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

Talia Backman[1], Sergio M. Latorre[2], Efthymia Symeonidi[1], Artur Muszyński[3], Ella Bleak[1], Lauren Eads[1], Paulina I. Martinez-Koury[1], Sarita Som[1], Aubrey Hawks[1], Andrew D. Gloss[4], David M. Belnap[1,5], Allison M. Manuel[6], Adam M. Deutschbauer[7], Joy Bergelson[4], Parastoo Azadi[3], 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]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA.

[4]Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA

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

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

[7]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 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

ES 0000-0002-4412-4596, u6030864@umail.utah.edu

AM 0000-0003-3497-9977, muszynski@ccrc.uga.edu

EB 0009-0009-3437-9163, u1346590@utah.edu

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

PM paulina.martinez@utah.edu

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

AH 0000-0001-6061-2584, aubrey.hawks@utah.edu

ADG 0000-0003-3649-1836, ag8612@nyu.edu

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

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

AD 0000-0003-2728-7622, amdeutschbauer@lbl.gov

JB 0000-0001-7893-7387, jb7684@nyu.edu

PA 0000-0002-6166-9432 azadi@ccrc.uga.edu

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

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

Abstract

Bacteriophages, the viruses of bacteria, are proposed to drive bacterial population dynamics, yet direct evidence of their impact on natural populations is limited. Here we identified viral sequences in a metapopulation of wild plant-associated *Pseudomonas* spp. genomes. We discovered that the most

abundant viral cluster does not encode an intact phage but instead encodes a tailocin - a phage-derived element that bacteria use to kill competitors for interbacterial warfare. Each pathogenic *Pseudomonas* sp. strain carries one of a few distinct tailocin variants, which target variable polysaccharides in the outer membrane of co-occurring pathogenic strains. Analysis of historic herbarium samples from the last 170 years revealed that the same tailocin and receptor variants have persisted in the *Pseudomonas* populations for at least two centuries, suggesting the continued use of a defined set of tailocin haplotypes and receptors. These results indicate that tailocin genetic diversity can be mined to develop targeted "tailocin cocktails" for microbial control.

One-Sentence Summary

Bacterial pathogens in a host-associated metapopulation use a repurposed prophage to kill their competitors.

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- and archaea- infecting viruses, are the most abundant biological entities on Earth (*5*). Phages can be highly specific in the bacteria they infect, targeting only a few strains of a bacterial species (*6*). The machinery of phages has also been repeatedly co-opted by bacteria to aid in killing competitors (*7*) or in colonizing hosts (*8*). Phages and phage-tail-like elements including bacteriocins or tailocins are hypothesized to be drivers of microbiome composition in host and non-host environments (*9*). To date, the impact of phages and their derived elements on plant microbiota has only been demonstrated in the lab under controlled conditions (*10*). Whether phages and their derived elements are a major suppressor of pathogen strains in the wild remains largely a matter of speculation (*9*) (but see (*11*, *12*)).

We previously found that wild populations of the plant *Arabidopsis thaliana* were colonized by a metapopulation of *Pseudomonas* spp. bacterial pathogens, a genetically-diverse population of *Pseudomonas* populations (*13*). Our prior results suggested that these pathogen populations do not undergo clonal expansions as is frequently observed in agricultural and clinical pathogens. Infections of *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

91 *Pseudomonas* composition (*14–16*). Because phages and their derived elements are common in
92 *Pseudomonas* populations (*17*) we hypothesized that differences in sensitivity to phage components could
93 suppress specific *Pseudomonas* spp. strains in our system.
94

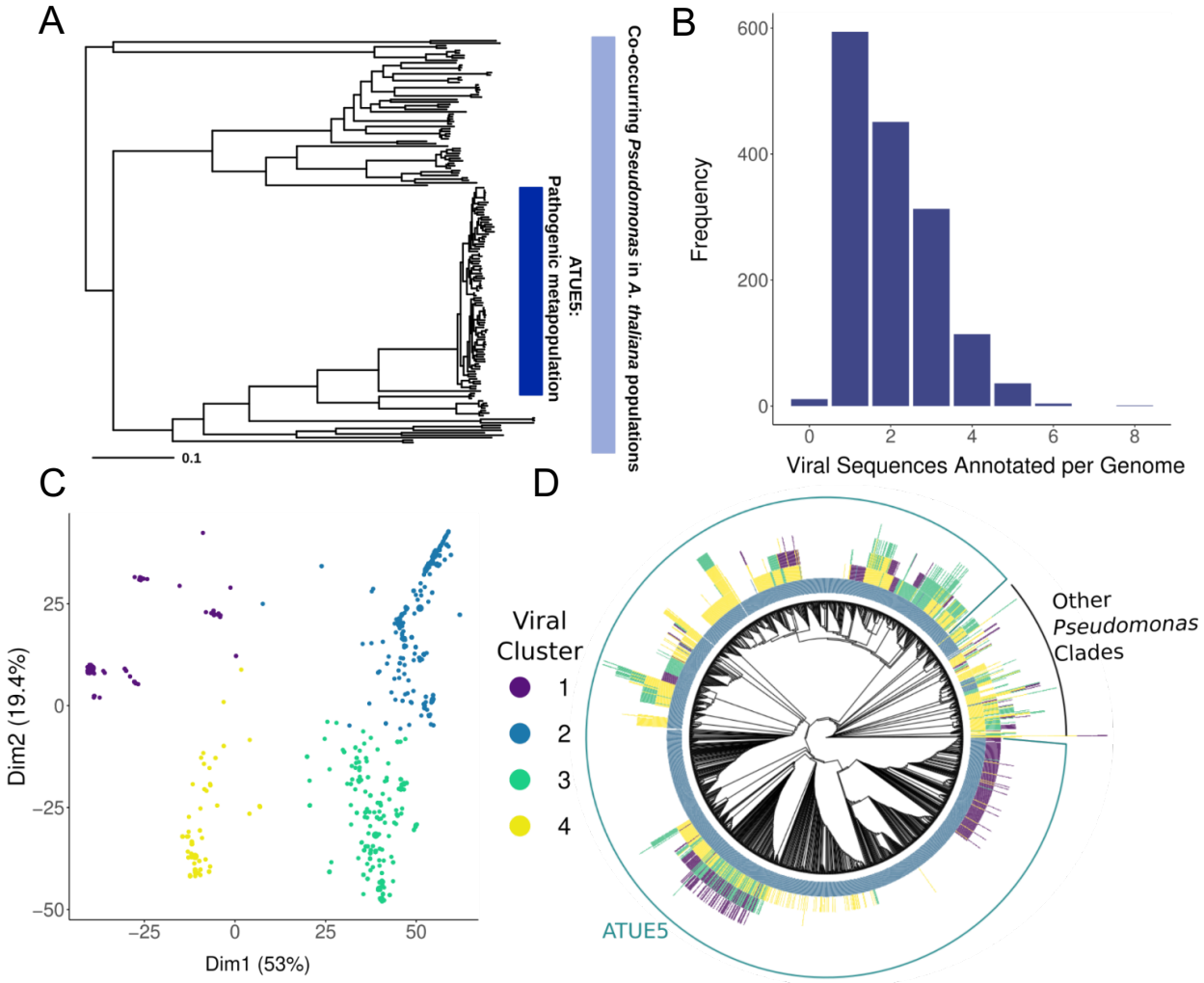


Figure 1. The pathogenic clade of *P. viridiflava* ATUE5 harbors a single highly conserved viral cluster (VC2). A. Maximum likelihood concatenated core genome phylogeny of 1,524 pseudomonads that co-occur in the *Arabidopsis thaliana* phyllosphere (light blue bar). The dark blue bar indicates a pathogenic 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).

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

To determine whether phages and their derived elements cause turnover in the *Pseudomonas* colonizing *A. thaliana*, we first sought to characterize the abundant phage sequences in a wild *Pseudomonas* metapopulation. Viral sequences integrated within a bacterial genome are the result of lysogeny with a compatible phage (*18*). Identifying compatible lysogeny events can reveal clade-specific phage enrichment. We characterized the presence and evolution of viral elements in 1,524 *Pseudomonas* spp. genomes all collected from *A. thaliana* in Southwestern Germany (Fig. S1) (*19*). 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 (*20*, *21*). We found viral sequences in 99.3% of all the genomes with an average of two viral sequences per genome (Fig. 1B). By using pairwise k-mer distances (*22*) 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 less frequent (29%) (Fig. 1D) in non-pathogenic isolates outside of the ATUE5 clade. Furthermore, the gene composition and synteny of VC2 is conserved across ATUE5 strains but has divergent sequences outside the closest relatives to ATUE5 (Fig. 2B). 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 (*23*, *24*). 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 and encapsulation such as capsid, terminase, integrase, and recombinase proteins, suggesting that the VC2 genomic region does not encode a fully functional phage (*25*, *26*). 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 (*27*). Tailocins are repurposed phages that bacteria use to kill co-occurring bacterial strains (*23*) using the targeting 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) (*28*), partially purified the lysate, and performed Transmission Electron Microscopy (TEM). Transmission electron micrographs (Fig. 2C, Fig. S2) 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. 2C). The rod-like structures suggest this tailocin is an R-type contractile tailocin (*29–31*). 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 intact phage or bacteriocin-like structures were observed in the TEM micrographs.

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 (*23*, *32*). Proteomics data confirmed the presence of tailocin proteins annotated in VC2 (Fig. 2B, Fig. S2 and Table S1). No tailocin proteins were found in the sample that was not induced with mitomycin C (MMC) and no other viral

140 proteins or proteinaceous toxins, such as S-bacteriocins, were detected in the tailocin lysates. Together,
141 these results suggest the highly conserved ATUE5 viral sequence is an active and inducible R-type tailocin.

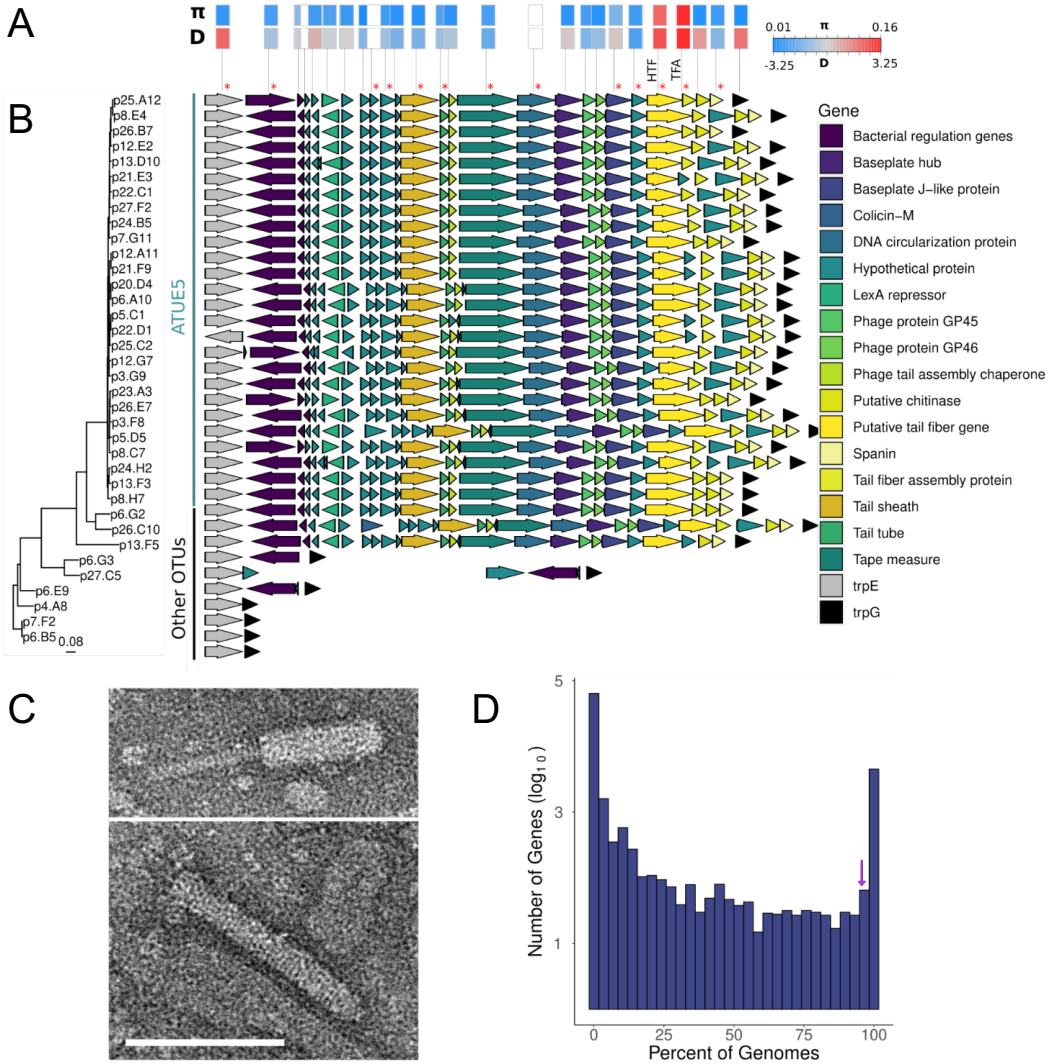


Figure 2. Viral cluster 2 encodes a structurally functional tailocin. A. Within-lineage measures of nucleotide diversity (measured as pairwise nucleotide differences (π)) and the site frequency spectrum measured as Tajima's D (D) for each tailocin gene. White boxes indicate genes that are present in less than 5% of strains (statistics are excluded). Red asterisks (*) indicate proteins found in proteomics analysis. **B.** 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). **C.** 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, 100nm, applies to all micrographs. **D.** 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 purple arrow in the histogram).

Co-occurring *Pseudomonas* spp. genomes encode highly diverse tailocin variants

The phylogenetic distribution of the VC2 cluster (Fig. 1, Fig. 2B) suggested that the VC2 genomic island is well-conserved across pathogenic ATUE5 strains. Across the 1,524 *Pseudomonas* spp. genomes analyzed (Fig. 2D) the full complement of the 27 tailocin genes is found in more than 95% of ATUE5 genomes suggesting the tailocin is in the core genome of ATUE5 strains along with essential bacterial genes (*33*). The conservation of the tailocin protein complex across all ATUE5 genomes, its conserved content and its consistent location between *trpE/trpG (27)* suggests a single integration of the VC2 tailocin in the common ancestor of ATUE5 that is likely important for fitness in its natural ecological context.

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. 2A, Fig. S3A), 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. 2A, Fig. S3B), 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 (*34*), 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* spp. strains. The HTF haplotypes exhibited polymorphism in terms of their sequence length, with the majority falling into four distinct length categories (Fig. 3A). The HTF and TFA haplotypes are significantly associated (Fig. S4, Fisher's Exact test, p-value = 0.0005). 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.

Different tailocin variants kill closely related pathogenic *Pseudomonas*

We hypothesized that like other tailocins, the tailocin variants encoded by the *Pseudomonas* spp. strains are used for interference competition against closely related competitor strains (*23*, *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 (ATUE5, <1% genome wide sequence divergence) to distant commensals (other ATUEs, >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

186 3 strains and 2 representative HTF length haplotypes were induced and partially purified, then concentrated
187 tailocin lysates were applied to 55 ATUE5 strains, 28 other ATUE strains, and over 50 other bacterial
188 strains isolated from the phyllosphere (taxonomy unknown). We found that, as expected, ATUE5 strains
189 were resistant to their own produced tailocin (*23*). The killing assay also revealed that non-ATUE5
190 *Pseudomonas* strains were sensitive in only 6% of ATUE5-derived tailocin treatments, while 40% of
191 ATUE5 treatments with ATUE5-derived tailocins exhibited sensitivity (Fig. 3B, Fig. S6). To test whether
192 other non-*Pseudomonas* community members were sensitive to the tailocin treatments, we tested fifty
193 colonies from the surrounding phyllosphere, and found that none were sensitive to tailocin treatment. An
194 agriculturally relevant *Pseudomonas* strain, DC3000, was also tested and found to be sensitive to one of the
195 tested tailocins (p25.A12). In conclusion, the different tailocin variants appear to exhibit partial killing
196 specificity to different subsets of other ATUE5 pathogenic *Pseudomonas*, only 6% of the other
197 *Pseudomonas* commensals, and not to other phyllosphere strains except for the above mentioned DC3000
198 case.
199
200 To verify that the tailocin is responsible for the observed killing activity, we generated a TFA-deficient
201 mutant. Tailocins collected from the mutant (p25.A12ΔTFA) lost killing activity against all strains tested
202 (Fig. 3C, Fig. S5), and the killing phenotype could be restored with an overexpression plasmid containing
203 the TFA gene. These results suggest the tailocin is necessary for the observed killing activity.
204
205 To determine whether the TFA or HTF haplotypes are associated with killing activity, we analyzed whether
206 the tailocin killing spectrum correlates with a bacterial tester strain's TFA or HTF haplotype. Rather than
207 a correlation, we found an array of different killing spectra, with the tail TFA/HTF haplotypes tested all
208 displaying broad killing against other ATUE5 strains (Fig. 3B and Fig. S6). Although, we found there to
209 be no significant associations between killing spectra and TFA or HTF haplotype, we did identify a
210 significant association between the length of the HTF and its killing spectrum (Fig. 3A and 3D, Fisher's
211 Exact Test $p = 10^{-8}$.). This suggests that tail fiber sequence length could be important in binding to target
212 cells and is likely associated with an outer membrane receptor gene.
213
214 We next aimed to determine whether tailocin killing is observed in the plant host. We performed
215 competition assays *in planta*. Strain p25.A12's tailocin was able to kill a particularly pathogenic strain,
216 p25.C2, in all three killing assays (Fig. 3B, 3rd row). To test whether the target strains would be killed by
217 p25.A12, we grew strains together using a variant of p25.C2 tagged with luciferase to measure its growth.
218 As a control, we also competed p25.C2 with p25.A12ΔTFA. Five ratios of p25.A12 and p25.C2 were used
219 to determine whether the concentration of the tailocin strain affected killing, while controlling the
220 abundance of the target strain. We found that p25.A12 significantly reduced the growth of p25.C2 in the
221 three highest ratios tested (Fig. 3E). The 10:10 and 3:10 ratios significantly differed from the p25.A12ΔTFA
222 competitive assay, suggesting that having a functional tailocin suppresses growth of p25.C2 and that
223 p25.A12 is no longer able to significantly reduce the growth of p25.C2 when lacking the tailocin. We found
224 that p25.C2 grown with p25.A12ΔTFA was never significantly different than p25.C2 grown alone,
225 suggesting that the tailocin is an effective competitive mechanism in the plant host. These results were
226 replicated in competition assays done in test tubes (Fig. S7), however in these trials p25.A12ΔTFA
227 competing with p25.C2 significantly differed from the control for the highest ratio. This result could
228 indicate that p25.A12 utilizes another competitive mechanism in the absence of tailocin, which proves

229 ineffective at lower starting abundances. Together, these results suggest that tailocin killing is important in
230 killing closely related competitors *in planta*.
231
232 **The tail fiber is co-evolving with its O-antigen receptor in natural populations**
233
234 The differences we observed between strains in their susceptibility to the tailocin (Fig. 3B)
235 indicated that these strains may encode an unidentified receptor that has different variants. To
236 identify the receptor, we conducted a transposon mutagenesis screen (*36*) in which we screened a
237 mutant library of a susceptible strain (p25.C2) for insertions that render the strain resistant to the
238 p25.A12 tailocin (Fig. 4A). Insertions (putative knockouts) in seventy genes were significantly
239 associated with resistance to the tailocin. Six of these genes lie within an O-antigen biosynthesis
240 gene cluster, a locus previously implicated in resistance to phage and tailocin tail fibers (*37*). These
241 results suggested that differences between strains in their O-antigen may explain differences in
242 tailocin susceptibility. Comparative DOC-PAGE analysis of the lipopolysaccharide (LPS) isolated
243 from two resistant and two susceptible strains showed the resistant strains lack a high molecular
244 weight polysaccharide chain that is present in the susceptible strains. Subsequent chromatography-
245 mass spectrometry (GC-MS) (Fig. 4C) revealed the susceptible strains had relatively high amounts
246 of rhamnose compared to the resistant strains (Table S2). Our finding of the higher content of
247 rhamnose, along with the expanded size of the LPS in tailocin resistant strains, agrees with the
248 previous studies suggesting the affected gene locus might be involved in the biosynthesis of the
249 O-antigen in *Pseudomonas* (*38–40*). Here, we propose the O-antigen is likely a receptor for the
250 focal tailocin.
251
252 Given that strains are resistant to their own tailocins (*41*), we posited that the tailocin and receptor
253 loci should co-evolve. To test this, we measured congruence between tail fiber and LPS haplotypes
254 (Fig. 4B) and found the tail fiber and LPS haplotypes to statistically co-vary (Fisher's exact test p-
255 values < 2.72e-10 for every association between presence/absence of the gene and the length-based
256 clusters; Fig. 4B). Knowing the tail fiber haplotype allows for prediction of the LPS haplotype. In
257 total, these experiments support a model in which the tail fiber and LPS of co-occurring
258 *Pseudomonas* strains are evolving in concert in wild *A. thaliana* populations (Fig. 4D).
259
260 **The divergent tailocin and LPS haplotypes have been maintained across time and space**
261
262 To ascertain the timescales of the emergence of the tailocin and different tailocin variants, we sought to
263 retrieve *Pseudomonas* genomes from historical *A. thaliana* herbarium specimens collected in the last two
264 centuries. Only historical *Pseudomonas* genomes can reveal whether extant populations are the result of
265 population continuity or population replacement. We employed state-of-the-art ancient DNA retrieval and
266 sequencing techniques (*42*) to sequence leaves of 35 *A. thaliana* herbarium specimens using shot-gun
267 genomics (Table S3). We mapped the herbarium-derived reads to the ATUE5 reference genome strain to
268 identify *Pseudomonas* reads and found seven samples with more than 60% of the genome covered by at
269 least one sequencing read (mean genome coverage 0.97-8.7x) (Fig. S8, Table S3). The *Pseudomonas*-
270 derived reads showed patterns of DNA damage and fragmentation typical of ancient DNA (*43*), which

authenticated their historical origin (Fig. S9). Our analysis revealed that herbarium specimens preserve the genomes of highly abundant bacterial genera of the *A. thaliana* phyllosphere.

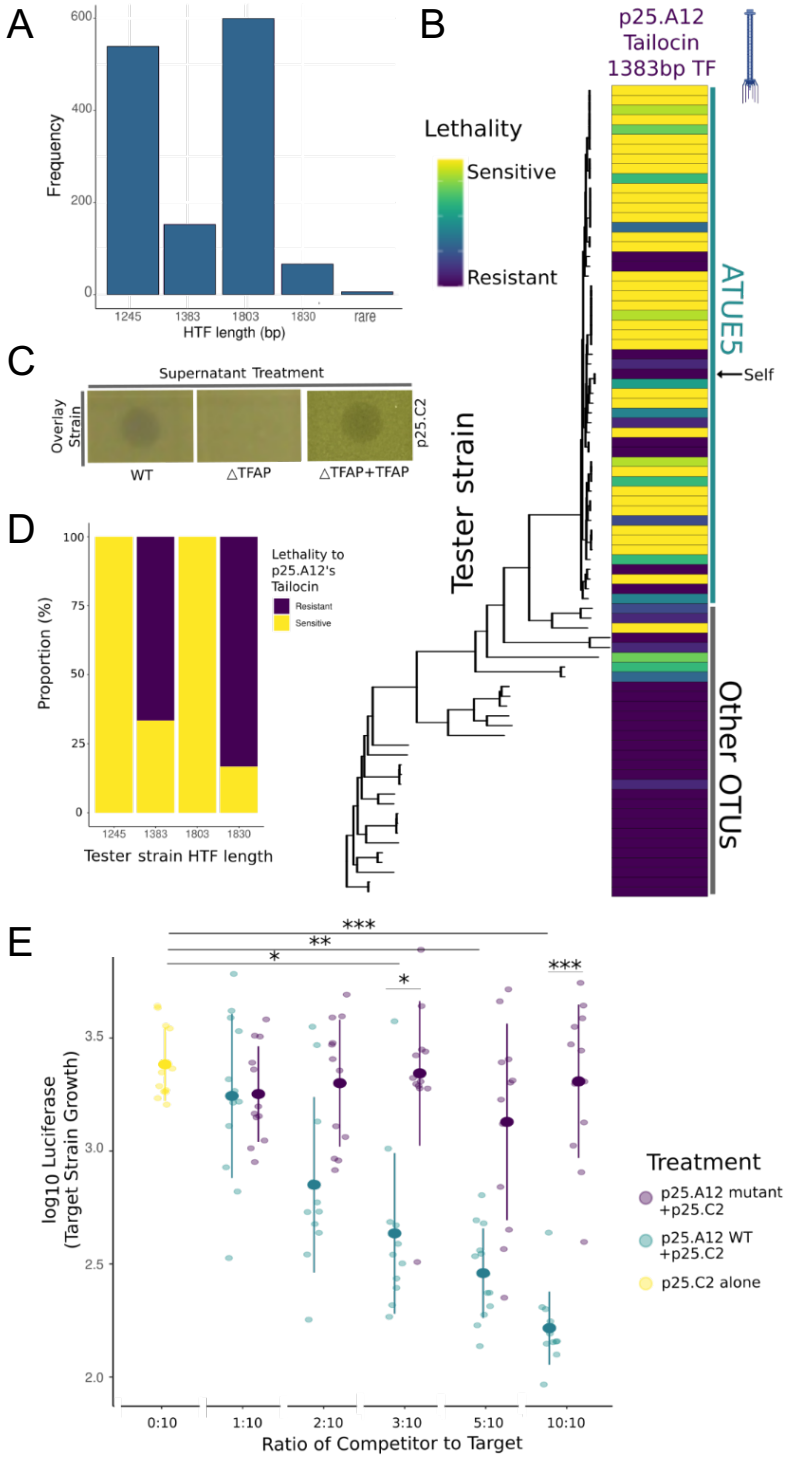


Figure 3. Tailocins target closely-related pathogens. A. Frequency of HTF nucleotide sequence lengths. There are four highly conserved lengths within the *Pseudomonas* populations. **B.** Tailocins are preferentially used for intra-lineage killing. Soft agar cultures of the *Pseudomonas* strains (rows) were challenged with viral particles extracted from cultures of one strain, p25.A12 (from the 1383bp hypothetical tail fiber length haplotype, column), in 3 technical afor each of 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 the strain's own tailocin are indicated by the black arrow pointing to self. For each replicate, a strain was given a score of 3 for clear zone of inhibition, a 2 for semi-clear, a 1 for opaque, or a 0 for no killing, and then added together after all 3 replicates. **C.** 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. Complementation of the gene on an overexpression plasmid in the knockout strain restores the killing phenotype. **D.** The proportion of tester strains sensitive or resistant to p25.A12's tailocin, 1383bp hypothetical tail fiber, significantly correlated with the tester strain's hypothetical tail fiber length haplotype (Fisher's Exact test, $p = 10^{-8}$). **E.** *In planta* coinfections of p25.A12 (competitor) and p25.C2 (target, known to be sensitive to p25.A12). Different ratios of competitor and constant amounts of target strain were used. p25.C2 grown alone as the control. ANOVA test, P-values are shown as $p < 0.001$ ***, $p < 0.01$ **, or $p < 0.05$ *.

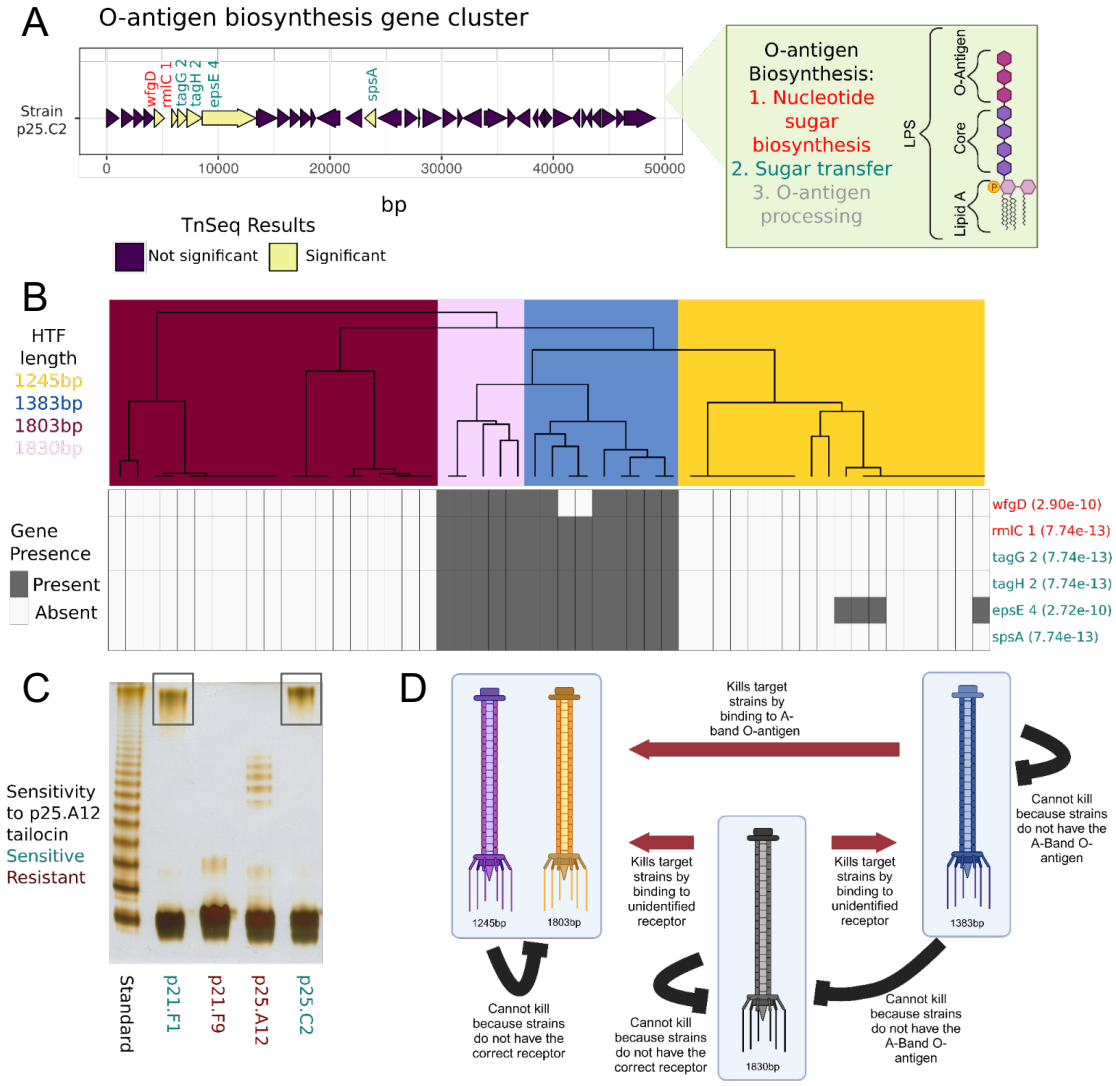


Figure 4. The O-antigen is important for tailocin lethality and is co-evolving with the tailocin HTFP. A. Gene plot of the a previously characterized O-antigen gene cluster and found in ATUE5 strains. Six of the seventy significant TnSeq genes in this study were found in this gene cluster, and are shown in yellow. Gene names are colored by the step in O-antigen biosynthesis pathway. **B.** The top dendrogram is from hierarchical clustering of the *PanKmer* output for the hypothetical tail fiber gene, colored by length. Rows represent the six significant O-antigen genes. Grey corresponds to gene presence and white to gene absence. Fisher's exact test *p*-values between genes and length-based clusters are shown in parenthesis **C.** DOC-PAGE profile of silver stained LPS (1 µg each lane) isolated from a subset of ATUE5 strains. The standard is of Salmonella enterica Ser. Typhimurium (S-type LPS). Grey boxes indicate high molecular weight O-chain. **D.** Our working model of tailocin killing and lethality in *P. viridiflava* isolates. Red arrows indicate known significant patterns of killing. Black lines show known significant patterns of non-killing. Figure was created with BioRender.com.

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

276 dominating *Pseudomonas* spp. strain (Fig. S10). This analysis allowed us to select three historical samples
277 (mean genome depth of 8.2 - 9.6x), which we combined with 65 present-day *Pseudomonas* genomes to
278 build a phylogenetic tree using whole-genome Single Nucleotide Polymorphisms (SNPs). The three
279 historical samples were placed within the present-day ATUE5 diversity (Fig. 5), showing the genetic
280 continuity of this clade for at least the last 177 years. Moreover, the mappings of the historical samples
281 against the ATUE5 reference genome identified the presence of the tailocin in all three historical bacterial
282 genomes (Fig. 5A). In order to ascertain the tailocin variants encoded in the historical genomes and to assess
283 whether the gene colinearity and the insertion location in the bacterial genome are conserved in the
284 historical genomes, we carried out *de novo* assembly of historical *Pseudomonas*-derived reads. We obtained
285 an average of 9 contigs per sample (average sequence length of ~3.4 Kb) that were sufficient to show that
286 both the insertion place of the tailocin and the colinearity of its genes are conserved in the historical
287 genomes (Fig. 5A). Moreover, we ascertained that each historical genome carries a different tailocin variant
288 and that all historic variants are segregating in present-day populations (Fig. 5B, Fig. S11A). The
289 polyphyletic distribution of the HTF and TFA gene haplotypes (Fig. 5B, Fig. S11A) is likely the result of
290 recombination at this locus (*34*).
291
292 We used our phylogenetic reconstruction to determine that the tail fiber assembly gene haplotypes of the
293 historical genomes are closely related to different modern haplotypes (Fig. 5C, Fig. S11B), suggesting that
294 the same haplotypes have circulated in the *Pseudomonas* metapopulation for the past 177 years.
295
296 The observed persistence of HTF length haplotypes within the bacterial metapopulation may result from a
297 potential artifact introduced by population structure at the plant, plant patch, or site level. At these levels,
298 specific haplotypes could reach fixation, creating the appearance of the maintenance of multiple haplotypes
299 when analyzing all populations collectively. To rule out this possibility, we looked at the co-occurence of
300 the HTF length haplotypes at the level of single leaves, all collected within a single site in the Germany
301 population as well as at broader geographic locations through analysis of 10 *Pseudomonas syringae*
302 genomes from NCBI (5 isolates collected from United States, 3 from Europe, and 2 from New Zealand).
303 We found that HTF lengths were co-occurred in single leaves and in all three German populations (Fig.
304 5D). Additionally, we found that the 1245bp, 1383bp, and 1803bp HTF lengths in other countries, in
305 addition to two other lengths. These results suggest that HTF lengths are conserved and maintained at the
306 microscale and over great geographic distances.
307
308 **Discussion**
309
310 Our results have two major implications. First, we discovered that a tailocin, a repurposed phage, is
311 conserved as part of the core genome of metapopulations of the pathogen *P. viridiflava*. Its conservation in
312 presence for at least tens of thousands of years and the sequence conservation of its genes (other than the
313 TFA/HTF) indicate that the tailocin is likely essential (or nearly so) for survival in *Pseudomonas* pathogen
314 metapopulations. This finding is consistent with previous work that suggests that antimicrobial
315 nanomachines such as the Type Six Secretion System are important for survival in the plant environment
316 (*44*). Despite its conservation in presence, multiple allelic variants of the tailocin have been maintained in
317 these populations. We found that the allelic variants target different strains of related pathogenic
318 *Pseudomonas* species/subspecies via differences in the LPS enriched outer leaflet of the outer membrane
319 (*23*). Similar to the coevolution of toxin-antitoxin systems, the tail fiber and assembly proteins of an ATUE5

strain are coevolving with the LPS membrane composition (*45*). The strong evolutionary conservation and correlation between loci suggests that inter-*Pseudomona*s-strain competition via the tailocin is a strong and persistent selective pressure on *Pseudomonas* pathogens.

Second, our results revealed that the same tail fiber/LPS variants have been maintained in the *Pseudomonas* populations for the past two hundred years. Bacteria are capable of rapid adaptation (*46*). 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 (*47*). The presence of a limited set of tail fiber haplotypes in the metapopulation could reflect a limited panel of resistance mechanisms.

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 (*48–50*). Proof-of-concept studies demonstrated that tailocins can be used to suppress specific pathogens in animal (*48*) and plant (*49*) models. With any antimicrobial treatment, there is a risk that the target bacterium will evolve resistance. Our findings indicate that, considering the probable constrained array of resistance mechanisms, the tailocin genetic diversity of a pathogen metapopulation can be mined to discover "tailocin cocktails". These cocktails have the potential to target the metapopulation in parallel, thereby reducing the likelihood of resistance evolution. In the future, the mining and characterization of the tailocin repertoire from diverse wild *Pseudomonas* populations will unveil the possible combinatorics between the tail fiber and LPS composition, as well as the timescales in natural settings over which tailocin resistance emerges.

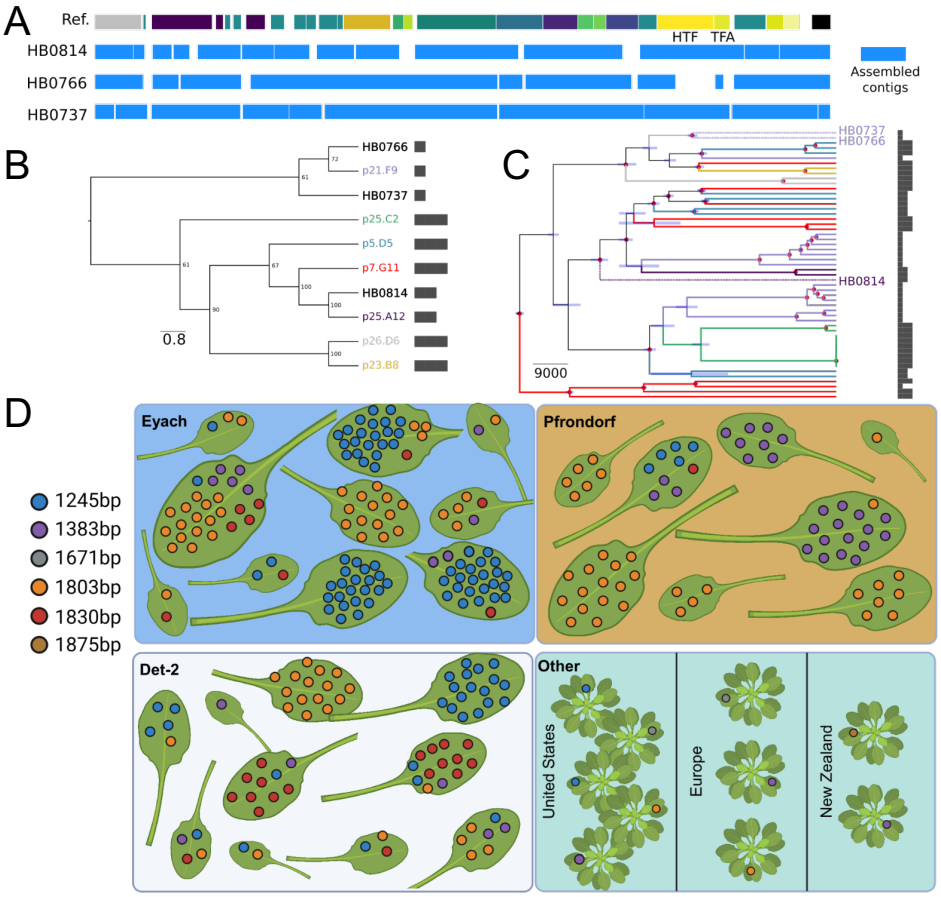


Figure 5. The same length haplotypes of the tail fiber (HTF) are present in contemporary and century-old historical samples and the HTF haplotype diversity is maintained at the leaf scale and broader geographical scales. A. *De-novo* assembled contigs spanning the tailocin genomic region for the historical samples. The reference track on the top represents the genes present in the ATUE5 reference genome. The blue boxes depict the *de-novo* assembled contigs. **B.** Maximum-Likelihood Neighbor-Joining tree of the hypothetical tail fiber gene translated sequence. Historical samples (HB) are placed in the context of the most common haplotypes. HTF lengths are shown with bars. **C.** 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 hypothetical tail fiber gene haplotype. Only historical samples labels 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. HTF lengths are shown with bars. **D.** A subset of eight to ten plants from the three German populations (blue, orange, and white panels), and 10 Pseudomonas syringae isolates collected from plants globally, downloaded from NCBI (green). Each leaf is a leaf from a single plant. Colored dots indicate a single bacterial isolate found and its corresponding HTF length. Plants in the bottom right panel indicate plants not found within the same population. Figure was created with BioRender.com.

Acknowledgements: We thank Sophien Kamoun, Patricia Lang, Gal Ofir, Dmitri Petrov and Detlef Weigel for comments that improved this manuscript.

Funding: This study was supported by startup funds from the University of Utah and by NIH grant R35 GM150722-01 (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. This work was supported in part by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences and Biosciences Division, under award #DE-SC0015662 (PA) at DOE Center for Plant and Microbial Complex Carbohydrates at the CCRC ; and by NIH grant R24GM137782 to (PA) at the National Glycoscience Resource-CCRC Service and Training.

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

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 Table S2. Bacteria were grown on nutrient lysogeny broth agar (LB agar) and in nutrient LB media at 28°C for *Pseudomonas* spp. 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 (*19*) was run on the 1,524 assembled *Pseudomonas* genomes and 10 NCBI 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 (*22*), 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 (*51*, *52*). Silhouette analysis (*53*) 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 (54)*. 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 50 mL fresh LB to extract and isolate tailocins from the various *Pseudomonas* spp. 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.

413
414 **Visualization of viral particles using transmission electron microscopy.**
415 One negatively stained specimen (from strain p25.A12) was prepared and imaged in the following manner:
416 3.5 µl of sample was placed on a carbon-coated, transmission EM grid that had been glow-discharged to
417 make it more hydrophilic. After approximately one minute, filter paper was used to remove excess solution
418 from the grid by touching the side of the grid. This was followed by two steps of brief (approx. 1-2 seconds)
419 wash in P-buffer and blotting with filter paper in the same manner. Next, this process was repeated twice
420 but with the negative stain solution (1% ammonium molybdate). Finally, the grid was incubated for 15-20
421 seconds in a droplet of 1% ammonium molybdate, blotted again, and allowed to air dry. Negatively stained
422 specimens were viewed on a ThermoFisher Tecnai 12 transmission electron microscope operated at 120
423 kV. Images were recorded on a Gatan UltraScan camera.
424
425 **Protein mass spectrometry.**
426 Tailocins partially purified as described above were treated with or without MMC and resuspended in 50
427 µM triethylammonium bicarbonate and 5% SDS (1x S-TRAP buffer) and frozen at -80°C until digestion.
428 Total protein content was quantified using the Pierce™ BCA Protein Assay Kit™ (ThermoFisher Scientific,
429 Waltham, MA) and 10 µg of protein was diluted into 25 µL of 1x S-TRAP buffer. The protein was reduced
430 with 20 mM dithiothreitol at 37°C for 30 minutes and alkylated with 40 mM iodoacetamide at room
431 temperature for 45 minutes in the dark. Samples were acidified with phosphoric acid and digested using
432 micro-S-TRAP™ columns (Protifi, Fairport, NY) with 0.5 µg of trypsin/Lys C per sample according to the
433 manufacturer's instructions. The peptides were dried to completion, resuspended in 300 µL of 0.1%
434 trifluoroacetic acid and desalted using Pierce™ Peptide Desalting Spin Columns (ThermoFisher Scientific,
435 Waltham, MA) according to the manufacturer's instructions. The peptides were resuspended in 40 µL of
436 0.1% formic acid for LC-MS/MS analysis.
437
438 **Mass spectrometry analysis of samples.**
439 Reversed-phase nano-LC-MS/MS was performed on an Dionex UltiMate 3000 RSLCnano system coupled
440 to a ThermoFisher Scientific Q Exactive-HF orbitrap mass spectrometer equipped with a nanoelectrospray
441 source. 1 µg of each sample were first trapped on a 2 cm Acclaim PepMap-100 column (ThermoFisher
442 Scientific, Waltham, MA) with 5% acetonitrile at 5 µl/min and at 5 minutes the sample was injected onto
443 the liquid chromatograph reverse-phase Acclaim™ PepMap™ 100 C18 2.0 µm nanocolumn
444 (ThermoFisher Scientific, Waltham, MA). A 500 mm long/ 0.075 mm inner diameter nanocolumn heated
445 to 35°C was employed for chromatographic separation. The peptides were eluted with a gradient of
446 reversed-phase buffers (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in 100%
447 acetonitrile) at a flow rate of 0.2 µL/min. The LC run lasted for 85 minutes with a starting concentration of
448 5% buffer B increasing to 28% buffer B over 75 minutes, up to 40% buffer B over 10 minutes and held at
449 90% B for 10 minutes. The column is allowed to equilibrate at 5% buffer B for 20 minutes before starting
450 the next data acquisition. The Q Exactive-HF mass spectrometer was operated in data-dependent acquisition
451 MS/MS analysis mode selecting the top 20 most abundant precursor ions between 375-1650 m/z at 60,000
452 resolution for fragmentation at 15,000 resolution.
453
454
455 **Mass spectrometry data analysis.**

456 The raw data was analyzed using Proteome Discoverer 3.0 software with the SEQUEST algorithm against
457 the uniprot_ref_pseudomonas_viridiflava database (1-18-2023 version with 4,389 proteins) or p25.A12
458 database. An allowance was made for 2 missed cleavages following trypsin/Lys C digestion. No fixed
459 modifications were considered. The variable modifications of methionine oxidation and cysteine
460 carbamidomethylation were considered with a mass tolerance of 15 ppm for precursor ions and a mass
461 tolerance of 0.02 Da for fragment ions. The results were filtered with a false discovery rate of 0.01. A
462 minimum of 1 unique peptide should be reported for all proteins identified.
463
464 **Allelic diversity and nucleotide sequence data analysis.**
465 Orthogroups that were previously identified using *panX* were included in the analysis. The nucleotide and
466 peptide sequences were aligned with *Clustal Omega* (*55*), and then a codon-based nucleic acid alignment
467 was generated with *pal2nal* (*56*). This is critical to ensure the nucleic acid alignment is aligned codon-by-
468 codon, to know whether substitutions result in a synonymous or nonsynonymous amino acid change. The
469 output aligned file was saved in fasta format and used as an input file for population summary statistic
470 calculations in R using the package "PopGenome" v2.7.5 (*57*). The level of sequence diversity among
471 ortho-groups was compared using summary statistics 'π' (average pairwise difference or average number
472 of nucleotide diversity per site) (*58*), 'θ' (population mutation parameter or number of segregating sites)
473 (*59*), whereas Tajima's D (*60*) was used as a proxy for the site frequency spectrum (SFS) and thus to
474 determine the demographic and selective forces shaping the SFS of tailocin genes.
475
476 **Ascertainment of tailocin haplotypes.**
477 To determine how many haplotypes of tail fiber assembly and hypothetical tail fiber genes there are in the
478 pathogenic strains in our study, nucleotide and peptide alignments from *Clustal Omega* (*55*) were analyzed
479 using the R package *pegas (61)*. We used the haplotype function to determine how many unique haplotypes
480 are in the ATUE5 strains. The most common haplotypes were used for downstream analyses.
481
482 **Mapping of reads from historical samples to ATUE5.**
483 A set of 35 herbarium *Arabidopsis thaliana* samples collected between 1817 - 1957 from Southern Germany
484 (Table S3) (*62*) were screened for *Pseudomonas* sp. presence. Adapter sequences from the raw reads were
485 trimmed and merged using *Adapterremoval* V.2 (*63*) and subsequently mapped to the *Pseudomonas*
486 p25.A12 assembly using *bwa aln* v.0.7.17 (*64*) with seed deactivation to allow the alignment of of
487 substitutions present at the termini of the reads, *e.g.* those arising from DNA deamination in historic
488 material (*42*). A total of 7 samples for which the *Pseudomonas* reference genome was covered in at least
489 60% were kept (Table S2). To authenticate the historic origin of the bacterial reads, the proportion of C-to-
490 T (or G-to-A) substitutions as well as the fragment size distribution of the reads were computed using
491 mapDamage v.2.2.1 (*65*) (Fig. S8). We analyzed patterns of homozygosity in the mapped reads to keep
492 only samples which were likely dominated by single *Pseudomonas* sp. lineages (Fig. S10). As a result, only
493 the samples HB0737, HB0766 and HB0814 were kept for downstream analyses.
494
495 **Ascertainment of tailocin variants in historical genomes.**
496 In order to identify reads aligning to the whole spectrum of tailocin genetic diversity, different haplotypes
497 of the highly polymorphic tail fiber assembly and hypothetical tail fiber genes (Fig. 3A) were added to the
498 p25.A12 reference genome, before the historical samples were mapped as described above. Reads covering
499 the Tailocin region surrounded by the flanking genes *trpE* and *trpG* as well as all reads covering any

500 haplotype of the mentioned genes were subset from the raw reads and used as input for *de-novo* assembly
501 using *SPAdes* v.3.13.0 (*66*). Samples HB0737, HB0766 and HB0814 yielded 10, 9 and 14 assembled
502 contigs, respectively (Table S4). To find collinearity between the *de-novo* assemblies and the reference
503 Tailocin region, we used *minimap* v.2.1 (*67*). Aminoacid translated sequences of the historical isolate-
504 specific haplotypes for the tail fiber assembly protein and Putative tail fiber protein were merged together
505 with the haplotypes segregating in the observed diversity. Finally, we used *Clustal Omega* v.1.2.4 (*68*) to
506 generate a multiple alignment and classify the haplotypes in the historical samples.
507
508 **Whole-genome phylogenetic reconstruction combined present-day and historical genomes.**
509 A set consisting of 50 ATUE5 modern and 3 historic *Pseudomonas* spp. strains were used to create a
510 phylogenetic reconstruction. We used the mapped reads to perform individual *de-novo* variant callings
511 using *bcftools mpileup* and *bcftools call* v1.11 (*69*) removing those reads with mapping quality values lower
512 than 30. We then merged all individual calls using *bcftools merge* v.1.11 (*69*) and removed those samples
513 with less than 60% of the called variants. We finally kept only biallelic SNPs and sites with a maximum
514 missing information of 10% of the remaining samples. The number of retrieved variants were 176,528 SNPs
515 segregating among 53 strains (50 modern and 3 historic).
516
517 Using the genomic SNPs we sought to reconstruct the phylogenetic relations among the 53 strains. We used
518 *IQ-TREE* V2.2.0.3 ((*70*)) in combination with *ModelFinder* ((*71*)), and ultrafast bootstrap *UFBoot* ((*72*))
519 to estimate a Maximum-Likelihood phylogenetic tree. Variant sites that are likely to arise due to putative
520 recombination were identified using *ClonalFrameML* v.1.12 ((*73*)) and the output was used to mask
521 putative recombinant positions in the original genomic SNP array. We finally used *BEAST2 (74)* to jointly
522 estimate the phylogenetic relations among the strains as well as the relative time of emergence for all the
523 tree nodes. For this purpose we used the collection dates (Table S3) and a previously estimated evolutionary
524 rate ((*75*)) as priors. In order to minimize parameter estimation we chose HYK as the evolutionary model
525 and to avoid demographic history assumptions we used a Coalescent Extended Bayesian Skyline
526 approach(*76*).
527
528 **Testing bacterial sensitivity to tailocins with spot test phenotypic assays.**
529 To test the sensitivity of the different bacterial strains to the tailocins, soft agar assays were performed using
530 an adaptation of a protocol from Vacheron et al. 2021 (*35*). 750 µL overnight cultures of each strain was
531 mixed with 25 mL of LB soft agar (0.8%). The mixture was poured into a square plate and left to harden.
532 Then, aliquots of 3-5 µL of concentrated viral particles suspension were applied to the agar, along with a
533 control of P-buffer alone and non-induced cultures in serial dilutions. The plates were incubated overnight
534 at 30°C. Bacterial sensitivity or resistance to the viral particles was assessed after 24 hours. When testing
535 serially diluted tailocins, no plaques were observed, suggesting the killing agent is nonreplicative and not
536 a phage. When testing the uninduced tailocin control samples (not induced with mitomycin C), if lethality
537 was observed, the interaction was deemed to be inconclusive. Spot tests were performed in three biological
538 and three technical triplicates.
539
540 **Testing bacterial sensitivity to tailocins in culture.**
541 To test the sensitivity of a known sensitive strain (p25.C2) to the mutant tailocin from 25.A12ΔTFA,
542 p25.C2 was grown overnight then diluted by a factor of 1:10 in fresh LB medium. Once the culture reached
543 growth phase, cultures were supplemented with 5% p25.A12 tailocin, 5% p25.A12ΔTFA tailocin, or a

544 buffer control. OD measurements at a wavelength of 600 nm were measured in a 96-well plate using a
545 microplate reader (TECAN Spark).
546
547 **Construction of the *Pseudomonas* p25.A12 tail fiber assembly gene deletion.**
548 The p25.A12ΔTFA mutant strain was cloned using the Gateway Cloning system with the donor vector
549 pDONR1K18ms (Addgene plasmid #72644) and the destination vector pDEST2T18ms (Addgene plasmid
550 #72647).
551
552 **Construction of the *Pseudomonas* p25.A12 tail fiber assembly gene overexpression rescue vector.**
553 The HTF gene was cloned into pMCSG11 (*77*) by restriction cloning (*78*). This rescue vector was
554 transformed into p25.A12ΔTFA, tailocins were induced and partially purified, and killing assays were
555 performed to assess rescue lethality.
556
557 ***In vivo* competition assays.**
558 Strain p25.C2 cultures tagged with a luciferase cassette were grown together with p25.A12 wild type or
559 p25.A12ΔTFA strain. Five different ratios were used for mixing the two strains (10:10, 5:10, 3:10, 2:10,
560 1:10) keeping the p25.C2 strain constant at a final OD_{600} of 0.005. Bacteria were grown overnight and were
561 diluted 1:10 the day of the experiment. Cells were grown for another 3 hours with shaking, and finally
562 collected and mixed accordingly. The co-inoculated cells were grown in a TECAN flat white plate
563 (TECAN, #30122300) in a final volume of 200 uL. The plates were placed in the TECAN plate reader
564 (TECAN Spark) for 21 hours at 28ºC for 55 kinetic cycles. Each cycle consisted of 15 minutes or orbital
565 shaking at 220 rpm, resting for 5 minutes and then the luciferase-tagged (p25.C2 cells) of each well was
566 measured. For the analysis the three highest values from each sample were used and the average luciferase
567 value was calculated for each condition.
568
569 ***In planta* co-infections.**
570 Columbia-0 wild type *A. thaliana* plants were grown in 24-well plates (Greiner Bio-One, #6621665) under
571 long day conditions (16 hours light, 8 hours dark) in an AR41L3 percival with 60% intensity of the SciWhite
572 LED lights. Thirteen day-old seedlings were used for the infections. Plants were infected with bacterial
573 suspension of the *Pseudomonas* strain p25.C2 (tagged with luciferase), and either the p25.A12 wild type or
574 the p25.A12ΔTFA strain. The same five ratios as the *in vivo* assays were used, keeping the p25.C2 constant
575 at a final OD_{600} of 0.005.
576 Bacteria were grown overnight and diluted 1:10 the day of the infection. Cells were grown for another three
577 hours and finally collected and resuspended in 10mM $MgSO_4$. Cells were mixed according to the ratios
578 described above, and the seedlings were submerged in 10 mL of cell suspension for 10 minutes in a
579 randomized manner. 850 µL were removed and the plants were placed back in the percival. Three days post
580 infection plants were collected in 2 mL deep-well plates (Thermo Fisher™ Nunc™, #12-565-605)
581 containing 1 mL $MgSO_4$. Plants were ground using the Qiagen tissue lyser II (QIAGEN) and 200µL of the
582 suspension were used to measure luciferase in the TECAN using flat white plates (TECAN, #30122300).
583 Plants infected with $MgSO_4$ or p25.C2 alone were used as controls. The experiment was repeated 3 times
584 with similar results.
585
586 **Constructing the BarSeq Mutant Fitness Library.**

587 Methods were adapted from Wetmore et al. (*36*). Briefly, we created the p25.C2 transposon mutant library
588 by conjugating p25.C2 with the *E. coli* conjugation donor (WM3064) harboring the pHLL250mariner
589 transposon vector library (AMD290)(*79*). Equal cell numbers of mid-log phase p25.C2 and AMD290 were
590 conjugated for 6 hours on 0.45 µm nitrocellulose filters (Millipore) overlaid on LB agar plates containing
591 Diaminopimelic acid. The resuspended cells were plated on LB plates with 50 g/mL kanamycin to select
592 for mutants. After 2 days, the kanamycin-resistant colonies were scraped into LB, the OD_{600} of the mixture
593 was measured, and the mutant library was diluted to a starting OD_{600} of 0.2 in 250 ml of LB with 50 g/ml
594 kanamycin. The diluted mutant library was grown at 28°C to a final OD_{600} of 1.0, added glycerol to a final
595 volume of 10%, and stored at -80°C freezer stocks. Cells were collected for genomic DNA extraction.
596
597 **Competitive Mutant Fitness Assays.**
598 Assays were adapted from (*23*). Briefly, assays were performed in glass tubes. Partially purified tailocins
599 were used as stressors at final concentrations of 0.05x of the stock preparation. P-buffer was used as a
600 control. LB + Km was supplied for the mutant library as the growth medium. The tubes were incubated
601 with shaking and mutants were harvested when the OD600 reached mid-log phase. Cells were pelleted by
602 centrifugation (8000g, 5 min) and stored at -20°C awaiting genomic DNA extraction. Each condition was
603 assayed in ten individual replicates.
604
605 **BarSeq and Analysis of BarSeq data.**
606 Genomic DNA was extracted and barcode PCR was performed as described in (*36*). DNA extractions were
607 quantified with NanoDrop 1000 (Thermo Fisher). Barcode sequence data were obtained by multiplexing
608 on a lane of HiSeq at the University of Utah High-Throughput Genomics core. Fitness data were calculated
609 and analyzed from these reads with DESeq2 (*80*) R package and scripts can be found on our GitHub page.
610
611 **LPS isolation.**
612 The cell culture was washed two times with PBS and inactivated using a solution of 1% phenol in PBS.
613 The cell material was suspended in water and pre-incubated to 68 °C with gentle stirring. The extraction
614 was carried out by using the corresponding volume of 90% (vol/vol) liquefied phenol following the standard
615 procedure (*81*) and extracted for 20 min, at 68 °C. The sample was cooled on an ice bath and centrifuged
616 for 20 min at 4 °C, 5000 x g. The aqueous upper phase was collected, and the remaining phenol phase and
617 cell debris were extracted two more times with water, following the same protocol. The combined aqueous
618 phases and the final phenol phase were dialyzed against several exchanges of water, for 4 days, (MWCO).
619
620 The freeze-dried dialysates were re-suspended in a sterile buffer (10 mg/mL - 50 mM MgCl2•6H2O and
621 20 mM NaOAc•3H2O) and the nucleic acids were digested with Benzonase (2 µl, 18 h at 37 °C) with gentle
622 agitation. Proteins were and digested with proteinase K (50 µg/mL - per mL of the LPS solution) by
623 overnight digestion at 37 °C. The buffer, nucleotides, and peptides were removed by dialysis against water
624 (12-14 kDa MWCO) at 4 °C, followed by ultracentrifugation at 100,000 x g for 16 h, at 4 °C. The LPS
625 pellets were suspended in water, freeze-dried, and used in this work.
626
627 **DOC-PAGE of LPS samples.**
628 The purified LPS samples were resolved in PAGE (4% stacking gel and 18% resolving gel) in the presence
629 of the deoxycholic acid buffer ((*82*)). The PAGE was visualized with a silver stain reagent kit (Bio-Rad).
630

631 **Glycosyl composition analysis by TMS method**

632 The analysis was performed by combined gas chromatography-mass spectrometry (GC-MS) of the O-
633 trimethylsilyl (TMS) methyl glycoside derivatives produced from the sample by acidic methanolysis
634 followed by trimethylsilyation(*83*) . GC-MS analysis of the TMS methyl glycosides was performed on an
635 Agilent AT 7890A GC system interfaced to a 5975B MSD, using an EC-1 fused silica capillary column
636 (30 m ´ 0.25 mm ID), and the temperature gradient was was 80°C for 2 min, then ramped to 140°C at
637 20°C/min with 2 min hold, and to 200°C at 2°C/min, followed by an increase to 250°C at 30°C/min with 5
638 min hold.
639
640

641 **References**

642 1. E. M. Goss, S. Timilsina, A bacterial epidemic in wild plants. *Nat Ecol Evol*. **2** (2018), pp. 1529–
643 1530.

644 2. T. L. Karasov, J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, D. S.
645 Lundberg, M. Neumann, J. Regalado, R. A. Neher, E. Kemen, D. Weigel, Arabidopsis thaliana and
646 Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host*
647 *Microbe*. **24**, 168–179.e4 (2018).

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

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

652 5. M. R. Clokie, A. D. Millard, A. V. Letarov, S. Heaphy, Phages in nature. *Bacteriophage*. **1**, 31–45
653 (2011).

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

656 7. P. G. Leiman, M. Basler, U. A. Ramagopal, J. B. Bonanno, J. M. Sauder, S. Pukatzki, S. K. Burley,
657 S. C. Almo, J. J. Mekalanos, Type VI secretion apparatus and phage tail-associated protein
658 complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences*.
659 **106**, 4154–4159 (2009).

660 8. B. T. Ho, T. G. Dong, J. J. Mekalanos, A view to a kill: the bacterial type VI secretion system. *Cell*
661 *Host Microbe*. **15**, 9–21 (2014).

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

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

667 11. B. Koskella, Phage-mediated selection on microbiota of a long-lived host. *Curr. Biol.* **23**, 1256–1260
668 (2013).

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

671 13. I. Hanski, Metapopulation dynamics. *Nature*. **396**, 41–49 (1998).

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

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

16. 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).

17. 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).

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

19. 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).

20. 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).

21. 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.

22. 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).

23. 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).

24. 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).

25. 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).

26. 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).

27. 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.

28. 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.

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

30. 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).

31. 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).

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

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

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

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

724 36. K. M. Wetmore, M. N. Price, R. J. Waters, J. S. Lamson, J. He, C. A. Hoover, M. J. Blow, J.
725 Bristow, G. Butland, A. P. Arkin, A. Deutschbauer, Rapid Quantification of Mutant Fitness in
726 Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *MBio*. **6** (2015),
727 doi:10.1128/mBio.00306-15.

728 37. J. Jayaraman, W. T. Jones, D. Harvey, L. M. Hemara, H. C. McCann, M. Yoon, S. L. Warring, P. C.
729 Fineran, C. H. Mesarich, M. D. Templeton, Variation at the common polysaccharide antigen locus
730 drives lipopolysaccharide diversity within the Pseudomonas syringae species complex. *Environ.*
731 *Microbiol.* **22**, 5356–5372 (2020).

732 38. D. A. Baltrus, S. Weaver, L. Krings, A. E. Nguyen, Genomic Correlates of Tailocin Sensitivity in
733 Pseudomonas syringae. *bioRxiv* (2023), p. 2023.04.24.538177.

734 39. J. D. King, D. Kocíncová, E. L. Westman, J. S. Lam, Review: Lipopolysaccharide biosynthesis in
735 Pseudomonas aeruginosa. *Innate Immun.* **15**, 261–312 (2009).

736 40. T. Köhler, V. Donner, C. van Delden, Lipopolysaccharide as shield and receptor for R-pyocin-
737 mediated killing in Pseudomonas aeruginosa. *J. Bacteriol.* **192**, 1921–1928 (2010).

738 41. M. G. K. Ghequire, R. De Mot, The Tailocin Tale: Peeling off Phage Tails. *Trends Microbiol.* **23**,
739 587–590 (2015).

740 42. S. M. Latorre, P. L. M. Lang, H. A. Burbano, R. M. Gutaker, Isolation, library preparation, and
741 bioinformatic analysis of historical and ancient plant DNA. *Curr. Protoc. Plant Biol.* **5**, e20121
742 (2020).

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

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

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

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

753 47. V. Druelle, R. A. Neher, Reversions to consensus are positively selected in HIV-1 and bias

754 substitution rate estimates. *bioRxiv* (2022), p. 2022.02.13.480259.

755 48. D. Gebhart, S. Lok, S. Clare, M. Tomas, M. Stares, D. Scholl, C. J. Donskey, T. D. Lawley, G. R.
756 Govoni, A modified R-type bacteriocin specifically targeting Clostridium difficile prevents
757 colonization of mice without affecting gut microbiota diversity. *MBio*. **6** (2015),
758 doi:10.1128/mBio.02368-14.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

790 65. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast
791 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–1684
792 (2013).

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

795 67. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–3100
796 (2018).

797 68. F. Sievers, D. G. Higgins, The Clustal Omega Multiple Alignment Package. *Methods Mol. Biol.*
798 **2231**, 3–16 (2021).

799 69. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane,
800 S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience*. **10**
801 (2021), doi:10.1093/gigascience/giab008.

802 70. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R.
803 Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic
804 Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

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

807 72. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the
808 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

809 73. X. Didelot, D. J. Wilson, ClonalFrameML: efficient inference of recombination in whole bacterial
810 genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).

811 74. A. J. Drummond, R. R. Bouckaert, *Bayesian Evolutionary Analysis with BEAST* (Cambridge
812 University Press, 2015).

813 75. E. P. McCann, K. L. Williams, J. A. Fifita, I. S. Tarr, J. O'Connor, D. B. Rowe, G. A. Nicholson, I.
814 P. Blair, The genotype-phenotype landscape of familial amyotrophic lateral sclerosis in Australia.
815 *Clin. Genet.* **92**, 259–266 (2017).

816 76. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian Coalescent Inference of Past
817 Population Dynamics from Molecular Sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).

818 77. W. H. Eschenfeldt, S. Lucy, C. S. Millard, A. Joachimiak, I. D. Mark, A family of LIC vectors for
819 high-throughput cloning and purification of proteins. *Methods Mol. Biol.* **498**, 105–115 (2009).

820 78. D. Nathans, H. O. Smith, Restriction endonucleases in the analysis and restructuring of dna
821 molecules. *Annu. Rev. Biochem.* **44**, 273–293 (1975).

822 79. B. A. Adler, A. E. Kazakov, C. Zhong, H. Liu, E. Kutter, L. M. Lui, T. N. Nielsen, H. Carion, A. M.
823 Deutschbauer, V. K. Mutalik, A. P. Arkin, The genetic basis of phage susceptibility, cross-resistance
824 and host-range in Salmonella. *Microbiology*. **167** (2021), doi:10.1099/mic.0.001126.

825 80. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq
826 data with DESeq2. *Genome Biol.* **15**, 550 (2014).

81. O. Westphal, Bacterial lipopolysaccharides : extraction with phenol-water and further applications of the procedure. *Methods Carbohydr. Chem.* **5**, 83 (1965).

82. M. A. Austin, J. L. Breslow, C. H. Hennekens, J. E. Buring, W. C. Willett, R. M. Krauss, Low-density lipoprotein subclass patterns and risk of myocardial infarction. *JAMA*. **260**, 1917–1921 (1988).

83. W. S. York, A. G. Darvill, M. McNeil, T. T. Stevenson, P. Albersheim, "Isolation and characterization of plant cell walls and cell wall components" in *Methods in Enzymology* (Academic Press, 1986), vol. 118, pp. 3–40.

Figure S1. The two modern and three historic collection sites in Southwestern Germany.

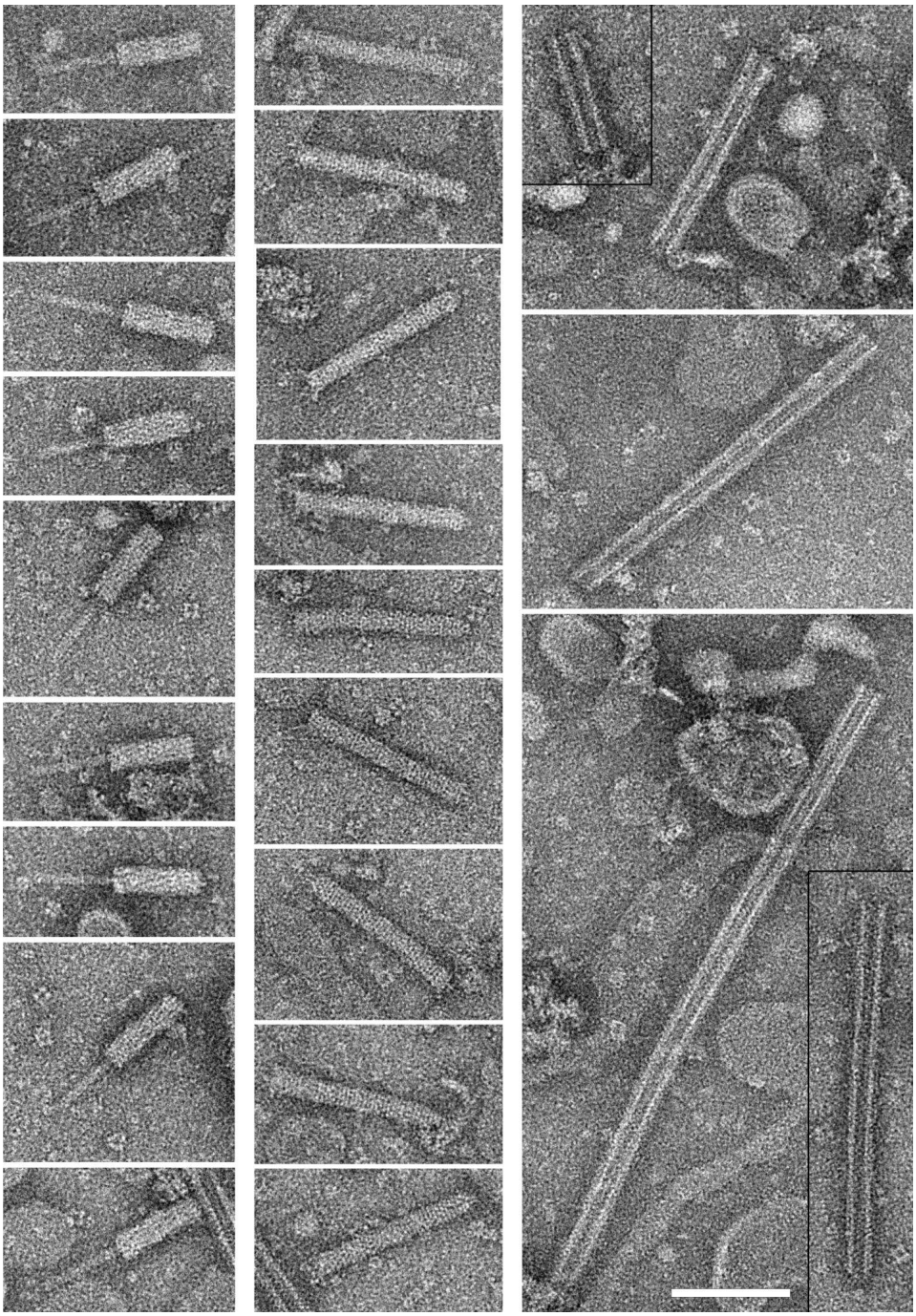


Figure S2. TEM images of p25.A12 contracted tailocins (left column), uncontracted tailocins (center column), and phage-like hollow tubes (right column). In the central column, the baseplate end of the tailocin faces left. The hollow tubes had highly variable lengths, average 404 (±153) nm (n = 14). Scale bar, 100 nm.

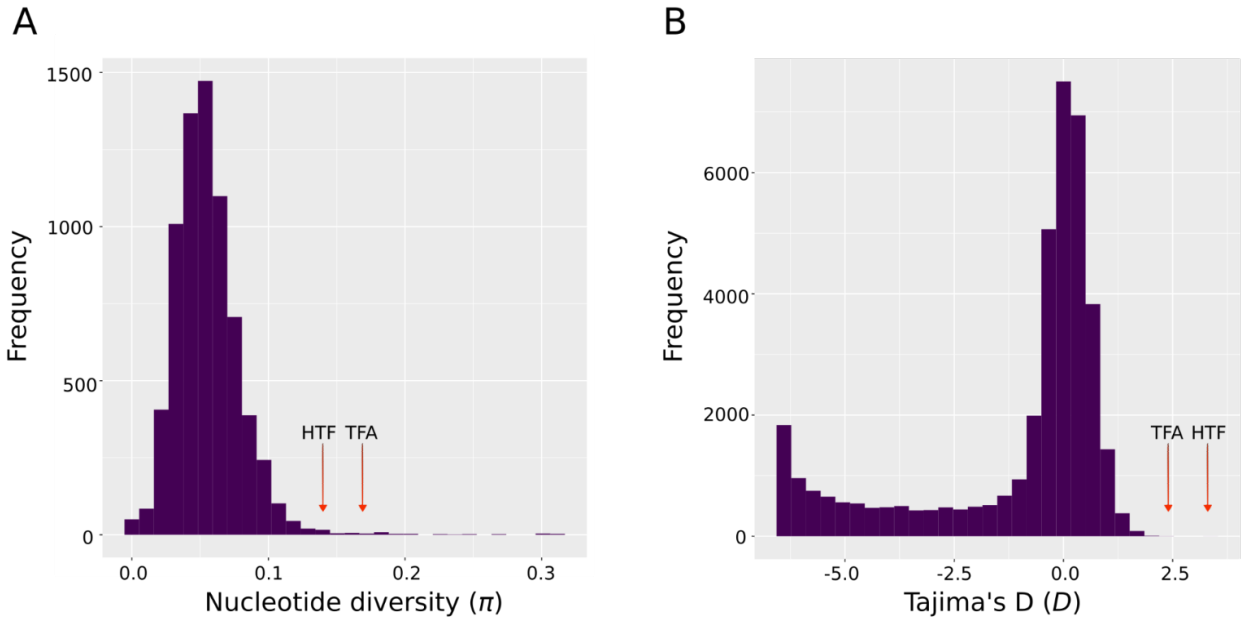


Figure S3 The tail fiber assembly gene and hypothetical tail fiber are evolving quicker than the rest of the bacterial genome. The distribution of **A.** nucleotide diversity (π) and **B.** Tajima's D (D) for all core genes in the 1,399 ATUE5 genomes. The TFA and HTF genes (red arrows) have significantly higher π and D compared to the rest of the core genes.

847
848
849
850
851
852

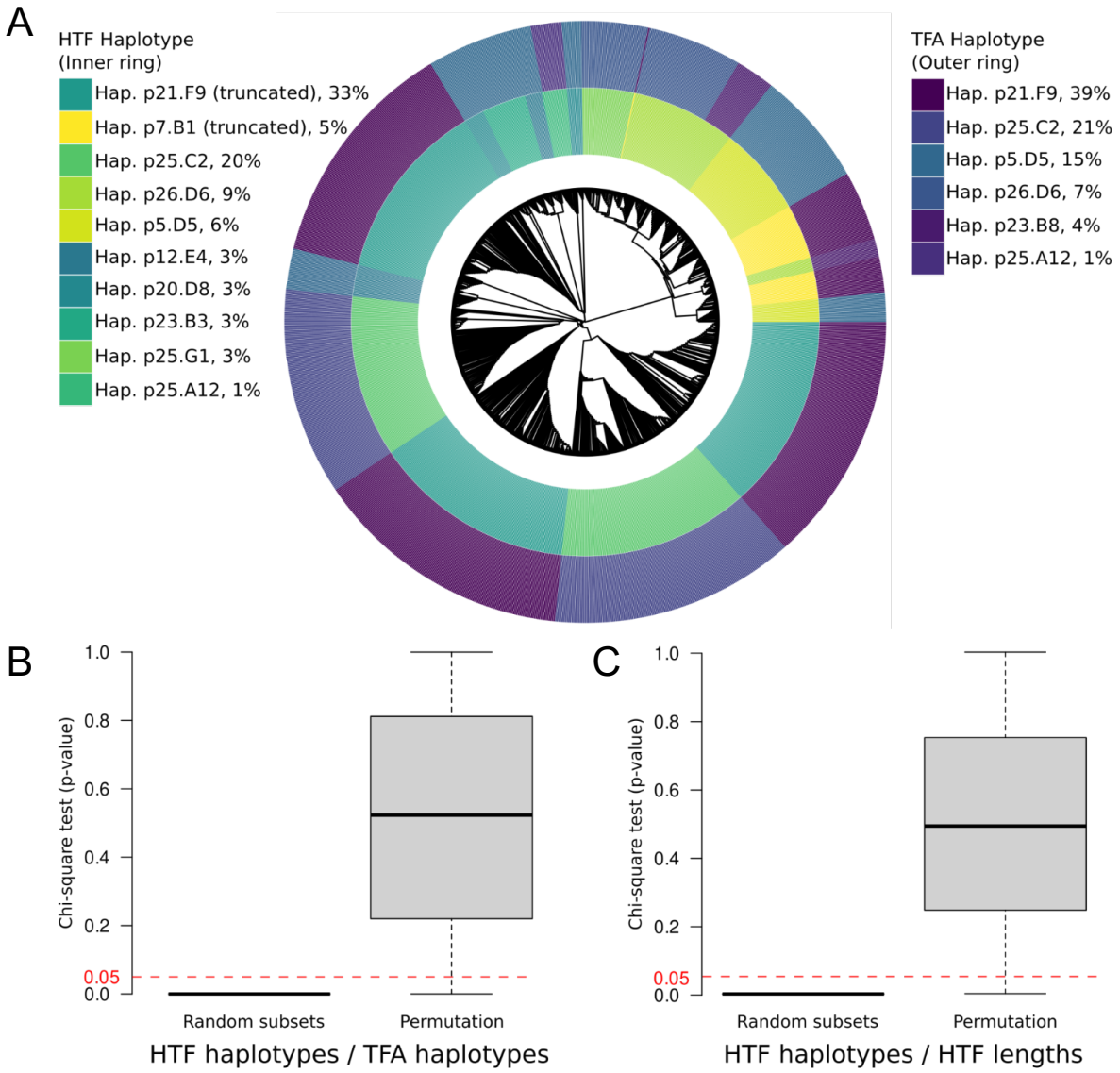
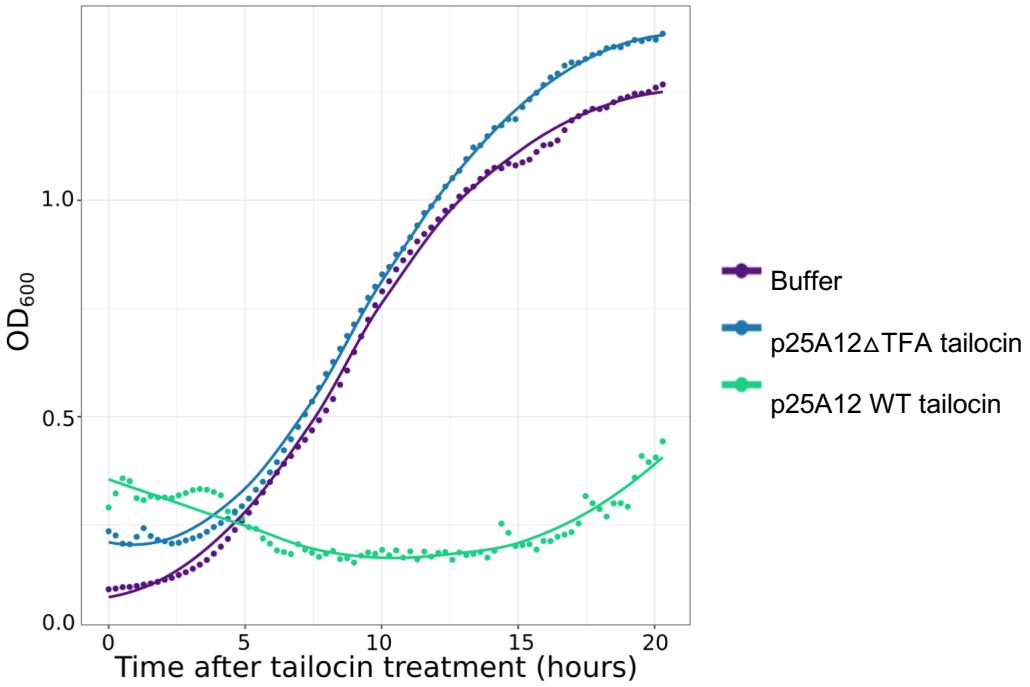


Figure S4. Associations between Tail Fiber Assembly (TFA) and Hypothetical Tail Fiber (HTF) haplotypes are significant. A. Maximum likelihood concatenated core genome phylogeny of 1,399 pathogenic pseudomonads that co-occur in the *Arabidopsis thaliana* phyllosphere with the HTF and TFA haplotypes plotted in the inner and outer rings. B-C. Each box plot represents a distribution of 1000 Chi-square test p-values drawn from random draws of pairs between TFA and HTF haplotypes and data sets for which HTF values were randomly permuted. Associations were tested between the two protein haplotypes (A) and between HTF haplotypes and TFA gene lengths (B).

853



854 **Figure S5. The tail fiber assembly gene was successfully knocked out and the knockout strain loses**
855 **killing ability.** A known sensitive strain to the tailocin (p25.C2) was grown in a Tecan for 20 hours.
856 Treatments included p25.A12 WT purified tailocin, p25.A12 mutant purified tailocin, and a buffer control.
857 The mean OD_{600} of two technical replicates were measured as a function of time.
858
859

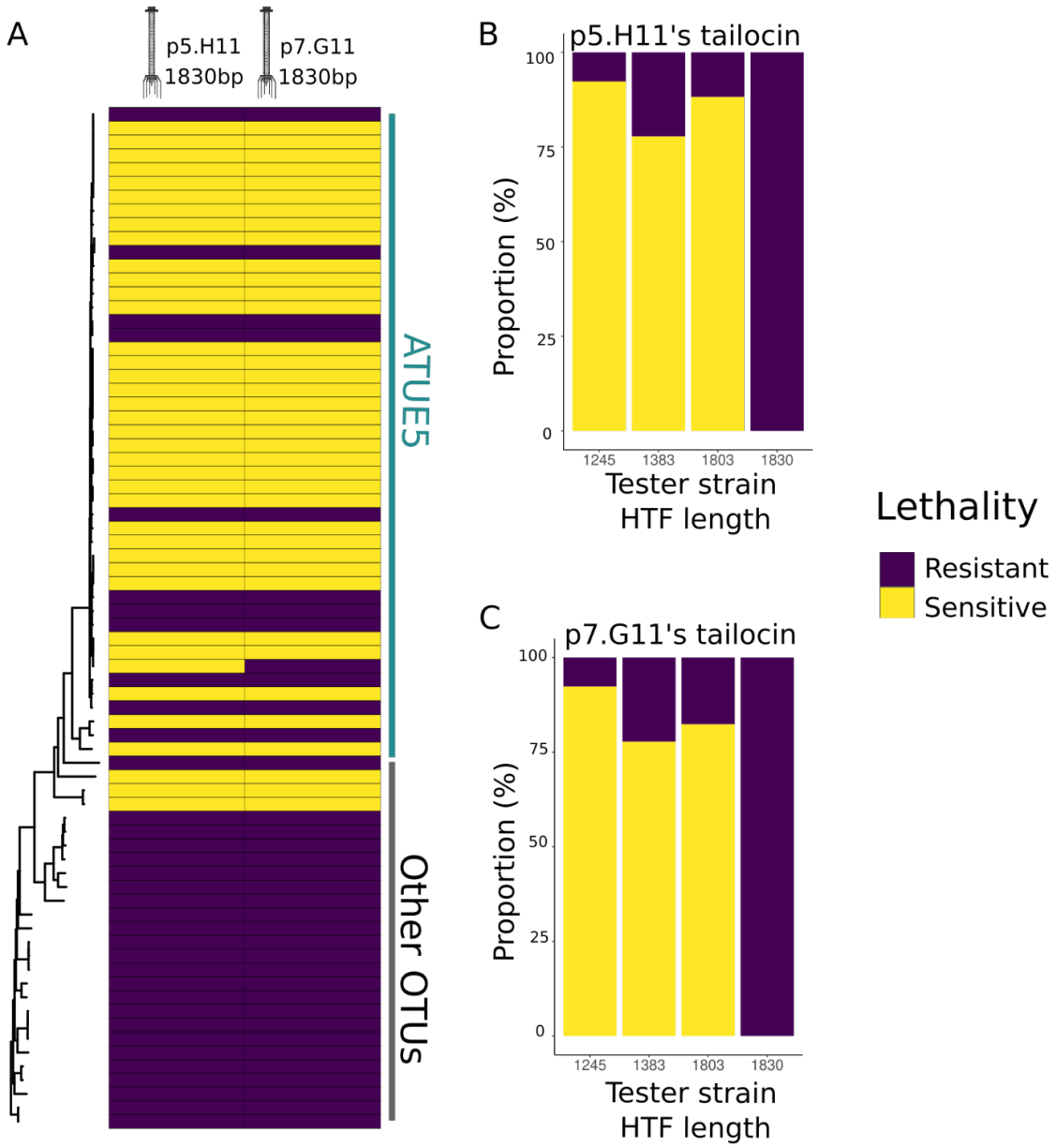


Figure S6. A. Tailocins are preferentially used for intra-lineage killing. Soft agar cultures of the *Pseudomonas* strains (rows) were challenged with viral particles extracted from cultures of two strains, p5.H11 and p7.G11 (from the 1830bp hypothetical tail fiber length haplotype, columns), in 3 technical replicates and 1 biological replicate. 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). **B-C**. The proportion of tester strains sensitive or resistant to p5.H11's tailocin (B) and p7.G11's tailocin (C). Lethality significantly correlated with the tester strain's hypothetical tail fiber length haplotype (Fisher's Exact test, $p = 10^{-4}$ (B), $p = 10^{-3}$ (C).

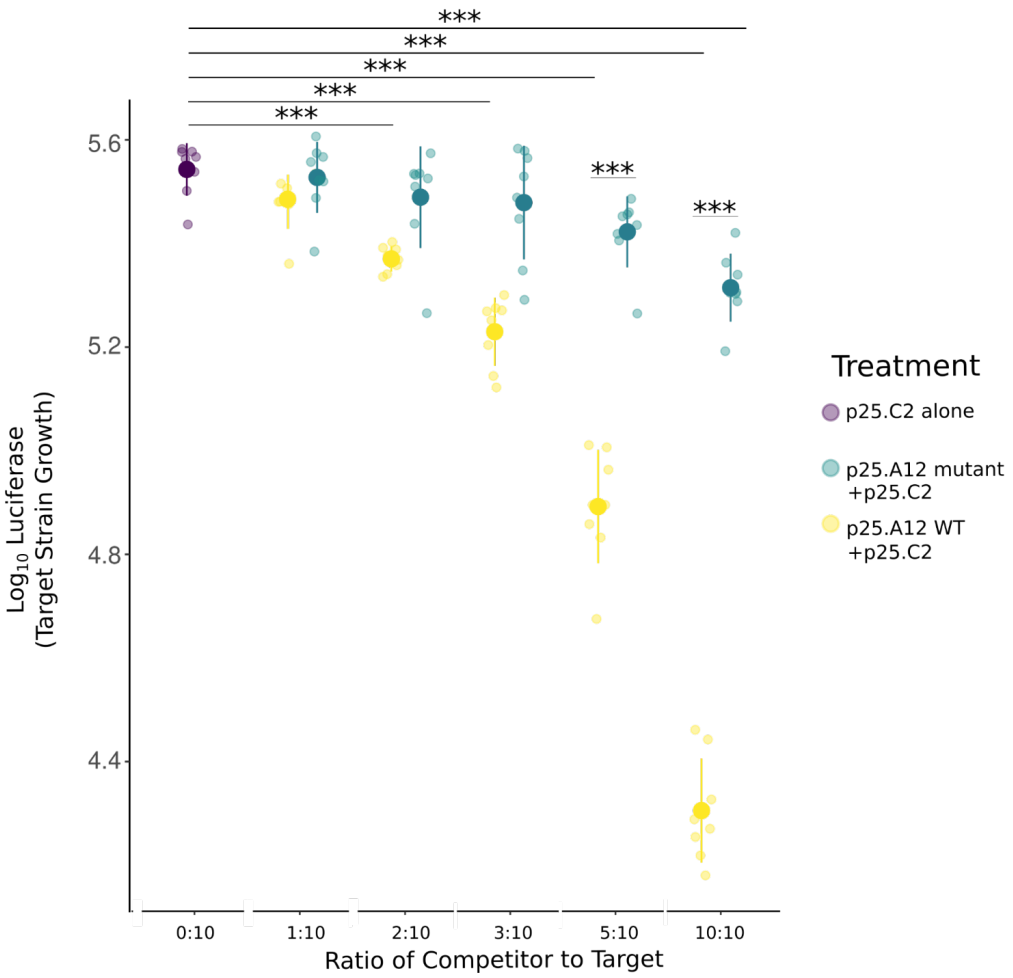


Figure S7. *In vivo* coinfections of p25.A12 (competitor) and p25.C2 (target, known to be sensitive to p25.A12). p25.A12 mutant competed with p25.C2 as well as p25.C2 grown alone were used as controls. Different ratios of competitor and constant amounts of target strain were used. p25.C2 grown alone as the control. ANOVA test, P-values are shown as $p < 0.001$ ***, $p < 0.01$ **, or $p < 0.05$ *.

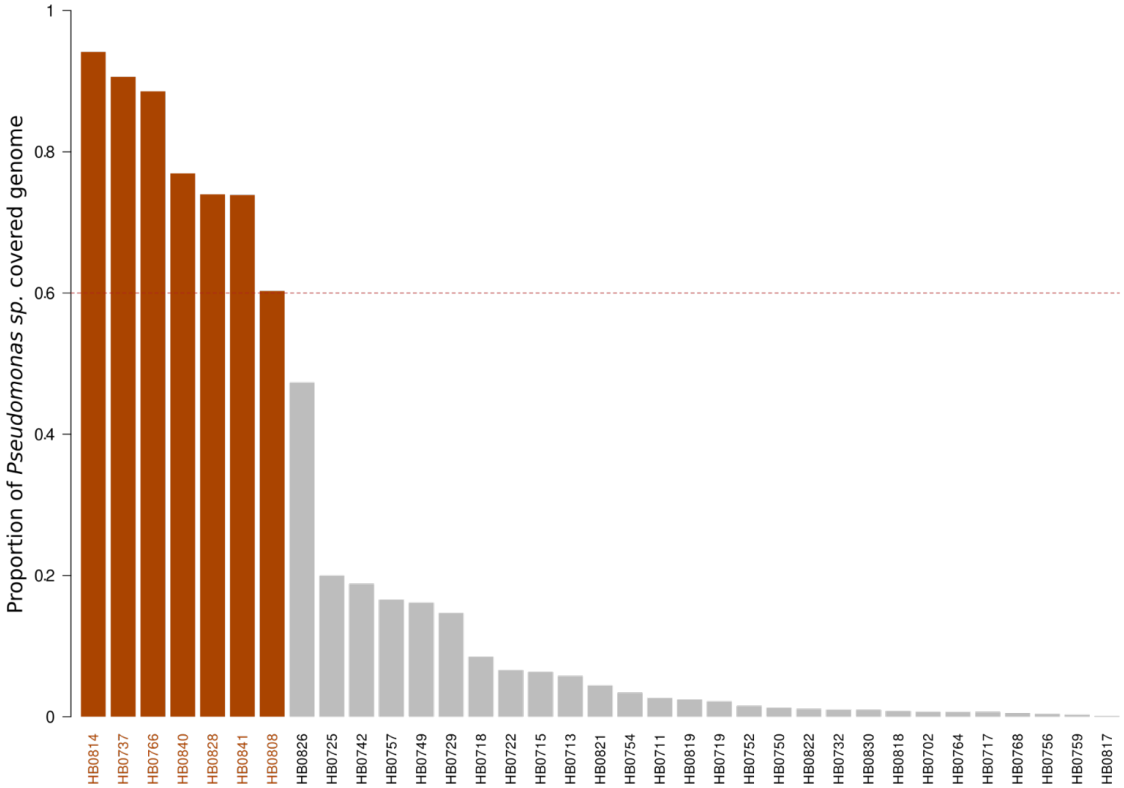


Figure S8. Seven historical samples cover at least 60% of the *Pseudomonas* genome. The height of the bars represent the proportion of the *Pseudomonas* sp. reference genome covered with at least one read after Mapping Quality filter value of 30. The horizontal dotted line depicts a threshold of 0.6.

878

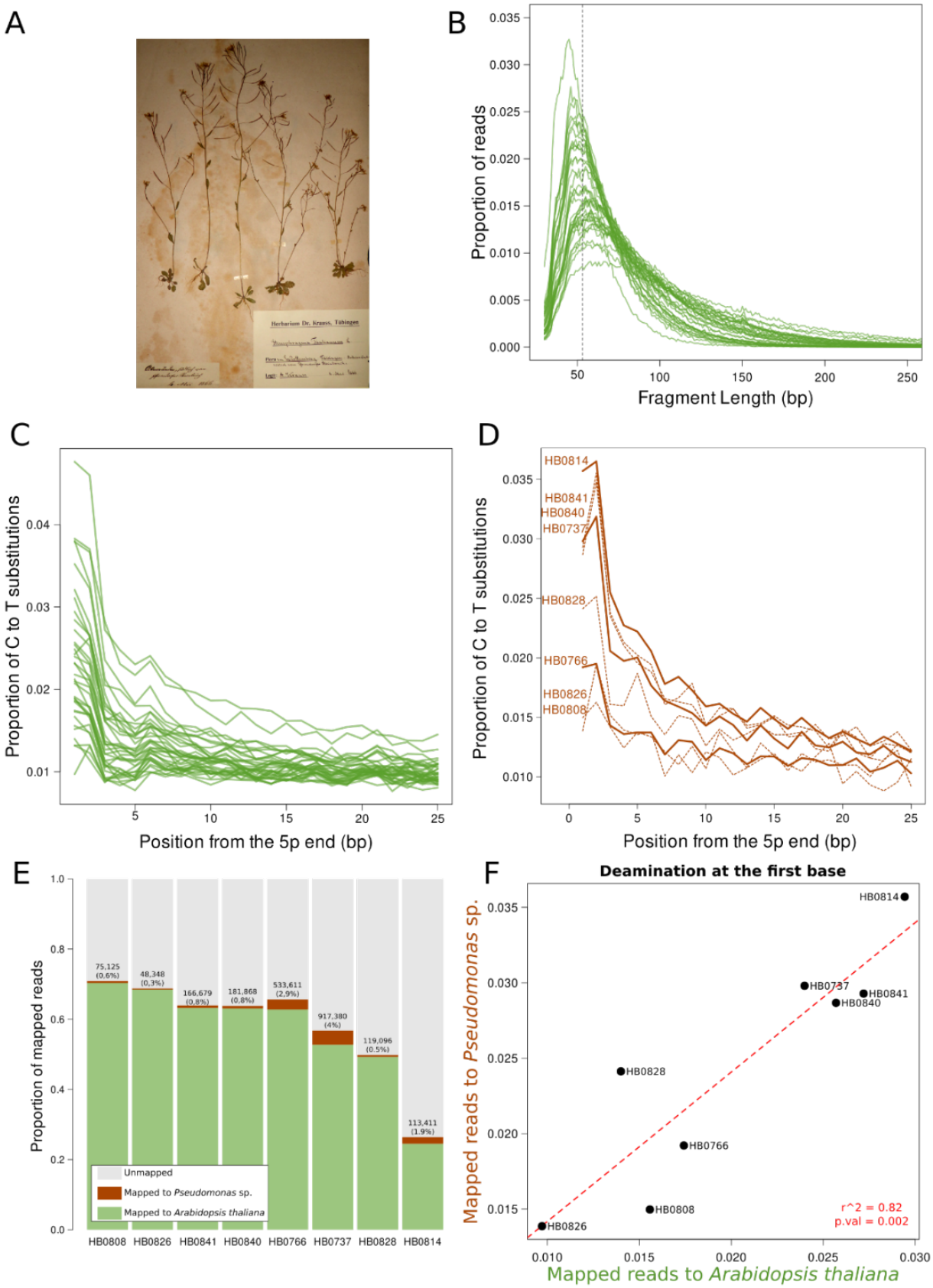

A
B
Proportion of reads
Fragment Length (bp)
C
Proportion of C to T substitutions
Position from the 5p end (bp)
D
HB0814
HB0841
HB0840
HB0737
HB0828
HB0766
HB0826
HB0808
Proportion of C to T substitutions
Position from the 5p end (bp)
E
Proportion of mapped reads
75,125 (0.6%)
48,348 (0.3%)
166,679 (0.8%)
181,868 (0.8%)
533,611 (2.9%)
917,380 (4%)
119,096 (0.5%)
113,411 (1.9%)
Unmapped
Mapped to Pseudomonas sp.
Mapped to Arabidopsis thaliana
HB0808 HB0826 HB0841 HB0840 HB0766 HB0737 HB0828 HB0814
F
Deamination at the first base
Mapped reads to Pseudomonas sp.
Mapped reads to Arabidopsis thaliana
r^2 = 0.82
p.val = 0.002

Figure S9. Herbarium-derived reads assigned to the host plant *Arabidopsis thaliana* and to *Pseudomonas* sp. show degradation patterns typical of ancient DNA. A. The photo shows of a representative *A. thaliana* herbarium specimen (sample HB0729) sampled in the Tübingen herbarium. **B.** Length distribution of merged sequenced reads of 35 historical herbarium samples after mapping to the *A. thaliana* reference genome. **C.** Cytosine-to-Thymine (C-to-T) substitutions at the 5' end of reads mapped to the *A. thaliana* reference genome. **D.** C-to-T substitutions at the 5' end of reads mapped to the *Pseudomonas viridiflava* reference genome. Only samples that covered more than 60% of the *P. viridiflava* genome were included in the analysis. The thick lines indicate samples included in the phylogenetic analyses. **E.** The bars indicate the proportion of reads assigned to *A. thaliana* or to *Pseudomonas* sp. **F.** The scatter plot displays the relationship between the proportion of C-to-T substitutions at the first base of the 5' end between reads assigned to *A. thaliana* or to *Pseudomonas* sp. The red dotted line represents a linear model.

881

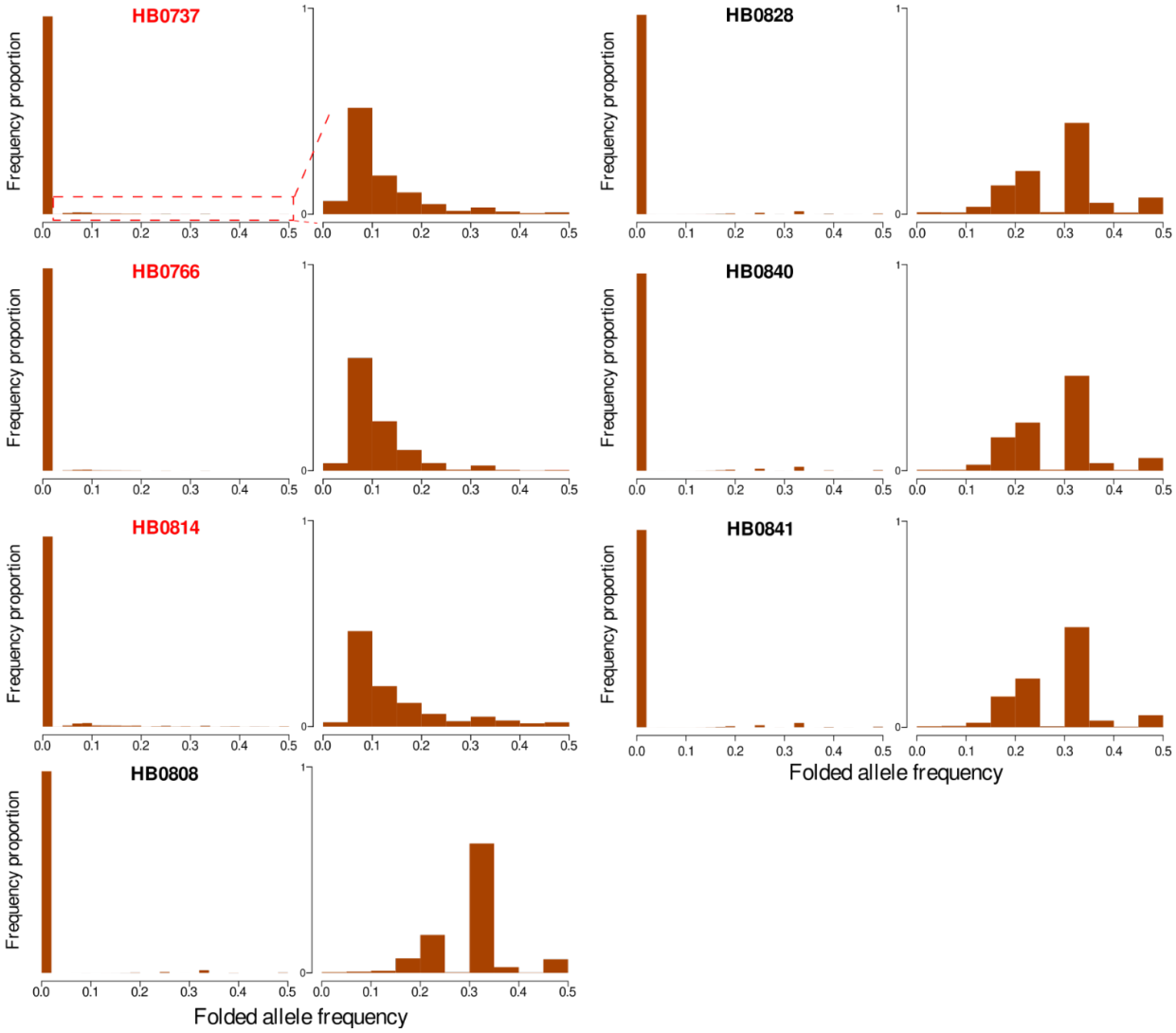


Figure S10. Historical *Arabidopsis thaliana* herbarium specimens are dominated by a single *Pseudomonas* sp. strain. The plots show the folded homozygosity support for each position along the *Pseudomonas viridiflava* genome with a coverage greater than 1x. A value of zero indicates that all reads support the same nucleotide (total homozygosity), where a value of 0.5 indicates maximum heterozygosity with only half of the reads supporting a particular nucleotide. Each row shows the folded homozygosity support for each sample. The left column shows all bins, whereas the bar indicating total homozygosity was removed in the right column.

882

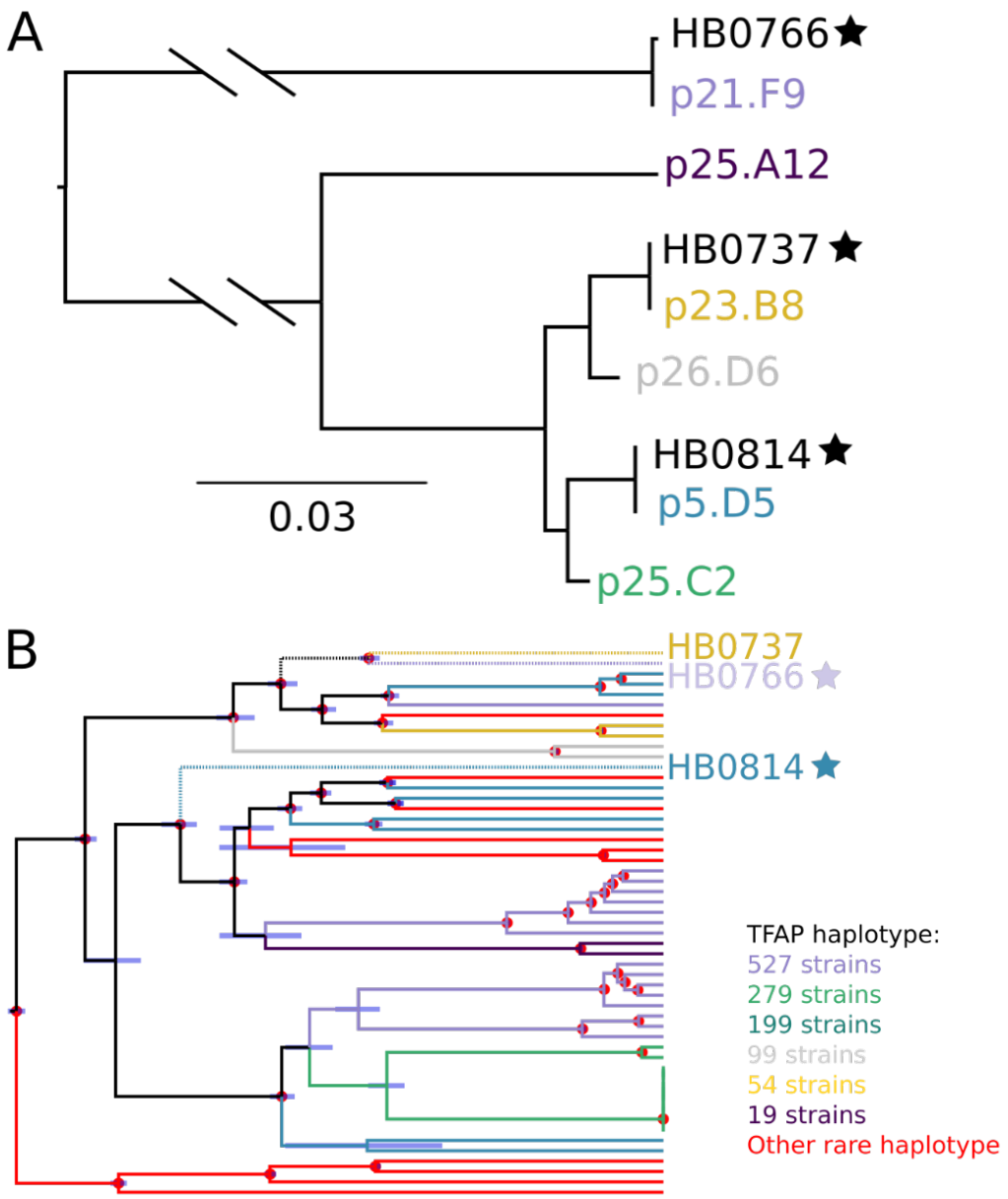


Figure S11. A. Neighbor-Joining tree of the tail fiber assembly gene translated sequence. Historical samples (HB) are placed in the context of the most common haplotypes. **B.** Bayesian tip-date calibrated phylogeny representing the evolutionary relationship between historical and modern *Pseudomonas* 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.

891 **Table S1. Untargeted data dependent LC/MS/MS Proteomics data for p25.A12 WT and**
892 **p25.A12ΔTFA tailocin partially purified extracts.**
893

Sample	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00762	Phage tail sheath protein	53.2	498	77	27	27	109		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00762	MAISFNNIPSDVRVPLFYAEMDNSAANSASAGMR	1xMet-loss [N-Term]	1	1	3515.66821	1172.56287	3	1.66	4.67
	JNHLCIDM_00762	RLIVAQVNDDVTGPEIGSLVLVPSVALAK		2	1	2973.6983	991.90564	3	1.37	4.62
	JNHLCIDM_00762	NDQHITIQAVEMAAPQPVWLQAAALAAR		3	0	3013.56763	1005.19531	3	1.24	4.39
	JNHLCIDM_00762	NAYGQADNSYLDSETMHQSAFIIR		12	0	2731.24167	911.08704	3	1.79	3.99
	JNHLCIDM_00762	ADPTGEVWCLPLLNTEGAK	1xCarbamidomethyl [C9]	7	0	2071.01647	1036.01257	2	0.68	3.74
	JNHLCIDM_00762	VQATVVNGATAAQAANALSVK		5	0	1984.08218	992.54694	2	2.23	3.45
	JNHLCIDM_00762	YGIATAYYEGGYVR		7	0	1582.75363	791.88129	2	1.05	3.38
	JNHLCIDM_00762	LIVAQVNDDVTGPEIGSLVLVPSVALAK		3	0	2817.59719	939.87231	3	1.85	3.12
	JNHLCIDM_00762	NLGGQGSMLAAMYETWR		3	0	1884.87311	942.94202	2	1.93	3.04
	JNHLCIDM_00762	SVVDAGVLTLSCK	1xCarbamidomethyl [C12]	4	0	1348.71408	674.86176	2	1.6	3.04
	JNHLCIDM_00762	LEEEGHVENAEVFAQHLIVER		5	0	2448.21538	816.74542	3	2.59	2.91
	JNHLCIDM_00762	TAVFISADASRPTQSGTMPGLDPAPASQR		6	0	2929.44724	977.15558	3	1.69	2.81
	JNHLCIDM_00762	FGAGQPIITPSTIR		4	0	1457.81109	729.40985	2	0.92	2.73
	JNHLCIDM_00762	RGTVGTLVAAGQLR		9	1	1398.81757	699.91278	2	0.51	2.46
	JNHLCIDM_00762	FTLTERESLLR		2	1	1364.75324	682.88141	2	1.69	2.33
	JNHLCIDM_00762	GELIAQYAR		1	0	1020.54727	510.77844	2	2.29	2.27
	JNHLCIDM_00762	INATPDLPVR		2	0	1095.61568	548.31262	2	2.08	2.06
	JNHLCIDM_00762	GTVGTLVAAGQLR		2	0	1242.71646	621.86285	2	1.59	1.98
	JNHLCIDM_00762	RLQGVITSK		2	1	1001.6102	501.30865	2	-0.17	1.82

	JNHLCIDM_00762	VFALLNQFR		10	0	1107.63094	554.31921	2	0.19	1.8
	JNHLCIDM_00762	MAISFNNIPSDVR	1xMet-loss [N-Term]	3	0	1332.69064	666.84991	2	1.44	1.79
	JNHLCIDM_00762	FGAGQPIITPSTIRGELIAQYAR		3	1	2459.34052	820.45215	3	0.56	1.34
	JNHLCIDM_00762	VPLFYAEMDNSAANSASAGMR		2	0	2201.99541	734.67078	3	1.07	1.27
	JNHLCIDM_00762	NDQHITIQAVEMAAPQPVWLQAAALAAR	1xOxidation [M12]	1	0	3029.56255	1010.52716	3	1.45	1.24
	JNHLCIDM_00762	QLYGHVYSAK		3	0	1165.60003	583.30603	2	4.08	1.23
	JNHLCIDM_00762	VNVMFPPDYINGLR		4	0	1634.83592	817.92236	2	0.93	1.14
	JNHLCIDM_00762	AGATVTVAGAATETGLLNLYVGGVR		2	0	2361.27725	787.76465	3	0.91	1.11
	JNHLCIDM_00762	DGNDPSRVNVMFPPDYINGLR		1	1	2376.1401	792.72034	3	2.67	0.86
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00761	Phage tail tube protein	12.5	115	68	7	7	24		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00761	QLTISGGGEAPL		6	0	1987.01647	994.01355	2	1.69	4.37
	JNHLCIDM_00761	TYVLSGAYLVEEPTSK		5	0	1756.90036	878.95569	2	2.13	3.02
	JNHLCIDM_00761	GTIDLRFDGSQG		1	1	1967.87297	984.44226	2	2.17	2.89
	JNHLCIDM_00761	AVHTPDMPLK		2	0	1108.58194	554.79468	2	0.12	2.53
	JNHLCIDM_00761	VDGTQLTISGGGEAPLMNVK	1xOxidation [M17]	1	0	2003.01138	668.34283	3	1.28	2.33
	JNHLCIDM_00761	ADDGTIDLR		2	0	975.47416	488.24094	2	0.44	2.27
	JNHLCIDM_00761	RDTVVPGYFKEVDK		1	2	1652.86425	551.62689	3	1.14	1.7
	JNHLCIDM_00761	RDTVVPGYFK		1	1	1181.63133	394.54852	3	-0.27	1.48
	JNHLCIDM_00761	AVHTPDMPLK	1xOxidation [M7]	5	0	1124.57686	375.53104	3	1.51	1.4
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00046	hypothetical protein JNHLCIDM_00046	48.5	460	52	17	17	28		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00046	AAVDSTDLGNLWYSR		2	0	1667.80237	834.4057	2	1.05	2.76
	JNHLCIDM_00046	IAALGLSEVGSYAFAR		5	0	1624.86933	812.93933	2	1.26	2.68

	JNHLCIDM_00046	AQNWIYLAAA		1	0	2318.09487	773.37079	3	1.27	2.58
	JNHLCIDM_00046	QVYNLVSEQG		1	0	1893.93009	947.46997	2	1.36	2.45
	JNHLCIDM_00046	GQVYNLVSEQG		1	1	2050.03121	684.01672	3	2.15	2.26
	JNHLCIDM_00046	SVPGVGLVSGK		1	0	999.58332	500.29584	2	1.08	2.21
	JNHLCIDM_00046	VQQGGGAGQL		3	0	1953.05524	977.0329	2	1.68	2.18
	JNHLCIDM_00046	QLTNTTAPVLGT		2	0	2632.37294	878.13116	3	2.28	2.12
	JNHLCIDM_00046	ANRGTTLEAYGITNAYTK		1	1	1943.98213	648.66626	3	1.08	1.79
	JNHLCIDM_00046	AILGLAASDYKK		2	1	1249.71506	625.36182	2	1.03	1.78
	JNHLCIDM_00046	NFNPDDKANR		1	1	1190.55487	397.52283	3	-0.8	1.77
	JNHLCIDM_00046	GSNPTTLAGYG		3	0	2505.17799	1253.09766	2	4.01	1.26
	JNHLCIDM_00046	VGMANNQPDA		1	0	2116.95388	706.32239	3	-0.6	1.25
	JNHLCIDM_00046	GTTLEAYGITNAYTK		1	0	1602.80098	801.90369	2	-0.55	1.25
	JNHLCIDM_00046	ESDNGVYYLQSR		1	0	1430.65465	715.83258	2	2.26	1.22
	JNHLCIDM_00046	SGVAAGNYSR		1	0	981.47483	491.24109	2	0.07	1.17
	JNHLCIDM_00046	IGWSGSSLK		1	0	934.49926	467.75381	2	1.17	1.04
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	estA	estA	69.5	643	26	10	10	19		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	estA	LYYLTGGGND		4	0	2142.02506	1071.51807	2	1.77	3.45
	estA	DAADRLADSAQ		4	1	2774.37037	925.46375	3	2.27	3.23
	estA	LADSAQVLQQAGAR		1	0	1427.76012	714.38489	2	1.67	2.75
	estA	DGSGEVYNLN		2	0	2427.17866	1214.09607	2	2.56	2.17
	estA	FNEQLVSR		1	0	992.51597	496.76288	2	2.53	1.97
	estA	TTQSDAVTPIT		3	0	2833.44254	945.15436	3	2.11	1.93
	estA	TSPVNAALGLP		1	0	3368.6692	1123.56396	3	2.42	1.93
	estA	STALTFSDQTR		1	0	1226.60116	613.80511	2	1.46	1.36
	estA	LFFNDR		1	0	811.40971	406.20889	2	0.98	1.03
	estA	GYNLTIGTSYR		1	0	1244.62698	622.81763	2	0.8	0.88
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00765	hypothetical protein JNHLCIDM_00765	12.2	112	15	1	1	1		

	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00765	VESAATVASEAAHTASR		1	0	1657.814	553.27771	3	2.76	1.17
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00764	hypothetical protein JNHLCIDM_00764	21.3	196	7	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00764	RGQEASEQLHAIR		1	1	1494.77716	498.93112	3	1.1	1.13
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00048	Chitinase class I	20	181	5	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00048	YDTGQLALR		1	0	1036.54218	518.77545	2	1.39	1.97
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00049	Bacteriophage lysis protein	18.3	169	4	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00049	LATADLR		1	0	759.43593	380.22162	2	0.04	1.1
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00039	hypothetical protein JNHLCIDM_00039	22.3	208	4	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00039	LNEFAEER		1	0	1007.47925	504.24396	2	1.38	1.05
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00044	Baseplate J-like protein	36.9	346	3	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00044	TQVDLAGDALR		1	0	1158.61133	579.80981	2	0.88	1.49
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	trpE	trpE	54.5	493	3	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	trpE	AILPAGTLSGAPK		1	0	1195.7045	598.35718	2	2.16	1.4

	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12 WT	JNHLCIDM_00040	Mu-like prophage FluMu DNA circularization protein	51.9	499	2	1	1	1		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00040	AVMATVDAVR		1	0	1032.55064	516.77795	2	-1.95	0.64
	Accession	Description	MW [kDa]	# AAs	Coverage [%]	# Peptides	# Unique Peptides	# PSMs		
p25.A12ΔTFA	JNHLCIDM_00762	Phage tail sheath protein	53.2	498	56	22	22	81		
	Master Protein Accessions	Sequence	Modifications	# PSMs	# Missed Cleavages	Theo. MH+ [Da]	Observed m/z [Da]	Charge	λM [ppm]	XCorr
	JNHLCIDM_00762	ADPTGEVWCLPLLNTEGAK	1xCarbamidomethyl [C9]	2	0	2071.01647	1036.01428	2	2.33	4.17
	JNHLCIDM_00762	SVVDAGVLTLSCK	1xCarbamidomethyl [C12]	5	0	1348.71408	674.86176	2	1.6	2.57
	JNHLCIDM_00762	FGAGQPIITPSTIR		5	0	1457.81109	729.41052	2	1.84	2.4

894
895
896 **Table S2: Relative mole percentage of monosaccharide residues detected by GC/MS.**
897

	Mole %			
Residue/Strain	**p21.F1**	**p21.F9**	**p25.A12**	**p25.C2**
Rha	**63.4**	**11.1**	**9.6**	**72.3**
Man	**0.1**	**0.2**	**0.1**	**0.1**
Gal	**0.0**	**0.0**	**0.1**	**0.0**
Glc	**21.7**	**56.5**	**75.4**	**14.5**
Hep	**0.2**	**Tr.**	**0.0**	**0.0**
Kdo	**3.0**	**11.4**	**1.2**	**0.3**
FucNAc	**5.8**	**0.0**	**0.0**	**5.1**
QuiNAc	**1.1**	**0.0**	**0.6**	**0.8**
ManNAc	**0.3**	**1.9**	**0.9**	**0.6**

GalNAc	**0.3**	**1.1**	**0.7**	**0.3**
GlcNAc	**4.2**	**17.6**	**11.6**	**6.0**
% CHO	**60.6%**	**22.5%**	**31%**	**52%**

Tr.- trace

Table S3: Historic and modern sample metadata.

* Type: (H)istoric ; (M)odern

ACCESSION_ID	ENA_ID	TYPE*	COLLECTION_YEAR	COUNTRY	LOCATION	LAT	LON	Pseudomonas_Genome_Coverage_Prop.	Pseudomonas_Genome_Average_Depth_MQ30	Used_for_authentication_analyses	_downstream_a
HB0702	SRR21814520	H	1905	DEU	Mellendorf_near_Buttlar	52.5470675	9.7299326	0.00701769	0.00629547	NO	NO
HB0711	SRR21814519	H	1840	DEU	NA	NA	NA	0.0266046	0.0199101	NO	NO
HB0713	SRR21814508	H	1826	DEU	Esslingen	48.7427584	9.3071685	0.057587	0.0515586	NO	NO
HB0715	SRR21814497	H	1828	DEU	Esslingen	48.7427584	9.3071685	0.0634669	0.0581094	NO	NO
HB0717	SRR21814491	H	1826	DEU	Alpirsbach	48.3456021	8.4033079	0.00664118	0.00581024	NO	NO
HB0718	SRR21814490	H	1890	DEU	Tuebingen	48.5236164	9.0535531	0.0850372	0.0709387	NO	NO
HB0719	SRR21814489	H	1890	DEU	Tuebingen	48.5236164	9.0535531	0.0216538	0.0139758	NO	NO
HB0722	SRR21814488	H	1891	DEU	Altburg	48.6619431	8.6926366	0.0659517	0.0522472	NO	NO
HB0725	SRR21814487	H	1888	DEU	Tuebingen	48.5236164	9.0535531	0.199974	0.197268	NO	NO
HB0729	SRR21814486	H	1866	DEU	Tuebingen	48.5236164	9.0535531	0.146878	0.136151	NO	NO
HB0732	SRR21814518	H	1932	DEU	Rastatt	49.0532906	8.5245089	0.0100316	0.0070049	NO	NO
HB0737	SRR21814517	H	1937	DEU	Tuebingen	48.5236164	9.0535531	0.906105	8.75871	YES	YES
HB0742	SRR21814516	H	1953	DEU	Lichtenstein	48.41996913	9.26630222	0.187988	0.162055	NO	NO
HB0749	SRR21814515	H	1849	DEU	near_Hohenheim	48.7118	9.2113	0.161478	0.156122	NO	NO
HB0750	SRR21814514	H	1957	DEU	Boeblingen	48.684969	9.0113444	0.012546	0.010024	NO	NO
HB0752	SRR21814513	H	1952	DEU	Eberspiel_in_the_district_of_Calw	48.7415506	8.6842699	0.0150715	0.0118002	NO	NO
HB0754	SRR21814512	H	1951	DEU	District_of_Calw	48.7153344	8.7381796	0.033677	0.028471	NO	NO
HB0756	SRR21814511	H	1951	DEU	Wendlingen	50.8021737	7.7130698	0.00412007	0.00351381	NO	NO
HB0757	SRR21814510	H	1951	DEU	Wendlingen	50.8021737	7.7130698	0.165843	0.100545	NO	NO
HB0759	SRR21814509	H	1893	DEU	Wohlmuthausen	50.5727241	10.21353915	0.00280539	0.00201518	NO	NO
HB0764	SRR21814507	H	1905	DEU	Mengen	48.049796	9.3316096	0.00683053	0.00540983	NO	NO

HB0766	SRR21814506	H	1846	DEU	Monastery_forest_Hohenzollern	47.9724751	10.3005044	0.885642	7.63767	YES	YES
HB0768	SRR21814505	H	1894	DEU	Todtnau	47.8303378	7.9452378	0.00528544	0.00427861	NO	NO
HB0808	SRR21814504	H	1817	DEU	Stuttgart_Moehringen	48.7784485	9.1800132	0.60285	0.974844	YES	NO
HB0814	SRR21814503	H	1882	DEU	Heidelberg	49.4093582	8.694724	0.941364	8.13508	YES	YES
HB0817	SRR21814502	H	1934	DEU	Gundelwangen_near_Bonndorf	47.8426163	8.2864051	0.00113424	0.00213973	NO	NO
HB0818	SRR21814501	H	1928	DEU	Langenau_near_Ulm	48.7276553	8.0150864	0.00812181	0.00744537	NO	NO
HB0819	SRR21814500	H	1928	DEU	Langenau_near_Ulm	48.7276553	8.0150864	0.0241737	0.0120816	NO	NO
HB0821	SRR21814499	H	1874	DEU	Winnenden	48.8754571	9.3978478	0.0443252	0.0396616	NO	NO
HB0822	SRR21814498	H	1936	DEU	Hohentwiel	47.7617515	8.8348709	0.0109147	0.0104388	NO	NO
HB0826	SRR21814496	H	1948	DEU	Abtsgmuend	48.8936471	10.0027636	0.473287	0.665255	NO	NO
HB0828	SRR21814495	H	1890	DEU	Schramberg	48.225478	8.3852168	0.739741	1.6142	YES	NO
HB0830	SRR21814494	H	1862	DEU	near_Beimerstetten	48.4840643	9.9836241	0.00970948	0.00894266	NO	NO
HB0840	SRR21814493	H	1851	DEU	Sulz_am_Neckar	48.3617509	8.6314329	0.769394	1.61748	YES	NO
HB0841	SRR21814492	H	1875	DEU	Freudenstadt_in_the_Black_Forest	48.4637727	8.4111727	0.738722	1.44545	YES	NO
p12.A11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.955211	43.2805	NA	YES
p12.A9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.385298	2.68517	NA	NO
p12.E2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.946364	124.361	NA	YES
p12.F2	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.941151	61.1632	NA	YES
p12.G7	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.942483	26.1727	NA	YES
p12.H7	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.703408	7.44366	NA	NO
p13.C1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.919667	91.4711	NA	YES
p13.C7	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.915668	30.2041	NA	NO
p13.D10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.917782	83.4618	NA	YES
p13.D5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.373545	3.2645	NA	NO
p13.F1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.548638	9.89508	NA	NO
p13.F3	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.930129	180.182	NA	YES
p13.F5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.486993	7.68403	NA	NO
p20.B10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.362051	6.0333	NA	NO
p20.D4	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.965996	26.275	NA	YES
p20.F10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.951949	27.3486	NA	YES

p20.G9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.951606	25.8164	NA	YES
p21.A8	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.951989	50.4423	NA	YES
p21.E3	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.912764	42.5976	NA	YES
p21.F1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.921125	54.2176	NA	YES
p21.F9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.951456	43.9553	NA	YES
p22.A8	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.910773	42.5346	NA	YES
p22.B5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.908795	31.8141	NA	YES
p22.C1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.910697	44.4703	NA	YES
p22.D1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.93509	9.86993	NA	YES
p22.D4	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.901368	6.48617	NA	YES
p23.A3	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.922147	90.3312	NA	YES
p23.A5	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.355851	5.69028	NA	NO
p23.B4	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.353754	10.3529	NA	NO
p23.B8	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.915323	25.8754	NA	YES
p24.B5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.906998	24.3191	NA	YES
p24.H2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.919194	36.961	NA	YES
p25.A12	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.904389	8.27369	NA	YES
p25.B2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.99527	9.73957	NA	YES
p25.C11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.936486	7.67613	NA	NO
p25.C2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.999957	38.9734	NA	YES
p25.D2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.997331	11.0366	NA	YES
p26.B7	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.918811	18.5831	NA	YES
p26.C10	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.532614	7.27294	NA	NO
p26.D6	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.907977	10.1768	NA	YES
p26.E7	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.915118	26.4647	NA	YES
p26.F6	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.263831	1.09767	NA	NO
p27.C5	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.33096	4.98267	NA	NO
p27.D6	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.915116	62.9111	NA	YES
p27.F2	NA	M	2018	DEU	Pfrondorf2	48.561087	9.109294	0.908864	30.4347	NA	YES
p3.A3	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.999479	12.1995	NA	YES
p3.F12	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.999766	64.6266	NA	YES

p3.F8	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.910145	14.5013	NA	NO
p3.G11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.405107	35.7165	NA	NO
p3.G9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.936757	40.7657	NA	YES
p4.A6	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.999359	11.257	NA	YES
p4.A8	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.317566	8.57523	NA	NO
p4.C5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.281958	2.19208	NA	NO
p4.D11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.329928	7.72875	NA	NO
p4.D2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.903089	43.5342	NA	YES
p4.E5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.905147	9.07888	NA	YES
p4.E6	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.907851	26.4274	NA	YES
p4.H3	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.359271	11.1964	NA	NO
p5.A5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.569164	20.9644	NA	NO
p5.C1	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.940631	83.5842	NA	YES
p5.C3	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.911686	61.8474	NA	YES
p5.D5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.908126	47.1102	NA	YES
p5.F2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.355109	10.8067	NA	NO
p5.H11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.912223	24.6014	NA	YES
p6.A10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.967914	86.1443	NA	YES
p6.B5	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.377432	8.17665	NA	NO
p6.B9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.915641	40.9034	NA	NO
p6.D10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.362251	6.29957	NA	NO
p6.E9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.348102	5.45088	NA	NO
p6.F1	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.441901	14.1066	NA	NO
p6.G2	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.591204	22.8581	NA	NO
p6.G3	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.374359	12.0216	NA	NO
p7.F2	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.420044	6.6193	NA	NO
p7.G11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.922476	63.4571	NA	YES
p8.B3	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.999766	72.1507	NA	YES
p8.B9	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.921803	46.3749	NA	YES
p8.C7	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.933317	51.4209	NA	YES
p8.D11	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.432979	9.92024	NA	NO

p8.E4	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.917628	39.6076	NA	YES
p8.G10	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.460138	12.253	NA	NO
p8.G2	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.450816	8.26752	NA	NO
p8.H7	NA	M	2018	DEU	Det-2	48.556255	9.135424	0.899835	67.821	NA	YES
p9.C4	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.946827	40.897	NA	NO
p9.H10	NA	M	2018	DEU	Eyach	48.446111	8.781611	0.469617	11.8808	NA	NO
BAV2572	NA	M	2013	USA	Virgina	NA	NA	NA	NA	NA	NO
CDRTc14	NA	M	2013	AUT	Steinberg-Dörfl	47.4N	16.47E	NA	NA	NA	NO
ICMP8820	NA	M	2010	NZL	NA	NA	NA	NA	NA	NA	NO
ICMP3272	NA	M	1975	NZL	Riverhead	NA	NA	NA	NA	NA	NO
CFBP1590	NA	M	1974	FRA	Vaucluse	NA	NA	NA	NA	NA	NO
DSM6694	NA	M	1930	CHE	NA	NA	NA	NA	NA	NA	NO
CH409132	NA	M	2013	USA	Boston, Massachusetts	42.22N	71.7W	NA	NA	NA	NO
LMCA8460	NA	M	2008	USA	Lake Michigan College	42.090114N	86.393408W	NA	NA	NA	NO
SV1779	NA	M	2017	USA	Scott Valley, California	NA	NA	NA	NA	NA	NO
T157	NA	M	2019	USA	Tulelake, California	NA	NA	NA	NA	NA	NO

Table S4: Key Resources Table

REAGENT or RESOURCE	**SOURCE**	**IDENTIFIER**
Deposited Data		
1524 *Pseudomonas* genomes	Karasov et al. 2018	ENA: PRJEB24450
Experimental Models: Organisms/Strains		
1524 Modern *Pseudomonas* strains	Karasov et al. 2018	NA
34 Historic strains		
10 NCBI genomes	NCBI	In Table S3 (Sample)
p25.A12 ΔTFA p25.A12ΔTFA +TFA	This study	NA
Software and Algorithms		
VIBRANT v1.2.1	Kieft et al., 2020	https://github.com/AnantharamanLab/VIBRANT
Mash v2.2.2	Ondov et al., 2016	https://github.com/marbl/Ma

		sh
panX	Ding et al., 2018	https://github.com/neherlab/pan-genome-analysis
pankmer	Aylward et al., 2023	https://pypi.org/project/pankmer/

906
907
908
909
910 **Table S5. De-novo assembly contigs statistics from reads mapped to the Tailocin region.**
911

MAETRIC\SAMPLE	HB0737	HB0766	HB0814
Total_n	10	9	14
Total_seq_bp	23769	21573	21574
Avg._seq_bp	2376.9	2397	1541
Median_seq_bp	2143	1665	1275.5
N_50_bp	2858	3405	1705
Min_seq_bp	583	494	415
Max_seq_bp	5543	7935	3337

912
913