3 replicates) percents of annealing quantified for each reaction. The error bars represent the minimum and maximum annealing measured.Supplementary.



Supplementary Figure 5: phage SSB proteins do not promote SSA. 10µM of complementary oligonucleotides were mixed with different concentration of SSB protein for 15 minutes at 30°C. SSB_{HK620} was used at decreasing concentrations of 8µM, 4µM, 2µM, 1µM and 0.5µM. SSB_{PA73} was used at 12µM, 6µM, 3µM, 1.5µM and 0.75µM.



Supplementary Figure 6: Inhibition of the SSA by the various SSB. The effect of SSB on each of the four recombinases was tested. Inhibition by SSB_{Ecoli} is strong, whereas that of SSB_{PA73} is absent.

Chapter IV: Discussion

I) Mosaic acquisition: causes and consequences

As explained in the introduction, temperate phage genomes are mosaic. Here I will discuss the contribution of homologous recombination to this mosaïcism as well as the consequences for both host and phage genomes. I will also address the question of phage classification and the notion of species under the homologous recombination point of view.

1) Phage recombinases, a major cause of mosaïcism

Phage recombinases seem to play an important role in mosaïcism. It was proposed that their relaxed fidelity during recombination might participate to mosaic acquisition in *Enterobacteria* phages, which principally encode Rad52-like recombinases (Martinsohn, Radman, Petit 2008). Interestingly, the prophages encoding a recombinase have more mosaic genomes than the others (Bobay, Rocha, Touchon 2013). These phage recombinases are grouped into three super-families: the Rad52-like, the Rad51-like and the Gp2.5-like (Lopes et al. 2010). Each of these families has its particularities and its restrictions, which are discussed below.

A) Rad52-like recombinases are permissive annealases.

The main activity of these recombinases is SSA, they cannot invade a dsDNA to search for homology, unless under particular conditions. The most studied recombinase of this super-family is Redβ, from the phage λ, which was used as a representative model during my thesis. We have shown that this recombinase, as well as other members of this family, recombines diverged sequences both *in vivo* and *in vitro* up to 20% of divergence ((Martinsohn, Radman, Petit 2008; Li et al. 2013), Chapter II and III).

As its main activity is SSA, the generally accepted hypothesis is that its action resides at the replication fork. Indeed, the only ssDNA present in a bacterium and protected from degradation is the template of the lagging strand at the replication fork. In agreement with this hypothesis, when DnaG, the primase that polymerizes RNA primers for the Okazaki fragments, is depleted, the free ssDNA at the replication fork is longer and the efficiency of recombination by Redβ is higher (Lajoie et al. 2012). Furthermore, mutations into the gene encoding Pol I, the polymerase that degrades RNA primers to polymerize DNA, and Pol III, the polymerase that polymerizes DNA from primers, decrease the efficiency of Redβ mediated recombination (Li et al. 2013; Poteete 2013). This property of annealing at the replication fork is taken into account when designing oligonucleotides for single

strand recombineering reactions *in vivo*. The aim of this technique is to anneal an oligonucleoide on the lagging strand template to bring a point mutation by assimilation of this oligonucleotide (Costantino, Court 2003). However, in the real life, Redβ and other recombinases of this super-family are able to recombine DNA fragments of much longer size than oligonucleotides. For example, in the Chapter II, we saw that several phage Rad52 recombinases are able to capture genes from the bacterial chromosome and incorporate them into the phage genome or recombine DNA from different phage genomes. How this happens exactly is not understood yet, but based on the idea that Redβ anneals at the replication fork, Poteete proposed a model of polymerase slippage where a ssDNA 3' overhang is wrapped around a Redβ ring and anneals on the lagging strand template of the replication fork. The polymerase of the leading strand slips to the annealed DNA and continues the DNA polymerization (see Figure6B in Chapter I, (Poteete 2008)). However, this model explains only half of the gene capture process as observed in Chapter II. Indeed, to complete the process, the replication fork has to perform two events of slippage promoted by Redβ in a range of 1 kb to capture a single gene.

Another model is more attractive and fits with these results. First, the dsDNA has to be prepared for integration by being turned into a ssDNA complementary to the lagging strand template of DNA replication fork. Then, the ssDNA fragment can be integrated by annealing of its two extremities, regardless of the sequence in between (see Figure 15 in Chapter I, (Maresca et al. 2010)). To explain gene capture with this model, we need to suppose that the bacterial chromosome has been truncated into pieces of dsDNA, and then converted into ssDNA by exonuclease digestion. These pieces of ssDNA would then anneal at the replication fork by Redβ as in the Maresca et al. model. How bacterial DNA is broken down to pieces during the phage replication is yet unknown.

However, this model is not sufficient to explain all our results. Indeed, the maximal length of the free ssDNA at the replication fork is equivalent to the length of Okazaki fragments, which measure between 1 and 2 kb, unless DnaG is mutated (Tougu, Marians 1996). DNA fragments of the size of a gene, as antibiotic resistance genes in Chapter II, or smaller, as the predominant little mosaics observed by bioinformatics in Chapter II, are probably exchanged this way. However, in the promoter inversion model (Chapter II), the fragment is somewhat longer (1.5 kb to 2 kb) and may be difficult to recombine this way. Furthermore, an exonuclease has to digest 2 kb to obtain a ssDNA substrate sufficient to be annealed at the replication fork (as proposed in Maresca et al. model), and induce promoter inversion.

Redβ is also capable of promoting some level of SI. First, it promotes a certain type of strand exchange (Figure 19): if the oligonucleotide has paired partially to a single strand template, the

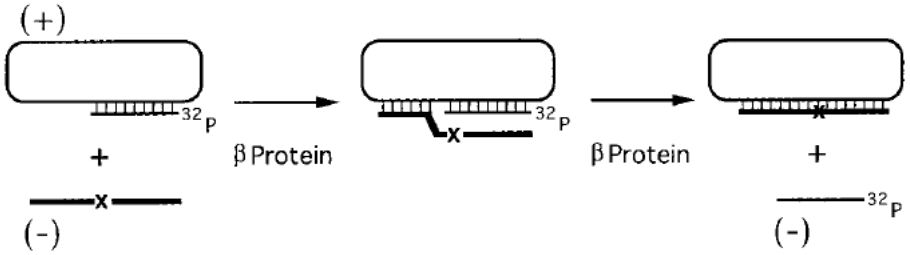


Figure 19: Design of the strand exchange reaction promoted by Redβ. A labelled 43-mer oligonucleotide that had been thermally annealed to single strand circular M13 DNA to form partial duplex was displaced in the presence of Redβ. By an overlapping 63-mer, called the donnor strand, that had 20 additional nucleotide residues entirely at its 3' or 5' end, relative to the 43-mer. (Li et al. 1998)

remaining duplex DNA is exchanged by the incoming oligonucleotide. This reaction has clear preference for a 5' first annealing (Li et al. 1998). Second, Redβ permits strand invasion under certain circumstances: at low magnesium concentration, the DNA breathes by repulsion of negative charges of the two DNA strands and Redβ, as RecT, could find the complementary strand during the breathing (Noirot, Kolodner 1998; Rybalchenko et al. 2004). Such breathing exists when DNA is supercoiled, and Redβ probably finds homology this way when it is possible. Other enzymes facilitating dsDNA opening, such as helicases could also play a role. Maybe some accessory factors of Redβ remain to be found. In favour of a SI mechanism, some recombinants of the phage promotor inversion assay annalysed by Martinsohn et al. have symetrical end points in the inverted repeat (Martinsohn, Radman, Petit 2008). Indeed, the SI mechanism leads to a four branches Holliday junction that gives rise to symmetrical resolution products.

Another mechanism may be MHEJ. As discussed in the introduction (Chapter I), a phage Rad52 recombinase, RecT, promotes this type of events with a frequency of 2.10^{-7} during phage induction (Shiraishi et al. 2002; Shiraishi et al. 2006). Furthermore, a group identified a major deletion (126 genes) in the bacterial chromosome of their *E. coli* strain during a recombineering experiment with Redβ (Vine et al. 2010). The sequencing of the deletion point revealed that at the recombination junction, only 1 nt was identical. 53% identity on a 17 bp region surrounding the junction was present (unpublished results of the team, Figure 20). This observation lets us think that Redβ has the same ability as RecT, and probably more phage Rad52 recombinases. MHEJ may explain a part of the non-symmetrical promoter inversion observed by Martinsohn et al. (Martinsohn, Radman, Petit 2008).

Redβ recombines diverged sequence, but to do so, it must (1) pair sequences with mismatches and (2) resist to the MMR system. We saw in Chapter III that Redβ anneals DNA of sequences up to 20% of divergence as efficiently as perfectly identical DNA *in vitro*. The Court group analyzed deeper this

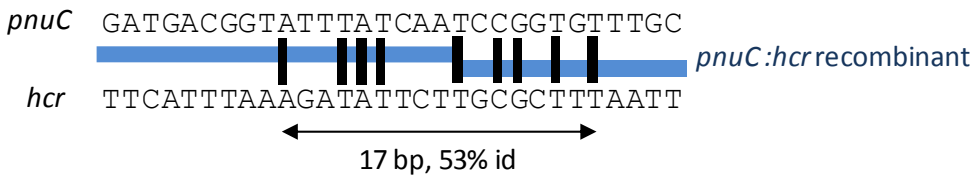


Figure 20: Representation of the *pnuC:hcr* junction in the strain cinstructed by Vine et al. (Vine et al. 2010). The *pnuC* and *hcr* sequences were aligned. Black bars represent the homologous bases. The final sequence after recombination is represented in blue.

phenomenon by placing mismatches at key points in an oligonucleotide recombineering assay. They saw that Redβ anneals with good efficiency oligonucleotides having diverged sequences up to 17% (1 mismatch every 6 pb) but looses dramatically in efficiency at 32% of divergence (1 mismatch per 3pb). The efficiency is retrieved when homologies are added at the oligonucleotide extremities. However, the homologies seem to be more important in the 5' part than in the 3' part of the incoming oligonucleotide (Li et al. 2013). In the recombineering experiment, the Court group reported that MutS and MutL, of the MMR mechanism, inhibit Redβ mediated ssDNA recombineering (Costantino, Court 2003). Our group saw, in the phage promoter inversion assay, that MMR does not affect the Redβ mediated recombination (Martinsohn, Radman, Petit 2008). The difference between these two assays is the potential number of recombination events in bacteria at the same time. In recombineering, only one to four recombination templates (chromosomes) are available for recombination, so that at most 4 recombination events occur at the same moment. In contrast, in phage promoter inversion assay, about 100 templates are available and certainly several recombination events occur together and saturate the MMR machinery. Very recent results in our team explored this hypothesis and seem to confirm it (J. Cornuault and M. De Paepe, personal communication).

B) Rad51-like recombinases need their SSB proteins

As explained in the introduction (Chapter I), two sorts of recombinases are distinguished in this super-family: the full size, similar to RecA, named UvsX, and the core-only, named Sak4. Each of these sub-families has its particularities.

The recombinase UvsX of T4 phage is the most studied. Its ability to promote SI and strand exchange is established and made the object of a recent review (Liu, Morrical 2010). Phages elaborate complex machinery for UvsX loading: Gp46/47, the phage heterodimer exonuclease, resects dsDNA to ssDNA on which Gp32, the phage SSB protein, loads. The UvsY mediator exchanges Gp32 to UvsX on the ssDNA, then UvsX contacts intact dsDNA to search for homologies and promote SI (see Figure 18 in Chapter I). Compared to other members of this super-family, *ie* RecA and Rad51, UvsX is faster to

degrade ATP and to promote a complete strand exchange (Liu et al. 2011). Nevertheless, we saw in Chapter III that RecA and UvsX both anneal diverged olinucleotides *in vitro*. Tests cannot be performed with RecA for ssDNA recombineering on diverged oligonucleotides, because RecA does not anneal *in vivo*, but UvsX does, albeit with a low efficiency (Lopes et al. 2010). Some accessory factor may be needed to fulfill efficient SSA *in vivo*. Once such factors are identified, it will be interesting to test the fidelity of UvsX in this test.

Sak4 was never characterized until this study. We saw in Chapter III that it promotes SSA but not SI, under the conditions used. We cannot exclude that Sak4 devoid of its GST tag promotes SI. Furthermore, Sak4 anneals diverged sequences both *in vivo* and *in vitro*. These capacities will be discussed deeper in paragraph II of this Chapter. Interestingly, Sak4 partially inhibits the activity of the host recombinase RecA *in vitro*. It could be interesting to test this inhibition in a classic UV assay with a $RecA^+$ *E. coli* strain carrying a plasmid expressing Sak4, with or without its SSB. If Sak4 inhibits RecA and does not promote SI, the UV induced DNA damages will not be repaired and the strain will have a UV sensitive phenotype.

SSB proteins are important for these recombinases. Indeed for the host recombinase RecA, SSB is necessary to promote efficient strand exchange ((Kowalczykowski, Krupp 1987) and Chapter III), but it can be replaced by T4 SSB protein Gp32 *in vitro* (Egner et al. 1987). Similarly, Gp32 is required for UvsX to promote efficient strand exchange. Furthermore, Gp32 presence enhances SSA and ATPase activities of UvsX (Formosa, Alberts 1986; Yonesaki, Minagawa 1989). Finally, Sak4 does not anneal at the replication fork without its SSB protein (Chapter III). It is probable that these phage SSB proteins load onto ssDNA at the same time as *E. coli* SSB and permit Sak4 loading. To test this idea, UvsX may be expressed together with Gp32 to see if its recombineering activity is enhanced. Maybe in this precise case, the UvsY mediator will also be required.

C) Gp2.5-like recombinases are poorly explored

The Gp2.5 protein looks like an SSB because of its OB-fold, but has a supplementary α-helix (Hollis et al. 2001). Until recently (Lopes et al. 2010), Gp2.5 of T7 was the only one characterized protein of this family, it is therefore the most studied of this super-family. It promotes SSA both *in vitro* and *in vivo* (Hyland, Rezende, Richardson 2003; Lopes et al. 2010). It also promotes SE in the presence of the phage helicase Gp4 *in vitro* (Kong, Richardson 1996). Gp2.5 cannot invade a dsDNA and needs the phage replicative helicase Gp4 to open the dsDNA. Contrary to other phage recombinases, the capacity of Gp2.5 to recombine diverged sequences has not been tested. It will be interesting to test

this capacity in our *in vitro* SSA assay involving diverged oligonucleotides as presented in Chapter III, and then *in vivo* with our ssDNA recombineering test as presented in Chapter III.

As I presented in the introduction (Chapter I), the Herpes Simplex Virus (HSV) caries a gene encoding ICP8, which is distantly related to Gp2.5. It forms a 2-rings structure closing over one another as a box where two ssDNA are compared by sliding (Figure 16 in Chapter I, (Tolun et al. 2013)). However, ICP8 shares with Gp2.5 only the middle segment of the protein, and it is not obvious that Gp2.5 has the same mode of action. The two quaternary structures of ICP8 and Gp2.5 should be compared to see if the ring structure could be found for Gp2.5.

2) Consequences on genomes

From a certain point a view, there is a tradeoff between fidelity and speed in the homologous recombination process. The relaxed fidelity of the phage recombinases permits to facilitate the exchanges between phages, but is damaging for the bacterial chromosome maintenance.

A) An advantage for phage genomes

As we saw above, all phage recombinases promote SSA *in vivo*, even UvsX, contrary to host recombinase RecA. Furthermore, most of them recombine diverged sequences (Gp2.5 unfaithful recombination activity is still to be demonstrated). These abilities change the way we imagine how phages protect their own genomes during replication compared to bacteria. Indeed homologous recombination permits bacteria to repair DNA damages with a high fidelity. In phages, recombinases probably repair faster their DNA at the expense of a lesser fidelity. Indeed, *in vitro* SSA between 81 nt long oligonucleotides is completed in less than a minute, even when sequences are diverged, whereas SI between 5 kb long DNA molecules is a slow process, needing some tens of minutes for completion and resolution. With SSA, phages can rapidly associate blocks of genomes on parts at least similar in term of homologies and functions to create as many viable phages as possible. This phenomenon separates clearly the way that phages stabilize their genomes, compared with bacteria. It may be that fast and low fidelity repair is linked with phage mosaïcism, whereas slow and faithful repair is assiociated with backbone/variable segment (VS) structures found in Bacteria, because of the necessity for them to conserve their chromosome integrity.

To support this theory, *Enterobacteria* prophages encoding for a recombinase are more mosaic than others (Bobay, Rocha, Touchon 2013). The authors proposed that the phages that do not encode for a recombination system use that of the host. To follow on the above reasoning, if phages use the

faithful RecA, they should have backbone/VS structure. However, these prophages have a mosaic structure. There are at least two other ways for phages to exchange DNA: (1) to encounter a phage encoding a recombinase in the same bacterium, or (2) to capture genes from prophages in the host genome, as we showed in Chapter II that RecA promotes this type of events.

It is to be noticed that genome comparisions show that virulent phages of the T4 family have a backbone/VS genome structure (Comeau et al. 2007; Ignacio-Espinoza, Sullivan 2012). It may be, as discussed in Chapter II, that virulent phages encounter each other with a lower frequency than temperate phages, and so have less opportunities to exchange genetic material. In addition, the fidelity of UvsX recombinase is unknown at present.

While the number and length of phage mosaics due to recent exchanges are now well explored, with an average of 2 mosaics per genome pair and a median mosaic length of 400 bp among temperate phage pairs (see Chapter II), we still do not have an overview on the particular genes that are being exchanged. This analysis can be performed with the data generated in the Chapter II. However, phage genomes are poorly annotated, preventing clear conclusion. A global annotation work is currently in progress in the team (H. Delattre, O. Souai, M.A. Petit, personal communication).

B) A danger for host genomes

When temperate phages are integrated into the bacterial chromosome, almost all prophage genes are repressed by a phage encoded repressor. However, leaks of transcription lead to a weak expression of prophage recombinases (as other genes), which could affect the host genome stability. This point is discussed in our recently submitted review (Menoui, Hutinet et al., *In Press*, FEMS).

In competent bacteria, prophage recombinases promote recombineering (Swingle et al. 2010). They insert oligonucleotides at the replication fork that could bring punctual mutations into the bacterial chromosome. Larger fragments may also be incorporated and bring several mutations at a time corresponding to the microdiversity regions previously observed in the team (Touzain et al. 2010). More dangerous for the bacteria, major delections could occur, without apparent need for long homologous region as mentioned above for the MMET activity of Redβ (Vine et al. 2010).

Rad52-like recombinase genes are absent in bacterial genomes, except in prophages but then essentially repressed. The ability of these recombinases to recombine diverged sequences urges bacteria to counter-select their presence into their genomes, in order to maintain a faithful method for the repair of DNA breaks by the SI ability of RecA, and to protect from aberrant recombination mediated by Rad52-like recombinases. However, Eukaryote genomes succeeded to domesticate

Rad52-like recombinases, without letting them directing the recombination process. Rad52 serves as a mediator in yeast between the SSB protein RPA and the recombinase Rad51 ((Krogh, Symington 2004) for review).

3) A new definition of phage species?

Classification of phages into species as done for the three kingdoms of life is difficult because (1) there is no gene shared by all phages, as the 16S-rDNA (18S in Eukaryotes), and (2) phages are highly divergent. This is why it is necessary to consider another way to define species. Indeed, Animal species were first defined as capable of producing a viable offspring by sexual reproduction. The hidden process behind is homologous recombination. Indeed, in order to be able to reproduce, Eukaryotic cells need to align their homologous chromosome by a mechanism of homologous recombination. This means that the homologous recombination process determines the species boundaries. In the same manner, different bacterial species cannot exchange core genetic material because of the necessity of homologous recombination in the process. This process is highly regulated, by the ability of RecA in a first place, and then by the MMR system (Stambuk, Radman 1998).

It may be convenient to consider a phage species definition based on homologous recombination. The fact that their recombinases have a relaxed fidelity may change our vision of species for phages. If the species membership can be circumscribed by the capacity to exchange DNA, two phages sharing a mosaic would belong to the same species. In the phage genomic tree, where phages are clusterised by their amount of shared genes (Rohwer, Edwards 2002), the closer phages infect the same host, while most distant phages infect different hosts. It is thus possible that these close phages are in the same species. In the same way, in the reticulate classification proposed by the Toussaint group (Lima-Mendez et al. 2008), each cluster including phages with high levels of shared genes infects related bacterial species and a cluster could be considered as a phage species. A clear example of phage species is T4-like phages. Their genome are so close from each other that the species definition is clearly visible by full genome alignments (Ignacio-Espinoza, Sullivan 2012). However, this way of thinking presents some difficulties. Indeed, in Chapter II we saw that active temperate *E. coli* phages are divided into two groups, probably two species, which do not exchange genome parts with each other: λ-likes and P2/Mu-likes. However when we look at prophages of these two “species”, they exchanged genome parts with each other. This probably means that phages of these two “species” might cross with each other. The resulting offspring might be insufficiently viable, so this case is found only in defective prophages.

II)The phage core-only RecA Sak4…

Sak4 has the particularity to share only one domain with RecA: the ATPase and DNA-binding domain. Below, Sak4 will be first compared with RecA and then with other core-only RecA.

1) … compared with RecA

In vitro, both RecA and GST-Sak4 promote SSA and both do it between diverged oligonucleotides. However, *in vivo* RecA cannot promote SSA while Sak4 does, but only in presence of its SSB. GST-Sak4 differs from RecA on other important aspects. Indeed, RecA hydrolyses ATP in a ssDNA-dependent way, while Sak4 does not. Whether it binds ATP without hydrolyzing it remains to be established. Finally, GST-Sak4 cannot promote SI as RecA does (Chapter III). A trivial explanation could be that GST-Sak4 is not functioning like Sak4. Further efforts of purification are needed to obtain a soluble Sak4. The differences between these two proteins reside into the Ct ssDNA-binding supplementary domain of RecA. In a recent model, this domain is supposed to participate actively to the SE process (De Vlaminck et al. 2012). Furthermore, deletions of part of this domain in RecA lead to a loss of SE activity, except if SE is performed at alkaline pH (Lusetti et al. 2003). It is not impossible that Sak4 promotes SE in these condition, and this remains to be tested. It may also be that an accessory factor remains to be discovered. For example, sak4 genes often co-occur with a nearby helicase gene on phage genomes (Supplementary Table 1 in Chapter III).

Another question must be asked. RecA is sensitive to the MMR system (Tham et al. 2013) and deletion of this system increases by 1000 to 10000-fold the efficiency of recombination during conjugation between *E. coli* and *Salmonella typhimurium* (Rayssiguier, Dohet, Radman 1991; Stambuk, Radman 1998). So if RecA is regulated by the MMR system, is Sak4 as well? Or does Sak4, similarly to Redβ, escape the MMR control, as there are too many recombination events during phage infections?

2) … compared with other core-only RecA

In Eukaryotes, core-only RecA form complexes with other Rad51-like proteins having a Nt domain. This is for instance the case in human, where Xrcc2 belong to the BCDX2 complex (Yokoyama et al. 2004; Chun, Buechelmaier, Powell 2013). This complex promotes SSA *in vitro*, as GST-Sak4 (Chapter III), but also permits to load Rad51 on ssDNA and to stabilize its filament, contrary to GST-Sak4 which inhibits partially RecA SE activity *in vitro* (Chapter III). In yeast, Csm2 and Psy3 belong to the SHU-

complex and, as BCDX2, permit to load and stabilize Rad51 on the ssDNA (She et al. 2012). At present, no evidence of comparable phenotypes between phage core-only Sak4 and Eukaryotic core-only Rad51 are available.

In Archaea, apart from RadA itself, the core-only RadB stabilizes the RadA filament (Komori et al. 2000; Graham, Rolfsmeier, Haseltine 2013). Interestingly, RadB does not degrade ATP like Sak4 (Komori et al. 2000). The other archaeal core-only aRadC is very ancestral and homologous to KaiC, a double tandem core-only RecA found in bacteria (see Chapter I, (Liang et al. 2013)).

III) Origin and distribution of recombinases

In this last part, I want to discuss the evolution and distribution of recombinases among the three (or four with viruses) kingdoms of life.

1) The ubiquitous RecA

RecA (Bacteria) and Rad51 (Eukaryotes) share a common domain, the core domain, capable of DNA binding and ATP hydrolysis, and have both a supplementary ssDNA-binding domain but not placed at the same position, in Ct for RecA and Nt for Rad51, which are not homologous to each other. However, UvsX (Phages) and RecA are perfectly homologous. This suggests that RecA and UvsX evolved from a common ancestor, and are more distantly related to Rad51. RadA (Archaea) is the perfect homolog of Rad51, with a similar Nt ssDNA binding domain (Haldenby, White, Allers 2009). Could it be possible that the ancestral RecA was a core-only RecA? And RecA/UvsX and Rad51/RadA diverged from each other genetically, by the acquisition of different supplementary ssDNA binding domains, and have converged phenotypically, as suggested by Yang et al. (Yang et al. 2001)? A tree and an evolution history can be drawn with some duplication of genes/genomes to join all RecA-likes (Figure 21 and (Lin et al. 2006)). A protein named KaiC, made of two RecA core domains in tandem, is homolog to aRadC, a core-only RadA in Archaea, which seems to branch very early in the RecA evolution tree (Haldenby, White, Allers 2009; Liang et al. 2013). The common ancestor to all Rad51-like proteins seems to be a aRadC-like core-only RecA. The other core-only RecA, collectively designated as Radβ, are domesticated to serve Rad51/RadA (Radα), and Radβ is close to Radα (Lin et al. 2006). It would be interesting to investigate where Sak4 would branch in such tree. It may either be the last phage representative of this aRadC-like ancestor that has its own recombination process or branch closer to RecA like Radβ and Radα.

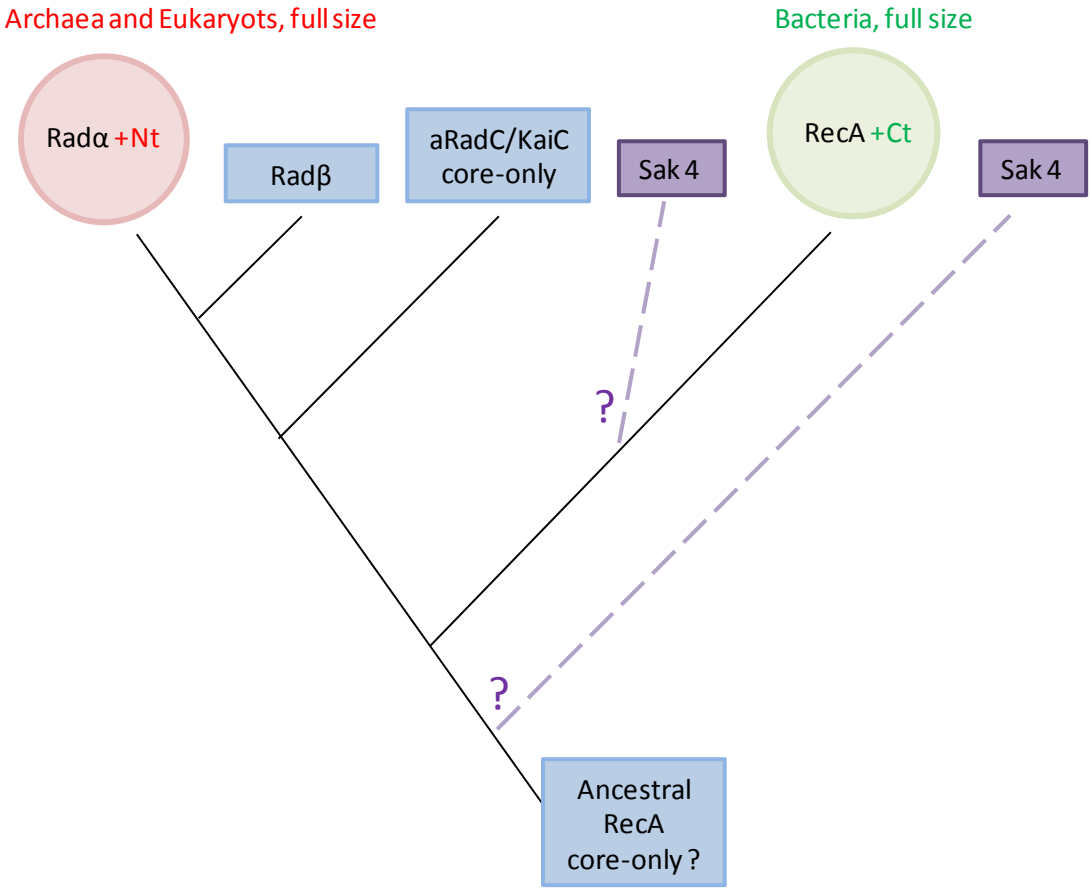


Figure 21: Where to place Sak4 in the RecA-like tree? A scheme summarizing various phylogenetic studies on RecA-like proteins is shown. On the right part, the bacterial RecA. On the left part, following Lin et al. nomenclature (Lin et al. 2006), the "Radα" which clusters both archaeal RadA and eukaryotic Rad51 proteins, that are full size and contain a similar Nt domain. More distantly related are the "Radβ" group proteins, many of them all core-only RecA (Xrcc2, RadB). Even more distant to the Radα are a group of recombination proteins found in Archeae, aRadC, which share homology to the circadian protein KaiC of cyanobacteria. We propose that the aRadC/KaiC core-only is ancestral (root of the tree), and ask whether Sak4, being a core-only RecA of phages, would be placed in such trees at the root, together with aRadC/KaiC, or rather next to RecA, in a position similar to RadB relative to RadA.

2) The phage Rad52 domesticated by Eukaryotes

Rad52-like recombinases are the predominant recombinases of phages. They were first classified into three groups: Redβ/RecT, Erf and Sak (Iyer, Koonin, Aravind 2002), and then clustered into one super family related to the Eukaryotic Rad52 (Lopes et al. 2010). No homologs were found in Archaea but Eukayotes have three Rad52 related proteins capable of promoting Rad51-independent SSA: (1) Rad52, having the Nt homologous to phage Rad52 and a supplementary Ct domain, that is a Rad51

and RPA (Replication Protein A, the SSB protein in Eukaryotes) binding domain (Park et al. 1996; Shen et al. 1996; Ploquin et al. 2008), (2) Rad59 (named Rad52B in human), homologous to phage Rad52 (Bai, Symington 1996; Mott, Symington 2011), and (3) Mgm101 (Mbantenkhu et al. 2011). This last protein acts into the mitochondria to repair DNA breaks. It is more related to phage Rad52 than to Eukaryotic ones on a structural point of view (Mbantenkhu et al. 2011). However, it is encoded in the nucleus. Two evolution histories are possible: (1) The α-Protobacteria that became mitochondria possessed a prophage encoding a Rad52, which was captured by the nucleus of the Eukaryote ancestor. This gene was then duplicated to obtain the three Eukaryotic Rad52. (2) Rad52 and Rad59 come from a Rad52-containing virus (currently disappeared or undiscovered) that was captured by an Eukaryote ancestor, and Mgm101 comes from a prophage trapped in the mitochondria.

3) The virus-only Gp2.5

Gp2.5-like recombinases of phages (Lopes et al. 2010) share homology with ICP8 from eukaryotic viruses (Kazlauskas, Venclovas 2012). No homologs were found in Archeal viruses yet, but it may be due to the difficulty of detecting them, as this super-family constitutes the most diverse super-family. This protein is found only in viruses and we have until now little knowledge on how it works. Other living organisms had probably rare occasions to capture these genes, or this protein could be deleterious for genome stability.

IV) General conclusion

In this thesis, I have contributed to show that relaxed fidelity is a general trait in Rad52-like and Sak4-like recombinases. The present discussion has tried to draw all consequences from this fact. At present, I propose that core-only RecA might be a very ancestral form of the Rad51-like super family of recombinases, that became more specialized thanks to the addition of domains. In the same way, Rad52-like recombinases, that are unfaithful, were domesticated by Eukaryotes by the acquisition of a supplementary domain, like in Eukaryotic Rad52, or by being used as cofactor of recombination, like Rad59. Finally, I propose a new method of classification of phages, excluding defective prophages, based on the relaxed fidelity of the phage recombinases: two phages capable of exchanging genetic material would belong to the same species.

References

Bai, Y, LS Symington. 1996. A Rad52 homolog is required for RAD51-independent mitotic recombination in Saccharomyces cerevisiae. Genes Dev 10:2025-2037.

Beaudoin, J, D Pratt. 1974. Antiserum Inactivation of Electrophoretically Purified M13 Diploid Virions Model for the F-Specific Filamentou. Journal of virology 13:466-469.

Bell, JC, JL Plank, CC Dombrowski, SC Kowalczykowski. 2012. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. Nature 491:274-278.

Better, M, D Freifelder. 1983. Studies on the replication of Escherichia coli phage lambda DNA. I. The kinetics of DNA replication and requirements for the generation of rolling circles. Virology 126:168-182.

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

Bouchard, JD, S Moineau. 2004. Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. J Bacteriol 186:3649-3652.

Bowater, R, AJ Doherty. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet 2:e8.

Briggs, GS, PA McEwan, J Yu, T Moore, J Emsley, RG Lloyd. 2007. Ring structure of the Escherichia coli DNA-binding protein RdgC associated with recombination and replication fork repair. J Biol Chem 282:12353-12357.

Brown, J, T Brown, KR Fox. 2001. Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. Biochem J 354:627-633.

Brussow, H, C Canchaya, WD Hardt. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68:560-602, table of contents.

Bryant, FR, IR Lehman. 1985. On the mechanism of renaturation of complementary DNA strands by the recA protein of Escherichia coli. Proc Natl Acad Sci U S A 82:297-301.

Chayot, R, B Montagne, D Mazel, M Ricchetti. 2010. An end-joining repair mechanism in Escherichia coli. Proc Natl Acad Sci U S A 107:2141-2146.

Chiapello, H, I Bourgait, F Sourivong, G Heuclin, A Gendrault-Jacquemard, MA Petit, M El Karoui. 2005. Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. BMC Bioinformatics 6:171.

Chun, J, ES Buechelmaier, SN Powell. 2013. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. Mol Cell Biol 33:387-395.

Clark, AJ, AD Margulies. 1965. Isolation and Characterization of Recombination-Deficient Mutants of Escherichia Coli K12. Proc Natl Acad Sci U S A 53:451-459.

Comeau, AM, C Bertrand, A Letarov, F Tetart, HM Krisch. 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. Virology 362:384-396.

Costantino, N, DL Court. 2003. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci U S A 100:15748-15753.

Court, R, N Cook, K Saikrishnan, D Wigley. 2007. The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. J Mol Biol 371:25-33.

D'Herelle, F. 1917. On an invisible microbe antagonistic toward dysenteric bacilli: brief note Res Microbiol 158:553-554.

d'Adda di Fagagna, F, GR Weller, AJ Doherty, SP Jackson. 2003. The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO reports 4:47-52.

Datta, S, N Costantino, X Zhou, DL Court. 2008. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc Natl Acad Sci U S A 105:1626-1631.

de Vega, M. 2013. The minimal Bacillus subtilis nonhomologous end joining repair machinery. PLoS One 8:e64232.

De Vlaminck, I, MT van Loenhout, L Zweifel, J den Blanken, K Hooning, S Hage, J Kerssemakers, C Dekker. 2012. Mechanism of homology recognition in DNA recombination from dual-molecule experiments. Mol Cell 46:616-624.

den Bakker, HC, CA Cummings, V Ferreira, P Vatta, RH Orsi, L Degoricija, M Barker, O Petrauskene, MR Furtado, M Wiedmann. 2010. Comparative genomics of the bacterial genus Listeria: Genome evolution is characterized by limited gene acquisition and limited gene loss. BMC Genomics 11:688.

Drees, JC, S Chitteni-Pattu, DR McCaslin, RB Inman, MM Cox. 2006. Inhibition of RecA protein function by the RdgC protein from Escherichia coli. J Biol Chem 281:4708-4717.

Egner, C, E Azhderian, SS Tsang, CM Radding, JW Chase. 1987. Effects of various single-stranded-DNA-binding proteins on reactions promoted by RecA protein. J Bacteriol 169:3422-3428.

Erler, A, S Wegmann, C Elie-Caille, CR Bradshaw, M Maresca, R Seidel, B Habermann, DJ Muller, AF Stewart. 2009. Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. J Mol Biol 391:586-598.

Forget, AL, SC Kowalczykowski. 2012. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. Nature 482:423-427.

Formosa, T, BM Alberts. 1986. Purification and characterization of the T4 bacteriophage uvsX protein. J Biol Chem 261:6107-6118.

Gajewski, S, MR Webb, V Galkin, EH Egelman, KN Kreuzer, SW White. 2011. Crystal structure of the phage T4 recombinase UvsX and its functional interaction with the T4 SF2 helicase UvsW. J Mol Biol 405:65-76.

Graham, WJt, ML Rolfsmeier, CA Haseltine. 2013. An archaeal RadA paralog influences presynaptic filament formation. DNA Repair (Amst) 12:403-413.

Haber, JE. 2007. Evolution of Models of homologous recombination.

Haldenby, S, MF White, T Allers. 2009. RecA family proteins in archaea: RadA and its cousins. Biochem Soc Trans 37:102-107.

Handa, N, K Morimatsu, ST Lovett, SC Kowalczykowski. 2009. Reconstitution of initial steps of dsDNA break repair by the RecF pathway of E. coli. Genes Dev 23:1234-1245.

Hatfull, GF. 2014. Mycobacteriophages: windows into tuberculosis. PLoS Pathog 10:e1003953.

Holliday, R. 1964. A mechanism for gene conversion in fungi. Genet Res 5:282-304.

Hollis, T, JM Stattel, DS Walther, CC Richardson, T Ellenberger. 2001. Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7. Proc Natl Acad Sci U S A 98:9557-9562.

Huang, CH, T Hsiang, JT Trevors. 2013. Comparative bacterial genomics: defining the minimal core genome. Antonie Van Leeuwenhoek 103:385-398.

Hyland, EM, LF Rezende, CC Richardson. 2003. The DNA binding domain of the gene 2.5 single-stranded DNA-binding protein of bacteriophage T7. J Biol Chem 278:7247-7256.

Ignacio-Espinoza, JC, MB Sullivan. 2012. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. Environ Microbiol 14:2113-2126.

Ivankovic, S, D Dermic. 2012. DNA end resection controls the balance between homologous and illegitimate recombination in Escherichia coli. PLoS One 7:e39030.

Iyer, LM, EV Koonin, L Aravind. 2002. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics 3:8.

Kagawa, W, H Kurumizaka, S Ikawa, S Yokoyama, T Shibata. 2001. Homologous pairing promoted by the human Rad52 protein. J Biol Chem 276:35201-35208.

Kagawa, W, H Kurumizaka, R Ishitani, S Fukai, O Nureki, T Shibata, S Yokoyama. 2002. Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. Mol Cell 10:359-371.

Kantake, N, MV Madiraju, T Sugiyama, SC Kowalczykowski. 2002. Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. Proc Natl Acad Sci U S A 99:15327-15332.

Karpenshif, Y, KA Bernstein. 2012. From yeast to mammals: recent advances in genetic control of homologous recombination. DNA Repair (Amst) 11:781-788.

Kazlauskas, D, C Venclovas. 2012. Two distinct SSB protein families in nucleo-cytoplasmic large DNA viruses. Bioinformatics 28:3186-3190.

Kim, YT, S Tabor, C Bortner, JD Griffith, CC Richardson. 1992. Purification and characterization of the bacteriophage T7 gene 2.5 protein. A single-stranded DNA-binding protein. J Biol Chem 267:15022-15031.

Kmiec, E, WK Holloman. 1981. Beta protein of bacteriophage lambda promotes renaturation of DNA. J Biol Chem 256:12636-12639.

Komori, K, T Miyata, J DiRuggiero, R Holley-Shanks, I Hayashi, IK Cann, K Mayanagi, H Shinagawa, Y Ishino. 2000. Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. J Biol Chem 275:33782-33790.

Kong, D, CC Richardson. 1996. Single-stranded DNA binding protein and DNA helicase of bacteriophage T7 mediate homologous DNA strand exchange. EMBO J 15:2010-2019.

Konstantinidis, KT, A Ramette, JM Tiedje. 2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci 361:1929-1940.

Kowalczykowski, SC, RA Krupp. 1987. Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. J Mol Biol 193:97-113.

Krogh, BO, LS Symington. 2004. Recombination proteins in yeast. Annu Rev Genet 38:233-271.

Kwan, T, J Liu, M DuBow, P Gros, J Pelletier. 2005. The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. Proc Natl Acad Sci U S A 102:5174-5179.

Lajoie, MJ, CJ Gregg, JA Mosberg, GC Washington, GM Church. 2012. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. Nucleic Acids Res 40:e170.

Leewenhoeck, A. 1684. An Abstract of a Letter from Mr. Anthony Leewenhoeck at Delft, Dated Sep. 17. 1683. Containing Some Microscopical Observations, about Animals in the Scurf of the Teeth, the Substance Call'd Worms in the Nose, the Cuticula Consisting of Scales. Philosophical Transactions of the Royal Society of London 14:568-574.

Lesterlin, C, G Ball, L Schermelleh, DJ Sherratt. 2014. RecA bundles mediate homology pairing between distant sisters during DNA break repair. Nature 506:249-253.

Lestini, R, B Michel. 2007. UvrD controls the access of recombination proteins to blocked replication forks. EMBO J 26:3804-3814.

Li, XT, LC Thomason, JA Sawitzke, N Costantino, DL Court. 2013. Bacterial DNA polymerases participate in oligonucleotide recombination. Mol Microbiol 88:906-920.

Li, Z, G Karakousis, SK Chiu, G Reddy, CM Radding. 1998. The beta protein of phage lambda promotes strand exchange. J Mol Biol 276:733-744.

Liang, PJ, WY Han, QH Huang, YZ Li, JF Ni, QX She, YL Shen. 2013. Knockouts of RecA-like proteins RadC1 and RadC2 have distinct responses to DNA damage agents in Sulfolobus islandicus. J Genet Genomics 40:533-542.

Lima-Mendez, G, J Van Helden, A Toussaint, R Leplae. 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol 25:762-777.

Lin, Z, H Kong, M Nei, H Ma. 2006. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. Proc Natl Acad Sci U S A 103:10328-10333.

Liu, J, KT Ehmsen, WD Heyer, SW Morrical. 2011. Presynaptic filament dynamics in homologous recombination and DNA repair. Crit Rev Biochem Mol Biol 46:240-270.

Liu, J, SW Morrical. 2010. Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. Virol J 7:357.

Lopes, A, J Amarir-Bouhram, G Faure, MA Petit, R Guerois. 2010. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res 38:3952-3962.

Lusetti, SL, JC Drees, EA Stohl, HS Seifert, MM Cox. 2004. The DinI and RecX proteins are competing modulators of RecA function. J Biol Chem 279:55073-55079.

Lusetti, SL, EA Wood, CD Fleming, MJ Modica, J Korth, L Abbott, DW Dwyer, AI Roca, RB Inman, MM Cox. 2003. C-terminal deletions of the Escherichia coli RecA protein. Characterization of in vivo and in vitro effects. J Biol Chem 278:16372-16380.

Maher, RL, SW Morrical. 2013. Coordinated Binding of Single-Stranded and Double-Stranded DNA by UvsX Recombinase. PLoS One 8:e66654.

Malkov, VA, L Sastry, RD Camerini-Otero. 1997. RecA protein assisted selection reveals a low fidelity of recognition of homology in a duplex DNA by an oligonucleotide. J Mol Biol 271:168-177.

Maresca, M, A Erler, J Fu, A Friedrich, Y Zhang, AF Stewart. 2010. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. BMC Mol Biol 11:54.

Martinsohn, JT, M Radman, MA Petit. 2008. The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. PLoS Genet 4:e1000065.

Matsubara, K, AD Malay, FA Curtis, GJ Sharples, JG Heddle. 2013. Structural and functional characterization of the Redbeta recombinase from bacteriophage lambda. PLoS One 8:e78869.

Mbantenkhu, M, X Wang, JD Nardozzi, S Wilkens, E Hoffman, A Patel, MS Cosgrove, XJ Chen. 2011. Mgm101 is a Rad52-related protein required for mitochondrial DNA recombination. J Biol Chem 286:42360-42370.

Mendel, g. 1866. Versuche über Plflanzen-hybriden. Verhandlungen des naturforschenden Ver-eines in Brünn Bs. IV für das Jahr 1865:3-47.

Meselson, M, JJ Weigle. 1961. Chromosome brekage accompanying genetic recombination in bacteriophage. Proc Natl Acad Sci U S A 47:857-868.

Meselson, MS, CM Radding. 1975. A general model for genetic recombination. Proc Natl Acad Sci U S A 72:358-361.

Morgan, TH. 1911. Chromosomes and Associative Inheritance. Science 34:636-638.

Morimatsu, K, SC Kowalczykowski. 2003. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell 11:1337-1347.

Mortensen, UH, C Bendixen, I Sunjevaric, R Rothstein. 1996. DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci U S A 93:10729-10734.

Mott, C, LS Symington. 2011. RAD51-independent inverted-repeat recombination by a strand-annealing mechanism. DNA Repair (Amst) 10:408-415.

Muller, HJ, E Altenburg. 1930. The Frequency of Translocations Produced by X-Rays in Drosophila. Genetics 15:283-311.

Murphy, KC. 2000. Bacteriophage P22 Abc2 protein binds to RecC increases the 5' strand nicking activity of RecBCD and together with lambda bet, promotes Chi-independent recombination. J Mol Biol 296:385-401.

Noirot, P, RD Kolodner. 1998. DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem 273:12274-12280.

Ochi, T, Q Wu, TL Blundell. 2014. The spatial organization of non-homologous end joining: from bridging to end joining. DNA Repair (Amst) 17:98-109.

Park, MS, DL Ludwig, E Stigger, SH Lee. 1996. Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. J Biol Chem 271:18996-19000.

Passy, SI, X Yu, Z Li, CM Radding, EH Egelman. 1999. Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. Proc Natl Acad Sci U S A 96:4279-4284.

Ploquin, M, A Bransi, ER Paquet, AZ Stasiak, A Stasiak, X Yu, AM Cieslinska, EH Egelman, S Moineau, JY Masson. 2008. Functional and structural basis for a bacteriophage homolog of human RAD52. Curr Biol 18:1142-1146.

Pope, WHD Jacobs-SeraDA Russell, et al. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One 6:e16329.

Poteete, AR. 2008. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. Mol Microbiol 68:66-74.

Poteete, AR. 2013. Involvement of DNA Replication Proteins in Phage Lambda Red-Mediated Homologous Recombination. PLoS One 8:e67440.

Proux, C, D van Sinderen, J Suarez, P Garcia, V Ladero, GF Fitzgerald, F Desiere, H Brussow. 2002. The Dilemma of Phage Taxonomy Illustrated by Comparative Genomics of Sfi21-Like Siphoviridae in Lactic Acid Bacteria. Journal of Bacteriology 184:6026-6036.

Radzimanowski, J, F Dehez, A Round, A Bidon-Chanal, S McSweeney, J Timmins. 2013. An 'open' structure of the RecOR complex supports ssDNA binding within the core of the complex. Nucleic Acids Res 41:7972-7986.

Ragunathan, K, C Liu, T Ha. 2012. RecA filament sliding on DNA facilitates homology search. Elife 1:e00067.

Rajagopala, SV, S Casjens, P Uetz. 2011. The protein interaction map of bacteriophage lambda. BMC Microbiol 11:213.

Rayssiguier, C, C Dohet, M Radman. 1991. Interspecific recombination between Escherichia coli and Salmonella typhimurium occurs by the RecABCD pathway. Biochimie 73:371-374.

Reddy, MS, MB Vaze, K Madhusudan, K Muniyappa. 2000. Binding of SSB and RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA polymerization on single-stranded DNA. Biochemistry 39:14250-14262.

Renkawitz, J, CA Lademann, M Kalocsay, S Jentsch. 2013. Monitoring homology search during DNA double-strand break repair in vivo. Mol Cell 50:261-272.

Renzette, N, N Gumlaw, SJ Sandler. 2007. DinI and RecX modulate RecA-DNA structures in Escherichia coli K-12. Mol Microbiol 63:103-115.

Resnick, MA. 1976. The repair of double-strand breaks in DNA; a model involving recombination. J Theor Biol 59:97-106.

Rezende, LF, S Willcox, JD Griffith, CC Richardson. 2003. A single-stranded DNA-binding protein of bacteriophage T7 defective in DNA annealing. J Biol Chem 278:29098-29105.

Rice, KP, AL Eggler, P Sung, MM Cox. 2001. DNA pairing and strand exchange by the Escherichia coli RecA and yeast Rad51 proteins without ATP hydrolysis: on the importance of not getting stuck. J Biol Chem 276:38570-38581.

Rohwer, F, R Edwards. 2002. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. Journal of Bacteriology 184:4529-4535.

Rossi, MJ, OM Mazina, DV Bugreev, AV Mazin. 2011. The RecA/RAD51 protein drives migration of Holliday junctions via polymerization on DNA. Proc Natl Acad Sci U S A 108:6432-6437.

Rotman, E, E Kouzminova, G Plunkett, 3rd, A Kuzminov. 2012. Genome of Enterobacteriophage Lula/phi80 and insights into its ability to spread in the laboratory environment. J Bacteriol 194:6802-6817.

Roux, S, M Krupovic, A Poulet, D Debroas, F Enault. 2012. Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One 7:e40418.

Rybalchenko, N, EI Golub, B Bi, CM Radding. 2004. Strand invasion promoted by recombination protein beta of coliphage lambda. Proc Natl Acad Sci U S A 101:17056-17060.

Sagi, D, T Tlusty, J Stavans. 2006. High fidelity of RecA-catalyzed recombination: a watchdog of genetic diversity. Nucleic Acids Res 34:5021-5031.

Sarker, SA, S McCallin, C Barretto, et al. 2012. Oral T4-like phage cocktail application to healthy adult volunteers from Bangladesh. Virology 434:222-232.

Seitz, EM, JP Brockman, SJ Sandler, AJ Clark, SC Kowalczykowski. 1998. RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. Genes Dev 12:1248-1253.

She, Z, ZQ Gao, Y Liu, WJ Wang, GF Liu, EV Shtykova, JH Xu, YH Dong. 2012. Structural and SAXS analysis of the budding yeast SHU-complex proteins. FEBS Lett 586:2306-2312.

Shen, Z, KG Cloud, DJ Chen, MS Park. 1996. Specific interactions between the human RAD51 and RAD52 proteins. J Biol Chem 271:148-152.

Shiraishi, K, K Hanada, Y Iwakura, H Ikeda. 2002. Roles of RecJ, RecO, and RecR in RecET-Mediated Illegitimate Recombination in Escherichia coli. Journal of Bacteriology 184:4715-4721.

Shiraishi, K, Y Imai, S Yoshizaki, T Tadaki, Y Ogata, H Ikeda. 2006. The role of UvrD in RecET-mediated illegitimate recombination in Escherichia coli. Genes Genet Syst 81:291-297.

Smith, GR. 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol Mol Biol Rev 76:217-228.

Smith, MC, RW Hendrix, R Dedrick, et al. 2013. Evolutionary relationships among actinophages and a putative adaptation for growth in Streptomyces spp. J Bacteriol 195:4924-4935.

Spies, M, SC Kowalczykowski. 2006. The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. Mol Cell 21:573-580.

Stahl, MM, L Thomason, AR Poteete, T Tarkowski, A Kuzminov, FW Stahl. 1997. Annealing vs. invasion in phage lambda recombination. Genetics 147:961-977.

Stambuk, S, M Radman. 1998. Mechanism and control of interspecies recombination in Escherichia coli. I. Mismatch repair, methylation, recombination and replication functions. Genetics 150:533-542.

Stanier, RYVN, C B. 1962. The concept of a bacterium. Archiv für Mikrobiologie 42:17-35.

Story, RM, IT Weber, TA Steitz. 1992. The structure of the E. coli recA protein monomer and polymer. Nature 355:318-325.

Swenson, KM, P Guertin, H Deschenes, A Bergeron. 2013. Reconstructing the modular recombination history of Staphylococcus aureus phages. BMC Bioinformatics 14 Suppl 15:S17.

Swingle, B, E Markel, N Costantino, MG Bubunenko, S Cartinhour, DL Court. 2010. Oligonucleotide recombination in Gram-negative bacteria. Mol Microbiol 75:138-148.

Szostak, JW, TL Orr-Weaver, RJ Rothstein, FW Stahl. 1983. The double-strand-break repair model for recombination. Cell 33:25-35.

Tham, KC, N Hermans, HH Winterwerp, MM Cox, C Wyman, R Kanaar, JH Lebbink. 2013. Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures. Mol Cell 51:326-337.

Thomas, CA, Jr. 1966. Recombination of DNA molecules. Prog Nucleic Acid Res Mol Biol 5:315-337.

Tolun, G, AM Makhov, SJ Ludtke, JD Griffith. 2013. Details of ssDNA annealing revealed by an HSV-1 ICP8-ssDNA binary complex. Nucleic Acids Res 41:5927-5937.

Touchon, M, C Hoede, O Tenaillon, et al. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 5:e1000344.

Tougu, K, KJ Marians. 1996. The extreme C terminus of primase is required for interaction with DnaB at the replication fork. J Biol Chem 271:21391-21397.

Touzain, F, E Denamur, C Medigue, V Barbe, M El Karoui, MA Petit. 2010. Small variable segments constitute a major type of diversity of bacterial genomes at the species level. Genome Biol 11:R45.

Touzain, F, MA Petit, S Schbath, M El Karoui. 2011. DNA motifs that sculpt the bacterial chromosome. Nat Rev Microbiol 9:15-26.

Twort, FW. 1915. An investivation on the nature of ultra-microscopic viruses. Bacteriophage 1:127-129.

van der Heijden, T, M Modesti, S Hage, R Kanaar, C Wyman, C Dekker. 2008. Homologous recombination in real time: DNA strand exchange by RecA. Mol Cell 30:530-538.

Van Dyck, E, AZ Stasiak, A Stasiak, SC West. 2001. Visualization of recombination intermediates produced by RAD52-mediated single-strand annealing. EMBO Rep 2:905-909.

Veaute, X, S Delmas, M Selva, J Jeusset, E Le Cam, I Matic, F Fabre, MA Petit. 2005. UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. EMBO J 24:180-189.

Viguera, E, D Canceill, SD Ehrlich. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J 20:2587-2595.

Vine, CE, MC Justino, LM Saraiva, J Cole. 2010. Detection by whole genome microarrays of a spontaneous 126-gene deletion during construction of a ytfE mutant: confirmation that a ytfE mutation results in loss of repair of iron-sulfur centres in proteins damaged by oxidative or nitrosative stress. J Microbiol Methods 81:77-79.

Watson, JD, FH Crick. 1953. The structure of DNA. Cold Spring Harb Symp Quant Biol 18:123-131.

Woese, CR, GE Fox. 1977. The concept of cellular evolution. J Mol Evol 10:1-6.

Woese, CR, O Kandler, ML Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576-4579.

Yang, S, MS VanLoock, X Yu, EH Egelman. 2001. Comparison of bacteriophage T4 UvsX and human Rad51 filaments suggests that RecA-like polymers may have evolved independently. J Mol Biol 312:999-1009.

Yokoyama, H, N Sarai, W Kagawa, R Enomoto, T Shibata, H Kurumizaka, S Yokoyama. 2004. Preferential binding to branched DNA strands and strand-annealing activity of the human Rad51B, Rad51C, Rad51D and Xrcc2 protein complex. Nucleic Acids Res 32:2556-2565.

Yonesaki, T, T Minagawa. 1989. Synergistic action of three recombination gene products of bacteriophage T4, uvsX, uvsY, and gene 32 proteins. J Biol Chem 264:7814-7820.

ANNEXES

Résumé du manuscrit de thèse :

Recombinaison homologue chez les bactériophages : moins de fidélité pour plus d'échanges.

Chapitre I : Introduction

Les êtres vivants sont divisés en trois règnes : les Eucaryotes, regroupant les animaux, les plantes, les champignons, et plus largement les espèces composées de cellules possédant un noyau ; les Archées, des espèces vivant dans les milieux extrêmes de températures ou de conditions salines par exemple ; et les Bactéries. C'est l'étude de ces dernières qui nous intéresse ici. Plus particulièrement l'étude de la structure de leurs génomes. Au sein de la même espèce, on observe un « squelette de génome » commun à toutes les souches, avec des parties variables propres à chaque souche. Ces parties variables sont sujettes à échanges par transfert horizontal. La partie stable du génome est conservée et réparée par recombinaison homologue si besoin.

Chaque règne du vivant est infecté par des virus, qui ciblent le plus souvent une ou un petit nombre d'espèces proches. Chez les bactéries, ces virus s'appellent des bactériophages, ou plus communément phages, et la plupart d'entre eux ont un génome à ADN. Il en existe deux types suivant les styles de vie qu'ils adoptent. Les phages dits virulents se reproduisent uniquement par cycles lytiques : ils entrent dans une bactérie, s'y répliquent et la lysent pour en sortir. Les phages dits tempérés font également des cycles lytiques, comme les virulents, mais peuvent aussi, après être entré dans la bactérie, s'intégrer dans son chromosome au lieu de se répliquer. On les appelle alors prophages. De temps en temps, les prophages sortent du cycle lysogène en s'excisant du chromosome suite à un stress et reprennent un cycle lytique.

La structure des génomes des phages virulents s'approche de celle des génomes des bactéries : un squelette constant, émaillé de régions variables. Les génomes de phages tempérés par contre sont plus complexes. Lorsqu'on les compare, chaque génome semble composé d'autres génomes, à l'image d'une mosaïque. La question se pose alors de comprendre par quel moyen des phages différents s'échangent des fragments d'ADN pour créer un phage mosaïque. Plusieurs hypothèses sont possibles. La première est que la polymérase, chargée de la réplication de l'ADN, peut sauter d'un substrat à l'autre, par exemple d'un génome de phage à l'autre. La deuxième est que deux morceaux d'ADN cassés soient simplement collés l'un à l'autre. La troisième consiste aussi

à lier deux molécules d'ADN cassées mais grâce à des petites zones de moins de 10 pb capable de s'apparier, situées au bout des extrémités cassées. Enfin, la dernière possibilité est la recombinaison homologue.

Chez les bactéries, la recombinaison homologue a été très étudiée. La protéine principale des événements de recombinaison s'appelle une recombinase. RecA est la recombinase chez les bactéries. Elle est capable de fixer un ADN simple brin et d'aller chercher de l'homologie dans un ADN double brin. Cet événement est appelé invasion de brin et consomme de l'ATP. La polymérisation de RecA sur un ADN simple est régulée par différentes protéines. Certaines, comme les complexes RecBCD ou RecOR, opèrent le chargement de RecA sur son substrat, qui forme ensuite un filament hélicoïdal sur l'ADN. D'autres, comme DinI, stabilisent le filament RecA, alors que RecX a une activité contraire. Ce système est très fidèle. RecA cherche l'homologue parfait pour la recombinaison, et s'il se trompe, le système de détection des mésappariements le corrige.

Les phages ont eux aussi des recombinases. Elles sont classées en trois groupes, de structures et d'activités différentes. Le premier regroupe les protéines de structure apparentée à Rad52, qui sont les plus fréquemment rencontrées chez les phages, il contient notamment les familles Sak, Erf, et Redβ / RecT. La plus étudiée est Redβ du phage λ infectant *Escherichia coli*. L'activité principale des protéines de ce groupe est l'appariement simple brin, une façon de faire de la recombinaison homologue différente de RecA. De plus, il a été montré que Redβ apparie même des séquences divergées jusqu'à 20%. Le deuxième groupe est celui des protéines de type Rad51. Ce sont des recombinases apparentées à celles de bactéries, archées et eucaryotes. Elles sont divisées en deux grandes familles : les UvsX et les Sak4. Tous les membres de ce groupe possèdent un domaine commun, le domaine cœur capable de lier l'ADN et d'hydrolyser l'ATP. Certaines ont un domaine supplémentaire, ayant également la faculté de lier l'ADN. Les UvsX ont la même activité que RecA, et les Sak4, qui ne possèdent pas de domaine supplémentaire n'ont pas été étudiées jusqu'à présent. Enfin la dernière famille est celle des Gp2.5, dont la plus étudiée est celle du phage d'*E. coli* T7. Elle est capable de faire de l'appariement simple brin.

Les mécanismes permettant l'acquisition de mosaïques chez les phages tempérés sont encore inconnus. Dans notre équipe, nous faisons l'hypothèse que les recombinases de phages participent à cette acquisition. En particulier, leur capacité à recombiner des séquences divergées, à l'inverse de RecA, pourrait expliquer que les échanges entre génomes de phages soient aussi abondants. Cependant la démonstration complète que la recombinaison entre séquences divergées dépend de ces recombinases n'a pas été obtenue. Nous avons donc mis au point différentes expériences de recombinaison où les recombinases de phages interviennent. Nous avons ainsi pu

mesurer précisément la contribution de différentes recombinases phagiques et de RecA aux échanges de matériel génétique, en fonction de la divergence des séquences recombinantes. Ce travail est présenté dans le chapitre II. Dans un deuxième temps, j'ai décidé de me tourner vers l'étude de Sak4, encore très peu étudiée. J'ai utilisé des approches de génétique et de biochimie pour caractériser cette protéine. La question principale a été de savoir si son activité était à l'image de RecA ou plus proche des Rad52. Ce travail est présenté au chapitre III

Chapitre II : Les phages tempérés acquièrent de l'ADN provenant des prophages défectifs par un mécanisme de recombinaison homologue relâchée : le rôle des recombinases de la famille des Rad52.

Les phages sont très divers. Des analyses de comparaison génomique suggèrent qu'ils ont un fort taux de transfert horizontal, surtout les phages tempérés à ADN double brin. Le mécanisme moléculaire sous-jacent est encore inconnu et sujet à plusieurs hypothèses. Notre hypothèse part de l'observation qu'une recombinase de la famille des Rad52, Redβ, encodée par le phage λ, a une fidélité diminuée par rapport à la recombinase de l'hôte, RecA. Nous proposons que les recombinases de phages,à cause de cette fidélité réduite sont capables de promouvoir le transfert de gènes entre phages même si les séquences sont éloignées. Les phages tempérés étant capables de s'insérer dans le chromosome de leur hôte, la probabilité de rencontre entre le phage infectant une bactérie et divers prophages déjà présents dans le chromosome bactérien, même défectifs, est augmentée, au-delà des échanges entre deux phages différents infectant simultanément la même bactérie. Enfin, nous nous sommes aussi demandés si les phages utilisaient uniquement leur propre recombinase ou s'ils étaient capables d'utiliser à la fois la leur et celle de l'hôte.

Dans un premier temps, nous avons cherché à comprendre quelles recombinases participaient à un événement de recombinaison entre le phage λ et divers prophages défectifs d'*E. coli* K12. Pour cela, nous avons identifié 3 régions homologues entre prophages défectifs et le phage λ : dans Dlp12, une région de 3500 pb autour du site cos présente 98% d'identité avec λ ; dans Qin, une région de 700 pb présente 95% d'identité avec le locus *tfa* de λ ; et enfin dans Rac une région de 700 pb présente 85% d'identité avec le locus *stf* de λ. Au centre de chacune de ces zones, un gène de résistance à un antibiotique (kanamycine ou chloramphénicol) a été introduit de telle manière à ce que λ puisse capturer ce gène par double événement de recombinaison homologue. Le principe est simple : un dérivé de λ possédant un marqueur de résistance à la phléomycine (λ_P), délété ou non de son gène de recombinason homologue *red*β*,* infecte une des souches d'*E. coli* ayant un prophage

modifié comme décrit plus haut. Cette souche est de plus délétée ou non de son gène de recombinaison homologue *recA*. Lors de l'infection, λ_P se multiplie par cycles lytiques, et a la possibilité de capturer le gène de résistance placé dans le prophage. Pour dénombrer les phages recombinants, le surnageant contenant les phages après la lyse des bactéries est récupéré, puis les phages sont lysogénisés dans une souche sensible aux antibiotiques. Les phages totaux sont comptabilisés par sélection sur phléomycine, et les phages recombinants par sélection sur chloramphénicol ou kanamycine, selon le marqueur placé dans le prophage. Le ratio de ces deux dénombrements permet de comptabiliser les fréquences de recombinaison lors du cycle lytique.

Aucun recombinant n'a été obtenu en l'absence des deux recombinases, Redβ et RecA, alors que des fréquences de recombinaison allant jusqu'à 10^{-4} sont observées en présence des recombinases (région de 3500 pb d'homologie). Ce résultat souligne l'essentialité des recombinases dans le processus de transfert horizontal entre phages et prophages et, par extension, entre phages. Un effet très net de la taille de la région homologue est observé : entre 700 pb et 3500pb, la fréquence augmente d'un facteur 100 (les deux régions sont à 98% identiques entre elles). Il est connu que la fréquence de recombinaison homologue augmente proportionnellement à la taille des séquences d'homologie. Il est également connu que la présence du site cos, un site de coupure reconnu par la terminase du phage, augmente la probabilité d'avoir des événements de recombinaison à ce locus précis. Sur des séquences longues à forte identité, RecA et Redβ semblent participer à parts égales aux événements de recombinaison : chaque simple mutant diminue d'un facteur deux environ la fréquence de recombinaison.

De manière intéressante, Il n'y a pas de différence entre la fréquence de recombinaison des séquences à 95% d'identité et celles entre séquences de 85% d'identité de même longueur (700 pb). Pour ces séquences plus courtes, divergées ou non, Redβ joue un rôle majeur, et RecA un rôle mineur : elle participe environ à 1/10ème des événements de recombinaison (Figure 1).

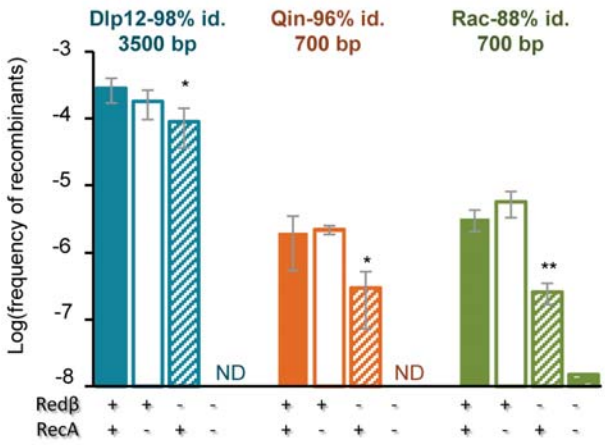


Figure 1 : Les recombinants entre λ et les prophages défectifs sont formés préférentiellement par Redβ, surtout si les séquences sont courtes. Moyenne des fréquences des recombinants avec Dlp12, Qin et Rac en fonction de la présence ou non de Redβ et RecA. Les barres d'erreur sont des écart-types calculé à partir de 3 tests de recombinaison indépendants.

Dans un deuxième temps, nous avons voulu généraliser les observations faites quant à la fidélité d'appariement de Redβ à toutes les recombinases de la famille des Rad52. Pour cela, des séquences inversées répétées de 800 pb ont été placées autour du promoteur P_L de λ, promoteur permettant la transcription des gènes de recombinaison du phage. Ces séquences sont soit identiques soit divergées à 22%. Lors du cycle lytique de λ, ces séquences peuvent recombiner et inverser le sens du promoteur, supprimant ainsi la transcription des gènes de recombinaison. Ce phénotype est quantifiable sur boite : un λ sauvage sera capable de se répliquer sur une souche *d'E. coli* délétée pour *recA* mais pas dans une souche lysogène pour P2. Inversement un λ n'exprimant pas sa recombinase ne se répliquera pas dans une souche d'*E. coli* délétée pour *recA* alors qu'elle en sera capable dans une souche lysogène pour P2. On calcule ensuite la fréquence de recombinaison lors d'un seul cycle lytique en faisait croitre les phages sur une souche *recA*. De cette manière seuls les λ non recombinés peuvent se reproduire. Les recombinants ne pouvant se propager, ceux produits lors du dernier cycle seront prédominants, ainsi la fréquence mesurée est le taux de recombinaison pour seul cycle. Dans ces phages, les gènes encodant la recombinase *red*β et son exonuclease associée *redα* ont été remplacés par ceux du prophage défectif d'*E. coli* Rac, *recT* et *recE* respectivement, ou ceux du phage D3 infectant *Pseudomonas aeruginosa*, *erf* et *exo*. Ces deux recombinases appartiennent à la famille des Rad52.

Les résultats concernant RecA et Redβ ont été publiés par Martinsohn et al. en 2008. Les auteurs ont montré que Redβ recombine ces séquences 10 fois mieux que RecA lorsque les séquences sont identiques, et 100 fois mieux lorsque les séquences sont divergées à 22%. De manière remarquable, Redβ recombine ces séquences divergées à 22% aussi bien que RecA recombine les séquences identiques. Le même résultat à été reproduit avec RecET de rac et ErfExo de D3. Ces deux recombinases recombinent au même niveau que Redαβ, à la fois sur séquences identiques et divergées (Figure 2). Ces résultats montrent qu'au moins 3 recombinases de la famille des Rad52 sont capables de recombiner des séquences divergées mieux que la recombinase de

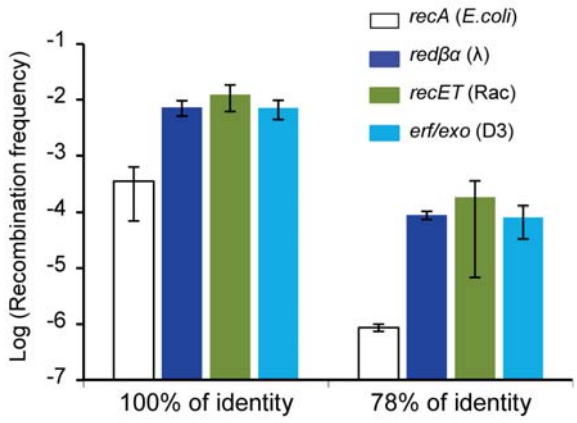


Figure 2 : Haute efficacité de recombinaison de deux autres Rad52. Moyenne des fréquences d'inversion du promoteur P_L suivant la présence de différentes recombinases précisées sur le coté. Les barres d'erreur sont des écart-types calculés à partir de 3 tests de recombinaison indépendants.

l'hôte, et ce pour des phages infectant au moins deux espèces bactériennes différentes. Ceci pourrait être généralisable à toute la superfamille Rad52, laissant supposer que cette famille de recombinases, la plus abondante dans les génomes de phages, pourrait participer au phénomène d'échange de matériel génétique entre phages grâce à cette fidélité diminuée pour la recombinaison.

Pour vérifier cette hypothèse, on devrait être capable de retrouver des traces de recombinaisons entre séquences divergées dans les génomes de phage par comparaison des génomes. En théorie (Figure 3), deux phages ancestraux ayant une identité moyenne de 60% sur tout leur génome avec deux zones de plus forte homologie, à environ 80% d'identité, peuvent s'échanger les gènes compris entre ces deux zones moins divergentes grâce à leurs recombinases. Plus tard, lorsqu'on compare le nouveau phage, mosaïque des deux phages ancestraux, à l'un des parents, on devrait observer une forte zone d'homologie (appelée mosaïque pour plus de simplicité), qui correspond à la zone récemment échangée, entourée de zones de moins forte homologie, mais supérieure à celle du reste de l'alignement des deux génomes, qui correspondrait aux traces de recombinaison homologue sur séquences divergées. Pour tester cette idée, 84 génomes de phages infectant *E. coli* ont été analysés, dont 24 tempérés, 26 virulents et 34 prophages défectifs. Ces génomes ont été comparés deux à deux par BLAST et les résultats ont été filtrés. Seules les régions ayant plus de 90% d'identité sur au moins 100 pb ont été conservées. Ces régions sont considérées comme des mosaïques, des zones d'échanges récents autour desquelles il serait possible de trouver des traces de recombinaison entre séquences divergées. Pour rechercher ces traces de recombinaison, les 2 kb flanquant la mosaïque sur chaque phage sont analysés : un alignement par programmation dynamique (Needlemann et Wunsch) est réalisé. Le pourcentage d'identité de cet alignement est ensuite déterminé par fenêtres successives de 50 pb. Il y a trace de recombinaison si l'on détecte au flanc de la mosaïque une des régions ayant un score d'identité intermédiaire entre la mosaïque et le reste du génome (un score d'identité inférieur d'au moins 10% à celui de la mosaïque et supérieur d'au moins 10% au score médian des identités trouvées pour toutes les fenêtres). Les

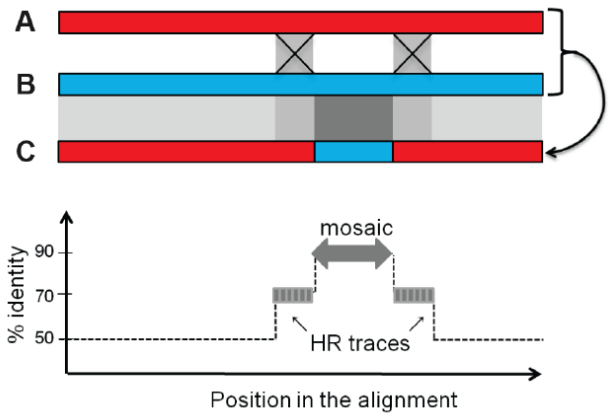


Figure 3 : Schéma de la détection bioinformatique des mosaïques. 2 phages ancestraux A et B partageant des zones d'identité, où il y a possibilité de recombiner, peuvent s'échanger du matériel génétique pour donner le phage C. Le pourcentage d'identité attendu le long de l'alignement de B contre C est présenté dans le graphique placé sous le schéma, la mosaïque est à 90% d'identité et des traces de recombinaisons entre séquences divergées autour sont à 70% d'identité.

résultats sont présentés dans le Tableau 1.

Paires de phages	T-T*	T-D	D-D	V-V	V- autres
N de paires de génomes analysés	266	807	549	289	1508
N de résultat de Blast	711	895	670	16	0
% d'IS	9.7	26.7	33.6	0.0	0
N de mosaïques analysées pour les traces de RH (sans IS, av id%<70%)	635	631	430	16	0
Densité mosaïques/paire	2.39	0.78	0.78	0.06	0
Taille médianes des mosaïques (bp)	404	494	458	131	0
% id médian	96.1	94.8	95.1	91.5	0
N de mosaïques sans traces de RH	432	411	298	5	0
N avec 1 traces de RH	169	171	106	6	0
N avec 2 traces de RH	34	49	26	5	0
P (≥ 1 trace de RH au hasard)	$<10^{-16}$	$<10^{-16}$	$<10^{-16}$	$1.8\ 10^{-14}$	

*T, tempérés, D, défectifs, V, virulents.

Tableau 1 : Données collectées sur les mosaïques par catégorie de paires de genomes compérées.

Avec notre définition des mosaïques, nous avons obtenu quelques paires de phages virulents ayant échangé récemment de l'ADN (mais ces mosaïques étaient bien moins nombreuses qu'entre autres catégories de phages). Cependant, les virulents ne partagent pas de mosaïques avec les tempérés et les défectifs. Chez les défectifs, on observe que 30% des mosaïques sont en fait des séquences d'insertion (IS), contre 10% chez les tempérés, donc des séquences qui recombinent par transposition, plutôt que par recombinaison homologue. Ces « mosaïques » sont retirées de l'analyse, car non pertinentes pour la recherche de traces de recombinaison homologue. Il reste alors moins d'une mosaïque par paire de génomes en moyenne lorsqu'on compare les défectifs entre eux, et plus de 2 mosaïques en moyenne lorsqu'on compare les tempérés entre eux. Deux tiers des mosaïques de virulents possèdent au moins une trace de recombinaison homologue sur séquences divergées à leurs bornes, contre environ 30% pour les tempérés et défectifs. Lorsqu'on regarde plus en détail comment se répartissent les mosaïques entre les paires de génomes de phages, deux groupes se distinguent clairement chez les tempérés : les phages similaires à λ et ceux semblables à P2 ou Mu. Cette distinction peut se faire aussi chez les défectifs avec des échanges supplémentaires entre ces deux groupes. De tels échanges semblent donc possibles mécanistiquement entre phages de familles distinctes, mais leur absence au sein des phages fonctionnels suggère que ces échanges ne sont pas viables pour les phages. Une plus faible proportion (30%) de traces de recombinaison entre séquences divergées ont été retrouvées chez les phages tempérés et défectifs, comparé aux phages virulents (66%). Cela peut s'expliquer par une probabilité de rencontre différente entre les

phages tempérés et virulents. En effet, les phages tempérés pouvant s'insérer dans le chromosome de leur hôte ont pu échanger plus facilement avec un autre phage infectant par la suite la même bactérie. Un échange récent pourrait alors cacher une trace de recombinaison d'un échange plus ancien.

Pour conclure sur ce chapitre, nous avons démontré que les recombinases de la famille des Rad52 ont la capacité de recombiner les séquences divergées, et qu'elles permettent la capture des gènes dans les prophages d'*E. coli*. De plus notre étude montre que les phages tempérés échangent beaucoup de matériel génétique entre eux mais aussi avec les prophages défectifs, et ce potentiellement grâce à leurs recombinases. On peut maintenant se demander si les autres familles de recombinases ont les mêmes capacités que les Rad52, ou si elles s'approchent plus de la recombinase de l'hôte en termes de fidélité et d'efficacité.

Chapitre III : La recombinase phagique Sak4, constituée du seul domaine central, a besoin de sa SSB partenaire pour accomplir l'appariement simple brin *in vivo*.

La recombinaison homologue est un mécanisme clé pour la survie et l'évolution des espèces vivantes. Elle participe à la réparation des dommages de l'ADN et à l'échange de matériel génétique. La protéine clé de ce mécanisme est la recombinase, qui s'appelle RecA chez les bactéries. Comme décrit en introduction, la diversité des recombinases est plus grande chez les bactériophages, puisqu'il existe au moins trois superfamilles décrites : les Rad51, les Rad52 et les Gp2.5. Comme nous l'avons vu au chapitre II, les recombinases de type Rad52 ont la faculté de recombiner les séquences divergées. Cette capacité peut-elle être étendue aux autres superfamilles de recombinases ? Dans ce chapitre, nous nous sommes intéressés aux recombinases de la superfamille des Rad51, et plus particulièrement à une famille encore inexplorée parmi celles-ci : les Sak4. Celles-ci ne sont composées que du domaine cœur commun à toutes les Rad51, permettant de dégrader l'ATP et de se lier à l'ADN. La question principale que nous nous sommes posée est : cette famille a-t-elle les mêmes capacités que les recombinases bactériennes, c'est-à-dire de recombiner par un mécanisme d'invasion de brin, ou est-elle plus proche des propriétés des Rad52, qui recombinent par un mécanisme d'appariement simple brin, et ce avec quelle fidélité ?

Dans un premier temps nous avons étudié le contexte génomique des Sak4 afin de choisir des candidats d'études. Nous avons trouvé qu'un gène codant pour une SSB était présent sur 80%

des génomes encodant une Sak4, et à proximité de ce gène. A l'inverse, tous les génomes présentant ces SSB codent pour une Sak4. Nous avons aussi cherché dans le voisinage local des partenaires de type exonucléase, comme retrouvés avec les Rad52. 25% des génomes codant pour une Sak4 codent aussi pour des exonucléases de la famille de RecE. Enfin, un génome code pour Abc2, une protéine qui détourne l'exonucléase principale de l'hôte RecBCD. Nous avons donc sélectionné le phage HK620, infectant *E. coli*, codant pour une Sak4, une SSB et un Abc2, et le phage PA73, infectant *Pseudomonas aeruginosa*, codant pour une Sak4, une SSB et une RecE.

Afin de tester la capacité d'appariement simple brin *in vivo*, nous avons utilisé un test de recombinaison par « recombineering », qui consiste à promouvoir l'appariement d'un oligonucléotide à la fourche de réplication du chromosome bactérien, dans le but de rétablir une résistance au chloramphénicol. Pour cela, les gènes des recombinases, seuls ou accompagnés des gènes des partenaires potentiels repérés par l'analyse bio-informatique, ont été clonés dans un vecteur sous le contrôle d'un promoteur inductible à l'arabinose. Le test de recombineering a été effectué dans une souche d'*E. coli* délétée pour *recA,* afin d'étudier l'activité de la seule recombinase phagique, et *mutS*, afin d'éviter l'inhibition du recombineering par le système de reconnaissance des mésappariements. Dans ce contexte génétique, chacun des plasmides codant pour une recombinase est introduit. Enfin, après induction à l'arabinose de la recombinase, les cellules sont préparées pour l'électroporation, et transformées par un oligonucléotide de 81 pb complémentaire à la région du gène *cat* muté, qui doit être réparée pour conférer la résistance au chloramphénicol. Si l'appariement de cet oligonucléotide sur la matrice du brin indirect au niveau de la fourche de réplication du chromosome bactérien réussit, la bactérie devient résistante au chloramphénicol, et donc sélectionnable sur boite. La fréquence d'appariement simple brin se mesure par le rapport du nombre de cellules résistantes au chloramphénicol au nombre total de cellules traitées (Figure 4).

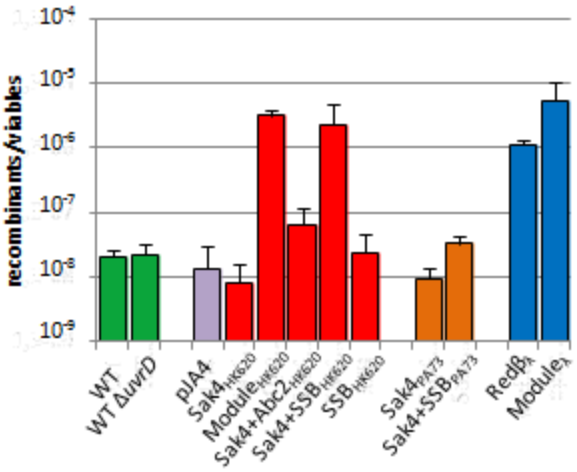


Figure 4 : Activité d'appariement à la fourche de réplication. Efficacité d'appariement de 4 rececombinases avec leurs partenaires : RecA d'*E.coli*, Sak4 de HK620, Sak4 de PA73 et Redβ de λ.

Dans un premier temps, les recombinases sans gène accessoire supplémentaire ont été testées. La recombinase de l'hôte, RecA, n'est pas capable de faire de l'appariement simple brin à la fourche de réplication. Redβ, du phage λ, apparie 100 fois mieux que l'appariement spontané sans recombinase. Les deux Sak4 testées, celle de HK620 et celle de PA73, ne favorisent pas l'appariement simple brin. Nous avons ensuite testé les Sak4 en présence de tous leurs partenaires. Malheureusement, la construction contenant les partenaires présents sur PA73 s'est avérée toxique pour la bactérie, et n'a pu être testée. Cependant la construction du plasmide contenant les trois gènes voisins de *sak4* présents sur HK620 a été obtenue et testée : l'activité d'appariement est augmenté de 100 fois. Nous avons cherché à savoir si un partenaire particulier était responsable de cette amélioration ou si plusieurs jouaient un rôle. Il s'est avéré que la présence de Sak4 et de sa SSB étaient suffisantes pour autoriser l'appariement simple brin à la fourche de réplication. La SSB seule ne dépasse pas le niveau d'appariement spontané et donc ne participe pas elle-même à l'appariement simple brin.

Toujours avec ce test, des oligonucléotides de plus en plus divergés par rapport à la matrice chromosomique ont été utilisés pour tester la fidélité de Sak4. En présence de tous ses partenaires, Sak4, à l'instar de Redβ, est capable d'apparier à la fourche de réplication des oligonucléotides divergés jusqu'à 12%.

Les Sak4 et les SSB provenant des deux phages choisis au départ ont ensuite été purifiées, ainsi que Redβ de λ et UvsX de T4. Pour cela, chacune des ces protéines a été clonée dans un vecteur d'expression, fusionnée avec la protéine GST, capable de fixer le glutathion. Après une première étape de purification sur une colonne de glutathion, la fusion est clivée grâce à la protéase du TEV (Tobaco Etch Virus) et la recombinase ou la SSB est récupérée. Les deux SSB, Redβ et UvsX ont été purifiées jusqu'au bout de ce procédé, mais la difficulté à solubiliser les Sak4 nous a limité. Seulement une fusion de GST avec la Sak4 en provenance de PA73 a été obtenue.

Chacune de ces protéines, ainsi que RecA et SSB d'*E. coli* qui ont été achetées dans le commerce, ont été étudiées biochimiquement. Tout d'abord, nous avons pu confirmer que les SSB de phages étaient capables de retarder un oligonucleotide lors de la migration dans un gel d'agarose, à l'instar de la SSB d'*E. coli*. Les recombinases ont ensuite été testées pour leur activité ATPasique : RecA et UvsX sont capables de dégrader l'ATP en présence d'ADN simple brin, contrairement à Redβ, qui est incapable de dégrader l'ATP. Ces résultats sont retrouvés dans un test classique de dégradation de l'ATP par détection de l'oxydation du NADH : chaque molécule d'ATP hydrolysée permet d'oxyder le NADH en NAD, changeant ainsi l'absorbance de la solution à une longueur d'onde de 340 nm. Cependant, Sak4 ne dégrade pas l'ATP malgré la présence du site de fixation de l'ATP

dans sa séquence. Des résultats similaires ont été rapportés dans la littérature pour RadB, une protéine d'archées composée uniquement du domaine cœur de RecA : elle est capable de fixer l'ATP, de le cliver, mais non de relâcher le produit de dégradation : l'ADP. Si tel est le cas pour GST-Sak4 dans notre réaction, l'ADP n'est pas relâché, le NADH ne sera pas oxydé et cette réaction sera invisible.

RecA et UvsX sont capables d'envahir un ADN double brin à partir d'un ADN simble brin, à condition qu'il soit l'exact complémentaire d'un des brins de l'ADN double brin. Le test classique de cette activité est l'échange de brin. Le but est d'apparier un ADN simple brin circulaire sur un ADN double brin linéaire grâce aux recombinases, et d'échanger ainsi un des deux brins du double brin par le simple brin. Ce test à été effectué avec RecA et GST-Sak4. Comme attendu, RecA est capable d'échanger les deux brins en présence de la SSB d'*E. coli*. En revanche, GST-Sak4 n'en est pas capable, même en présence de la SSB de l'hôte ou de la SSB de PA73. Cependant GST-Sak4 semble avoir un effet inhibiteur sur RecA s'il capture l'ADN simple brin avant RecA. Ce résultat est à nouveau comparable à l'activité publiée de RadB, provenant des Archées. Elle ne fait pas d'invasion de brin non plus et inhibe légèrement RadA, l'homologue de RecA chez les Archées. Cependant dans le cas de Sak4, nous ne pouvons pas exclure que l'échange de brin soit empéché par le domaine GST fusionné à Sak4. Il est donc difficile de conclure sur cette activité pour Sak4.

Enfin, l'activité d'appariement simple brin a été testée *in vitro*. Pour cela, deux oligonucléotides complémentaires de 81 nucléotides, dont un marqué au Cy5, sont mis en solution, en présence ou non de recombinases, et incubés 7 minutes à 30°C, dans le but qu'il s'apparie l'un à l'autre. Le simple brin et le double brin ne migrent pas à la même vitesse sur gel d'accrylamide et peuvent donc être séparés et quantifiés. Redβ, notre contrôle positif, fait effectivement de l'appariement simple brin. RecA, que nous pensions être un contrôle négatif, en fait aussi, ainsi qu'UvsX et GST-Sak4, les recombinases testées. Des mésappariements ont été ajoutés entre les deux oligonucléotides pour tester la fidélité des recombinases dans la reconnaissance de séquences complémentaires. Toutes les recombinases testées, à savoir Redβ, RecA, UvsX et GST-Sak4, sont capables d'apparier des séquences divergées aussi bien que des séquences identiques, et ce jusqu'à 20% de divergence pour Sak4, 28% pour les autres. Des tests d'inhibition par les SSB lors de l'appariement simple brin ont aussi été mis en œuvre après avoir vérifié que ces SSB n'étaient pas capables à elles seules d'apparier deux oligonucléotides. Celle provenant d'*E. coli* inhibe toutes les recombinases, à l'exception de Redβ pour laquelle l'inhibition est partielle. De manière intéressante, contrairement à SSB d'*E. coli*, la SSB de PA73 n'a aucun effet sur l'appariement simple brin dû aux recombinases.

Il est difficile de conclure sur l'activité de Sak4 en général, puisque nous avons caractérisé celle de HK620 seulement *in vivo* et celle de PA73 seulement *in vitro*. Cependant, ces différentes expériences nous on permis de constater qu'à l'instar de toute les recombinases de phages, Sak4 fait de l'appariement simple brin. Pour RecA, cette activité bien qu'observée *in vitro*, est inhibée *in vivo*. Nous avons aussi vu qu'*in vivo*, la SSB accompagnant Sak4 sur HK620 est suffisante pour que Sak4 puisse faire de l'appariement simple brin, alors qu'*in vitro* pour la Sak4 de PA73, la SSB ne stimule en rien l'appariement simple brin, qui est efficace, même pour la protéine fusion GST-Sak4.

Chapitre IV : Discussion

Je replacerai ici mon travail de thèse dans le cadre de différents travaux qui ont tenté de tracer l' histoire évolutive et les activités de chaque famille de recombinases. Je reviendrai également sur les effets que pourraient avoir ces recombinases sur la structure des génomes, à la fois des bactéries et des phages.

Les recombinases de type Rad52 sont présentes dans les génomes de phages, mais pas chez les bactéries. Chez les Eucaryotes, on retrouve cette protéine fusionnée à un domaine C-terminal, qui réoriente la fonction de la protéine vers une activité de chargement de Rad51. Je propose que cette famille de recombinases soit une famille spécifique des génomes de phages, spécialisée dans la recombinaison entre séquences divergées, qui a été domestiquée par les Eucaryotes à partir d'une copie apportée par un génome de bactérie endocytée par ces derniers, bactérie qui contenait peut-être un prophage. Les plus étudiées des recombinases de cette famille sont la recombinase du phage λ, Redβ, et la recombinase du prophage *rac* d'*E. coli*, RecT. Ces recombinases ont plusieurs activités, dont une est plus efficace que les autres. La plus efficace de ces activités est l'appariement simple brin, qui consiste à apparier deux ADN simple brin complémentaires d'origines différentes. Le modèle le plus approprié pour expliquer nos résultats consiste à proposer qu'à partir d'un fragment linéaire d'ADN double brin, une exonucléase dégrade ce fragment de sorte qu'il soit rendu entièrement simple brin. Redβ peut alors se fixer à ce substrat, et chercher de l'homologie à la fourche de réplication, le fragment d'ADN simple brin prend alors le rôle de fragment d'Okazaki. Cependant, ce modèle ne permet pas d'expliquer les recombinaisons allant jusqu'à un échange de grands fragments, d'une taille supérieure à la région laissée simple brin avant synthèse des fragments d'Okasaki (1 à 2 kb), comme ceux qui s'échangent dans le test de capture de gène de résistance, décrit plus haut. La deuxième activité, une certaine forme d'échange de brin, pourrait mieux expliquer des échanges de grands fragments. En effet Redβ, une fois qu'elle a apparié deux

ADN simple brin, est capable de déplacer le brin excédentaire pour continuer l'appariement du brin invasif. De plus, il est courant que l'ADN baille, c'est-à-dire qu'il s'ouvre de lui-même, permettant ainsi à Redβ de commencer l'appariement sur un substrat simple brin. Cette propriété de l'ADN est surtout retrouvée dans les ADN circulaires super-enroulés. Enfin la dernière activité de ces recombinases pourrait être la recombinaison illégitime avec des zones de micro-homologie. Cela consiste à relier ensemble deux ADN double brin cassés grâce à des homologies très petites, en dessous de 10 pb. Il est enfin important de rappeler une autre particularité des Rad52 phagiques , leur capacité 10 à 100 fois plus élevée que RecA à apparier des séquences divergées *in vivo*. Ce qui veut dire qu'elles arrivent à échapper au système de réparation des mésappariements.

Les recombinases phagiques de type Rad51 appartiennent à la famille ubiquitaire de recombinases, qui se retrouve dans les trois règnes de la vie. Chez les phages, elles se déclinent en deux catégories : les Sak4, plus présentes dans les génomes de phages tempérés, et les UvsX, plus présentes dans les génomes de phages virulents. Toutes les recombinases de cette superfamille partagent un domaine cœur, capable d'hydrolyser l'ATP et de lier l'ADN, et la plupart d'entre elles ont un domaine supplémentaire, en C-terminal pour RecA et UvsX, ou en N-terminal pour celle provenant des eucaryotes et des archées. Ce domaine a le plus souvent une activité de liaison à l'ADN. Ceci laisse penser que l'ancêtre commun à toutes ces recombinases n'aurait que le domaine cœur, comme Sak4 ou encore d'autres recombinases, comme RadB chez les archées, ou XRCC2 chez l'homme. RecA, RadA, Rad51, DMC1 et UvsX sont toutes capables de faire de l'invasion de brin, et recombinent de manière fidèle. Cependant, Sak4 ne fait pas d'invasion de brin, son activité serait plutôt de type appariement simple brin. Nous avons vu au chapitre III que RecA et UvsX sont elles aussi capables d'appariement simple brin *in vitro*, cependant RecA n'en fait pas *in vivo*. De plus Sak4 apparie *in vivo* des séquences divergées jusqu'à 12%. Peut-être que cette fidélité moindre pourrait être liée à l'activité d'appariement simple brin, alors que l'invasion de brin oblige à une recombinaison fidèle.

Les recombinases de type Gp2.5 sont pour l'instant très peu étudiées. Elles se trouvent uniquement chez les Herpès virus et les phages, laissant supposer que : elle tire son origine d'un virus infectant l'être vivant à la racine de l'arbre du vivant, ou qu'il y a eu transfert horizontal entre ces deux entités. A l'heure actuelle, du côté des phages, seule celle en provenance du phage T7, infectant *E. coli*, a été étudiée. Elle est capable de faire de l'appariement simple brin. De plus, en présence d'une hélicase, une protéine ouvrant la double hélice d'ADN, cette Gp2.5 fait de l'invasion de brin. Il n'y a aucune étude quant à la fidélité de cette famille de recombinases.

Si on considère que l'activité d'appariement simple brin des recombinases de phages est liée à leur capacité à recombiner les séquences divergées, on peut considérer qu'elles sont à la fois un atout majeur pour l'évolution des génomes de phages et un désavantage pour les génomes bactériens. En effet les phages sont très divers et, lorsqu'on les compare au niveau de la séquence de leur génome en acides nucléiques, ils sont très divergés les uns par rapport aux autres, y compris au sein d'une même famille. Ils échangent cependant du matériel génétique, surement grâce à la faible fidélité et haute efficacité de leur recombinases, pour ceux qui en portent une. Quant aux génomes bactériens, il est dangereux pour eux d'avoir ce genre de recombinases. En effet les recombinases de l'hôte préservent l'intégrité du chromosome bactérien grâce à leur activité fidèle. Les recombinases qui apparaissent dans les prophages ne sont pas, ou très peu, exprimées afin d'éviter les dommages, comme par exemple des délétions majeures ou des inversions chromosomiques. Enfin, il serait peut-être envisageable de dériver de ces travaux une notion d'espèces pour les phages. En effet, si des phages sont capables d'échanger du matériel génétique, grâce aux recombinases ou par un autre procédé, de manière à créer une progéniture viable, alors sans doute ces phages appartiennent à la même espèce.

Dans cette thèse, j'ai montré que la fidélité relâchée est un trait général des recombinases des familles des Rad52 et des Sak4. La discussion ci-dessus a tenté de tracer les conséquences de ce fait. Je propose que l'ancêtre commun aux Rad51 ne soit composé que du domaine cœur commun à toutes les protéines de la superfamille, qui par la suite a acquis des domaines supplémentaires pour se spécialiser. De la même manière, les recombinases de la famille des Rad52, qui ont une fidélité relâchée, ont été domestiquées par les eucaryotes par l'acquisition de domaines supplémentaires, comme la Rad52 eucaryote, ou utilisés comme cofacteur de recombinaison, comme Rad59. Enfin, je propose une nouvelle méthode de classification des phages, en excluant les prophages défectifs, en se basant sur la fidélité diminuée des recombinases phagiques : deux phages capables d'échanger du matériel génétique appartiendraient à la même espèce.

Résumé

La diversité des génomes de virus infectant les bactéries, les bactériophages (ou phage en abrégé), est telle qu'il est difficile de les classer de manière satisfaisante, la notion d'espèce elle-même ne faisant pas accord dans la communauté scientifique. A la racine de cette diversité, un des facteurs clé est la recombinaison de l'ADN, qui est élevée chez les bactériophages, et permet des échanges de gènes entre entités parfois fort différentes. Mes travaux se sont centrés sur la recombinaison homologue chez les bactériophages, et en particulier sur la protéine centrale de ce processus, la recombinase. J'ai montré pour deux grands types de recombinases phagiques, de type Rad52 et Sak4, que celles-ci étaient beaucoup moins fidèles dans le processus de recombinaison, comparées à la recombinase bactérienne RecA. De plus, pour Sak4, j'ai observé que cette recombinaison se produisait par appariement simple brin, et qu'elle dépendait entièrement *in vivo* d'une SSB phagique, dont le gène est situé à proximité du gène *sak4* sur le chromosome du phage. Les échanges génétiques sont donc grandement facilités pour les phages contenant ce type de recombinases, mais ils ne sont pas non plus anarchiques : la recombinaison s'observe jusque 22% de divergence, mais deux séquences à 50% de divergence ne peuvent recombiner. Tout se passe donc comme si la notion d'espèce devait être élargie chez les phages par rapport aux bactéries, pour inclure dans un même groupe des génomes portant des traces d'échanges récents de matériel génétique par recombinaison homologue (ce que l'on appelle le mosaïcisme).

Summary

The diversity of the viruses infecting bacteria (bacteriophages, or phages for short) is so important that it is difficult to classify them in a pertinent way, and the species notion itself is a matter of debate among specialists. At the root of this diversity, one of the key factors is DNA recombination, which occurs at high levels among phages, and permits gene exchanges among entities that are sometimes very distant. My research has focused on homologous recombination in phages, and in particular on the protein that is key to the process, the recombinase. I have shown, for two different types of recombinases, Rad52-like and Sak4-like, that their fidelity was relaxed, compared to the bacterial recombinase, RecA. Moreover, for Sak4, a protein that had not been studied before, I showed that recombination occurs by single strand annealing, and that it is strictly dependent *in vivo* on the co-expression of its cognate SSB protein, whose gene is often encoded nearby in phage genomes encoding *sak4.* Genetic exchanges are therefore greatly facilitated for phages encoding these types of recombinases. Nevertheless, exchanges are not anarchical: recombination is seen up to 22% diverged substrates, but 50% diverged DNA sequences will not recombine. It may be that the species notion should be enlarged for phages, so as to include into a same group all phages exhibiting traces of recent exchanges of genetic material (the so-called mosaicism).