SBM SOCIEDADE BRASILEIRA DE MICROBIOLOGIA

ENVIRONMENTAL MICROBIOLOGY - RESEARCH PAPER

Check for updates

Identification and characterization of integrated prophages and CRISPR-Cas system in *Bacillus subtilis* RS10 genome

Sajid Iqbal[1,2] · **Farida Begum[3]**

Abstract

Bacteriophages have been extensively investigated due to their prominent role in the virulence and resistance of pathogenic bacteria. However, little attention has been given to the non-pathogenic *Bacillus* phages, and their role in the ecological bacteria genome is overlooked. In the present study, we characterized two *Bacillus* phages with a linear DNA genome of 33.6 kb with 44.83% GC contents and 129.3 kb with 34.70% GC contents. A total of 46 and 175 putative coding DNA sequences (CDS) were identified in prophage 1 (P1) and prophage 2 (P2), respectively, with no tRNA genes. Comparative genome sequence analysis revealed that P1 shares eight CDS with phage Jimmer 2 (NC-041976), and phage Osiris (NC-028969), and six with phage phi CT9441A (NC-029022). On the other hand, P2 showed high similarity with Bacill_SPbeta_NC_001884 and *Bacillus* phage phi 105. Further, genome analysis indicates several horizontal gene transfer events in both phages during the evolution process. In addition, we detected two CRISPR-Cas systems for the first time in *B. subtilis*. The identified CRISPR system consists of 24 and 25 direct repeats and integrase coding genes, while the *cas* gene which encodes Cas protein involved in the cleavage of a target sequence is missing. These findings will expand the current knowledge of soil phages as well as help to develop a new perspective for investigating more ecological phages to understand their role in bacterial communities and diversity.

Keywords *Bacillus subtilis* · Prophages · CRISPR-Cas, Ecological role · Application · Biotechnology

Introduction

Bacteriophages are viruses of bacteria and are an important and integral part of bacterial ecology. They exist in a wide variety of ecosystems including marine, freshwater, soil, and sewage plants [1]. They infect bacteria and use host metabolism to proliferate by means of two different ways. The direct proliferation is called the lytic cycle, while in some cases, the phages integrate their genetic material into the host bacterial genome, deactivate their lytic genes, and reproduce with the host. This process leads to the lysogenic bacterium, and the phages are called temperate phages. Phages are also powerful predators of bacteria in extreme environments and are considered candidates for biotechnological tools and antimicrobial agents [2]. Prophages can offer new capabilities to the host bacterium via the additional genetic material and in exceptional cases make non-pathogenic into pathogenic bacteria. The phages that infect *Bacillus* spp. largely belong to Myoviridae, Siphoviridae, Tectiviridae, and Podoviridae. Siphoviridae is a subfamily of Caudoviridae, characterized as tail bacteriophages. Myoviridae is a family of tail bacteriophages, but unlike *Siphoviridae*, they are known to contract and extent their tail. Both families have an icosahedral head where double-stranded DNA is stored [3].

Bacillus subtilis is a Gram-positive bacteria, rod-shaped, and aerobic bacteria abundantly found in soil. *B. subtilis* is currently the best-known laboratory model for

1 Department of Industrial Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), H-12, Islamabad, Pakistan

2 Oujiang Laboratory (Zhejiang Laboratory for Regenerative Medicine, Vision and Brain Health), Wenzhou 325000, China

3 Department of Biochemistry, Abdul Wali Khan University Mardan (AWKUM), Mardan, Pakistan

Gram-positive bacteria. Its capacity to effectively release proteins into the media, as well as its status as generally regarded as safe, makes it appealing for biotechnological applications [4, 5]. *B. subtilis* is utilized in the commercial manufacture of enzymes, vitamins, and antibiotics, as well as in the food sector for the fermentation of various foods [5]. However, majority of the fermentation industries are struggling with bacteriophage contamination as *B. subtilis* is vulnerable to phage infection [15]. In the current study, we identified a CRISPR array and two bacteriophages for the first time in the *B. subtilis* strain RS10 (CP046860.1). The strain RS10 harbors plant growth–promoting traits and was previously isolated from the rhizosphere [6].

Materials and methods

Strain isolation and prophage identification

Bacillus subtilis strain RS10 (accession number CP046860.1) was isolated from the rhizosphere region. The strain was demonstrated as a plant growth–promoting strain, and several horizontal gene transfer events were witnessed. The strain RS10 genome is highly diverse and identified as a novel sequence type (ST176) [6]. The prophages were identified in the genome of *B. subtilis* RS10 using PHASTER (https://phaster.ca/) and VirSorter [7]. PHASTER searches against a phage database (https://phagesdb.org/) and a prophage database [8]. The phage-like genes are grouped using DBSCAN []. PHASTER was employed due to its ability to determine the completeness of the predicted prophages via the identification of specific indicators such as attachment sites. In contrast, VirSorter does not locate attachment sites, but it performs better than other tools in identifying prophages in fragmented genomic data. The custom application programming interface was utilized to predict prophages from PHASTER web server, while VirSorter prophage identification was conducted locally using a command line interface.

Genome annotation and CRISPR-Cas system identification

The prophage regions were extracted from the *B. subtilis* genome and were annotated using the NCBI domain conserved database [9] and analyzed for tRNA using tRNA scan [10]. The PHASTER-predicted prophages with overlapping regions between the two tools were considered. Multiple sequence alignment with the related phages was made using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). CRISPR-Cas systems were predicted using CRISPRCas-Finder online server (https://crisprcas.i2bc.paris-saclay.fr/), and the key genes that play important role in bacterial

adaptive immunity such as spacer integrase and *cas* genes were manually searched in the RS10 genome.

Phylogenetic analysis

A phylogenetic tree was constructed based on whole genome using MEGA-X [11]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The constructed tree was edited using the iTOL web-based server (https://itol.embl.de/).

Result and discussion

Prophages in *B. subtilis* strain RS10

PHASTER identified five prophage regions in the *B. subtilis* RS10 genome (Fig. 1). Among these, two each were identified as complete and incomplete prophages, while one was identified as questionable prophage. VirSorter identified seven prophages with numeric value 5 (category 2). Herein, we characterized and discussed the two complete prophages P1 and P2 overlapping between the two tool predictions. The genome of P1 is a dsDNA of 33.6 kb with a GC content of 44.83% while P2 carries a genome of 129.3 kb with a GC content of 34.70% (Fig. 1). P1 showed maximum similarity with *Rhizobium* phage vB RleS L338C (NC 023502.1),

Fig. 1 Prophages identified in *Bacillus subtilis* RS10 genome. The green color indicates intact prophages, the red color indicates incomplete prophages, and the blue color represents a questionable prophage

Jimmer 1, and Osiris phages while P1 exhibits high similarity with *Bacillus* phage phi 105 and SPbeta-like prophage. The P1 and relative prophage are defective and, upon induction, package random chromosomal fragments inside phage particles. It acts as a killing factor for non-related strains, similar to a bacteriocin. Phage P2 is a SPbeta-like prophage, a type likewise widespread in *B. subtilis*. SPbeta phage in *B. subtilis* strain 168 is an intact prophage with an interesting developmentally regulated excision from the chromosome during sporulation [12].

Phylogenetic analysis of identified prophages

To infer whether the identified prophages are derived from the same origin and homology, the phylogenetic tree was constructed based on whole-genome sequences of P1 with 14 related phages. These results are also in agreement with the BLAST search and P1 cluster with *rhizobium* phage vB RleS L338C (NC 023502.1) and P2 shared a clade with *Bacillus* phage phi 105 (NC 004167.1) (Fig. 2). Phylogenetic analysis reveals that these phages belong to distinct lineages within the family of bacteriophages, indicating their diverse evolutionary origins. Despite this, there are certain similarities and shared features between P1 and P2, suggesting potential evolutionary connections or gene exchange events. Moreover, we explore the intriguing phenomenon of rhizobium phage transfer into *Bacillus*, highlighting the potential mechanisms and implications of horizontal gene transfer between distantly related bacterial hosts. Previous studies revealed that the infection of these phages can lead to various adaptive changes in the bacterial host, including the acquisition of novel genetic material, alterations in gene expression patterns, and increased resistance to environmental stresses [13]. The bacteriophage phi 105 is a temperate *Bacillus subtilis*–derived phage that integrates into the host genome at a special location that is between the *pheA* and *ilvC* bacterial markers [14]. In contrast, the phage vB RleS L338C is a *rhizobium*-derived phage and was identified in *B. subtilis* RS10 isolated from the rhizosphere region. Therefore, it is hypothesized that this phage transfected *B. subtilis* RS10 in the rhizosphere region where *rhizobium* are ubiquitously found.

Genomic characterization of identified prophages

A total of 46 CDS and no tRNA coding genes were detected in P1 (Supplementary file 1). The genome of P1 carries genes for DNA replication, membrane-associated initiation of head vertex, tail sheath protein and capsid protein, immunity protein D and putative tail spike, and beta-helical glycoside (Fig. 3). The annotated genome of P1 accounted for 80.4% (phage plus hypothetical protein) of the total genome representing a more compact genome than other *Bacillus*
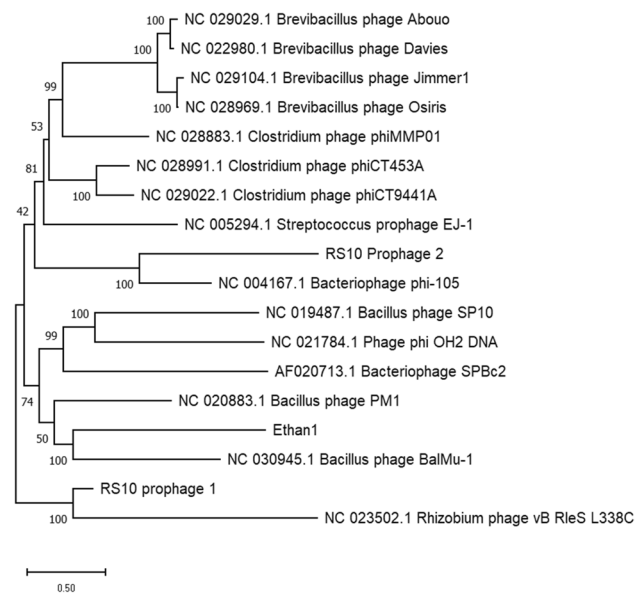

Fig. 2 The evolutionary history was inferred based on whole-genome sequences using MEGA-X. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method [3] and are in the units of the number of base substitutions per site. This analysis involved 18 nucleotide sequences. Codon positions included were $1st + 2nd + 3rd + noncoding$. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 152,061 positions in the final dataset

phages. Few of the coding genes such as the XkdB gene overlap a hypothetical gene that allows a short nucleotide region to encode the maximum amount of information [15].

P2 encodes a total of 177 CDS, and no tRNA genes were identified. The P2 genome harbors genes for DNA polymerase, helicase, phage portal protein, putative tail spike, beta helical glycoside, putative methionine sulfoxide reductase, and several hypothetical proteins with unknown functions (Supplementary file 2). The annotated genome of P2 represented 97.1% of phage and hypothetical protein indicating a compact genome. The genome of P2 encodes small acid-soluble protein C which plays a vital role in resistance to heavy ionizing radiation such as X-rays [16]. Ionized radiations are lethal and mutagenic to all types of living organisms. However, *B. subtilis* is reported to be highly resistant and indicates that prophages contribute to maintain ecological adaptation and evolution. Since these phages act as a vehicle for horizontal gene transfer and encode several adaptability factors that allow the host to survive and adapt to the harsh environment.

Extreme resistance to ionizing radiation is important in medical sterilization, food preservation, and
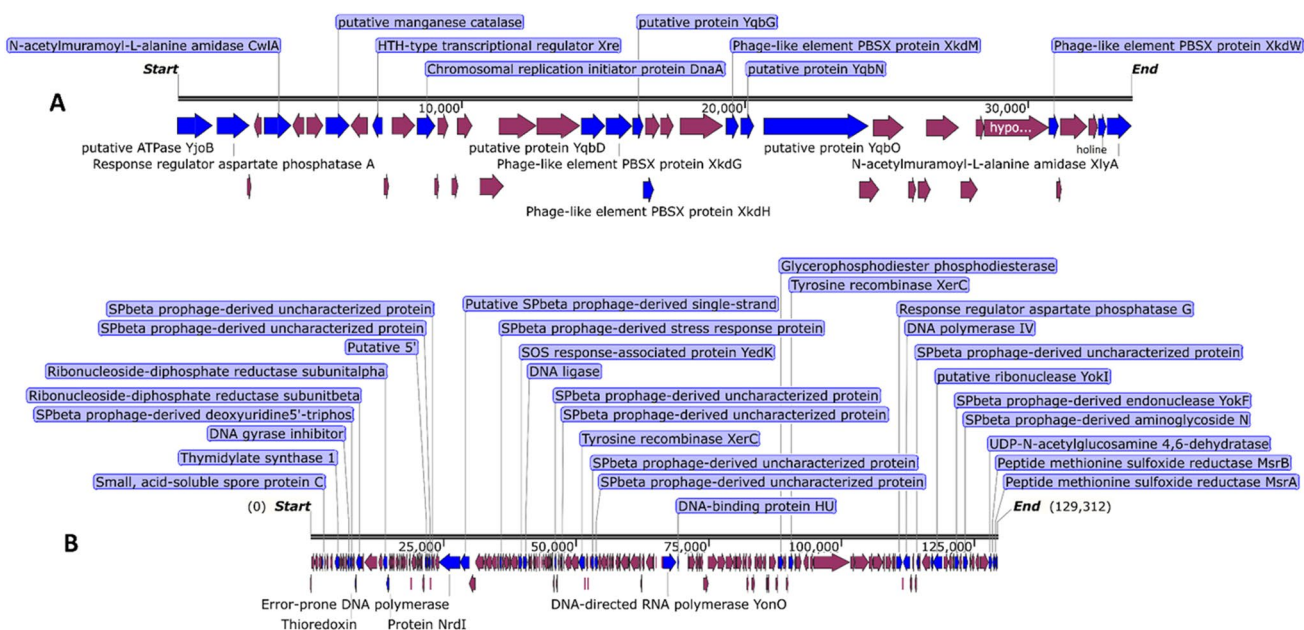
Fig. 3 Genome organization of coding DNA sequence (CDS) in the **A** prophage P1 and **B** prophage P2. The hypothetical protein modules are shown in color and phage-related modules are shown in blue color with corresponding label protein. The double line numbers show the genomic location, and CDS orientations are represented by a horizontal arrow

decontamination from a bioterror attack. Similarly, physical stability is a pre-requirement for the commercial application of phages as biocontrol agents and in vivo immune function measuring [17, 18]. Nevertheless, majority of the phages are sensitive to extreme condition, and their successful application is potentially affected by altering the phage genome structure. Even a single non-synonym mutation in the viral genome leads to altering the phenotype.

Identification of CRISPR-Cas system

The cluster regularly interspaced short palindromic repeats (CRISPR)-Cas (CRISPR-associated *cas*) systems are constituent of defense mechanisms in Bacteria and Archaea, which provide resistance against bacteriophage infection and other invasive mobile genetic elements [19]. It is made up of CRISPR repeat-spacer arrays and a collection of CRISPR-associated (cas) genes and spacer integrase that are associated with endonuclease activity adaptation period, respectively [20]. When prokaryotes are invaded by foreign genetic material, Cas proteins can cut the invading DNA into short fragments, which are subsequently incorporated into the CRISPR array as new spacers. When the same invader returns, crRNA quickly recognizes and pairs with foreign DNA, guiding Cas protein to break specific regions of foreign DNA and so safeguarding the host [21]. The current study identified two CRISPR-Cas systems in the *B. subtilis* RS10 genome (Table 1). In addition, integrase genes were also identified, but no cas gene was detected in RS10 genome. This may be due to the incomplete (draft) genome. Both the identified CRISPR systems consist of 100% conserved spacer regions, while 87% and 96% conserved direct repeats, respectively. We found no similarity between prophages and spacer sequence, the characteristic that allows integration of the prophages in recipient bacterium since CRISPR-Cas systems were unable to recognize them. The repeated sequence length of CRISPR system 1 is 24 bp, and CRISPR 2 is 25 bp (Table 2).

Table 1 Prophages identified in *Bacillus subtilis* strain RS10 genome

Phage	Size	Completeness	Score	CDS	Start–end	Closest phage	GC (%)
1	33.6 kb	Intact	120	46	364363–398003	Jimmer2 (NC-041976)	44.83
2	129.3 kb	Intact	93	175	1195877–1325188	SPbeta (NC-001884)	34.70
3	27.1 kb	Incomplete	20	13	2578211–2605349	Thermu (OH2-NC-021784)	43.89
4	45 kb	Questionable	90	66	3934164–3979207	SPP1 (NC-004166)	41.84
5	25.3 kb	Incomplete	30	33	4290850–4316212	Bacill_G (NC-023719)	43.26

Table 2 CRISPR-Cas systems found in *B. subtilis* RS10 genome

CRISPR-Cas	Direct repeats sequence	Repeated length (bp)	Repeats ratio	Conservation repeats identity	CRISPR length (bp)	Orientation
1	TGATGGGAATCGAACCCACGACAT	24	0.40	87.5%	107	Unknown
2	CAGCTGATTGCTGGTTTTGTTTTCT	25	0.44	96%	105	Reverse

Association between prophages and CRISPR-Cas systems

Bacteriophages are the major thread for Bacteria from where spacer in the CRISPR-Cas luci originated. If a bacteriophage invades bacteria, the spacer sequences in the strain carry a fragment that is corresponding to the phage genetic material. Therefore, the current study attempts to identify bacteriophages present in *B. subtilis* RS10 genome to infer the interaction between host strain and prophages. To determine the origin of foreign DNA (invaders), BLAST search of the extracted spacer sequences in the virus RefSeq database was conducted. The spacer sequence 1 was matched with three sequences (Siphoviridae (BK041997), Caudovirales (BK049247), and virus AG-345-E08 (MH319740)) in the RefSeq database, while no match was observed for the spacer sequence 2. A strain containing CRISPR-Cas with more spacer is expected to be matched with a greater number of prophages suggesting that such a strain possesses a promising adaptive immunity. To investigate the association of CRISPR-Cas system with the lysogeny of the prophages, the spacer sequences were BLAST against both prophages in the RS10 genome. The results showed no significant similarity between the spacer sequences and identified prophages. These results are in agreement with a previous study where prophages in *Bifidobacterium pseudocatenulatum* were analyzed and indicate that the number of prophages and CRISPR array are not associated with a number of CRISPR spacer and prophage region, respectively [22]. Overall, these results indicate that the strain RS10 has adaptive immunity against several viruses but not against these that are identified in the current study.

Conclusion

The current study identified two complete prophages and two CRISPR-Cas systems for the first time in the *B. subtilis* species. These phage genomes are mosaic and are capable to serve as a potential phage system to investigate the evolution and adaptation of *B. subtilis.* The prophages P1 and P2 exhibit high similarity with Myoviridae and Siphoviridae families, respectively, and encode biotechnologically important enzymes such as thermostable enzymes and ionizing radiation-resistant protein. Further, the genes related to DNA polymerase and holin were identified in prophages which can be used as biotechnological tools. On the other hand, numerous genes were identified with unknown functions, indicating a vast reservoir for new information to be explored. Further research is warranted to elucidate the molecular mechanisms underlying these phage-host interactions and their potential applications in agriculture and biotechnology.

Declarations

Ethics approval This research article does not include any experiments related to humans or animals.

Conflict of interest The authors declare no competing interests.

References

1. Quaiser A et al (2015) Diversity and comparative genomics of *Microviridae* in Sphagnum- dominated peatlands. Front Microbiol 6:375. https://doi.org/10.3389/fmicb.2015.00375

2. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV (2018) Viruses of archaea: structural, functional, environmental and evolutionary genomics. Virus Res 244:181–193. https://doi.org/10.1016/j.virusres.2017.11.025

3. Xu J, Hendrix RW, Duda RL (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell 16(1):11–21. https://doi.org/10.1016/j.molcel.2004.09.006

4. van Dijl JM, Hecker M (2013) *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microb Cell Factories 12:3. https://doi.org/10.1186/1475-2859-12-3. (England)

5. Iqbal S et al (2023) Classification and multifaceted potential of secondary metabolites produced by Bacillus subtilis group: a comprehensive review. Molecules 28(3). https://doi.org/10.3390/molecules28030927

6. Iqbal S, Ullah N, Janjua HA (2021) In vitro evaluation and genome mining of *Bacillus subtilis* strain RS10 reveals its biocontrol and plant growth-promoting potential. Agriculture 11(12):1273. https://doi.org/10.3390/agriculture11121273

7. Roux S, Enault F, Hurwitz BL, Sullivan MB (2015) VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985

8. Srividhya KV et al (2006) Database and comparative identification of prophages BT - intelligent control and automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006

9. Sayers EW et al (2020) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 48(D1):D9–D16. https://doi.org/10.1093/nar/gkz899

10. Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

12. Abe K et al (2014) Developmentally-regulated excision of the SPβ prophage reconstitutes a gene required for spore envelope maturation in *Bacillus subtilis*. PLoS Genet 10(10):e1004636. https://doi.org/10.1371/journal.pgen.1004636

13. Chevallereau A, Pons BJ, van Houte S, Westra ER (2022) Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol 20(1):49–62. https://doi.org/10.1038/s41579-021-00602-y

14. Ellis DM, Dean DH (1986) Location of the *Bacillus subtilis* temperate bacteriophage phi 105 attP attachment site. J Virol 58(1):223–224. https://doi.org/10.1128/JVI.58.1.223-224.1986

15. Pavesi A (2006) Origin and evolution of overlapping genes in the family Microviridae. J Gen Virol 87(Pt 4):1013–1017. https://doi.org/10.1099/vir.0.81375-0

16. Moeller R et al (2007) Role of DNA repair by nonhomologous-end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. J Bacteriol 189(8):3306–3311. https://doi.org/10.1128/JB.00018-07

17. Fogelman I et al (2000) Evaluation of CD4+ T cell function in vivo in HIV-infected patients as measured by bacteriophage phiX174 immunization. J Infect Dis 182(2):435–441. https://doi.org/10.1086/315739

18. Yin Y, Ni P, Deng B, Wang S, Xu W, Wang D (2019) Isolation and characterisation of phages against *Pseudomonas syringae* pv. Actinidiae. Acta Agric Scand Sect B — Soil Plant Sci 69(3):199–208. https://doi.org/10.1080/09064710.2018.1526965

19. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327(5962):167–170. https://doi.org/10.1126/science.1179555

20. Koonin EV, Makarova KS (2009) CRISPR-Cas: an adaptive immunity system in prokaryotes. F1000 Biol Rep 1:95. https://doi.org/10.3410/B1-95

21. Makarova KS et al (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9(6):467–477. https://doi.org/10.1038/nrmicro2577. (England)

22. Wang G et al (2020) The diversity of the CRISPR-Cas system and prophages present in the genome reveals the co-evolution of *Bifidobacterium pseudocatenulatum* and phages. Front Microbiol 11:1088. https://doi.org/10.3389/fmicb.2020.01088