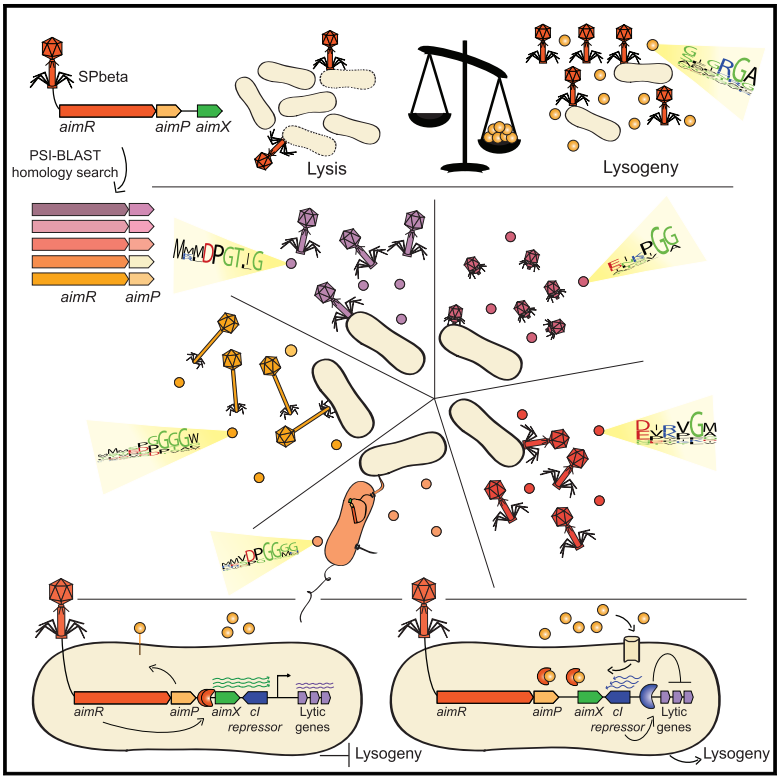
Article

Cell Host & Microbe

Widespread Utilization of Peptide Communication in Phages Infecting Soil and Pathogenic Bacteria

Graphical Abstract



Highlights

- Arbitrium-like systems are common in phages infecting pathogens and soil bacteria
- Peptide-based arbitrium communication regulates lysogeny in numerous phage types
- Arbitrium systems are frequently encoded on mobile pathogenicity elements
- A peptide-responsive non-coding RNA controls the regulator of the lysogenic state

Authors

Avigail Stokar-Avihail, Nitzan Tal, Zohar Erez, Anna Lopatina, Rotem Sorek

Correspondence

rotem.sorek@weizmann.ac.il

In Brief

Stokar-Avihail et al. discovered that numerous types of phages and mobile genetic elements encode peptide-based communication systems that guide lysogeny decisions. These elements infect many soil and pathogenic *Bacilli*. Peptide-based lysogeny decisions are likely executed by a non-coding RNA that controls the regulator of lysogeny.

Stokar-Avihail et al., 2019, Cell Host & Microbe *25*, 746–755
May 8, 2019
https://doi.org/10.1016/j.chom.2019.03.017

CellPress

CellPress

Cell Host & Microbe
Article

Widespread Utilization of Peptide Communication in Phages Infecting Soil and Pathogenic Bacteria

Avigail Stokar-Avihail,[1,2] Nitzan Tal,[1,2] Zohar Erez,[1] Anna Lopatina,[1] and Rotem Sorek[1,3,*]
[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel
[2]These authors contributed equally
[3]Lead Contact
*Correspondence: rotem.sorek@weizmann.ac.il
https://doi.org/10.1016/j.chom.2019.03.017

SUMMARY

Temperate phages can adopt either a lytic or lysogenic lifestyle within their host bacteria. It was recently shown that *Bacillus-subtilis*-infecting phages of the SPbeta group utilize a peptide-based communication system called arbitrium to coordinate the lysogeny decision. The occurrence of peptide-based communication systems among phages more broadly remains to be explored. Here, we uncover a wide array of peptide-based communication systems utilized by phages for lysogeny decisions. These arbitrium-like systems show diverse peptide codes and can be detected in numerous genetically distant phage types and conjugative elements. The pathogens *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* are commonly infected by arbitrium-carrying mobile elements, which often carry toxins essential for pathogenicity. Experiments with phages containing these arbitrium-like systems demonstrate their involvement in lysogeny decisions. Finally, our results suggest that the peptide-based decision is executed by an antisense RNA that controls the regulator of the lysogenic state.

INTRODUCTION

Temperate viruses are viruses capable of two possible life cycles—lytic and dormant (also referred to as "latent" or "lysogenic") (Howard-Varona et al., 2017). The temperate lifestyle is common among viruses that infect bacteria, termed phages. When infecting its host bacterium, a temperate phage must make a choice between replicating in the cell and eventually killing it (lysis) or entering a lysogenic state, in which its genome, termed a prophage, either integrates into the host genome or replicates as an episome (Howard-Varona et al., 2017; Ofir and Sorek, 2018). Prophage-encoded factors protect the host bacterium from further lytic infection by virions of the same phage (Abedon, 2015; Oppenheim et al., 2005).

The lysis-lysogeny decision has been extensively studied for the *Escherichia coli* phage Lambda and was found to be mainly influenced by the metabolic state of the infected cell and the number of co-infecting phage particles (Oppenheim et al., 2005; Zeng et al., 2010). Genetically, the lysis-lysogeny decision in phage Lambda is controlled by an interplay between the phage-encoded transcriptional regulators CI, CII, and Cro as well as a number of host-encoded factors (Oppenheim et al., 2005; Ptashne, 2004). The master regulator of the lysogenic state is the CI transcriptional repressor, which binds the promoters of the lytic genes and inhibits their expression (Oppenheim et al., 2005).

It was recently shown that *Bacillus* phages of the SPbeta group use a peptide-based communication system termed "arbitrium" to coordinate their lysis-lysogeny decisions (Erez et al., 2017). During lytic infection, each phage produces a measured amount of a communication peptide, which is released into the growth medium and internalized by nearby bacteria. In subsequent infections, progeny phages sense the concentration of this peptide and preferably enter the lysogenic cycle when its concentration is high. Such communication allows the presently infecting phage to evaluate the extent of recent bacterial lysis events by predecessor phage infections and switch to the lysogenic cycle when the bacterial population dwindles.

The arbitrium communication system has been characterized for phi3T (Erez et al., 2017), a phage from the SPbeta group that infects *Bacillus subtilis*. This system consists of three phage-encoded genes: *aimP*, producing the communication peptide; *aimR*, encoding the intracellular peptide receptor; and *aimX*, a non-coding RNA that is a negative regulator of lysogeny. During lytic infection, *aimP* encodes a short protein that is released as a pro-peptide into the growth medium where it is then further processed by extracellular proteases to produce a 6 amino acid (aa) mature communication peptide. This 6-aa peptide accumulates in the growth medium and is imported by nearby bacteria through the bacterial oligopeptide permease (OPP) channel. In the absence of mature peptides in the cell, AimR binds to the promotor of *aimX* and activates its transcription, resulting in the inhibition of lysogeny by AimX through a mechanism yet unknown. When bound by the peptide, AimR dissociates from the DNA and no longer activates the expression of AimX, thereby leading to lysogeny (Erez et al., 2017).

Close homologs of the arbitrium system were found mainly in genomes of SPbeta-like phages and prophages (Erez et al., 2017), where the system is always organized as an *aimR-aimP-aimX* gene cluster. Interestingly, different phages encode diverse

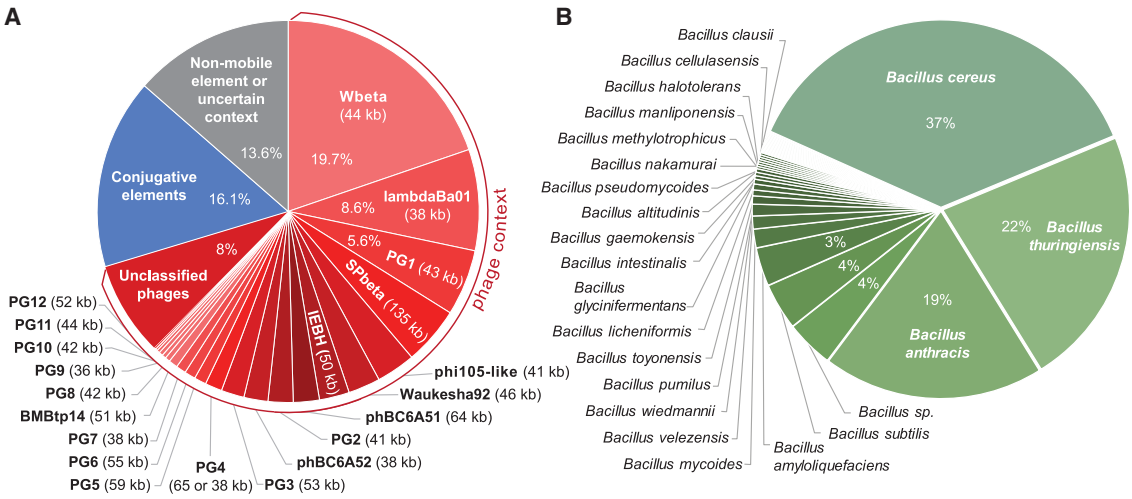


Figure 1. Arbitrium Systems Are Found in Diverse Mobile Elements of *Bacilli*

(A) Distribution of arbitrium systems in different phages and conjugative elements. Phage genome sequences were clustered to form distinct groups (STAR Methods). A representative phage name was designated for the group in cases where the group contained a previously described phage or phages with strong homology to a known phage (Table S1). PG designates phage groups with no strong homology to a known phage. Median genome size for each phage type is in parenthesis. Cases where the AimR was found near genes required for conjugation were designated as conjugative elements.

(B) Distribution of arbitrium-carrying mobile elements in host bacteria. Presented are species for which at least two elements with arbitrium systems have been detected.

communication peptides that vary in the first three amino acids (for example, the peptide sequence of phage phi3T is SAIRGA and that of phage SPbeta is GMPRGA). Each phage responds specifically only to its own peptide. Consistently, the structure of the AimR peptide receptor bound to its cognate peptide has recently been determined for phages phi3T and SPbeta, showing high specificity between the peptide and its binding pocket within the receptor (Dou et al., 2018; Gallego del Sol et al., 2019; Wang et al., 2018; Zhen et al., 2019).

While phages from the SPbeta group were shown to frequently encode arbitrium, it is currently unknown how common such communication is in other phages. In addition, the mechanism by which the signal propagates to establish lysogeny and the role of the AimX non-coding RNA in executing the lysogeny decision are still unresolved. In this study, we found that multiple types of phages and conjugative elements encode arbitrium systems that display a diverse set of peptide codes. We further show evidence that the AimX non-coding RNA controls the master regulator of lysogeny through direct RNA antisense interactions.

RESULTS

Arbitrium-like Systems Are Common in Phages Infecting Pathogens and Soil Bacteria

In our previous study (Erez et al., 2017), a BLAST search was used to identify close homologs of the AimR protein within phages and prophages. This search revealed 112 homologs, the majority of which are found within prophages that belong to the SPbeta group of phages. In an attempt to identify more distantly related communication systems, we now performed profile-based homology searches using iterative PSI-BLAST (Altschul et al., 1997) across a database of >60,000 bacterial, archaeal, and viral genomes (STAR Methods). This search retrieved 1,180 homologs of AimR, with 96% (1,135 of 1,180) of them located directly upstream of an *aimP*-like gene that contained an N-terminal signal peptide sequence (Table S1).

The vast majority of these arbitrium-like systems were located in phages and prophages, and these belong to numerous phage groups that vary in genome size (23–142 kb) and architecture (Figure 1A; Table S1). Apart from SPbeta, these phages include Wbeta, which infects *B. cereus* and *B. anthracis*, a variant of which has been widely used for clinical diagnosis of *B. anthracis* infections (Abshire et al., 2005); phage phIS3501, which is utilized by *B. thuringiensis* var. israelensis as a genetic switch for the regulation of toxin production (Moumen et al., 2012); and at least 15 other types of phages (Figure 1A). In addition, 16% of these arbitrium homologs reside within conjugative elements, consistent with previous observations that mobile elements can use peptide communication to regulate their transfer between hosts (Singh and Meijer, 2014).

Overall, mobile genetic elements containing arbitrium and arbitrium-like systems were found within the genomes of 41 species, and the vast majority of them represent bacteria from the genus *Bacillus* (Figure 1B; Table S1). In fact, over 50% (442/782) of all genomes belonging to this genus in the database we analyzed contained at least one arbitrium-carrying prophage or conjugative element, with up to 8 such prophages in a single genome (*B. thuringiensis* YBT-1518). Arbitrium-carrying mobile elements were especially prevalent in the sequenced genomes of the pathogens *B. anthracis* (97%, 89/92), *B. thuringiensis* (89%, 51/57), and *B. cereus* (76%, 167/224). Notably, arbitrium systems are found in plasmids that carry Bt insecticidal toxins in *B. thuringiensis* (Figures 2A and S1) as well as in the virulence plasmid pXO2, which is essential for full pathogenicity of *B. anthracis* (Dixon et al., 1999) (Figure 2B). This implies that peptide-based communication may play a role in regulating horizontal transfer of virulence traits among pathogenic bacteria (see Discussion).

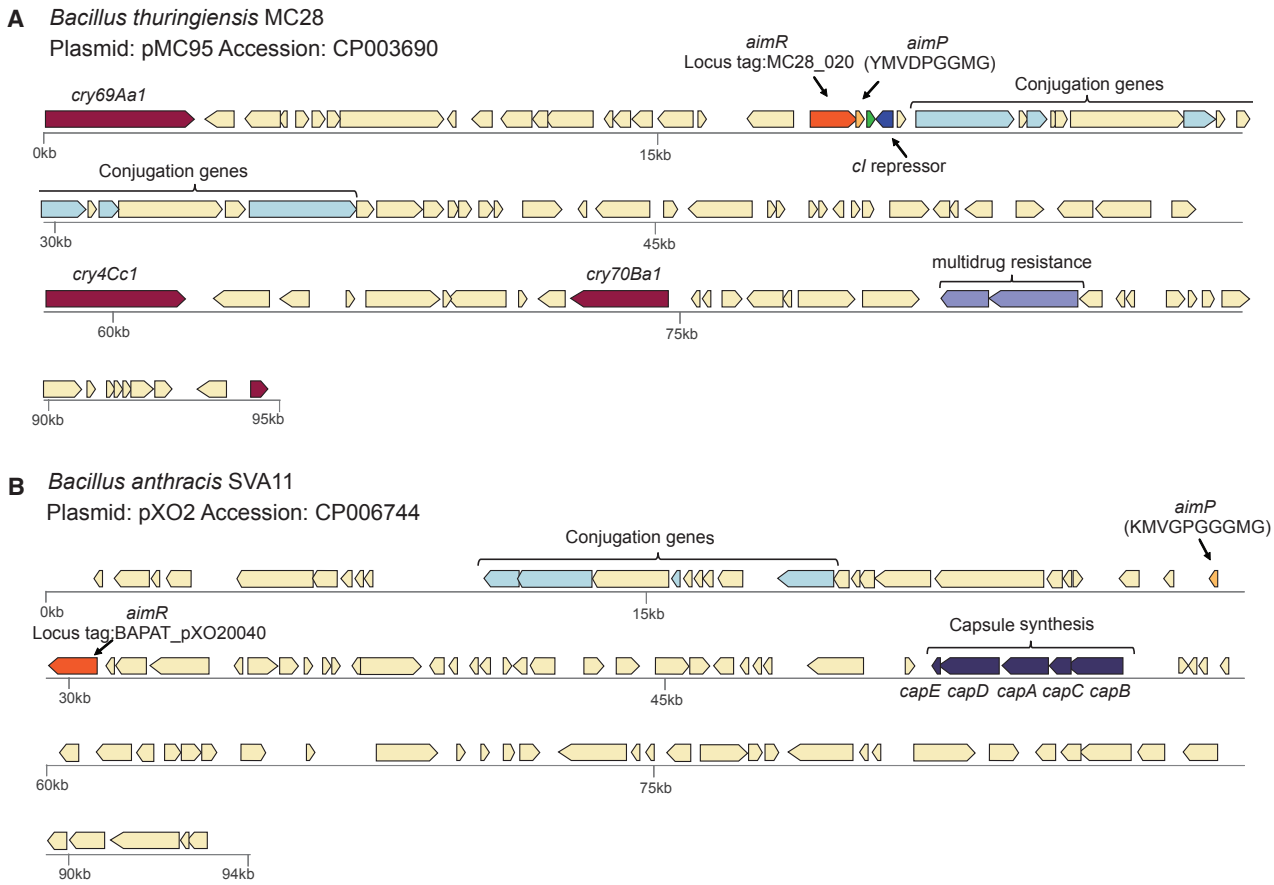


Figure 2. Arbitrium Systems on *Bacillus* Virulence and Insecticidal Plasmids

(A and B) Schematic representation of the virulence plasmids pMC95 of *B. thuringiensis* (A) and pXO2 of *B. anthracis* (B). The *aimR* gene is colored dark orange; *aimP* is light orange, with the putative mature peptide in parenthesis; predicted *cI* repressor dark blue; insecticidal Crystal (Cry) genes in burgundy; capsule synthesis genes in purple; and conjugation genes in light blue. The bacterial strain and the GenBank accession of the plasmids are specified on the upper left corner. See also Figure S1.

Phylogenetic analysis based on multiple sequence alignment of all AimR homologs showed that they cluster into 10 different clades (Figure 3). Five of these clades (clades 1, 2, 3, 4, and 9) are found almost exclusively in phages, one clade (5) is specific to conjugative elements, and three clades (6, 7, and 8) can be found both in phages and conjugative elements. An additional clade (colored gray in Figure 3) included predicted AimR homologs that are not found within a mobile element; we consider this clade as a possible non-arbitrium outgroup (see Discussion). The arbitrium systems of phages phi3T and SPbeta localized to clade 2, indicating that the originally described arbitrium systems form only a small subset of the actual genetic diversity of phage communication systems found in nature.

The mature communication peptide of the phi3T and SPbeta phages consists of the C-terminal end of AimP, formed by cleavage of a longer pro-peptide, similar to the Phr quorum-sensing peptides of *B. subtilis* (Pottathil and Lazazzera, 2003). We therefore examined the C terminus of the AimP protein associated with each of the AimR homologs. These C-terminal peptides were characterized by high diversity, with 168 unique sequences found. The peptides corresponding to each of the clades in the AimR tree form distinct sequence signatures, indicating that arbitrium systems from different clades correspond to different peptide codes (Figure 3). Our experimental data (see below) show that the size of the mature peptide can reach up to 10 aa in some clades (Figure 3).

Arbitrium Systems Regulate Lysogeny in Multiple Phages

We next examined the location of each arbitrium system within its respective phage genome. Arbitrium systems were frequently found at the edge of the integrated prophage (Figure 4), and in 70% of the cases (579 of 830 phages), they localized near a phage integrase gene, suggesting that these systems belong to the lysogeny module in most phages (Table S1).

To examine whether arbitrium-like systems from clades other than the SPbeta clade also function in guiding phage lysogeny decisions, we performed infection experiments with phage Wbeta, in which we detected an arbitrium-like system from clade 3 (Figure 5A). The arbitrium system in this phage is located in a genomic region previously defined as the lysogeny control module (Schuch and Fischetti, 2006). If the arbitrium-like system in phage Wbeta indeed guides its lysogeny decision, then we expect that addition of the arbitrium peptide to

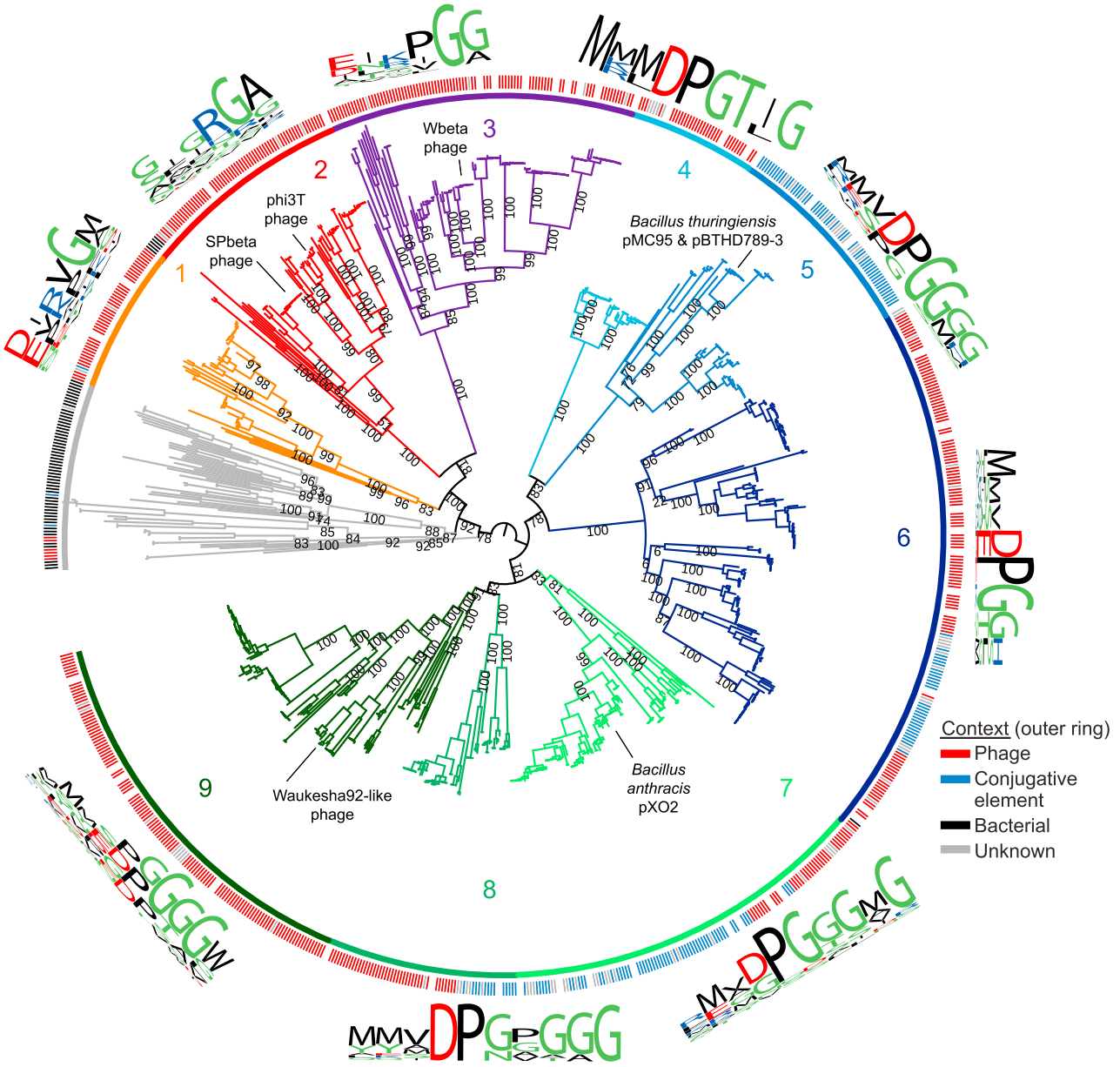


Figure 3. Phylogenetic Tree of 1,180 AimR Protein Homologs

Tree visualization was performed using iTOL (Letunic and Bork, 2016). The context of the arbitrium system within the bacterial genome is designated by colored lines in the tree circumference. Bootstrap support values, obtained from 1,000 iterations using IQ-TREE (Nguyen et al., 2015), are presented for major branches. Identical or near-identical sequences were collapsed into a single node. The tree is artificially rooted between the gray-colored clade and clades 4–9 for illustrative purposes. Sequence profile of the C terminus of the respective AimP, representing the predicted mature arbitrium peptide, is presented for each clade. See also Figure S3.

the growth medium during phage infection will lead to its lysogeny. Based on the observation that the C-terminal 6 aa in the AimP proteins of phages phi3T and SPbeta form the mature, active arbitrium peptide, we synthesized the peptide EIKPGG that represents the last 6 aa of the predicted AimP protein of phage Wbeta (Figure 5A). *B. cereus* bacteria infected by phage Wbeta in the presence of 1 μM of this peptide were completely protected from lysis (Figure 5B), whereas shorter forms of the peptide (EIKPG, IKPGG, and KPGG) had no effect (Figures 5C–5E), suggesting that the 6-aa peptide indeed guides Wbeta lysogeny in a highly specific manner. Semi-quantitative PCR using primers spanning the junction of Wbeta integration within the host *B. cereus* genome confirmed that the presence of the 6-aa peptide induces lysogeny as early as 30 min following initial infection (Figure 5F).

We next tested a more distantly related arbitrium-like system from clade 9. This system was found on a prophage that resides within *B. thuringiensis* sv. kurstaki HD-1. This prophage is almost

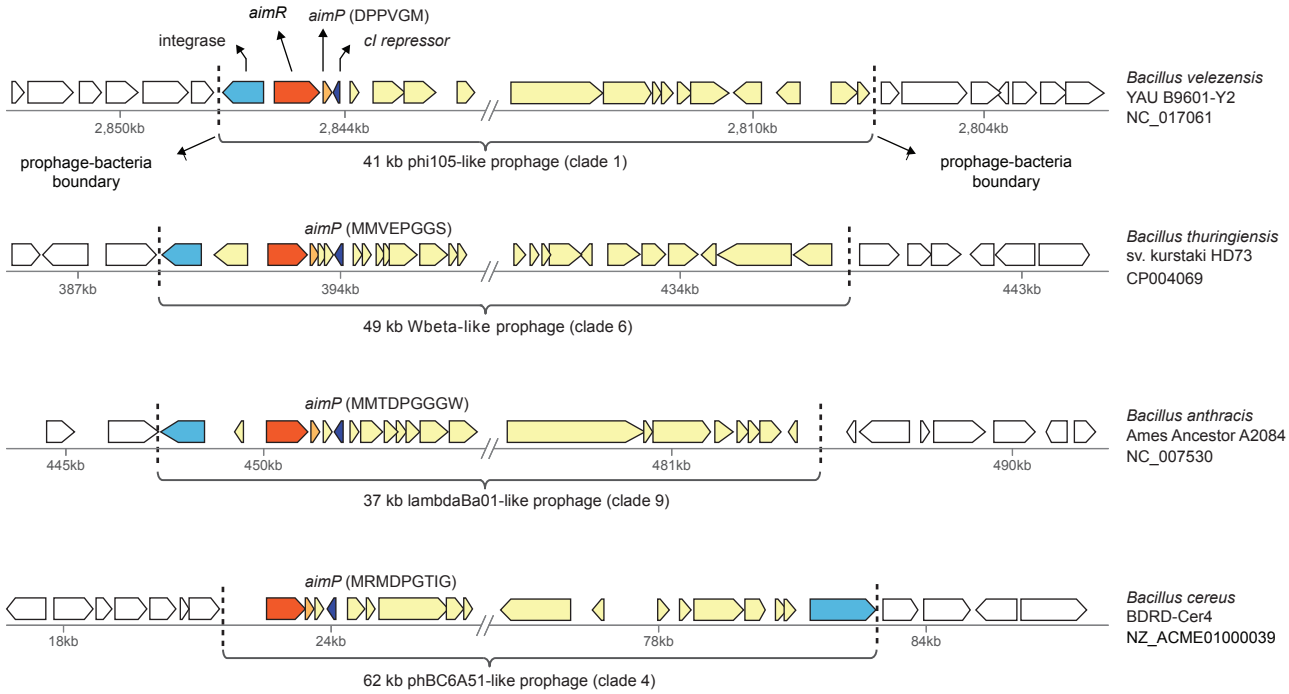


Figure 4. Arbitrium Systems Are Frequently Found in the Lysogeny Module of Diverse Phages

Schematic representation of the integration site of several arbitrium-containing prophages. The *aimR* gene is colored dark orange; *aimP* is light orange, with the putative mature peptide in parenthesis; phage integrase light blue; and predicted phage *cI* repressor dark blue. The rest of the phage genes are colored light yellow. Locus tag accessions for the presented *aimR* genes are BANAU_2649, HD73_0378, GBAA0429, and bcere0015_17290 from top to bottom. The putative prophage integration site is marked by a dashed line, and the prophage length is designated below, with the AimR clade number indicated in parenthesis. The bacterial host strain in which the prophage resides and the scaffold accession are specified to the right of each prophage.

identical to a *Myoviridae Bacillus* phage called Waukesha92 (Sauder et al., 2014) (Figure 6A). We induced the Waukesha92-like prophage from its original host and found that it can also infect *B. thuringiensis* sv. galleriae BGSC:4G5 (STAR Methods). However, in contrast to the cases of Wbeta, phi3T, and SPbeta, a peptide composed of the last 6 aa of the Waukesha92-like AimP protein did not affect its lytic behavior (Figure 6B).

A closer examination of the AimP of this Waukesha92-like phage revealed a sequence that resembles the consensus cleavage site for peptide maturation encoded immediately before the phi3T and SPbeta mature peptides (Figures 6A and S2). This motif is thought to be identified by *Bacillus* extracellular proteases that cleave immediately after it and release the mature short peptide (Erez et al., 2017; Pottathil and Lazazzera, 2003). Interestingly, the position of this predicted motif within the phage AimP suggested that the mature arbitrium peptide in this phage is 10 aa long with the sequence MMSEPGGGGW. Indeed, addition of 1 μM of this 10-aa peptide to the growth medium resulted in infection dynamics suggestive of increased lysogeny (Figure 6C). Shorter versions of the peptide (MSEPGGGGW, 9 aa; SEPGGGGW, 8 aa; and EPGGGGW, 7 aa) also induced lysogeny, albeit to a lesser extent (Figures 6D–6F), possibly because of lower affinity to the cognate AimR. These results show that some arbitrium systems utilize peptides longer than 6 aa, demonstrating diverse peptide codes for these phage communication systems.

Based on these results, we predicted the consensus sequences of the mature arbitrium peptides for clades 4–8 (Figure 3). The arbitrium consensus signatures in these clades resemble the mature arbitrium that we verified in the Waukesha92-like phage that belongs to clade 9. These peptides are characterized by an invariant acidic residue (aspartate or glutamate) followed by proline and glycine. In the Waukesha92-like arbitrium, these conserved residues are found at positions 4–6 of the mature peptide, and we hence predicted the mature peptides of the other clades accordingly. However, we note that these predictions are yet to be verified experimentally and that the actual mature peptides may be shorter than what we predicted. In addition, as opposed to most clades in which the peptide size seems to be conserved across the clade, we observed fluctuations in the sizes of the mature peptides in clade 9, possibly reflecting relaxation of the structural constraints imposed by AimR on the evolution of peptide sizes (Figure S3).

The AimX Effector Transcript

It was previously demonstrated that in phage phi3T, the non-coding RNA AimX functions as a negative regulator of lysogeny (Erez et al., 2017). However, the molecular mechanism by which AimX executes the downstream lysogeny decision is unknown. To search for a cognate *aimX* in phage Wbeta, we performed whole-transcriptome RNA sequencing (RNA-seq) on total RNA extracted from cells infected by Wbeta in the presence or absence of externally added arbitrium peptide. As in the case of phi3T, the most notable change in expression was that of a transcript encoded immediately downstream to the *aimR-aimP* locus. This transcript showed substantial expression in the absence of arbitrium peptide, but its expression was completely

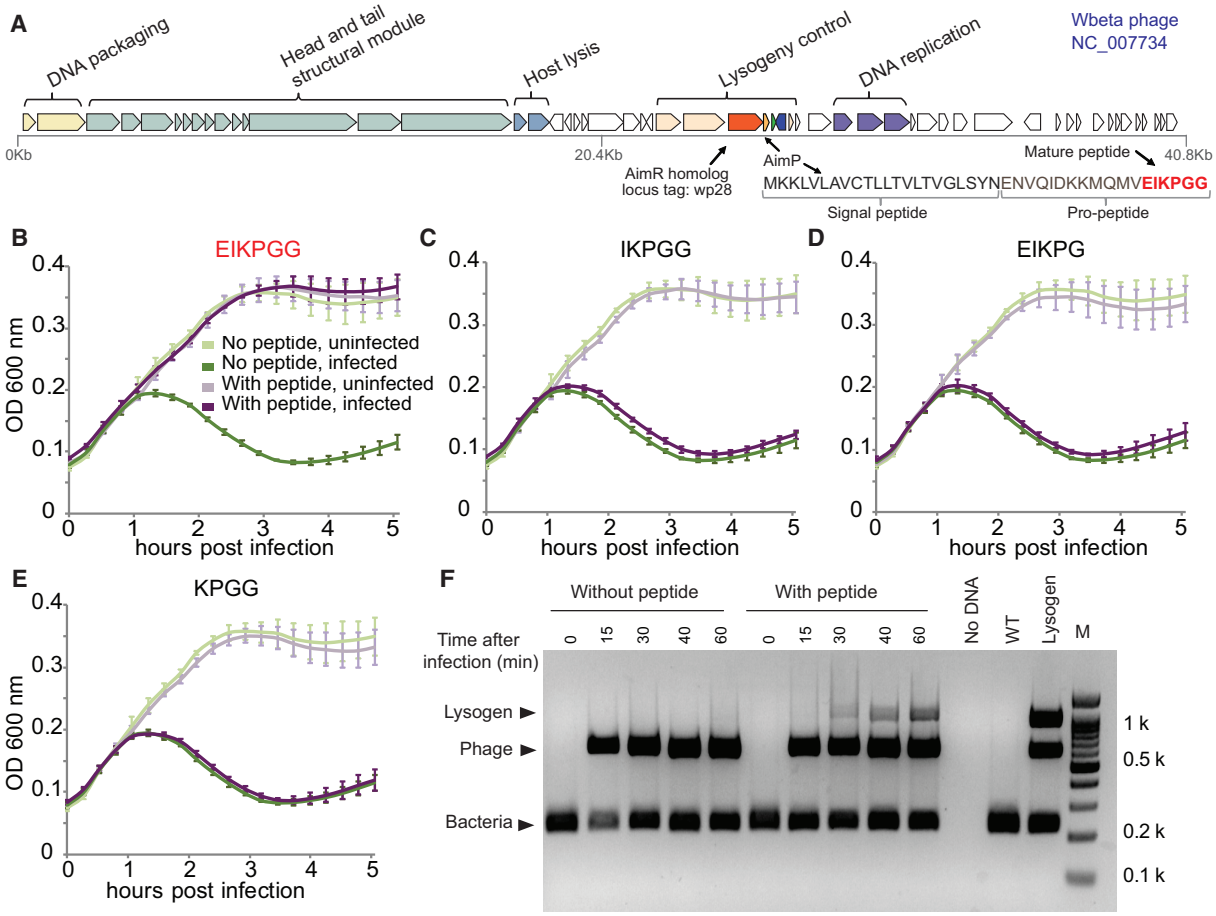


Figure 5. Arbitrium Controls Phage Wbeta Lysogeny in *B. cereus*

(A) A schematic representation of the phage Wbeta genome. Different colors represent functional modules, with the *aimR* homolog colored dark orange, *aimP* light orange, *aimX* green, and the *cI* repressor dark blue. Sequence of the AimP pre-pro-peptide is indicated, with the signal peptide position predicted using TMHMM (Krogh et al., 2001).

(B–E) Growth curves of *Bacillus cereus* RSVF1 infected by Wbeta with multiplicity of infection (MOI) of 10. Phages were added at t = 0. The sequence of the synthetic peptide supplemented to the medium (1 μM) is indicated above each graph. Data represent an average of three replicates using separate colonies, each with two technical replicates. Error bars are standard deviation of the means. All experiments were performed simultaneously with single "no peptide, uninfected" and "no peptide, infected" controls. (B) External addition of the synthetic peptide EIKPGG. (C) External addition of the synthetic peptide IKPGG. (D) External addition of the synthetic peptide EIKPG. (E) External addition of the synthetic peptide KPGG.

(F) Semi-quantitative PCR assay for phage lysogeny during an infection time course of *B. cereus* RSVF1 with phage Wbeta with or without external addition of the synthetic peptide EIKPGG. No DNA, control without DNA; WT, DNA from uninfected culture; lysogen, genomic DNA of a Wbeta lysogen; and M, DNA ladder with sizes indicated on the right.

silenced when the medium was supplemented with 1 μM of the EIKPGG peptide (Figure 7). This transcript therefore represents the AimX of Wbeta.

Similar to the case of phi3T, the Wbeta AimX transcript contains a short (40 aa) open reading frame (ORF), followed by a long stretch of non-coding RNA sequence (Figure 7A). Interestingly, the Wbeta *aimX* non-coding transcript overlaps the phage *cI* repressor gene, which is encoded in the same locus on the opposite DNA strand (Figure 7A). In the lytic condition (no peptide), the expression of AimX was 15- to 20-fold higher than that of the phage *cI* repressor, while in the lysogenic condition (with peptide), the expression of the phage repressor dominated (Figure 7B). The same expression patterns were observed for the Waukesha92-like phage (Figure S4).

When examining the genomic vicinity of the arbitrium systems we identified in phages and other mobile elements, we found that 80% of these systems were encoded directly upstream to a *cI* repressor-like gene encoded on the opposite strand, similar to the genomic organization of the Wbeta and Waukesha92-like phages (Table S1; also see Figure 4). This suggests that most of the arbitrium systems execute the lysogeny decision through AimX-mediated regulation of the *cI* repressor, which is the master regulator of lysogeny (Figures 7C and 7D; see Discussion).

DISCUSSION

In this study, we have shown that arbitrium phage communication systems are highly diverse and are present in numerous phages

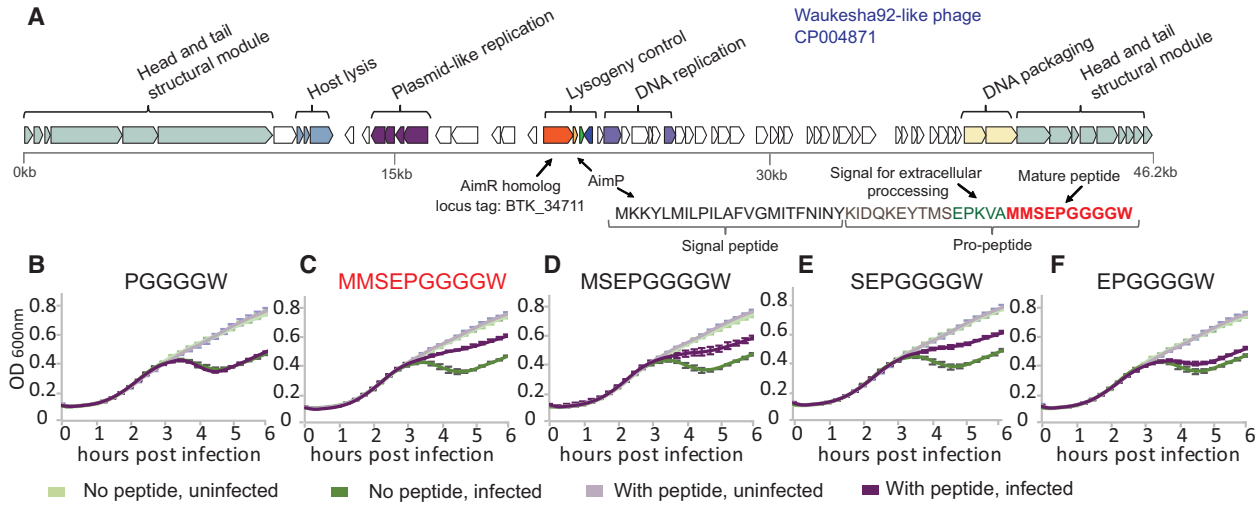


Figure 6. Arbitrium Controls Lysogeny in a Waukesha92-like Phage

(A) A schematic representation of the Waukesha92-like prophage genome. This prophage is found as a circular episome (and is therefore annotated as a plasmid) in the sequenced genome of *B. thuringiensis* sv. kurstaki HD-1. Genes in the lysogeny control module are colored according to their function, with the *aimR* homolog colored dark orange, *aimP* light orange, *aimX* green, and the *cI* repressor dark blue. Sequence of the AimP pre-pro-peptide is indicated, with the signal peptide position predicted using TMHMM (Krogh et al., 2001).

(B–F) Growth curves of *B. thuringiensis* sv. galleriae BGSC:4G5 infected by the Waukesha92-like phage with MOI of 0.1. Phages were added at t = 0. The sequence of the synthetic peptide supplemented to the medium (1 μM) is indicated above each graph. Data represent an average of three replicates using separate colonies, each with two technical replicates. Error bars are standard deviation of the means. All experiments were performed simultaneously with single "no peptide, uninfected" and "no peptide, infected" controls. (B) External addition of the synthetic peptide PGGGGW. (C) External addition of the synthetic peptide MMSEPGGGGW. (D) External addition of the synthetic peptide MSEPGGGGW. (E) External addition of the synthetic peptide SEPGGGGW. (F) External addition of the synthetic peptide EPGGGGW.

See also Figure S2.

and prophages. We found arbitrium and arbitrium-like systems within phages infecting many bacterial species, most of which are pathogenic and soil *Bacilli*. While it is possible that communication-based lysis-lysogeny decisions are broadly utilized by phages in nature, curiously, homologs of AimR were found only within the phylogenetic premises of the *Bacillales* order. Quorum sensing in this group of bacteria is based on small peptides (Pottathil and Lazazzera, 2003), and the phages seem to have taken advantage of specific characteristics of this quorum sensing, including the utilization of the OPP channel and extracellular proteases, for implementing their own communication system. It is possible that the arbitrium system is limited to phages infecting *Bacillales* because of its dependence on these characteristics. Alternatively, perhaps arbitrium systems are present in phages infecting other bacteria, but these systems have evolved beyond the detection capabilities of homology-based searches because of the rapid evolution of phages (Duffy and Turner, 2008).

In 16% of the cases, arbitrium systems were found to be encoded by conjugative elements. Such elements are known to utilize quorum-sensing molecules, including N-acylhomoserine lactones (Fuqua and Winans, 1994; Ramsay et al., 2009) and short peptides (Auchtung et al., 2005; Dunny and Berntsson, 2016; Singh et al., 2013), to control their transfer among bacteria. For example, in the peptide-based communication systems encoded by the *B.subtilis* integrative-conjugative element ICEBs1 (Auchtung et al., 2005) and conjugative plasmid pLS20 (Singh et al., 2013), conjugation is repressed when the peptide accumulates. The peptide is used as a signal that the majority of the population already contains the conjugative element, such that transfer will be repressed when the conjugative element is abundantly present in the nearby bacteria (Auchtung et al., 2005; Singh and Meijer, 2014; Singh et al., 2013). This logic is somewhat similar to that of the arbitrium system, in which lytic propagation of the phage is inhibited upon accumulation of the peptide. We hypothesize that the arbitrium systems that we found on conjugative elements function to serve a similar purpose, especially as these systems are found in the genomic vicinity of genes of the conjugation machinery as well as genes resembling the *cI* repressor (Table S1; Figures 2 and S1).

Notably, arbitrium systems are especially abundant in phages and mobile elements infecting pathogens such as *B. thuringiensis* and *B. anthracis*. *B. thuringiensis* is a common pathogen of insects and produces a family of insecticidal proteins called Bt toxins (Schnepf et al., 1998). These toxins play an important biotechnological role, with key crops genetically engineered to express such toxins for protection against insect pests (Romeis et al., 2006). We found several mobile elements in *B. thuringiensis* that carry Bt toxins as well as an arbitrium system (Figures 2A and S1). It is therefore likely that arbitrium communication is important for controlling horizontal transfer of these plasmids among bacteria. In *B. anthracis*, one of the two key virulence plasmids, called pXO2, encodes an arbitrium system with the predicted peptide KMVGPGGGMG. This plasmid carries genes responsible for synthesis of the anti-phagocytic poly-D-glutamic acid capsule that is essential for *B. anthracis* virulence (Dixon et al., 1999) (Figure 2B). In the reference strain *B. anthracis* Ames Ancestor (Ravel et al., 2009), AimR

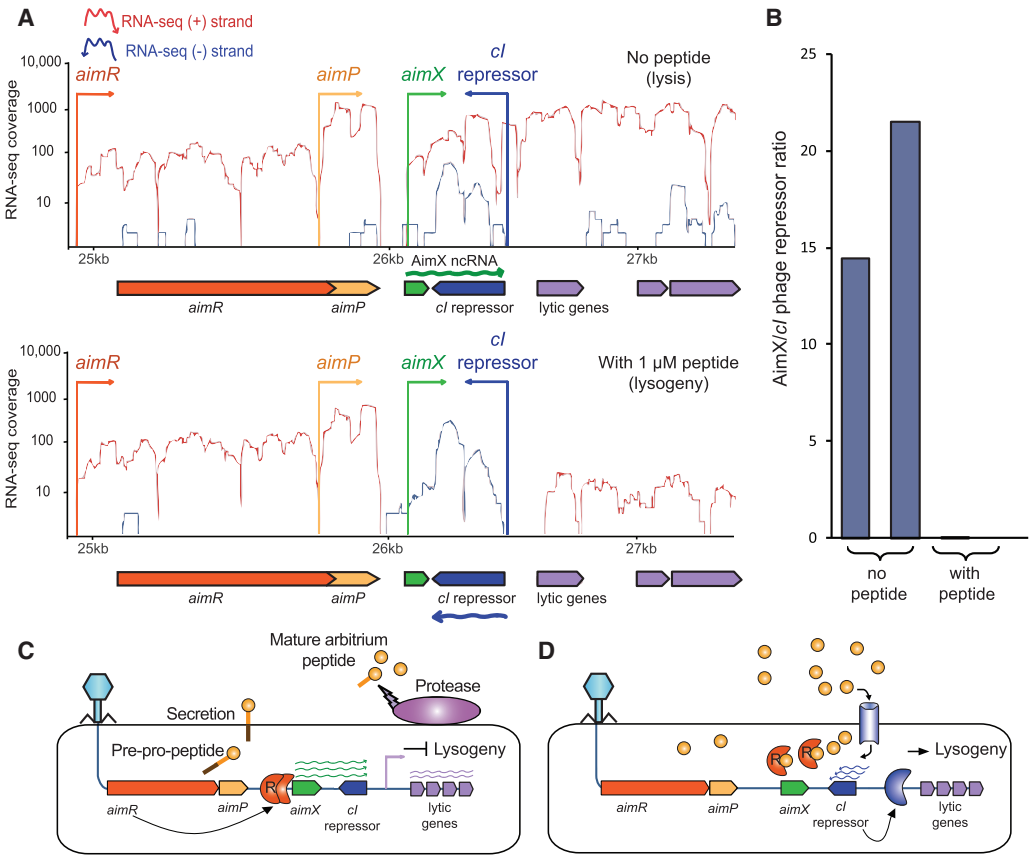


Figure 7. The AimX Effector Transcript Overlaps the Phage *cI* Repressor

(A) RNA-seq coverage of the arbitrium locus in phage Wbeta, 20 min post infection of *Bacillus cereus* RSVF1 at MOI = 10. Top, no peptide supplemented; bottom, with 1 μM synthesized EIKPGG peptide added to the growth medium.

(B) Ratio between the expression of *aimX* and the *cI* repressor, calculated as the ratio between the RPKM values of the sense and antisense strands. Data for two replicates are shown.

(C and D) Mechanistic model for the arbitrium-mediated lysis-lysogeny decision via peptide communication controlling *cis*-antisense regulation of the phage *cI* repressor. (C) At the first encounter of the phage with a bacterial population, when the concentration of the mature arbitrium peptide is low, AimR is free to activate the transcription of *aimX*, and AimX silences the expression of the phage *cI* repressor. (D) At later stages, when the concentration of the mature arbitrium peptide is high, AimR binds the arbitrium peptide, thus preventing the expression of *aimX*. In the absence of AimX the phage repressor is expressed, leading to suppression of lytic genes and establishing stable lysogeny.

See also Figure S4.

contains a frameshift mutation that likely inactivates it; yet in other *B. anthracis* strains, the arbitrium system in pXO2 is intact (Figure 2B). If the arbitrium system indeed regulates the transfer of these virulence plasmids among *B. anthracis* strains, then the arbitrium peptide could be potentially used for limiting such transfer in clinical or field settings.

About 5% of the AimR homologs we detected in this study showed only distant homology to AimR and localized within the gray-colored clade in the AimR phylogenetic tree (Figure 3). These homologs are found in diverse bacterial genera, including *Virgibacillus*, *Halobacillus*, and *Gracilibacillus*, and most of them are not found within mobile elements (Figure 3; Table S1). We consider these homologs as possibly encoding bacterial-peptide-based quorum-sensing systems that share a distant common ancestry with phage arbitrium systems.

Our transcriptional RNA-seq data (Figures 7A, 7B, and S4) may suggest that AimX in the Wbeta and Waukesha92-like phages is a *cis*-antisense RNA transcript that overlaps the transcript of the *cI* repressor. Such *cis*-antisense transcripts are known to silence the genes that they overlap, either by suppressing translation or by stimulating RNA degradation (Sesto et al., 2013). In the lytic condition, transcription of AimX dominates that of the phage *cI* repressor (Figure 7B), and we hypothesize that through *cis*-antisense regulation, AimX silences the expression of the *cI* repressor. This silencing is, in turn, alleviated in the presence of the arbitrium peptide, which leads to the suppression of AimX expression.

These results suggest a more complete model for the arbitrium-mediated lysogeny decision (Figures 7C and 7D). At the first encounter of the phage with a bacterial population, *aimR* and *aimP* are expressed immediately upon infection. AimR functions as a transcriptional regulator and activates the transcription of *aimX*, which blocks lysogeny by directly inhibiting the phage *cI* repressor using a *cis*-antisense mechanism. AimP is released into the medium and processed into the mature arbitrium peptide, which is internalized into neighboring cells via the

OPP channel. At later stages, phages will infect cells in which the intracellular concentration of the mature arbitrium peptide is high, and in these cases their AimR binds the arbitrium peptide, thus preventing the expression of *aimX*. Subsequently, the phage repressor will be expressed and will bind the promoters of lytic genes, suppressing their expression and establishing stable lysogeny.

Antisense-based silencing is also known to play an important role in the lysogeny decision of phage Lambda, where, for example, the stability of the *cII* transcript is regulated by a short antisense RNA (OOP) that promotes RNaseIII-mediated degradation (Krinke and Wulff, 1990; Oppenheim et al., 2005; Spiegelman et al., 1972), highlighting a common role for antisense regulation in executing phage lysogeny decisions. Our data imply that most of the arbitrium-carrying phages execute their lysogeny decisions using the *aimX-cI* direct *cis*-antisense regulation because of the similar organization of the lysogeny module in their genomes (*aimR-aimP-aimX-cI*). However, in clade 2 arbitrium systems found in SPbeta-like phages, the *cI* repressor is found elsewhere in the genome, distant from the arbitrium locus. We hypothesize that in these cases, the AimX non-coding RNA regulates the *cI* repressor in *trans*, possibly through a base-pairing mechanism typical to non-coding RNA regulation in bacteria (Waters and Storz, 2009).

Curiously, the AimX transcript also contains, in most cases, a conserved short reading frame (Figures 7A and S4). Such a small protein-coding gene is also present on the phi3T AimX transcript, although it is lacking from the SPbeta AimX (Erez et al., 2017). While the role of these small genes is currently unknown, we hypothesize that they may serve as antirepressors, which are frequently used by diverse phages to sequester the CI repressor protein by binding it and causing it to dissociate from the DNA (Lemire et al., 2011; Susskind and Botstein, 1975). If this hypothesis is correct, then *aimX* encodes two separate functions that together enforce tight regulation of the *cI* repressor, both on the RNA level and on the protein level.

The arbitrium system was initially shown to regulate the lysis-lysogeny decision of phages from the SPbeta group infecting *B. subtilis* (Erez et al., 2017). In the current study, we show that such communication systems are highly abundant among phages that infect soil and pathogenic *Bacilli*, utilizing a diverse space of peptide communication codes. Recently, host-derived quorum-sensing molecules were shown to regulate the lysogeny decisions of phages infecting *Vibrio cholera* (Silpe and Bassler, 2019), highlighting quorum sensing as an important aspect in phage biology that has so far been overlooked. We anticipate that future studies may expose more roles for small-molecule communication in coordinating group behavior of phages.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2019.03.017.

ACKNOWLEDGMENTS

We thank M. Voichek for assistance in the RNA-seq experiments, V. Fischetti for kindly contributing the Wbeta lysogen and its host, T. Dagan for advice in phylogenetic analyses, and A. Eldar and the Sorek lab members for fruitful discussion. R.S. was supported, in part, by the Israel Science Foundation (personal grant 1360/16 and I-CORE grant 1796/12), the European Research Council (ERC) (grant ERC-CoG 681203), the German Research Council (DFG) priority program SPP 2002 (grant SO 1611/1-1), the David and Fela Shapell Family Foundation, and the Benoziyo Advancement of Science grant.

AUTHOR CONTRIBUTIONS

A.S.-A. performed the computational analyses. N.T. performed the experimental assays and their analysis. Z.E. and A.L. assisted in designing experiments and reviewed the manuscript. R.S. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 30, 2018
Revised: February 4, 2019
Accepted: March 25, 2019
Published: May 8, 2019

REFERENCES

Abedon, S.T. (2015). Bacteriophage secondary infection. Virol. Sin. *30*, 3–10.

Abshire, T.G., Brown, J.E., and Ezzell, J.W. (2005). Production and validation of the use of gamma phage for identification of Bacillus anthracis. J. Clin. Microbiol. *43*, 4780–4788.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

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.

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.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Chen, I.-M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res. *45*, D507–D516.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P., and Sorek, R. (2016). Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. Science *352*, aad9822.

Dixon, T.C., Meselson, M., Guillemin, J., and Hanna, P.C. (1999). Anthrax. N. Engl. J. Med. *341*, 815–826.

van Dongen, S. (2000). A Cluster Algorithm for Graphs. Technical Report INS-R0010 (National Research Institute for Mathematics and Computer Science in the Netherlands).

van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. Methods Mol. Biol. *804*, 281–295.

Dou, C., Xiong, J., Gu, Y., Yin, K., Wang, J., Hu, Y., Zhou, D., Fu, X., Qi, S., Zhu, X., et al. (2018). Structural and functional insights into the regulation of the lysis–lysogeny decision in viral communities. Nat. Microbiol. *3*, 1285–1294.

Duffy, S., and Turner, P.E. (2008). Phage evolutionary biology. In Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses (Cambridge University Press), pp. 147–176.

Dunny, G.M., and Berntsson, R.P.A. (2016). Enterococcal sex pheromones: evolutionary pathways to complex, two-signal systems. J. Bacteriol. *198*, 1556–1562.

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.

Fuqua, W.C., and Winans, S.C. (1994). A LuxR-LuxI type regulatory system activates Agrobacterium Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. J. Bacteriol. *176*, 2796–2806.

Gallego del Sol, F., Penadés, J.R., and Marina, A. (2019). Deciphering the molecular mechanism underpinning phage arbitrium communication systems. Mol. Cell. *74*, 59–72.e3.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. *11*, 1511–1520.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. https://doi.org/10.1093/bib/bbx108.

Krinke, L., and Wulff, D.L. (1990). RNase III-dependent hydrolysis of lambda cII-O gene mRNA mediated by lambda OOP antisense RNA. Genes Dev. *4*, 2223–2233.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Pottathil, M., and Lazazzera, B.A. (2003). The extracellular Phr peptide-Rap phosphatase signaling circuit of bacillus subtilis. Front. Biosci. *8*, d32–d45.

Lemire, S., Figueroa-Bossi, N., and Bossi, L. (2011). Bacteriophage crosstalk: coordination of prophage induction by trans-acting antirepressors. PLoS Genet. *7*, e1002149.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44*, W242–W245.

Moumen, B., Nguen-The, C., and Sorokin, A. (2012). Sequence analysis of inducible prophage phIS3501 integrated into the haemolysin II gene of Bacillus thuringiensis var israelensis ATCC35646. Genet. Res. Int. *2012*, 543286.

Nguyen, L.-T.L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

Ofir, G., and Sorek, R. (2018). Contemporary phage biology: from classic models to new insights. Cell *172*, 1260–1270.

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in bacteriophage Lambda development. Annu. Rev. Genet. *39*, 409–429.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Ptashne, M. (2004). A Genetic Switch: Phage Lambda Revisited (Cold Spring Harbor Laboratory Press).

Ramsay, J.P., Sullivan, J.T., Jambari, N., Ortori, C.A., Heeb, S., Williams, P., Barrett, D.A., Lamont, L.L.I.L., and Ronson, C.W. (2009). A LuxRI-family regulatory system controls excision and transfer of the Mesorhizobium loti strain R7A symbiosis island by activating expression of two conserved hypothetical genes. Mol. Microbiol. *73*, 1141–1155.

Ravel, J., Jiang, L., Stanley, S.T., Wilson, M.R., Decker, R.S., Read, T.D., Worsham, P., Keim, P.S., Salzberg, S.L., Fraser-Liggett, C.M., et al. (2009). The complete genome sequence of Bacillus anthracis Ames "ancestor". J. Bacteriol. *191*, 445–446.

Romeis, J., Meissle, M., and Bigler, F. (2006). Transgenic crops expressing Bacillus thuringiensis toxins and biological control. Nat. Biotechnol. *24*, 63–71.

Sauder, A.B., Carter, B., Langouet Astrie, C., and Temple, L. (2014). Complete genome sequence of a mosaic bacteriophage, Waukesha92. Genome Announc. *2*, e00339–14.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., and Dean, D.H. (1998). Bacillus thuringiensis and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. *62*, 775–806.

Schuch, R., and Fischetti, V.A. (2006). Detailed genomic analysis of the Wbeta and gamma phages infecting Bacillus anthracis: implications for evolution of environmental fitness and antibiotic resistance. J. Bacteriol. *188*, 3037–3051.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev. Microbiol. *11*, 75–82.

Silpe, J.E., and Bassler, B.L. (2019). A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. Cell *176*, 268–280.e13.

Singh, P.K., and Meijer, W.J.J. (2014). Diverse regulatory circuits for transfer of conjugative elements. FEMS Microbiol. Lett. *358*, 119–128.

Singh, P.K., Ramachandran, G., Ramos-Ruiz, R., Peiró-Pastor, R., Abia, D., Wu, L.J., and Meijer, W.J.J. (2013). Mobility of the native Bacillus subtilis conjugative plasmid pLS20 is regulated by intercellular signaling. PLoS Genet. *9*, e1003892.

Spiegelman, W.G., Reichardt, L.F., Yaniv, M., Heinemann, S.F., Kaiser, A.D., and Eisen, H. (1972). Bidirectional transcription and the regulation of phage Lambda repressor synthesis. Proc. Natl. Acad. Sci. USA *69*, 3156–3160.

Susskind, M.M., and Botstein, D. (1975). Mechanism of action of Salmonella phage P22 antirepressor. J. Mol. Biol. *98*, 413–424.

Wang, Q., Guan, Z., Pei, K., Wang, J., Liu, Z., Yin, P., Peng, D., and Zou, T. (2018). Structural basis of the arbitrium peptide-AimR communication system in the phage lysis-lysogeny decision. Nat. Microbiol. *3*, 1266–1273.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136*, 615–628.

Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell *141*, 682–691.

Zhen, X., Zhou, H., Ding, W., Zhou, B., Xu, X., Perčulija, V., Chen, C.-J.C.J., Chang, M.-X.M.X., Choudhary, M.I., and Ouyang, S. (2019). Structural basis of AimP signaling molecule recognition by AimR in Spbeta group of bacteriophages. Protein Cell *10*, 131–136.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Bacillus cereus W ATCC11950	ATCC	ATCC: 11950
Bacillus thuringiensis sv. kurstaki HD-1 BGSC: 4D1	Bacillus Genetic Stock Center (BGSC)	BGSC: 4D1
Bacillus cereus RSVF1 ATCC4342	ATCC	ATCC: 4342
Bacillus thuringiensis sv. galleriae BGSC: 4G5	Bacillus Genetic Stock Center (BGSC)	BGSC: 4G5
Wbeta phage	This paper	N/A
Waukesha92-like phage	This paper	N/A
Chemicals, Peptides, and Recombinant Proteins		
Synthetic peptide EIKPGG	Genscript Corp	N/A
Synthetic peptide EIKPG	Genscript Corp	N/A
Synthetic peptide IKPGG	Genscript Corp	N/A
Synthetic peptide KPGG	Genscript Corp	N/A
Synthetic peptide PGGGGW	Genscript Corp	N/A
Synthetic peptide MMSEPGGGGW	Genscript Corp	N/A
Synthetic peptide MSEPGGGGW	Genscript Corp	N/A
Synthetic peptide SEPGGGGW	Genscript Corp	N/A
Synthetic peptide EPGGGGW	Genscript Corp	N/A
Phosphomycin disodium salt	Sigma-Aldrich	P5396; CAS: 26016-99-9
Mitomycin C	Sigma-Aldrich	M0503; CAS: 50-07-7
Critical Commercial Assays		
Turbo DNA free kit	Life Technologies	Cat#AM2238
Fastprep RNA protect solution	MP Biomedicals	Cat#6055050
Fastprep homogenizer	MP Biomedicals	Cat#6911050
FastRNA PRO blue kit	MP Biomedicals	Cat#116025050
Ambion fragmentation buffer	Invitrogen	Cat#10136824
RiboZero rRNA Removal Kit	Epicentre	Cat#MRZB12424
NEBNext Ultra Directional RNA Library Prep Kit	NEB	Cat#E7420
Illumina NextSeq500	Illumina	N/A
Deposited Data		
RNA-seq data	This paper	ENA: PRJEB31642; http://www.ebi.ac.uk/ena/data/view/PRJEB31642
Oligonucleotides		
GCCTTTTTTACGCAAATCTCC	Sigma-Aldrich	Forward primer for detecting the Waukesha92-like phage genome
GCCCTCTAAAATCAAGAAATGC	Sigma-Aldrich	Reverse primer for detecting the Waukesha92-like phage genome
GTCCTAGGAAAAAGAGGAGG	Sigma-Aldrich	Forward primer for detecting *B. cereus* RSVF1 genome
CTCGTCATACGTCTGTCGAG	Sigma-Aldrich	Reverse primer for detecting *B. cereus* RSVF1 genome
TCCTTATGTACTTGGCGGAG	Sigma-Aldrich	Forward primer for detecting Wbeta phage genome
AATCCGTTTGTAACGAATGG	Sigma-Aldrich	Reverse primer for detecting Wbeta phage genome
AATTGTATCTCCAACAGCAG	Sigma-Aldrich	Forward primer for detecting Wbeta prophage integration
TCTCGGATTGTTCCAGTTTC	Sigma-Aldrich	Reverse primer for detecting Wbeta prophage integration

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
PSI-BLAST version 2.5.0 of BLAST+	Altschul et al., 1997; Camacho et al., 2009	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
PHASTER URLAPI	Arndt et al., 2016	http://phaster.ca/instructions
The Integrated Microbial Genomes (IMG)	Chen et al., 2017	https://img.jgi.doe.gov/cgi-bin/m/main.cgi
IQ-TREE version 1.6.5	Nguyen et al., 2015	http://www.iqtree.org
MAFFT online server	Katoh et al., 2002; Katoh et al., 2017	https://mafft.cbrc.jp/alignment/server/
iTOL	Letunic and Bork, 2016	https://itol.embl.de/
SignalP 4.1	Petersen et al., 2011	http://www.cbs.dtu.dk/services/SignalP/
Weblogo 2.8.2	Crooks et al., 2004	https://weblogo.berkeley.edu/
MCL algorithm version 14-137	van Dongen, 2000; van Dongen and Abreu-Goodger, 2012	https://micans.org/mcl/
NovoAlign (Novocraft) v3.02.02		http://www.novocraft.com/documentation/novoalign-2/novoalign-ngs-quick-start-tutorial/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rotem Sorek (rotem.sorek@weizmann.ac.il).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The strains and phages used in this study are listed in the Key Resources Table. The bacterial strains and media used in this study are: *Bacillus cereus* W ATCC11950, cultivated in brain heart infusion (BHI, BD cat. #237500) at 30°C; *Bacillus cereus* RSVF1 ATCC4342, cultivated in BHI or nutrient broth at 30°C or 37°C (as stated below); *Bacillus thuringiensis* sv. kurstaki HD-1 BGSC:4D1, cultivated in LB at 30°C; and *Bacillus thuringiensis* sv. galleriae BGSC:4G5 (cured of a prophage-like plasmid, see below), cultivated in MMB (LB media supplemented with 0.1 mM MnCl2 and 5 mM MgCl2) at 30°C. The phages used in this study are phage Wbeta and the Waukesha92-like phage.

METHOD DETAILS

Identification of AimR Homologs

AimR homologs were identified using PSI-BLAST version 2.5.0 of BLAST+ (Altschul et al., 1997; Camacho et al., 2009) with the phi3T phage AimR protein ("GenBank: APD21232.1") as the query sequence. The search was performed in four iterations with an e-value cutoff of 1e-10 against a database of proteins from bacterial, archaeal and viral genomes downloaded from the NCBI FTP site on April 2016 (bacteria and archaea) and May 2018 (viruses) (ftp://ftp.ncbi.nih.gov/genomes/genbank/bacteria/, ftp://ftp.ncbi.nih.gov/genomes/genbank/archaea/ and ftp://ftp.ncbi.nih.gov/genomes/genbank/viral/, respectively). The resulting 1327 hits were cross-referenced against the Integrated Microbial Genomes (IMG) database (Chen et al., 2017) using BLASTN to identify genes with 100% identity and 100% query coverage in the same strain. This resulted in 1257 hits for which a corresponding gene was identified in the IMG database. Proteins shorter than 300 amino acids were removed from the dataset, as possibly encoding truncated forms of the AimR protein. Finally, 12 proteins found in viral genomes that were not in IMG were added to our dataset, overall resulting in a final list of 1180 genes that were used for further analyses.

Clustering of AimR Homologs

Multiple sequence alignment (MSA) of the amino acid sequences of the 1180 AimR homologs was performed using the MAFFT online server (Katoh et al., 2002; Katoh et al., 2017) version 7 with default parameters. The resulting MSA was used to create a maximum likelihood phylogenetic tree using IQ-TREE (Nguyen et al., 2015) version 1.6.5 including ultrafast bootstrap analysis (Hoang et al., 2018) with 1000 alignments ("-bb 1000") and using the "-m TEST" parameter to test for the best substitution model (Kalyaanamoorthy et al., 2017). The tree was visualized using the iTOL program (Letunic and Bork, 2016) and rooted between clades 4-9 and the gray clade for display purposes. Clades with an average branch length distance to leaves below 10^{-6} were collapsed.

Identification of *aimP* Genes

To identify the putative *aimP* genes associated with each of the *aimR* homologs, we searched for an ORF harboring a putative signal peptide sequence immediately downstream to the *aimR* gene. To this end, the 100 terminal bases of each *aimR* homolog and 500 additional bases downstream to it were extracted. Then, ORFs of 15 or more amino acids were searched within the sequence, including ORFs that begin with alternative start codons (GTG or TTG). The identified ORFs were submitted to SignalP 4.1 (Petersen et al., 2011) using the ''gram-positive bacteria organism group'' option and a minimal D-cutoff value of 0.3, to identify possible N-terminal signal peptide sequences. If several ORFs with a signal peptide were identified, the one with the highest D-cutoff value was chosen. The C-terminal end of the putative AimP sequences were collected and unique sequences were used to generate a sequence logo using the Weblogo 2.8.2 web server (Crooks et al., 2004). For 13 of the AimP sequences, which belong to clade 1, an internal peptide (rather than a C-terminal one) was predicted based on homology to the C-terminal end of the rest of the AimP sequences of the clade.

The Genomic Context of Arbitrium Systems

To check whether the arbitrium systems are located within a prophage, the PHASTER URL API (Arndt et al., 2016) was used. The entire DNA sequence of the contig in which the *aimR* homolog was located was submitted as a query to identify prophage regions, and cases where the *aimR* was included in such a region were recorded. In addition, the genomic neighborhood for each *aimR* homolog was examined using the gene neighborhoods function in IMG (Chen et al., 2017), and the *aimR* was considered to be found in a prophage if it was located next to genes annotated as phage genes. In cases where we were unable to detect phage structural genes and the gene was found near genes annotated as encoding conjugation proteins, the gene was considered to be found in a conjugative element. Genes with the following annotations in IMG were considered as conjugation proteins: Type IV secretory pathway proteins (VirD4, VirB2, VirB4 and VirB6), conjugative transposon proteins (TcpC and TcpE), pilus assembly proteins (CpaB and CpaF), TraG/TraD conjugation proteins or TraM recognition site of TraD and TraG.

To calculate the gene distance of each *aimR* homolog to a phage integrase gene, five neighboring genes from each side of the *aimR* homolog were inspected. An integrase gene was identified as a gene annotated in IMG as an Integrase or Recombinase in one of the following fields: Gene Product Name, COG, Pfam, KEGG Orthology ID, Name or Definition. Putative prophage edges (marking its integration site within the bacterial genome) were identified by genomic comparison to a closely related bacterial strain that lacks the integrated prophage. As the integrase gene is typically present at the edge of the integrated prophage, if a phage integrase was located at the edge of the prophage it was considered the *bona fide* integrase, and others were disregarded.

To identify putative *cI* repressor genes associated with the arbitrium system, three neighboring genes downstream to the *aimR* homolog were inspected. A repressor was identified as a protein of up to 250 aa on the opposite strand of the *aimR* gene that was annotated as a transcriptional regulator in IMG. This includes annotations of helix-turn-helix (HTH), transcriptional regulator, or DNA-binding in one of the following fields: Gene Product Name, COG, Pfam, IMG Term or the gene name of the top hit obtained using a BLASTP search against the nr DB.

Phage Type Determination

Prophage DNA sequences were extracted from the bacterial genomes and sequences longer than 10 kb were taken for further analysis. The extracted prophage sequences, as well as phage sequences that harbor arbitrium, were compared using an ''all vs. all'' tBLASTX with an e-value of 0.001 and default parameters. Hits with less than 40% query coverage were removed. The query coverage of the top hit for each phage pair was used as a similarity score between the pair, using the best value found between each pair. The phage sequences were then clustered, based on these similarity scores, into groups using the MCL algorithm (van Dongen, 2000; van Dongen and Abreu-Goodger, 2012) version 14-137 with an inflation value of 1.5.

Homology between arbitrium-containing prophages to known phages was searched using BLASTN against a database of viral genomes downloaded from NCBI on May 2018 with an e-value cutoff of 0.05. Hits with over 85% identity and at least 40% query coverage were taken into account for group identity determination. In addition, the ''Most Common Phage Name'' field in the PHASTER output for each prophage was used if the number of genes (hits) in that field was 18 or more. MCL-derived phage groups for which the majority of the members showed homology to a known group of phages were designated accordingly. Phage groups for which a strong enough hit to a known phage was not found were designated as PG (for Phage Group) with a serial number (Figure 1A).

Prophage Induction and Propagation

Bacillus cereus W ATCC11950, containing the Wbeta prophage, was obtained from ATCC. Phage Wbeta was induced by plating the lysogenic strain on a BHI agar plate containing 20 μg/ml phosphomycin, as described previously (Schuch and Fischetti, 2006). After overnight incubation at 30°C, ring-shaped colonies formed. The clearing zones were picked and streaked on a BHI soft agar (0.3%) plate containing *Bacillus cereus* RSVF1 ATCC4342 obtained from ATCC. Plates were incubated overnight at 30°C. Lysis zones were picked into phage buffer (50 mM Tris pH 7.4, 100 mM MgCl2, 10 mM NaCl). Further infection cycles were carried out using a liquid culture grown in nutrient broth (meat extract 1 g/L, peptone 5 g/L, sodium chloride 5 g/L, yeast extract 2 g/L) as described previously (Schuch and Fischetti, 2006) until sufficient titer was achieved.

The Waukesha92-like phage was induced from a liquid culture of *Bacillus thuringiensis* sv. kurstaki HD-1 BGSC:4D1 obtained from the Bacillus Genetic Stock Center (BGSC). The bacteria were streaked on an LB agar plate and a single colony was picked and grown in liquid LB medium at 30°C. When the bacterial culture reached an OD_{600} of 0.3, mitomycin C was added to a final concentration of

0.5 μg/ml. After overnight incubation, the culture was centrifuged for 10 minutes at 3200 g. Supernatant was collected and filtered through 0.2 μm filters (GE Healthcare Life Sciences, Whatman, cat. #10462200). Drop assay was used to test for plaque formation on 0.3% agar plates with MMB containing *B. thuringiensis* sv. galleriae BGSC:4G5 (obtained from BGSC) cured of a prophage-like plasmid. For curing, a number of *B. thuringiensis* sv. galleriae BGSC:4G5 colonies were picked, and whole genome sequencing was performed to confirm the strain identity and the absence of the prophage-like plasmid ("GenBank: CP010095.1"). A single plaque was picked from the drop assay into 50 μl phage buffer, and passaged two additional times through a single plaque to isolate the phage. Identity of phage was verified using PCR (primers are listed in the Key Resources Table). Phages were then propagated as follows: An overnight starter of the bacteria grown in MMB medium was diluted 1:100 into 2 ml of fresh MMB medium and grown at 30°C with shaking until reaching an OD_{600} of 0.3. The low-titer phage stock was added to the liquid culture and the culture was incubated for 5 hours with shaking at 30°C. The culture was centrifuged for 10 minutes at 3200 g and the supernatant was collected and then filtered with a 0.2 μm filter. This procedure was repeated until sufficient titer of the phage was obtained. Phage working stocks were stored at 4°C.

Synthetic Peptides

Lyophilized peptides were ordered from Genscript Corp, with purity ranging from 89.9% to 99.1%. The peptides of phage Wbeta were suspended in 100% DMSO and then gradually diluted with ultrapure water (UPW) to reach 2% DMSO and a peptide concentration of 10 mM. The peptides of the Waukesha92-like phage were suspended in UPW to a peptide concentration of 10 mM. Peptides were further diluted in UPW to obtain working stocks of 100 μM.

Phage Infection Dynamics

An overnight starter of the bacteria was diluted 1:100 into 5 ml of medium (MMB medium for *B. thuringiensis* sv. galleriae BGSC:4G5 when infected by the Waukesha92-like phage, and nutrient broth for *B. cereus* RSVF1 ATCC4342 when infected by Wbeta) and incubated at 30°C or 37°C (for *B. thuringiensis* sv. galleriae and *B. cereus* RSVF1, respectively) until reaching an OD_{600} of 0.1. The liquid culture was centrifuged for 10 minutes at 3200 g at room temperature, the supernatant was discarded, and the pellet was re-suspended in 0.5 ml of fresh medium. The concentrated bacterial culture was further resuspended in fresh medium at a ratio of 1:9. For the experiments that contained peptides, the synthetic peptide was added at a ratio of 1:100 (final volume) to a final concentration of 1 μM. Bacteria were incubated in the presence of the peptide for 1 hour at room temperature prior to infection.

Infection was performed for phage Wbeta at MOI=10 and for the Waukesha92-like phage at MOI=0.1, added in a volume that constituted 10% of the final volume. The OD_{600} was measured using a TECAN Infinite 200 plate reader in a 96-well plate at 30°C or 37°C for 10 hours. Throughout the infection experiments, *B. cereus* RSVF1 ATCC4342 was grown in nutrient broth at 37°C and *B. thuringiensis* sv. galleriae BGSC:4G5 was grown in MMB medium at 30°C.

Lysogeny Assay

For the experiment described in Figure 5F, an overnight culture of *B. cereus* RSVF1 ATCC4342 was diluted 1:100 into fresh nutrient broth and grown with shaking at 37°C until reaching OD_{600} of 0.1. The bacterial culture was centrifuged at 3200 g for 10 minutes and the pellet was re-suspended in an equivalent amount of fresh nutrient broth. The culture was subsequently divided into two flasks, one of which was supplemented with the synthetic peptide to reach a final concentration of 1 μM. Both flasks were incubated for 1 hour at room temperature with gentle shaking. Following incubation, both liquid cultures were infected with Wbeta at MOI=10. 2 ml was collected from each sample at different time points: 0, 15, 30, 40, and 60 minutes. Each sample was centrifuged immediately for 5 minutes at 3200 g and at 4°C. The pellet was frozen using dry ice and ethanol.

DNA extraction was done using the Qiagen DNeasy blood and tissue kit (cat. #69504). The extracted DNA of each sample was diluted to a final concentration of 8 ng/μl. Semi-quantitative PCR was performed using three pairs of primers (Key Resources Table), targeted to amplify the bacterial DNA, the viral DNA and the junction between the integrated phage genome and the bacterial DNA (Key Resources Table). PCR mix was made using KAPA HIFI HotStart ReadyMix. Annealing temperature was set to 60°C, elongation to 30 seconds and 26 cycles of amplification were performed.

RNA-seq

Whole transcriptome RNA sequencing was performed for *B. cereus* RSVF1 ATCC4342 (nutrient broth, 37°C, infected with the Wbeta phage at MOI=10) and *B. thuringiensis* sv. galleriae 4G5 (MMB medium, 30°C, infected with the Waukesha92-like phage at MOI=1). An overnight culture was diluted 1:100 into 40 ml of fresh medium (either nutrient medium or MMB, depending on the bacterium) and grown with shaking until reaching an OD_{600} of 0.1. The bacterial culture was centrifuged at 3200 g for 10 minutes and the pellet was re-suspended in 40 ml of fresh medium. The culture was then divided into two flasks, one of which was supplemented with the synthetic peptide to reach a final concentration of 1 μM. Both flasks were incubated with gentle shaking for 1 hour at room temperature. Then, both liquid cultures were infected with the relevant phage. 5 ml were collected from each sample (at 20 minutes post infection for the Wbeta phage; and at 120 minutes post infection for the Waukesha92-like phage). Each sample was centrifuged immediately for 5 minutes at 3200 g and at 4°C. The pellet was flash frozen using dry ice and ethanol.

RNA extraction was performed as described previously (Dar et al., 2016). Briefly, pellets were re-suspended in 1 ml of RNA protect solution (FastPrep) and lysed by Fastprep homogenizer (MP Biomedicals). RNA was extracted using the FastRNA PRO blue kit (MP Biomedicals, 116025050) according to the manufacturer's instructions. DNase treatment was performed using the Turbo DNA free kit

(Life Technologies, AM2238). RNA was subsequently fragmented using fragmentation buffer (Ambion-Invitrogen, cat. #10136824) at 72°C for 1 min 45 sec. The reactions were cleaned by adding ×2.5 SPRI beads (Agencourt AMPure XP, Beckman-Coulter, A63881). The beads were washed twice with 80% ethanol and air dried for 5 minutes. The RNA was eluted using H_2O. Ribosomal RNA was depleted by using the Ribo-Zero rRNA Removal Kit (epicentre, MRZB12424). Strand-pecific RNA-seq was performed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, E7420) with the following adjustments: all cleanup stages were performed using ×1.8 SPRI beads, and only one cleanup step was performed after the end repair step. Following sequencing on an Illumina NextSeq500, sequenced reads were demultiplexed and adapters were trimmed using 'fastx clipper' with default parameters. Reads were mapped to the phage genome (Gene annotations and sequences were downloaded from "GenBank: ABCZ02000003.1" for the Wbeta prophage within the genome of *B. cereus* W, and ''GenBank: CP004871.1" for the Waukesha92-like prophage within the genome of *B. thuringiensis* sv. kurstaki HD-1 BGSC:4D1) by using NovoAlign (Novocraft) v3.02.02 with default parameters, discarding reads that were nonuniquely mapped as previously described (Dar et al., 2016). Reads per gene were calculated for each condition and biological repeat, and normalized using RPKM values.

QUANTIFICATION AND STATISTICAL ANALYSIS

All of the statistical details of the experiments can be found in the figure legends. Average +/- standard deviation of the mean of biological triplicates is shown throughout unless stated otherwise. For the phylogenetic analysis presented in Figure 3, ultrafast bootstrap support values (Hoang et al., 2018) were calculated based on 1000 iterations using IQ-TREE (Nguyen et al., 2015) and the substitution model was chosen using the ''-m TEST'' parameter (Kalyaanamoorthy et al., 2017).

DATA AND SOFTWARE AVAILABILITY

The RNA-sequencing raw data have been deposited in the European Nucleotide Archive (ENA) under the study accession number "ENA: PRJEB31642". Characteristics of arbitrium systems identified in this study are reported in Table S1.