



HAL
open science

Expanding Diversity of Firmicutes Single-Strand Annealing Proteins: A Putative Role of Bacteriophage-Host Arms Race

Kamil Steczkiewicz, Eric Prestel, Elena Bidnenko, Agnieszka Szczepankowska

► **To cite this version:**

Kamil Steczkiewicz, Eric Prestel, Elena Bidnenko, Agnieszka Szczepankowska. Expanding Diversity of Firmicutes Single-Strand Annealing Proteins: A Putative Role of Bacteriophage-Host Arms Race. *Frontiers in Microbiology*, 2021, 12, 10.3389/fmicb.2021.644622 . hal-03343061

HAL Id: hal-03343061

<https://hal.inrae.fr/hal-03343061v1>

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Expanding Diversity of *Firmicutes* Single-Strand Annealing Proteins: A Putative Role of Bacteriophage-Host Arms Race

Kamil Steczkiewicz¹, Eric Prestel², Elena Bidnenko² and Agnieszka K. Szczepankowska^{1*}

¹ Institute of Biochemistry and Biophysics PAS, Warsaw, Poland, ² Micalis Institute, INRAE, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

OPEN ACCESS

Edited by:

Frank T. Robb,
University of Maryland, Baltimore,
United States

Reviewed by:

Xu Peng,
University of Copenhagen, Denmark
Eric Miller,
North Carolina State University,
United States

*Correspondence:

Agnieszka K. Szczepankowska
agaszczep@ibb.waw.pl

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 19 January 2021

Accepted: 25 March 2021

Published: 20 April 2021

Citation:

Steczkiwicz K, Prestel E, Bidnenko E and Szczepankowska AK (2021) Expanding Diversity of *Firmicutes* Single-Strand Annealing Proteins: A Putative Role of Bacteriophage-Host Arms Race. *Front. Microbiol.* 12:644622. doi: 10.3389/fmicb.2021.644622

Bacteriophage-encoded single strand annealing proteins (SSAPs) are recombinases which can substitute the classical, bacterial RecA and manage the DNA metabolism at different steps of phage propagation. SSAPs have been shown to efficiently promote recombination between short and rather divergent DNA sequences and were exploited for *in vivo* genetic engineering mainly in Gram-negative bacteria. In opposition to the conserved and almost universal bacterial RecA protein, SSAPs display great sequence diversity. The importance for SSAPs in phage biology and phage-bacteria evolution is underlined by their role as key players in events of horizontal gene transfer (HGT). All of the above provoke a constant interest for the identification and study of new phage recombinase proteins *in vivo*, *in vitro* as well as *in silico*. Despite this, a huge body of putative *ssap* genes escapes conventional classification, as they are not properly annotated. In this work, we performed a wide-scale identification, classification and analysis of SSAPs encoded by the *Firmicutes* bacteria and their phages. By using sequence similarity network and gene context analyses, we created a new high quality dataset of phage-related SSAPs, substantially increasing the number of annotated SSAPs. We classified the identified SSAPs into seven distinct families, namely RecA, Gp2.5, RecT/Red β , Erf, Rad52/22, Sak3, and Sak4, organized into three superfamilies. Analysis of the relationships between the revealed protein clusters led us to recognize Sak3-like proteins as a new distinct SSAP family. Our analysis showed an irregular phylogenetic distribution of *ssap* genes among different bacterial phyla and specific phages, which can be explained by the high rates of *ssap* HGT. We propose that the evolution of phage recombinases could be tightly linked to the dissemination of bacterial phage-resistance mechanisms (e.g., abortive infection and CRISPR/Cas systems) targeting *ssap* genes and be a part of the constant phage-bacteria arms race.

Keywords: single strand annealing proteins (SSAP), phage recombinase, *Firmicutes* bacteriophages, CLANS, CRISPR/cas, abortive infection, Sak3/DUF1071, phage-bacteria arms race

INTRODUCTION

The great genetic diversity and abundance of bacteriophages (phages) on earth determine their profound impact on the bacterial world and the biosphere on the whole (Srinivasiah et al., 2008; Kristensen et al., 2010; Hatfull and Hendrix, 2011). Rapid bacteria-phage co-evolution was suggested in order to explain the structure and dynamics of different bacterial populations and the extensive horizontal gene transfer (HGT) between different bacterial and phage (prophage) genomes (Canchaya et al., 2003; Pal et al., 2007). Thus, phages are commonly considered as a major driving force in bacterial evolution in different natural environments, including the human or animal body (Gómez and Buckling, 2011; De Paepe et al., 2014b; De Sordi et al., 2019).

DNA recombination is a keystone of HGT; the almost universal RecA recombinase encoded by nearly all bacterial genomes sequenced so far plays a vital role in this process (Rocha et al., 2005). Phages rarely encode classic RecA proteins but rather used to own recombinases—DNA single strand annealing proteins (SSAPs), which are diverse in sequence and phyletic distribution. Phage SSAPs substitute for RecA and manage the DNA metabolism at different steps of phage infection (Kuzminov, 1999; Szczepańska, 2009). The importance of phage-encoded SSAPs for HGT is now well documented and cannot be underestimated (De Paepe et al., 2014a).

Among the characteristic features of phage SSAPs is their efficient pairing activity of short DNA sequences (40–50 bp) and the ability to promote recombination between divergent DNA sequences (Murphy et al., 2000; Court et al., 2002; Noirot et al., 2003; Martinsohn et al., 2008). Another distinguishing and intriguing trait of phage SSAPs is their great diversity, presenting a deep contrast to the conserved bacterial RecA protein.

Initially three evolutionary distinct SSAP superfamilies (RecT/Red β , Erf and Rad52) have been identified by bioinformatic analyses (Iyer et al., 2002) and a fourth one (Sak4) by experimental studies (Bouchard and Moineau, 2004). Later it has been proposed that the majority of 465 SSAPs identified within the annotated phage genomes archived in the ACLAME database (version 0.3) of mobile genetic elements can be classified into one of these three superfamilies: Rad52 (incorporating RecT/Red β and Erf superfamilies, and the Sak-like protein family), Rad51 and Gp2.5 (Leplae et al., 2010; Lopes et al., 2010). Six phage SSAP families, namely UvsX, Gp2.5, Sak, RecT/Red β , Erf, and Sak4 have been identified within these superfamilies (Lopes et al., 2010). At present, the total amount of annotated SSAPs continues to increase, not only because of the rising number of available phage/prophage sequences, but also due to the growing performance of methods of sequence analysis. Yet, a huge body of putative *ssap* genes is not taken into account and escapes classifications as it is not annotated as phage-related. Thus, a comprehensive updated classification of SSAPs comprising the current data and determination of the evolutionary relationships between different SSAPs are required.

Special attention in this subject was exercised by the use of some *ssap* genes for recombineering, an efficient method of *in vivo* genetic engineering first developed in *Escherichia coli* (Datsenko and Wanner, 2000; Ellis et al., 2001; Court et al., 2002;

Sawitzke et al., 2007) and adapted for other, mainly Gram-negative, bacterial species (Derbise et al., 2003; Ranallo et al., 2006; Sawitzke et al., 2007; Lesic and Rahme, 2008; van Kessel et al., 2008). Up to now, the use of SSAPs for recombineering is limited to proteins from the RecT/Red β SSAP family which co-occur with their native exonuclease. It is noteworthy that although RecT/Red β recombinases of different origin can function in *E. coli* to catalyze oligonucleotide recombination, yet some at albeit barely detectable thresholds, the double-strand DNA recombination activities with their exonuclease partners have been shown to be inefficient (Datta et al., 2008). Information concerning other SSAPs, especially those encoded by phages and prophages of Gram-positive bacteria, is in short supply. All above justify the constant interest for the identification and study of new phage recombinase proteins *in vivo*, *in vitro* as well as *in silico*.

Although the extreme diversity of phage SSAPs and their large phyletic distribution are now generally recognized and documented, a rational explanation for this has never been proposed. A number of studies have analyzed the role of gene flow in the adaptation of phages to their hosts (Paterson et al., 2010), but the mechanisms and the driving force of evolution as well as the diversity of phage genes have been investigated to a lesser extent. At the same time, there is a significant amount of experimental data supporting the role of phage resistance mechanisms in phage-bacteria co-evolution and phage diversity (Buckling and Rainey, 2002; Labrie and Moineau, 2007; Sorek et al., 2008; Enikeeva et al., 2010; Koskella and Brockhurst, 2014; Broniewski et al., 2020).

Bacteria have developed different effective anti-phage defense strategies: both generalized and highly specific (Sorek et al., 2008; Labrie et al., 2010; Makarova et al., 2011). Among these, Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR)/Cas system and phage Abortive infection (Abi) are known to target specific phage functions essential for intracellular phage development. The most elaborated adaptive defense CRISPR/Cas systems provide immunity against phages and invasive genetic elements. The defense mechanism is based on the incorporation of short DNA sequences (proto-spacers) from the infecting phage into the CRISPR locus. Subsequently, CRISPR transcripts are processed into small interfering RNAs that guide the bacterial protein complex to cleave matching foreign DNA. Spacer sequences homologous to various phage genes were identified in sequenced bacterial genomes (Sorek et al., 2008). Abis are characterized by the blockage of a crucial step of phage intracellular development associated with the premature onset of cell death (Chopin et al., 2005; Labrie et al., 2010). The *abi* genes have been identified in different bacterial phyla, even if the majority of them were characterized for Gram-positive bacteria *Lactococcus lactis* (Chopin et al., 2005; Labrie et al., 2010).

In the present study, we focused on the identification, classification and analysis of SSAPs encoded by the *Firmicutes* (low G+C Gram-positive) bacteria and their phages. The use of a sequence similarity network and gene context analyses allowed us to propose a new dataset of phage-related SSAPs, to significantly increase the amount of annotated SSAPs and to demonstrate the relationships between them. We recognized the Sak3-like proteins as a new distinct SSAP family. In result, the identified

SSAPs were classified into seven distinct families, namely RecA, Gp2.5, RecT/Red β , Erf, Rad52/22, Sak3 and Sak4, organized into three superfamilies. Here, we also suggest that bacterial phage-defense systems specifically targeting phage *ssap* genes could contribute to evolution and dissemination of the *ssap* genes among different phages and bacteria.

MATERIALS AND METHODS

SSAPs Identification and Clustering

Sequences of SSAPs belonging to seven families: RecA (PF00154), DUF2815 (Gp2.5, PF10991), RecT/Red β (PF03837), Erf (PF04404), Rad52/Rad22 (PF04098), DUF1071 (PF06378), and AAA_24 (Sak4, PF13479) were identified using Psi-BLAST searches (3 iterations, inclusion threshold *E*-value 0.001) against the non-redundant (nr) protein sequence database with corresponding family consensus sequences as queries. Partial protein sequences and hits from unknown organisms (e.g., Prokaryotic virus, *Siphovirus* phage, *Firmicutes* bacteria, etc.) were removed after manual inspection. In order to study detailed taxonomic distribution of identified proteins, also redundant hits were assessed. Redundant sequences present in both phage and bacterial genomes have a high chance to indicate a prophage and were separated into individual entries. The resultant set of SSAPs was clustered using CLuster ANalysis of Sequences (CLANS) to identify subfamilies of closely related SSAP sequences and elucidate the relationships between and within the SSAP families (Frickey and Lupas, 2004). CLANS runs BLAST on given sequences, all-against-all, and clusters them in 3D according to their similarity. A 2D-representation was obtained by seeding sequences randomly in the arbitrary distance space.

Gene Context Analysis

In order to confirm the functional annotation of identified SSAP proteins [as suggested by Saha et al. (2020)], immediate genomic neighborhood (three upstream and three downstream genes) for each phage-encoded hit was retrieved from the NCBI database and manually inspected. Bacterial SSAPs were omitted because curation of such a vast dataset would be unfeasible, not to mention that many bacterial hits come from whole-genome sequencing (WGS) data resulting in a number of scattered contigs which hinders reconstruction of genomic neighborhood. Protein domains encoded by the neighboring genes were identified using *pfam_scan.pl* script that wraps HMMER searches against Pfam database (ver. 32).

Identification of Abi- and P2-EC30 RT-Related Proteins

Homologs of the *L. lactis* AbiA (AAA65072.1) and AbiK (WP_011117213.1) and *Enterobacteria* phage P2-EC30 reverse transcriptase (CAJ43154.1) proteins were retrieved using Psi-BLAST searches (five iterations, inclusion threshold *E*-value 0.05) against the non-redundant (nr) protein sequence database. Partial protein sequences were removed manually. The remaining sequences were clustered using CLANS in order to assess the

distinction between highly homologous AbiA-, AbiK-, and P2-EC30-like proteins. For hits obtained after the first iteration, the gene context was inferred from the NCBI to confirm that they are functional homologs of the query proteins.

CRISPR Spacer Analysis

Spacers corresponding to the sequences (proto-spacers) of SSAP-encoding genes were identified by BLASTn searches (at default parameters) against the CRISPRdb database¹ (Grissa et al., 2007). Based on previous reports (Anderson et al., 2011), spacers “matches” were considered given 100% of identity along at least 20 bp. Sequences complying with less stringent thresholds (92–99% identity corresponding to one or two mismatches over the spacer length) were retained for further analyses, but regarded as degenerated. Each *ssap* proto-spacer sequence was compared for the number of matching spacers and the organism (phylum) from which they derive.

RESULTS AND DISCUSSION

Construction and Analysis of the *Firmicutes* SSAP Dataset

In total, we identified more than 35,100 SSAP proteins from *Firmicutes* bacteria (bacterial subset) and their specific phages (phage subset) (Table 1). Notably, less than half of the identified phage SSAPs is annotated in the updated version of the ACLAME database comprising all known prophages and phage genomes (Leplae et al., 2010). The majority of SSAP sequences present in the newly constructed dataset (Supplementary Data Sheet 1) is encoded by chromosomal genes. However, we suggested that the bacterial subset includes unannotated prophages and phage remnant elements allocated, due to their localization, in this fraction of the SSAP dataset.

The sequences from the SSAP dataset were compared all-against-all with BLAST to find clusters of similar sequences and visualize their sequence relationships, as implemented in the CLANS (CLuster ANalysis of Sequences) program (Frickey and Lupas, 2004). This global analysis showed that SSAPs from *Firmicutes* and their phages belong to seven well-defined protein families: RecA (PF00154), Gp1.2 (DUF2815; PF10991), RecT/Red β (PF03837), Erf (PF04404), Rad52/22 (PF04098), Sak4 (AAA_24; PF13479), and Sak3 (DUF1071; PF06378) (Figure 1). Importantly, our analysis revealed Sak3 as a distinct SSAP family, which carries representatives previously classified to Rad52/22 or Sak protein clusters (Lopes et al., 2010).

Thus, all actually identified phage SSAPs are present in our dataset. Not surprisingly, they do not display any regularity in the phyletic distribution. The SSAP families and subgroups within them are formed by proteins from unrelated phages infecting different bacterial species (Supplementary Data Sheet 1). These observations indicate an interesting particularity in the evolution of the SSAP-encoding genes as sequence similarity between phage proteins from different species and infecting diverse hosts tends to be very limited (Hendrix et al., 1999;

¹<https://crispr.i2bc.paris-saclay.fr/>

TABLE 1 | *Firmicutes* SSAP families identified in this study.

Superfamily (Pfam Clan)	SCOP fold	Pfam ID	Family names	No. of Bacterial ssap sequences	No. of Phage ssap sequences
OB	OB-fold	PF10991	Gp2.5 Gp1.2 DUF2815	5,861	126
P-loop_NTPase	P-loop containing nucleoside triphosphate hydrolases	PF00154	RecA	nd	229
DSRM	dsRBD-like	PF13479	Sak4	6,737	339
		PF04404	Erf	5,358	216
		PF06378	Sak3 (DUF1071)	3,499	233
		PF04098	Rad52/22	440	6
		PF03837	RecT Red β	10,517 1,375	212 80

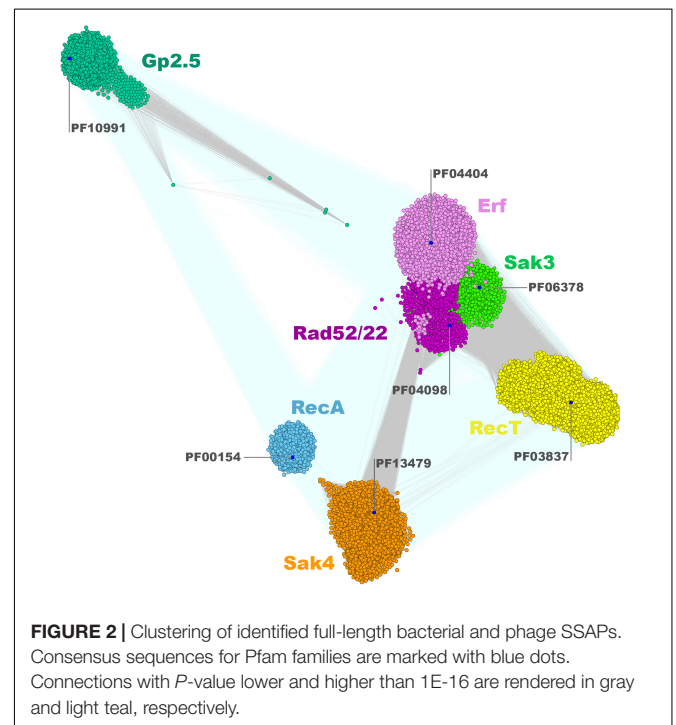
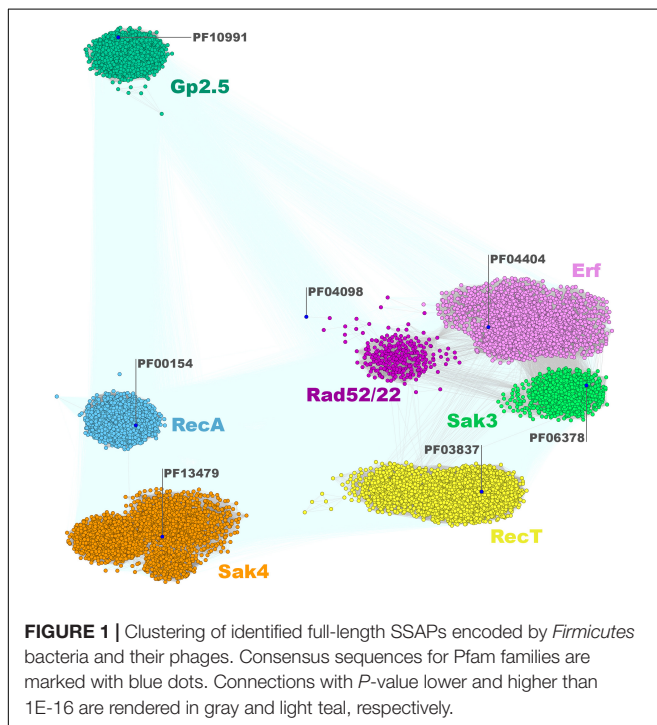
Cresawn et al., 2011). On the other hand, similar phages infecting related hosts encode SSAPs belonging to different families. Thus, *Firmicutes* SSAPs constitute an example of an almost ubiquitous but extremely diverse and randomly distributed group of proteins.

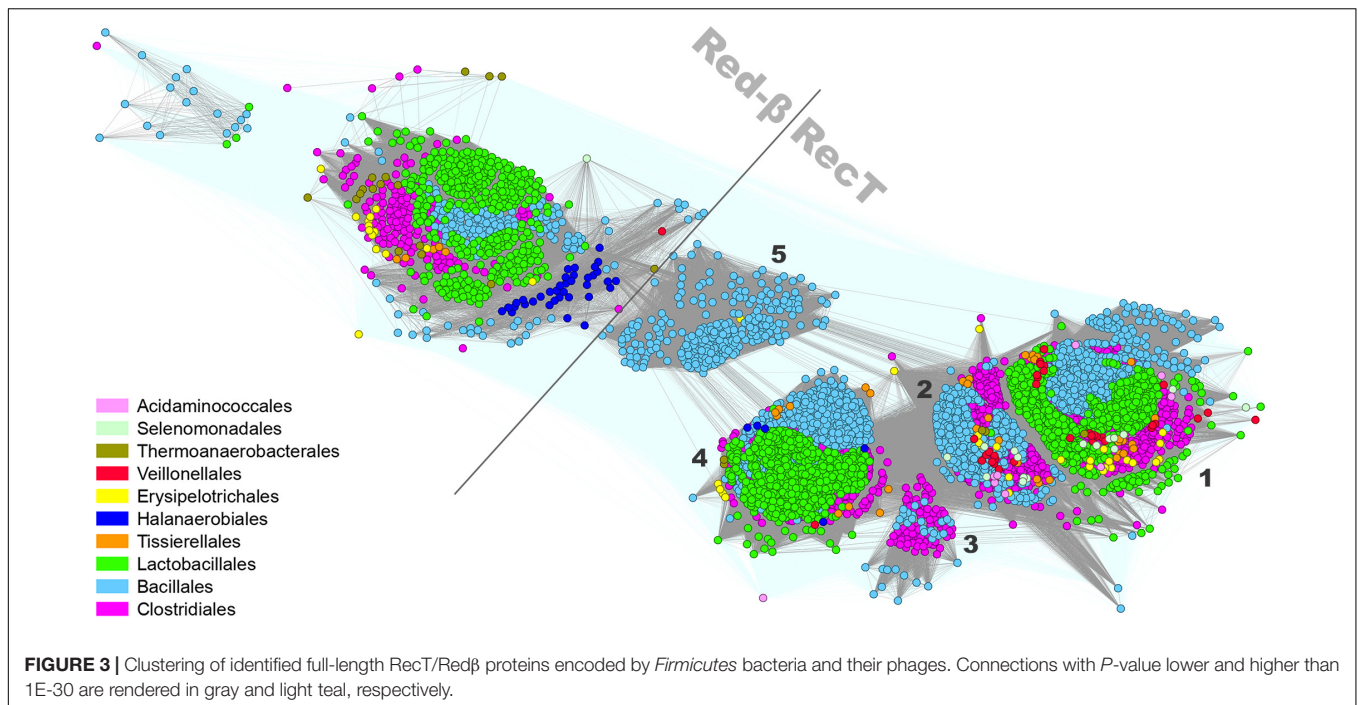
The clustering of the *Firmicutes* SSAP subset matches the overall SSAP sequence similarity network calculated for all Gram-positive and Gram-negative bacterial proteins in general (Figure 2). When considering SSAPs from the Bacteria kingdom, Erf, Sak3, and Rad52/22 families are much more tightly connected, if not overlapping to some extent. In turn, those three families are more pronounced within the *Firmicutes*-only datasets. Moreover, more connections with better *P*-value emerge between Sak4, RecT, and Erf/Sak3/Rad52/22 for the Bacteria subset. Connected representatives from these distinct clusters

share conserved, additional C-terminal DNA-binding domain (HTH) remotely homologous to the Red β C-terminal domain [pdb| 6m9k (30624736)] and ParD DNA binding protein [pdb| 2an7 (17656583)]. This might be a consequence of protein domain shuffling between distinct phages that employ different SSAPs but co-infected the same host. The observed overall agreement between bacterial clustering and the *Firmicutes* subset confirms that *Firmicutes* SSAPs follow the evolutionary history of all SSAPs found in Bacteria. This allows us to assume that results obtained for the *Firmicutes* subset would be also valid for SSAPs in general.

RecT/Red β Superfamily

Proteins belonging to the *Firmicutes* RecT/Red β family (Iyer et al., 2002; Lopes et al., 2010) form six well-defined





subgroups: one Red β -like subfamily of around 1,450 proteins and five RecT-like clusters altogether spanning 10,729 homologs (**Figure 3**). Most of the Red β -like proteins were identified in the genomes of *Streptococcus* sp. (*Lactobacillales*), *Listeria* sp. (*Bacillales*), and *Clostridiales*. The relatively limited occurrence of Red β -like proteins within *Firmicutes* Red β -RecT family presents a contrast to the entirety of *Proteobacteria* where these proteins over count RecT (**Supplementary Figure 1**). This could suggest a relatively recent HGT of the *red* genes from Gram-negative to Gram-positive bacteria. Five subgroups of the RecT family are formed mainly by proteins encoded by: cluster 1 – *Bacillales* (*Bacillus* sp. and *Listeria* sp.), *Lactobacillales* (mainly *Enterococcus* and *Streptococcus* sp.) and *Clostridiales*, cluster 2 – *Bacillales* (mainly *Bacillus* sp.) and *Clostridiales*, cluster 3 – *Clostridiales* (*Clostridium* sp.) and *Bacillales* (*Paenibacillus* sp.), cluster 4 – *Bacillales* (mainly *Staphylococcus* sp.) and *Lactobacillales*, cluster 5 – *Bacillales* (mainly *Listeria* sp.) (**Supplementary Data Sheet 2**). Remarkably, cluster 5 is well separated from the other RecT clusters and localized in-between Red β and RecT both when analyzing *Firmicutes* and *g-Proteobacteria* protein sets. Only three clusters (1, 2, and 4) contain SSAPs from virulent phages while all other members of the RecT/Red β protein family derive from recognized prophages or are chromosomally-encoded.

In *Firmicutes*, genes coding for phage RecT/Red β proteins are localized in proximity to genes engaged in DNA replication, e.g., phage replisome organizer with N-terminal replication initiation domain (Phage_rep_org_N; PF09681) associated with either its C-terminal, low complexity helical domain (rep_org_C, PF06926), replication initiation, and membrane attachment domain (DnaB_2, PF07261) or replication helicases (IstB_IS21-like domain). Three most common direct genomic neighbors include: the YqaJ-like (PF09588) nuclease encoding-gene,

adjacent to *recT* found mostly in *Bacillales* and *Lactobacillales* phages (cluster 1 and few records in cluster 2 and 4), *duf3799*, encoding a PDDEXK-like protein (PF12684), which accompanies *recT* genes mainly in the *Lactobacillales* phage subset (cluster 1 and 4), or *duf1351* (PF07083), located next to either *red β* or *recT* genes of *Lactobacillales* phages (4 cluster). Other functions encoded in direct proximity of the phage-encoded *recT/red β* genes include SSB (PF00436, single-strand binding protein), SMC (PF02463, structural maintenance of chromosome-like protein) or further uncharacterized AAA_ATPases (**Supplementary Data Sheet 1**). In Gram-negative bacteria and their phages, *recT* is found adjacent to gene encoding a *recE*-like exonuclease, together forming the *recET* module involved in homologous recombination (Chang and Julin, 2001), used for recombineering, and functionally analogous to the *red β /exo* module of phage λ (Muyrers et al., 2000). RecE exonucleases assist RecT in catalyzing recombination by digesting DNA extremity in 5'–3' direction (Court et al., 2002). The co-occurrence of the *recT* gene either with YqaJ-like or DUF3799 *exo* is typical for all phage RecT subgroups in *Firmicutes*, except for certain lactococcal and *Streptococcus* phages with the *duf1351* gene instead. The combination *recT-lactamase B2* is characteristic specifically for *Bacillales* prophages (of *Paenibacillus* sp., *Bacillus* sp., *Staphylococcus* sp. hosts) that lack a recognizable *exo* in the proximity of the *ssap* genes. This finding together with the confirmed role of phages in dissemination of drug resistance genes, such as lactamase genes, via the HGT (Muniesa et al., 2004) suggests the spread of the whole *recT-lactamase B2* by this manner. The topological arrangement within the Red β subgroup shows the *ssap-duf1351* gene order to be the most common. Analysis of the present *Firmicutes* SSAP dataset did not reveal any member of the EHAP1-like protein family previously identified

in *E. coli* and *Salmonella* as a distinct subgroup of the RecT/Red β superfamily (Iyer et al., 2002). Given all of the identified SSAP clusters examined throughout this study (see below), the proteins from RecT/Red β family can be detected in over 100 genera and seem to be the most widely distributed among *Firmicutes* taxa.

Rad52/22-Erf-Sak3 Superfamily

Both the Erf-like and Rad52/22 proteins have been recognized as members of the second SSAP superfamily, Rad52/22-Erf (Lopes et al., 2010). Erf proteins are abundant among the identified SSAPs (216 viral subset members; **Table 1**). Based on CLANS comparison, the Erf protein family is also one of the most diffused clades of the Rad52/22-Erf superfamily (**Figure 1**). The greater fraction of these proteins is encoded by temperate phages, prophages or phage remnants residing in *Firmicutes* bacterial genomes of *Lactobacillales* (mainly *Lactobacillus* spp. and *Enterococcus* spp.), *Bacillales* (mainly *Staphylococcus* spp.) and *Clostridiales*. Seventy-seven proteins were found to be encoded by virulent phages, most of which infect *Lactococcus* species (**Supplementary Data Sheet 1**). This revises the previous view (Lopes et al., 2010) that Erf-like proteins are strictly specific to temperate phages.

Examination of gene neighborhoods of the phage subset of *Firmicutes* *erfs* revealed the most frequent association (almost 80%) with *ssb* genes. Other possible downstream genes that may assist *erf* vary and include *rep*, *hnh endo*, *phg_2220_C* (**Supplementary Data Sheet 1**). In the majority of cases, the *erf-ssb* module is neighbored upstream by a gene encoding: (i) the homolog of phage Mu DNA binding protein Gam (Akroyd and Symonds, 1986), (ii) DUF2483- or (iii) DUF1351-type protein, or (iv) the homolog of *S. thermophilus* phage ϕ Sfi21 *orf157* product (*sipho_gp157*) (Foley et al., 1998). The common localization of these genes in respect to the *erf-ssb* module implies their analogous function. The *duf2483-erf* module seems to be restricted to *Staphylococcus* phages, while *duf1351-erf* to *Lactobacillus* phages.

In contrast to *erfs*, which are encoded in the genomes of various *Firmicutes* bacterial hosts, temperate and lytic phages, Rad52/22 sequences were found only in a few genera. Rad52/22 homologs were identified within temperate phages and prophage sequences of bacterial hosts from *Lactobacillales* (*Lactococcus lactis*), *Bacillales* (mainly *Bacillus* spp.), and *Clostridiales* (mainly *Clostridium* spp.) order. The only virulent phages encoding a Rad52/22-like protein are members of the lactococcal phage P335 group, which carry a non-functional, lysogeny module (Labrie and Moineau, 2002; Mahony et al., 2013). Firm clustering of Rad52/22 members indicates strong sequence conservation within this protein family and suggests their close phylogenetic relationships. Remarkably, the *Firmicutes* Rad52/22 cluster is shifted away from the consensus sequence for this Pfam family (**Figure 1**), suggesting sequence divergence of this subgroup from proteins found in other, non-Gram positive bacteria.

We found that the genetic context of the phage subset of *rad52/22* genes is similar to the organization of the *erf* recombination modules (**Supplementary Data Sheet 1**), which are adjacent to *ssb* gene (*rad52/22-ssb* module), followed directly downstream by *rep* functions. The third distinct SSAP clan

(**Figure 1**) identified by our CLANS analysis regroups proteins related to Sak3 SSAP (*gporf35*) from the virulent lactococcal phage p2 (Bouchard and Moineau, 2004; Scaltriti et al., 2011). This protein was classified earlier as a member of the Rad52/22 SSAP family on the basis of the prediction of their structural folds (Lopes et al., 2010). However, we have established that Sak3-like proteins (formally a DUF1071 family of proteins of unknown function) form an individual distinct SSAP clan. Majority of them constitute the bacterial SSAP subset and are encoded in large part in *Bacillales* (mainly *Staphylococcus* spp.) genomes (**Supplementary Data Sheet 1** and **Figure 1**). In turn, 90% of the DUF1071 sequences representing the phage subset are associated with the genomes of small isometric-headed, lytic lactococcal phages. In general, proteins from the Sak3/DUF1071 phage subset seem to be characteristic for lytic *Firmicutes* phages and, besides them, were found in phages of high G+C Gram-positive bacteria (*Mycobacteria* – 88 phages and *Actinobacteria* – 1 phage) and in few vibrio- and cyanophages infecting Gram-negative hosts (**Supplementary Data Sheet 1**). The less frequent occurrence (~10-fold) of DUF1071-encoding genes in Gram-negative bacteria and phages might indicate a recent expansion of these proteins uprooting from Gram-positive organisms.

We found that like *erf* and *rad52/22*, *sak3/duf1071* sequences are often associated with the same genes. The *duf1071-ssb* organization is prevalent, except for lactobacilli phages which possess the *duf1071-phg_2220* module (**Supplementary Data Sheet 1**). In the genomes of virulent lactococcal phages, *duf1071* is associated with a gene encoding a particular type of SSB protein (Szczepankowska et al., 2011). In staphylococcal phage genomes, the *duf1071-ssb* modules are typically followed by genes encoding proteins with the HNHc_6 domain, followed by *rep* genes; thus, forming a *duf1071-ssb-hnhc_6-rep* module. We observed a similar organization in some staphylococcal phages encoding the Erf protein. Yet, in contrast to staphylococcal *erf* modules, *duf1071* is always preceded by *duf2483*. In turn, in lactococcal phages, the *duf1071-ssb* module is most often neighbored upstream by genes encoding homing nucleases (NUMOD1/HNH_3/NUMOD4). Interestingly, the gene context typical for Gram-positive bacteria is not conserved in Gram-negative species where the *duf1071* gene was detected. These genes seem not to be located in the vicinity of other prophage genes, but rather within the transposon-like regions (data not shown).

Sak3/DUF1071 proteins exhibit relevant sequence similarity with both Erf and Rad52/22 proteins, which is supported and elegantly visualized by our CLANS analysis. The finding that *sak3/duf1071* genes are often positioned in the same genetic context as *erf* and *rad52/22* genes emphasizes the orthology between these three types of SSAPs and implies coinciding roles of all three proteins in recombination processes. It probably provides an explanation for frequent HGT detected for members of this superfamily.

Sak4 and RecA Families

Genes encoding Sak4-like proteins, initially identified in the genomes of lactococcal phages phi31 (*orf245*) and TPW22 (Bouchard and Moineau, 2004), represent a novel type of SSAPs

TABLE 2 | CRISPR spacers that match to the *Firmicutes* ssap genes.

Spacer	ssap match	Spacer match (nt)	Source organism
Sak4-like			
NC_007432_1_6	<i>Streptococcus</i> phage Javan17 <i>Streptococcus</i> phage Javan31 <i>Streptococcus</i> phage Javan46	30	<i>S. agalactiae</i> A909
NC_007432_1_6	<i>Streptococcus</i> phage Javan47	30	<i>S. agalactiae</i> A909
NC_008022_1_1	<i>Streptococcus</i> phage 6180.1 <i>Streptococcus</i> phage Javan474 <i>Streptococcus</i> phage Javan484 <i>Streptococcus</i> phage Javan516	30	<i>S. pyogenes</i> MGAS10270
NC_008022_1_1	<i>Streptococcus</i> phage Javan7 <i>Streptococcus</i> phage LF4	30	<i>S. pyogenes</i> MGAS10270
NC_008022_1_1	<i>Streptococcus</i> phage JX01	30	<i>S. pyogenes</i> MGAS10270
NC_008022_1_1	<i>Streptococcus</i> phage LF1	30	<i>S. pyogenes</i> MGAS10270
NC_017568_1_15	<i>Staphylococcus</i> phage phiSP15-1	28	<i>S. pseudintermedius</i> ED99
NZ_CP007573_2_32	<i>Streptococcus</i> phage Javan59	30	<i>S. anginosus</i> GCF_000831165
NZ_CP007573_2_32	<i>Streptococcus</i> phage Javan68	30	<i>S. anginosus</i> GCF_000831165
NZ_CP010874_2_3	<i>Streptococcus</i> phage Javan17	30	<i>S. agalactiae</i> GCF_001683515
NZ_CP010875_2_2	<i>Streptococcus</i> phage Javan31		
NZ_CP012419_2_2	<i>Streptococcus</i> phage Javan46		
NZ_CP016391_2_6			
NZ_CP010874_2_3	<i>Streptococcus</i> phage Javan253	20	<i>S. agalactiae</i> GCF_001275545
NZ_CP010875_2_2			
NZ_CP012419_2_2			
NZ_CP016391_2_6			
NZ_CP010874_2_3	<i>Streptococcus</i> phage Javan460	30	<i>S. agalactiae</i> GCF_001592425
NZ_CP010875_2_2	<i>Streptococcus</i> phage Javan488		
NZ_CP012419_2_2	<i>Streptococcus</i> phage Javan506		
NZ_CP016391_2_6	<i>Streptococcus</i> phage 370.1		
NZ_CP010874_2_3	<i>Streptococcus</i> phage Javan47	30	<i>S. agalactiae</i> GCF_001683515
NZ_CP010875_2_2			
NZ_CP012419_2_2			
NZ_CP016391_2_6			
NZ_CP019794_2_3	<i>Paenibacillus</i> phage LincolnB	34	<i>P. larvae</i> subsp. <i>pulvificiens</i> GCF_002007765
NZ_CP020327_9_3			
NZ_CP019794_2_3	<i>Paenibacillus</i> phage Wanderer	34	<i>P. larvae</i> subsp. <i>pulvificiens</i> GCF_002007765
NZ_CP020327_9_3			
Erf-like			
NC_011134_4_10	<i>Streptococcus</i> phage Javan179	30	<i>S. equi</i> subsp. <i>zooepidemicus</i> MGCS10565
NC_011134_4_13			
NC_017174_12_24	<i>Clostridium</i> phage phiCDHM19	37	<i>C. difficile</i> M120
NC_017576_2_18	<i>Streptococcus</i> phage Javan226	30	<i>S. gallolyticus</i> subsp. <i>gallolyticus</i> ATCC 43143
NZ_CP019687_3_4	<i>Paenibacillus</i> phage Harrison <i>Paenibacillus</i> phage Paisley	36	<i>P. larvae</i> subsp. <i>larvae</i> GCF_002003265
Gp2.5-like			
NC_014377_4_7	<i>Clostridium</i> phage phiCT453A	22	<i>T. oceanii</i> DSM 16646
NC_014377_4_7	<i>Erysipelothrix</i> phage phi1605	22	<i>T. oceanii</i> DSM 16646
RecT/Redβ-like			
NC_004368_1_9	<i>Streptococcus</i> phage Javan384	26	<i>S. agalactiae</i> NEM316
NC_012891_2_9	<i>Streptococcus</i> phage 370.2 <i>Streptococcus</i> phage Javan459 <i>Streptococcus</i> phage Javan489 <i>Streptococcus</i> phage Javan525	25	<i>S. dysgalactiae</i> subsp. <i>equisimilis</i> GGS_124
NC_012891_2_9	<i>Streptococcus</i> phage Javan451	25	<i>S. dysgalactiae</i> subsp. <i>equisimilis</i> GGS_124
NC_014206_6_10	<i>Bacillus</i> phage vB_BspS_SplendidRed	20	<i>Geobacillus</i> sp. C56-T3
NC_017174_6_5	<i>Clostridium</i> phage CDKM9	37	<i>C. difficile</i> M120

(Continued)

TABLE 2 | Continued

Spacer	ssap match	Spacer match (nt)	Source organism
NC_018712_3_12	<i>Streptococcus</i> phage Javan141	36	<i>S. dysgalactiae</i> subsp. <i>equisimilis</i> RE378
NC_021900_2_6	<i>Streptococcus</i> phage Javan202	21	<i>S. lutetiensis</i> 033
NC_021900_2_6	<i>Streptococcus</i> phage Javan215	25	<i>S. lutetiensis</i> 033
NC_021900_2_6	<i>Streptococcus</i> phage Javan263	20	<i>S. lutetiensis</i> 033
NZ_CP015408_1_1	<i>Lactobacillus</i> phage LR1	23	<i>Lb. reuteri</i> GCF_001688685
NZ_CP017267_2_7	<i>Brochothrix</i> phage NF5	26	<i>V. teuberi</i> GCF_001870205
NZ_CP017267_2_7	<i>Streptococcus</i> phage Javan384	23	<i>V. teuberi</i> GCF_001870205
RecA			
NZ_CP009687_2_45	<i>Staphylococcus</i> phage phiSA_BS1	22	<i>C. acetivum</i> GCF_001042715
NZ_CP009687_2_45	<i>Staphylococcus</i> phage phiSA_BS2	22	<i>C. acetivum</i> GCF_001042715

Spacers are named by “genome. locus. spacer-number” (counted from the leader end). SSAP family and origin (ssap match) name are given along with the length of the spacer match and source organism. Only spacers derived from bacteria that match ssap genes with 100% identity are listed.

not related to Rad52/22-Erf-SAK3-like superfamily. Instead, Sak4 proteins show much higher similarity with phage-encoded RecA recombinases and retain signatures characteristic for the RecA superfamily, e.g., catalytic glutamate in the P-loop domain and an arginine-finger from a C-terminal β -hairpin (Leipe et al., 2000). However, they were recognized as a distinct group from the classical RecA proteins based on the presence of a minimal version of the ATPase domain (de Souza et al., 2010). In our database, Sak4 family pools more than 330 members encoded both by lytic and temperate Firmicutes-infecting phages and around 6,700 by Firmicutes bacteria, mainly Bacillales (*Staphylococcus* spp.) and Lactobacillales (*Lactobacillus* spp.) (Figure 1 and Supplementary Data Sheet 1). We found that the phage *sak4* genes are most commonly localized upstream of genes encoding SSB proteins, including DUF669—a distant homolog of SSB of phages N4, Lc-Nu, and PhiAT3 (Hutinet et al., 2018) and DNA helicases (resIII/helicase_C domain). Similar to *erf*, the upstream gene neighbors of *sak4* include: phage Mu *gam*-like genes, *S. thermophilus* phage ϕ Sfi21 *orf157*-like genes and *duf2483* or *duf1351*. The *duf2483-sak4* module seems to be restricted to *Staphylococcus* phages, while *duf1351-sak4* to *Streptococcus* phages. In several cases, *sak4* of streptococcal phages is accompanied by the *sbcC* exonuclease gene.

As for the canonical RecA, we found that the phage proteins are encoded in great majority by large virulent broad-host range Herelleviridae phages specific for Bacillales (mainly *Staphylococcus* spp., but also *Bacillus* spp. and *Listeria* spp.), recognized as SPO1-like viruses (Klump et al., 2010; El-Arabi et al., 2013; Barylski et al., 2014). These phages retain common morphological properties, their genomes consist of a terminally redundant, non-permuted dsDNA molecule of 127–200 kb in size and share considerable amino acid homology. Yet, it seems that the RecA recombinase is not a primary feature of large SPO1-related phages as some other large virulent phages infecting *Bacillus* sp. and *Staphylococcus* sp. deposited in GenBank do not encode for this protein. We determined that the genetic context of phage *recA* genes is characterized by the proximity of a putative *ssb* and a gene encoding phage specific RNA polymerase sigma factors. Other genes that can be found in close vicinity of *recA* vary and include *orfs* with mostly

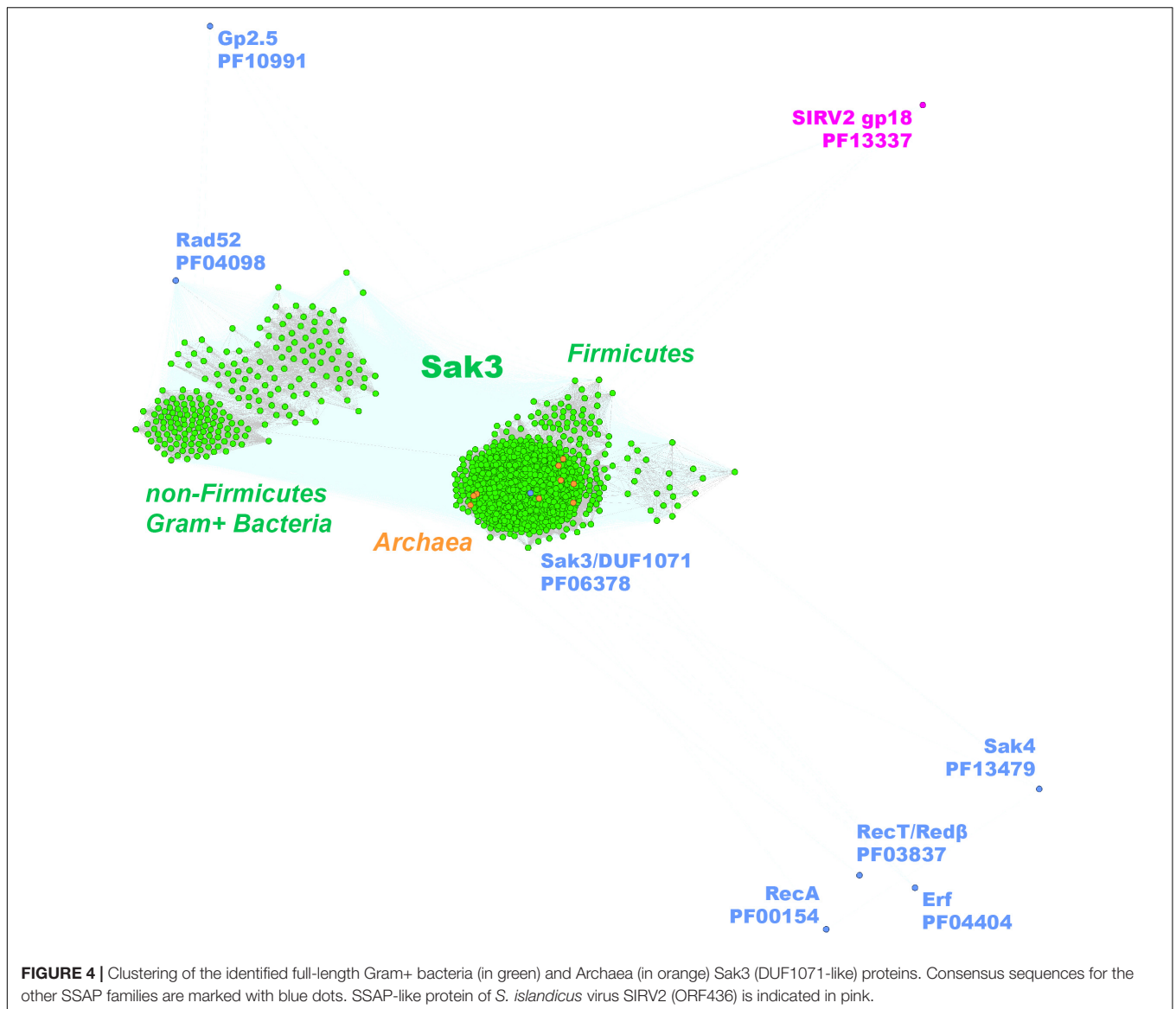
unknown functions, but also phage DNA *polA* and ligase genes (Supplementary Data Sheet 1).

Gp2.5 Family

Based on our analyses, we determined that protein members homologous to phage T7 Gp2.5 protein form a very compact clan distinct from other SSAPs (Figure 1). Within Firmicutes bacteria, Gp2.5-like SSAPs are mostly present in Bacillales (mainly *Staphylococcus* spp.), Clostridiales (mainly *Clostridium* spp.) and, to a lesser extent, in Lactobacillales (mainly *Streptococcus* spp.). In the phage subset, we detected the highest number of Gp2.5 homologs in the genomes of streptococcal and staphylococcal phages (Supplementary Data Sheet 1). In phages, *gp2.5* is often located in proximity to DNA *polA* and *duf2800* (PD-(D/E)-XK nuclease-like gene) (Steczkiewicz et al., 2012), altogether forming the *duf2800-gp2.5-polA* module. Interestingly, an identical gene context was also detected in the majority of phages of Gram-negative bacteria (data not shown). The distinctive feature of the *gp2.5* genetic neighborhood is the absence of *ssb* genes. High level of sequence similarity, restricted phyletic distribution and conserved genetic context, altogether indicate that the Gp2.5 family has emerged recently or displays low/stable evolution rates.

Diversity of Phage SSAPs and Bacterial Phage Defense Mechanisms

The present study demonstrates that Firmicutes bacteria and their phages possess diverse SSAPs which are distributed irregularly across different phyla suggesting frequent horizontal transfer of the corresponding genes. The high diversity of SSAPs is in stark contrast to the ubiquitous and highly conserved RecA recombinase which is one of the most slowly evolving proteins (Eisen, 1995) and shows a low rate of horizontal transfer (Rocha et al., 2005). High rates of molecular evolution have been found for genes, which are constantly subjected to strong selective pressure, especially genes associated with infection and resistance to infection (Buckling and Rainey, 2002; Labrie and Moineau, 2007; Paterson et al., 2010). Thus, one can suggest that phage *ssap* genes, which are crucial for phage multiplication, represent a target for commonly distributed bacterial resistance mechanisms.

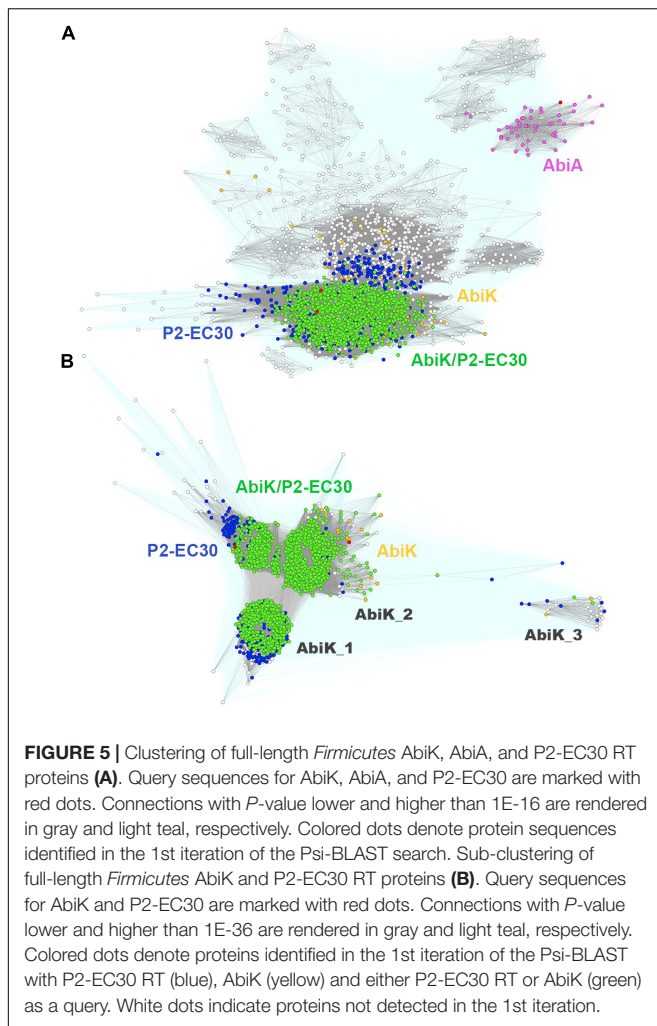


Consequently, this could stimulate the generation of the high variety and sequence diversity among *ssap* genes.

In order to counterfeit phage infection, bacteria developed multiple defense strategies, including CRISPR/Cas and Abortive infection (Abi) systems. CRISPR/Cas systems efficiently prevent phage multiplication by using spacer sequences homologous to regions in the phage genomes. A specific spacer guides the Cas nuclease to cleave the complementary phage nucleic acid. The CRISPR spacers are apt to change in time so that the system loses its specificity toward a particular phage sequence (Sorek et al., 2008). Thus, one could not expect to detect a significant number of spacers identical to the *ssap* sequences if they are not regularly re-acquired.

To investigate whether *ssap* genes could be targeted by CRISPR/Cas, we compared 1,388 sequences from the *Firmicutes* phage SSAP dataset to spacer sequences from the CRISPRdb database (Grissa et al., 2007). Fifty-one *ssaps* sequences from the

phage Gp2.5, Sak4, Erf, RecT/Red β , and RecA SSAP subgroups matched sequences of spacers from Bacteria and Archaea (**Supplementary Data Sheet 3**). Most of the spacers derived from *Firmicutes* bacteria and matched sequences from phages that infect them (**Table 2**). Remarkably, no spacers from *Firmicutes* bacteria matching the *duf1071*-like or *rad52/22* *ssap* genes could be detected. This was quite surprising, especially for *duf1071*, which is a prevalent *ssap* type among lytic phages of *Lactococcus lactis* species. CRISPR spacers may degenerate, but still can be regarded as a sign of past infections (Andersson and Banfield, 2008). Thus, we also searched for sequence “matches” exhibiting lower identity (up to two mismatches). After loosening the search criteria, we established that *duf1071* genes from lactococcal phages were most frequently detected as matches to spacer sequences, whereas the spacers themselves derived from archaeal genomes (**Supplementary Data Sheet 3**). This finding opens the discussion of the possible evolutionary link between lactococcal



phages and Archaea. Our previous study provided evidence that SSB proteins encoded by lytic lactococcal phages might have archaeal origin (Szczebankowska et al., 2011). Genes encoding SSAP and SSB proteins define the recombination modules in many phage genomes (Szczebankowska et al., 2007; Szczebankowska et al., 2011; Neamah et al., 2017). Also, the only functionally characterized SSAP, ORF436 (gp18), of an archaeal phage *Sulfolobus islandicus* SIRV2 is encoded directly downstream of an *ssb* gene (Guo et al., 2015). Despite the presence of different domains within SIRV2_ORF436 (Lon_2 protease-like domain, PF13337) and DUF1071-like proteins (dsRBD-like domain, PF06378), the conserved position of *ssap-ssb* gene module in both, Bacteria and Archaea suggests similar function. Our idea on the possible archaeal origin of now deteriorated, matching spacer sequences necessitates further studies but seems to be reinforced by identification of *duf1071*-like genes in several archaeal genomes (Supplementary Data Sheet 1), which all cluster tightly with the *Firmicutes* Sak3/DUF1071 family (Figure 4).

Overall, our analysis shows that spacer sequences matching *ssap* genes represent a non-negligible quantity, especially taking into account that a bulk of almost 9.000 CRISPR spacers from

the CRISPRdb database does not match any known sequences (Grissa et al., 2007; Shmakov et al., 2017). In accordance with previous communications, this may reflect (repeated) acquisition of specific spacers based on their effectiveness (Paez-Espino et al., 2013).

Given that only one CRISPR/Cas system has been identified in *Lactococcus lactis* (Millen et al., 2012), we asked whether other phage resistance mechanisms could specifically target the *ssap* genes of *Firmicutes* bacteria. Several SSAPs encoded by lactococcal phages, hereby classified as members of Erf, Rad52/22, Sak3/DUF1071, and Sak4 protein families, were already experimentally confirmed as targets of the phage abortive infection mechanism, AbiK (Emond et al., 1997; Boucher et al., 2000, 2001; Bouchard and Moineau, 2004; Fortier et al., 2005; Scaltriti et al., 2010, 2011). One of them, the Sak4 protein encoded by phage phi31, has also been recognized as a target for another abortive infection mechanism—AbiA (Garvey et al., 1995; Dinsmore and Klaenhammer, 1997).

Both Abi mechanisms are active against lactococcal phages from three most abundant and genetically unrelated groups (936-like small isometric-headed phages, c2-like prolate-headed phages, and mostly temperate P335-like phages) (Deveau et al., 2006; Mahony et al., 2013). The AbiA system has also been demonstrated to block multiplication of six tested *Streptococcus thermophilus* phages (Tangney and Fitzgerald, 2002), of which one (phage 7202) was annotated to possess the *erf* gene (NC_002185). Up to now, only four proteins highly homologous to the *L. lactis* AbiK have been annotated as putative abortive infection agents (Wang et al., 2011). However, both lactococcal *abiK* and *abiA* are plasmid-encoded genes which might facilitate their wide distribution among related Gram-positive bacteria, such as the aforementioned *S. thermophilus* species. Here, we identified *abiK* and *abiA* homologous genes on over forty and five plasmids, respectively, found in several different *Firmicutes* species (Supplementary Data Sheet 4). Yet, the majority of *abiK* and *abiA* genes are not plasmid-encoded but are rather located on chromosomes or come from WGS data (Supplementary Data Sheet 4). Phage defense systems (e.g., restriction-modification, toxin-antitoxin modules, and CRISPR elements) tend to be clustered into genomic islands (Makarova et al., 2011) known to spread *via* horizontal transfer events (Juhás et al., 2009). By gene context analyses we revealed that *abiK*- and *abiA*-like genes are associated with diverse genes involved in phage resistance, which is suggestive for their HGT-mediated acquisition (Supplementary Data Sheet 4).

Both, AbiK and AbiA exhibit high sequence similarity (23% identity and 44% of similarity) and contain the RVT_1 reverse transcriptase domain (PF00078), which has been shown to interfere with phage DNA replication (Boucher et al., 2000; Wang et al., 2011). Interestingly, the classical *L. lactis* AbiK shares homology (32% sequence identity) with yet another reverse transcriptase encoded by the *E. coli* P2-EC30 prophage (Odegrip et al., 2006). Both protein subfamilies overlap significantly, but remain clearly isolated from AbiA (Figure 5A). More detailed sequence clustering allowed to separate AbiK from P2-EC30 and reveal the presence of several AbiK subclusters consisting of proteins with probably slightly different anti-phage

specificities (**Figure 5B**). Each of these AbiK subgroups is shown to have different phylogenetic distribution (**Supplementary Data Sheet 4**). The canonical AbiK proteins are present mainly in staphylococci. The biggest newly identified clusters of AbiK-like proteins include AbiK_1, mainly from *Bacillales*, *Clostridia*, and *Lactobacillales*, AbiK_3 proteins most of which derive from *Lactobacillus* species and AbiK_2 proteins (mainly *Bacillales* and *Clostridia*) closely related to the canonical AbiK-like group. Interestingly, despite the initial identification of *abiK* and *abiA* genes as phage resistance mechanisms active in *L. lactis*, we found that AbiA, AbiK or P2-EC30-like systems are under-represented in multiple *Lactococcus* sp. hosts. This observation may support the earlier notion that DUF1071-encoding genes, prevalent in lytic lactococcal phages, have emerged recently within *Firmicutes* and that Abi phage resistance mechanisms for these SSAPs are less frequently encountered in their bacterial hosts.

The wide distribution of the Abi systems and SSAPs across *Firmicutes* phylum implies efficient HGT of both. We suggest that the spread of AbiK- and AbiA-like protein clusters among diverse *Firmicutes* bacteria is a factor affecting the diversity and irregular dissemination of their potential phage targets, namely the members of Erf, Rad52/22, Sak3/DUF1071 and Sak4 SSAP families. Such notion is in agreement with the well-documented ability of the different phage resistance mechanisms to contribute to the evolution of lactococcal phages (Bouchard and Moineau, 2000, 2004; Durmaz and Klaenhammer, 2000; Odegrip et al., 2006; Labrie and Moineau, 2007).

We speculate that, in general, the diverging evolution of phage recombinases could be a consequence of phage-bacteria arms race, linked to the evolution of phage resistance mechanisms targeting the *ssap* genes. Certainly, experiments are required to verify this hypothesis. Phages could evade Abi- and/or CRISPR-mediated defense systems through single mutations within the target gene or *via* HGT, which may provide an alternative way to escape phage defense mechanisms directed against such genes.

CONCLUSION

We significantly broadened and manually inspected the dataset of SSAP proteins encoded by *Firmicutes* and their phages. By applying detailed sequence clustering, we have defined numerous known and newly identified SSAPs that have not been annotated as such.

Based on thorough gene context analyses, we established signature functions that accompany the phage-encoded *ssap* genes. This resulted in a high quality dataset that can be further used for rational selection of individual *ssap* genes or recombination modules for application in recombineering techniques.

Results of our analyses show that phyletic distribution of genes encoding specific SSAP types is not phage- or bacteria-related. We suggest that the high genetic variability and irregular

dissemination of phage SSAP-encoding genes occurs *via* HGT as a consequence of the co-evolutionary arms race between bacteria and phages. Bacterial-phage defense mechanisms, either highly specialized (Abi systems) or adaptive (CRISPR/Cas systems) that target SSAP functions can make a significant contribution to the diversity and evolution of the corresponding (*ssap*) genes. On the other hand, the presence of defense mechanisms directed against *ssaps* may interfere with viability and/or stability of strains carrying phage recombination genes and should be taken in consideration when choosing the bacteria-*ssap* pair used for recombineering.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

KS collected SSAP sequences, performed bioinformatics analyses, prepared the figures, and wrote sections of the manuscript. EP designed the experiments and performed initial database analyses. AS manually analyzed all data, performed Abi and CRISPR spacer analyses, prepared the tables, and the main text of the manuscript. EB was responsible for the experimental concept and drafted the main text of the manuscript. KS retrieved all data from public databases and prepared **Supplementary Figure 1**. AS prepared the excel spread sheets and performed manual inspection of all data contained in the files. All authors have read and approved the manuscript.

FUNDING

This work was in part supported by the joint research project between Institut National de la Recherche Agronomique (INRA), France and the Ministry of Science and Higher Education (MSHE), Poland (number P260/N/INRA/2008/0).

ACKNOWLEDGMENTS

We would like to thank Anna Muszewska and Marcin Grynberg for their help in the initial steps of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.644622/full#supplementary-material>

REFERENCES

- Akroyd, J., and Symonds, N. (1986). Localization of the gam gene of bacteriophage Mu and characterisation of the gene product. *Gene* 49, 273–282.
- Anderson, R. E., Brazelton, W. J., and Baross, J. A. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133. doi: 10.1111/j.1574-6941.2011.01090.x
- Andersson, A. F., and Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050. doi: 10.1126/science.1157358
- Barylski, J., Nowicki, G., and Goździcka-Józefiak, A. (2014). The discovery of phiAGATE, a novel phage infecting *Bacillus pumilus*, leads to new insights into the phylogeny of the subfamily Spounavirinae. *PLoS One* 9:e86632. doi: 10.1371/journal.pone.0086632
- Bouchard, J. D., and Moineau, S. (2000). Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. *Virology* 270, 65–75. doi: 10.1006/viro.2000.0226
- 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
- Boucher, I., Emond, E., Dion, E., Montpetit, D., and Moineau, S. (2000). Microbiological and molecular impacts of AbiK on the lytic cycle of *Lactococcus lactis* phages of the 936 and P335 species. *Microbiology* 146, 445–453. doi: 10.1099/00221287-146-2-445
- Boucher, I., Emond, E., Parrot, M., and Moineau, S. (2001). DNA sequence analysis of three *Lactococcus lactis* plasmids encoding phage resistance mechanisms. *J. Dairy Sci.* 84, 1610–1620. doi: 10.3168/jds.S0022-0302(01)74595-X
- Broniewski, J. M., Meaden, S., Paterson, S., Buckling, A., and Westra, E. R. (2020). The effect of phage genetic diversity on bacterial resistance evolution. *ISME J.* 14, 828–836. doi: 10.1038/s41396-019-0577-7
- Buckling, A., and Rainey, P. B. (2002). Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. Biol. Sci.* 269, 931–936. doi: 10.1098/rspb.2001.1945
- Canchaya, C., Founous, G., Chibani-Chennoufi, S., Dillmann, M. L., and Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* 6, 417–424. doi: 10.1016/s1369-5274(03)00086-9
- Chang, H. W., and Julin, D. A. (2001). Structure and function of the *Escherichia coli* RecE protein, a member of the RecB nuclease domain family. *J. Biol. Chem.* 276, 46004–46010. doi: 10.1074/jbc.M108627200
- Chopin, M.-C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Curr. Opin. Microbiol.* 8, 473–479. doi: 10.1016/j.mib.2005.06.006
- Court, D. L., Sawitzke, J. A., and Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annu. Rev. Genet.* 36, 361–388. doi: 10.1146/annurev.genet.36.061102.093104
- Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., and Hatfull, G. H. (2011). Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12:395. doi: 10.1186/1471-2105-12-395
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PRC products. *Proc. Natl. Acad. Sci. U S A.* 97, 6640–6645. doi: 10.1073/pnas.120163297
- 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. (2014a). 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 Paepe, M., Leclerc, M., Tinsley, C. R., and Petit, M. A. (2014b). Bacteriophages: an underestimated role in human and animal health? *Front. Cell. Infect. Microbiol.* 4:39. doi: 10.3389/fcimb.2014.00039
- De Sordi, L., Lourenço, M., and Debarbieux, L. (2019). The battle within: interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host Microbe.* 25, 210–218. doi: 10.1016/j.chom.2019.01.018
- de Souza, R. F., Iyer, L. M., and Aravind, L. (2010). Diversity and evolution of chromatin proteins encoded by DNA viruses. *Biochim. Biophys. Acta* 1799, 302–318. doi: 10.1016/j.bbagr.2009.10.006
- Derbise, A., Lesic, B., Dacheux, D., Ghigo, J. M., and Carniel, E. (2003). A rapid and simple method for inactivating chromosomal genes in *Yersinia*. *FEMS Immunol. Med. Microbiol.* 38, 113–116. doi: 10.1016/S0928-8244(03)00181-0
- Deveau, H., Labrie, S. J., Chopin, M.-C., and Moineau, S. (2006). Biodiversity and classification of lactococcal phages. *Appl. Environ. Microbiol.* 72, 4338–4346. doi: 10.1128/AEM.02517-05
- Dinsmore, P. K., and Klaenhammer, T. R. (1997). Molecular characterization of a genomic region in a lactococcal bacteriophage that is involved in its sensitivity to the phage defense mechanism AbiA. *J. Bacteriol.* 179, 2949–2957. doi: 10.1128/jb.179.9.2949-2957.1997
- Durmaz, E., and Klaenhammer, T. R. (2000). Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. *Appl. Environ. Microbiol.* 66, 895–903. doi: 10.1128/aem.66.3.895-903.2000
- Eisen, J. A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* 41, 1105–1123. doi: 10.1007/BF00173192
- El-Arabi, T. F., Griffiths, M. W., She, Y. M., Villegas, A., Lingohr, E. J., and Kropinski, A. M. (2013). Genome sequence and analysis of a broad-host range lytic bacteriophage that infects the *Bacillus cereus* group. *Virol. J.* 10:48. doi: 10.1186/1743-422X-10-48
- Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001). High efficiency mutagenesis, repair and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U S A.* 98, 6742–6746. doi: 10.1073/pnas.121164898
- Emond, E., Holler, B. J., Boucher, I., Vandenberg, P. A., Vedamuthu, E. R., Kondo, J. K., et al. (1997). Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from *Lactococcus lactis*. *Appl. Environ. Microbiol.* 63, 1274–1283. doi: 10.1128/AEM.63.4.1274-1283
- Enikeeva, F. N., Severinov, K. V., and Gelfand, M. S. (2010). Restriction-modification systems and bacteriophage invasion: who wins? *J. Theor. Biol.* 266, 550–559. doi: 10.1016/j.jtbi.2010.07.006
- Foley, S., Lucchini, S., Zwahlen, M.-C., and Brüssow, H. (1998). A short noncoding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus*. *Virology* 250, 377–387. doi: 10.1006/viro.1998.9387
- Fortier, L. C., Bouchard, J. D., and Moineau, S. (2005). Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. *J. Bacteriol.* 187, 3721–3730. doi: 10.1128/JB.187.11.3721-3730
- Frickey, T., and Lupas, A. N. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702–3704. doi: 10.1093/bioinformatics/bth444
- Garvey, P., Fitzgerald, G. F., and Hill, C. (1995). Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Appl. Environ. Microbiol.* 61, 4321–4328. doi: 10.1128/AEM.61.12.4321-4328
- Gómez, P., and Buckling, A. (2011). Bacteria-phage antagonistic coevolution in soil. *Science* 332, 106–109. doi: 10.1126/science.1198767
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 23:172. doi: 10.1186/1471-2105-8-172
- Guo, Y., Kragelund, B. B., White, M. F., and Peng, X. (2015). Functional characterization of a conserved archaeal viral operon revealing single-stranded DNA binding, annealing and nuclease activities. *J. Mol. Biol.* 427, 2179–2191. doi: 10.1016/j.jmb.2015.03.013
- Hatfull, G. F., and Hendrix, R. W. (2011). Bacteriophages and their genomes. *Curr. Opin. Virol.* 1, 298–303. doi: 10.1016/j.coviro.2011.06.009
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *PNAS* 96, 2192–2197. doi: 10.1073/pnas.96.5.2192
- Hutinet, G., Besle, A., Son, O., McGovern, S., Guerois, R., Petit, M.-A., et al. (2018). Sak4 of phage HK620 is a RecA remote homolog with single-strand annealing activity stimulated by its cognate SSB protein. *Front. Microbiol.* 9:743. doi: 10.3389/fmicb.2018.00743
- Iyer, L. M., Koonin, E. V., and Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Red β , ERF and RAD52. *BMC Genomics* 3:8. doi: 10.1186/1471-2164-3-8
- Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.x

- Klumpp, J., Lavigne, R., Loessner, M. J., and Ackermann, H. W. (2010). The SPO1-related bacteriophages. *Arch. Virol.* 155, 1547–1561. doi: 10.1007/s00705-010-0783-0
- Koskella, B., and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. doi: 10.1111/1574-6976.12072
- Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2010). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19. doi: 10.1016/j.tim.2009.11.003
- Kuzminov, A. (1999). Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* 63, 751–813.
- Labrie, S., and Moineau, S. (2002). Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. *Virology* 296, 308–320. doi: 10.1006/viro.2002.1401
- Labrie, S. J., and Moineau, S. (2007). Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *J. Bacteriol.* 189, 1482–1487.
- Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327. doi: 10.1038/nrmicro2315
- Leipe, D. D., Aravind, L., Grishin, N. V., and Koonin, E. V. (2000). The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.* 10, 5–16.
- Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010). ACLAME: a CLAssification of mobile genetic elements, update 2010. *NAR* 38, D57–D61. doi: 10.1093/nar/gkp938
- Lesic, B., and Rahme, L. G. (2008). Use of the lambda Red recombinase system to rapidly generate mutants in *Pseudomonas aeruginosa*. *BMC Mol. Biol.* 9:20. doi: 10.1186/1471-2199-9-20
- 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. *NAR* 38, 3952–3962. doi: 10.1093/nar/gkq096
- Mahony, J., Martel, B., Tremblay, D. M., Neve, H., Heller, K. J., Moineau, S., et al. (2013). Identification of a new P335 subgroup through molecular analysis of lactococcal phages Q33 and BM13. *Appl. Environ. Microbiol.* 79, 4401–4409. doi: 10.1128/AEM.00832-13
- Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/JB.05535-11
- 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. doi: 10.1371/journal.pgen.1000065
- Millen, A. M., Horvath, P., Boyaval, P., and Romero, D. A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* 7:e51663. doi: 10.1371/journal.pone.0051663
- Muniesa, M., García, A., Miró, E., Mirelis, B., Prats, G., Jofre, J., et al. (2004). Bacteriophages and diffusion of beta-lactamase genes. *Emerg. Infect. Dis.* 10, 1134–1137. doi: 10.3201/eid1006.030472
- Murphy, K. C., Campellone, K. G., and Poteete, A. R. (2000). PCR-mediated gene replacement in *Escherichia coli*. *Gene* 246, 321–330. doi: 10.1016/s0378-1119(00)00071-8
- Muyrers, J. P., Zhang, Y., Buchholz, F., and Stewart, A. F. (2000). RecE/RecT and RedA/RedB initiate double-stranded break repair by specifically interacting with their respective partners. *Gen. Dev.* 14, 1971–1982. doi: 10.1101/gad.14.15.1971
- Neamah, M. M., Mir-Sanchis, I., López-Sanz, M., Acosta, S., Baquedano, I., Haag, A. F., et al. (2017). Sak and Sak4 recombinases are required for bacteriophage replication in *Staphylococcus aureus*. *NAR* 45, 6507–6519. doi: 10.1093/nar/gkx308
- Noirot, P., Gupta, R. C., Radding, C. M., and Kolodner, R. D. (2003). Hallmarks of homology recognition by RecA-like recombinases are exhibited by the unrelated *Escherichia coli* RecT protein. *EMBO J.* 22, 324–334. doi: 10.1093/emboj/cdg027
- Odegrip, R., Nilsson, A. S., and Haggård-Ljungquist, E. (2006). Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. *J. Bacteriol.* 188, 1643–1647. doi: 10.1128/JB.188.4.1643-1647
- Paez-Espino, D., Morovic, W., Sun, C., Thomas, B. C., Ueda, K., Stahl, B., et al. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat. Commun.* 4:1430. doi: 10.1038/ncomms2440
- Pal, C., Maciá, M. D., Oliver, A., Schachar, I., and Buckling, A. (2007). Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450, 1079–1081. doi: 10.1038/nature06350
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., et al. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature* 464, 275–278. doi: 10.1038/nature08798
- Ranallo, R. T., Barnoy, S., Thakkar, S., Urlick, T., and Venkatesan, M. M. (2006). Developing live *Shigella* vaccines using lambda Red recombineering. *FEMS Immunol. Med. Microbiol.* 47, 462–469. doi: 10.1111/j.1574-695X.2006.00118.x
- Rocha, E. P., Cornet, E., and Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genetics*. 1:e15. doi: 10.1371/journal.pgen.0010015
- Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M., and Atkinson, G. C. (2020). Predicting functional associations using flanking genes (FlaGs). *BioRxiv [preprint]* doi: 10.1093/bioinformatics/btaa788
- Sawitzke, J. A., Thomason, L. C., Costantino, N., Bubunenko, M., Datta, S., and Court, D. L. (2007). Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol.* 421, 171–199. doi: 10.1016/S0076-6879(06)21015-2
- Scaltriti, E., Launay, H., Genois, M. M., Bron, P., Rivetti, C., Grolli, S., et al. (2011). Lactococcal phage p2 ORF35-Sak3 is an ATPase involved in DNA recombination and AbiK mechanism. *Mol. Microbiol.* 80, 102–116. doi: 10.1111/j.1365-2958.2011.07561.x
- Scaltriti, E., Moineau, S., Launay, H., Masson, J. Y., Rivetti, C., Ramoni, R., et al. (2010). Deciphering the function of lactococcal phage ul36 Sak domains. *J. Struct. Biol.* 170, 462–469. doi: 10.1016/j.jsb.2009.12.021
- Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* 8, e01397-17. doi: 10.1128/mBio.01397-17
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793
- Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K. E. (2008). Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res. Microbiol.* 159, 349–357. doi: 10.1016/j.resmic.2008.04.010
- Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., and Ginalski, K. (2012). Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *NAR* 40, 7016–7045. doi: 10.1093/nar/gks382
- Szczepankowska, A. K., Prestel, E., Mariadassou, M., Bardowski, J., and Bidnenko, E. (2011). Phylogenetic and complementation analysis of a single-stranded DNA binding protein family from lactococcal phages indicates a non-bacterial origin. *PLoS One* 6:e26942. doi: 10.1371/journal.pone.0026942
- Szczepańska, A. K. (2009). Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit. Rev. Microbiol.* 35, 197–220.
- Szczepańska, A. K., Bidnenko, E., Plochocka, D., McGovern, S., Ehrlich, S. D., Bardowski, J., et al. (2007). A distinct single-stranded DNA-binding protein encoded by the *Lactococcus lactis* bacteriophage bIL67. *Virology* 363, 104–112. doi: 10.1016/j.viro.2007.01.023
- Tangney, M., and Fitzgerald, G. F. (2002). AbiA, a lactococcal abortive infection mechanism functioning in *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* 68, 6388–6391. doi: 10.1128/aem.68.12.6388-6391
- van Kessel, J. C., Marinelli, L. J., and Hatfull, G. F. (2008). Recombineering mycobacteria and their phages. *Nat. Rev. Microbiol.* 6, 851–857. doi: 10.1038/nrmicro2014
- Wang, C., Villion, M., Semper, C., Coros, C., Moineau, S., and Zimmerly, S. (2011). A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro. *NAR* 39, 7620–7629. doi: 10.1093/nar/gkr397

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Steczkiewicz, Prestel, Bidnenko and Szczepankowska. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.