The evolutionary tale of tailocins in *Pectobacterium* show genus-wide homologous recombination and inversions of tail fiber genes

Lakhansing A. Pardeshi[1,2,*], Anne Kupczok[1], Dick de Ridder[1], Sandra Smit[1], and Theo A. J. van der Lee[2,*]

[1]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands

[2]Biointeractions and Plant Health, Wageningen Plant Research, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands

[*]Corresponding authors: lakhansing.pardeshi@wur.nl, theo.vanderlee@wur.nl

Abstract

Pectobacterium species comprise globally present phytopathogens known to co-inhabit infected plants. During such co-infection, these pathogens can have cooperative and/or antagonistic interactions to prevent intrusion, dominate a niche, displace established strains, or release nutrients. The phage tail-like bacteriocins, also called tailocins, play an important role in shaping such bacterial communities because of their ability to kill closely related strains of the same species. Adopting a comparative pangenomics approach, we performed a phylogenetically structured study on the diversity and evolution of the carotovoricin, the only tailocin present in the *Pectobacterium* species. We discovered that the carotovoricin biosynthesis gene cluster was domesticated by the ancestor of the *Pectobacterium* genus, and its overall retention and purifying selection indicate its significance in bacterial community interactions. We identified differential conservation within the carotovoricin locus with exceptional variability and phylogenetic discord in the genes encoding the tail fibers, known to play a major role in host range specificity. We identified an evolutionary mechanism involving recombination-mediated exchange of these tail fiber genes across the *Pectobacterium* genus, which complements the existing DNA sequence inversion mechanism to achieve tailocin polymorphism. In particular, the genus-wide exchange of tail fiber locus discovered here is a novel strategy adopted by the plant pathogen.

Keywords: Tailocins, Horizontal gene transfer, *Pectobacterium*, Pangenomics

Introduction

Over the past decades, the genomes of tens of thousands of bacteria have become available, which offers unprecedented insight into bacterial genome evolution. Comparative genomics has shown that bacteria evolve by point mutations, structural variations such as deletions, and horizontal gene transfer (HGT) (*1*, *2*). Transduction mediated by bacteriophages is one of the mechanisms for HGT in bacteria. In particular, bacterial chromosomes can harbor integrated genomes of temperate bacteriophages, called prophages (*3–6*). Apart from these complete prophages, bacterial genomes frequently contain defective prophages and various molecular systems of presumed prophage origin, namely gene transfer agents (*7*), type VI secretion systems and phage tail-like bacteriocins (*8–11*). Phage tail-like bacteriocins, also known as tailocins, resemble "headless phages" wherein the tail and tail fibers form a functional unit. Tailocins are widespread

throughout the gram-negative bacteria, such as *Pseudomonas* spp. (*12*), *Burkholderia cenocepacia* (*13*), *Allorhizobium vitis* (*14*), *Dickeya dadantii* (*15*) and *Pectobacterium carotovorum* (*16*). A contractile phage tail-like bacteriocin was identified in *P. carotovorum* and named carotovoricin (*16–18*). The genetic locus of carotovoricin consists of 18 genes arranged in tandem, flanked by *tolC_2* and *ybiB* genes. Similar to other tailocin-producing bacteria (*15*, *19*, *20*), carotovoricin's origin was traced back to a common ancestor of the *Salmonella typhi* phage (*21*).

Tailocins are divided into two classes, based on having a rigid (R-type) or flexible (F-type) core structure. The F-type tailocins are noncontractile and flexuous. In contrast, R-type tailocins, such as carotovoricin, have a rigid contractile tail, formed of a hollow tube encapsulated by a contractile sheath, attached to a baseplate along with tail fibers (*22*, *23*). These fibers recognize and bind the target bacterial cell surface receptors, followed by contraction of the sheath and extension of the core tube. This creates an ungated channel in the cell membrane, causing cell death by membrane depolarization (*24*).

Bacteria that produce tailocins are generally resistant to their own tailocins (*25*). As the tail fibers enable the recognition of specific target cell membrane receptors, tailocins offer an effective and targeted mechanism of killing co-occurring bacteria (*26*). Normally, the target range is very narrow and limited to strains of the same species or phylogenetically close relatives (*15*). However, increasing the variability of tail fiber genes may extend the recognition spectrum. For example, in *Pseudomonas* species, the gene clusters responsible for tailocin synthesis show highest sequence divergence in the genes encoding these tail fiber genes compared to the rest of tailocin encoding region (*12*). There are various mechanisms to increase the target range, such as recombination of tail fiber genes to produce new sequence combinations (*27*), and the presence of multiple tail fiber genes within tailocin loci (*28*). In *P. carotovorum* yet another mechanism is found, where a DNA-invertase gene *ein* is located immediately downstream of the fibA/fibB tail fiber gene. This DNA-invertase can alter the C-terminal region of tail fiber genes, again leading to variations in carotovoricin host range specificity (*29*). These variations in the tail fibers yield a diverse pool of tailocins being maintained in a species (*30*).

The widespread presence of tailocins and maintenance of their tail fiber diversity underscore their role in non-self recognition and interbacterial interactions. The inter- and intra-species interactions can influence broader as well as local microbial communities. For example, a species-level replacement of *P. atrosepticum* and *P. parmentieri* by *P. brasiliense* was observed at a wide geographical scale in Europe (*31–33*). Strain-specific competitive interactions are also observed in the *Pectobacterium* species, where certain strains can inhibit other strains but none can inhibit the entire species (*34*). The mechanisms underlying such competitive replacement are not yet understood for the *Pectobacterium* species. These competitive interactions may be attributed to non-genetic factors such as climate change (*31*) and changing cultivation methods in agriculture (*35*), and to genetic factors, mainly encoded by accessory genes, such as toxin production, resistance genes and tailocins (*34*). Tailocins also play an important role in intra-genus competitive interactions of *Pectobacterium* species, and is produced at the basal level by most of the species in this genus (*36*).

Our previous pangenome study showed widespread presence of carotovoricin in the *Pectobacterium* genus

(*33*). Here, we elaborate our pangenomics approach to systematically study the extent of conservation of the carotovoricin biosynthesis gene cluster in the *Pectobacterium* genus. Furthermore, we study local variation within the carotovoricin locus to understand the polymorphic nature of tailocins. Subsequently, we evaluate the discordance between the evolutionary histories of the two loci within the carotovoricin cluster to help understand the underlying mechanism generating polymorphisms. Finally, we discuss the potential role of carotovoricin in the intra- and inter-species community dynamics observed in *Pectobacterium* species.

Results

Carotovoricin conservation in the *Pectobacterium* genus

We recently constructed a *Pectobacterium* genus pangenome with PanTools (*37*, *38*) and discovered the contribution of prophages to pangenome growth and the emergence of new lineages (*33*). This pangenome comprises 454 *Pectobacterium* genomes (g_1 to g_454 in Supplementary Table S1), where 1,977,865 protein-coding genes were grouped into 30,156 homology groups. To study the dynamics of 1,319 putative prophages identified in these genomes, we represented prophages as a sequence of homology groups and traced these homology group signatures across the pangenome. We identified orthologous prophages by quantifying syntenic conservation of homology group signatures between prophages and dereplicated the prophages to 436 clusters (see Methods, Supplementary Tables S2,S3). The largest of these clusters, containing 428 orthologous prophages, was identified as the phage tail-like bacteriocin carotovoricin, based on the functional annotation of homology groups. Each carotovoricin region contained genes encoding the core structural proteins necessary for tailocin production: holin, lytic transglycosylase, tail sheath protein, tail tube protein, tail assembly protein, tail tape measure protein, baseplate proteins, tail spike protein, and tail fiber proteins (Figure 1a).

We observed a single copy of the carotovoricin biosynthesis gene cluster in 428 *Pectobacterium* genomes, whenever it was identified (Supplementary Table S4). Of these, 365 were present on a single contig and the remaining were split onto two contigs due to fragmented genome assemblies. Because of the presence of such fragmented assemblies in the public repository, a 5kb region flanking the carotovoricin was used to determine the flanking gene structural conservation rather than a fixed number of neighbouring genes. This gene neighborhood flanking carotovoricin, comprising three to four genes, is syntenically conserved across the *Pectobacterium* genus, suggesting that a common ancestor acquired the carotovoricin cluster. Genes *tolC_2* and *ybiB* mark the upstream and downstream boundary of the carotovoricin biosynthesis gene clusters in all *Pectobacterium* species.

Complete carotovoricins encoding all core structural genes were found to be present in all but three *Pectobacterium* species (Figure 1b). Species with a complete carotovoricin contained 18 to 27 genes. Most of these genes were conserved, as evidenced by a single homology group per gene across all species. A partial carotovoricin cluster was observed in the *P. betavasculorum* and *P. atrosepticum* genomes, where only the first five and seven genes from the 5' end were found to be present, respectively. The genes in these partial carotovoricins are part of the first transcriptional unit out of the four identified in the carotovoricin

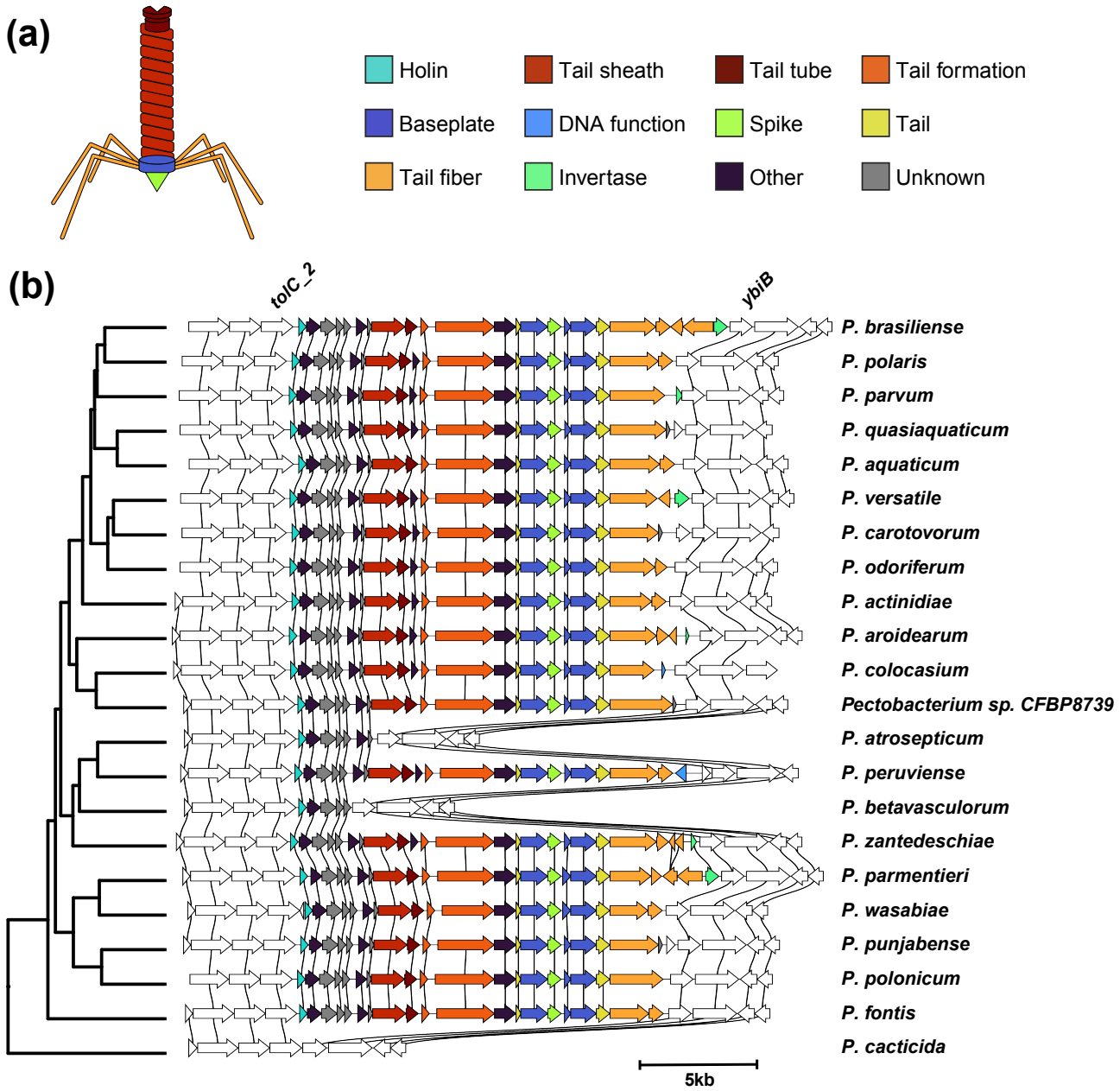


Figure 1: Carotovoricin cluster across the *Pectobacterium* genus. (a) An illustration of a carotovoricin particle where structural units are colored to match the carotovoricin biosynthesis cluster alignment shown in (b). (b) Carotovoricin biosynthesis gene clusters are ordered along the *Pectobacterium* genus phylogeny. Genes in 5kb regions flanking carotovoricin clusters are shown in white. Genes belonging to the same homology group are connected with vertical black lines.

locus (*21*). Transcriptomics data (*39*) suggests that the genes within this partial carotovoricin cluster are expressed and regulated in *P. atrosepticum*; this could not be verified for *P. betavasculorum*, due to a lack of publicly available gene expression data. Based on a phylogeny built on 1,949 core genes in the pangenome (Supplementary Figure S1), *P. betavasculorum* and *P. atrosepticum* share a common ancestor

with *P. peruviense* and *P. zantedeschiae* respectively, both with intact carotovoricins. Given the difference in the deleted carotovoricin loci and the phylogenetic distance to the closest carotovoricin-carrying species, it seems that carotovoricin was lost independently in *P. betavasculorum* and *P. atrosepticum*. In contrast, *P. cacticida* lacks carotovoricin and shows no remnants. A recent phylogenomic analysis and detailed phenotypic characterization proposed a reclassification of *P. cacticida* into a new genus, *Alcorniella comb. nov* (*40*). Therefore, the ancestor of the *Pectobacterium* genus likely domesticated and repurposed the carotovoricin cluster. In addition to these species-wide losses, there were seven strains from various *Pectobacterium* species that showed disruption of the carotovoricin cluster (Supplementary Results). In summary, carotovoricin is broadly conserved across the *Pectobacterium* species, suggesting its crucial role in community interactions.

Differential conservation within the Carotovoricin cluster

As carotovoricin was present in all *Pectobacterium* species except for two, we investigated if it has further evolved independently in each species and understand the inter-species variation in the carotovoricin cluster. To avoid analysis artifacts due to assembly quality, we selected carotovoricin loci from 365 genomes in which it was present on a single contig together with the flanking genes. Next, we optimized the sequence identity threshold at which genes are grouped together in homology groups. For example, a strict threshold can result in orthologous genes being distributed into multiple homology groups, and eventually in a higher number of homology groups in the pangenome. To study the carotovoricin evolutionary dynamics, an optimal sequence identity threshold for the whole pangenome was estimated by evaluating various settings based on BUSCO gene sets. The PanTools homology grouping setting of 4 (equivalent to >65% sequence identity) optimally captures the sequence diversity of BUSCO gene sets across the *Pectobacterium* genus. This setting also resulted in a single homology group for each of the carotovoricin core structural genes across the *Pectobacterium* genus. In line with this, we anticipated a small set of carotovoricin-related homology groups. However, we found a total of 130 homology groups mapped to carotovoricin regions across *Pectobacterium* genus, indicating significant gene-level diversity between different carotovoricins (Supplementary Table S5).

Closer inspection of the carotovoricin loci across genomes showed conserved regions in which the genes were represented by a single homology group, and variable regions in which genes were found in multiple homology groups (Figure 1b). First, we observed presence-absence variation of the nitric oxide deoxygenase gene (homology group hg_22427623 in Supplementary Table S5) in *Pectobacterium* genus (present in 88/365 the genomes) (Figure 2a). The gene encoding the tail tape measure protein, located between the tail assembly chaperon (hg_22427622) and P2 GpU family protein (hg_22427616), also showed variation across the *Pectobacterium* genus and was represented by 9 homology groups. Finally, the 3' region flanked by the tail protein (homology group hg_22427604) and the *ybiB* gene was represented by multiple homology group combinations across the pangenome annotation graph (Figure 2a). Interestingly, the invertase *ein* also showed presence-absence variation: it was found in 47% (172/365) of the carotovoricins. A similar degree of diversity in carotovoricin cluster was also observed within a species, as exemplified by *P. brasiliense* (Supplementary Figure S4) and *P. versatile* species (Supplementary Figure S5). In summary, the carotovoricin cluster shows patchy conservation of loci, both at the genus and species levels.

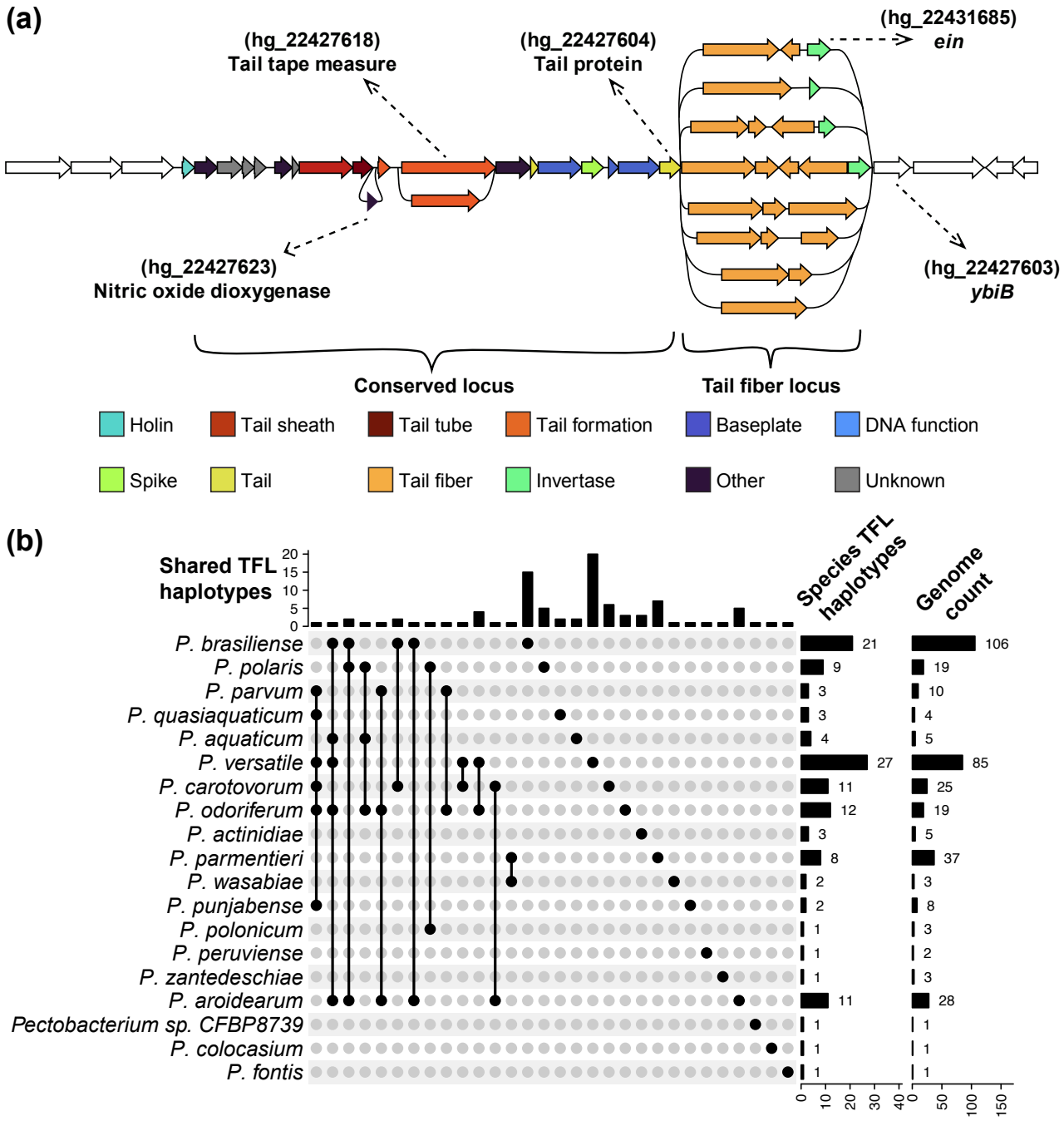


Figure 2: Carotovoricin locus variation. (a) An illustration of gene-level variation in carotovoricin cluster, with gene presence-absence variation shown as loops. Flanking genes are represented as white arrows. The conserved locus of carotovoricin is defined as genes encompassing holin and tail protein (hg_22427604). The variable tail fiber locus is defined as the genes between the tail protein (hg_22427604) and *ybiB*. The unique homology group combinations in the tail fiber loci (TFLs) across the *Pectobacterium* genus are used to define haplotypes. (b) The UpSet plot shows the number of *Pectobacterium* species that share a TFL haplotype.

For subsequent analysis, we refer to the carotovoricin region spanning from the holin (hg_22427640) to the tail protein (hg_22427604) as the conserved locus and the remaining region as the TFL. Since the TFL includes multiple tail fiber genes, inversion by Ein can result in multiple combinations of genes at this locus across strains. As our homology grouping adequately captured this sequence-level diversity, we further used the uniquely ordered combinations of homology groups to define TFL haplotypes. In total, there were 93 TFL haplotypes at the genus level, some of which were present in multiple *Pectobacterium* species (Supplementary Table S6). The species *P. brasiliense* and *P. versatile* showed the most haplotypes, 21 and 27 respectively, with only one shared haplotype (Figure 2b). Five species had 8 or more TFL haplotypes, and the remaining species had less than 5 haplotypes, which might be because of the lower number of genome assemblies for these species. Interestingly, intra-species average nucleotide identity (ANI), is not related to the number of TFL haplotypes. For example, *P. brasiliense* is more diverse than the *P. versatile* (based on broader intra-species ANI (Supplementary Figure S6)), however, we observed more TFL haplotypes in the ecologically and geographically broadly distributed *P. versatile* (*41*) compared to *P. brasiliense*. Together, these results suggest that the effect of spatio-temporal interactions experienced by *Pectobacterium* species likely determines the tailocin locus diversity of a species.

Tail fiber locus haplotypes are shared across the genus

Comparison of the TFLs revealed cases where two or more *Pectobacterium* species had the same TFLs (Figure 2b, Supplementary Table S6). For example, among the samples collected at the watershed niche along the Durance River in the southeast of France (*42*), *P. carotovorum* (g_106, g_57) had the same TFL as *P. versatile* (g_53). In the Netherlands, we found that *P. brasiliense* (g_145, g_194, g_429) and *P. aroidearum* (g_421, g_442) have the same TFL haplotype. *P. versatile* (g_150, g_447) and *P. odoriferum* (g_434) from the Netherlands likewise shared the same TFL haplotype. In contrast, some species did not share TFL haplotypes, for example *P. parmentieri* and *P. brasiliense*. To further understand the complete extent of TFL haplotype sharing, we mapped all TFL haplotypes on the *Pectobacterium* genus phylogeny (Figure 3a-c). We observed that these haplotypes do not follow the phylogeny (Figure 3a-c) and found widespread evidence of shared TFL haplotypes, both intra- and inter-species.

To further investigate the intra- and inter-species variation of conserved and variable carotovoricin loci at a higher resolution, we compared DNA sequences between the conserved loci and the TFL. The presence of multiple structural variants in the TFL region, including inversions, insertions, and deletions, impedes the reconstruction of a reliable multiple sequence alignment, which would be required for phylogeny reconstruction. Therefore, we used a k-mer approach to calculate pairwise distance between carotovoricin loci. Pairwise species distance was calculated as the cophenetic distance in the pangenome core gene phylogenetic tree. Under the assumption of uniform evolutionary pressure across the genome, we expect the locus distance to correlate with the species distance. In line with this, pairwise locus distances for the conserved carotovoricin locus showed a strong positive correlation ($R = 0.86$, $p < 2.2 \times 10^{-16}$) with species distance (Figure 4a). Additionally, the intra-species pairwise locus distances for the conserved locus were smaller than the inter-species distances. A positive, albeit weaker, correlation was observed between TFL locus distance and species distance ($R = 0.55$, $p < 2.2 \times 10^{-16}$) (Figure 4b). However, the locus distance range

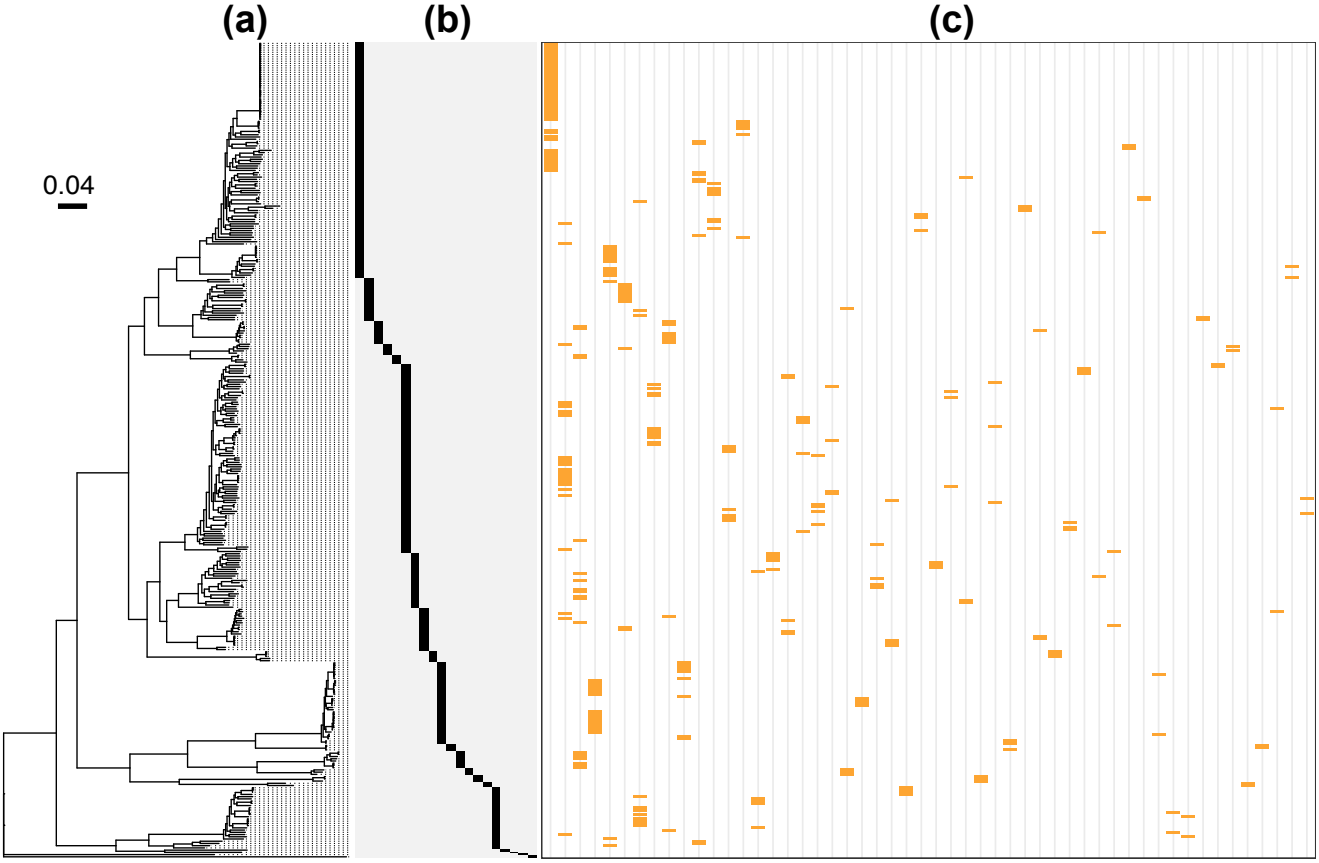


Figure 3: Carotovoricin polymorphism due to tail fiber locus diversity. (a) A core-genome phylogenetic tree is shown for 365 *Pectobacterium* genomes that have intact carotovoricin. (b) A species key marking species from left to right: *P. brasiliense*, *P. polaris*, *P. parvum*, *P. aquaticum*, *P. quasiaquaticum*, *P. versatile*, *P. carotovorum*, *P. odoriferum*, *P. actinidiae*, *P. parmentieri*, *P. wasabiae*, *P. punjabense*, *P. polonicum*, *P. zantedeschiae*, *P. peruviense*, *P. aroidearum*, *Pectobacterium sp. CFBP8739*, *P. colocasium*, *P. fontis* and *P. cacticida*. (c) The presence of TFL haplotypes is shown as yellow bands. Only the haplotypes that are present in two or more genomes are shown.

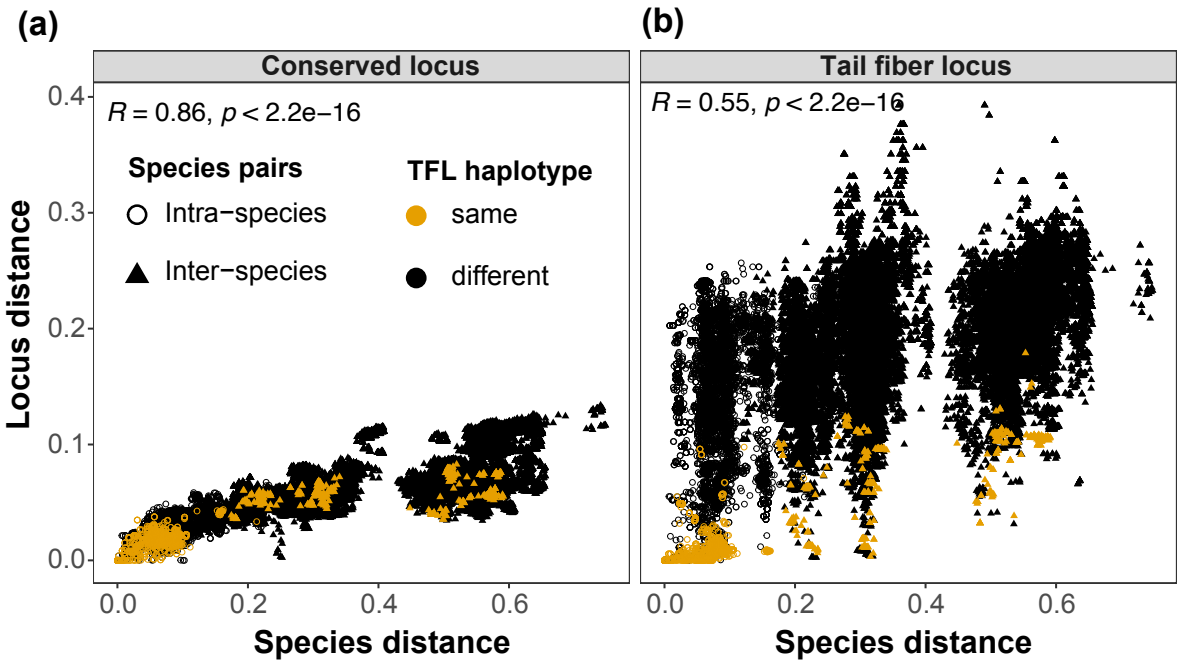


Figure 4: Sequence diversity of the tail fiber locus. (a) Pairwise locus distances for carotovoricin regions, estimated using Mash, were compared against the species evolutionary distance, estimated from the *Pectobacterium* pangenome core gene phylogeny. These comparisons are shown for two loci in the carotovoricin: (a) the conserved locus, and (b) the variable TFL. The comparisons in which both loci belong to the same TFL haplotype are colored in yellow in both scatter plots for reference. Such matching haplotypes show smaller genomic distances, also at higher evolutionary distances, for inter-species comparisons of TFLs. This trend is not observed for the conserved locus pairwise comparison in (a).

was significantly broader for both intra- and inter-species TFL compared to the conserved locus (Paired Wilcoxon test, $p < 2.2 \times 10^{-16}$). This overall broad range of pairwise distances for TFLs underlines the sequence level diversity of the tail fiber pool (Figure 4b). As expected, TFL pairs with the same haplotypes had lower locus distances, irrespective of intra- or inter-species comparisons. Interestingly, some of the inter-species TFL pairs had zero or very small locus distances, indicating (nearly) identical TFLs in different *Pectobacterium* species (Figure 4b). Together, the above results suggest that a diverse collection of TFLs is maintained by the *Pectobacterium* species. These phytopathogens likely have a common mechanism to maintain similar carotovoricin structures across species, or perhaps are exchanging carotovoricin loci.

Recombination mediated exchange of tail fiber loci

The presence of identical TFLs across different species suggests a possible HGT of these loci. To examine this, we analyzed phylogenetic incongruence between the conserved locus and TFL using 16 genomes from 10 *Pectobacterium* species. These genomes encompassed three different TFL haplotypes that lacked inversions, thus allowing us to reconstruct a multiple sequence alignment and phylogenetic tree. In the absence of HGT, the expectation is that both phylogenetic trees should show a consensus and follow the *Pectobacterium* species tree. The conserved locus phylogeny showed strains clustering by *Pectobacterium* species. On the

contrary, the tree derived from the variable TFL region alignment did not follow the *Pectobacterium* species phylogeny (Figure 5a-b). This therefore supports that the TFLs are exchanged among *Pectobacterium* species, independent of the rest of the carotovoricin locus.

A lateral transfer of such loci between two species and strains is mediated by homologous recombination. A key requirement for successful DNA recombination is the presence of a homologous region. To assess the existence of such compositional similarity in the regions flanking the TFLs, we identified locally collinear blocks in selected TFLs using Mauve (*43*, *44*). First, the intra-species comparison of *P. brasiliense* TFLs with *ein* showed the inversions in this region across several genomes, flanked by positionally conserved locally collinear blocks with high conservation across the genomes (Figure 5c). These positionally conserved collinear blocks imply a common mechanism of DNA inversion, most likely Ein-mediated. The upstream conserved collinear block in TFL overlaps with the first half of the tail fiber gene, implying a conserved N-terminal domain of the protein encoded by this gene. The remaining region of the TFL shows complex structural rearrangements and results in variable C-terminal tail fiber domains. Another comparison involving TFLs from multiple *Pectobacterium* species also showed a complex reordering and inversion of locally collinear blocks (Figure 5d). The flanking collinear blocks in TFLs without *ein* also showed positional conservation (Figure 5d), although it was less pronounced. The overall conservation of positions of collinear blocks in both sets supports our hypothesis of recombination-mediated horizontal gene transfer. Finally, the observed complex rearrangements extend the previously reported modular structure of tail fibers in *P. carotovorum* (*21*) to the broader *Pectobacterium* genus.

Discussion

Multiple *Pectobacterium* species share ecological niches and even co-occur on both symptomatic and asymptomatic plant material (*45–47*). The success and stability of such a heterogeneous community depend on having the means to overcome competition from other closely related strains or species. Carotovoricin, a tailocin, is constitutively expressed and produced by *Pectobacterium* species during host infection, both on symptomatic and surrounding healthy tissue (*36*). As tailocins play an important role in strain-level interactions, we systematically studied the conservation and dynamics of carotovoricin in the *Pectobacterium* genus using a pangenomics approach. We detected differential conservation of various loci within the carotovoricin biosynthesis gene cluster. We found evidence for both intra- and inter-species HGT in one of these variable loci in carotovoricin. Here, we discuss the possible implications of the observed tailocin dynamics in *Pectobacterium* species and compare them with other tailocin-producing species.

Our pangenome-wide analysis shows that carotovoricin is overall conserved in the *Pectobacterium* genus and was likely acquired by the common ancestor of *Pectobacterium* species. Conservation of both carotovoricin core functional genes and their chromosomal location across the *Pectobacterium* genus underscores the important role of this tailocin. Additionally, carotovoricin was found to be present in a single copy and no other tailocin was identified in *Pectobacterium* species genomes. This is different from *Pseudomonas* species, where multiple tailocins exist with various genome integration sites, evolutionary history and regulatory genes (*48*, *49*). We observed species-wide partial degradation of carotovoricin in *P. betavasculorum* and *P.*

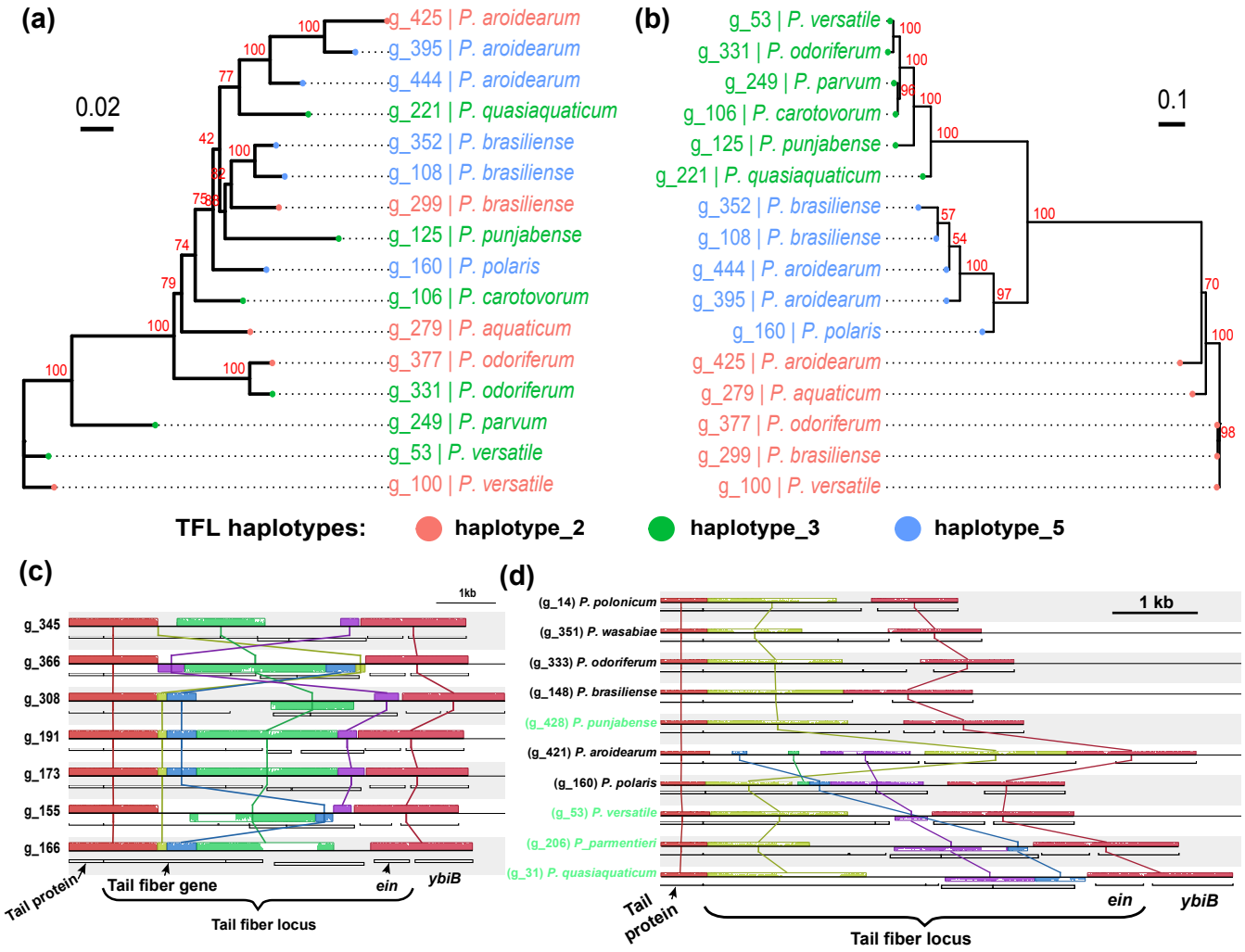


Figure 5: Horizontal gene transfer of tail fiber loci. (a) An unrooted maximum-likelihood phylogenetic tree for the conserved locus of carotovoricin from 16 genomes. (b) An unrooted maximum-likelihood phylogenetic tree generated from the TFLs from 16 genomes comprising 3 different haplotypes. For both the trees, bootstrap values are shown next to internal nodes and tips are colored as per the TFL haplotype. (c) Complex structural rearrangements of tail fiber locus (TFL) haplotypes are shown in the form of Mauve collinear blocks for carotovoricin loci in *P. brasiliense* accessions with the invertase gene *ein*. (d) Mauve collinear blocks in TFLs of various *Pectobacterium* species are shown and labels for TFLs that have invertase *ein* are colored green. Locally collinear blocks are shown as colored boxes above (forward) and below (reverse) the black horizontal line that represents the TFL haplotype region in a genome. Conserved blocks across the genomes are linked by vertical lines. Inside each block, a similarity profile is drawn which represents the average level of conservation across the TFL haplotypes. White areas represent sequences specific to a TFL and are not aligned.

atrosepticum. Although we found transcription evidence for the degraded carotovoricin in *P. atrosepticum*, the function of such partial tailocin biosynthesis gene clusters remains unclear, as they do not code for a complete functional unit. Nonetheless, this suggests that the phage tail-like bacteriocins are not immune to dynamic degradation, a phenomenon observed in their prophage ancestors where a rapid initial deactivation is followed by gradual decay (*50*). We hypothesize that their role in inter-bacterial interactions likely imposes a purifying selection pressure that protects tailocins from rapid initial deactivation, resulting in different domestication dynamics of tailocins compared to prophages.

We observed that *P. brasiliense* strains isolated from aquatic niches lacked both carotovoricin and type III secretion system (T3SS) gene clusters, systems involved in inter-bacterial competition and bacteria-host interaction, respectively. These distinct ecological niches with lower bacterial densities and lack of plant hosts provide different competition profiles. Given that tailocin production is dependent on environmental cues, with its induction occurring in the infected plant host but absent in nutrient-deficient environments such as river water (*36*), a strictly aquatic niche may not exert strong purifying selection pressure to maintain a function that is not used. Therefore, we hypothesize that prolonged growth of these *P. brasiliense* strains in less competitive and host-deprived niches may have resulted in the loss of both carotovoricin and T3SS.

Although carotovoricin was broadly conserved in the *Pectobacterium* genus, we observed varying degrees of conservation for multiple loci in the carotovoricin-encoding region, namely the nitric oxide deoxygenase gene, the tape measure gene and the tail fiber locus. Multiple sequence alignment has been used to estimate TFL haplotypes for *Pseudomonas* species tailocins (*30*). However, generating a multiple sequence alignment for genomic regions with complex structural variation, such as the TFLs, is challenging. Our pangenome approach to identify TFL haplotypes overcomes this limitation. We identified 93 haplotypes for the *Pectobacterium* carotovoricin TFL. Indeed, variation in the number and composition of tail fiber genes is a generic property of tailocins that determines their host specificity (*51*). We hypothesize that the *Pectobacterium* community in a plant host niche maintains a pool of carotovoricin tail fiber diversity to resist invasion by other closely related *Pectobacterium* species pathogens. This maintenance of TFL diversity in *Pectobacterium* might be over both space and time, as reported for another phytopathogen *Pseudomonas* (*30*). Such a mechanism could be beneficial for the established colonizers in case of multi-species soft-rot infections, which are commonly observed in *Pectobacterium* species pathogens (*33*, *45*, *47*). Indeed, such a competitive exclusion mechanism is noted for R-tailocin producing *Pseudomonas chlororaphis* in mixed-species rhizosphere communities (*28*, *52*).

Remarkably, the mechanism to achieve TFL diversity, Ein-mediated inversion (*29*), was absent in approximately half of the carotovoricins within the *Pectobacterium* genus. However, we show existence of an additional mechanism to maintain TFL diversity, horizontal acquisition of TFLs, across the *Pectobacterium* genus. Phylogenetic incongruence showed that HGT occurs in the tail fiber encoding region of the carotovoricin and is likely mediated by site-specific recombination. Although the recombinational exchange of TFLs is known in other tailocin-producing bacteria (*27*), two mechanisms complementing each other to maintain a pool of TFLs make *Pectobacterium* phytopathogens a unique case, to our knowledge. Based on the observed TFL diversity and sharing of TFL haplotypes across species, innovations in terms of functional

diversity appear to be shared within the *Pectobacterium* genus.

Given the selection pressure for carotovoricin in *Pectobacterium*, and species-wide partial deletion of carotovoricin in *P. atrosepticum* and *P. betavasculorum*, it is intriguing why these two species never reacquired a functional copy of carotovoricin by HGT from other *Pectobacterium* species. One explanation could be that HGT of a complete carotovoricin locus is not possible due to its size or required homology as we only observed transfer of TFLs in the *Pectobacterium* genus. Another possibility is that these species do not co-exist with other *Pectobacterium* species, because of niche adaptation or other antagonistic interactions. This hypothesis is supported by our TFL haplotype analysis of *P. parmentieri* and *P. brasiliense*, which are known to not co-exist on the same plant (*45*). Specifically, *P. parmentieri* does not share any of the 8 TFL haplotypes with *P. brasiliense*.

Sugar residues from the O-antigen of the lipopolysaccharide capsule are the main receptors of tailocin tail fibers (*25*, *53–55*). These O-antigen biosynthesis genes are typically organized in a cluster and are known to undergo HGT with closely related strains (*56–58*). Consistent with this, *Pectobacterium* species exhibit diverse O-antigen chemical structures and instances of similar O-antigen structure between different species have been observed (*59*). A recent study in *Pseudomonas* species revealed maintenance of tailocin tail-fiber and O-antigen diversity across space and time and showed their co-evolution. Based on our results, we anticipate that the horizontal gene transfer of carotovoricin TFLs is also driving the co-evolution of O-antigen biosynthesis gene clusters in *Pectobacterium* pathogens.

Several studies have shown application of phages for biocontrol of *Pectobacterium* species and other plant pathogens (*60–66*). Recently, tailocin was used to target a *Clostridium difficile* infection without affecting the gut microbiota diversity in the mouse model (*67*). Tailocins are also used in controlling phytopathogens; for example, rhizoviticin from *Allorhizobium vitis* to treat grapevine crown gall disease (*14*), *Burkholderia gladioli* tailocin to treat Rice seed-borne disease (*68*) and *P.syringae* tailocin for biocontrol of *Nicotiana benthamiana* (*69*). Furthermore, genetic engineering and synthetic biology allow the use of phages as scaffolds to design a broad range of tailocins, as well as strain-specific versions (*70*, *71*). Another application, developing a broad-range virulent phage using tailocin tail fibers, is also possible. One such example is the broad-range chimeric phage Milagro targeting *Burkholderia*, designed by combining virulent phage and tail fiber receptor binding domains from *Burkholderia* tailocin (*72*). The pangenome approach taken here allows a detailed characterization of the heterogeneity of tailocins, especially in the variable tail fiber encoding region, which is essential for developing such synthetic tailocin applications for biocontrol.

This study presented a comprehensive characterization of tailocins in the *Pectobacterium* genus. In a genus-wide pangenome approach, we studied the conservation of tailocins as well as tail fiber diversity. Exploiting verified settings to detect homology group signatures to define tail fiber haplotypes uncovered large-scale HGT of tail fibers in *Pectobacterium*, which is supported by a traditional phylogenetic approach. We further zoomed in into the tail fiber locus to study the modular nature of tail fiber proteins and the genus-wide HGT that must be mediated by site-specific recombination. These findings underline the active interactions between members of the *Pectobacterium* pathogen community mediated by tailocins. It supports a role for horizontal gene transfer in achieving tail fiber diversity and maintaining a tailocin pool in phytopathogen

communities.

Methods

Genome collection and pangenome construction

A *Pectobacterium* genus pangenome comprising 454 genomes was recently constructed by us (*33*). Briefly, we used PanTools (v4.1.1) (*37*, *38*) to construct a De Bruijn graph pangenome, including gene annotations, and functional annotations using InterProScan (v5.56-89.0) (*73*, *74*) and EggNOG-mapper (v2.1.10) (*75*, *76*). The protein-coding genes in the pangenome were clustered using PanTools' *group* command with relaxation setting of 4, which corresponds to 65% sequence similarity, 0.05 intersection rate for the k-mer set comparison, a 7.2 inflation rate for Markov clustering and a contrast factor of 5.

Phylogenetic analysis

Multiple sequence alignment of genes in each of the core homology groups was performed using the PanTools *msa* function. These alignments were used to construct a maximum likelihood phylogeny using the *core_phylogeny* function (*77*), which uses a concatenation-based phylogenomic approach. Phylogenetic trees were processed using R (v4.2.1) packages Ape (v5.7.1) (*78*, *79*) and treeio (v1.22.0) (*80*) and visualized using ggtree (v3.6.2) (*81*) and ComplexHeatmap (v2.15.1) (*82*).

Carotovoricin identification

Prophages were predicted using the geNomad (v1.5.0) pipeline (*83*) and completeness was assessed using checkV (v1.0.1) (*84*). Any bacterial genomic regions at the prophage borders were identified using CheckV and subsequently trimmed. Putative prophages without any gene of viral origin were not considered. Each prophage was represented as a sequence of pangenome homology groups, referred as prophage signatures. Two or more prophage signatures from one genome matching together a single larger prophage signature from another genome with $\geq 80\%$ coverage were considered as fragmented prophages due to poor quality genome assembly. Such fragmented prophages were merged into a single one and labeled as fragmented (110 fragmented prophages were merged into 50). All putative prophages with length < 5kb were further excluded, resulting into a final collection of 1,369 prophages.

The pairwise similarity between prophage homology group signatures was quantified as the syntenic conservation of homology groups: the syntenic Jaccard index. To calculate the syntenic Jaccard index, dynamic programming was used to align two prophage signatures, with a score of $+5$ for a match and -2 for a mismatch of homology groups between two prophage signatures. The longest common subsequence between two prophage signatures, with at most two gaps and at least five matching homology groups, was considered a valid alignment of homology groups. The syntenic matching homology groups from such valid alignments were used to determined the intersection size while calculating the Jaccard index. The syntenic Jaccard index matrix for all-vs-all comparisons was clustered using complete linkage hierarchical clustering with the 'hclust' function in R. Next to various small clusters, this yielded one large group of 428 orthologous prophages. Among these, prophages from *P. carotovorum* matched the phage tail-like bacteriocin

encoding region, called "Carotovoricin Er". All other prophages belonging to this cluster showed similar gene structures and hence were denoted as carotovoricin.

Transcriptomics evidence for carotovoricin in *P. atrosepticum*

RNAseq normalized count data for *P. atrosepticum* was downloaded from the GEO record GSE196675 (*39*). Genes from the partial carotovoricin locus were located in the region NC_004547.2:3113731-3119520. These genes were found to be differentially expressed upon polyphenol treatment.

Carotovoricin region synteny visualization

A 5kb flanking region for each prophage was included to provide broader context to understand integration at same or different site in the genome. An R script was written to extract the homology groups for carotovoricin regions and the data was stored in the JSON format required by clustermap.js (https://github.com/gamcil/clustermap.js). For species wise visualization for *P. brasiliense* and *P. versatile*, carotovoricins for the respective species were clustered based on the homology group present-absence matrix and the dendrogram was cut at the appropriate to select representatives for clinker plotting.

Comparative phylogenetics analysis of TFLs

Carotovoricin region from tail protein (hg_22427604) to *ybiB* (hg_22427603) was defined as the tail fiber locus and the remaining upstream region as the conserved locus. Haplotypes for the tail fiber locus were determined based on the unique ordering of homology groups in this region across all *Pectobacterium* genomes. Locus distance, estimated as a sequence-based distance between carotovoricin pairs was calculated using the MinHash approach in Mash (v2.3) (*85*). Pairwise locus distance was calculated separately for the conserved carotovoricin locus and tail fiber locus. Pairwise species distances were calculated as the cophenetic distance using 'cophenetic.phylo()' function from the R package Ape. Cophenetic distances are calculated as pairwise distances between the pairs of tips of a phylogenetic tree using its branch lengths.

Carotovoricins from 16 genomes spanning three TFL haplotypes and representing 10 *Pectobacterium* species were selected for performing comparative phylogenetics analysis. The DNA sequences for conserved loci and TFLs for these carotovoricin regions were independently aligned using MAFFT (v7.508) (*86*). Maximum-likelihood phylogenetic trees for these two loci were estimated using iqtree2 (2.2.0.3) (*77*). The trees were visualized using R (v4.2.1) package ggtree (v3.6.2) (*81*).

Visualization of structural variation in tail fiber locus

DNA sequences of carotovoricin regions from tail protein (hg_22427604) to *ybiB* (hg_22427603) were extracted as TFL sequences. The homology group level annotations from the PanTools' pangenome and DNA sequences were combined into a GenBank file format using the *seqret* tool from the EMBOSS software suite (v6.6.0.0) (*87*). These GenBank format sequences were aligned and visualized using the 'progressiveMauve' command of the Mauve v(2.4.0) program (*43*, *44*).

Data and code availability

All scripts used for preprocessing and analysis are available in the GitHub repository https://lakhanp1.

github.io/Pectobacterium_pangenome. Zenodo DOI for the GitHub repository: https://doi.org/10.5281/zenodo.16585541. Metadata for 454 genome assemblies used for the analysis is listed in Supplementary Table S1.

Supplementary information

Additional file 1: Supplementary_Tables.xlsx

Additional file 2: Supplementary_Material.pdf

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Dutch Ministry of Economic Affairs in the Topsector Program "Horticulture and Starting Materials" under the theme "Plant Health" (project number: LWV20.235) and its partners (NVWA, NAK, Naktuinbouw and BKD).

Authors' contributions

LP, SS, TL, DR: Conceptualization. LP: Data curation, Formal analysis, Investigation, Software, Writing – original draft. SS, TL, DR: Supervision, Funding acquisition, Project administration. LP, SS, TL, DR, AK: Writing - review and editing. All authors reviewed the manuscript.

Acknowledgements

We would like to acknowledge Inge van Duivenbode from Dutch General Inspection Service for agricultural seeds and seed potatoes (NAK) and Michiel J. C. Pel from the Netherlands Institute for Vectors, Invasive plants and Plant health (NIVIP) for providing *Pectobacterium* isolates and their valuable discussion in this research.

References

(*1*) McInerney, J. O., McNally, A., and O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nature Microbiology 2*, 1–5.

(*2*) Treangen, T. J., and Rocha, E. P. C. (2011). Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. *PLoS Genetics 7*, e1001284.

(*3*) Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology 49*, 277–300.

(*4*) Varani, A. M., Monteiro-Vitorello, C. B., Nakaya, H. I., and Van Sluys, M.-A. (2013). The Role of Prophage in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology 51*, 429–451.

(*5*) Sharma, V., Hünnefeld, M., Luthe, T., and Frunzke, J. (2023). Systematic analysis of prophage elements in actinobacterial genomes reveals a remarkable phylogenetic diversity. *Scientific Reports 13*, 4410.

(*6*) Costa, A. R., Monteiro, R., and Azeredo, J. (2018). Genomic analysis of Acinetobacter baumannii prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness. *Scientific Reports 8*, 15346.

(*7*) Lang, A. S., and Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends in Microbiology 15*, 54–62.

(*8*) Heiman, C. M., Vacheron, J., and Keel, C. (2023). Evolutionary and ecological role of extracellular contractile injection systems: from threat to weapon. *Frontiers in Microbiology 14*, 1264877.

(*9*) Czajkowski, R. (2016). Bacteriophages of Soft Rot Enterobacteriaceae—a minireview. *FEMS Microbiology Letters 363*, fnv230.

(*10*) Czajkowski, R. (2019). May the Phage be With You? Prophage-Like Elements in the Genomes of Soft Rot Pectobacteriaceae: Pectobacterium spp. and Dickeya spp. *Frontiers in Microbiology 10*, 138.

(*11*) Miroshnikov, K. A., Evseev, P. V., Lukianova, A. A., and Ignatov, A. N. (2021). Tailed Lytic Bacteriophages of Soft Rot Pectobacteriaceae. *Microorganisms 9*, 1819.

(*12*) Ghequire, M. G., and De Mot, R. (2014). Ribosomally encoded antibacterial proteins and peptides from Pseudomonas. *FEMS Microbiology Reviews 38*, 523–568.

(*13*) Yao, G. W., Duarte, I., Le, T. T., Carmody, L., LiPuma, J. J., Young, R., and Gonzalez, C. F. (2017). A Broad-Host-Range Tailocin from Burkholderia cenocepacia. *Applied and Environmental Microbiology 83*, e03414–16.

(*14*) Ishii, T. et al. (2024). Rhizoviticin is an alphaproteobacterial tailocin that mediates biocontrol of grapevine crown gall disease. *The ISME Journal 18*, wrad003.

(*15*) Borowicz, M., Krzyżanowska, D. M., Narajczyk, M., Sobolewska, M., Rajewska, M., Czaplewska, P., Węgrzyn, K., and Czajkowski, R. (2023). Soft rot pathogen Dickeya dadantii 3937 produces tailocins resembling the tails of Peduovirus P2. *Frontiers in Microbiology 14*, DOI: 10.3389/fmicb.2023.1307349.

(*16*) Kamimiya, S., Izaki, K., and Takahashi, H. (1977). Bacteriocins in *Erwinia aroideae* with Tail Like Structure of Bacteriophages. *Agricultural and Biological Chemistry 41*, 911–912.

(*17*) Itoh, Y., Izaki, K., and Takahashi, H. (1978). PURIFICATION AND CHARACTERIZATION OF A BACTERIOCIN FROM *ERWINIA CAROTOVORA*. *The Journal of General and Applied Microbiology 24*, 27–39.

(*18*) Arizala, D., and Arif, M. (2019). Genome-Wide Analyses Revealed Remarkable Heterogeneity in Pathogenicity Determinants, Antimicrobial Compounds, and CRISPR-Cas Systems of Complex Phytopathogenic Genus Pectobacterium. *Pathogens (Basel, Switzerland) 8*, E247.

(*19*) Stice, S. P., Jan, H.-H., Chen, H.-C., Nwosu, L., Shin, G. Y., Weaver, S., Coutinho, T., Kvitko, B. H., and Baltrus, D. A. (2023). Pantailocins: phage-derived bacteriocins from *Pantoea ananatis* and *Pantoea stewartii* subsp. *indologenes*. *Applied and Environmental Microbiology 89*, ed. by Alexandre, G., e00929–23.

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

(*21*) YAMADA, K., HIROTA, M., NIIMI, Y., NGUYEN, H. A., TAKAHARA, Y., KAMIO, Y., and KANEKO, J. (2006). Nucleotide Sequences and Organization of the Genes for Carotovoricin (Ctv) from Erwinia carotovora Indicate That Ctv Evolved from the Same Ancestor as Salmonella typhi Prophage. *Bioscience, Biotechnology, and Biochemistry 70*, 2236–2247.

(*22*) Ge, P., Scholl, D., Leiman, P. G., Yu, X., Miller, J. F., and Zhou, Z. H. (2015). Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nature Structural & Molecular Biology 22*, 377–382.

(*23*) Desfosses, A., Venugopal, H., Joshi, T., Felix, J., Jessop, M., Jeong, H., Hyun, J., Heymann, J. B., Hurst, M. R. H., Gutsche, I., and Mitra, A. K. (2019). Atomic structures of an entire contractile injection system in both the extended and contracted states. *Nature Microbiology 4*, 1885–1894.

(*24*) Scholl, D. (2017). Phage Tail–Like Bacteriocins. *Annual Review of Virology 4*, 453–467.

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

(*26*) Bhattacharjee, R., Nandi, A., Sinha, A., Kumar, H., Mitra, D., Mojumdar, A., Patel, P., Jha, E., Mishra, S., Rout, P. K., Panda, P. K., Suar, M., and Verma, S. K. (2022). Phage-tail-like bacteriocins as a biomedical platform to counter anti-microbial resistant pathogens. *Biomedicine & Pharmacotherapy 155*, 113720.

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

(*28*) Dorosky, R. J., Pierson, L. S., and Pierson, E. A. (2018). Pseudomonas chlororaphis Produces Multiple R-Tailocin Particles That Broaden the Killing Spectrum and Contribute to Persistence in Rhizosphere Communities. *Applied and Environmental Microbiology 84*, e01230–18.

(*29*) Nguyen, H. A., Tomita, T., Hirota, M., Kaneko, J., Hayashi, T., and Kamio, Y. (2001). DNA Inversion in the Tail Fiber Gene Alters the Host Range Specificity of Carotovoricin Er, a Phage-Tail-Like Bacteriocin of Phytopathogenic Erwinia carotovora subsp. carotovora Er. *Journal of Bacteriology 183*, 6274–6281.

(*30*) Backman, T. et al. (2024). A phage tail-like bacteriocin suppresses competitors in metapopulations of pathogenic bacteria. *Science 384*, eado0713.

(*31*) van der Wolf, J. M., de Haan, E. G., Kastelein, P., Krijger, M., de Haas, B. H., Velvis, H., Mendes, O., Kooman-Gersmann, M., and van der Zouwen, P. S. (2017). Virulence of Pectobacterium carotovorum subsp. brasiliense on potato compared with that of other Pectobacterium and Dickeya species under climatic conditions prevailing in the Netherlands. *Plant Pathology 66*, 571–583.

(*32*) de Werra, P., Debonneville, C., Kellenberger, I., and Dupuis, B. (2021). Pathogenicity and Relative Abundance of Dickeya and Pectobacterium Species in Switzerland: An Epidemiological Dichotomy. *Microorganisms 9*, 2270.

(*33*) Pardeshi, L. A., van Duivenbode, I., Pel, M. J., Jonkheer, E. M., Kupczok, A., de Ridder, D., Smit, S., and van der Lee, T. A. J. (2025). Pangenomics to understand prophage dynamics in the Pectobacterium genus and the radiating lineages of Pectobacterium brasiliense. *Microbial Genomics 11*, 001392.

(*34*) Barny, M.-A., Thieffry, S., Gomes De Faria, C., Thebault, E., and Pédron, J. (2024). Bacterial pathogens dynamic during multi-species infections. *Peer Community Journal 4*, e49.

(*35*) De Boer, S. H., Li, X., and Ward, L. J. (2012). Pectobacterium spp. Associated with Bacterial Stem Rot Syndrome of Potato in Canada. *Phytopathology® 102*, 937–947.

(*36*) Borowicz, M., Krzyzanowska, D. M., Sobolewska, M., Narajczyk, M., Mruk, I., Czaplewska, P., Pedron, J., Barny, M.-A., Canto, P. Y., Dziadkowiec, J., and Czajkowski, R. Tailocin-mediated interactions in Soft Rot Pectobacteriaceae, 2024.

(*37*) Sheikhizadeh, S., Schranz, M. E., Akdel, M., de Ridder, D., and Smit, S. (2016). PanTools: representation, storage and exploration of pan-genomic data. *Bioinformatics (Oxford, England) 32*, i487–i493.

(*38*) Jonkheer, E. M., van Workum, D.-J. M., Sheikhizadeh Anari, S., Brankovics, B., de Haan, J. R., Berke, L., van der Lee, T. A. J., de Ridder, D., and Smit, S. (2022). PanTools v3: functional annotation, classification and phylogenomics. *Bioinformatics 38*, 4403–4405.

(*39*) Kang, J. E., Hwang, S., Yoo, N., Kim, B. S., and Chung, E.-H. (2022). A resveratrol oligomer, hopeaphenol suppresses virulence activity of Pectobacterium atrosepticum via the modulation of the master regulator, FlhDC. *Frontiers in Microbiology 13*, 999522.

(*40*) Jonca, J., Pirhonen, M., Waleron, M. M., Gawor, J., Mrozik, A., Smoktunowicz, M., Waleron, K., and Waleron, M. (2024). Comprehensive phenomic and genomic studies of the species, Pectobacterium cacticida and proposal for reclassification as Alcorniella cacticida comb. nov. *Frontiers in Plant Science 15*, DOI: 10.3389/fpls.2024.1323790.

(*41*) Portier, P., Pédron, J., Taghouti, G., Dutrieux, C., and Barny, M.-A. (2020). Updated Taxonomy of Pectobacterium Genus in the CIRM-CFBP Bacterial Collection: When Newly Described Species Reveal "Old" Endemic Population. *Microorganisms 8*, 1441.

(*42*) Ben Moussa, H., Bertrand, C., Rochelle-Newall, E., Fiorini, S., Pédron, J., and Barny, M.-A. (2022). The diversity of soft rot Pectobacteriaceae along the Durance River stream in the south-east of France revealed by multiple seasonal surveys. *Phytopathology®*, DOI: 10.1094/PHYTO-12-21-0515-R.

(*43*) Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research 14*, 1394–1403.

(*44*) Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE 5*, e11147.

(*45*) Smoktunowicz, M., Jonca, J., Stachowska, A., May, M., Waleron, M. M., Waleron, M., and Waleron, K. (2022). The International Trade of Ware Vegetables and Orna-Mental Plants—An Underestimated Risk of Accelerated Spreading of Phytopathogenic Bacteria in the Era of Globalisation and Ongoing Climatic Changes. *Pathogens 11*, 728.

(*46*) Motyka-Pomagruk, A., Zoledowska, S., Sledz, W., and Lojkowska, E. (2021). The occurrence of bacteria from different species of Pectobacteriaceae on seed potato plantations in Poland. *European Journal of Plant Pathology 159*, 309–325.

(*47*) Zhou, J., Hu, M., Hu, A., Li, C., Ren, X., Tao, M., Xue, Y., Chen, S., Tang, C., Xu, Y., Zhang, L., and Zhou, X. (2022). Isolation and Genome Analysis of Pectobacterium colocasium sp. nov. and Pectobacterium aroidearum, Two New Pathogens of Taro. *Frontiers in Plant Science 13*.

(*48*) Ghequire, M. G., Dillen, Y., Lambrichts, I., Proost, P., Wattiez, R., and De Mot, R. (2015). Different Ancestries of R Tailocins in Rhizospheric Pseudomonas Isolates. *Genome Biology and Evolution 7*, 2810–2828.

(*49*) Hockett, K. L., Renner, T., and Baltrus, D. A. (2015). Independent Co-Option of a Tailed Bacteriophage into a Killing Complex in Pseudomonas. *mBio 6*, 10.1128/mbio.00452–15.

(*50*) Bobay, L.-M., Touchon, M., and Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences 111*, 12127–12132.

(*51*) Patz, S., Becker, Y., Richert-Pöggeler, K. R., Berger, B., Ruppel, S., Huson, D. H., and Becker, M. (2019). Phage tail-like particles are versatile bacterial nanomachines - A mini-review. *Journal of Advanced Research 19*, 75–84.

(*52*) Dorosky, R. J., Yu, J. M., Pierson, L. S., and Pierson, E. A. (2017). Pseudomonas chlororaphis Produces Two Distinct R-Tailocins That Contribute to Bacterial Competition in Biofilms and on Roots. *Applied and Environmental Microbiology 83*, e00706–17.

(*53*) Köhler, T., Donner, V., and van Delden, C. (2010). Lipopolysaccharide as Shield and Receptor for R-Pyocin-Mediated Killing in Pseudomonas aeruginosa. *Journal of Bacteriology 192*, 1921–1928.

(*54*) Kocíncová, D., and Lam, J. S. (2013). A deletion in the promoter in many serotypes of Pseudomonas aeruginosa accounts for the lack of a terminal glucose residue in the core oligosaccharide and resistance to killing by R3-pyocin. *Molecular Microbiology 89*, 464–478.

(*55*) Heiman, C. M., Maurhofer, M., Calderon, S., Dupasquier, M., Marquis, J., Keel, C., and Vacheron, J. (2022). Pivotal role of O-antigenic polysaccharide display in the sensitivity against phage tail-like particles in environmental Pseudomonas kin competition. *The ISME Journal 16*, 1683–1693.

(*56*) Wang, L., Wang, Q., and Reeves, P. R. In *Endotoxins: Structure, Function and Recognition*, Wang, X., and Quinn, P. J., Eds.; Springer Netherlands: Dordrecht, 2010, pp 123–152.

(*57*) Wyres, K. L., Lambertsen, L. M., Croucher, N. J., McGee, L., von Gottberg, A., Liñares, J., Jacobs, M. R., Kristinsson, K. G., Beall, B. W., Klugman, K. P., Parkhill, J., Hakenbeck, R., Bentley, S. D., and Brueggemann, A. B. (2013). Pneumococcal Capsular Switching: A Historical Perspective. *The Journal of Infectious Diseases 207*, 439–449.

(*58*) Lang, C., Fruth, A., Campbell, I. W., Jenkins, C., Smith, P., Strockbine, N., Weill, F.-X., Nübel, U., Grad, Y. H., Waldor, M. K., and Flieger, A. (2023). O-Antigen Diversification Masks Identification of Highly Pathogenic Shiga Toxin-Producing Escherichia coli O104:H4-Like Strains. *Microbiology Spectrum 11*, e00987–23.

(*59*) Szulta, S., and Kornicka, A. (2023). Pectobacterium and Dickeya genus—A review on structural variations of O-polysaccharides and their role in the pathogenic process of plants. *Plant Pathology 72*, 998–1010.

(*60*) Czajkowski, R., Ozymko, Z., de Jager, V., Siwinska, J., Smolarska, A., Ossowicki, A., Narajczyk, M., and Lojkowska, E. (2015). Genomic, proteomic and morphological characterization of two novel broad host lytic bacteriophages ΦPD10.3 and ΦPD23.1 infecting pectinolytic Pectobacterium spp. and Dickeya spp. *PloS One 10*, e0119812.

(*61*) Muturi, P., Yu, J., Maina, A. N., Kariuki, S., Mwaura, F. B., and Wei, H. (2019). Bacteriophages Isolated in China for the Control of Pectobacterium carotovorum Causing Potato Soft Rot in Kenya. *Virologica Sinica 34*, 287–294.

(*62*) Lee, S., Vu, N.-T., Oh, E.-J., Rahimi-Midani, A., Thi, T.-N., Song, Y.-R., Hwang, I.-S., Choi, T.-J., and Oh, C.-S. (2021). Biocontrol of Soft Rot Caused by Pectobacterium odoriferum with Bacteriophage phiPccP-1 in Kimchi Cabbage. *Microorganisms 9*, 779.

(*63*) Zaczek-Moczydłowska, M. A., Young, G. K., Trudgett, J., Plahe, C., Fleming, C. C., Campbell, K., and Hanlon, R. O. (2020). Phage cocktail containing Podoviridae and Myoviridae bacteriophages inhibits the growth of Pectobacterium spp. under in vitro and in vivo conditions. *PLOS ONE 15*, e0230842.

(*64*) McKenna, F., El-Tarabily, K. A., Hardy, G. E. S. J., and Dell, B. (2001). Novel in vivo use of a polyvalent Streptomyces phage to disinfest Streptomyces scabies-infected seed potatoes. *Plant Pathology 50*, 666–675.

(*65*) Czajkowski, R., Ozymko, Z., and Lojkowska, E. (2014). Isolation and characterization of novel soil-borne lytic bacteriophages infecting Dickeya spp. biovar 3 ('D. solani'). *Plant Pathology 63*, 758–772.

(*66*) Carstens, A. B., Djurhuus, A. M., Kot, W., Jacobs-Sera, D., Hatfull, G. F., and Hansen, L. H. (2018). Unlocking the Potential of 46 New Bacteriophages for Biocontrol of Dickeya Solani. *Viruses 10*, 621.

(*67*) Gebhart, D., Lok, S., Clare, S., Tomas, M., Stares, M., Scholl, D., Donskey, C. J., Lawley, T. D., and Govoni, G. R. (2015). A Modified R-Type Bacteriocin Specifically Targeting Clostridium difficile Prevents Colonization of Mice without Affecting Gut Microbiota Diversity. *mBio 6*, 10.1128/mbio.02368–14.

(*68*) Kouzai, Y., Sagehashi, Y., Watanabe, R., Kajiwara, H., Suzuki, N., Ono, H., Naito, K., and Akimoto-Tomiyama, C. (2024). BglaTNB6, a tailocin produced by a plant-associated nonpathogenic bacterium, prevents rice seed-borne bacterial diseases. *PLOS Pathogens 20*, e1012645.

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

(*70*) Williams, S., Gebhart, D., Martin, D., and Scholl, D. (2008). Retargeting R-type pyocins to generate novel bactericidal protein complexes. *Applied and Environmental Microbiology 74*, 3868–3876.

(*71*) Woudstra, C., Sørensen, A. N., Sørensen, M. C. H., and Brøndsted, L. (2024). Strategies for developing phages into novel antimicrobial tailocins. *Trends in Microbiology*, DOI: 10.1016/j.tim.2024.03.003.

(*72*) Yao, G., Le, T., Korn, A. M., Peterson, H. N., Liu, M., Gonzalez, C. F., and Gill, J. J. (2023). Phage Milagro: a platform for engineering a broad host range virulent phage for Burkholderia. *Journal of Virology 97*, e0085023.

(*73*) Jones, P. et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics 30*, 1236–1240.

(*74*) Blum, M. et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research 49*, D344–D354.

(*75*) Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution 38*, 5825–5829.

(*76*) Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C. P., Giner-Lamia, J., Mende, D. R., Kirsch, R., Rattei, T., Letunic, I., Jensen, L. J., Bork, P., von Mering, C., and Huerta-Cepas, J. (2023). eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research 51*, D389–D394.

(*77*) Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution 37*, 1530–1534.

(*78*) Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics 20*, 289–290.

(*79*) Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics 35*, 526–528.

(*80*) Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., and Yu, G. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution 37*, 599–603.

(*81*) Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T. T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta 1*, e56.

(*82*) Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics 32*, 2847–2849.

(*83*) Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., and Kyrpides, N. C. (2023). Identification of mobile genetic elements with geNomad. *Nature Biotechnology*, 1–10.

(*84*) Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology 39*, 578–585.

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

(*86*) Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics 20*, 1160–1166.

(*87*) Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics: TIG 16*, 276–277.

Supplementary data, belonging to Pardeshi *et al.*, "The evolutionary tale of tailocins in *Pectobacterium* show genus-wide homologous recombination and inversions of tail fiber genes"

Supplementary Results

Rare events of carotovoricin disruption

In addition to the species-wide loss of carotovoricin, there were seven strains from various *Pectobacterium* species that showed disruption of the carotovoricin cluster. One *P. brasiliense* isolated from potato plant (g_149) showed an insertion of a bacteriophage genome within the carotovoricin cluster. One *P. versatile* isolated from the potato plant (g_313) showed a deletion of five genes in carotovoricin: tail sheath, tail core, ferredoxin I, tail assembly chaperon and a partial tail tape measure gene. One *P. carotovorum* isolate (g_15) sampled from the Japanese angelica tree, recently reclassified as *P. araliae* (*1*), showed complete loss of the carotovoricin locus. Finally, we identified four *P. brasiliense* isolates lacking carotovoricin, all sampled in Europe after 2017. A homology group signature comparison of these isolates showed that the flanking gene neighborhood was intact, ruling out assembly artifacts (Supplementary Figure S2). These four isolates were sampled from fresh water in the Netherlands and France (g_182, g_185 and g_236), which has shown to be a rare ecological niche for *P. brasiliense* (*2*), and from insect traps (g_177). Based on the core gene phylogenetic tree, these four isolates formed a monophyletic, yet non-clonal group (Supplementary Figure S1). We also assessed whether other important systems involved in interaction are lost in these genomes. A detailed comparison with the other 134 *P. brasiliense* genomes showed that the T3SS, involved in the injection of bacterial effector proteins into the eukaryotic host cells to modulate host response during infection (*3*, *4*), was also missing in these four isolates (Supplementary Figure S3). With the existing data, it is difficult to ascertain either independence or association in the deletion of these two systems.

References

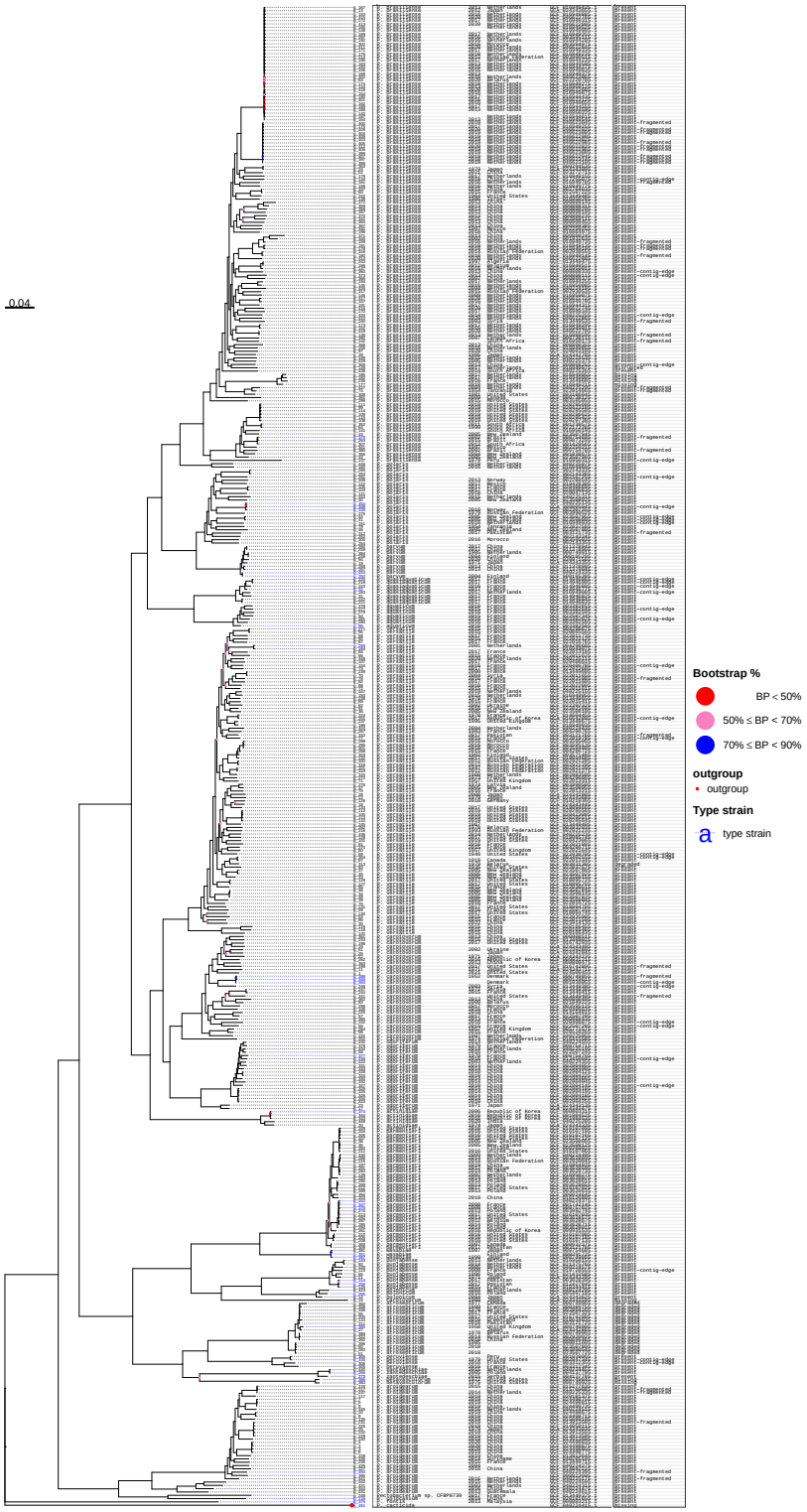
(*1*) Sawada, H., Someya, N., Morohoshi, T., Ono, M., and Satou, M. (2024). Pectobacterium araliae sp. nov., a pathogen causing bacterial soft rot of Japanese angelica tree in Japan. *International Journal of Systematic and Evolutionary Microbiology 74*, 006326.

(*2*) Ben Moussa, H., Bertrand, C., Rochelle-Newall, E., Fiorini, S., Pédron, J., and Barny, M.-A. (2022). The diversity of soft rot Pectobacteriaceae along the Durance River stream in the south-east of France revealed by multiple seasonal surveys. *Phytopathology®*, DOI: 10.1094/PHYTO-12-21-0515-R.

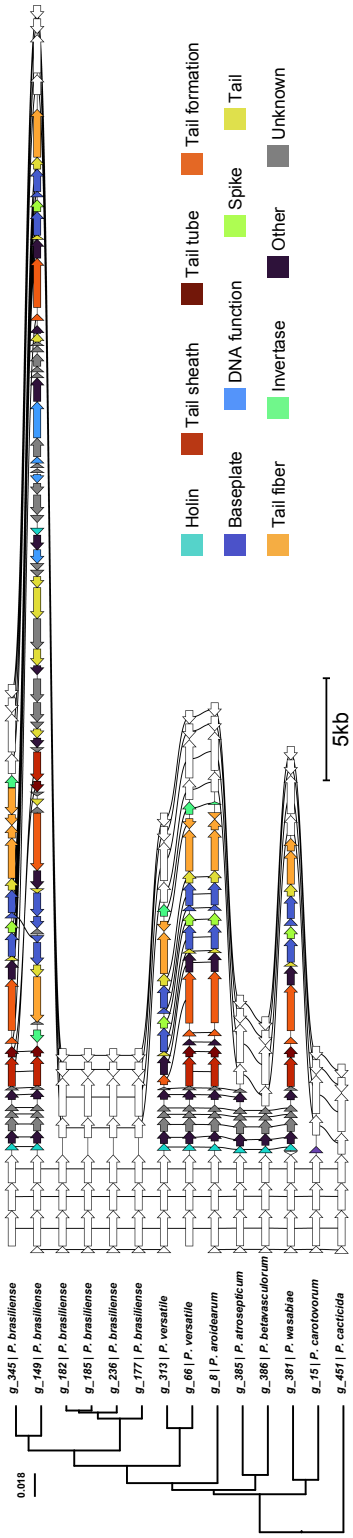
(*3*) Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology 13*, 343–359.

(*4*) Green, E. R., and Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology Spectrum 4*, 4.1.13.

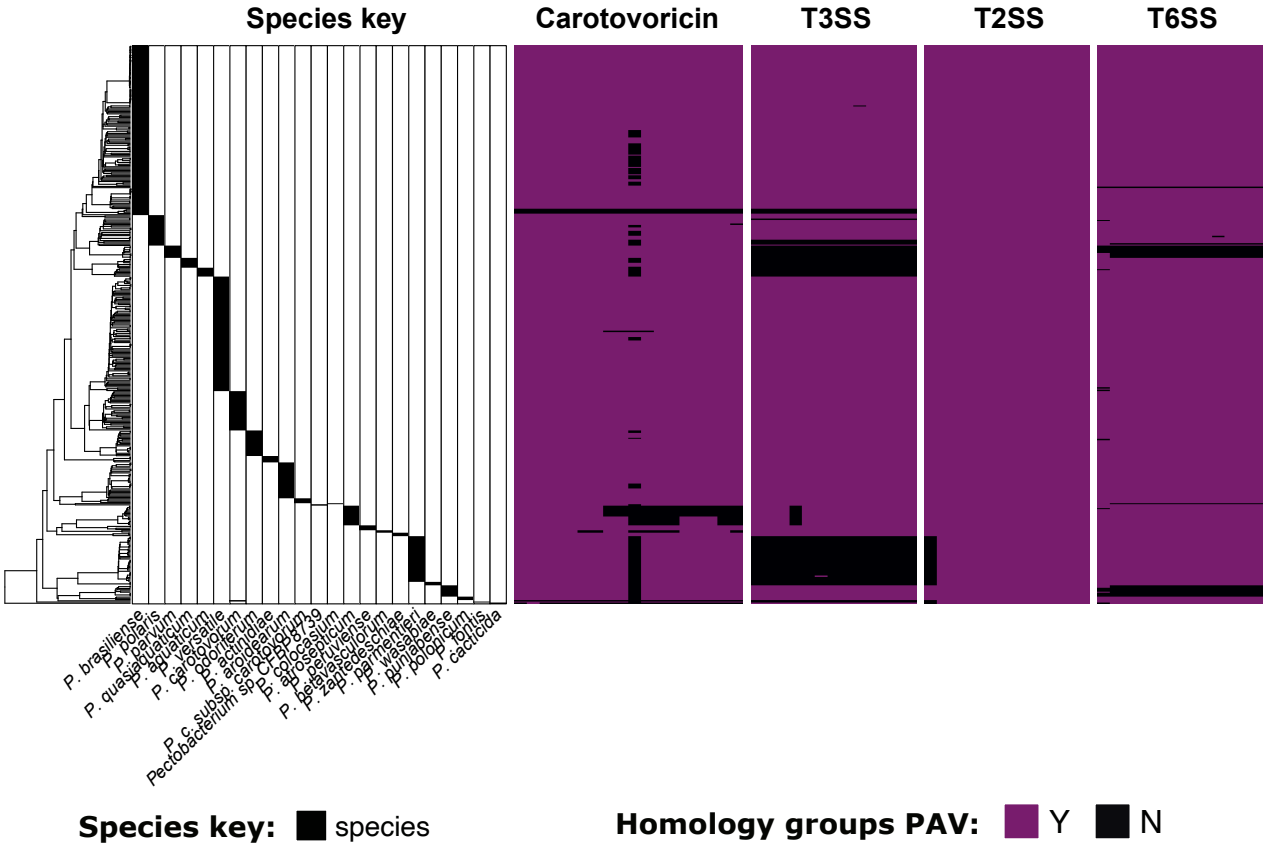
Supplementary Figures



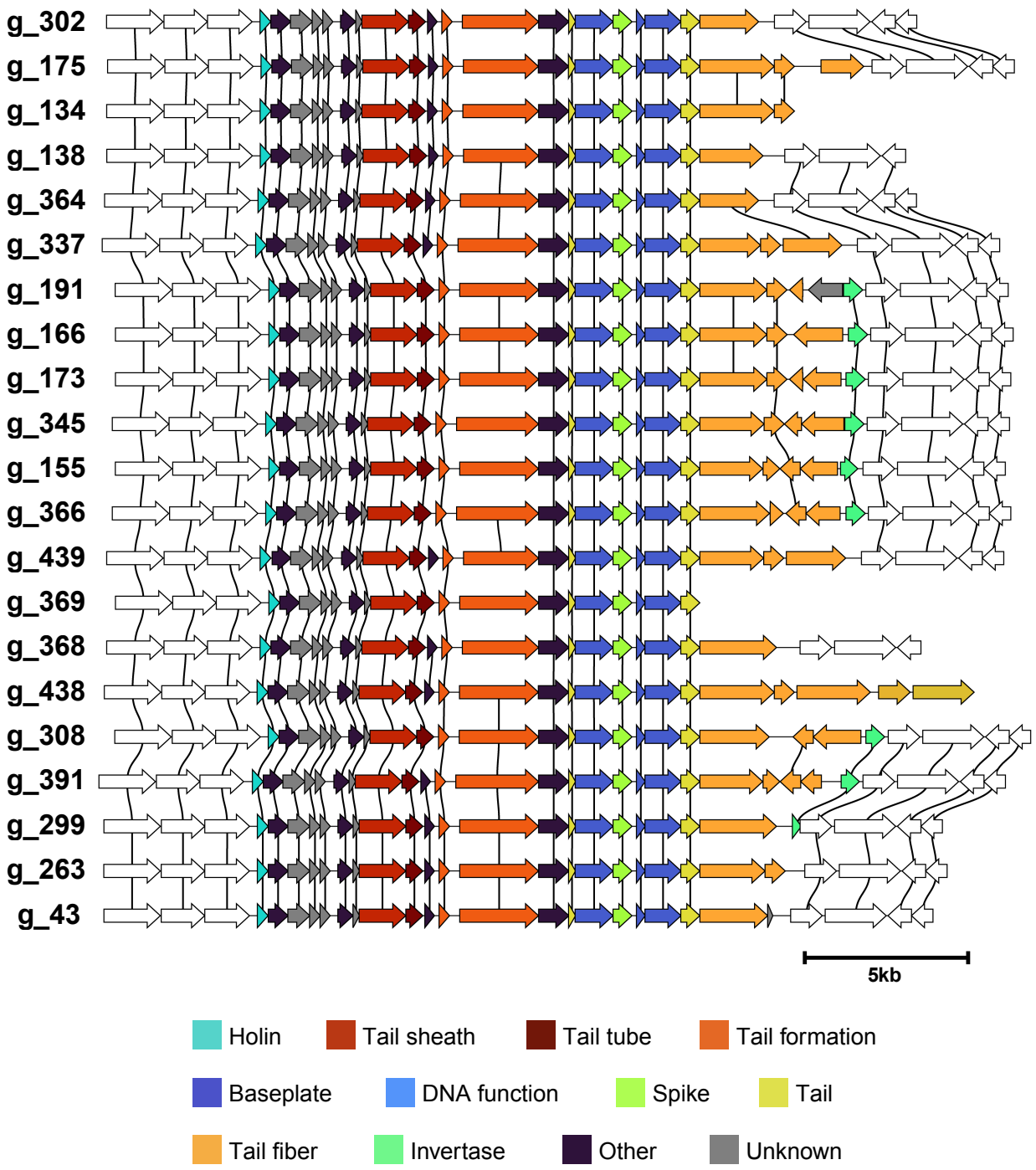
Supplementary Figure S1: *Pectobacterium* genus phylogeny. Core-genome phylogenetic tree for *Pectobacterium* genomes used to construct the pangenome. The tree is rooted using *P. cacticida* as the outgroup. The tip labels include the following data: genome identifier in the pangenome, *Pectobacterium* species name, year of isolation, country of origin, NCBI assembly identifier and carotovoricin presence-absence status.



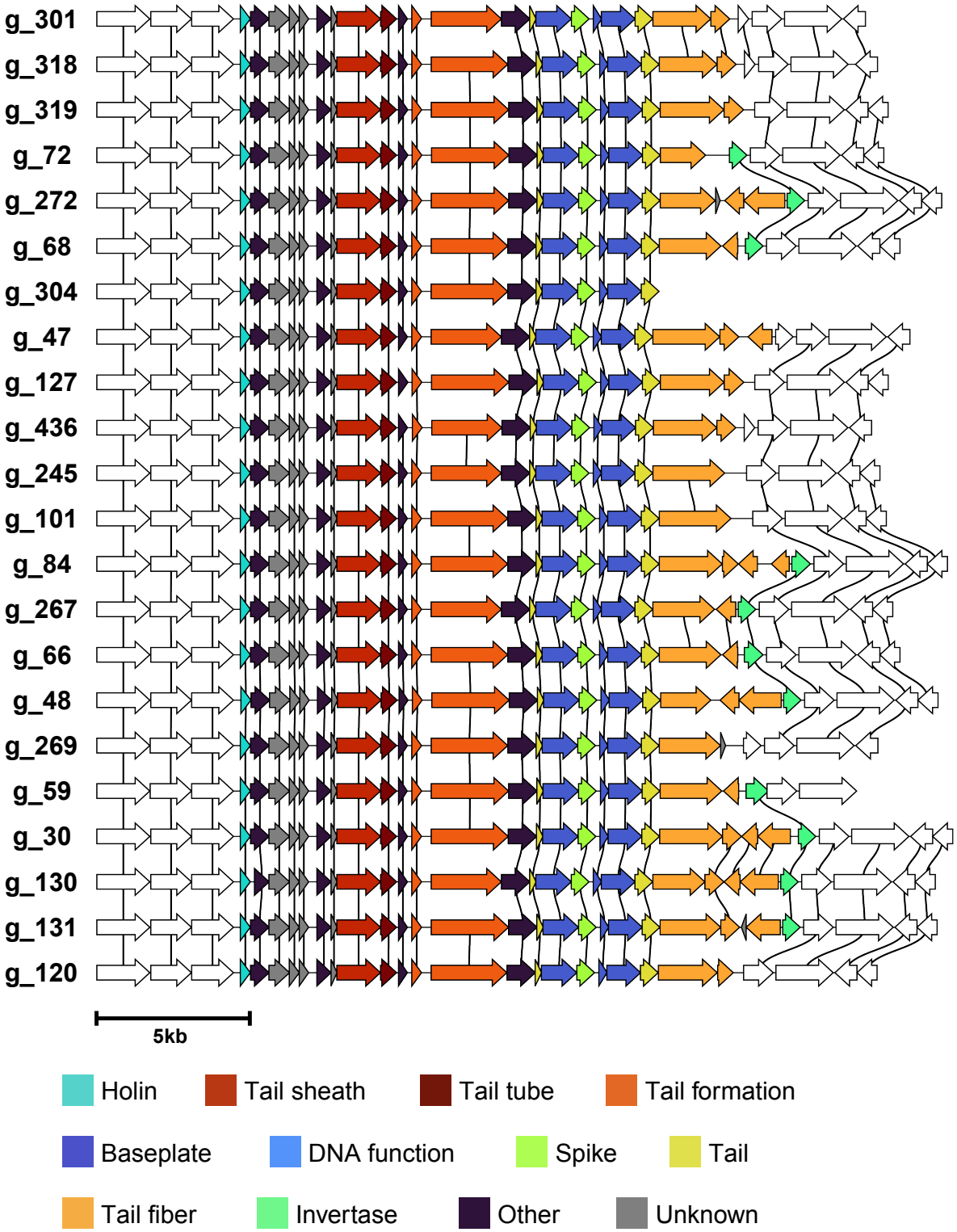
Supplementary Figure S2: Carotovoricin deletion. Genomic regions for strains in which carotovoricin was degraded are shown along with the phylogenetic tree. Genes in 5kb regions flanking carotovoricin clusters are shown in white. Genes belonging to the same homology group are connected with vertical black lines.



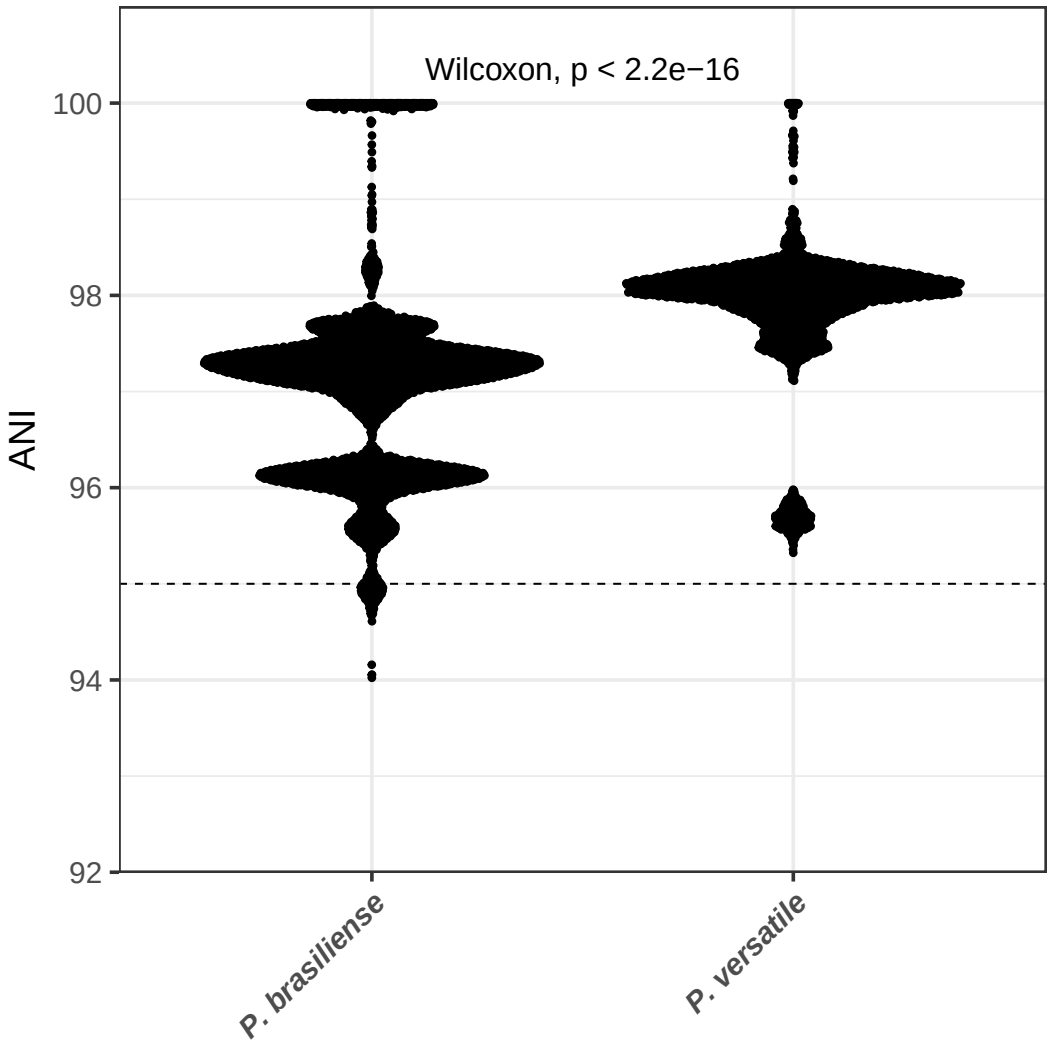
Supplementary Figure S3: Carotovoricin and T3SS deletion. *Pectobacterium* phylogeny for 454 genomes with a species key marking respective species column is shown on the left. The subsequent four heatmaps show core genes of carotovoricin, type 3 secretion system (T3SS), type 2 secretion system (T2SS) and type 6 secretion system (T6SS).



Supplementary Figure S4: Carotovoricin cluster variation in *P. brasiliense* species. Genes in 5kb regions flanking carotovoricin clusters are shown in white. Genes belonging to the same homology group are connected with vertical black lines.



Supplementary Figure S5: Carotovoricin cluster variation in *P. versatile* species. Genes in 5kb regions flanking carotovoricin clusters are shown in white. Genes belonging to the same homology group are connected with vertical black lines.



Supplementary Figure S6: Pairwise intra-species ANI comparison for *P. brasiliense* and *P. versatile* genomes.