A weaponized phage suppresses competitors in historical and modern metapopulations of pathogenic bacteria

Talia Backman[1], Sergio M. Latorre[2], Lauren Eads[1], Sarita Som[1], David Belnap[3], Allison M. Manuel[4], Hernán A. Burbano[2*†], Talia L Karasov[1*†]

Affiliations:

[1]School of Biological Sciences, University of Utah 257S 1400E Salt Lake City, UT USA 84112

[2]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London, UK

[3]Department of Biochemistry, University of Utah 15N Medical Drive Salt Lake City, UT USA 84112

[4]Mass Spectrometry and Proteomics Core, The University of Utah, Salt Lake City, UT, USA

*Equal contribution

†Corresponding author

ORCID:

TB 0000-0002-1967-270X, talia.backman@utah.edu

SML 0000-0002-5889-0670, s.latorre@ucl.ac.uk

LE 0000-0003-0966-739X, lauren.eads@utah.edu

SS 0000-0001-8717-3817, u1263424@utah.edu

DB 0000-0002-0619-7487, david.belnap@utah.edu

AM 0000-0003-2853-7005, allison.manuel@cores.utah.edu

HAB 0000-0003-3433-719X, h.burbano@ucl.ac.uk

TLK 0000-0002-3592-6597, t.karasov@utah.edu

Abstract

Bacteriophages are proposed to drive bacterial population dynamics, yet direct evidence of their impact on natural populations is limited. Here we identified phages in a metapopulation of wild plant-associated *Pseudomonas* genomes, including historical herbarium specimens. We discovered a viral cluster shared amongst pathogenic *Pseudomonas* isolates, which encodes a tailocin - a phage-derived element utilized for interbacterial warfare. Each pathogenic *Pseudomonas* strain carries one of a few distinct tailocin variants, which selectively kill other co-occurring pathogenic strains. Herbarium-derived genomes revealed that the same tailocin variants have persisted in the *Pseudomonas* populations for at least two centuries, suggesting the continued use of a defined set of variants. These results indicate that tailocin genetic diversity can be mined to develop targeted "tailocin cocktails" for microbial control.

Introduction

Understanding the factors that influence the spread of pathogens is fundamental for developing disease prevention and treatment strategies. To colonize a plant, bacterial pathogens need to overcome not only the plant's immune system, but also must compete with surrounding microbes and must defend against bacteriophages. Microbial disease is a frequent outcome in genetically homogeneous agricultural fields, but rarer in genetically diverse wild plant populations (*1*). These wild plant populations are instead typically colonized at low levels by diverse pathogens, with no strain spreading widely (*2*). What prevents the dominance and spread of single lineages as is so often seen in agriculture? The prominent role of host genetics in suppressing pathogenic lineages is well-documented (*3*), but how the surrounding microbiota affect disease dynamics remains largely unknown (though of increasing interest (*4*)).

Bacteriophages (henceforth phages), bacteria-infecting viruses, are the most abundant biological entities on earth (*5*). Phages can be highly specific in the bacteria they infect, often targeting only a few strains of a bacterial species (*6*) and are hypothesized to be drivers of microbiome composition in host and non-host environments (*7*). To date, the impact of phages on plant pathogens has been demonstrated in the lab under controlled conditions (*8*). Whether phages are a major suppressor of pathogen strains in the wild remains largely a matter of speculation (*7*) (but see (*9*, *10*)).

We previously found that wild populations of the plant *Arabidopsis thaliana* were colonized by a metapopulation of *Pseudomonas* bacterial pathogens. Our prior results suggested that these pathogen populations do not undergo clonal expansions as is frequently observed in agricultural and clinical pathogens. Infecting *Pseudomonas viridiflava* populations, even within an infection of a single plant, consisted of several co-occurring strains and no single strain rising to prominence (*2*). What prevents single pathogenic lineages from spreading? *A. thaliana* immune diversity certainly plays a role in maintaining *Pseudomonas* pathogen diversity; however, our previous work indicated that other members of the plant microbiome also affect the *Pseudomonas* composition (*11–13*). Because phages are common in *Pseudomonas* populations (*14*) we hypothesized that differences in phage sensitivity could suppress specific *Pseudomonas* strains in our system.

The viral cluster VC2 is strongly associated with a pathogenic clade of *P. viridiflava*

To determine whether phages cause turnover in the *Pseudomonas* colonizing *A. thaliana*, we first sought to characterize the abundant phages in a wild *Pseudomonas* metapopulation. Viral sequences integrated within a bacterial genome are the result of lysogeny with a compatible phage (*15*). Identifying compatible lysogeny events can reveal clade-specific phage enrichment.

We characterized the presence and evolution of viral elements in over 1500 *Pseudomonas* genomes all collected from *A. thaliana* in Southwestern Germany (Fig. S1) (*16*). More than 85% of these genomes are classified as the ATUE5 clade of *P. viridiflava* (Fig. 1A), an opportunistic pathogen that colonizes *A. thaliana* throughout Europe and the USA (*17*, *18*). We found viral sequences in 99.3% of the genomes with an average of two viral sequences per genome (Fig. 1B). By using pairwise k-mer distances (*19*) and subsequent k-means clustering, we identified four viral sequence clusters (Fig. 1C).

Phylogenetic associations between *Pseudomonas* clades and viral sequence clusters were immediately apparent (Fig. 1D) (*2*), with viral cluster 2 (VC2) found in all ATUE5 pathogenic strains, but infrequent (29%) (Fig. 1C) in primarily non-pathogenic isolates outside of the ATUE5 clade. Furthermore, the gene composition and synteny of VC2 is conserved across ATUE5 strains but divergent outside ATUE5 (Fig. 2A) suggesting that non-ATUE5 isolates encode a different viral element. The VC2 sequences in ATUE5 consist of 24 genes that are colinearly integrated in all genomes at the same location — between the *trpE* and *trpG* bacterial genes (*20*, *21*). These VC2 sequences are predicted to encode the structural components typical of a prophage, including baseplates, tail spikes, tubes, sheaths, tail fibers, putative assembly chaperones, and transcriptional regulation proteins.

The viral cluster VC2 encodes a structurally functional tailocin

The analysis of the VC2 gene content revealed that this cluster lacks components necessary for phage replication such as capsid, terminase, integrase, and recombinase proteins, suggesting that the VC2 genomic region does not encode a fully functional phage (*22*, *23*). While the structure of VC2 was not consistent with a functional phage, it did resemble previously described bacteriocins that were derived from phages, dubbed tailocins (*24*). Tailocins are repurposed phages that bacteria use to kill co-occurring bacterial strains (*20*) using the infiltration machinery of the phage ancestor.

To test whether the tailocin genomic island produced the predicted tailocin protein complex, we induced tailocin production in a representative strain of ATUE5 (p25.A12) using treatment with mitomycin C (MMC) (*25*), partially purified the lysate, and performed Transmission Electron Microscopy. Transmission electron micrographs (TEM, Fig. 2B) revealed one rod-like tailocin in the MMC-induced lysates of p25.A12 existing in both the extended (sheath uncontracted) and contracted (sheath contracted, and tube ejected) forms (Fig. 2B). The rod-like structure suggests this tailocin is an R-type tailocin (*26–28*). Image comparison of the two tailocin forms indicated that the tailocins differed in average length: the contracted tailocins were on average 130 nm,

while the extended tailocins were on average 144 nm in length. No other phage or bacteriocin-like structures were observed in the TEM micrographs.

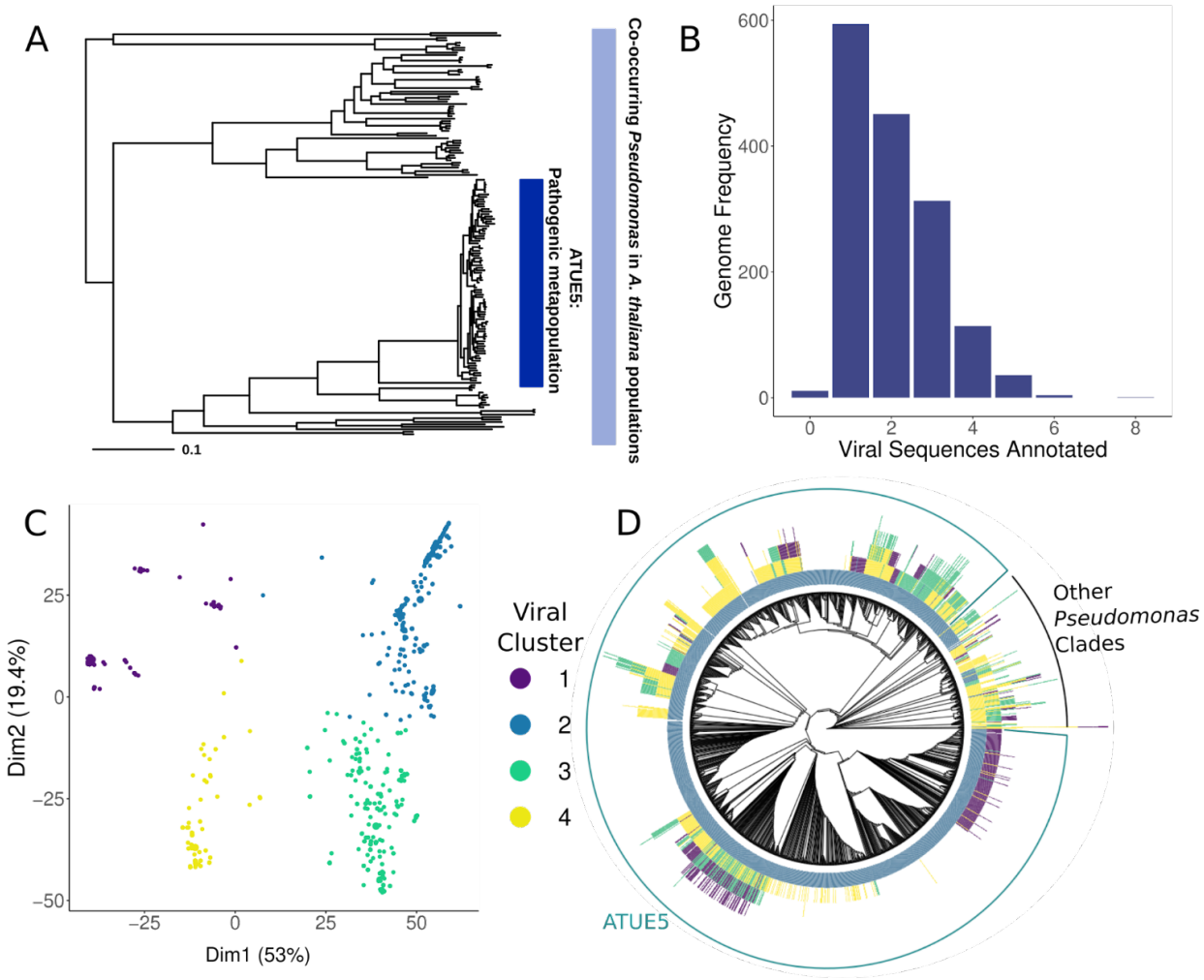

Figure 1. The pathogenic clade of *Pseudomonas viridiflava* ATUE5 harbors a single highly conserved viral cluster (VC2). A. Maximum likelihood concatenated core genome phylogeny of 1,524 pseudomonads that co-occur in the *Arabidopsis thaliana* phyllosphere (light blue bar). The dark blue bar indicates a pathogenic clade of *P. viridiflava* ATUE5 (adapted from Karasov et al. 2018)**. B.** Frequency of viral elements identified in each *Pseudomonas* genome. Each genome harbors between one and eight viral elements. **C.** Principal Component Analysis of viral elements based on pairwise k-mer Mash distances (Ondov et al. 2016) and silhouette analysis reveals four defined clusters. Points are colored as indicated in the inset. **D.** Viral elements of Cluster 2 are present in all genomes of the pathogenic operational taxonomic unit (OTU) ATUE5. The phylogeny shows the relationships among 1,524 *Pseudomonas* strains. Each stacked bar around the phylogeny indicates the presence of a viral element (with a maximum height of eight viral elements). The bars are colored as in C. The outer circle shows strains that belong to ATUE5 (colored in green) and other OTUs (colored in black).

To validate that the observed tailocin in the TEM image corresponded to our predicted tailocin from VC2, we performed untargeted data dependent LC/MS/MS proteomics analysis (*20*, *29*). Proteomics data confirmed the presence of tailocin proteins annotated in VC2 (Fig. 2A and Table

S1). No tailocin proteins were found in the sample that was not induced with mitomycin C (MMC) and no other viral proteins or proteinaceous toxins, such as S-bacteriocins, were detected in the tailocin lysates. Together, these results suggest the highly conserved ATUE5 viral sequence is an active and expressed R-type tailocin.

Pseudomonas genomes encode highly diverse tailocin variants

The phylogenetic distribution of the VC2 cluster (Fig. 1, Fig. 2A) suggested that the VC2 genomic island is well-conserved across pathogenic ATUE5 strains. Across the 1,524 *Pseudomonas* genomes analyzed (Fig. 2C) the full complement of the 27 tailocin genes is found in more than 95% of ATUE5 genomes (Fig. 2C). The other essential bacterial genes found in more than 95% of ATUE5 genomes make up approximately 54% of the average ATUE5 genome, suggesting the tailocin is in the core genome of ATUE5 strains along with essential bacterial genes (*30*). The conservation of the tailocin protein complex across all ATUE5 genomes, its conserved content and its consistent location between *trpE/trpG* suggests a single integration of the VC2 tailocin in the common ancestor of ATUE5 that is likely important for ATUE5 survival.

In addition to the conservation of its gene content and collinearity, we found that the VC2 cluster harbors average levels of intra-lineage nucleotide diversity comparable with bona fide *Pseudomonas* core genes (Fig. 3A, Fig. S2A), indicating that the VC2 cluster has likely been evolving under similar levels of selective constraint. However, through calculating nucleotide diversity for each of the VC2 genes individually, we identified greater levels of nucleotide diversity in two tailocin genes, the hypothetical tail fiber (HTF) and the tail fiber assembly (TFA) genes. Moreover, these two genes were also outliers for positive values of Tajima's D (Fig. 3A, Fig. S2B), a measure of the skewness of the site frequency spectrum, which in this case might indicate gene genealogies with deep divergence times, and the likely existence of distinct haplotypes. It has been previously shown that the HTF and the TFA genes together are key determinants of bacterial host range (*31*), which could explain their high level of diversity and the likely presence of multiple haplotypes.

We formally ascertained 40 and 25 haplotypes for HTF and TFA, respectively. In both cases, the five most common haplotypes accounted for about 80% of the *Pseudomonas* strains. The HTF and TFA haplotypes were nearly completely linked (Fig. S3), thus the presence of a given HTF haplotype is sufficient to accurately predict the corresponding TFA haplotype in a *Pseudomonas* genome. This near perfect linkage of haplotypes is very likely a result of the protein-protein interactions between the two proteins since it has been suggested that the HTF protein (HTFP) directly binds to the targeted bacterial cell, whereas the TFA protein (TFAP) acts as a chaperone facilitating the correct folding of the former (*31*, *32*). These haplotypes were found in all

populations surveyed at all time points. The high positive values of Tajimas's D suggest that the divergent HTF and TFA haplotypes might have been maintained in the ATUE5 clade for long evolutionary times. Our population genomic analyses suggest that while the tailocin is conserved in presence in pathogenic ATUE5 populations, the different ATUE5 strains encode one of a few divergent tailocin variants, with possible differences in binding and killing activity between the variants.

Historical *Pseudomonas* genomes derived from herbarium specimens reveal the long-term maintenance of different tailocin haplotypes

To ascertain the timescales of the emergence of the tailocin and different tailocin variants, we sought to retrieve *Pseudomonas* genomes from historical *A. thaliana* herbarium specimens collected in the last two centuries. Only historical *Pseudomonas* genomes can reveal whether extant populations are the result of population continuity or population replacement. We employed state-of-the-art ancient DNA retrieval and sequencing techniques (*33*) to sequence leaves of 35 *A. thaliana* herbarium specimens using shot-gun genomics (Table S2). We mapped the herbarium-derived reads to the ATUE5 reference genome strain to identify *Pseudomonas* reads and found seven samples with more than 60% of the genome covered by at least one sequencing read (mean genome coverage 0.97-8.7x) (Fig. S4, Table S2). The *Pseudomonas*-derived reads showed patterns of DNA damage and fragmentation typical of ancient DNA (*34*), which authenticated their historical origin (Fig. S5). Our analysis revealed that herbarium specimens preserve the genomes of highly abundant bacterial genera of the *A. thaliana* phyllosphere.

To place the historical *Pseudomonas* genomes in the context of extant genetic diversity, we first used patterns of homozygosity support to identify herbarium specimens most likely to be colonized by a single dominating *Pseudomonas* strain (Fig. S6). This analysis allowed us to select three historical samples (mean genome depth of 8.2 - 9.6x), which we combined with 65 present-day *Pseudomonas* genomes to build a phylogenetic tree using whole-genome Single Nucleotide Polymorphisms (SNPs). The three historical samples were placed within the present-day ATUE5 diversity (Fig. 3D), showing the genetic continuity of this clade for at least the last 177 years. Moreover, the mappings of the historical samples against the ATUE5 reference genome identified the presence of the tailocin in all three historical bacterial genomes (Fig. 3B). In order to ascertain the tailocin variants encoded in the historical genomes and to assess whether the gene colinearity and the insertion location in the bacterial genome are conserved in the historical

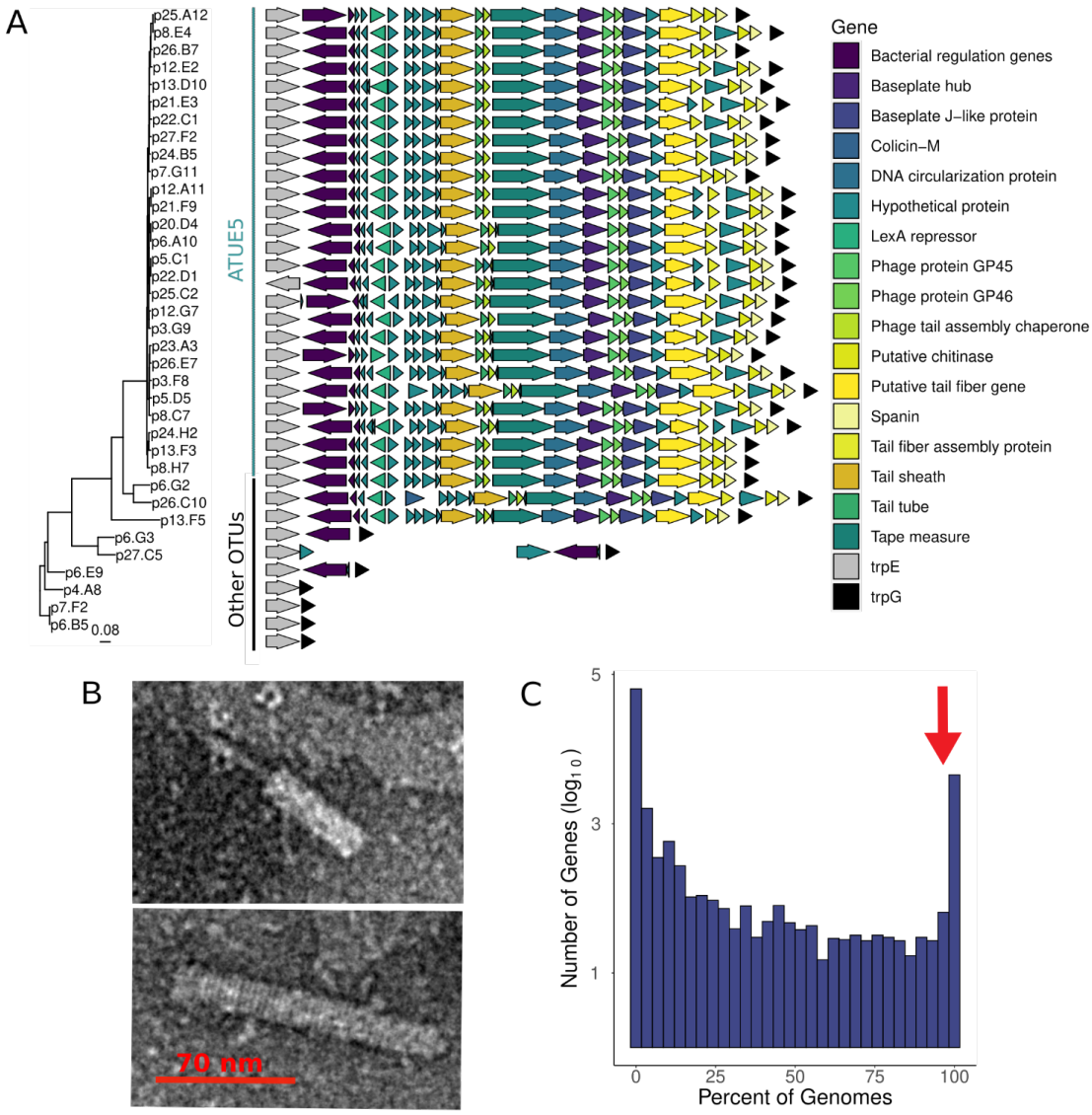

Figure 2. Viral cluster 2 encodes a structurally functional tailocin. A. Genes of Cluster 2 viral elements are syntenic in genomes of ATUE5. The arrows represent tailocin genes and surrounding bacterial genes (gene names are color-coded in the inset). Each row is organized according to its phylogenetic placement. The phylogeny includes 36 *Pseudomonas* representative strains. The vertical lines indicate strains that belong to ATUE5 (colored in green) and other OTUs (colored in black). **B.** Transmission Electron Microscopy (TEM) demonstrates the presence of an assembled tailocin. TEM micrographs show the induced and partially purified tailocin from one representative ATUE5 strain (p25.A12) in both its contracted (top) and (bottom) uncontracted forms. The scale bar shown applies to all micrographs. **C.** The tailocin is part of the ATUE5 core genome. The histogram shows the frequency of each of the pangenome genes within 1,399 ATUE5 genomes. The eleven most conserved tailocins genes are present in more than 90% of the ATUE5 genomes (marked with a red arrow in the histogram).

genomes, we carried out *de novo* assembly of historical *Pseudomonas*-derived reads. We obtained an average of 9 contigs per sample (average sequence length of ~3.4 Kb) that were sufficient to show that both the insertion place of the tailocin and the colinearity of its genes are conserved in the historical genomes (Fig. 3C). Moreover, we ascertained that each historical

genome carries a different tailocin variant and that all historic variants are segregating in present-day populations (Fig. 3B). The polyphyletic distribution of the tail fiber assembly gene haplotypes (Fig. 3D) is likely the result of recombination at this locus.

Finally, we used our phylogenetic reconstruction to determine that the tail fiber assembly gene haplotypes of the historical genomes are closely related to different modern haplotypes (Fig. 3D), suggesting that the same haplotypes have circulated in the *Pseudomonas* metapopulation for the past 177 years.

Different tailocin variants kill closely related pathogenic *Pseudomonas*

We hypothesized that like other tailocins, the tailocin variants encoded by the *Pseudomonas* strains are used for interference competition against closely related competitor strains (*20*, *35*). Our system provides a unique opportunity to characterize killing activity of tailocins produced by co-occurring microbes in the same wild plant populations. The relatedness of these microbes spans the spectrum from closely related (<1% genome wide sequence divergence) to distant commensals (>20% genome wide sequence divergence).

We hypothesized the ATUE5 tailocin is used for interference competition either against (i) other pathogenic pseudomonads that also produce a tailocin, (ii) commensal pseudomonads that mostly do not produce the VC2 tailocin, or (iii) other bacterial strains in the phyllosphere community. To test this, the tailocins from 4 representative TFAP and HTFP haplotypes were induced and partially purified, then concentrated tailocin lysates were applied to 55 ATUE5 strains, 28 other plant-associated *Pseudomonas* strains, and over 50 other phyllosphere bacterial strains isolated from the phyllosphere (taxonomy unknown) in three replicates. We found that, as expected, ATUE5 strains were resistant to their own produced tailocin (*20*). The killing assay also revealed that non-ATUE5 strains were sensitive in only 6% of ATUE-5 derived tailocin treatments, while 40% of ATUE5 treatments with ATUE5-derived tailocins exhibited sensitivity (Fig. 4A). Additionally, none of the other non-*Pseudomonas* bacterial community members were sensitive to the tailocin treatments (data not shown). An agriculturally relevant *Pseudomonas* strain, DC3000, was also tested and found to be sensitive to one of the tested tailocins (hap_p21.F9). In conclusion, the different tailocin variants appear to exhibit partial killing specificity to different subsets of other ATUE5 pathogenic *Pseudomonas*, not to the other *Pseudomonas* commensals, and not to other phyllosphere strains except for the above mentioned DC3000 case.

To verify that the tailocin is responsible for the observed killing activity, we generated a TFAP-deficient mutant. Tailocins collected from the mutant (p25.A12ΔTFAP, (Fig. S7A)) lost killing

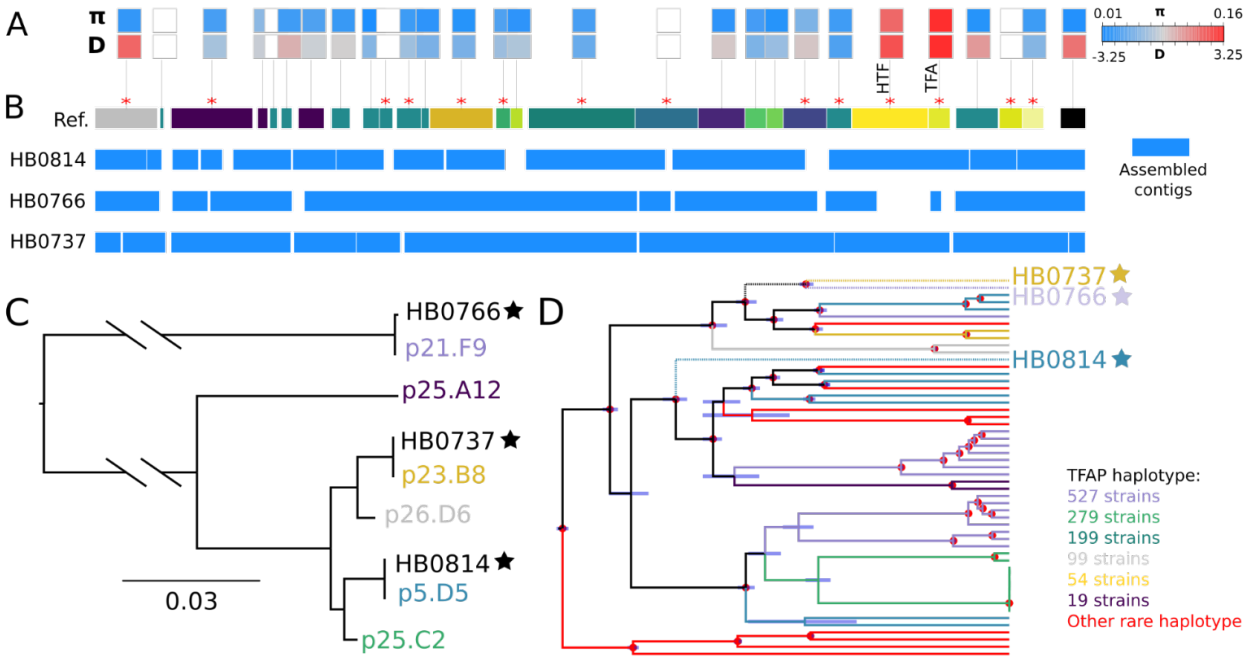

Figure 3. The tail fiber and its assembly protein are hypervariable, and the same alleles are found in historic and modern samples. A. Within-lineage measures of nucleotide diversity (measured as pairwise nucleotide differences (π)) and the site frequency spectrum measured as Tajima's D (D) for each tailocin gene. White boxes indicate genes that are present in less than 5% of strains (statistics are excluded). **B.** *De-novo* assembled contigs spanning the Tailocin region for the historical samples. The reference track on the top represents the genes present on the ATUE5 reference genome. The blue boxes depict the *de-novo* assembled contigs. Red asterisks (*) indicate proteins found in proteomics analysis. **C.** Neighbor-Joining tree of the tail fiber assembly gene translated sequence. Historical samples (HB) are placed in the context of the most common haplotypes. **D.** Bayesian tip-date calibrated phylogeny representing the evolutionary relationship between historical and modern *Pseudomonas* sp. strains. The tips and branches are colored based on the tail fiber assembly gene haplotype. Historical samples are shown with stars. The node bars represent the 95% Highest Posterior Density Intervals of the estimated time and the nodes marked with red dots represent those with posterior probability of 1.

activity against all strains tested (Fig. 4B, Fig. S7B). These results suggest the tailocin is necessary for the observed killing activity, not something else from the tailocin induction lysate.

To determine whether the TFA or HTF haplotypes are associated with killing activity, we analyzed whether the tailocin killing spectrum correlates with a bacterial tester strain's TFA or HTF haplotype (Fig. 4C, Fig. S8). We found an array of different killing spectra. There were TFA/HTF haplotypes with broad killing (hap. p25.A12) and TFA/HTF haplotypes with narrow killing spectra (hap. p23.B8), and a range in between of tailocin haplotypes with divergent, but unique killing activity (Fig. 4C, Fig. S8). Still, we found there were no significant associations between killing spectra and TFA or HTF haplotype. Associating TFA/HTF haplotype with the bacterial tester strain tailocin's TFA/HTF haplotype is not sufficient to explain the killing spectra of the tailocins in our panel. We hypothesize that killing is more complicated than TFA/HTF haplotype alone, and that future work should aim to identify TFA/HTF haplotype associations with bacterial tester

strain tailocin's outer membrane protein composition (lipopolysaccharie and O-antigen, specifically (*20*)).

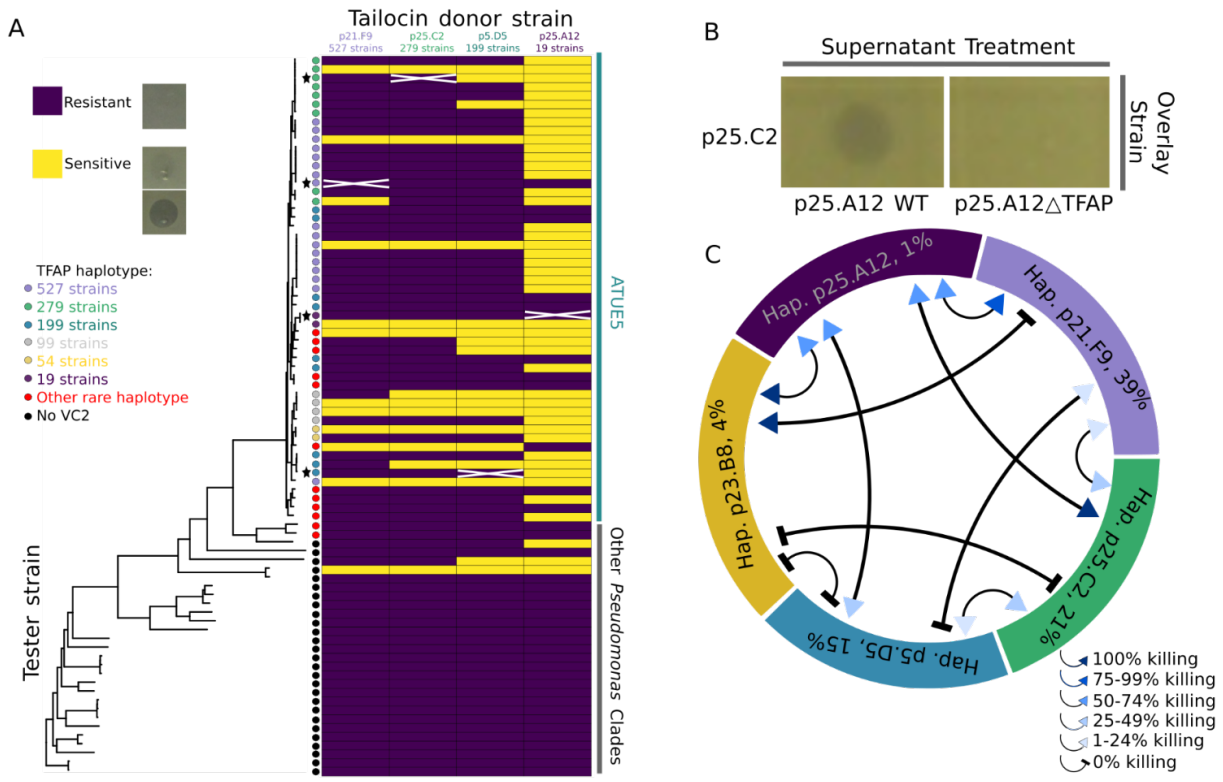

Figure 4. Tailocins target closely-related pathogens. A. Tailocins are preferentially used for intra-lineage killing. Soft agar cultures of the *Pseudomonas* strains (rows) were challenged with viral particles extracted from cultures of one strain from each tail fiber assembly protein haplotype (columns) in 3 biological replicates. The phylogeny includes 83 *Pseudomonas* representative strains and are displayed according to their phylogenetic placement. Vertical lines indicate strains that belong to ATUE5 (colored in green) and other OTUs (colored in black). Interactions with a strain's own tailocin are indicated by white X's. Strains producing clear to opaque lysis zones 2/3 replicates were scored sensitive to the extracted viral particles, while strains producing no lysis zones for 2/3 replicates were scored resistant. Pictures of resistant and sensitive interactions were modified to enhance contrast. **B.** Knockout of the tail fiber assembly gene disrupts tailocin bactericidal activity. Killing activity is indicated by a clearing in the lawn of the overlay strain. **C.** Patterns of killing defined by the tailocin's tail fiber assembly protein haplotype. The colored arrow tips show how often each haplotype kills all other tester strain's haplotypes.

Discussion

Our results have three major implications. First, we discovered that a tailocin, a repurposed phage, is conserved as part of the core genome of metapopulations of the pathogen *P. viridiflava*. Its conservation in presence for at least tens of thousands of years and the sequence conservation of its genes (other than the TFA/HTF) indicate that the tailocin is likely essential (or

nearly so) for survival in *Pseudomonas* metapopulations. This finding is consistent with previous work that suggests that antimicrobial nanomachines such as the Type Six Secretion System are important for survival in the plant environment (*36*). Despite its conservation in presence, multiple allelic variants of the tailocin have been maintained in these populations. We found that the allelic variants target different strains of related pathogenic *Pseudomonas* species/subspecies suggesting that inter-*Pseudomonas*-strain competition is a strong and persistent selective pressure on *Pseudomonas* pathogens.

Second, our results revealed that the same tail fiber/chaperone variants have been maintained in the *Pseudomonas* populations for the past two hundred years. Bacteria are capable of rapid adaptation (*37*). The fact that a defined set of variants is maintained in these populations suggests selective constraints on the evolution of this system, consistent with what has been found in viral evolution (*38*).

Lastly, our results indicate that these pathogen populations are also polymorphic in their susceptibility to tailocins. The mechanism underlying this variability has not yet been identified, but previous work suggests that the mechanisms are likely to involve differences in lipopolysaccharide composition (LPS) (*20*). The presence of a limited set of tail fiber haplotypes in the metapopulation could reflect a limited panel of resistance mechanisms. Similar to the coevolution of toxin-antitoxin systems, the tail fiber and assembly proteins of an ATUE5 strain may be coevolving with the membrane composition (*39*).

Together, these findings provide a roadmap for identifying tail fiber specificities to different strains and the possibility of determining the mechanism of this specificity. Tailocin therapy, like phage therapy, has been proposed as a possible alternative to small molecule antibiotics (*40*–*42*). Proof-of-concept studies demonstrated that tailocins can be used to suppress specific pathogens in animal (*40*) and plant (*41*) models. With any antimicrobial treatment, there is a risk that the target bacterium will evolve resistance. Here we demonstrate that the tailocin genetic diversity of a pathogen metapopulation can be mined to discover "tailocin cocktails" that will target the metapopulation in parallel.

Acknowledgements: We thank Joy Bergelson, Kevin Hockett, Sophien Kamoun, Patricia Lang, Gal Ofir, Dmitri Petrov and Detlef Weigel for useful discussions and comments that improved this manuscript.

Funding: This study was supported by startup funds from the University of Utah (TLK), The Leverhulme Trust (Philip Leverhulme Prize) (HAB), the Royal Society (RSWF\R1\191011) (HAB),

and an equipment grant from the UK Biological Sciences Research Council - BBSRC (BB/R01356X/1) to University College London. Proteomics mass spectrometry analysis was performed at the Mass Spectrometry and Proteomics Core Facility at the University of Utah. Mass spectrometry equipment was obtained through a Shared Instrumentation Grant #S10 OD018210 01A1.

Author contributions: Biological discovery and conceptualization: TB, TLK, HAB. Methodology: TB, TLK, HAB, SL. Experiments: TB, DB, AMM, SS, LE. Computation: TB, SL. Writing: TB, TLK, HAB, SL.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data and code are available in the supplementary materials or at the git pages:
https://github.com/talia-backman/Ps1524_tailocin
https://github.com/smlatorreo/Pseudomonas_tailocin.

Materials and Methods

Bacterial strains, plasmids and growth conditions.

The strains used in this study are detailed in Supplementary Table 2. Bacteria were grown on nutrient agar (LB agar) and in nutrient lysogeny broth (LB) at 30°C for *Pseudomonas* strains and 37°C for *Escherichia coli* strains. All liquid cultures were incubated with shaking at 180 rpm. Growth media were supplemented with kanamycin (50 mg mL−1, Km50), nitrofurantoin (100 mg mL -1, NFN100), or tetracycline (20 mg mL -1, TET20). All strains were stored at −80°C in 20% glycerol [v/v]. For plasmid preparation, *E. coli* transformants were cultured in LB. TET20 was used to select *E. coli* transformants. LB supplemented with sucrose, and nitrofurantoin (10 g/L tryptone, 5 g/L yeast extract, 10% (W/V) filtered sucrose, 15 g/L agar, NFN100) plates were used in the selection of the second *P. syringae* homologous recombination event.

Viral Identification and Clustering.

VIBRANT v1.2.1 (*16*) was run on the 1,524 assembled *Pseudomonas* genomes to identify viral sequences. To group the 3,104 annotated viral sequences from *VIBRANT* into genetically similar clusters the pairwise mutation distance was calculated between all 3,104 viral sequences with *Mash* v2.2.2 (*19*), a MiniHash-based technique that reduces sequences to compressed k-mer representations (sketches). *Mash* was run ignoring single copy k-mers that are more likely the result of sequencing errors, and a sketch size of 10,000 which corresponds to the number of non-redundant min-hashes that are kept for the pairwise distance calculation. This resulted in a 3,104 x 3,104 pairwise mutation distance matrix. Next, the viral sequences were clustered from the pairwise mutation distance matrix with principal component analysis and K-means clustering (*43*, *44*). Silhouette analysis (*45*) was performed to determine the optimal number of clusters for the data.

Annotation of the viral cluster 2 (VC2) gene cluster.

A highly conserved viral sequence, the VC2 gene cluster, was identified in assembled genomes using VIBRANT v1.2.1 and groups of orthologous genes (orthogroups) were identified using *panX (46)*. Insertion sites for VC2 gene clusters were identified by searching for orthologs of known viral sequence flanking genes: *mutS*, *cinA*, *trpE*, and *trpG* genes. VC2 was always found between *trpE* and *trpG*. Functional gene annotations were determined by orthologous viral genes previously characterized in *VIBRANT* and *BLAST*. Tailocin gene clusters were visualized with the R package "gggenes".

Viral particle extraction and partial purification.

Overnight cultures were back-diluted into fresh LB to extract and isolate tailocins from the various *Pseudomonas* strains. When cultures reached an exponential growth phase (an OD 600nm of 0.4-0.6) 5 µg mL−1 of mitomycin C (MMC, Selleck Catalog No. S8146) was added to induce the bacterial SOS response and tailocin induction. The cultures were then incubated at 28°C for a minimum of 18 hours. Cultures were then centrifuged at 4°C for 1 hour at 1,400g to pellet cell debris. The supernatants were sterilized by filtration with 0.2 µm cellulose acetate filters. To precipitate viral particles, 40% [w/v] of ammonium sulfate was slowly added to the filter-sterilized tailocin lysate while stirring on ice. The lysate was left stirring on ice for at least 18 hrs. After 18 hrs at 4°C, the ammonium sulfate was pelleted by centrifugation at 2,090g for 2 hrs at 4°C. The supernatant was discarded, and the pellet was resuspended in 500 µL cold P-buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris-HCl, pH 7.5) and stored at 4°C. Tailocin lysates were used fresh or for up to two weeks.

Visualization of viral particles using transmission electron microscopy.

Two negatively stained specimens (from samples p25.A12 and p25.C2) were prepared and imaged in the following manner: 3.5 µl of sample was placed on a carbon-coated, transmission EM grid that had been glow-discharged to make it more hydrophilic. After approximately one minute, filter paper was used to remove excess solution from the grid by touching the side of the grid. This was followed by two steps of brief (approx. 1-2 seconds) wash in P-buffer and blotting with filter paper in the same manner. Next, this process was repeated twice but with the negative stain solution (1% ammonium molybdate). Finally, the grid was incubated for 15-20 seconds in a droplet of 1% ammonium molybdate, blotted again, and allowed to air dry. Negatively stained specimens were viewed on a ThermoFisher Tecnai 12 transmission electron microscope operated at 120 kV. Images were recorded on a Gatan UltraScan camera.

Protein mass spectrometry.

Tailocins partially purified as described above were treated with or without MMC and resuspended in um bicarbonate(1x S-trap buffer and frozen at -80°C until digestion. Total protein content was quantified using the Pierce™ BCA Protein Assay Kit. rotein ™ was diluted into 25 µL of 1x S-TRAP buffer and the protein was reduced with 20 mM at 37°C temperature for 30 minutes. Samples were alkylated with 40 mM at room temperature for 45 minutes in the dark and acidified phosphoric acid. (10% 100 mM TEABpH 7.5 90% methanol)four 150 µL of binding buffer. 10 µg of protein sample was digested with 0.5 µg of trypsin for 2.5 hours at 47°C. The peptides were eluted from the column with 40 µL of 50 mM TEAB pH 8.5 followed by 40 µL of 0.2% formic acid and 40 µL of elution buffer (50 mM TEAB pH 8.5 + 0.2% formic acid + 50 % acetonitrile). The peptides were dried to completion, resuspended in 300 µL of 0.1% TFA and

desalted using Pierce™ Peptide Desalting Spin Columns according to the manufacturer's instructions. The peptides were resuspended in 40 µL of 0.1% formic acid for LC-MS/MS analysis.

Mass spectrometry analysis of samples.

Reversed-phase nano-LC-MS/MS was performed on an UltiMate 3000 RSLCnano system (Dionex) coupled to a Thermo Scientific Q Exactive-HF mass spectrometer equipped with a nanoelectrospray source. 2 µL of each sample were first trapped on a 2 cm Acclaim PepMap-100 column (Thermo Scientific) with 5% acetonitrile at 5 µl/min and at 5 minutes the sample was injected onto the liquid chromatograph reverse-phase Acclaim™ PepMap™ 100 C18 2.0 µm nanocolumn (Thermo Scientific). A 500 mm long/ 0.075 mm inner diameter nanocolumn heated to 35°C was employed for chromatographic separation. The peptides were eluted with a gradient of reversed-phase buffers (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in 100% acetonitrile) at a flow rate of 0.2 µL/min. The LC run lasted for 85 minutes with a starting concentration of 5% buffer B increasing to 28% buffer B over 75 minutes, up to 40% buffer B over 10 minutes and held at 90% B for 10 minutes. The column is allowed to equilibrate at 5% buffer B for 20 minutes before starting the next data acquisition. The Orbitrap was operated in data-dependent acquisition (DDA) MS/MS analysis mode selecting the top 20 most abundant precursor ions between 375-1650 m/z at 60,000 resolution for fragmentation at 15,000 resolution.

Mass spectrometry data analysis.

The raw data was analyzed using Proteome Discover 3.0 software with the SEQUEST algorithm against the uniprot_ref_pseudomonas_viridiflava database (1-18-2023 version with 4,389 proteins) or p25.A12 database. An allowance was made for 2 missed cleavages following trypsin/Lys C digestion. No fixed modifications were considered. The variable modifications of methionine oxidation and cysteine carbamidomethylation (IAA) were considered with a mass tolerance of 15 ppm for precursor ions and a mass tolerance of 0.02 Da for fragment ions. The results were filtered with a false discovery rate of 0.01 (Percolator node). A minimum of 1 unique peptide should be reported for all proteins identified.

Allelic diversity and nucleotide sequence data analysis.

Orthogroups that were previously identified using *panX* were included in the analysis. The nucleotide and peptide sequences were aligned with *Clustal Omega* (*47*), and then a codon-based nucleic acid alignment was generated with *pal2nal* (*48*). This is critical to ensure the nucleic acid alignment is aligned codon-by-codon, to know whether substitutions result in a synonymous or nonsynonymous amino acid change. The output aligned file was saved in fasta format and used as an input file for population summary statistic calculations in R using the package

"PopGenome" v2.7.5 (*49*). The level of sequence diversity among ortho-groups was compared using summary statistics 'π' (average pairwise difference or average number of nucleotide diversity per site) (*50*), '**θ**' (population mutation parameter or number of segregating sites) (*51*), whereas Tajima's D (*52*) was used as a proxy for the site frequency spectrum (SFS) and thus to determine the demographic and selective forces shaping the SFS of tailocin genes.

Ascertainment of tailocin haplotypes.

To determine how many haplotypes of tail fiber assembly and hypothetical tail fiber genes there are in the pathogenic strains in our study, nucleotide and peptide alignments from *Clustal Omega* (*47*) were analyzed using the R package *pegas (53)*. We used the haplotype function to determine how many unique haplotypes are in the ATUE5 strains. The most common haplotypes were used for downstream analyses.

Mapping of reads from historical samples to ATUE5.

A set of 35 herbarium *Arabidopsis thaliana* samples collected between 1817 - 1957 from Southern Germany (Supp. Table 2) (*54*) were screened for *Pseudomonas* sp. presence. Adapter sequences from the raw reads were trimmed and merged using *Adapterremoval* V.2 (*55*) and subsequently mapped to the *Pseudomonas* p25.A12 assembly using *bwa aln* v.0.7.17 (*56*) with seed deactivation to allow the alignment of of substitutions present at the termini of the reads, *e.g.* those arising from DNA deamination in historic material (*33*). A total of 8 samples for which the *Pseudomonas* reference genome was covered in at least 60% were kept (Supp Table 2). To authenticate the historic origin of the bacterial reads, the proportion of C-to-T (or G-to-A) substitutions as well as the fragment size distribution of the reads were computed using mapDamage v.2.2.1 (*57*) (Supp Fig S5).

Ascertainment of tailocin variants in historical genomes.

In order to identify reads aligning to the whole spectrum of tailocin genetic diversity, different haplotypes of the highly polymorphic tail fiber assembly and hypothetical tail fiber genes (Fig 3A) were added to the p25.A12 reference genome, before the historical samples were mapped as described above. Reads covering the Tailocin region surrounded by the flanking genes *trpE* and *trpG* as well as all reads covering any haplotype of the mentioned genes were subset from the raw reads and used as input for *de-novo* assembly using *SPAdes* v.3.13.0 (*58*). Due to their low coverage, five out of the eight historical samples did not produce any assembly output. In contrast, samples HB0737, HB0766 and HB0814 yielded 10, 9 and 14 assembled contigs (Supp Table 4). To find collinearity between the *de-novo* assemblies and the reference Tailocin region, we used *minimap* v.2.1 (*59*). Aminoacid translated sequences of the historical isolate-specific haplotypes for the Tail Fibre assembly protein and Putative Tail Fibre protein were merged

together with the haplotypes segregating in the observed diversity. Finally, we used *Clustal Omega* v.1.2.4 (*60*) to generate a multiple alignment and classify the haplotypes in the historical samples.

Whole-genome phylogenetic reconstruction combined present-day and historical genomes.

A set consisting of 50 ATUE5 modern and 3 historic *Pseudomonas* strains were used to create a phylogenetic reconstruction. We used the mapped reads to perform individual *de-novo* variant callings using *bcftools mpileup* and *bcftools call* v1.11 (Danecek et al. 2021) removing those reads with mapping quality values lower than 30. We then merged all individual calls using *bcftools merge* v.1.11 (Danecek et al. 2021) and removed those samples with less than 60% of the called variants. We finally kept only biallelic SNPs and sites with a maximum missing information of 10% of the remaining samples. The number of retrieved variants were 176,528 SNPs segregating among 53 strains (50 modern and 3 historic).

Using the genomic SNPs we sought to reconstruct the phylogenetic relations among the 53 strains. We used *IQ-TREE* V2.2.0.3 (Minh et al. 2020) in combination with *ModelFinder* (Kalyaanamoorthy et al. 2017), and ultrafast bootstrap *UFBoot* (Hoang et al. 2018) to estimate a Maximum-Likelihood phylogenetic tree. Variant sites that are likely to arise due to putative recombination were identified using *ClonalFrameML* v.1.12 (Didelot and Wilson 2015) and the output was used to mask putative recombinant positions in the original genomic SNP array. We finally used *BEAST2* (Drummond and Bouckaert 2015) to jointly estimate the phylogenetic relations among the strains as well as the relative time of emergence for all the tree nodes. For this purpose we used the collection dates (Supp Table 2) and a previously estimated evolutionary rate (McCann et al. 2017) as priors. In order to minimize parameter estimation we chose HYK as the evolutionary model and to avoid demographic history assumptions we used a Coalescent Extended Bayesian Skyline approach (Drummond et al. 2005).

Testing bacterial sensitivity to tailocins with spot test phenotypic assays.

To test the sensitivity of the different bacterial strains to the tailocins, soft agar assays were performed using an adaptation of a protocol from Vacheron et al. 2021 (*35*). 750 µL overnight cultures of each strain was mixed with 25 mL of LB soft agar (0.8%). The mixture was poured into a square plate and left to harden. Then, aliquots of 3-5 µL of concentrated viral particles suspension were applied to the agar, along with a control of P-buffer alone and non-induced cultures in serial dilutions. The plates were incubated overnight at 30°C. Bacterial sensitivity or resistance to the viral particles was assessed after 24-48 hours. When testing serially diluted tailocins, no plaques were observed, suggesting the killing agent is nonreplicative and not a phage. When testing the uninduced tialocin control samples (not induced with mitomycin C), if

lethality was observed, the interaction was deemed to be inconclusive. Spot tests were performed in biological triplicates.

Testing bacterial sensitivity to tailocins in culture.

To test the sensitivity of a known sensitive strain (p25.C2) to the mutant tailocin from 25.A12ΔTFAP, p25.C2 was grown overnight at 30°C with shaking at 180 r.p.m. then diluted by a factor of 1:10 in fresh LB medium. Once the culture reached growth phase, cultures were supplemented with 5% p25.A12 tailocin, 5% p25.A12ΔTFAP tailocin, or a buffer control. OD measurements at a wavelength of 600 nm were measured in a 96-well plate using a microplate reader (TECAN Spark).

Construction of the p25.A12 tail fiber assembly gene deletion.

The p25.A12ΔTFAP mutant strain was created using the Gateway Cloning system with the donor vector pDONR1K18ms (Addgene plasmid #72644) and the destination vector pDEST2T18ms (Addgene plasmid #72647). Plasmids and primers are listed in Supplementary Table 3.

Construction of the *Pseudomonas* tail fiber assembly gene deletion.

A. **Amplification of the upstream and downstream sequences of the locus of interest.**
   Partial overlapping flanking gene primers were designed to amplify the upstream and the downstream 200 bp nucleotide sequences flanking the tail fiber assembly protein (TFAP) with attB site and kanamycin resistance cassette overlap. The flanking primers for the TFAP deletion construct are as follows: upstream forward, 5′ ggggacaagtttgtacaaaaaagcaggctCGCGCGCGTCATCAATTCTA 3′ and upstream reverse 5′ TTTCAAAGAACCAGTTAAGGGTATTCATgatcctctag 3′; downstream forward, 5′ tgcaagcttgCAGGAGTTCACATGCCTATCG 3′ and downstream reverse 5′ ggggaccactttgtacaagaaagctgggtGTGCGTGAGCTGGGCAAC 3′, with overlapping bases shown in lower case, which facilitate fusion in the overlap PCR of the deletion construct. The colony PCR reactions contained 10 μM dNTPs, 1.0 μM forward and reverse primers, DNA template picked from colonies, 0.02 U/μl 5X Q5 high fidelity DNA polymerase (New England Biolabs) in 1x reaction buffer containing 2 mM MgCl2 supplied by New England Biolabs. The thermo cycles were programmed as followed: initial colony PCR denaturation at 98°C for 5 minutes, followed by 42 cycles of 98°C for 10 seconds, annealing temperature (58°C) for 20 seconds, and 72°C for 30 seconds, and a final extension at 72°C for 5 minutes. The PCR products were analyzed by DNA gel electrophoresis and purified by gel extraction (GeneJET Gel Extraciton Kit) before use.
B. **Overlap PCR of the deletion construct.**

The upstream and downstream sequences of the TFAP flanked by attB sites and with a kanamycin resistance cassette in place of the TFAP was constructed by overlap PCR. Briefly, the reaction mix contained 10 nM of each amplified upstream, downstream, and antibiotic resistance cassette DNA fragments, 10 μM forward and reverse oligos with the desired overlap fragments, 0.0125 U/μl of Q5 DNA polymerase, and 10 μM dNTPs in 1x Q5 DNA polymerase reaction buffer containing 2 mM MgCl2. The thermo cycles were programmed as followed: initial colony PCR denaturation at 98°C for 1 minute, followed by 30 cycles of 98°C for 10 seconds, annealing temperature of the overlap regions (52°C) for 1 minute, and 72°C for 45 seconds. The PCR products were analyzed by DNA gel electrophoresis and overlap products were purified by gel extraction (GeneJET Gel Extraciton Kit) before use in Gateway Cloning.

C. **Construction of the deletion suicide vector plasmid.**

The attB-flanked overlap PCR product was directly transferred into an expression clone with a Gateway Cloning one tube reaction (*61*). Briefly, a reaction mixture of 50-100ng attB DNA, 195 ng attP DNA (Invitrogen pDONR vector), 3.0 μl BP Clonase II enzyme mix, and TE Buffer, pH 8.0 added to a final volume of 15 μl were incubated at 25°C for at least 4 hours. 5 μl of the reaction was removed to a separate tube and transformed into competent *E. coli* DH5**α** cells to assess the efficiency of the BP reaction. Then, to the remaining 10 μl reaction, 300ng of the Destination vector and 3.0 μl of the LR Clonase II enzyme mix were mixed and incubated at 25°C for at least 2 hours. 2 μl of proteinase K solution was added and the reaction mixture was incubated at 37°C for 10 minutes. The resulting expression clone plasmid was transformed into high efficiency chemically competent *E. coli* DH5**α**. Transformants were selected for with TET20 and screened by colony PCR using primers which include the flanking arms cloning site: forward, 5′ CGCGCGCGTCATCAATTCTA 3′ and reverse, 5′ GTGCGTGAGCTGGGCAAC 3′ (annealing temperature: 58 °C). Positive clones were further verified by plasmid DNA extraction (GeneJET Plasmid Miniprep Kit) followed by DNA sequencing of the plasmid. A colony containing the deletion suicide vector was inoculated in LB broth and used for mating with *Pseudomonas* strain p25.A12.

D. **Mating competent *E. coli* DH5α with the constructed suicide vector with *Pseudomonas* strain p25.A12 and selection of the first homologous recombination.**

One colony of *Pseudomonas* strain p25.A12 and *E. coli* HB101 (ATCC 33694) with the helper plasmid pRK2013 (*62*) were grown overnight in 5mL LB broth. The next morning, the overnight cultures were back-diluted 1:10 in LB and grown to exponential growth. Then 4mL of each culture (the DH5**α** suicide vector, the *Pseudomonas* strain p25.A12, and the *E. coli* strain with the helper plasmid) were mixed in a 15 mL centrifuge tube and cells were immediately harvested at 3,000 rpm for 7 minutes at room temperature.

Supernatant was removed and washed in 1 mL LB three times. The mixture was resuspended in 100 uL LB and cells were pipetted in 5ul dots onto LB plates with no antibiotic to allow for conjugation of the suicide vector. Mating was allowed to occur overnight at 30°C and for at least 24 hours. The cells were scraped into 500uL 10mM sterile MgSO4 and the cells were harvested at 8,000 rpm for 2 minutes. The supernatant was discarded and the cells were resuspended in 1mL 10 mM MgSO4. The first homologous recombination was selected for by plating 100 ul of the resuspended cells on selective plates with NFN100 and TET20 and grown for 1-3 days at 30°C and until there are colonies, but before a lawn grows. After NFN100 and TET20 selection, colonies were screened by colony PCR using the same primers including the flanking arms cloning site.

E. SacB counter-selection of the second homologous recombination.

Positive clones were purified by streaking on new NFN100 and TET20 plates, then inoculated in LB broth with no antibiotic to allow for recombination events. Colonies were screened by colony PCR using a set of primers flanking the genomic sequence of the deleted locus of the TFAP (forward, 5′ CGCGCGCGTCATCAATTCTA 3′, and reverse, 5′ GTGCGTGAGCTGGGCAAC 3′). Colonies that resolved to the wild-type sequence were 946bp while the deletion with the kanamycin resistance cassette were 1816bp. Colonies with the confirmed deletion were purified by streaking on NFN100 plates and further verified by sanger sequencing and induction of the knockout strain's tailocin with complementary killing assays on a known sensitive strain.

F. Genotypic verification of ΔTFAP.

The flanking arms to the TFAP were amplified in the mutant strain p25.A12ΔTFAP and sent for Sanger Sequencing.

G. Phenotypic verification of ΔTFAP.

The tailocin from the mutant strain p25.A12ΔTFAP was induced with MMC and partially purified. Bacterial sensitivity to the mutant tailocins were tested with the spot test.

Tailocin sensitivity phenotypes were validated through marked deletions of the tail fiber assembly gene. Disruption of tailocin production was verified by spot test phenotypic assays.

References

1. E. M. Goss, S. Timilsina, A bacterial epidemic in wild plants. *Nat Ecol Evol.* **2** (2018), pp. 1529–1530.

2. T. L. Karasov, J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, D. S. Lundberg, M. Neumann, J. Regalado, R. A. Neher, E. Kemen, D. Weigel, Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary

Timescales. *Cell Host Microbe*. **24**, 168–179.e4 (2018).

3. E. A. Van der Biezen, J. D. Jones, Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* **23**, 454–456 (1998).

4. T. Ruiz-Bedoya, P. W. Wang, D. Desveaux, D. S. Guttman, Cooperative virulence via the collective action of secreted pathogen effectors. *Nature Microbiology*, 1–11 (2023).

5. M. R. Clokie, A. D. Millard, A. V. Letarov, S. Heaphy, Phages in nature. *Bacteriophage*. **1**, 31–45 (2011).

6. S. L. Díaz-Muñoz, B. Koskella, Bacteria-phage interactions in natural environments. *Adv. Appl. Microbiol.* **89**, 135–183 (2014).

7. B. Koskella, M. A. Brockhurst, Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* (2014) (available at https://academic.oup.com/femsre/article-abstract/38/5/916/497097).

8. M. A. Brockhurst, A. Buckling, P. B. Rainey, The effect of a bacteriophage on diversification of the opportunistic bacterial pathogen, Pseudomonas aeruginosa. *Proc. Biol. Sci.* **272**, 1385–1391 (2005).

9. B. Koskella, Phage-mediated selection on microbiota of a long-lived host. *Curr. Biol.* **23**, 1256–1260 (2013).

10. E. A. Dewald-Wang, N. Parr, K. Tiley, A. Lee, B. Koskella, Multiyear time-shift study of bacteria and phage dynamics in the phyllosphere. *Am. Nat.* **199**, 126–140 (2022).

11. T. L. Karasov, J. M. Kniskern, L. Gao, B. J. DeYoung, J. Ding, U. Dubiella, R. O. Lastra, S. Nallu, F. Roux, R. W. Innes, L. G. Barrett, R. R. Hudson, J. Bergelson, The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature*. **512**, 436–440 (2014).

12. O. Shalev, T. L. Karasov, D. S. Lundberg, H. Ashkenazy, P. Pramoj Na Ayutthaya, D. Weigel, Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant. *Nature Ecology & Evolution*, 1–14 (2022).

13. D. S. Lundberg, R. de Pedro Jové, P. Pramoj Na Ayutthaya, T. L. Karasov, O. Shalev, K. Poersch, W. Ding, A. Bollmann-Giolai, I. Bezrukov, D. Weigel, Contrasting patterns of microbial dominance in the Arabidopsis thaliana phyllosphere. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2211881119 (2022).

14. J. De Smet, H. Hendrix, B. G. Blasdel, K. Danis-Wlodarczyk, R. Lavigne, Pseudomonas predators: understanding and exploiting phage--host interactions. *Nat. Rev. Microbiol.* **15**, 517 (2017).

15. C. B. Silveira, A. Luque, F. Rohwer, The landscape of lysogeny across microbial community density, diversity and energetics. *Environ. Microbiol.* **23**, 4098–4111 (2021).

16. K. Kieft, Z. Zhou, K. Anantharaman, VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome.* **8**, 90 (2020).

17. K. Jakob, E. M. Goss, H. Araki, T. Van, M. Kreitman, J. Bergelson, Pseudomonas viridiflava and P. syringae—natural pathogens of Arabidopsis thaliana. *Mol. Plant. Microbe. Interact.* **15**, 1195–1203 (2002).

18. T. L. Karasov, M. Neumann, G. Shirsekar, G. Monroe, D. Weigel, R. Schwab, Drought selection on Arabidopsis populations and their microbiomes. *bioRxiv* (2022), p. 2022.04.08.487684.

19. B. D. Ondov, T. J. Treangen, P. Melsted, A. B. Mallonee, N. H. Bergman, S. Koren, A. M. Phillippy, Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).

20. S. Carim, A. L. Azadeh, A. E. Kazakov, M. N. Price, P. J. Walian, L. M. Lui, T. N. Nielsen, R. Chakraborty, A. M. Deutschbauer, V. K. Mutalik, A. P. Arkin, Systematic discovery of pseudomonad genetic factors involved in sensitivity to tailocins. *ISME J.* **15**, 2289–2305 (2021).

21. K. Nakayama, K. Takashima, H. Ishihara, T. Shinomiya, M. Kageyama, S. Kanaya, M. Ohnishi, T. Murata, H. Mori, T. Hayashi, The R-type pyocin of Pseudomonas aeruginosa is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* **38**, 213–231 (2000).

22. W. H. Roos, I. L. Ivanovska, A. Evilevitch, G. J. L. Wuite, Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. *Cell. Mol. Life Sci.* **64**, 1484–1497 (2007).

23. X. Shen, M. Li, Y. Zeng, X. Hu, Y. Tan, X. Rao, X. Jin, S. Li, J. Zhu, K. Zhang, F. Hu, Functional identification of the DNA packaging terminase from Pseudomonas aeruginosa phage PaP3. *Arch. Virol.* **157**, 2133–2141 (2012).

24. K. L. Hockett, T. Renner, D. A. Baltrus, "Independent Co-option of a Tailed Bacteriophage into a Killing Complex in Pseudomonas" (2014), , doi:10.1101/011486.

25. K. L. Hockett, D. A. Baltrus, Use of the Soft-agar Overlay Technique to Screen for Bacterially Produced Inhibitory Compounds. *J. Vis. Exp.* (2017), doi:10.3791/55064.

26. S. I. Ishii, Y. Nishi, F. Egami, The fine structure of a pyocin. *J. Mol. Biol.* **13**, 428–431 (1965).

27. J. R. Govan, Studies on the pyocins of Pseudomonas aeruginosa: morphology and mode of action of contractile pyocins. *J. Gen. Microbiol.* **80**, 1–15 (1974).

28. T. B. Higerd, C. A. Baechler, R. S. Berk, Morphological studies on relaxed and contracted forms of purified pyocin particles. *J. Bacteriol.* **98**, 1378–1389 (1969).

29. E. C. Schirmer, J. R. Yates 3rd, L. Gerace, MudPIT: A powerful proteomics tool for discovery. *Discov. Med.* **3**, 38–39 (2003).

30. D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).

31. D. A. Baltrus, M. Clark, C. Smith, K. L. Hockett, Localized recombination drives diversification of killing spectra for phage-derived syringacins. *ISME J.* **13**, 237–249 (2019).

32. A. J. Salazar, M. Sherekar, J. Tsai, J. C. Sacchettini, R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold. *PLoS One.* **14**, e0211432 (2019).

33. S. M. Latorre, P. L. M. Lang, H. A. Burbano, R. M. Gutaker, Isolation, library preparation, and bioinformatic analysis of historical and ancient plant DNA. *Curr. Protoc. Plant Biol.* **5**, e20121 (2020).

34. C. L. Weiß, V. J. Schuenemann, J. Devos, G. Shirsekar, E. Reiter, B. A. Gould, J. R. Stinchcombe, J. Krause, H. A. Burbano, Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R Soc Open Sci.* **3**, 160239 (2016).

35. J. Vacheron, C. M. Heiman, C. Keel, Live cell dynamics of production, explosive release and killing activity of phage tail-like weapons for Pseudomonas kin exclusion. *Commun Biol.* **4**, 87 (2021).

36. P. Bernal, L. P. Allsopp, A. Filloux, M. A. Llamas, The Pseudomonas putida T6SS is a plant warden against phytopathogens. *ISME J.* **11**, 972–987 (2017).

37. E. R. Moxon, P. B. Rainey, M. A. Nowak, R. E. Lenski, Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).

38. V. Druelle, R. A. Neher, Reversions to consensus are positively selected in HIV-1 and bias substitution rate estimates. *bioRxiv* (2022), p. 2022.02.13.480259.

39. D. J. Rankin, L. A. Turner, J. A. Heinemann, S. P. Brown, The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. *Proceedings of the Royal Society B: Biological Sciences.* **279**, 3706–3715 (2012).

40. D. Gebhart, S. Lok, S. Clare, M. Tomas, M. Stares, D. Scholl, C. J. Donskey, T. D. Lawley, G. R. Govoni, A modified R-type bacteriocin specifically targeting Clostridium difficile prevents colonization of mice without affecting gut microbiota diversity. *MBio.* **6** (2015),

doi:10.1128/mBio.02368-14.

41. A. Príncipe, M. Fernandez, M. Torasso, A. Godino, S. Fischer, Effectiveness of tailocins produced by Pseudomonas fluorescens SF4c in controlling the bacterial-spot disease in tomatoes caused by Xanthomonas vesicatoria. *Microbiol. Res.* **212-213**, 94–102 (2018).

42. D. A. Baltrus, M. Clark, K. L. Hockett, M. Mollico, C. Smith, S. Weaver, Prophylactic Application of Tailocins Prevents Infection by Pseudomonas syringae. *Phytopathology*. **112**, 561–566 (2022).

43. K. Pearson, LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. **2**, 559–572 (1901).

44. S. Lloyd, Least squares quantization in PCM. *IEEE Trans. Inf. Theory*. **28**, 129–137 (1982).

45. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).

46. W. Ding, F. Baumdicker, R. A. Neher, panX: pan-genome analysis and exploration. *Nucleic Acids Res.* **46**, e5 (2018).

47. F. Sievers, D. G. Higgins, Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).

48. M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12 (2006).

49. B. Pfeifer, U. Wittelsbürger, S. E. Ramos-Onsins, M. J. Lercher, PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).

50. M. Nei, W. H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5269–5273 (1979).

51. Y. X. Fu, Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**, 172–197 (1995).

52. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. **123**, 585–595 (1989).

53. E. Paradis, pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. **26**, 419–420 (2010).

54. S. M. Latorre, P. L. M. Lang, H. A. Burbano, *Historical Arabidopsis thaliana genomes from Germany* (2022; https://zenodo.org/record/7156189).

55. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes.* **9**, 88 (2016).

56. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754–1760 (2009).

57. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* **29**, 1682–1684 (2013).

58. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinformatics.* **70**, e102 (2020).

59. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**, 3094–3100 (2018).

60. F. Sievers, D. G. Higgins, The Clustal Omega Multiple Alignment Package. *Methods Mol. Biol.* **2231**, 3–16 (2021).

61. F. Katzen, Gateway(®) recombinational cloning: a biological operating system. *Expert Opin. Drug Discov.* **2**, 571–589 (2007).

62. D. H. Figurski, D. R. Helinski, Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1648–1652 (1979).