Phages carry interbacterial weapons encoded by biosynthetic gene clusters

Dragoš, Anna; Andersen, Aaron J.C.; Lozano-Andrade, Carlos N.; Kempen, Paul J.; Kovács, Ákos T.; Strube, Mikael Lenz

Published in:
Current Biology

Link to article, DOI:
10.1016/j.cub.2021.05.046

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Dragoš, A., Andersen, A. J. C., Lozano-Andrade, C. N., Kempen, P. J., Kovács, Á. T., & Strube, M. L. (2021). Phages carry interbacterial weapons encoded by biosynthetic gene clusters. Current Biology, 31(16), 3479-3489. https://doi.org/10.1016/j.cub.2021.05.046

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Phages carry interbacterial weapons encoded by biosynthetic gene clusters

Highlights

- BGCs can be found in temperate phages infecting certain human-associated bacteria
- Almost all phage BGCs are bacteriocins, e.g., sublancin in Bacillus phage SPβ
- A BGC within SPβ provides competitive fitness advantage for the lysogenized host

Authors

Anna Dragoš, Aaron J.C. Andersen, Carlos N. Lozano-Andrade, Paul J. Kempen, Ákos T. Kovács, Mikael Lenz Strube

Correspondence

anna.dragos@bf.uni-lj.si (A.D.), milst@dtu.dk (M.L.S.)

In brief

Bacteria carry biosynthetic gene clusters (BGCs) encoding for biosynthesis of specialized metabolites. Dragos et al. perform high-throughput BGC mining within all sequenced phage and prophage genomes. Bioinformatics analysis, supported by lab experiments, suggests that phages may equip bacteria with weapons, hence providing benefits for both.
Phages carry interbacterial weapons encoded by biosynthetic gene clusters

Anna Dragoš,1,4,5,10,* Aaron J.C. Andersen,1 Carlos N. Lozano-Andrade,1,6 Paul J. Kempen,2,3,7 Ákos T. Kovács,1,8 and Mikael Lenz Strube1,9,*

1Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads bldg. 221, DK-2800 Kgs Lyngby, Denmark
2Department of Health Technology, Technical University of Denmark, Produktionsstorvet bldg. 423, DK-2800 Kgs Lyngby, Denmark
3National Center for Nano Fabrication and Characterization, Technical University of Denmark, Fysikvej bldg. 307, DK-2800 Kgs Lyngby, Denmark
4Present address: Biotechnical faculty, University of Ljubljana, Večna pot 111, 1000 Ljubljana, Slovenia
5Twitter: @Anna_Dragos
6Twitter: @neftaly1
7Twitter: @pkempen
8Twitter: @EvolvedBiofilm
9Twitter: @Mikael_lenz
10Lead contact
*Correspondence: anna.dragos@bf.uni-lj.si (A.D.), milst@dtu.dk (M.L.S.)
https://doi.org/10.1016/j.cub.2021.05.046

SUMMARY

Bacteria produce diverse specialized metabolites that mediate ecological interactions and serve as a rich source of industrially relevant natural products. Biosynthetic pathways for these metabolites are encoded by organized groups of genes called biosynthetic gene clusters (BGCs). Understanding the natural function and distribution of BGCs provides insight into the mechanisms through which microorganisms interact and compete. Further, understanding BGCs is extremely important for biocontrol and the mining of new bioactivities. Here, we investigated phage-encoded BGCs (pBGCs), challenging the relationship between phage origin and BGC structure and function. The results demonstrated that pBGCs are rare, and they predominantly reside within temperate phages infecting commensal or pathogenic bacterial hosts. Further, the vast majority of pBGCs were found to encode for bacteriocins. Using the soil- and gut-associated bacterium *Bacillus subtilis*, we experimentally demonstrated how a temperate phage equips a bacterium with a fully functional BGC, providing a clear competitive fitness advantage over the ancestor. Moreover, we demonstrated a similar transfer of the same phage in prophage form. Finally, using genetic and genomic comparisons, a strong association between pBGC type and phage host range was revealed. These findings suggest that bacteriocins are encoded in temperate phages of a few commensal bacterial genera. In these cases, lysogenic conversion provides an evolutionary benefit to the infected host and, hence, to the phage itself. This study is an important step toward understanding the natural role of bacterial compounds encoded by BGCs, the mechanisms driving their horizontal transfer, and the sometimes mutualistic relationship between bacteria and temperate phages.

INTRODUCTION

Alongside molecules that are directly involved in basic metabolism, microbes produce a plethora of specialized compounds, or metabolites. These specialized metabolites mediate microbial ecological interactions, but also supply medical1 and biotechnological industries2 with novel biological activities. Despite a long-standing tradition of industrial exploitation, the natural role of most specialized metabolites remains largely unknown.3 However, there is evidence that some of these metabolites act as chemical weapons in competitive interactions.4,5

The enzymatic pathways that are required to produce specialized metabolites are encoded by biosynthetic gene clusters (BGCs). A single BGC encodes for a single type of specialized metabolite, but multiple structural derivatives of this metabolite can occasionally be synthesized.6 Although BGCs differ both structurally and functionally, major groups with similar biosynthetic functions have been found. For example, BGCs encode polyketide synthases (PKS BGCs) and non-ribosomal peptide synthetases (NRPSs), which are extremely large (10–100 kb), multi-enzyme complexes that synthesize peptides with antibiotic properties and signaling, immunosuppressive, and biosurfactant activities.7,8 Ribosomal peptide natural products (RPNPs or RiPPs) are mostly represented by bacteriocins. Bacteriocins are recognized for their role in interbacterial competition9 and their BGCs are smaller than PKS and NRPS BGCs.

The presence and expression of a BGC within a bacterial genome can determine the pathogenicity of the bacterium10,11 and its evolutionary success in an ecological niche.12 In addition, more recently, the importance of specialized metabolites in...
human microbiomes for the balance between health and disease has become increasingly apparent.\(^{13}\) With the exception of certain intracellular pathogens that have retained BGCs despite reductive genome evolution,\(^{14}\) most bacteria are locked in a perpetual state of chemical warfare with other bacteria sharing the same niche.\(^{15}\) Therefore, given the antibiotic activity of many specialized metabolites, it can be assumed that specialized metabolites are crucial for the assembly of microbiomes. It has been proposed that competitive interactions (such as those between community members harboring BGCs that encode for antagonistic compounds) are widespread within microbiomes, and they can be linked to community stability.\(^{16}\) Bacteriocins are particularly interesting in this context as they are uniformly bactericidal; other classes of specialized metabolites may have entirely divergent effects depending on the concentration.\(^{17}\)

Progress in metagenomics, combined with excellent computational BGC prediction tools,\(^{18}\) has allowed us to characterize the genetic potential of microbiomes to produce specialized metabolites to identify the main bacteria hosting BGCs.\(^{17,18}\) Moreover, the abundance of BGCs in soil microbiomes has been found to vary depending on vegetation and soil depth.\(^{18}\) In the healthy human gut microbiome, the Bacteroides genus contains many BGCs, with those encoding for saccharide biosynthesis being the most abundant. Meanwhile, common gut-associated genera, such as like Escherichia, Lactobacillus, Haemophilus, and Enterococcus, appear to carry a limited number of BGCs.\(^{17}\) In the human oral microbiome, Firmicutes are the main carriers of BGCs, with oligosaccharide-encoding clusters and RfPPs as the major BGC types.\(^{15}\) Current interdisciplinary approaches allow us to identify the key BGC types and their producers. However, the natural functions of specialized metabolites and the eco-evolutionary forces driving their distribution and prevalence in microbial communities remain poorly understood.

In certain bacterial genera, BGCs are commonly associated with mobile genetic elements (MGEs)\(^{19-21}\) within the so-called flexible genome. A major component of MGEs in bacteria are phages, which are considered to be the most abundant biological agents on Earth. Lytic phages hijack the DNA and protein synthesis machinery of the host, so that they rapidly multiply and eventually lyse the host cell to release progeny virions. Temperate phages can integrate their DNA into the host chromosome using specific or unspecific attachment sites (att) and replicate with the host bacterium as a prophage.

Phages can encode virulence factors,\(^{21,22}\) toxins,\(^{23}\) and other compounds that are potentially valuable for the host. For instance, the SP\(\beta\) prophage present in some strains of Bacillus subtilis, which is a soil, plant, and gut bacterium,\(^{24}\) encodes for the S-linked glycocin, sublancin.\(^{25,26}\) The sublancin cluster (5 kb) contains a precursor peptide (SunA), posttranslational modification enzymes (SunS, BdbA, and BdB), a self-immunity protein (SunI), and a transporter (SunT) that ensures the export of the bacteriocin into the extracellular space.\(^{26}\) Sublancin presumably enters cells using the glucose-specific phosphoenolpyruvate:sugar phosphotransferase system (PTS).\(^{28}\) Upon entry, sublancin negatively affects DNA replication, transcription, and RNA translation.\(^{29}\) Bacterial mutants that are deprived of SP\(\beta\), the entire sublancin cluster, or just the immunity gene sunI are inhibited by isogenic strains that carry intact SP\(\beta\).\(^{30}\) This prophage can be found in closely related but geographically dispersed B. subtilis isolates, and it always occupies the same position in the chromosome.\(^{24}\) The contribution of sublancin to the ecological success of SP\(\beta\) remains an open question. It is not known whether other phages or prophages carry BGCs, or whether the products of such BGCs may share structural or functional features. This knowledge would facilitate the mining and utilization of new specialized metabolites.

The aim of this study was to explore all high-quality phage and prophage genomes for the presence of BGCs using a large-scale bioinformatics approach. Findings of this analysis, supported by experimental work, contribute significantly to understanding the ecological role of specialized metabolites and the sometimes mutualistic relationship between phages and their bacteria.

**RESULTS**

**BGCs are extremely rare within virion-derived genomes**

To assess the contribution of phages to the genetic pool of microbial specialized metabolites, we addressed the following fundamental question: can BGCs that encode for specialized metabolites be found in phage genomes? To this end, all complete phage genomes that were available in the PATRIC 3.6.2 database (10,063 in total) were subjected to BGC detection using antiSMASH 4. Only 0.07% of all phages (69 in total) were found to carry BGCs (Figure 1A). These phage-encoded BGCs are referred to as pBGCs (phage-encoded biosynthetic gene clusters) hereafter.

The vast majority (64/69) of the pBGCs were found in temperate phages (this lifestyle was confirmed experimentally for these phages in the literature). Only five pBGCs were carried by a lytic phage (lifestyle confirmed experimentally in the literature; Data S1). It was also observed that phages infecting certain bacterial genera were overrepresented among the pBGC hits, compared to relative distribution of phage hosts in the PATRIC database (Data S1; Figure S1A). These genera were Escherichia (represented solely by Escherichia coli species), Mannheimia, Enterobacter, and Shigella (Figures 2A and S1A).

It is important to note that the number of BGC hits per phage genome was restricted to 1 or none in all cases. This initial analysis suggested that pBGCs are extremely rare, and that they are more likely to reside within host-associated (temperate) phages that infect particular bacterial species.

**BGCs are more abundant in prophage elements than in virions**

As the first analysis suggested that pBGCs may be more common in temperate phages, the pBGCs encoded by prophage elements were investigated. As the pBGC-positive phage pool is largely overrepresented by Escherichia coli phages (Data S1), the aim was to analyze a more diverse set of (pro)phage genomes. Prophage regions were predicted and extracted from all the high-quality bacterial genomes available in the National Center for Biotechnology Information (NCBI) database. These regions were then subjected to BGC detection using antiSMASH 4 (Figure 1B). The majority of the analyzed bacterial genomes (67.5% of 15,184) was found to carry at least one prophage element. Only approximately 3% of these prophage regions (307 in total) contained a BGC (Figure 1B; Data S2), which
corresponded to approximately 3% of all complete bacterial genomes.

It was again observed that the abundance of pBGC-carrying prophage elements was overrepresented in certain host species, considering the relative abundance of these genera in the entire genome database (Data S2; Figure S1B). These particular hosts included three of the four genera that were also common hosts of pBGC-carrying virions. Specifically, nearly 90% of all pBGC-carrying prophages were associated with 11 genera of human/animal commensals and pathogens: Escherichia (all but one isolate belonged to E. coli), Mannheimia (all Mannheimia haemolytica), Bacillus, Pasteurella, Haemophilus, Lactobacillus, Enterococcus, Enterobacter, Listeria, Aeromonas, and Klebsiella. All of these genera, apart from Aeromonas and Klebsiella, seemed to be several-fold overrepresented within the pBGC dataset (Figures 2A and S1B; Data S2). However, not all representatives of these genera carried pBGCs. In addition, pBGC presence was clearly over-represented in the prophages of certain species. B. subtilis comprises a small proportion (20%) of all Bacillus genomes in NCBI, but it contributed to the majority (74%) of pBGC hits within the Bacillus genus. Lactobacillus brevis comprises only 4% of all Lactobacillus but the majority (67%) of pBGC hits within the Lactobacillus genus. Furthermore, Enterococcus faecalis comprises 22% of all enterococci, but over 90% of pBGC hits within the corresponding genus (Data S2; Figure S1B).

In the vast majority of pBGC-carrying bacteria, the prophage carried only one BGC. Notable exceptions to this included some isolates of M. haemolytica (8 isolates), E. coli (7 isolates), and L. brevis (1 isolate), which were found to carry 2 pBGCs. A single genome belonging to Alkaliphilus metalliredigenus carried three pBGCs, all of which were identical (Data S2).

Since the total numbers of phages were substantially higher in pBGC hosts (Figure 2B; Mann-Whitney U test, p = 2.2 × 10⁻¹⁶), it needed to be questioned whether prophages with BGC cargo are simply more abundant among species that are generally rich in BGCs. The number of BGCs encoded outside of the prophage regions (referred to as gBGCs for core-genome-encoded biosynthetic gene clusters) were thus assessed (Data S2). This revealed the opposite scenario; the total gBGC count was slightly, but significantly, lower in pBGC hosts (Figure 2B; Mann-Whitney U test, p = 0.0011). Eliminating the most abundant genus Escherichia from this analysis, however, removed the negative association between gBGCs and pBGCs (p = 0.8342). Meanwhile, the relationship between phage count and pBGC positivity still held true (p = 2.2 × 10⁻¹⁶).

Still, in certain species, such as Listeria monocytogenes, E. faecalis, and L. brevis, which have low numbers of gBGCs, phages appeared to represent a major, and sometimes even the only, source of BGCs (Figure 2B). These results confirm that, although relatively rare, pBGCs are most prevalent within temperate phages infecting specific commensal or pathogenic species, and whose bacterial hosts only harbor a low number of gBGCs.

The vast majority of prophage BGCs encode for bacteriocins

The next step was to investigate whether prophage BGCs encode for compounds that share structural or functional features. Indeed, according to the antiSMASH prediction, nearly all pBGCs were assigned to bacteriocins (99.3%), with some stratification onto glyocins (n = 22, 6.8%) and lantipeptides (n = 2, 0.6%; Table S4). To investigate how many of the antiSMASH-predicted pBGCs encoded for experimentally confirmed or studied bacteriocins, the core biosynthetic genes (as determined by antiSMASH) were compared with the BACTIBASE database. BACTIBASE contains bacteriocin peptide sequences collected from UniProt and scientific literature along with information on their mode of action, structure, and physicochemical properties.

The vast majority of prophage BGCs displayed poor matches in BACTIBASE; most had protein similarities and alignment lengths well below 60% (Figure 3). The pBGCs that displayed high similarity to experimentally confirmed bacteriocin BGCs encoded bacteriocins including encode: sublancin (22 hits in B. subtilis; 100% similarity), enterocin (5 hits in L. monocytogenes; >88% similarity), and carocin (Pectobacterium carotovorum; RefSeq: GCF_000294535.1; >98% similarity). One pBGC was found to be close in length, but highly dissimilar to linocin BGCs in Nocardia terpenica (RefSeq: GCF_0002568625.1; >56% similarity and >74% length) and E. coli (RefSeq: GCF_001901005.1 and RefSeq: GCF_900036075.1 (>2% similarity and 100% length).

Although the vast majority of pBGCs were predicted to encode for bacteriocins, several exceptions were also found. For instance, two pBGCs that were found in two
separate *Pseudomonas aeruginosa* genomes contained a partial NRPS-like cluster with a single AMP-binding domain, which according to BLASTP analysis to the NR database is likely to be part of the pyoverdine BGC. The two pBGCs also contained a siderophore receptor and a sigma-70 polymerase sigma factor (Table S2). Meanwhile, two pBGCs that were found within two *Mycobacterium* temperate phages encoded for ectoine, which may play a role in osmoadaptation. A pBGC that was found in the pseudo-lysogenic *Streptomyces* phage, ZL12, was annotated as a Type 3 PKS. In addition, a pBGC that was found in a lytic *Roseobacter* phage encoded for homo-serine lactone, which is a molecule involved in cell-to-cell communication (Table S2).

These results suggest that phages can be the source of different BGC types. However, the vast majority of pBGCs encode for bacteriocins, putative weapons for use in competition against closely related bacteria. Phages weaponize their hosts with BGCs through lysogeny

As pBGCs are more commonly found in temperate phages, it was hypothesized that they may play an important role during lysogeny. As the vast majority of pBGCs encode for bacteriocins, they could serve as biological weapons for use by the host bacterium when competing with its closely related neighbors. The pBGCs would thereby increase the fitness benefits associated with lysogeny for both the host and phage. This hypothesis was tested using a *B. subtilis* model system: a pBGC-carrying phage was transferred from one natural isolate to another and fitness benefits associated with the pBGC-phage infection were monitored (Figure 4A). The isolate MB8_B7 (RefSeq: GCF_009662215.1) served as the phage donor, as it is a natural host of an SPβ phage. The SPβ phage carries the BGC for sublancin, which is an antimicrobial glycocin (Figures 4A and S2A). The pBGC-negative isolate P9_B1 (RefSeq: GCF_009662455.1) was used as the receiver strain, as it lacks the SPβ phage in the genome. The two strains (MB8_B7 and P9_B1) are very closely related based on their core genome phylogeny and share all BGCs except for the sublancin pBGC.

The genome sequence of the MB8_B7 prophage was identical to the SPβ-phage found in the lab strain *B. subtilis* 168. An exception was that MB8_B7 does not contain the integrative and conjugative element, ICEBs1. ICEBs1 has been shown to interfere with the lytic cycle of SPβ. It was confirmed that deletion of SPβ from MB8_B7 provoked sensitivity toward its ancestor, the wild-type MB8_B7 (Figure 4A). Importantly, sublancin could be detected in the spent medium of MB8_B7, but not in that of its ΔSPβ derivative (Figures 4C, S2B, and S2C).

Next, SPβ virions were isolated from MB8_B7 culture and their identity confirmed via Nanopore sequencing (Figure S2D). The SPβ virions were then used to infect the naturally pBGC-free isolate, P9_B1, to create P9_B1SPβ (Figures 4A, S2E, and S2F). The phage viability was clearly evident from the plaque assay performed on a lawn of Δ6 strain, which is a derivative of *B. subtilis* laboratory strain 168 that is deprived of all prophage elements (Figure S2E).

As anticipated, P9_B1SPβ gained the ability to kill its P9_B1 ancestor (Figure 4A). P9_B1SPβ was also able to produce sublancin (Figures 4A–4C, S2B, and S2C). A similar transfer of antimicrobial activity from MB8_B7 (the natural SPβ host) to P9_B1 (the recipient) was also observed to occur spontaneously during co-cultivation (Figure 4B). Isolates of P9_B1 showing both antimicrobial activity and PCR-confirmed SPβ integration were detected after 24 h of exposure to MB8_B7 (Figure S2F). This suggests that initially SPβ-negative P9_B1 bacteria eventually become SPβ-positive. This partly explains the non-zero final abundance of P9_B1 seen in Figure 4D.

These results clearly demonstrated that the pBGC and its biological activity can readily be transferred between two strains via phage infection. Such transfer can occur spontaneously (without

---

Figure 2. Distribution of pBGC across genera

(A) Relative distribution of bacteria that serve as hosts to pBGC-carrying phages colored by genus. For color interpretation, see (B). Pink indicates *Shigella* genus. (B) Number of prophages in each genome versus number of genomic BGCs colored by genus. Genomes with only one pBGC are marked by triangles and genomes containing multiple pBGCs are marked by diamonds. Points have been jittered to avoid overlap. See also Figure S1 and Data S1 and S2.
addition of prophage-inducing agents) when the lysogen and the susceptible non-lysogen interact.

To further examine the benefits of pBGC acquisition, P9_B1 and P9_B1^{spm} were labeled with fluorescent reporters (STAR Methods; Figures S3A–S3D), and their growth was monitored in co-cultures that started with a 1:1 ratio.

Clearly, P9_B1^{spm} outcompeted P9_B1, regardless of the fluorescent marker used (Student’s t test, p = 0.694 × 10^{-13} and p = 0.160 × 10^{-7}, for P9_B1_{GFP} versus P9_B1^{spm}_{mKate} and P9_B1^{mKate} versus P9_B1^{spm}_{mKate}, respectively; Figures 4D and S4A). To confirm that the competitive advantage of P9_B1^{spm} was due to the pBGC (sublancin), we deleted the entire pBGC (sublancin production and immunity gene) or only the core biosynthetic gene sunA from P9_B1^{spm}. The resulting mutants were challenged against P9_B1 or P9_B1^{spm}. Both the P9_B1^{spm}_{sunA} and P9_B1^{spm}_{sunA} strains lost their competitive advantage over P9_B1 (Figures 4D, S4A, and S4B). In addition, P9_B1^{spm}_{sunA}, but not P9_B1^{spm}_{sunA}, was less competitive than P9_B1^{spm}. This was most likely due to the lack of immunity to sublancin (Student’s t test, p = 0.664 × 10^{-10} and p = 0.301 × 10^{-17}, for P9_B1_{GFP} versus P9_B1^{spm}_{mKate} and P9_B1^{mKate} versus P9_B1^{spm}_{mKate}, respectively; STAR Methods) (Figures 4D and S4B). As a control for the potential fitness effects of the fluorescent markers, the co-cultures of isogenic strains were also examined; the ratios remained the same as for the inoculated strains (Figure S4C).

Lysogeny by the pBGC-containing phage, SPβ, provides B. subtilis with a competitive advantage. Therefore, we further hypothesized that the bacterium could also benefit from carrying an attachment site (att), spsM, on the chromosome for the pBGC-containing phage on the chromosome. The spsM gene encodes for a polysaccharide synthase. Not all B. subtilis strains carry spsM, and the lack of this SPβ attachment gene would greatly reduce the frequency of SPβ integration into the host chromosome (Figure 5A). This would effectively render the spsM mutant less competitively fit compared to the wild type upon SPβ invasion. We designed a subsequent inoculation-based competition assay involving P9_B1 and P9_B1^{spsM}. The assay was performed without SPβ (control) and with a single episode of exposure to SPβ (Figure 5B). We started this assay from a control experiment, in which we showed similar susceptibility to lysis by SPβ, based on growth curves and plaque assays (Figure S5A).

In the absence of the phage, both strains could co-exist at comparable relative proportions for at least four rounds of cultivation (Figures 5A and S5B). However, a single exposure to SPβ at the beginning of the experiment dramatically changed the competition outcome, resulting in a higher frequency of P9_B1, regardless of the fluorescent reporter used (p < 0.125 × 10^{-3}, p < 0.423 × 10^{-3} for P9_B1_{GFP} versus P9_B1^{spsM}_{mKate} and P9_B1^{mKate} versus P9_B1^{spsM}_{GFP}, respectively; STAR Methods) (Figures 5A and S5B). To test whether P9_B1* (the asterisk indicates a single exposure to SPβ) was more competitive due to a higher rate of lysogenic conversion and pBGC acquisition, the strains were separated to single colonies and were tested for antagonistic activity against the P9_B1 ancestor (Figure 5B). Indeed, the vast majority of P9_B1* colonies were antagonistic against the ancestor, which was not the case for P9_B1^{spsM}. Finally, from among the few P9_B1^{spsM} colonies that showed antagonistic activity, two colonies were screened by PCR. This screening confirmed this antagonistic activity correlated with the integration of SPβ-specific regions (including sublancin regions) into the chromosome (Figure S5C). Using Nanopore sequencing, SPβ integration was confirmed at an alternative locus (αzL) in the absence of the main att site in the bacterial chromosome (Figure S5D).

These results demonstrate that a phage carrying a pBGC can confer a fitness benefit to its host bacterium. In addition, bacteria can also benefit from carrying attachment sites in their genomes where BGC-containing phages can integrate.

**BGC-carrying phages are more promiscuous than the BGCs they carry**

The subsequent experiments involved assessing the role of phages in the horizontal gene transfer (HGT) of BGCs between closely and more distantly related bacteria. First, all BGC-carrying prophages and virions along with the core biosynthetic genes of their corresponding pBGCs were compared using average nucleotide identity (ANI) values (Figure S6). Next, all pBGCs were subjected to BiGSCAPE clustering, which allowed us to parse the 393 pBGCs into 60 gene cluster families (referred to as BiGSCAPE families hereafter; Figure 6). This enabled the assessment of the pBGC conservation level within host genera, and of the potential of phages to transfer pBGCs across genera and families. When comparing ANI clustering of phage genomes to BiGSCAPE families (Figure 6) and ANI clustering of core biosynthetic pBGC genes (Figure S6), clustering according to...
bacterial host genera was evident. Most pBGC-carrying virions retrieved from the PATRIC database had high similarity to prophages retrieved from their PATRIC-annotated hosts (Figures 6 and S6). Thus, highly similar genetic material is found both in the prophages and virions. This confirms the presence of a functional lytic cycle in certain pBGC-carrying prophages that target *Escherichia*, *Mannhemia*, *Listeria*, and *Lactobacillus*. This opens up the possibility of pBGCs being transferred via phage transduction in nature, as we demonstrated in vitro in the *B. subtilis* model.

By comparing the sequence similarity of the pBGC-carrying prophages with the BIGSCAPE family of the pBGCs, it was possible to identify three main patterns that were dependent on the bacterial genera: (1) phage relatedness matches pBGC relatedness, (2) distinct pBGCs can be found in similar phages, and (3) the sequences of both pBGC carrying prophages and virions can be entirely different (including the pBGC) despite having similar hosts.

The first pattern was represented by *Bacillus*, *Lactobacillus*, and *Listeria*. In these genera the distribution of pBGC-carrying phages overlaps with the BIGSCAPE families of the pBGCs (Figure 6). For *Listeria*, the pBGC-carrying phages, BIGSCAPE family, and pBGC core biosynthetic gene ANI (Figure S6) are nearly identical within the genus. Although the remaining genera (*Bacillus* and *Lactobacillus*) contain some outlier pBGCs, that do not belong to the main pBGC clusters; these were always associated with outlier prophages found within the corresponding genera. This suggests that in *Bacillus*, *Lactobacillus*, and *Listeria*, potential HGT of pBGCs may be restricted to the genus level.

The second pattern was represented by the two large clusters of the members of Enterobacteriaceae (*Escherichia*, *Shigella*, *Dickeya*, and *Enterobacter*) and Pasturellaceae (*Mannhemia*,...
Bacteria benefit from attachment gene for pBGC-carrying phages
(A) Graphics above: SP\(\text{B}\) typically integrates into spsM gene splitting it. It was hypothesized that the frequency of lysogenic conversion will be greatly reduced if spsM is removed. Below: long-term competition assay between initially SP\(\text{B}\)-negative P9_B1\(\text{mKate}\) versus P9_B1\(\Delta\text{spsM}\) without and with single exposure to SP\(\text{B}\) at the exponential growth phase during first co-cultivation round. All assays were initiated at 1:1 ratios. Relative abundance of two competing strains was monitored by changes in ratios of GFP and mKate fluorescence signal, shown in yellow and purple, respectively. Error bars (light gray) indicate standard error (n = 8).

(B) Mixed populations of P9_B1\(^*\) and P9_B1\(\Delta\text{spsM}\) (asterisk indicates the strains were exposed to SP\(\text{B}\)) after long-term cultivation experiment were screened for antagonistic activity against the P9_B1 ancestor. All co-cultures exposed to SP\(\text{B}\) acquired antagonistic activity toward the P9_B1, which was not observed in any of the control co-cultures (no phage solution added); see Petri-dish images on top. P9_B1\(^*\) and P9_B1\(\Delta\text{spsM}\) were separated into single colonies and again screened for antagonistic activity against the P9_B1 ancestor by streaking the colonies on top of the P9_B1 lawn. The blue rectangles (right corner) are example images obtained during the screening assay, showing LB-agar plates covered with P9_B1 lawns and with streaks of P9_B1\(^*\) and P9_B1\(\Delta\text{spsM}\) single colonies, illuminated by UV lamp. The plates were imaged by iPhone\(7\) and the fraction of P9_B1\(^*\) and P9_B1\(\Delta\text{spsM}\) colonies inhibitory toward P9_B1 was calculated (boxplot). Green streaks indicate GFP-labeled strains and pink streaks indicate mKate-labeled strains. ***p < 0.001 (n = 4).

See also Figure S5.

Pasteurella, and Haemophilus). Here, highly similar phages may carry distinct pBGCs. In the cluster containing Mannheimia, Pasteurella, and Haemophilus, this was especially evident as the latter two carry Mannheimia-like prophages, but the BiGSCAPE families remain distinct within each genus (Figure 6). The same is true for Enterobacter and Klebsiella, which carry Escherichia-like prophages, but not Escherichia-like pBGC BiGSCAPE families (Figures 6 and S6). This may indicate that pBGCs play a role in the early diversification of phages toward different host species or genera. BLAST analysis confirmed that most pBGCs (290 of 307) were similar in sequence (>70% similarity and >70% length) to BGCs within the genome of another pBGC. Only 17 were unique pBGCs. This was as also observed from the ANI values (Figure S6).

The third pattern was represented by outlier prophages belonging to Bacillus and the virions of Escherichia and Aeromonas among others (Figures 6 and S6).

BLAST comparison also revealed that 19 pBGC carriers hosted an additional copy of a highly similar (80%–90%) bacteriocin BGC outside the original prophage region (Table S3). It was also found that 5.1% of genomes that were originally identified as non-pBGC carriers (14,877 in total; Figure 1B) had at least one analog to the pBGC core biosynthetic genes. In 56.4% of these cases, the pBGC analogs were found in the same species as the original pBGC host. In 38.0% of cases, homologs were found in different species within the same genus as the pBGC host. Meanwhile, in 7.3% of cases, homologs were found in different genera, but the same family (e.g., Escherichia > Shigella, Manheimia > Glaesserella).

Overall, these results indicate that despite the promiscuity of certain pBGC-carrying phages, and the high potential for within-genus transfer of pBGCs, pBGCs are rarely shared between different host genera.

Importantly, the 60 BiGSCAPE families in our analysis, which were mostly associated with RiPP bacteriocins, represent only 5.80% of the 1,023 RiPP families derived from complete bacterial genomes in the BiGFAM database. This indicates that pBGC RiPPs only constitute a small fraction of the total diversity of RiPPs.

DISCUSSION

Specialized metabolites play an important role in microbial interactions and are exploited by the pharmaceutical industry. However, understanding the role of these compounds in nature has only recently become a research focal point. In the present study, we explored the novel concept that phages may function as vectors for transferring the potential for specialized metabolite biosynthesis, thereby contributing to their natural distribution and ecological impact. Although BGCs are rarely found within phage genomes, when they are, they mostly reside within temperate phages or prophages. Given the relative rarity and limited diversity of pBGCs, phage propagation of secondary metabolites likely constitutes a minor, but ecologically important
role of BGC exchange. This study demonstrates that phage-mediated BGC exchange can occur spontaneously during interaction between two bacterial strains over relatively short timescales (days).

The bioinformatics approach used in this study has several drawbacks that are inherent to the scale of our analysis. The ProphET algorithm was used for prophage recognition as it is currently the best-performing stand-alone prophage finder. However, the reliance on a prophage database may underestimate novel prophages. AntiSMASH is currently the gold standard for profiling genomes for all types of BGCs, including genes that encode for bacteriocins.

Almost all of the pBGCs detected in this study encoded for bacteriocins. Considering the relatively high abundance of other phage-encoded compounds destined for secretion (such as toxins and other virulence factors) that impact bacterial ecology, one could expect phages to carry BGCs encoding for bacteriocins. On the other hand, hosting large multi-gene assemblies like BGCs is probably constrained by the efficiency of the phage DNA packing process, which translates to virion stability. Such size constraints could also explain why phages accommodate the smallest BGCs—those encoding bacteriocins, which rarely exceed 10 kb.

It has previously been shown that a host-associated lifestyle tends to correlate with higher prophage cargo. This is in line with our findings that there is a higher abundance of pBGCs in commensal and pathogenic species (Figure 2A). However, host lifestyle cannot be the sole predictor for pBGC abundance, since some recurrent facultative pathogens, such as *Acinetobacter*, *Burkholderia*, *Pseudomonas*, or *Staphylococcus*, for which there is a high abundance of genomes in NCBI, contained no, or very few pBGCs, despite the presence of prophages.

The fact that pBGCs were overrepresented in bacterial genera with lower numbers of gBGCs, especially *Escherichia*, could suggest that the presence of gBGCs reduces the likelihood of phage infections. This matches previous data on the role of gBGCs in anti-phage defense. Alternatively, lysogeny could reduce the likelihood of gBGC acquisition, for instance by negatively influencing genetic competence for transformation. Nonetheless, our results were in line with those of a survey of BGCs within human microbiomes, where the leading pBGC carriers *Escherichia*, *Lactobacillus*, *Haemophilus*, and *Enterococcus* were identified as having the lowest numbers of BGCs (less than 2).

The present work supports previous findings on temperate phages as carriers of putative weapons in inter-bacterial competition. Previously, it was demonstrated that phages can repair BGCs by transducing missing genes between strains. In the present study, a complete and functional BGC was gained by prophage integration. This work demonstrates, in line with previous findings, that sublancin alone can be sufficient to explain the fitness benefits of lysogeny. These benefits are present within a clonal population, with the maintenance of the bacteriocin-carrying MGE (in this case a prophage) being ensured through addiction module principle. This principle has been well described for toxin-antitoxin systems and has been previously demonstrated for plasmid-encoded bacteriocins.

Figure 6. Diversity of phages and pBGCs
(A) Similarity network of full-length prophages and virions carrying pBGCs, labeled by major genera. Each edge is weighted by the ANI value of each pair.
(B) Same as (A) but colored in a gradient according to BigScape family. Note that BigScape families with closely related colors are not related unless connected through an edge. Black nodes have BGCs not recognized by BigScape.

See also Figure S6 and Table S3.
Recently, so-called MuF polymorphic toxins have been predicted within multiple prophages of Firmicutes and have been proposed to serve as weapons in microbial warfare.23 Only minimal overlap can be observed between MuF toxin carriers and pBGC-carrying phages from our dataset (nine bacterial genomes and only five prophages overlap; compare Data S1 and S2 and Figure S6). This confirms their differences in genetic architecture but does not exclude similar ecological function, which in the case of MuF toxins requires experimental validation.

In the present study, it was observed that the pBGC family correlated with host taxonomy more than the entire genomes of pBGC-carrying phages. This result is in line with the results of a previous study that compared phage genomes and single protein networks with host phylogeny.55 Specifically, this previous study revealed that the best predictors of bacterial host phylogeny are phage proteins that play a role in bacterial surface recognition.55 This leads us to question of whether bacteriocin-encoding pBGCs could be a good predictor of phage host range, and what the potential underlying mechanism might be. Phage specificity depends on recognition of phage receptor-binding proteins (RBPs) by specific receptors on the surface of host cells. Therefore, it is possible that phage-encoded bacteriocins could be recognized by the same receptors. The next question to address is whether phages are more specific than bacteriocins they carry, or vice versa. From comparing the narrow host range of SPb24 with reports on the anti-staphylococcal activity of sublancin,56 it may appear that the phages are more specific. However, a systematic study addressing this fascinating evolutionary question still awaits.

Recent work on virome transplants57 has indicated the tremendous role of phages in human health. It remains to be discovered to what extent this role can be assigned to pBGCs. First, the fitness benefits provided to the bacterial host should increase the vertical transmission of pBGC-carrying phages. In addition, the horizontal transmission of pBGCs would be facilitated by the induction of phage lytic cycles, which can be easily triggered by antibiotic therapy or even due to diet choices.58 Here, we demonstrate a natural role of pBGCs in interbacterial warfare and show that they can be transferred, and confer benefits to, another bacterial host. Thus, these findings unveil the potential for new avenues in the biotechnical applications of specialized metabolites. In summary, encoding of bacteriocins could be an evolutionary strategy adapted by some phages to weaponize their host, thereby conferring a competitive advantage to both the host and the phage infecting it.

STAR Methods

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Strain construction
- **METHOD DETAILS**
  - Phage isolation and purification
  - Sublancin activity assay
  - Competition assays
  - Sublancin purification
  - Chemical confirmation of sublancin
  - Sample preparation and LC-MS
  - Transmission electron microscopy
  - pBGC mining
  - Profiling of pBGCs
  - Genetic comparisons
  - Whole genome sequencing by MiniON Nanopore

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.05.046.

**ACKNOWLEDGMENTS**

The authors thank Lone Gram for feedback during ongoing project and constructive comments on the final version of this manuscript. We also thank Anika Kofod-Petersen for valuable discussions. We thank Elisa Granato and two anonymous reviewers for extensive help in shaping the manuscript. This work was supported by the Danish National Research Foundation (DNRF137) for the Center for Microbial Secondary Metabolites.

**AUTHOR CONTRIBUTIONS**

A.D. and M.L.S. designed the study. M.L.S. performed bioinformatic analysis, A.D. performed the experiments, A.J.C.A. performed chemical analysis, C.N.L.-A. performed Nanopore sequencing and data processing, P.J.K. performed electron microscopy, and A.T.K. provided molecular tools for the study. A.D. and M.L.S. wrote the manuscript. All authors contributed to the final version of the manuscript.

**DECLARATION OF INTERESTS**

The authors declare no conflicts of interest.

Received: November 12, 2020
Revised: April 16, 2021
Accepted: May 20, 2021
Published: June 28, 2021

**REFERENCES**

1. Belknap, K.C., Park, C.J., Barth, B.M., and Andam, C.P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. Sci. Rep. 10, 2003.
2. Hug, J.J., Krug, D., and Müller, R. (2020). Bacteria as genetically programmable producers of bioactive natural products. Nat. Rev. Chem. 4, 172–193.
3. Traxler, M.F., and Kolter, R. (2015). Natural products in soil microbe interactions and evolution. Nat. Prod. Rep. 32, 956–970.
4. Granato, E.T., Meiller-Legrand, T.A., and Foster, K.R. (2019). The evolution and ecology of bacterial warfare. Curr. Opin. Chem. Biol. 50, R521–R537.
5. Coyte, K.Z., Schluter, J., and Foster, K.R. (2015). The ecology of the microbial warfare and competition axes and evolution. Nat. Prod. Rep. 10, 925–936.
6. Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. 3, 925–936.
8. Süssmuth, R.D., and Mainz, A. (2017). Nonribosomal peptide synthesis—principles and prospects. Angew. Chem. Int. Ed. Engl. 56, 3770–3821.
9. Li, Y., and Rebuffat, S. (2020). The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. J. Biol. Chem. 295, 34–54.
10. Biggs, J.B., Termei, M.A., and Brady, S.F. (2012). Malleilactone, a polyketide-synthesized virulence factor encoded by the cryptic secondary metabolome of Burkholderia pseudomallei group pathogens. J. Am. Chem. Soc. 134, 13192–13195.
11. Quadri, L.E.N., Sello, J., Keating, T.A., Weinreb, P.H., and Walsh, C.T. (1998). Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem. Biol. 5, 631–645.
12. Tyc, O., Song, C., Dickschat, J.S., Vos, M., and Garbeva, P. (2017). The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. Trends Microbiol. 25, 280–292.
13. Aleti, G., Baker, J.L., Tang, X., Alvarez, R., Dinis, M., Tran, N.C., Melnik, A.V., Zhong, C., Ernst, M., Dorrestein, P.C., and Edlund, A. (2019). Identification of the bacterial biosynthetic gene clusters of the oral microbe that illuminates the unexplored social language of bacteria during health and disease. MBio 10, e00321-e19.
14. Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C., Mavrommatis, K., Patti, A., Godfrey, P.A., Koehrens, M., Clardy, J., and et al. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158, 412–421.
15. Davies, J., Spiegelman, G.B., and Yim, G. (2006). The world of subinhibitory antibiotic concentrations. Curr. Opin. Microbiol. 9, 445–455.
16. Blin, K., Kim, H.U., Medema, M.H., and Weber, T. (2019). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. Brief. Bioinform. 20, 1103–1113.
17. Donia, M.S., Cimermancic, P., Schulze, C.J., Wieland Brown, L.C., Marty, J., Mitreva, M., Clardy, J., Lintoning, R.G., and Fischbach, M.A. (2014). A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158, 1402–1414.
18. Sharrar, A.M., Crits-Christoph, A., Meheust, R., Diamond, S., Starr, E.P., and Banfield, J.F. (2020). Bacterial secondary metabolic biosynthesis potential in soil varies with phylum, depth, and vegetation type. MBio 17, e00416-e00420.
19. Libis, V., Antonovsky, N., Zhang, M., Shang, Z., Montiel, D., Maniko, J., Termei, M.A., Calle, P.Y., Lemetre, C., Owen, J.G., and Brady, S.F. (2019). Uncovering the biosynthetic potential of rare metagenomic DNA by soil bacteria. Curr. Biol. 39, 3479–3489, August 23, 2021.
20. Ruzzini, A.C., and Clardy, J. (2016). Gene flow and molecular innovation in Bacillus subtilis. Mol. Biol. Evol. 20, 2076–2090.
21. Abe, K., Kawano, Y., Iwamoto, K., Arai, K., Maruyama, Y., Eichenberger, P.M., and Sato, T. (2014). Developmentally-regulated excision of the SP prophage reconstitutes a gene required for spore envelope maturation in Bacillus subtilis. PLoS Genet. 10, e1004636.
22. Reis-Cunha, J.L., Bartholomeu, D.C., Manson, A.L., Earl, A.M., and Cerqueira, G.C. (2019). ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-updating reference data. mSystems 6, e00770, e20.
23. 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. bioRxiv. https://doi.org/10.1101/2020.12.13.422588.
24. Westers, H., Dorenbos, R., van Dijl, J.M., Kabel, J., Planagan, T., Devine, K.M., Jude, F., Seror, S.J., Beekman, A.C., Darmon, E., et al. (2013). Genome engineering reveals large dispensable regions in Bacillus subtilis. Mol. Biol. Evol. 20, 2076–2090.
25. Abe, K., Kawano, Y., Iwamoto, K., Arai, K., Maruyama, Y., Eichenberger, P.M., and Sato, T. (2014). Developmentally-regulated excision of the SP prophage reconstitutes a gene required for spore envelope maturation in Bacillus subtilis. PLoS Genet. 10, e1004636.
26. Reis-Cunha, J.L., Bartholomeu, D.C., Manson, A.L., Earl, A.M., and Cerqueira, G.C. (2019). ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-updating reference database. mSystems 6, e00770, e20.
44. Maqueda, M., Sánchez-Hidalgo, M., Fernández, M., Montalbán-López, M., Valdivia, E., and Martínez-Bueno, M. (2008). Genetic features of circular bacteriocins produced by Gram-positive bacteria. FEMS Microbiol. Rev. 32, 2–22.
45. Touchon, M., Berneche, A., and Rocha, E.P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J. 10, 2744–2754.
46. Costa, A.R., Monteiro, R., and Azeredo, J. (2018). Genomic analysis of Staphylococcus aureus—prevalence, chromosome location and major genes involved. Viruses 10, 297.
47. Goerke, C., Pantucek, R., Hoftdreser, S., Schulte, B., Zink, M., Grumann, D., Bröker, B.M., Doksar, J., and Wolz, C. (2009). Diversity of prophages in dominant Staphylococcus aureus clonal lineages. J. Bacteriol. 191, 3462–3468.
48. Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A.I., Mantel, I., Nodwell, J.R., and Maxwell, K.L. (2018). A chemical defence against phage infection. Nature 564, 283–286.
49. Brooks, M.R., Padilla-Vélez, L., Khan, T.A., Qureshi, A.A., Pieper, J.B., Maddox, C.W., and Alam, M.T. (2020). Prophage-mediated disruption of genetic competence in Staphylococcus pseudintermedius. mSystems 5, e00684, e19.
50. Mashburn-Warren, L., Goodman, S.D., Federle, M.J., and Prenha, G. (2018). The conserved mosaic prophage protein paraxin inhibits the natural competence regulator ComI in Streptococcus. Sci. Rep. 8, 16535.
51. Matilla, M.A., and Salmond, G.P.C. (2014). Bacteriophage eSM1, a viu-alkalivirus, is a broad-host-range, high-efficiency generalized transducer that infects environmental and clinical isolates of the enterobacterial genera Serratia and Kluyvera. Appl. Environ. Microbiol. 80, 6446–6457.
52. Engelberg-Kulka, H., and Glaser, G. (1999). Addiction modules and programmed cell death and antiedath in bacterial cultures. Annu. Rev. Microbiol. 53, 43–70.
53. Inglis, R.F., Bayramoglu, B., Gillor, O., and Ackermann, M. (2013). The role of bacteriocins as selfish genetic elements. Biol. Lett. 9, 20121173.
54. Shapiro, J.W., and Putonti, C. (2018). Gene co-occurrence networks reflect bacteriophage ecology and evolution. MBio 9, e01870, e17.
55. Wang, S., Wang, Q., Zeng, X., Ye, Q., Huang, S., Yu, H., Yang, T., and Qiao, S. (2017). Use of the antimicrobial peptide sublancin with combined antibacterial and immunomodulatory activities to protect against methicillin-resistant Staphylococcus aureus infection in mice. J. Agric. Food Chem. 65, 8955–8965.
56. Rasmussen, T.S., Mentzel, C.M.J., Kot, W., Castro-Mejía, J.L., Zuffa, S., Swann, J.R., Hansen, L.H., Vogensen, F.K., Hansen, A.K., and Nielsen, D.S. (2020). Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model. Gut 69, 2122–2130.
57. 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 Microbes 11, 721–734.
58. Dubois, J.Y.F., Kouwen, T.R., Schurich, A.K., Reis, C.R., Ensing, H.T., Trip, E.N., Zweers, J.C., and van Dijl, J.M. (2009). Immunity to the bacteriocin sublancin 168 Is determined by the Suni (YoF) protein of Bacillus subtilis. Antimicrob. Agents Chemother. 53, 651–661.
59. Sanchez-Vizuete, P., Le Coq, D., Bridier, A., Herry, J.M., Aymerich, S., and Briandet, R. (2015). Identification of ypoP 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.
60. Lane, D.J. (1991). 16S/23S rRNA sequencing. In Nucleic Acid Techniques in Bacterial Systematics, E. Stackebrandt, and M. Goodfellow, eds. (John Wiley & Sons), pp. 115–175.
61. van Gestel, J., Weissing, F.J., Kuipers, O.P., and Kovács, Á.T. (2014). Density of founder cells affects spatial pattern formation and cooperation in Bacillus subtilis biofilms. ISME J. 8, 2069–2079.
62. Dragóš, A., Kiesewalter, H., Martin, M., Hsu, C.Y., Hartmann, R., Wechsler, T., Eerksen, C., Brix, S., Drescher, K., Stanley-Wall, N., et al. (2018). Division of labor during biofilm matrix production. Curr. Biol. 28, 1903–1913.e5.
63. Schindelin, J., Arganda-Carreras, I., Frise, E., Kayning, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schindl, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.
64. Navaarro-Muñoz, J.C., Selén-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. Nat. Chem. Biol. 16, 60–68.
65. 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, W347–52.
66. 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 (W1), W16–21.
67. Tóth, I., Sváb, D., Bálint, B., Brown-Jaque, M., and Marót, G. (2016). Comparative analysis of the Shiga toxin converting bacteriophage first detected in Shigella sonnei. Infect. Genet. Evol. 37, 150–157.
68. 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. mSystems 5, 651–661.
## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Chemicals, peptides, and recombinant proteins** | | |
| Lysogeny broth (LB), Lennox | Carl Roth GmbH | Catalog # X964.1 |
| Agar-Agar | Carl Roth GmbH | Catalog # 5210.3 |
| PEG-8000 | Carl Roth GmbH | Catalog # 0263.2 |
| Magnesium (II) sulfate | Carl Roth GmbH | Catalog # P027.1 |
| Tris-HCl | VWR | Catalog # 103157P |
| Tris base | Sigma – Aldrich | Catalog # T1503-1KG |
| HCl (37%) | Sigma – Aldrich | Catalog # 258148-500mL |
| Tris(2-carboxyethyl)phosphine | Sigma – Aldrich | Catalog # C4706-2G |
| NaCl | Carl Roth GmbH | Catalog # 2326.1 |
| Trifluoroacetic acid | MilliporeSigma | Catalog # T6508-500ML |
| Acetonitrile | VWR | Catalog # 60-046-514 |
| Methanol | VWR | Catalog # 85661.320 |
| Formic acid | Fisher Chemicals | Catalog # A117-50 |
| Uranyl acetate | EMS | Catalog # 22400 |
| Glycerol | Carl Roth GmbH | Catalog # 7533.1 |
| EDTA | Carl Roth GmbH | Catalog # X986.1 |
| SDS | Carl Roth GmbH | Catalog # 2326.1 |
| Phenol | Sigma | Catalog # P4557-400 ml |
| Chlorophorm | VWR | Catalog # 83626.290P |
| Potassium acetate | Carl Roth GmbH | Catalog # 4986.1 |
| Isopropanol | Sigma – Aldrich | Catalog # 2413-2.5L-R |
| Ethanol | VWR | Catalog # 153386F |
| Spectinomycin dihydrochloride | Alfa Aesar | Catalog # J61820 |
| Kanamycin | Carl Roth GmbH | Catalog # T832.2 |
| Chloramphenicol | Carl Roth GmbH | Catalog # 3886.2 |
| Mitomycin C | Carl Roth GmbH | Catalog # 4150.1 |
| **Enzymes** | | |
| DNase I | Thermo Fisher | Catalog # EN0521 |
| RNase A | EurX Bacterial & Yeast Genomic DNA Purification Kit | Catalog # E3580 |
| Proteinase K | EurX Bacterial & Yeast Genomic DNA Purification Kit | Catalog # E3580 |
| Chymotrypsin | Thermoscientific | Catalog # 90056 |
| **Experimental models: Organisms/strains** | | |
| *Bacillus subtilis* P9_B1 derivatives (listed in Table S1) | This study | N/A |
| *Bacillus subtilis* MB8_B7 derivatives (listed in Table S1) | This study | N/A |
| Δ6 | This study | N/A |
| *Bacillus subtilis* strains used as gDNA donors (listed in Table S1) | 38, 39, 59, 60 | N/A |
| *Bacillus phage* SPβ | This study | N/A |
| **Oligonucleotides** | | |
| TATTGAGTTGCGCAAACCTCATAGAAGTGA | This study | oTB122 (F) |
| CTGCCTGGAAGGAAGGCAGAGTTA | This study | oAD2 (R) |
| GCAGGGCCCTACACCGTGAGGAAA | This study | oAD28 (F) |

(Continued on next page)
RESOURCE AVAILABILITY

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

Materials availability
All bacterial strains generated during this work are freely available from A. Dragoš (anna.dragos@bf.uni-lj.si). The study did not generate new unique plasmids or new reagents.

Data and code availability
Raw sequencing data have been deposited to SRA database under SUB8228942 (for ΔspbMSP) and SUB8237787 (for SPβ) accession numbers. Raw data obtained during growth assays and competition assays using Synergy XHT multi-mode reader, have been deposited into Zenodo platform, and are available under: https://zenodo.org/record/4449746#.YAbG0OhKg2w. The complete code used for mining of phage-encoded BGCs is available at https://github.com/mikaells/Find-pBGCs.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All bacterial strains used in this study as well as strains used as genomic DNA donors belong to Bacillus subtilis species. Strains were routinely maintained in lysogeny broth (LB) (LB-Lennox, Carl Roth; 10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, and 5 g l⁻¹ NaCl) at 37°C with shaking at 200 rpm

Strain construction
All strains that were used in this study, or strains that were used as gDNA donors, are listed in Table S1. Strain DTUB231 (P9_B1 SPβ) was obtained by infecting P9_B1 with a Bacillus phage SPβ isolated from strain MB8_B7. Isogenic lysogeny was confirmed by colony PCR using primers pairs TB122/oAD2 and oAD28/oAD3 binding to flanking prophage regions and host chromosome close to attL or attR, respectively. Colonies that were positive for oTB122/oAD2 and oAD28/oAD3, but negative for oAD2/oAD3 (intact att) were selected as stable SPβ lysogens. Strain DTUB208 was obtained by transforming MB8_B7 with gDNA of SPmini. DTUB43 and DTUB222 were obtained by transforming P9_B1 with phy_sgFP and phy_mKATE2 plasmids, respectively. Strains DTUB235 and DTUB236 emerged from infecting of DTUB43 and DTUB222 with SPβ phage, respectively. Stable lysogens were confirmed by colony PCR as described above. Strains DTUB233 and DTUB234 were obtained by transforming DTUB43 and DTUB222 with gDNA isolated from GM3248 and selecting for kanamycin-resistant colonies. Strains DTUB244 and DTUB245 were obtained by transforming DTUB43 and DTUB222 with gDNA isolated from ANC3, also selecting for kanamycin resistance. Deletions were confirmed using oAD61/oAD62 binding to sunA-flanking regions. Working concentrations of antibiotics were 5 µg ml⁻¹ for erythromycin, kanamycin and chloramphenicol, and 100 µg ml⁻¹ for spectinomycin. Primers 27F/1492R targeting 16S rRNA were used as a positive PCR control.
**METHOD DETAILS**

**Phage isolation and purification**
Strain MB8_B7 was cultivated at 37°C with shaking at 200 rpm in mid-exponential phase. Mitomycin C was added (1.5 μg ml⁻¹) to trigger prophage induction, following cultivation for another 6 h. Cells were pelleted down (8000 g, 10 min), supernatant was filtered-sterilized and diluted in order to obtain well-separated single plaques. Plaque assays were performed using Δ6 strain as a host. Plaque giving PCR product for SPΔ-specific primer pairs oTB88/oTB89 and oAD61/oAD62, was carefully removed from the soft agar using a sterile scalpel, resuspended in 200 μL of SM buffer and used to infect exponentially growing phage-free host Δ6 to allow propagation. Phage was subsequently propagated in soft agar and liquid host suspension until the titer reached at least 10⁸ pfu/mL. Such culture supernatants were collected, adjusted to pH of 7.0, filter-sterilized and mixed at a 1:4 rate with PEG-8000 solution (PEG-8000 20%, 116 g l⁻¹ NaCl). After overnight incubation at 4°C, the solutions were centrifuged for 60 min at 12000 rpm to obtain phage precipitates. The pellets were resuspended in 1% of the initial volume in SM buffer (5.8 g l⁻¹ NaCl, 0.96 g l⁻¹ MgSO₄, 6 g l⁻¹ Tris-HCl, pH 7.5) to obtain concentrated solution of phage particles.

**Sublancin activity assay**
Sublancin activity assays were performed as previously described. Briefly, overnight cultures of selected strains were diluted 100 x, and 100 μL of the diluted cultures were transferred onto the LB-agar (1.5%) plate using a plastic spreader to serve as a lawn (target strain). For the focal strains, undiluted overnight cultures were spotted (5 μl) on top of the lawn. Plates were incubated at 37°C for 24 h. Clearing zone around the focal colonies indicated their antimicrobial activity toward the target strain.

**Competition assays**
Prior to competition assays the potential fitness costs of introducing fluorescent reporters, were examined and it was noted that mKate-labeled strains may have a slight disadvantage in competition (Figure S3). Therefore, all competition assays were performed with controls, where fluorescent reporters were swapped. As the increase in GFP and mKate fluorescence in monoculture matched the growth pattern of corresponding strains (Figure S3), they were used as relative strain ratio indicators in co-cultures. In co-cultures compensation correction was applied, to subtract the effect of GFP background (0.044%) in mKate channel. Although fluorescence serves as a good indicator of population density, at lower optical densities, mKate-labeled strains produces slightly lower fluorescence compared to GFP-labeled strains, which results in a slight overestimation of GFP-labeled strain in a co-culture at the start of the experiment (Figure S3D). Swapping the reporters in each competition assay alleviates this effect.

Overnight cultures of selected strains were obtained, optical density was measured and cultures were pelleted down (8000 g, 5 min) and re-suspended in 0.9% NaCl, to reach optical density of 5. Next, 1:1 co-cultures were created by mixing equal volumes of selected strains suspensions. Such co-cultures were then inoculated at 1% into 200 μl of LB distributed in 96-well microtiter plates. Cultivation was performed in Synergy XHT multi-mode reader (Biotek Instruments, Winooski, VT, US), at 37°C with linear continuous shaking (3 mm), monitoring the optical density (600 nm) as well as GFP (Ex: 482/20; Em:528/20; Gain: 35) and mKate (Ex: 590/20; Em: 635/32; Gain: 35) fluorescence every 10 min. For long-term competition assays, co-cultures were transferred to fresh LB medium (2.5% inoculum), every 24 h.

Growth rates were calculated by linear regression on log transformed OD-values in exponential phase.

**Sublancin purification**
In order to confirm the identification of sublancin, a semi-purified sample was obtained from 10 pooled cultures (50 mL) of Bacillus subtilis P9_B1SP. Small scale testing indicated that an initial QuEChERS-like extraction prior to SPE resulted in a greater yield of sublancin compared to direct SPE, therefore this modified method was utilized. Sodium sulfate (40 g) and sodium chloride (40 g) were added to 1 L of acetonitrile with 0.1% TFA in a 3 L glass bottle. The solution was shaken and the 1 L of culture was then added. The combined solution was then shaken for approximately 1 min and left to settle. The liquid phases were decanted from the undisolved salts into a 3 L separation funnel. This solution was left to separate and the aqueous phase was removed into a round bottom flask, and the organic phase was discarded. The aqueous phase was placed on a rotary evaporator to remove solubilized acetonitrile. The aqueous phase was removed from the evaporator once approximately 700 mL of remained (approximately 200 mL acetonitrile removed). The 700 mL aqueous extract was loaded onto a hydrated and equilibrated 10 g HP20SS column on a Biotage Isolera flash chromatography system. The column was washed with H₂O (150 mL), and the sample was fractionated using a stepwise gradient of MeCN and H₂O, both containing 0.05% TFA. A stepped gradient with 10% intervals, and fraction volumes of 150 mL, was used. The 30% MeCN fraction contained the highest proportion of sublancin (determined my LC-MS), and was dried on a rotary evaporator, prior to final purification by HPLC.

The fraction containing sublancin was further fractionated on an Agilent 1290 HPLC-DAD equipped with a Phenomenex Luna Phenyl Hexyl column (250x10 mm, 5 μm, 100 Å). Elution utilized two eluents, MeCN with 50 ppm TFA, and H₂O with 50 ppm TFA. Separation was achieved with gradient elution from 20%-40% MeCN over 30 min. Fractionation was undertaken in 1 min intervals, and a fraction containing a high abundance of sublancin eluted between 4 and 5 min (determined by LC-MS). Although not completely purified, the sample was sufficiently concentrated for chemical investigations in order to confirm the identity of sublancin. To a dried subsample of the purified sublancin, 100 μL of TCEP (50 mM in H₂O) was added, and left at room temperature to react (2 h). Afterward, the sample was desalted using C₁₈ packed 250 μL pipette tips, and then submitted to LC-MS. A portion of the TCEP
reduced sublancin was dried under a nitrogen stream and redissolved in chymotrypsin solution (100 mM Tris-base, adjusted to pH 8.6 with HCl) and left at 25°C for 5 h to react. In parallel, a subsample of the native sublancin sample was also redissolved in chymotrypsin solution and left to react 25°C for 5 h. After 5 h, both samples were acidified with formic acid, and desalted with C18 packed 250 μL pipette tips before being submitted to analysis by LC-MS/MS.

Chemical confirmation of sublancin
Sublancin was identified based on mass spectrometric characteristics expected of the published structure and molecular formula (C162H254N50O51S5) of sublancin. The identification of the sublancin chemical feature was based on accurate mass, isotope pattern analysis, reduction of disulphide bonds using TCEP, targeted fragmentation of the reduced molecular ion, digestion of the sublancin by chymotrypsin, and subsequent fragmentation of products of the digest. These characteristics, together, were used to identify sublancin.

Initial analysis by LC-MS revealed an accurate mass and isotope pattern that conformed closely to that which would be expected of sublancin, with an accurate mass of the monoisotopic mass < 1 ppm, and an isotope pattern with a deviation of < 5% relative to the principle ion (Figure S2B).

Upon reduction by TCEP, the sublancin molecular feature was no longer observed, and a new feature with an additional 4 Da became present, as expected for sublancin, having 2 disulphide bonds (Figure S2B). This reduced feature was then fragmented at multiple energies to investigate the sequence of amino acids. Analysis of the product ions revealed the expected amino acid sequence for all expected b ions up to b20, and all y ions up to y27 (Figure S2C). In addition, this fragmentation analysis indicated the positioning of a posttranslational modification corresponding to 162 Da (C6O5H10) on the y17 ion (Figure S2C). This modification is in alignment with the positioning and mass of the carbohydrate moiety on sublancin, published previously. To investigate the positioning of the disulphide bonds a subsample of the TCEP reduced sublancin sample, as well as sample of the native form, were digested with chymotrypsin. Possibly due to the high polarity of some of the expected digestion products, only one expected fragment was observed in both samples. The observed digest fragment was the Leu12-Try32 section, and significantly for this comparison, the Leu12-Try32 peptide in the TCEP reduced sample had an additional 2 Da compared to that of the non-reduced sublancin sample, indicating the presence of one disulphide bond, congruent with the published structure. Fragmentation of the digested peptide also confirmed the positioning of the terminal amino acids within this section of the peptide, results that are consistent with the previously report structure of sublancin.

Sample preparation and LC-MS
To a 15 mL falcon tube magnesium sulfate (280 mg) and sodium chloride (140 mg) was added. Acidified (0.1% TFA) acetonitrile (7 mL) was added, and shaken vigorously by hand for a short time. Culture media (7 mL) was then added to the solution. The mixture was extracted by shaking for 30 min at room temperature. The sample was centrifuged (4200 CFU, 5 min) to allow two liquid phases to form, an organic, acetonitrile-rich, upper phase, and an aqueous, salt-rich, lower phase. The aqueous phase was removed and partially evaporated to approximately 5.5 mL. The partially evaporated aqueous phase was then separated by solid phase extract (SPE), as follows: the aqueous extract was loaded onto a reversed-phase SPE cartridge (Phenomenex Strata-X: 3 mL, 30 mg) after addition of magnesium sulfate (280 mg) and sodium chloride (140 mg). The mixture was allowed to adsorb for 1 min. The grids were rinsed 3 times on droplets of milliQ water and subjected to staining with 2% uranyl acetate. Specifically, with the help of EM grid-grade tweezers, the grids were placed sequentially on droplets of 2% uranyl acetate so-

Transmission electron microscopy
Before use, 400 mesh nickel grids with a 3-4 nm thick carbon film, CF400-Ni-UL EMS (Electron Microscopy Sciences, Hatfield, PA, USA) were hydrophilized by 30 s of electric glow discharging. Next, 5 μL of purified phage solutions were applied onto the grids and allowed to adsorb for 1 min. The grids were rinsed 3 times on droplets of milliQ water and subjected to staining with 2% uranyl acetate. Specifically, with the help of EM grid-grade tweezers, the grids were placed sequentially on droplets of 2% uranyl acetate solution for 10 s, 2 s and 20 s. Excess uranyl acetate was wicked away using filter paper and the grids were allowed to dry overnight and submitted for analysis by transmission electron microscopy.
stored in a desiccator until analysis. Transmission electron microscopy was performed utilizing a FEI Tecnai T12 Biotwin TEM (Thermo Fisher Scientific, Hillsboro, OR, USA) operating at 120 kV located at the Center for Electron Nanoscopy at the Technical University of Denmark, and images were acquired using a Bottom mounted CCD, Gatan Orius SC1000WC (Gatan, Pleasanton, CA, USA).

pBGC mining
All complete bacterial refseq genomes were downloaded using ncbi-genome-download (https://github.com/kblin/ncbi-genome-download). Next, all genomes were profiled using a custom script in which prophages were found by using ProphET,40 in which BLASTP was substituted for DIAMOND for performance reasons. Next, total genomic biosynthetic gene clusters (gBGCs) were found using antiSMASH41 with the –minimal option enabled, again for performance reasons. Third, prophage encoded BGCs (pBGCs) were found by rerunning antiSMASH4 on the predicted prophages. The genomes of all phages were downloaded from PATRIC and subjected to antiSMASH analysis as above. All the BGC hits within phage and prophage genomes were listed in Data S1 and S2, respectively.

Certain species are overrepresented in the genome database and many of the genomes are highly similar which may give both additional false positives and negatives. Dereplication of 15,000+ genomes was considered unfeasible.

Profiling of pBGCs
The class of pBGCs were extracted from the antiSMASH output, which in the case of bacteriocins, is predicted by profile Hidden Markov Models (pHMMs) complemented by the RODEO algorithm.16 Core biosynthetic genes of bacteriocins were extracted from the antiSMASH output and were further annotated using BLASTP against the BACTIBASE database46 with default settings. Presence of pBGC core biosynthetic genes outside the prophage regions were investigated by using BLASTN against the host and non-host genomes with identity of 70% and a coverage of 50%. BiGSCAPE46 was used to profile the BGCs and cluster them into BiGSCAPE families.

Genetic comparisons
Prophages and virions positive for pBGCs were further analyzed by average nucleotide identity using pyani (https://github.com/widdowquinn/pyani), and the resulting nucleotide identity matrix was clustered by hierarchical clustering using hclust() from R on the corresponding distance matrix.

A network was then made by considering the ANI as an adjacency matrix using the igraph package in R and building a weighted and undirected graph.

The Phaster software66,67 was used to detect prophage elements in natural B. subtilis isolates. Next, the sequence of SPβ-like prophage detected in MB8_B7 was compared with SPβ sequence of B. subtilis 168 using BLASTN. Lack of ICEBs1 (or its fragments) in MB8_B7 was confirmed by BLASTN.

Whole genome sequencing by MinION Nanopore
Bacterial genomic DNA was isolated using EURex Bacterial and Yeast Genomic DNA Kit, while phage DNA was extracted using the phenol:chloroform method as described previously.68 The qualities of the extracted gDNA were evaluated by NanoDrop DS-11+ Fluorometer (Thermo Fisher Scientific). For multiplex MinION sequencing, the Rapid Barcoding gDNA Sequencing kit was used to allow pooling of samples on a single MinION Flow Cell (FLO-MINSP6). Genomic DNA library was prepared using the SQK-RBK004 rapid barcoding kit and subsequently loaded into the MinION Flow Cell. The local base calling was performed automatically in real-time using the MinKnow ONT software (v3.1.19). De-multiplexing and adaptor trimming was carried out using EP-I2ME Agent software (v2020.2.10) using the Fastq Barcoding r2020.03.10 function and the genomes were then assembled using Unicycler v0.4.8.69

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis
The association between pBGC positivity versus prophage or gBGC counts was tested for significance using Mann-Whitney U-tests. To calculate the individual effects of the SPβ prophage, the intact pBGC or only the sunA gene on strain performance during competition, the change in proportion from 0 h to 24 h in a strain with a given gene content was compared to the change in proportion of the non-modified strain with the same fluorescent marker. Statistical differences between two experimental groups were then identified using a two-tailed Student’s t tests assuming equal variance. To assess the effect of phage att site (spsM) on competition in the presence and absence of SPβ, changes in P9_B1 proportion throughout the competition (start versus 3rd at transfer 24 h) in the absence of SPβ, where compared to changes of its proportion during competition with single exposure to SPβ. Relative abundance of strains in co-culture was calculated based on fluorescence values (Figure S3). Normal distributions within the above datasets were confirmed by Kolmogorov–Smirnov (p > 0.05). Differences in growth rates between P9_B1, P9_B1GFP and P9_B1mKate were assessed by One-Way ANOVA and mean comparisons by Tukey test. No outliers were removed from the dataset. No statistical methods were used to predetermine sample size and the experiments were not randomized.