Ecology of prophage-like elements in Bacillus subtilis at global and local geographical scales

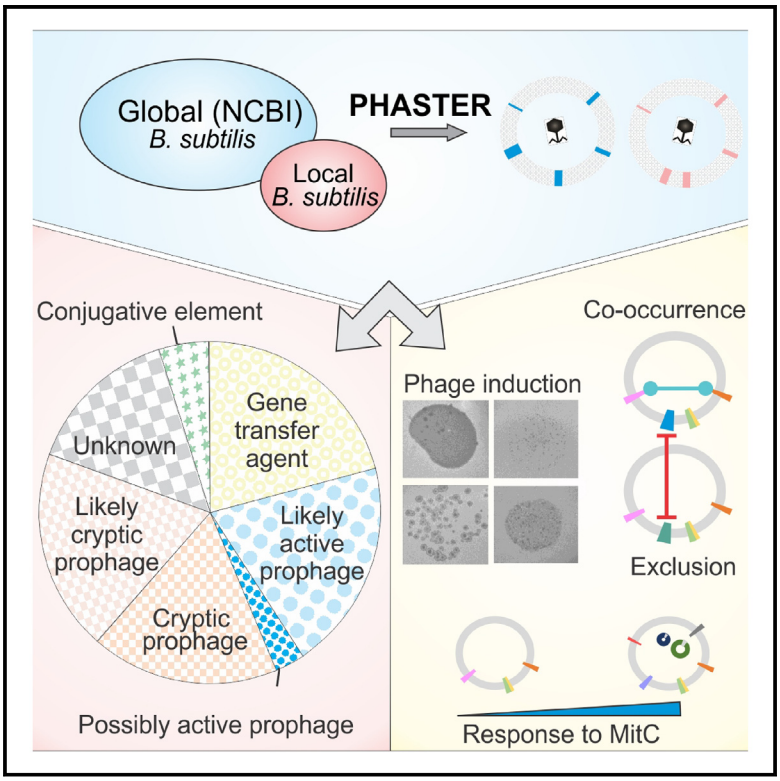
Stefanic, P.; Stare, E.; Floccari, V.A.; Kovac, J.; Hertel, R.; Rocha, U.; ... ; Dragoš, A.

Cell Reports

Article

Ecology of prophage-like elements in *Bacillus subtilis* at global and local geographical scales

Graphical abstract



Authors

Polonca Stefanic, Eva Stare, Valentina A. Floccari, ..., Ines Mandić-Mulec, Mikael Lenz Strube, Anna Dragoš

Correspondence

anna.dragos@bf.uni-lj.si

In brief

Stefanic et al. explore prophage diversity and distribution in *Bacillus subtilis* at global and local scales. Their findings reveal links between host phylogeny, isolation sites, and distinct phages, offering insights into prophage-host interactions, antagonism, and co-dependence. This study enhances understanding of bacterial genome evolution and prophage roles in ecology.

Highlights

- Local *Bacillus subtilis* isolates carry both common and distinct prophage elements
- Phylogeny predicts prophage presence better than geographic origin
- Prophages show mutual exclusion and co-dependence within host genomes
- Most predicted prophages are not stress induced but increase host stress sensitivity

Stefanic et al., 2025, Cell Reports *44*, 115197
January 28, 2025
https://doi.org/10.1016/j.celrep.2024.115197

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Ecology of prophage-like elements in *Bacillus subtilis* at global and local geographical scales

Polonca Stefanic,[1] Eva Stare,[1] Valentina A. Floccari,[1] Jasna Kovac,[2] Robert Hertel,[3] Ulisses Rocha,[4] Ákos T. Kovács,[5,6] Ines Mandić-Mulec,[1] Mikael Lenz Strube,[6] and Anna Dragoš[1,7,*]

[1]Department of Microbiology, Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia
[2]Department of Food Science, Pennsylvania State University, University Park, PA 16802, USA
[3]Department of Genomic and Applied Microbiology, Georg-August-University of Göttingen, 37077 Göttingen, Germany
[4]Department of Applied Microbial Ecology, Helmholtz Center for Environmental Research, 04318 Leipzig, Germany
[5]Institute of Biology, Leiden University, 2333 Leiden, the Netherlands
[6]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs Lyngby 2800, Denmark
[7]Lead contact
*Correspondence: anna.dragos@bf.uni-lj.si
https://doi.org/10.1016/j.celrep.2024.115197

SUMMARY

Prophages constitute a substantial portion of bacterial genomes, yet their effects on hosts remain poorly understood. We examine the abundance, distribution, and activity of prophages in *Bacillus subtilis* using computational and laboratory analyses. Genome sequences from the NCBI database and riverbank soil isolates reveal prophages primarily related to mobile genetic elements in laboratory strains. Distinct and previously unknown prophages in local isolates prompt an investigation into factors shaping prophage presence, with phylogenetic relatedness predicting the prophage repertoire slightly better than geographical origin. Data also show that prophages exhibit strong co-occurrence and exclusion patterns within genomes. Laboratory experiments indicate that most predicted prophages are cryptic, as they are not induced under DNA-damaging conditions. Importantly, stress responses increase with the number of predicted prophages, suggesting their influence on host physiology. This study highlights the diversity, integration patterns, and potential roles of prophages in *B. subtilis*, shedding light on bacterial genome evolution and phage-host dynamics.

INTRODUCTION

Bacteriophages (phages) are the most widespread biological agents on planet Earth.[1] They not only act as bacterial predators but can also profoundly impact host bacteria when integrated into their genome. During lysogeny, the host genome integrates phage elements, called prophages, that can influence a wide array of bacterial traits, such as virulence,[2,3] dormancy,[4] interference with neighbors,[5] resistance to lytic phages,[6] metabolic pathways,[7,8] and cell-cell communication.[9–11] A phenomenon where bacterium acquires a new trait due to the presence of a prophage is known as lysogenic conversion.[12] Importantly, once lysogenized, a bacterial cell may become resistant to subsequent infections by other, similar phages, via so-called superinfection exclusion.[12] The effects of prophage elements on their bacterial hosts can further influence higher organisms colonized by lysogenized bacteria, which can translate to global impact phenomena such as disease outbreaks[13,14] or mating patterns in insects.[15] Given the key role of prophages and their presence in nearly 70% of all bacterial genomes,[5,16] it is important to understand the effects of particular prophage elements on their bacterial host, their global distribution, their potential for transmission, and the factors controlling their abundance, transmission, and stability within host genomes.

The abundance of prophage-like elements differs across bacterial phyla, genera, and species,[5,16,17] as well as within species.[18,19] The diversity of mobile genetic elements (MGEs), including prophages, across different species hinders scientific progress in understanding the broader ecological roles of prophages. This is because most experimental studies, even for well-established model species of medical, agricultural, or industrial significance, are typically conducted on just a few well-characterized strains, such as *Bacillus subtilis* 168,[20,21] *Escherichia coli* K-12,[22] and *Pseudomonas aeruginosa* PAO1.[23]

Consequently, we mostly explore prophages and MGEs that specifically occur in these laboratory strains, while the significance and dynamics of these mobile elements are not recognized in a broader ecological context.

B. subtilis is a versatile soil-dwelling bacterium capable of thriving in a wide range of environments, including terrestrial and aquatic environments, as well as in humans and other animals. *B. subtilis* is widely used in the industrial production of enzymes

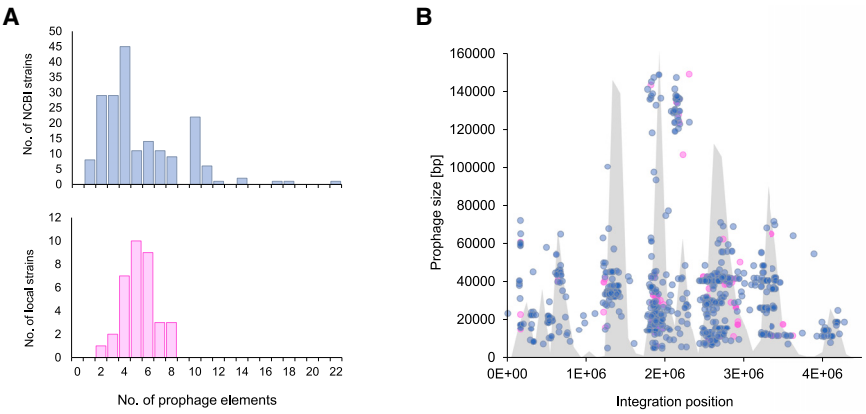


Figure 1. Predicted prophage elements and prophage integration hotspots in *B. subtilis*

(A) Histograms showing the distribution of prophage elements in analyzed *B. subtilis* genomes, where global-scale NCBI isolates (blue) and local isolates (pink) are represented separately.

(B) Chromosomal prophage elements are plotted according to integration position (x axis) and genome size (y axis). Data obtained for NCBI isolates are presented in blue, and data for local isolates are presented in pink. The filled histogram in gray depicts the distribution of prophage elements across genomes to indicate the locations of integration hotspots.

and antibiotics,[24] controlling plant diseases and promoting plant growth,[25] improving gut health,[26–28] boosting immunity in mammals,[29] and promoting the overall wellness of humans and animals.[30–32] *B. subtilis* species form highly resistant spores that can withstand exposure to UV, heat, desiccation, and lack of nutrients.[33] Furthermore, *B. subtilis* can travel via air over long distances through dust storms.[34,35] Despite global mixing, a recent study revealed that the relatedness of the *B. subtilis* strains is driven by habitat heterogeneity and sustained at different spatial scales within the environment.[36] Niche-driven diversification seems to be faster than the rate of dispersal, but how habitat-driven diversification affects the acquisition dynamics and evolution of the phage arsenal is unknown.

As a member of the Bacillota (previously Firmicutes) phylum, *B. subtilis* is known to be rich in prophage elements.[17] Recent experimental studies demonstrated the impacts of these elements on intraspecific competition,[5,37] dormancy,[4] surface colonization,[38] and genome stability.[37,39] It was also shown that lytic cycle induction of certain prophages can be inhibited by other MGEs, such as conjugative plasmids,[40] and that the presence of such inhibitory plasmids can correlate with reduced genetic diversity within prophage group.[39] Furthermore, prophages and defective phage plasmids of this species impair the genetic competence of *B. subtilis*.[41,42] The above findings suggest that prophage elements play an important role in host adaptation and that the presence of certain MGEs may influence the stability of prophages within host genomes. However, only a small number of *B. subtilis* temperate phages and prophage-like elements have been described and experimentally characterized, and those examined have been mostly derived from laboratory strains.[39,43,44] As a result, we know very little about the general abundance, phage activity, and ecological relevance of already described temperate phages or the presence, diversity, and ecology of other, unrelated prophages in non-domesticated isolates. Furthermore, the influence of its environmental niche on the phage repertoire of *B. subtilis* strains remains unknown.

This work investigated prophages and prophage-like elements within *B. subtilis* strains at global (publicly available genome sequence databases) and local (our collection of soil isolates) geographical scales. We confirm that *B. subtilis* species are rich in prophage elements and that these elements integrate into a few specific genomic hotspots. Based on encoded protein similarity, we clustered prophages of *B. subtilis* into dozens of groups, some globally distributed and others restricted to local geographical areas. We also demonstrate clear positive or antagonistic interactions in terms of genomic co-occurrence between certain prophage elements that likely drive their diversification. As isolates from global and local scales showed similar levels of prophage diversity, we experimentally tested the potential of prophages from our local *B. subtilis* strain library for horizontal transmission. Only a minority of prophages could be activated with a common inducing agent. Our work highlights the ecological dynamics of prophage elements within a bacterial species based on public genome sequence database information and a locally collected isolate library, with insights into the genetic interactions (coexistence, exclusion) between MGEs within host genomes.

RESULTS

B. subtilis species are rich in prophages mainly located in nine integration hotspots

We predicted prophages in a large set of *B. subtilis* genomes, including all 191 complete genomes available on NCBI (hereafter referred to as global-scale isolates) and 40 in a recently sequenced isolate library from a defined 1 cm^3 soil sample from the Sava riverbank,[45] referred to as local isolates. It is important to note that functional categories assigned by the prediction software were considered unreliable and hence ignored, and some of the predicted elements may be cryptic. However, for simplification, we will refer to all predicted prophage-like elements as prophages thorough this manuscript. Collectively, these genomes contained 1,243 prophages (Data S2), and although each genome contained at least one prophage element, most strains carried up to four prophages. Genomes from local-scale strains had a slightly different distribution of predicted prophages, with the mode shifted toward a higher average number of prophages (five) per genome but with up to a maximum of eight (Figure 1A). Among global-scale strains, 35 carried 10 or more prophages (Figure 1A). Next, we mapped the predicted prophages on a graph according to their integration position in the host chromosome, revealing the presence of at least nine prophage-integration hotspots distributed across the chromosome. The integration hotspots nearly overlapped

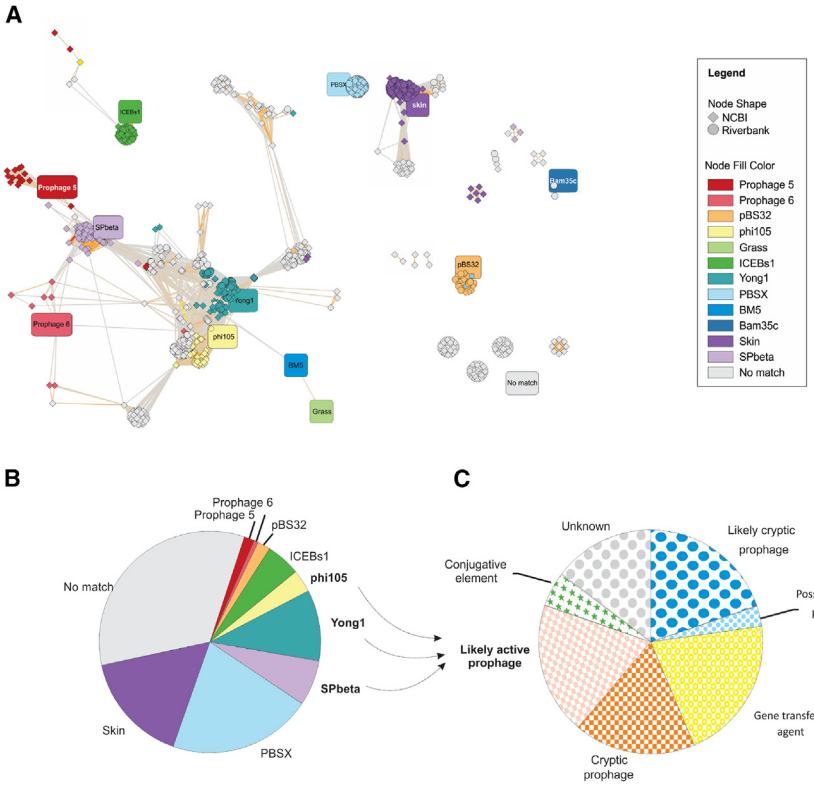


Figure 2. Clustering of predicted prophage elements

(A) Prophages colored according to clustering. Reference phages or known mobile genetic elements (MGEs) are depicted as rounded squares, with each color representing one phage cluster. Prophages found within genomes available from the NCBI are shown as diamonds, while prophages from riverbank soil local-scale samples are shown as circles. The main phage clusters are depicted in the colored rectangles shown next to the clusters.

(B) Distribution of predicted prophage elements according to reference matching via nucleotide BLAST.

(C) Distribution of predicted prophage elements according to their predicted function, based on similarity to active phages, other MGEs, and known *B. subtilis* 168 chromosomal regions. Phages were classified according to the known function of their reference match. In addition, phages matching *B. subtilis* 168 non-MGE chromosomal regions with >30% coverage and >88% identity were classified as likely cryptic. Prophages with lower coverage (<30%) or with no match to references were classified as unknown.

between local- and global-scale isolates (Figure 1B; Data S2). The abundance of prophage elements varied among the identified hotspots, with the highest abundance of prophages at hotspots near the replication terminus and the lowest abundance close to the origin of replication (Figure 1B). Moreover, the sizes of prophage elements differed depending on their respective integration position. For example, relatively small prophages (averaging 14.6 ± 2.8 kb) were predominantly found at position ~4 Mb, while two hotspots around position ~2 Mb exhibited a bimodal distribution of prophage sizes (Figure 1B), with a distinct group of large prophages exceeding 120 kb (Figure 1B; Data S2).

Most prophages are associated with known *B. subtilis* phages or MGEs, but two prophage groups are unique to local-geographical-scale isolates

Based on total prophage numbers, fragment sizes, and integration sites, broad phylogenetic diversity of prophage elements was expected, which led us to use vContact2[46] to build a local phylogeny for our phage collection. This network derived from vContact2 was subjected to clustering using clusterONE, resulting in 62 clusters (Data S2 and S3), which were subsequently used as vContact2-derived viral operational taxonomic units (vOTUs) in downstream analysis. To determine the taxonomy of the predicted prophages, we used discontinuous mega-BLAST to match them to the INPHARED phage database[47] of confirmed active phages, which we supplemented with sequences of previously described MGEs within *B. subtilis* 168 (see STAR Methods). Finally, prophages from vOTUs that did not have a match within the aforementioned references were additionally compared against the genome of *B. subtilis* 168 without MGEs (see STAR Methods) to check for matches to annotated *B. subtilis* chromosomal regions that are not functional prophages. This approach allowed us to classify most of the predicted prophages according to their relatedness to known functional phages or previously described *B. subtilis* MGEs as well as according to the predicted phage activity (Figures 2A–2C; Data S3).

The most prominent group of predicted prophages (n = 238) was phylogenetically related to defective phage PBSX, followed by prophages similar to cryptic prophage *skin* (n = 182), recently isolated phage Yong1 (n = 116), and temperate phage SPβ (n = 74; Figures 2A and 2B; Data S3). Other relatively abundant groups were prophages with similarity to conjugative element ICEBs1 (n = 56), previously described temperate phage phi105 (n = 37), a cryptic prophage found on the pBS32 plasmid present in *B. subtilis* 3610 (n = 21), or cryptic prophage elements 5 (n = 17) and 6 (n = 7), which are also present in *B. subtilis* 168 (Figures 2A and 2B; Data S3). While PBSX, *skin*, ICEBs1, and pBS32 clusters are distant from each other and from other vOTUs, clusters of SPβ, elements 5 and 6, phi105, and Yong1 are genetically interconnected, which suggests a substantial sharing of proteins between most phages with little phylogenetic conservation (Figure 2A). In addition to taxonomy, the vContact2 network revealed substantial differences in proteome conservation between phage groups, with some groups (e.g., PBSX and ICEBs1) being rather conserved and others (e.g., *skin*, Yong1, phi105, and SPβ) being more diverse, displaying lower similarity within vOTUs and even grouped into several different vOTUs (Data S3; Figure S1).

We also observed that prophage elements that clustered together tended to reside in similar integration sites within

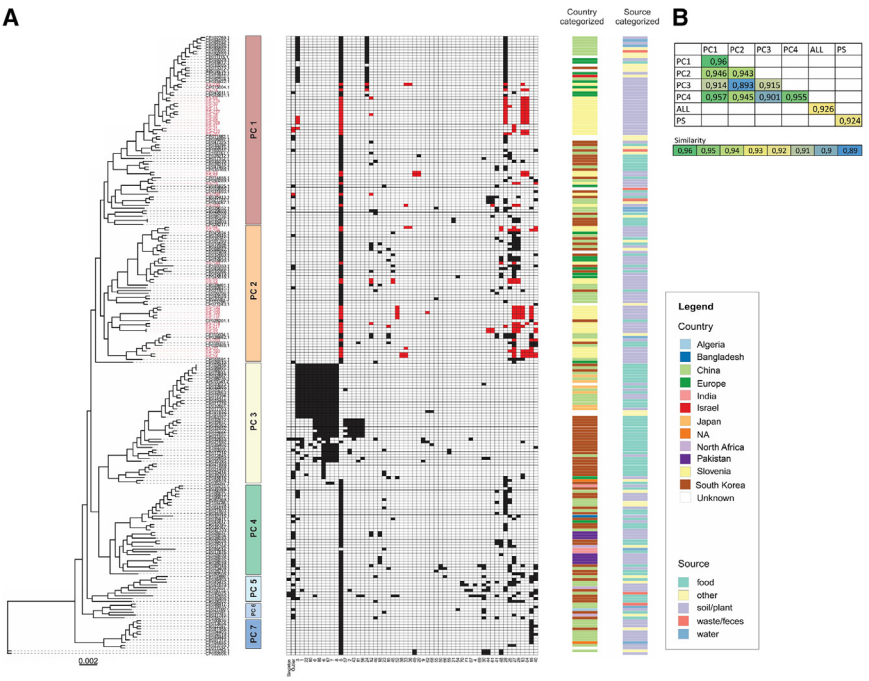


Figure 3. Relationship between phylogeny, prophage cargo, and the country/isolation source of *B. subtilis* strains

(A) From left to right: phylogenetic tree illustrating the phylogeny (core gene alignment) among all analyzed *B. subtilis* strains; phylogenetic cluster demarcation, with each color representing a different phylogenetic clade; and matrix of presence/absence of prophage elements (red indicating the presence of local strains, black indicating the presence of global strains, and an empty space indicating an absence) associated with specific clusters, country of origin, and the source of strain isolation. Singletons are unique prophages having no similarity to any vOTU, and outliers are prophages with too little similarity to a vOTU to be considered a member: in both cases, these are collapsed into one group.

(B) Prophage presence/absence similarity scores calculated within or between phylogenetic clusters or within local-scale isolates (see STAR Methods). Briefly, the similarity scores within strains for each phage vOTU were calculated, and scores of all vOTUs were averaged to obtain similarity at the level of the entire prophage repertoire (see also Data S4).

B. subtilis genomes (Figure S2). However, there were certain exceptions, where phages classified into different phylogenetic groups were found at similar chromosomal locations (e.g., *skin*-, Yong1-, and phi105-related prophages). Prophages of these three groups were among the less conserved in terms of chromosomal position, as they were found in multiple different chromosomal positions (Figure S2). Finally, we gained further insights into larger clusters of completely unknown prophages, for example, vOTU 8 that always seems to integrate near Yong1, vOTU 0 that always integrates close to the origin of replication, and vOTU 16 that may replace the *skin* element in certain genomes (Figure S2).

A minor portion of *B. subtilis* strains also carried plasmids, and a minority of those also contained prophages, which clustered into a few different vOTUs, entirely unrelated to all other phages (Figure 2A). Finally, certain relatively large clusters were found (e.g., vOTUs 8, 16, 0, and 40) that had no matches with known phages or MGEs or matched poorly with *B. subtilis* 168 chromosomal regions. These were assigned as completely unknown (Figures 2A–2C; Data S3). Based on similarity to references and chromosomal regions, we classified ~23% of predicted prophages as likely or possibly active, ~63% as cryptic, likely cryptic, or other MGEs that are not functional phages, and ~15% as completely unknown (Figure 2C; Data S3).

Finally, we compared prophage clustering networks of database (NCBI) and local-scale bacterial strains, specifically asking whether there are any distinct prophage groups associated solely with strains from local-scale samples. While the vast majority of prophage groups found among riverbank isolates were represented also within global-scale isolates, we found groups that were unique to local-scale isolates, specifically two plasmid prophages assigned as outliers associated with the Bam35c reference bacteriophage and a unique cluster of Yong1-related phages (vOTU 64) that were unique to local-scale isolates (Figure 2A).

Specific host PCs are defined by distinct sets of prophages

To better understand the connection between host phylogeny and the similarity of prophage elements, a phylogenetic tree based on all analyzed *B. subtilis* strains was constructed using core gene alignment (see STAR Methods). *B. subtilis* strains were clustered into seven phylogenetic clades (PCs), with the local strains belonging to only two clades (PC1 and PC2), which were also more closely related to each other than to other clades (Figure 3).

Mapping prophage elements to the phylogeny of strains revealed that phylogeny correlated with the presence of specific prophage elements to a certain extent (Figure 3; see STAR Methods). This was confirmed after we determined prophage repertoire similarity scores within (0.96–0.915) and between (0.955–0.893) the largest clades (Data S4; Figure 3; see STAR Methods). For all except PC3, prophage repertoires were more similar within clades than between clades when strains from all four clades were included (Figure 3). Clade PC3 was rather unique, as it was characterized by the highest diversity of prophage elements (0.915) and by strains carrying a distinct set of prophage elements compared to other strains (Figure 3). Interestingly, PC3 strains were mostly missing the most abundant prophage vOTU 5 (PBSX) and carried vOTU 6 (also PBSX) instead. There was no obvious distinction in phage repertoire between clusters containing fewer strains (clusters PC5–PC7). It was difficult to estimate the effect of local environment on the diversity of prophage repertoire among isolates, mostly because local strains clustered into PC1 and PC2, which also exhibited rather high similarity of prophages between clades (Figure 3).

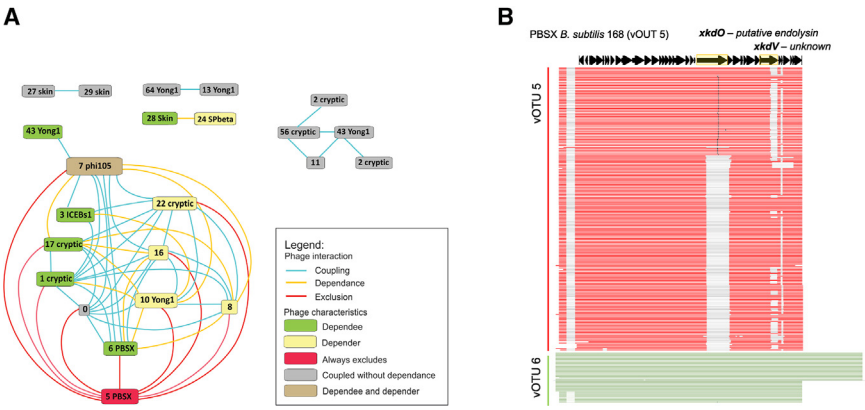


Figure 4. Map of prophage interactions and characteristics considering their co-occurrence within a single host chromosome

(A) Prophage vOTU labels colored according to their relationships with other vOTUs within a single chromosome: dependee, prophages of other vOTUs tend to occur only when this vOTU is present; depender, this vOTU occurs only in the presence of another vOTU; always excludes, if this vOTU is present, certain other vOTUs never occur; coupled without dependence, tends to co-occur with another vOTU; depender and dependee, its presence depends on a certain vOTU but also determines the presence of another vOTU. Lines connecting squares are colored according to interactions: coupling (blue), exclusion (red), and dependence (yellow).

(B) Graphic summary of alignments of all PBSX genomes from vOTU 5 (colored red) and vOTU 6 (colored green) obtain by BLAST. Map of PBSX from *B. subtilis* 168 (GenBank: NC_000964.3) is shown above the alignment. The two regions that are commonly distinct between vOTU 5 and vOTU 6 overlap with open reading frames of *xkdO* and *xkdV* of PBSX *B. subtilis* 168.

Nevertheless, within local strains, the prophage similarity score (0.942) was lower than prophage similarity scores within PC1 (0.96) and PC2 (0.943), suggesting that phylogeny was a better predictor of prophage repertoire than strain isolation site. This was also in line with a rather random distribution of countries and isolation sites across the phylogenetic tree (Figure 3). Together, these results indicate that closely related strains tend to carry similar prophages.

Both mutual inclusion and exclusion of prophages are common within bacterial genomes

Next, we tested the hypothesis that prophage elements shape the prophage repertoire of the host, for instance, by either preventing or promoting acquisition of prophages. In particular, this hypothesis was suggested by several unique prophage elements within PC3 (Figure 3) that seemed to co-occur within genomes belonging to this PC. Our analysis revealed both negative and positive co-occurrence between predicted prophage elements (Figures 4 and S4; Data S1).

Among all vOTUs, roughly a third (21 of all 62 vOTUs) were involved either in positive or negative co-occurrences with other prophage elements within the host chromosome. From all types of interdependencies detected, most vOTUs showed coupling (statistically significant co-occurrence of both vOTUs within the host chromosome), followed by dependence (where the presence of one vOTU depends on the presence of another, which is not dependent on the first) and exclusion (where a certain vOTU is never found in the presence of another vOTU), as shown in Figure S4. Prophage elements that seemed to have the largest effect on the overall prophage arsenal were PBSX vOTU 5, which showed exclusion toward nine other vOTUs; PBSX vOTU 6, which showed coupling with five vOTUs and determined the presence of another three vOTUs; and phi105 vOTU 7 and vOTU 0 and Yong1 vOTU 10, which also showed a significant number of positive interactions (Figures 4 and S4; Data S1).

To delve deeper into the factors contributing to these correlations, we focused on two groups of PBSX elements, vOTU 5 and vOTU 6, which never co-occurred within the host chromosome and which exerted a profound impact on the presence/absence of other prophages (Figures 4 and S4). PBSX of vOTU 5, encompassing ~83% of all identified PBSX prophages, exhibited a strong antagonistic relationship with PBSX in cluster 6 (constituting ~17%) and with phage phi105. Consequently, if an isolate carried PBSX of cluster 5, it was highly unlikely to be lysogenic for phi105 vOTU 7, Yong1 vOTU 10, or a few other unknown or likely cryptic prophages. Conversely, PBSX of vOTU 6 displayed a strong positive correlation with these prophage elements (Figures 4 and S4). Furthermore, we noticed coupling between certain types of SPβ prophages and the *skin* element (Figures 4 and S4).

Given that PBSX of vOTU 5 is the most widespread prophage, we hypothesized that protection from attack and lysogenization by multiple other phages may select for its maintenance within the species. An interesting observation for this prophage was the existence of an alternative PBSX variant differing in the cell wall hydrolase *xkdO* gene region and the end fragment of the hypothetical protein-encoding gene *xkdV* (Figure 4B).

Overall, the results of prophage co-occurrence analysis clearly show that the likelihood of the acquisition or loss of prophage elements belonging to certain group (vOTUs) can be predicted based on the presence or absence of other specific prophage group.

Strains having more predicted prophages show a stronger DNA damage response, but only some prophages appear to be induced by a DNA-damaging agent

Previous experimental data suggest that the majority of computationally predicted prophage elements are non-functional temperate phages,[48] meaning they are not capable of forming phage particles and propagating via a lytic cycle. To challenge bioinformatic prophage prediction with experimental data, we tested 40 local isolates from the soil microscale for spontaneous or stress-induced release of active phage particles using mitomycin C. We found that prophages could be induced in 6 of 40 *B. subtilis* strains (PS-11, PS-24, PS-25, PS-233, PS-237, and

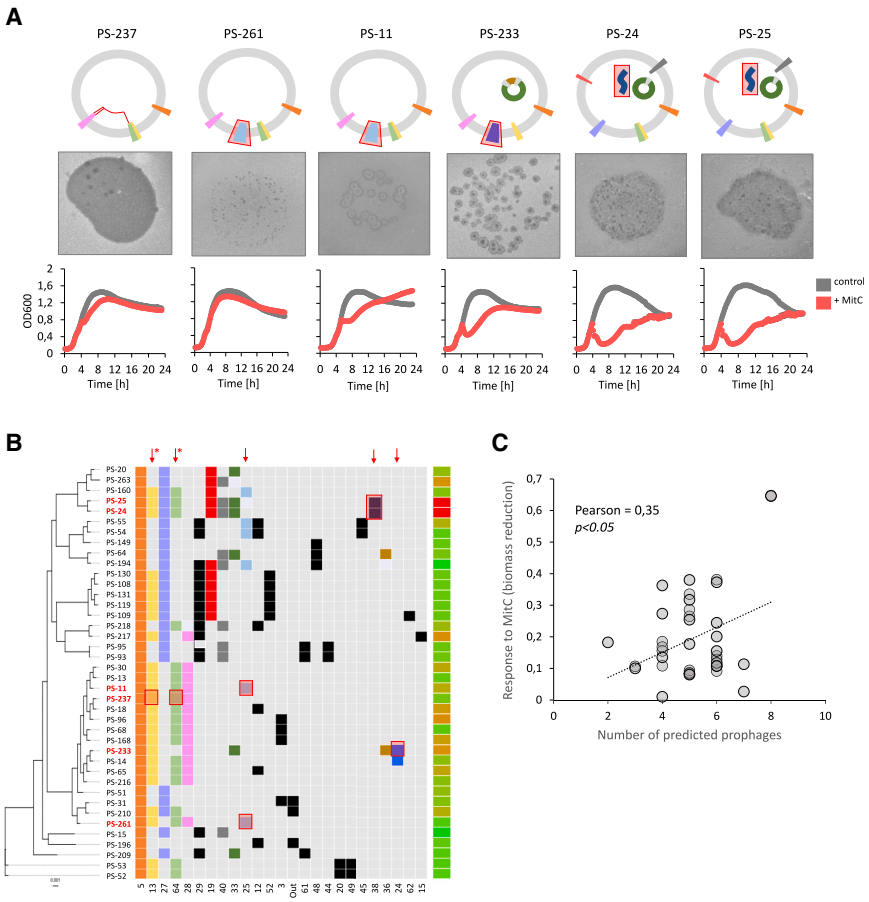


Figure 5. Screening for active prophages within local-scale isolates

(A) Of 40 strains treated with mitomycin C, six showed the presence of active phages in their supernatants (PS-237, PS-261, PS-11, PS-233, PS-24, and PS-25). Prophage elements predicted within their genomes are schematically labeled on circular graphs. Elements that were identified as active by phage DNA sequencing are labeled in red. Graphs underneath plaque images represent growth curves with and without the addition of mitomycin C (0.1 μg/mL). The results are the means of six biological replicates, and error bars represent ± standard error of the mean. Error bars are plotted but not visible due to being smaller than the data markers.

(B) Phylogenetic tree of local strains, with the presence of prophage elements that belong to certain similarity clusters represented by squares of different colors. Supernatants of strains labeled in red produced plaques on the indicator strain when collected post-prophage induction with mitomycin C. The heatmap on the right represents the strength of the response to mitomycin C, measured as biomass reduction. Active prophage elements as identified by phage DNA sequencing are labeled by red rectangles, and entire vOTUs within which active prophages were identified are labeled by red arrows. Arrows with an asterisk signify partial coverage of phage DNA reads.

(C) Number of predicted prophage elements versus strength of response to mitomycin C measured as biomass reduction compared to non-treated controls. P, Pearson correlation coefficient. Boxes represent the first and third quartiles. Lines represent the median, and bars span maximum to minimum values. All data points are overlaid on boxes.

PS-261), producing titers in the range of 10^2–10^7/mL using *B. subtilis* Δ6 (ΔSPβ Δ*skin* ΔPBSX Δprophage 1 pkscat Δprophage 3) as an indicator strain, and in 2 of these 6 strains, the induction was spontaneous (without mitomycin C treatment), as indicated by the presence of plaques on the lawn of the indicator *B. subtilis* strain Δ6 (Figure 5A; Data S5).

Spontaneous induction rates were too low to determine active prophages from sequencing coverage (Figure S5). Therefore, phage DNA was extracted from post-induction lysates and subjected to Illumina sequencing (see STAR Methods). The obtained sequencing reads clearly mapped onto predicted prophage sequences within the genomes of PS strains: in the case of PS-24 and PS-25, reads mapped onto linear phage plasmids (vOTU 38); for PS-11, PS-233, and PS-261, they mapped onto SPβ-related elements (vOTU 25 and vOTU 24, respectively); and for PS-233, they distributed among Yong-related elements vOTU 13 and vOTU 64 (Figures 5A and S6; Data S5). As our previous analysis suggested a positive association between vOTU 13 and vOTU 64 (Figure 4A), these data strongly suggest the co-dependence of these two prophage elements and the assembly of a composite phage.

Regardless of plaquing on the indicator strain, we also tested how all 40 strains responded to the optimized concentration of mitomycin C (0.1 μg/mL), which triggered strong biomass reduction mainly in strains carrying active prophages (Figure S7). However, not all strains that carried predicted active phages also showed strong biomass reduction in the presence of mitomycin C (Figure 5B; Data S5), and vice versa. For example, strains PS-18 and PS-96 showed strong biomass reduction but did not produce active phages (Figure S8). In theory, strains PS-18 and PS-96 (and other strains that showed strong responses to mitomycin C but did not plaque) could release phages that are not specific to the indicator strain. Nevertheless, there was no relationship between host phylogeny and the detection of plaques on the indicator strain (Figure 5B; Data S5), suggesting that a growth response to mitomycin C does not always reflect the presence of active prophages in the host genome.

Next, we compared the total number of predicted prophages based on the strength of the response to mitomycin C (calculated as the reduction in biomass compared to untreated controls; Figure 5C). Strains with more predicted prophages within their genome were found to respond more strongly to mitomycin C (Figure 5C; Data S5). These results indicate that although most of the predicted prophage elements were no longer able to transmit and replicate in a new bacterial host, they still triggered cell lysis in response to DNA damage. The results also show that vOTU membership cannot predict the activity of a prophage element (e.g., the majority of strains carrying vOTU 25 did not

show prophage activity) and that even elements with high induction rates can be restricted to only a minority of strains within the microscale (e.g., prophages of vOTU 38). The results also suggest that coupling detected between certain prophages within a single host chromosome (Figure 4A) can be due to the assembly of composite phages from two closely related and distantly integrated prophages upon lytic cycle induction.

DISCUSSION

Here, we assessed prophage elements, their diversity, and potential interactions within complete *B. subtilis* genomes available on NCBI and within a collection of *B. subtilis* strains isolated from 1 cm^3 of riverbank soil. Recently, prophage elements were assessed in 194 *B. subtilis* genomes available on NCBI,[49] where diversity analysis was limited to prophage elements categorized as intact.[49] In our approach, we ignored the functionality categories assigned by the PHASTER software to remove many relevant MGEs (such as gene transfer agents or plasmids with phage features) from the analysis. A genomic comparison identified 55 distinct types of prophages with varying genome sizes, integration sites, and gene content, as well as several groups of prophage elements with homology to temperate phi105.[49] Our protein-based clustering analysis confirms these results, as well as the presence of two groups of SPβ-like phages.

We found that strains from local-geographical-scale samples had a prophage frequency distribution distinct from that of global isolates, with a higher mode value but a lack of extreme cases with >8 prophage elements, which were relatively common among global isolates. We could also identify certain groups of prophage elements that were solely found within local-scale samples. Unique within-genome distribution combined with unique prophage cargo of local-scale isolates may suggest the rapid acquisition and spread of temperate phages from and within local microbiomes, which also aligns with previous experimental studies.[50] Our work also clearly demonstrates barriers against phage transmission; even strains that coexist within the local scale and are closely phylogenetically related differ profoundly in the prophage repertoire they carry. Moreover, even though some of these prophages showed high potential for lytic cycle induction and horizontal transmission (like plasmid prophages present in strains PS-24 and PS-25), they were restricted to only 2 of 40 isolates within the local isolate library.

In both global and local isolates, the integration of prophage elements was restricted to a few specific genomic hotspots, with increasing numbers of integrated prophages close to the replication terminus. This observation is in line with previous work[18] and seems to be a general cross-species phenomenon, likely linked to transcriptional regulation and easier maintenance of stable lysogeny in the chromosomal region where gene expression (including lytic cycle genes) is lowest.[51]

As most of the present knowledge about *B. subtilis* MGEs is derived from laboratory strain *B. subtilis* 168 and its ancestor strain, *B. subtilis* NCIB 3610, we attempted to group the predicted prophage elements according to their similarities to MGEs residing in these strains. Our work confirmed the conservation of defective prophage PBSX in all isolates of local and global scale. The second most abundant group included phage elements that clustered with cryptic prophage *skin*, recently characterized *Bacillus* phage Yong1, temperate phage SPβ, and functional temperate phage phi105. Meanwhile, within natural isolates, we found PBSX and *skin* that were identical to those present in *B. subtilis* 168 but no strains carrying full-length Yong1 or phi105 in a prophage form. This may suggest a strong selection for the domestication of this temperate phage within the host genome.[52]

How and why PBSX is strictly maintained within *B. subtilis* is unclear. Although it was successfully deleted from the *B. subtilis* genome and classified as dispensable,[53] its conservation within species implies a strong ecological advantage. It was recently shown that PBSX is involved in extracellular vesicle formation during stress-induced cell lysis, a process that depends on expression of the PBSX endolysin gene.[54] Previous studies also suggest that PBSX may be constantly induced in a small subpopulation of cells,[55] which could potentially infer toxicity toward PBSX-negative strains and cause a rapid vertical spread of this element. Importantly, we show that PBSX exists in two taxonomic subgroups, which are strictly exclusive toward each other and where one strictly antagonizes phi105-related prophage elements. This suggests that this dominant PBSX type could protect *B. subtilis* from other MGEs, like phi105, especially since the latter was classified as a "naive" phage, entering the lytic cycle during *B. subtilis* genetic competence development.[42] This observation is also in line with previous reports on competition between phages for intracellular resources during lytic induction[56] and superinfection immunity encoded within prophage elements.[57]

In addition to the strict antagonism between the two subgroups of PBSX, we found a strict co-occurrence between certain Yong 1 and phi105 "relatives," as well as unknown and cryptic prophage groups. It is likely that phi105 vOTU 7, vOTU 17 (cryptic), and vOTU 22 (cryptic) are remnants of a single prophage because they have neighboring integration sites (*att* at 3.2 Mb). The fact that they always co-occur with distantly integrated elements such as Yong1 vOTU 10 (*att* at 1.7 Mb) may indicate that all three are needed for the maintenance of lysogeny and host survival. Another possibility is molecular piracy, in which one defective element cannot spread without the other helper element[58]—a hypothesis that is strongly supported by phage DNA sequencing results, indicating that sequences from two topologically distant prophage elements showing statistically significant co-occurrence within a single chromosome (both related to Yong1) can indeed assemble into a composite phage.

Interestingly, certain prophage elements that were abundant among wild isolates clustered together with well-described non-prophage regions of domesticated *B. subtilis* 168. For example, elements clustered with conjugative element ICEBs1 were found in 54 strains spanning local- and global-scale isolates. Although this element was previously shown to block lytic cycle induction in SPβ,[59] we did not find any antagonism or statistically significant patterns of co-occurrence between these two elements in *B. subtilis* genomes.

We also found that some prophage groups present in wild strains partially or almost completely matched chromosomal regions of *B. subtilis* 168. This may suggest either a misprediction

by the software or the presence of prophage remnants in the lab strain. Some of these regions, such as those encoding biosynthetic gene clusters, deserve special attention, as in addition to sublancin encoded by SPβ, other biosynthetic gene clusters may also be transferred within prophage genomes. Finally, we found that certain genomic regions of *B. subtilis* 168, proposed to be prophage regions, had very few or no matches to the collection of natural isolates (e.g., prophage element 1, 2, or 3), strongly suggesting they are not part of functional prophages.

While very few *B. subtilis* strains carried plasmids, many of these appear to carry prophage-like elements. We discovered very potent functional phages on small linear plasmids within PS-24 and PS-25 local-scale isolates. Further studies are needed to better understand their role in bacteria and/or host range. Our experimental data support recent findings[48] showing that despite the abundance of predicted prophage regions within *B. subtilis* chromosomes, only a fraction are induced by mitomycin C. Importantly, it remains to be verified whether some of these prophages, which we failed to induce with mitomycin C, could still be inducible in the presence of other chemical or physical agents.[48,60] Our work also shows that the most active phages are not necessarily widespread within the local community. Further studies are needed to determine whether the presence of plasmid prophages with a high spontaneous induction rate may provide certain benefits to host bacteria. We also found that not all prophage elements within the same vOTU were inducible by mitomycin C. Further work is required to determine whether there are consistent genomic differences between the inducible and non-inducible variants or if there are notable variations in the prophage repertoires of their host strains. Although *B. subtilis* is a model organism and an industrially relevant species, knowledge of the natural diversity of its prophages remains limited. Our study uncovers this diversity and novel interactions between MGEs, as well as their level of conservation at the global scale and their level of uniqueness at the local scale. To what extent these elements drive interactions within the species and adaptation to certain growth conditions, such as fermented food, remains to be discovered. Our work opens multiple possible directions for further experimental research on the molecular mechanisms of antagonistic and positive interactions between MGEs, as well as host range, transmission, and conversion studies on active MGEs.

Limitations of the study

This study faced several limitations related to the bioinformatics tools used and the experimental procedures. One of the primary limitations is the reliance on PHASTER, which uses an internal database of known phages and phage genes. This tool, while useful, has inherent limitations in both sensitivity and specificity. The PHASTER database, last updated in December 2020, may be incomplete, which could result in missed detections. Furthermore, PHASTER may generate false positives, particularly if it relies on conserved regions of known phages, leading to an overestimation of prophage elements. The study also employed the INPHARED database, which depends on dc-megaBLAST for detecting active phages. This approach introduces a bias toward well-characterized phages, potentially missing less-studied or novel phages (false negatives) or detecting false positives by matching only highly conserved regions. Additionally, the INPHARED database is skewed toward lytic phages, representing 70% of its entries, which may limit the detection of other prophage types. Another limitation stems from the use of incomplete or inconsistent bacterial genome metadata, which can hinder the accurate classification and ecological distribution of prophages. Despite efforts to eliminate low-quality genomes, errors in the assembly of highly mosaic prophage regions could impact prophage detection and classification. The varying accuracy of sequencing technologies further complicates this issue, as differences in base calling could lead to incorrect assembly of prophage regions. In the experimental section, the use of mitomycin C as the sole prophage induction agent may have underestimated the number of inducible phages. Some prophages may have formed particles without producing detectable plaques. Moreover, the reliance on Illumina sequencing rather than long-read technologies limits the precision of active prophage identification. Therefore, we cannot be certain about the prophages in strain PS-237, where sequencing reads mapped to two prophages in the genome. Future work should incorporate long-read sequencing methods and more comprehensive databases to improve the accuracy of prophage detection.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents will be fulfilled by the lead contact, Anna Dragoš (anna.dragos@bf.uni-lj.si).

Materials availability

Primary materials generated during this study are available upon request through the lead contact.

Data and code availability

Bacteria and phage genome sequencing data have been deposited to NCBI under accession numbers PRJNA437002 and PRJNA1187528, respectively. All extracted prophages sequences arranged according to clustering performed in this work are available at the Zenodo repository under accession number Zenodo Data: https://doi.org/10.5281/zenodo.13834859. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors thank students Agnieszka Surowka, Mercedes Polajnar, Ema Florjančič, and Mia Babič for help with prophage induction, enrichment, and DNA extraction experiments. A.D. and V.A.F. were funded by the European Union (ERC, PHAGECONTROL, 101041421). The views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. I.M.-M., P.S., E.S., and A.D. were supported by the Slovenian Agency for Research and Innovation (grant no. P4-0116). I.M.-M., P.S., and E.S. were also supported by the Slovenian Agency for Research and Innovation (J4-4550). P.S. was additionally supported by J1-4411 and BI-US/22-24-161. A.T.K., M.L.S., and A.D. were supported by the Danish National Research Foundation (DNRF137) for the Center for Microbial Secondary Metabolites. J.K. was supported by the USDA NIFA Hatch Appropriations under project PEN04853 and accession 7005519 and multistate project 4666. Sequencing of some genomes was funded by the CDC (BAA 75D301-21-C-12111). U.R. was supported by German Research Foundation (DFG) project 460129525.

AUTHOR CONTRIBUTIONS

A.D., P.S., I.M.-M., and R.H. conceived the project; V.A.F. performed the experiments; P.S., E.S., J.K., and U.R. performed genome sequencing and assembly analysis; A.D., M.L.S., E.S., V.A.F., and P.S. contributed to the methods; A.D. and P.S. produced the figures; A.D., P.S., M.L.Z., and I.M.-M. wrote the manuscript; and all authors contributed to the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Bacterial genomes
  - Identification of prophage-like elements and integration loci
  - Classification of phages into virus operational taxonomic units (vOTUs) and into functional categories
  - Host phylogeny and prophage repertoire similarity scores
  - Ecological distribution
  - Correlations and anticorrelations between the presence of prophage elements within a single host genome
  - Prophage induction and plaque assay
  - Phage DNA extraction and sequencing
  - Analysis of predicted prophage activity from illumina reads coverage
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.115197.

Received: June 28, 2024
Revised: September 27, 2024
Accepted: December 20, 2024

REFERENCES

1. Clokie, M.R., Millard, A.D., Letarov, A.V., and Heaphy, S. (2011). Phages in nature. Bacteriophage *1*, 31–45. https://doi.org/10.4161/BACT.1.1.14942.
2. Castillo, D., Kauffman, K., Hussain, F., Kalatzis, P., Rørbo, N., Polz, M.F., and Middelboe, M. (2018). Widespread distribution of prophage-encoded virulence factors in marine Vibrio communities. Sci. Rep. *8*, 9973. https://doi.org/10.1038/s41598-018-28326-9.
3. Rodríguez-Rubio, L., Haarmann, N., Schwidder, M., Muniesa, M., and Schmidt, H. (2021). Bacteriophages of Shiga toxin-producing *Escherichia coli* and their contribution to pathogenicity. Pathogens *10*, 404. https://doi.org/10.3390/PATHOGENS10040404.
4. Butala, M., and Dragoš, A. (2022). Unique relationships between phages and endospore-forming hosts. Trends Microbiol. *31*, 498–510. https://doi.org/10.1016/J.TIM.2022.11.009.
5. Dragoš, A., Andersen, A.J.C., Lozano-Andrade, C.N., Kempen, P.J., Kovács, Á.T., and Strube, M.L. (2021). Phages carry interbacterial weapons encoded by biosynthetic gene clusters. Curr. Biol. *31*, 3479–3489. https://doi.org/10.1016/j.cub.2021.05.046.
6. Owen, S.V., Wenner, N., Dulberger, C.L., Rodwell, E.V., Bowers-Barnard, A., Quinones-Olvera, N., Rigden, D.J., Rubin, E.J., Garner, E.C., Baym, M., and Hinton, J.C.D. (2021). Prophages encode phage-defense systems with cognate self-immunity. Cell Host Microbe *29*, 1620–1633.e8. https://doi.org/10.1016/J.CHOM.2021.09.002.
7. Papaianni, M., Cuomo, P., Fulgione, A., Albanese, D., Gallo, M., Paris, D., Motta, A., Iannelli, D., and Capparelli, R. (2020). Bacteriophages promote metabolic changes in bacteria biofilm. Microorganisms *8*, 480. https://doi.org/10.3390/MICROORGANISMS8040480.
8. De Smet, J., Zimmermann, M., Kogadeeva, M., Ceyssens, P.J., Vermaelen, W., Blasdel, B., Bin Jang, H., Sauer, U., and Lavigne, R. (2016). High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. ISME J. *10*, 1823–1835. https://doi.org/10.1038/ismej.2016.3.
9. Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., et al. (2017). Communication between viruses guides lysis-lysogeny decisions. Nature *541*, 488–493. https://doi.org/10.1038/nature21049.
10. Bernard, C., Li, Y., Lopez, P., and Bapteste, E. (2021). Beyond arbitrium: identification of a second communication system in Bacillus phage phi3T that may regulate host defense mechanisms. ISME J. *15*, 545–549. https://doi.org/10.1038/s41396-020-00795-9.
11. Aframian, N., Omer Bendori, S., Kabel, S., Guler, P., Stokar-Avihail, A., Manor, E., Msaeed, K., Lipsman, V., Grinberg, I., Mahagna, A., and Eldar, A. (2021). Dormant phages communicate via arbitrium to control exit from lysogeny. Nat. Microbiol. *7*, 145–153. https://doi.org/10.1038/s41564-021-01008-5.
12. Abedon, S.,T. (2008). Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses (Cambridge University Press). https://doi.org/10.1017/CBO9780511541483.
13. Wu, C.J., Hsueh, P.R., and Ko, W.C. (2011). A new health threat in Europe: Shiga toxin–producing *Escherichia coli* O104:H4 infections. J. Microbiol. Immunol. Infect. *44*, 390–393. https://doi.org/10.1016/J.JMII.2011.07.001.
14. Khalil, R.K.S., Skinner, C., Patfield, S., and He, X. (2016). Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic *Enterobacter* and *Escherichia coli* strains. Pathog. Dis. *74*, ftw037. https://doi.org/10.1093/FEMSPD/FTW037.
15. Le Page, D.P., Metcalf, J.A., Bordenstein, S.R., On, J., Perlmutter, J.I., Shropshire, J.D., Layton, E.M., Funkhouser-Jones, L.J., Beckmann, J.F., and Bordenstein, S.R. (2017). Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. Nature *543*, 243–247. https://doi.org/10.1038/NATURE21391.
16. López-Leal, G., Camelo-Valera, L.C., Hurtado-Ramírez, J.M., Verleyen, J., Castillo-Ramírez, S., and Reyes-Muñoz, A. (2022). Mining of thousands of prokaryotic genomes reveals high abundance of prophages with a strictly narrow host range. mSystems *7*, e00326-22. https://doi.org/10.1128/MSYSTEMS.00326-22.
17. Kim, M.S., and Bae, J.W. (2018). Lysogeny is prevalent and widely distributed in the murine gut microbiota. ISME J. *12*, 1127–1141. https://doi.org/10.1038/s41396-018-0061-9.
18. Bobay, L.M., Rocha, E.P.C., and Touchon, M. (2013). The adaptation of temperate bacteriophages to their host genomes. Mol. Biol. Evol. *30*, 737–751. https://doi.org/10.1093/MOLBEV/MSS279.
19. Castillo, D., Andersen, N., Kalatzis, P.G., and Middelboe, M. (2019). Large phenotypic and genetic diversity of prophages induced from the fish pathogen vibrio anguillarum. Viruses *11*, 983. https://doi.org/10.3390/v11110983.
20. McLoon, A.L., Guttenplan, S.B., Kearns, D.B., Kolter, R., and Losick, R. (2011). Tracing the domestication of a biofilm-forming bacterium. J. Bacteriol. *193*, 2027–2034. https://doi.org/10.1128/JB.01542-10.

21. Gallegos-Monterrosa, R., Mhatre, E., and Kovács, Á.T. (2016). Specific *Bacillus subtilis* 168 variants form biofilms on nutrient-rich medium. Microbiology *162*, 1922–1932. https://doi.org/10.1099/mic.0.000371.

22. Browning, D.F., Hobman, J.L., and Busby, S.J.W. (2023). Laboratory strains of *Escherichia coli* K-12: things are seldom what they seem. Microb. Genom. *9*, mgen000922. https://doi.org/10.1099/MGEN.0.000922.

23. Klockgether, J., Munder, A., Neugebauer, J., Davenport, C.F., Stanke, F., Larbig, K.D., Heeb, S., Schöck, U., Pohl, T.M., Wiehlmann, L., and Tümmler, B. (2010). Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. J. Bacteriol. *192*, 1113–1121. https://doi.org/10.1128/JB.01515-09.

24. Hohmann, H.-P., Dijl, J.M. van, Krishnappa, L., and Prágai, Z. (2017). Host Organisms: *Bacillus subtilis*. Ind. Biotechnol. *1*, 221–297. https://doi.org/10.1002/9783527807796.CH7.

25. Blake, C., Christensen, M.N., and Kovacs, A.T. (2021). Molecular Aspects of Plant Growth Promotion and Protection by *Bacillus Subtilis* (American Phytopathological Society). https://doi.org/10.1094/MPMI-08-20-0225-CR.

26. Erega, A., Stefanie, P., Dogsa, I., Danevčič, T., Simunovic, K., Klančnik, A., Možina, S.S., and Mulec, I.M. (2021). Bacillaene mediates the inhibitory effect of *Bacillus subtilis* on *Campylobacter jejuni* biofilms. Appl. Environ. Microbiol. *87*, e02955-20. https://doi.org/10.1128/AEM.02955-20.

27. Podnar, E., Erega, A., Danevčič, T., Kovačec, E., Lories, B., Steenackers, H., and Mandic-Mulec, I. (2022). Nutrient availability and biofilm polysaccharide shape the bacillaene-dependent antagonism of *Bacillus subtilis* against *Salmonella* Typhimurium. Microbiol. Spectr. *10*, e01836-22. https://doi.org/10.1128/SPECTRUM.01836-22.

28. Šimunović, K., Stefanic, P., Klančnik, A., Erega, A., Mandic Mulec, I., and Smole Možina, S. (2022). *Bacillus subtilis* PS-216 antagonistic activities against *Campylobacter jejuni* NCTC 11168 are modulated by temperature, oxygen, and growth medium. Microorganisms *10*, 289. https://doi.org/10.3390/MICROORGANISMS10020289/S1.

29. Guo, M., Wu, F., Hao, G., Qi, Q., Li, R., Li, N., Wei, L., and Chai, T. (2017). *Bacillus subtilis* improves immunity and disease resistance in rabbits. Front. Immunol. *8*, 354. https://doi.org/10.3389/fimmu.2017.00354.

30. Jones, S.E., Paynich, M.L., Kearns, D.B., and Knight, K.L. (2014). Protection from intestinal inflammation by bacterial exopolysaccharides. J. Immunol. *192*, 4813–4820. https://doi.org/10.4049/jimmunol.1303369.

31. Donato, V., Ayala, F.R., Cogliati, S., Bauman, C., Costa, J.G., Leñini, C., and Grau, R. (2017). *Bacillus subtilis* biofilm extends *Caenorhabditis elegans* longevity through downregulation of the insulin-like signalling pathway. Nat. Commun. *8*, 14332. https://doi.org/10.1038/ncomms14332.

32. Goodarzi Boroojeni, F., Vahjen, W., Männer, K., Blanch, A., Sandvang, D., and Zentek, J. (2018). *Bacillus subtilis* in broiler diets with different levels of energy and protein. Poultry Sci. *97*, 3967–3976. https://doi.org/10.3382/ps/pey265.

33. Setlow, P. (2006). Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. J. Appl. Microbiol. *101*, 514–525. https://doi.org/10.1111/J.1365-2672.2005.02736.X.

34. Kellogg, C.A., and Griffin, D.W. (2006). Aerobiology and the global transport of desert dust. Trends Ecol. Evol. *21*, 638–644. https://doi.org/10.1016/J.TREE.2006.07.004.

35. Griffin, D.W. (2007). Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. Clin. Microbiol. Rev. *20*, 459–477. https://doi.org/10.1128/CMR.00039-06.

36. Štefanič, P. (2020). Environment Shapes the Intra-species Diversity of *Bacillus subtilis* Isolates. Microb. Ecol. *79*, 853–864. https://doi.org/10.1007/S00248-019-01455-Y.

37. Martin, M., Dragoš, A., Hölscher, T., Maróti, G., Bálint, B., Westermann, M., and Kovács, Á.T. (2017). De novo evolved interference competition promotes the spread of biofilm defectors. Nat. Commun. *8*, 15127. https://doi.org/10.1038/ncomms15127.

38. Sanchez-Vizuete, P., Le Coq, D., Bridier, A., Herry, J.-M., Aymerich, S., and Briandet, R. (2015). Identification of *ypqP* as a new *Bacillus subtilis* biofilm determinant that mediates the protection of *Staphylococcus aureus* against antimicrobial agents in mixed-species communities. Appl. Environ. Microbiol. *81*, 109–118. https://doi.org/10.1128/AEM.02473-14.

39. Dragoš, A., Priyadarshini, B., Hasan, Z., Strube, M.L., Kempen, P.J., Maróti, G., Kaspar, C., Bose, B., Burton, B.M., Bischofs, I.B., et al. (2020). Pervasive prophage recombination occurs during evolution of spore-forming Bacilli. ISME J., 1–15. https://doi.org/10.1038/s41396-020-00854-1.

40. Johnson, C.M., Harden, M.M., and Grossman, A.D. (2022). Interactions between mobile genetic elements: An anti-phage gene in an integrative and conjugative element protects host cells from predation by a temperate bacteriophage. PLoS Genet. *18*, e1010065. https://doi.org/10.1371/JOURNAL.PGEN.1010065.

41. Konkol, M.A., Blair, K.M., and Kearns, D.B. (2013). Plasmid-encoded ComI inhibits competence in the ancestral 3610 strain of *Bacillus subtilis*. J. Bacteriol. *195*, 4085–4093. https://doi.org/10.1128/JB.00696-13.

42. Mcveigh, R.R., and Yasbin, R.E. (1996). Phenotypic differentiation of "smart" versus "naive" bacteriophages of *Bacillus subtilis*. J. Bacteriol. *178*, 3399–3401. https://doi.org/10.1128/JB.178.11.3399-3401.1996.

43. Anderson, L.M., and Bott, K.F. (1985). DNA packaging by the Bacillus subtilis defective bacteriophage PBSX. J. Virol. *54*, 773–780. https://doi.org/10.1128/JVI.54.3.773-780.1985.

44. Kimura, T., Amaya, Y., Kobayashi, K., Ogasawara, N., and Sato, T. (2010). Repression of *sigK* intervening (skin) element gene expression by the CI-like protein SknR and effect of SknR depletion on growth of *Bacillus subtilis* cells. J. Bacteriol. *192*, 6209–6216. https://doi.org/10.1128/JB.00625-10.

45. Štefanič, P., and Mandić-Mulec, I. (2009). Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. J. Bacteriol. *191*, 1756–1764. https://doi.org/10.1128/JB.01290-08.

46. Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. *37*, 632–639. https://doi.org/10.1038/s41587-019-0100-8.

47. Cook, R., Brown, N., Redgwell, T., Rihtman, B., Barnes, M., Clokie, M., Stekel, D.J., Hobman, J., Jones, M.A., and Millard, A. (2021). INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of cultured phage genomes. PHAGE (New Rochelle, N.Y.) *2*, 214–223. https://doi.org/10.1089/PHAGE.2021.0007.

48. Dahlman, S., Avellaneda-Franco, L., Kett, C., Subedi, D., Young, R.B., Gould, J.A., Rutten, E.L., Gulliver, E.L., Turkington, C.J., Nezam-Abadi, N., et al. (2023). Temperate gut phages are prevalent, diverse, and predominantly inactive. Preprint at bioRxiv, 553642. https://doi.org/10.1101/2023.08.17.553642.

49. Abraha, H.B., Choi, Y., Hyun, W., Lee, J.W., Kang, H.S., So, M.S., Shin, D., Jung, J.H., Sbhatu, D.B., and Kim, K.P. (2022). Prevalence, diversity and UV-light inducibility potential of prophages in *Bacillus subtilis* and their possible roles in host properties. Viruses *14*, 483. https://doi.org/10.3390/V14030483.

50. Frazão, N., Sousa, A., Lässig, M., and Gordo, I. (2019). Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. Proc. Natl. Acad. Sci. USA *116*, 17906–17915. https://doi.org/10.1073/pnas.1906958116.

51. Lato, D.F., and Golding, G.B. (2020). Spatial patterns of gene expression in bacterial genomes. J. Mol. Evol. *88*, 510–520. https://doi.org/10.1007/S00239-020-09951-3.

52. Bobay, L.-M., Touchon, M., and Rocha, E.P.C. (2014). Pervasive domestication of defective prophages by bacteria. Proc. Natl. Acad. Sci. USA *111*, 12127–12132. https://doi.org/10.1073/pnas.1405336111.

53. Westers, H., Dorenbos, R., Van Dijl, J.M., Kabel, J., Flanagan, T., Devine, K.M., Jude, F., Séror, S.J., Beekman, A.C., Darmon, E., et al. (2003). Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol. Biol. Evol. *20*, 2076–2090. https://doi.org/10.1093/molbev/msg219.

54. Toyofuku, M., Cárcamo-Oyarce, G., Yamamoto, T., Eisenstein, F., Hsiao, C.C., Kurosawa, M., Gademann, K., Pilhofer, M., Nomura, N., and Eberl, L. (2017). Prophage-triggered membrane vesicle formation through peptidoglycan damage in *Bacillus subtilis*. Nat. Commun. *81*, 1–10. https://doi.org/10.1038/s41467-017-00492-w.

55. Shingaki, R., Kasahara, Y., Inoue, T., Kokeguchi, S., and Fukui, K. (2003). Chromosome DNA fragmentation and excretion caused by defective prophage gene expression in the early-exponential-phase culture of *Bacillus subtilis*. Can. J. Microbiol. *49*, 313–325. https://doi.org/10.1139/W03-041.

56. Refardt, D. (2011). Within-host competition determines reproductive success of temperate bacteriophages. ISME J. *5*, 1451–1460. https://doi.org/10.1038/ISMEJ.2011.30.

57. Patel, P.H., and Maxwell, K.L. (2023). Prophages provide a rich source of antiphage defense systems. Curr. Opin. Microbiol. *73*, 102321. https://doi.org/10.1016/J.MIB.2023.102321.

58. Schmidt, A.K., Faith, D.R., and Secor, P.R. (2023). PICI thieves: Molecular piracy and cooperation. Cell Host Microbe *31*, 3–5. https://doi.org/10.1016/j.chom.2022.12.008.

59. Johnson, C.M., Harden, M.M., and Grossman, A.D. (2020). An integrative and conjugative element encodes an abortive infection system to protect host cells from predation by a bacteriophage. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.13.422588.

60. Boling, L., Cuevas, D.A., Grasis, J.A., Kang, H.S., Knowles, B., Levi, K., Maughan, H., McNair, K., Rojas, M.I., Sanchez, S.E., et al. (2020). Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome. Gut Microb. *11*, 721–734. https://doi.org/10.1080/19490976.2019.1701353.

61. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. *13*. https://doi.org/10.1371/journal.pcbi.1005595.

62. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. Microb. Genom. *3*, e000132. https://doi.org/10.1099/MGEN.0.000132/CITE/REFWORKS.

63. Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. *44*, W16–W21. https://doi.org/10.1093/nar/gkw387.

64. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011). PHAST: A Fast Phage Search Tool. Nucleic Acids Res. *39*, W347–W352. https://doi.org/10.1093/nar/gkr485.

65. Bolduc, B., Jang, H.B., Doulcier, G., You, Z.Q., Roux, S., and Sullivan, M.B. (2017). vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ *5*, e3243. https://doi.org/10.7717/PEERJ.3243/SUPP-9.

66. Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. *11*. https://doi.org/10.1186/1471-2105-11-119/TABLES/5.

67. Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods *9*, 471–472. https://doi.org/10.1038/nmeth.1938.

68. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/BIOINFORMATICS/BTU153.

69. Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics *31*, 3691–3693. https://doi.org/10.1093/BIOINFORMATICS/BTV421.

70. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. *26*, 1641–1650. https://doi.org/10.1093/MOLBEV/MSP077.

71. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

72. Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. *46*, W537–W544. https://doi.org/10.1093/NAR/GKY379.

73. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature (London, U. K.) *390*, 249–256. https://doi.org/10.1038/36786.

74. Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., and Grossman, A.D. (2005). Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proc. Natl. Acad. Sci. USA *102*, 12554–12559. https://doi.org/10.1073/pnas.0505835102.

75. Wood, H.E., Dawson, M.T., Devine, K.M., and McConnell, D.J. (1990). Characterization of PBSX, a defective prophage of *Bacillus subtilis*. J. Bacteriol. *172*, 2667–2674. https://doi.org/10.1128/JB.172.5.2667-2674.1990.

76. Kunkel, B., Losick, R., and Stragier, P. (1990). The *Bacillus subtilis* gene for the developmental transcription factor σ(K) is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. Genes Dev. *4*, 525–535. https://doi.org/10.1101/gad.4.4.525.

77. Abe, K., Kawano, Y., Iwamoto, K., Arai, K., Maruyama, Y., Eichenberger, P., and Sato, T. (2014). Developmentally-regulated excision of the SPβ prophage reconstitutes a gene required for spore envelope maturation in *Bacillus subtilis*. PLoS Genet. *10*, e1004636. https://doi.org/10.1371/journal.pgen.1004636.

78. Burton, A.T., and Kearns, D.B. (2020). The large pBS32/pLS32 plasmid of ancestral *Bacillus subtilis*. J. Bacteriol. *202*, e00290-20. https://doi.org/10.1128/JB.00290-20.

79. Zeigler, D.R. (2013). Complete genome sequence of *Bacillus subtilis* phage ϕ105. Genome Announc. *1*, 641–654. https://doi.org/10.1128/GENOMEA.00641-13.

80. Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F.M., and Mandic-Mulec, I. (2012). The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. Environ. Microbiol. *14*, 1378–1389. https://doi.org/10.1111/j.1462-2920.2012.02717.x.

81. Stefanic, P., Kraigher, B., Lyons, N.A., Kolter, R., and Mandic-Mulec, I. (2015). Kin discrimination between sympatric *Bacillus subtilis* isolates. Proc. Natl. Acad. Sci. USA *112*, 14042–14047.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
B. subtilis Δ6	Westers et al.[53]	CP015975.1
PS-11	Štefanič et al.[45]	SAMN08637079
PS-13	Štefanič et al.[45]	SAMN08637080
PS-15	Štefanič et al.[45]	SAMN08637081
PS-18	Štefanič et al.[45]	SAMN08637082
PS-20	Štefanič et al.[45]	SAMN08637083
PS-24	Štefanič et al.[45]	SAMN08637084
PS-25	Štefanič et al.[45]	SAMN08637085
PS-31	Štefanič et al.[45]	SAMN08637086
PS-52	Štefanič et al.[45]	SAMN08637087
PS-53	Štefanič et al.[45]	SAMN08637088
PS-65	Štefanič et al.[45]	SAMN08637089
PS-68	Štefanič et al.[45]	SAMN08637090
PS-108	Štefanič et al.[45]	SAMN08637091
PS-160	Štefanič et al.[45]	SAMN08637092
PS-196	Štefanič et al.[45]	SAMN08637093
PS-209	Štefanič et al.[45]	SAMN08637094
PS-210	Štefanič et al.[45]	SAMN08637095
PS-216	Štefanič et al.[45]	SAMN08637096
PS-217	Štefanič et al.[45]	SAMN08637097
PS-218	Štefanič et al.[45]	SAMN08637098
PS-261	Štefanič et al.[45]	SAMN08637099
PS-14	Štefanič et al.[45]	SAMN36922878
PS-30	Štefanič et al.[45]	SAMN36922879
PS-51	Štefanič et al.[45]	SAMN36922880
PS-54	This work*	SAMN36922881
PS-55	Štefanič et al.[45]	SAMN36922882
PS-64	Štefanič et al.[45]	SAMN36922883
PS-93	Štefanič et al.[45]	SAMN36922884
PS-95	Štefanič et al.[45]	SAMN36922885
PS-96	Štefanič et al.[45]	SAMN36922886
PS-109	Štefanič et al.[45]	SAMN36922887
PS-119	Štefanič et al.[45]	SAMN36922888
PS-130	Štefanič et al.[45]	SAMN36922889
PS-131	Štefanič et al.[45]	SAMN36922890
PS-149	Štefanič et al.[45]	SAMN36922891
PS-168	Štefanič et al.[45]	SAMN36922892
PS-194	Štefanič et al.[45]	SAMN36922893
PS-233	Štefanič et al.[45]	SAMN36922894
PS-237	Štefanič et al.[45]	SAMN36922895
PS-263	Štefanič et al.[45]	SAMN36922896
Chemicals, peptides, and recombinant proteins		
Mitomycin C	Apollo Scientific, United Kingdom	CAS 50-07-7

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PEG-8000	Merck & Co., Inc, New Jersey, USA	CAS 25322-68-3
Critical commercial assays		
Genomic Mini AX Phage kit	A&A Biotechnology, Poland	011–20
MiSeq Reagent Kit v2	Illumina, Inc., California, USA	MS-102-2001
NEBNext®Ultra™ II DNA Library Prep Kit	New England Biolabs, Inc., Massachusetts, USA	E7645S
Deposited data		
Bacillus subtilis prophages_clustered to vOTUs	This work	[Zenodo]: [10.5281/zenodo.13834859]
Bacillus subtilis genome sequencing and assembly	This work	[SRA]: [PRJNA437002]
Phage genome sequencing data	This work	[SRA]: [PRJNA1187528]
Software and algorithms		
Unicycler v0.5.0 (on Galaxy platform)	Wick et al.[61] Wick et al.[62]	–
Snapgene	www.snapgene.com	–
Phaster	Arndt et al.[63] Zhou et al.[64]	–
vContact2 v0.11.3	Bin Jang et al.[46] Bolduc et al.[65]	–
Prodigal v2.6.3	Hyatt et al.[66]	–
ClusterOne v1.0	Nepusz et al.[67]	–
Prokka v1.14.5	Seemann et al.[68]	–
Roary v3.11.2	Page et al.[69]	–
FastTree v3.2.1	Price et al.[70]	–
R v.4.3.3	www.r-project.org	–
Python v.3.10.12	www.python.org	–
Open cage Geocoding API	www.opencagedata.com/api	–
Bbcl2fastq v2.17.1.14	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html	–
Geneious Prime v2024.0.2	www.geneious.com	–
BBMap v38.84	www.sourceforge.net/projects/bbmap/	–
MAFFT	Katoh et al.[71]	–

*this strain was previously misidentified, therefore originally a collection of 39 microscale isolates was published.[45] In this work the strain is included for the first time.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacillus subtilis, a Gram-positive bacterium, was used as a microbial model in this study. The strain utilized was a derivative of B. subtilis 168, a widely studied laboratory strain, as well as set of natural *B. subtilis* isolates from microscale, obtained at Sava Riverbank (Slovenia). Cultures were maintained under standard laboratory conditions, with growth in LB medium at 37°C with constant agitation. Institutional approval for microbial work was obtained, ensuring compliance with biosafety regulations. The influence of sex or gender is not applicable to this study due to the nature of the microbial model.

METHOD DETAILS

Bacterial genomes

A dataset of *B. subtilis* genomes consisting of complete genome sequences available on NCBI (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/, last accessed March 2023) was used along with complete genomes of local soil microscale isolates (See below; Data S1). Raw sequencing data for microscale isolates were obtained using various sequencing platforms including PacBio, Nanopore, and Illumina (Nextera and Truseq PCR-free libraries) from single isolate DNA extracts. Detailed information about genome sequences, libraries, and platforms used for each strain can be found in the NCBI database under Bioproject accession number PRJNA437002. Unicycler (accessed via Galaxy platform[72]) was used to hybrid assemble genomes using short and long reads.[61,62]

NCBI genomes were carefully curated based on metadata, which specifically meant that laboratory strains and their derivatives were excluded, while isolates with data for isolation sites and sources were retained. All genomes were oriented relative to *B. subtilis* 168 (AP011541.2), if necessary, using the 'linearize' function of SnapGene (www.snapgene.com). The final dataset contained 191 NCBI genomes and 40 genomes of soil microscale isolates.

Identification of prophage-like elements and integration loci

Prophage coordinates in bacterial genomes were determined using the web-based Phaster tool.[63,64] Functional categories assigned by the software (complete, questionable, incomplete) were ignored, and all predicted prophage elements were retrieved from bacterial genomes. Functional categories were ignored as we were interested in all prophage-like elements within bacterial genomes, regardless of their activity/domestication level. As genome orientations were corrected prior to prophage extraction, prophage coordinates were treated as integration loci. Prophages were retrieved from their parent genomes according to their coordinate using biopython and used for further analysis. The pipeline allows to retrieve selected sequences from genomes in high throughput, using sequence coordinates and genome files as an input. As references, a collection of sequences consisting of prophages and other MGEs was individually retrieved from *B. subtilis* 168 (NC_000964) and contained the following: prophage elements 1−6,[73] conjugative element ICEBs1,[74] defective phage PBSX,[75] *skin* element,[76] SPβ,[77] and plasmid pBS32 (KF365913.1) known to carry a defective prophage.[78] In addition, phi105 (NC_048631.1) was added as a reference, since this is a known temperate phage targeting *B. subtilis*.[79] Finally, *Bacillus* phage Yong1 (OP918669.1) was added as a reference after initial analysis of a substantial number of matches.

Classification of phages into virus operational taxonomic units (vOTUs) and into functional categories

All prophage elements predicted by Phaster (regardless of software-assigned functional category) were included in the classification. In order to organize predicted prophages into clusters, vContact2 (v0.11.3)[46,65] was used. In brief, prophage genes were predicted with Prodigal (v2.6.3),[66] translated into proteins, and these were used for reference-free phylogenetic clustering with vContact2, which builds a network based on the number of shared proteins between phages. ClusterOne (v1.0.)[67] was then used to infer clusters of related phages based on this network. Reference sequences of selected MGEs of *B. subtilis*, as mentioned above, were manually added to this set. These clusters were then considered as viral Operational Taxonomic Units (vOTUs) and used for further analysis. Prophages not confidently assigned to a vOTUs where categorized as either Outliers or Singletons: Outliers have too few shared genes to be part of a vOTU, whereas Singletons are unique and have no similarity to any other clusters. Due to their sparsity, members of each group were collapsed into either Outlier or Singletons for analysis. The entire dataset of prophage sequences grouped into vOTUs is available at Zenodo platform: https://zenodo.org/records/13834859, To further examine the taxonomy of these phages, each phage was aligned to a database of known phages. Here, we used discontiguous megablast (dc-megablast) to match them to the INPHARED phage database[47] of confirmed active phages, allowing for large gaps, which was necessary due to the high level of phage recombination. We considered a phage to match a reference if the coverage of the phage was at least 50% at $\geq$70% identity. Prophages from vOTUs that did not have a match within reference sequences were additionally blasted against the *B. subtilis* 168 genome without MGE sequences to check for potential matches to annotated *B. subtilis* chromosomal regions that do not carry functional prophages. To predict the functional categories of prophage groups (vOTUs), we used a combination of sequence coverage/similarity of experimentally proven active phages and of *B. subtilis* host chromosomal regions. Specifically, vOTUs sequences were compared with 1) the entire NCBI database, 2) the *B. subtilis* 168 genome [NC_000964.3] and 3) the *B. subtilis* 168 genome with artificially removed MGE (mobile genetic element) regions. Then, vOTUs were classified based on the following criteria: likely active prophages - if the majority of sequences within a vOTU matched functional reference phages that had been isolated as phage particles, and showed no, or limited coverage or similarity with host chromosomal regions; possibly active prophages - if a minority of sequences within a vOTU matched functional reference phages, or if the majority matched functional phages but had more than 50% coverage overlap with non-MGE *B. subtilis* chromosomal regions; prophages with known MGE functions - if the majority of sequences within vOTU matched known MGEs within *B. subtilis*, their functionality was inferred based on the known functions of those MGEs; likely cryptic prophages - if majority of sequecnes within vOTU had more than 30% coverage and over 88% identity to non-MGE chromosomal regions of *Bacillus subtilis* 168; unknown - if no matching reference was found and the prophages had less than 30% coverage with non-MGE chromosomal regions of *B. subtilis* 168. Outliers and singletons were excluded from this analysis.

Host phylogeny and prophage repertoire similarity scores

All bacterial host genomes were phylogenetically evaluated using pan-genome analysis. Specifically, bacterial genomes were annotated with PROKKA (v1.14.5)[68] using the *B. subtillis* 168 annotation as a reference, and the resulting annotations were used to build a core genome using ROARY (v3.11.2)[69] by considering genes to be core if they had >95% protein similarity in >99% of genomes. The resulting core gene set was then aligned with MAFFT and then used to construct a maximum likelihood phylogenetic tree with FastTree (v3.2.1)[70] using the generalized time reversible (GTR) substitution rate model. Phylogenetic clusters containing more than five *B. subtilis* subspecies *subtilis* strains were determined arbitrarily. Prophage repertoire similarity scores were calculated using a descriptive, rule-based scoring system. Specifically, a selected group of strains (e.g., members of the same phylogenetic cluster) was assessed for similarity in the presence or absence of prophages from each of 56 vOTUs, one by one, using a majority rule. If the prophage from a given vOTU was present in the majority of strains, the number of lysogens (strains carrying the prophage) was

counted. Conversely, if the prophage was present in the minority, the number of non-lysogens (strains without the prophage) was counted. This approach allowed the strains to score high in similarity either by consistently carrying or not carrying a given prophage. Next, the proportion of lysogens (or non-lysogens) was calculated by dividing their number by the total number of strains in the group, yielding the percentage of strains with (or without) the prophage of interest. A higher value indicated greater similarity among the strains with respect to the given prophage. This process was repeated for prophages from all vOTUs, and the results were averaged across all vOTUs to produce the overall prophage repertoire similarity score.

Ecological distribution

Information on strain geography and habitat was derived from NCBI metadata. Furthermore, isolation sites were limited to include the most abundant geographical locations (Data S2), and isolation habitats were divided into five main categories: soil/plant, food, waste/feces, water, and others. Host prophage cargo and isolation habitats of hosts were mapped onto a phylogenetic tree using R (v.4.3.3). The locations of strain isolation were mapped onto the world map using Python 3.10.12 and Open cage Geocoding API (Figure S9). Briefly, global strains were isolated from North America, Africa, Europe, and Asia. Local strains were isolated from two 1 cm^3 soil samples from the riverbank of Sava river in Slovenia[45] and have been previously described and phenotypically analyzed.[36,45,80,81]

Correlations and anticorrelations between the presence of prophage elements within a single host genome

To investigate if some phages are mutually exclusive/inclusive within the same genome, we noted the presence of each vOTU, as defined above, within each host genome, resulting in a binary presence/absence matrix. Logistic regression was then performed to statistically test whether or not vOTUs co-occurred within genomes (i.e., which vOTUs were mutually inclusive or conversely, mutually exclusive). In more detail, we here fit a logistic equation describing the relationship among pairs of prophages, i.e., the probability of a given vOTU being present in genome as a function of the presence of another vOTU, which is then tested using ANOVA with a c^2 distribution. A positive β then denotes an association between the two vOTUs, and vice versa, both of which can be turned into a odds-ratios by natural exponentiation.

Prophage induction and plaque assay

Bacterial strains were routinely cultivated in LB broth (LB-Lennox, Carl Roth; 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl). To assess phage activity in culture spent media, overnight cultures were initiated from −80°C glycerol stocks and cultivated overnight at 37°C with shaking at 200 rpm for ~16 h. Cultures were transferred to fresh LB medium at 1% inoculum and further cultivated at 37°C with shaking at 200 rpm. In mid-exponential phase (OD_{600} ~0.5), mitomycin C was added (1.5 μg/mL) to trigger prophage induction, followed by cultivation for another 4 h. Cells were pelleted by centrifugation (8000 g, 10 min), supernatants were filtered-sterilized, and diluted in order to obtain well-separated single plaques. Plaque assays were performed using *B. subtilis* Δ6 (ΔSPβ Δ*skin* ΔPBSX Δprophage 1 pkscat Δprophage 3)[53] as an indicator strain. A similar experiment, but without addition of mitomycin C, was performed to assess spontaneous prophage induction (SPI). The effect of mitomycin C on growth dynamics was also analyzed in a separate assay on a microtiter plate using two different concentrations (1.5 μg/mL and 0.1 μg/mL). The mitomycin C effect was quantified by comparing areas under the growth curves between controls (without addition of mitomycin C) and treated samples, and expressed as % of biomass reduction.

Phage DNA extraction and sequencing

Phages were propagated on soft agar to obtain at least 20mL of lysates with titer of 10^6 pfu/mL. Lysates were filter-sterilized and mixed at a 1:4 rate with PEG-8000 solution (PEG-8000 20%, 116 g/L NaCl). After overnight incubation at 4°C, the solutions were centrifuged for 60 min at 11000 rpm to obtain phage precipitates. The pellets were resuspended in 500μL of SM buffer (5.8 g/L NaCl, 0.96 g/L MgSO4, 6 g/L Tris-HCl, pH 7.5), followed by phage DNA extraction using Genomic Mini AX Phage kit (A&A Biotechnology, Poland), using modified protocol with nuclease incubation step extended to 60 min. Phage sequencing was performed by Illumina MiSeq instrument and a 2 × 250 nt paired-end chemistry (MiSeq Reagent Kit v2 (500-cycles). Primary data analysis (base-calling) was carried out with Bbcl2fastq^ software (v2.17.1.14, Illumina). *In vitro* fragment libraries were prepared using the NEBNextUltra II DNA Library Prep Kit for Illumina.

Analysis of predicted prophage activity from illumina reads coverage

For the sequenced genomes (*B. subtilis* PS-11, PS-24, PS-25, PS-233, PS-237, PS-261), predicted prophage activity was assessed by visually analyzing coverage discrepancies between prophage sequences and the rest of the genome. This involved mapping Illumina reads to their corresponding reference genome in Geneious Prime (2024.0.2) using BBMap (v38.84) with a normal sensitivity setting, and comparing the average prophage sequence coverage with the average coverage of the rest of the genome. Enrichment of prophage regions indicates more phage DNA and lytic phages.

QUANTIFICATION AND STATISTICAL ANALYSIS

Information about statistical methods used in bioinformatics analysis can be found in the above 'Method details' section, next to description of specific analysis. Statistics related to experiments, are denoted in figure legends. Briefly, means of at least 6 biological

replicates are shown on the graphs, with error bars showing standard errors. If boxplots are plotted, all datapoints are shown on the graph. The Pearson correlation coefficient was used to estimate the correlation between the number of prophages and the magnitude of the stress response.

ADDITIONAL RESOURCES

None. All relevant links have been provided in 'Method details' or 'Key Resource Table'.