Pangenomics to understand prophage dynamics in the *Pectobacterium* genus and the radiating lineages of *P. brasiliense*

Lakhansing A. Pardeshi[1,2,*], Inge van Duivenbode[3], Michiel J. C. Pel[4], Eef M. Jonkheer[1], Anne Kupczok[1], Dick de Ridder[1], Sandra Smit[1], and Theo A. J. van der Lee[2,*]

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

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

[3]Dutch General Inspection Service for agricultural seeds and seed potatoes (NAK), Randweg 14, 8304 AS Emmeloord, The Netherlands

[4]Netherlands Institute for Vectors, Invasive plants and Plant health (NIVIP), National Plant Protection Organization (NPPO), Netherlands Food and Consumer Product Safety Authority (NVWA), Geertjesweg 15, 6706 EA Wageningen, The Netherlands.

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

Abstract

Bacterial pathogens of the genus *Pectobacterium* are responsible for soft rot and blackleg disease in a wide range of crops and have a global impact on food production. The emergence of new lineages and their competitive succession is frequently observed in *Pectobacterium* species, in particular in *P. brasiliense*. With a focus on one such recently emerged *P. brasiliense* lineage in the Netherlands that causes blackleg in potatoes, we studied genome evolution in this genus using a reference-free graph-based pangenome approach. We clustered 1,977,865 proteins from 454 *Pectobacterium* spp. genomes into 30,156 homology groups. The *Pectobacterium* genus pangenome is open and its growth is mainly contributed by the accessory genome. Bacteriophage genes were enriched in the accessory genome and contributed 16% of the pangenome. Blackleg-causing *P. brasiliense* isolates had increased genome size with high levels of prophage integration. To study the diversity and dynamics of these prophages across the pangenome, we developed an approach to trace prophages across genomes using pangenome homology group signatures. We identified lineage-specific as well as generalist bacteriophages infecting *Pectobacterium* species. Our results capture the ongoing dynamics of mobile genetic elements, even in the clonal lineages. The observed lineage-specific prophage dynamics provide mechanistic insights into *Pectobacterium* pangenome growth and contribution to the radiating lineages of *P. brasiliense*.

Keywords: *Pectobacterium*, Pangenome, Prophages

Introduction

In plants, soft-rot disease is mainly caused by bacteria from the *Pectobacterium* and *Dickeya* species, belonging to the *Pectobacteriaceae* family. *Pectobacterium* spp. is among the top 10 plant pathogens and infects potatoes, ornamental plants and vegetable crops (*1*). Although a number of *Pectobacterium* species, such as *P. brasiliense*, *P. atrosepticum*, *P. versatile*, *P. parmentieri*, *P. polaris*, and *P. punjabense* cause soft-rot disease in plants (*2*), *P. brasiliense* is one of the dominant species, known to infect at least 19 different plant species across the world (*3*). The *P. brasiliense* was first detected and characterized as a

pathogen associated with blackleg (BL) disease of potato (*Solanum tuberosum L.*) in Brazil and initially named *Erwinia carotovora subsp. brasiliensis* (*4*). Subsequent molecular, phylogenetic, *in silico* DNA-DNA hybridization and average nucleotide identity (ANI) analysis led to the elevation of this taxon to the species level (*5–8*).

Pectobacterium species are known to co-infect (*9–11*) and show all three inter-species interaction behaviors - competition, cooperation and commensalism (*12*). Until 2013, *Dickeya* spp. and *P. atrosepticum* were dominant in western European countries, like the Netherlands and Switzerland. After the introduction of *P. brasiliense* in western Europe in 2013, its abundance increased rapidly and it is now detected in almost 70-80% of the blackleg diseased plants (*13–15*). Strain specificity towards infecting certain host plants (*3*) and the ability to cause latent infections (*16*) are characteristics of *P. brasiliense*. This species shows regular emergence of new lineages, also evident from the observed high heterogeneity and a broad range intra-species ANI (*17*). Recent pangenome studies on *Pectobacterium* genus have shown a growing pangenome indicating that not all diversity is captured in this genus yet (*17*, *18*).

In prokaryotes, various mechanisms are responsible for the evolution of such heterogeneous strains, such as mutations and horizontal gene transfer (HGT), for example via plasmids and bacteriophages (short phages). Genomes for many species belonging to the *Pectobacterium* genus were assembled at single contig level with minimal evidence of autonomous plasmids (*19–21*) and yet, computational analyses predicted the presence of plasmids in other 89 accessions (*22*). Some phages can integrate into a prokaryotic genome through the lysogenic pathway and are called prophages. A study on a small collection of 54 *Pectobacterium* spp. genomes showed the presence of prophage-like sequences (*23*). These prophages can influence physiological functions and provide fitness advantages to the host bacteria by mechanisms such as superinfection exclusion and harboring virulence factors and antibiotic resistance genes (*24–28*), also noted in *Pectobacterium* pathogens (*29*). From a wider perspective of population expansion, prophages are known to contribute to the competitive success and dispersal of a prokaryotic strain from the environmental reservoir (*30*, *31*). Therefore, it is critical to understand the role of prophages in species and strain diversity, and its contribution to population structure, evolution and epidemiology.

Two distinct groups of *P. brasiliense* strains were detected during annual surveys from 2016 to 2018 in the Netherlands: a clonal lineage with a high ability to cause BL and another, relatively diverse group unable to cause BL in potatoes in field trials (*17*). A subsequent comparative pangenomic analysis enabled the design of new functional assays to detect these BL-causing isolates (*32*). However, a new group of BL-causing isolates that escaped these assays were detected during ongoing annual surveys. A common observation is the emergence of new lineages of *P. brasiliense* in the Netherlands followed by their dominance over the existing lineages. We therefore extended the *Pectobacterium* genus pangenome by combining a genome collection from local epidemiological surveillance with global data to understand the mechanism of appearance and succession of the *P. brasiliense* strains. Further, we investigated the contribution of mobile genetic elements, and prophages in particular, in the growing *Pectobacterium* genus pangenome. Our results showcase an application of the pangenome approach to trace such mobile genetic elements, delineate their dynamics and provide mechanistic insights into pangenome growth.

Results

Emergence of blackleg-causing *P. brasiliense* isolates that evade diagnostics

Yearly inspection, coupled with PCR-based identification of plant pathogens, indicates that after 2013 *P. brasiliense* has emerged as the dominant pathogen in the Netherlands, replacing *Dickeya* spp. In recent years, we have observed a higher prevalence of *P. brasiliense* on blackleg and softrot infected plants, wherein 95% of symptomatic plant material show infection by *P. brasiliense* (Figure 1a). From 2018 to 2020, *P. brasiliense* isolates were sampled from symptomatic and asymptomatic potato (*Solanum tuberosum*) plants in the Netherlands. These isolates were genotyped as BL-causing or BL-non-causing based on diagnostic quantitative PCR assays (*32*). Subsequent field trials showed that some of the presumed non-causing *P. brasiliense* isolates did induce the blackleg phenotype, implying false negative PCR diagnosis. These isolates are henceforth referred to as 'FN-Pbr'.

We also detected multiple *Pectobacterium* species on the same infected plant material during the inspection in the Netherlands. To assess the relation between the newly identified FN-Pbr isolates and the existing BL-causing isolates and to understand potential inter-species interactions, we sequenced 58 *Pectobacterium* spp. isolates collected in the Netherlands. This collection included 28 *P. brasiliense* isolates, 10 of which were FN-Pbr. Genome assemblies from public databases were leveraged to study FN-Pbr evolution at high resolution and further capture the genomic diversity in *Pectobacterium* genus. At the time of analysis, the NCBI database contained 450 *Pectobacterium* spp. genomes. We added 396 high-quality genomes of these after curation (see Methods), resulting in a collection of 454 genomes representing 22 *Pectobacterium* species (Supplementary Table S1, Supplementary Figure S1).

A growing *Pectobacterium* genus pangenome

To study the overall sequence diversity in our collection of genomes, we first performed pairwise ANI comparisons. Within the *P. brasiliense* clade, we observed a broad range of intra-species ANI scores: 9.8% of *P. brasiliense* genome pairs had ANI scores of 96% or less, 1.8% of pairs were with scores of 95% or less. Additionally, a genus level ANI comparison showed at least five distinct *Pectobacterium* species for which inter-species ANIs were closer to 95%, a commonly used species delineation cutoff (*33*, *34*) (Supplementary Figure S2). For example, a clade with *P. versatile*, *P. carotovorum* and *P. odoriferum*; and another clade with *P. polaris* and *P. parvum*. Together, these results demonstrate a high genetic variability in *Pectobacterium* species, in particular for *P. brasiliense* in the Netherlands. Additionally, BL-causing *P. brasiliense* isolates had significantly larger genomes (Wilcoxon-test, $p = 3.6 \times 10^{-7}$) and more protein-coding genes ($p = 1.1 \times 10^{-5}$) than non-causing ones (Figure 1b). This was not due to differences in assembly quality, as there was no significant difference ($p = 0.076$) in median assembly N50 values between the two groups and all genomes had BUSCO completeness $\geq 99\%$.

For a reference-free comparison of this heterogeneous genus, we constructed a genus-level pangenome of 454 *Pectobacterium* spp. genomes using PanTools (*35*). In our pangenome 1,977,865 protein-coding genes were clustered into 30,156 homology groups, which were further categorized into 1,949 core (present in all

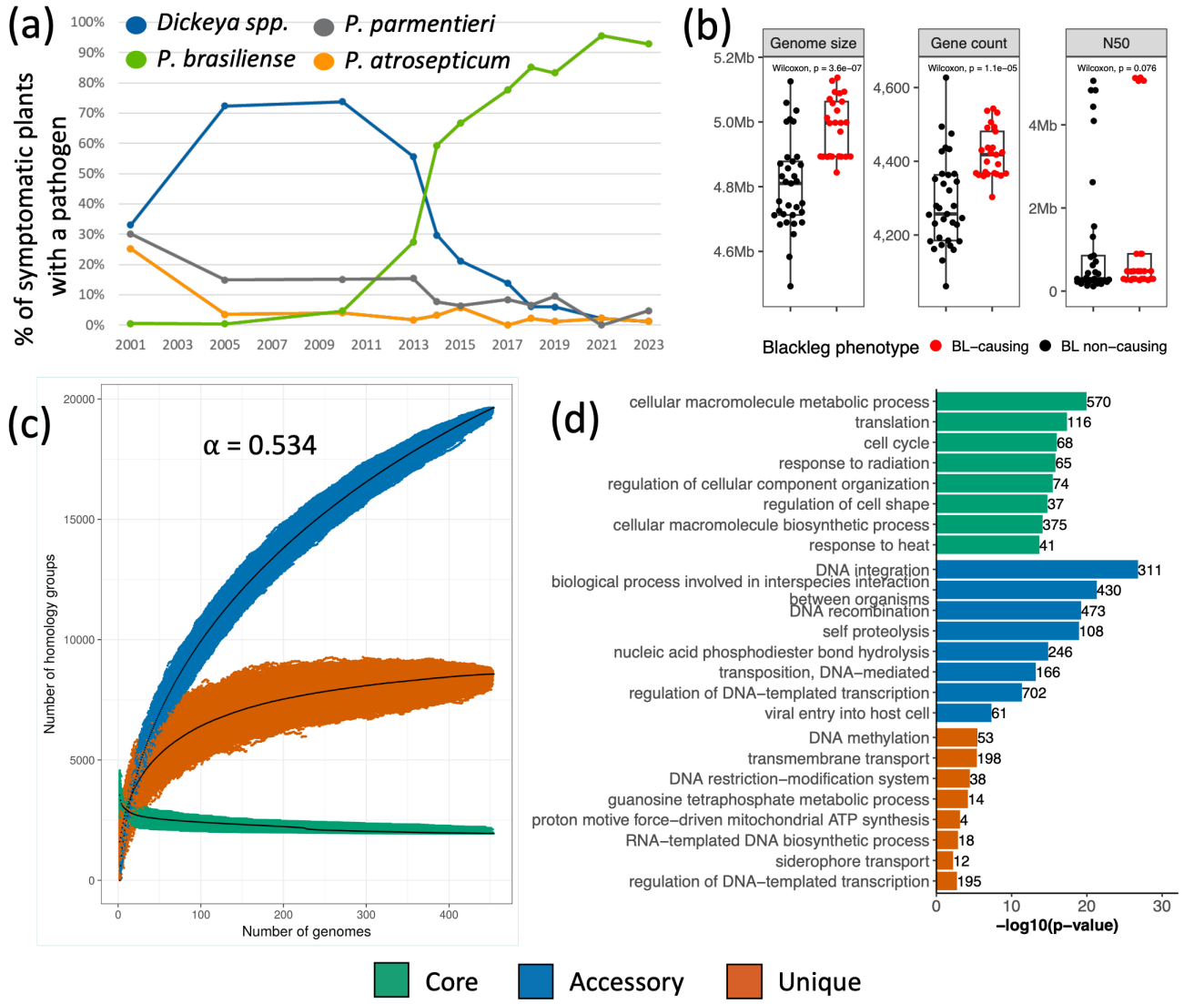


Figure 1: (a) Percentage of different pathogens identified from the plants showing BL symptoms during the annual inspection in the Netherlands. (b) In the Dutch collection, BL-causing isolates have larger genomes and more protein-coding genes than BL-negative isolates. (c) The *Pectobacterium* pangenome has a stable core but increasing accessory genome. Heaps' law $alpha = 0.534$ is < 1 and implies an open pangenome. (d) Functional enrichment of Gene Ontology biological processes in the core, accessory and unique pangenome.

genomes), 8,571 unique (exclusive to individual genomes) and 19,642 accessory groups. Fitting a Heaps' law model to homology group counts as a function of newly added genomes provides an estimate of the pangenome openness, where a decay rate $\alpha < 1$ implies an open pangenome and $\alpha > 1$ a closed one (*36*). In line with the earlier, smaller pangenomes (*17*, *18*), the *Pectobacterium* spp. pangenome is still open, with a decay rate α of 0.534 (Figure 1c). In the species-level sub-pangenomes extracted from the genus-level pangenome, *P. versatile* and *P. brasiliense* are represented by significantly more genomes than others but surprisingly had the lowest decay rates of 0.501 and 0.538 respectively (Supplementary Figure S3, Supplementary Table S2). After more than doubling the number of genomes in the genus-level pangenome compared to the previous version, the core remained stable with a small decrease of 83 homology groups (Supplementary Figure S4). The increase in the overall pangenome size by 7,815 homology groups was mainly attributed to the accessory category, with 6,474 additional homology groups. This increase was mostly contributed by *P. brasiliense* and *P. versatile*, with 7,290 and 6,384 accessory homology groups respectively (approx. 50% of the homology groups in the respective sub-pangenomes) (Supplementary Table S2).

A new blackleg-causing *P. brasiliense* lineage

A maximum likelihood phylogenetic tree based on the alignment of 1,949 core genes revealed that all 25 BL-causing *P. brasiliense* isolates were part of a monophyletic clade of 48 isolates (Figure 2). This BL-causing clade was split into two sub-clades: one of 35 isolates enriched for TP-Pbr, the other of 13 isolates enriched for FN-Pbr. Intra-clade pairwise comparison of the isolates within these sub-clades showed 100% ANI. Based on the core phylogeny and ANI comparison, it is evident that FN-Pbr and TP-Pbr represent clonal lineages radiating from a BL-causing *P. brasiliense* common ancestor (Supplementary Figure S2). Only one strain (NAK399) from the FN-Pbr sub-clade did not cause any BL phenotype in the field trials (0.5% infected plants).

To gain broad functional insights of our pangenome, we performed a gene ontology enrichment analysis on core, accessory and unique gene sets. The *Pectobacterium* core genome was enriched for elementary processes such as cell cycle, translation, heat response, macromolecule metabolism, ribosome assembly, etc. The unique genome was enriched in DNA methylation, siderophore transport and lipid transmembrane transport. The accessory genome was enriched for DNA integration, recombination, transposition, viral entry into host, inter-species interaction, etc. (Figure 1d, Supplementary Table S3).

Contribution of prophage-like elements in the *Pectobacterium* pangenome

Intrigued by our observation of the heterogeneity in *P. brasiliense* and enrichment of biological processes related to mobile genetic elements, we hypothesized such elements as an underlying factor in pangenome growth. Therefore, we systematically analyzed mobile genetic elements in *Pectobacterium* genomes using the geNomad pipeline (*37*) (see Methods for details). A total of 1,369 prophage-like regions were identified in the 454 genomes, where all but one genome contained at least one prophage. Cumulatively, genes from prophages accounted for 4,801 (15.9%) of the pangenome homology groups. Among these prophage

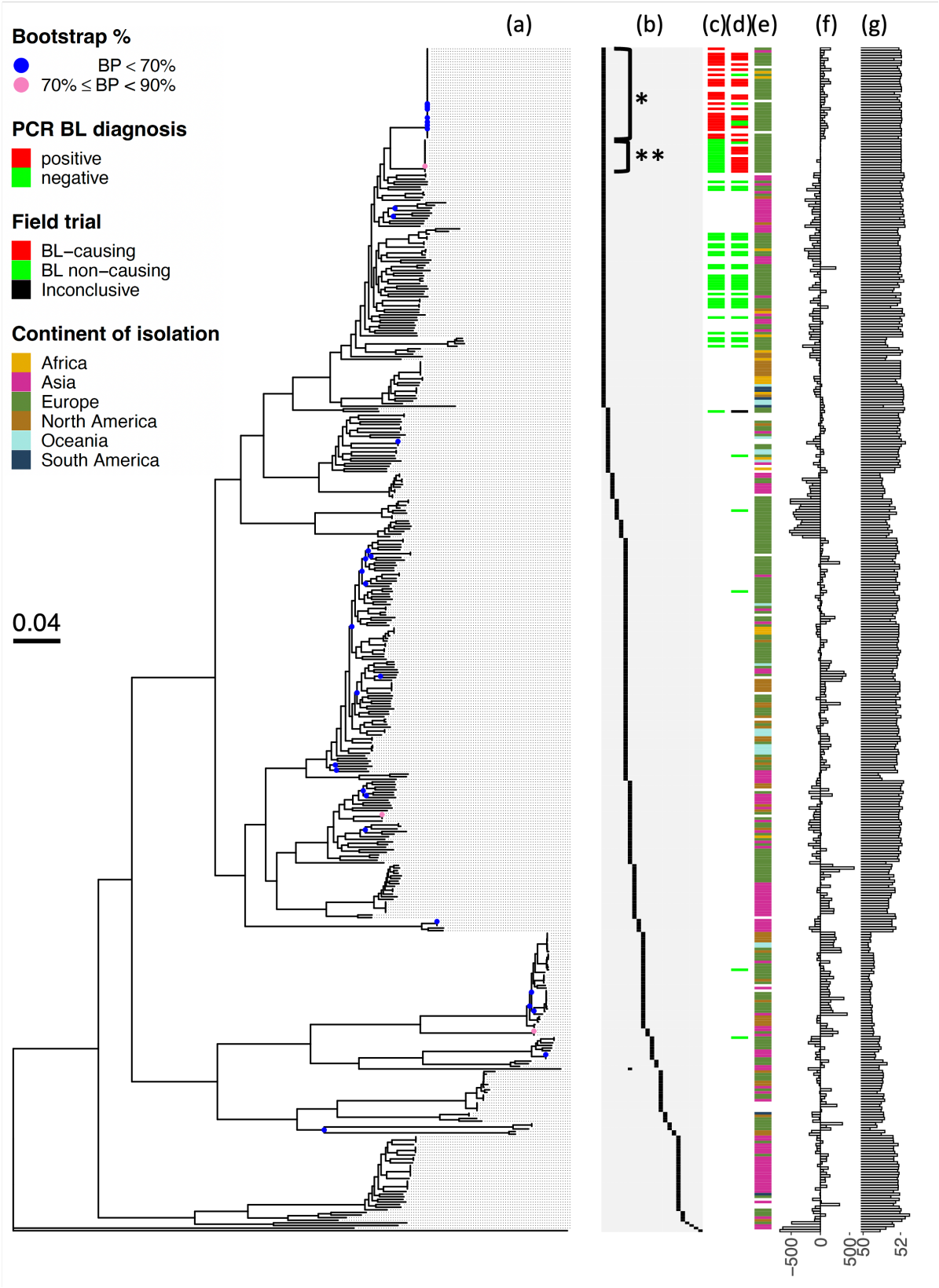

Bootstrap %
BP < 70%
70% ≤ BP < 90%
PCR BL diagnosis
positive
negative
Field trial
BL−causing
BL non−causing
Inconclusive
Continent of isolation
Africa
Asia
Europe
North America
Oceania
South America
0.04
(a)
(b)
(c)(d)(e)
(f)
(g)
*
**
−500
0
500
50
52

Figure 2

Figure 2: (a) A robust maximum likelihood tree generated using 1,949 core gene alignments. Nodes with bootstrap values $< 90\%$ and $< 70\%$ are colored pink and blue respectively. Metadata (from left to right): a species key (b), where each column denotes a species ordered from left to right: *P. brasiliense*, *P. polaris*, *P. parvum*, *P. quasiaquaticum*, *P. aquaticum*, *P. versatile*, *P. carotovorum*, *P. odoriferum*, *P. actinidiae*, *P. parmentieri*, *P. wasabiae*, *P. punjabense*, *P. polonicum*, *P. atrosepticum*, *P. peruviense*, *P. zantedeschiae*, *P. betavasculorum*, *P. aroidearum*, *P. carotovorum subsp. carotovorum*, *Pectobacterium* sp. CFBP8739, *P. colocasium*, *P. fontis* and *P. cacticida*. *: Clade enriched for TP-Pbr isolates. **: Clade enriched for FN-Pbr isolates. PCR based diagnosis of BL-causing ability of the isolates (c), BL phenotype during field trials (d), continent on which the isolate was collected (e), difference in number of protein-coding genes with respect to the median of 4,363 genes across *Pectobacterium* genus (f), and GC% (lower limit of the bar plot shown is 50%) (g).

homology groups, 3,286 (16.7%) were accessory, 1,477 (17.2%) unique and only 38 belonged to the core genes in *Pectobacterium* genus pangenome (Figure 3a). Similarly, predicted plasmids were represented by 10,544 homology groups. A comparison of clonal isolate genomes ($ANI > 99.9\%$) from the TP-Pbr clade ($n = 35$) showed that geNomad predicted plasmids in fragmented genome assemblies ($n = 26$) but not in the single chromosome assemblies ($n = 9$). Furthermore, the plasmid homology groups from the isolates with fragmented genome assemblies could be mapped in tandem onto the chromosome-level assembly, suggesting a false-positive identification. Alternatively, these can be integrated plasmids in the single chromosome genomes, also known as episomes; this is supported by the consistent observation of contigs of length 293kb and 89kb as putative plasmids in the fragmented genome assemblies. Nevertheless, this renders the plasmid prediction results unreliable.

We used CheckV (*38*) to assess the completeness of the putative prophages and found 731 prophage-like elements with a completeness score of 50% or lower and 423 prophages with 90% or higher completeness (Figure 3b, Supplementary Table S5). The shortest prophage with 95% or more completeness was 5.4kb long. The prophage length distribution shows three distinct peaks, with the highest peak at 16kb mostly representing incomplete prophages; a peak at 35kb representing prophages with completeness $> 90\%$; and a peak at 50kb comprising complete prophages. At the species level, *P. brasiliense* and *P. versatile* prophages contributed to $\sim 14\%$ of their respective pangenomes (Supplementary Table S4). This distribution is similar to the prophage fraction in the accessory genome of other highly dynamic bacterial pangenomes like *E. coli* and *S. enterica* (*39*). Since the BL-causing *P. brasiliense* isolates have larger genomes, we compared the total prophage length between the two groups for a potential explanation. We observed that the BL-causing isolates had significantly larger median total prophage length (t-test, $p = 1.1 \times 10^{-4}$). This 50kb difference observed in total prophage length accounts for 25% of the observed 200kb disparity in median genome sizes between these two categories (Supplementary Figure S5).

The tailed bacteriophage class *Caudoviricetes*, with a double stranded DNA genome, was represented most abundantly ($n = 1,357$) in the *Pectobacterium* pangenome. Although much less abundant, other classes of viruses included filamentous phages of the *Inoviridae* ($n = 8$) and *Microviridae* ($n = 2$) families, a virus from *Patatavirales* order ($n = 1$) and a 5.2kb unclassified prophage (Supplementary Figure S6). *Inoviridae*

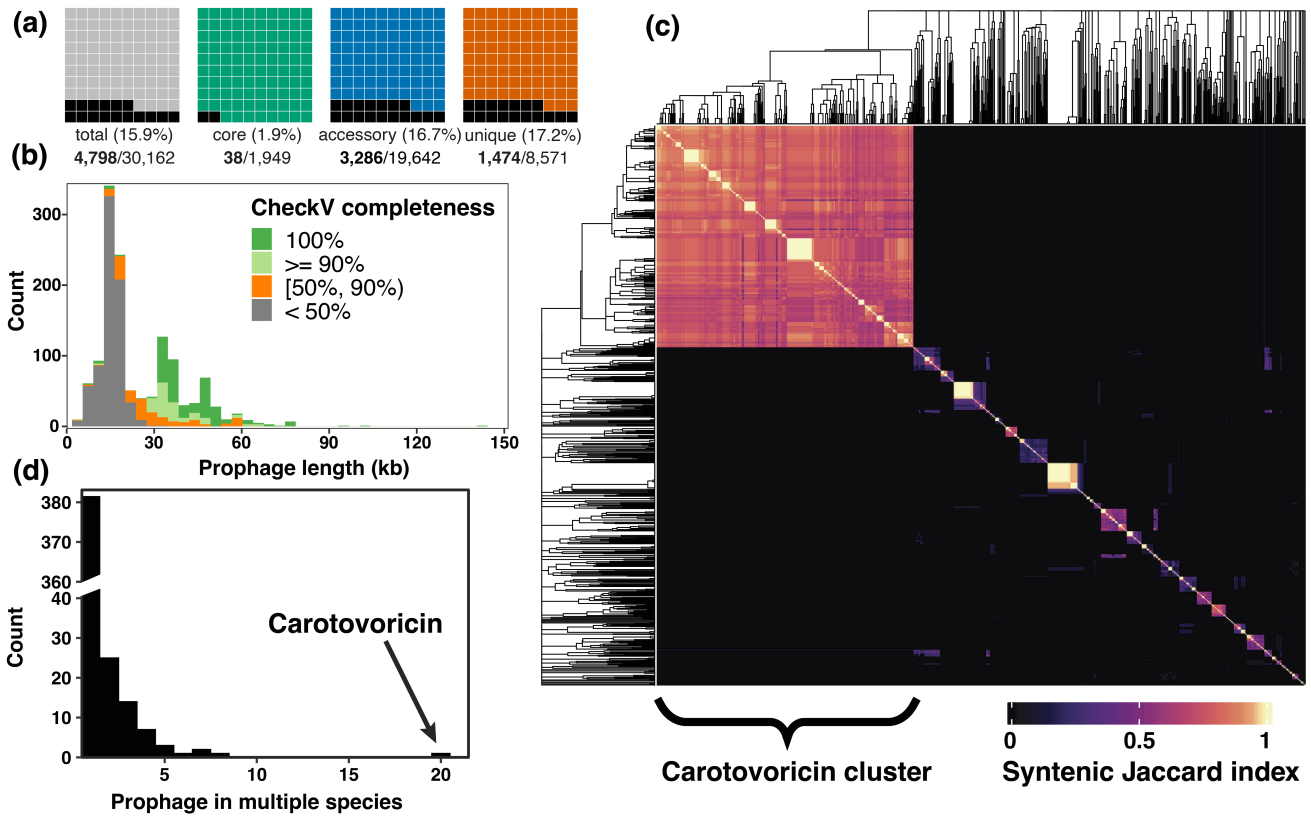


Figure 3: Prophage analysis. (a) Fraction of pangenome homology groups corresponding to prophages. (b) Prophage length distribution. (c) Clustering based on homology group content, represented by a heatmap of syntenic Jaccard index values. The largest cluster of carotovoricin (CTV) is annotated in the heatmap. Complete linkage dendrogram was cut at the node height of CTV cluster to identify groups of orthologous prophages. (d) A histogram showing prophage backbone present in one or multiple *Pectobacterium* species. The CTV, a prophage-like element present in 20 species is marked with an arrow.

and *Microviridae* are non-enveloped single-stranded DNA viruses, whereas *Patatavirales* is the largest order of non-enveloped positive-strand RNA viruses of plants and contains a single family *Potyviridae*, accounting for 30% of all known plant viruses (*40*). While it was interesting to identify a plant virus of the order *Patatavirales* in the *P. parvum* genome assembly GCF_011378945.1, caution must be taken, as phages detected on independent contigs are difficult to ascertain as true integrations. Among the less abundant phages, host DNA integration evidence was observed for *Inoviridae* family prophages (Supplementary Table S5), thus these can be considered true prophages. The *Microviridae* and *Patatavirales* prophages were detected on an independent contig without flanking host genes and are likely the result of lytic viruses or sample contamination.

Generalist prophages of *Pectobacterium* species

Bacteriophages have high genetic diversity and mosaicism is a hallmark of their genomes, primarily driven by horizontal gene transfer (*41–44*). Given the variation in prophage length and the dominance of a single bacteriophage class, finding orthologous prophages would help in the characterization of prophage diversity

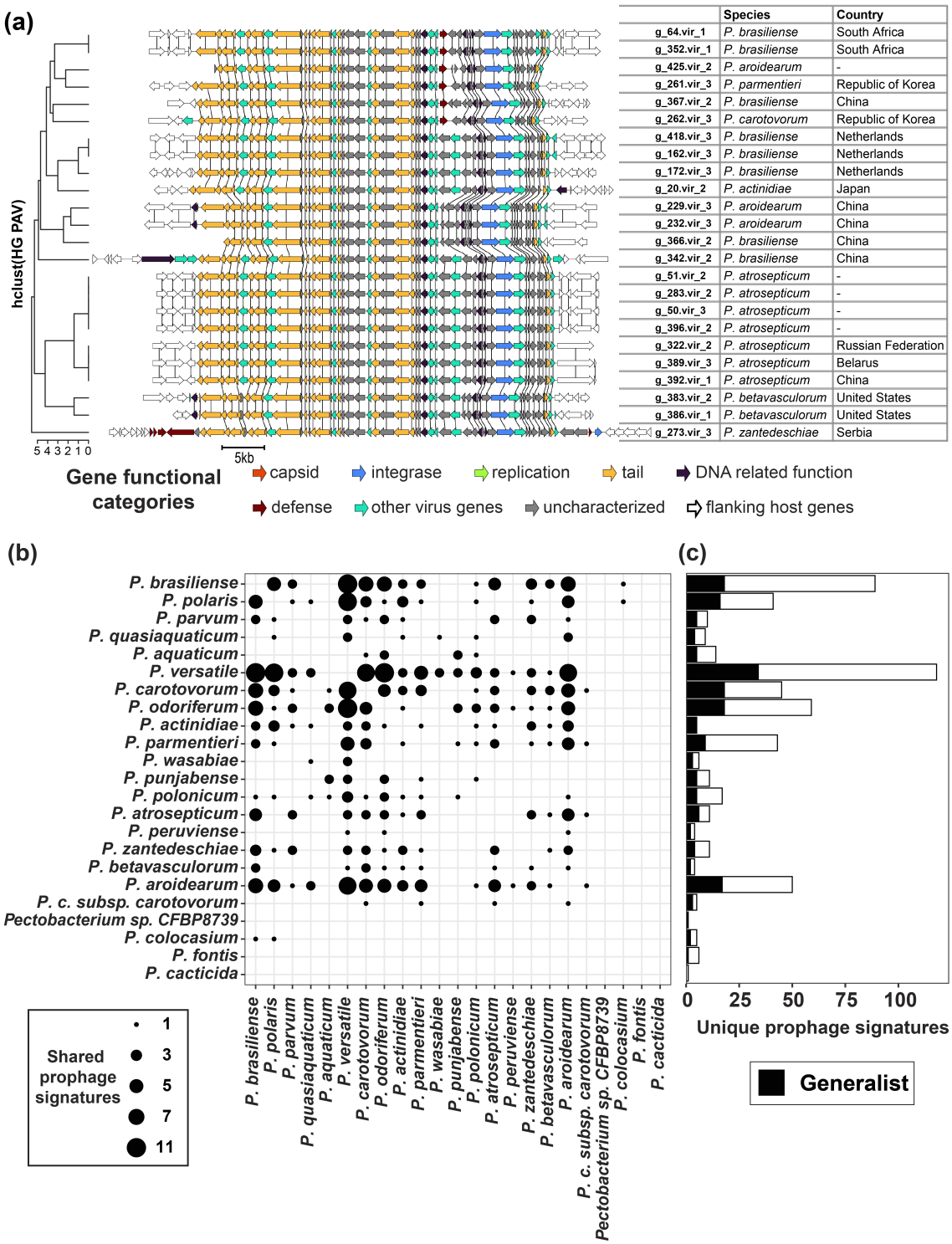
and dynamics in the *Pectobacterium* pangenome. Therefore, we compared putative prophages by calculating syntenic Jaccard indices of the prophage homology group signatures. CTV, a conserved 16kb phage tail-like bacteriocin occurring in 428 genomes was used to decide the prophage grouping threshold (Figure 3c) (see Methods). As a result, we identified 436 clusters of orthologous prophages (named phage_grp_1 to phage_grp_436) and their respective cluster representative (Supplementary Figure S7, Supplementary Table S6). These representative prophages covered a total of 4,493 homology groups (94% of the total 4,801), implying minimal loss of information during clustering.

Most phages are species-specific and even strain/serovar/pathovar specific (*45–47*), though broad range phages targeting different species also exist (*48*, *49*). To identify such generalist phages infecting multiple *Pectobacterium* species, we checked prophage clusters for the presence of multiple species. We identified 54 generalists and 382 species specialists (Figure 3d). For example, clusters phage_grp_71 (Figure 4a) and phage_grp_36 with 24 and 21 prophages respectively, included complete prophages targeting multiple *Pectobacterium* species and their hosts were collected from a wide spatiotemporal context (Supplementary Table S6). No association was found between the *Pectobacterium* isolates harboring phage_grp_71 phages and plant infection, as isolates were collected from both symptomatic and asymptomatic plants. Furthermore, none of the isolates harboring phage_grp_36 phages were sampled from infected plants. *P. versatile* was the most frequent host of the generalist prophages, targeted by 34 out of 54 prophages (Figure 4b-c). The diverse ecological niches of *P. versatile* isolates, spanning water, soil and infected plants, likely contributed to their susceptibility to generalist phages. In contrast, the closely related species *P. aquaticum* and *P. quasiaquaticum* collected from freshwater bodies in France did not share any prophages, yet did share prophage signatures with other *Pectobacterium* species.

Prophage dynamics in the BL-causing clonal lineages

We further explored the prophages of isolates in the BL-causing clade, specifically to understand the intra-lineage variation and horizontal transfer with other *Pectobacterium* species. We detected four clusters with intact prophages exclusive to the BL-causing clade (Figure 5a). First, a 49kb intact prophage was found in all but two isolates from this clade (phage_grp_45 in Supplementary Table S6). This prophage carried a YdaS/YdaT toxin-antitoxin (TA) system and a DNA methyltransferase gene, important elements of bacterial defense systems (Figure 5b). The same flanking genes are found across all genomes, which suggests a single insertion event of this prophage in the common ancestor of TP-Pbr and FN-Pbr isolates followed by vertical transmission.

A single FN-Pbr clade isolate sampled in 2019 in the Netherlands showed an insertion of a 50kb prophage (phage_grp_343 in Supplementary Table S6). Interestingly, this prophage had a very similar homology group signature to phage_grp_45 prophages, with some rearrangements, and was integrated at the same site in the host genome (Figure 5b), implying a mosaic relationship. These prophages could potentially be clustered together when a relaxed threshold is applied. Nevertheless, they are distinct in gene content and structure and our approach effectively captures phage mosaicism and dynamics across multiple genomes. Another 44kb prophage was detected in all of the FN-Pbr isolates (phage_grp_107 in Supplementary Table



	Species	Country
g_64.vir_1	*P. brasiliense*	South Africa
g_352.vir_1	*P. brasiliense*	South Africa
g_425.vir_2	*P. aroidearum*	-
g_261.vir_3	*P. parmentieri*	Republic of Korea
g_367.vir_2	*P. brasiliense*	China
g_262.vir_3	*P. carotovorum*	Republic of Korea
g_418.vir_3	*P. brasiliense*	Netherlands
g_162.vir_3	*P. brasiliense*	Netherlands
g_172.vir_3	*P. brasiliense*	Netherlands
g_20.vir_2	*P. actinidiae*	Japan
g_229.vir_3	*P. aroidearum*	China
g_232.vir_3	*P. aroidearum*	China
g_366.vir_2	*P. brasiliense*	China
g_342.vir_2	*P. brasiliense*	China
g_51.vir_2	*P. atrosepticum*	-
g_283.vir_2	*P. atrosepticum*	-
g_50.vir_3	*P. atrosepticum*	-
g_396.vir_2	*P. atrosepticum*	-
g_322.vir_2	*P. atrosepticum*	Russian Federation
g_389.vir_3	*P. atrosepticum*	Belarus
g_392.vir_1	*P. atrosepticum*	China
g_383.vir_2	*P. betavasculorum*	United States
g_386.vir_1	*P. betavasculorum*	United States
g_273.vir_3	*P. zantedeschiae*	Serbia

Figure 4: (a) Prophage signatures of 22 members of the phage_grp_71 cluster. Flanking genes are colored white and the genes belonging to the same homology groups are connected with vertical lines. (b) Generalist prophages shown as the overlap of prophage signatures between *Pectobacterium* species pairs. (c) Number of unique prophage signatures per species. The fraction of generalist prophages is indicated in black.

S6, Figure 5c). In fragmented genome assemblies of seven FN-Pbr isolates, this prophage was split over two or more contigs. However, our pangenomic homology group signature matching allowed merging these fragments into a single prophage signature, demonstrating the use of a pangenomic approach in studying mobile genetic element dynamics. Interestingly, this prophage harbored a methyltransferase and antitoxin gene of the TA-system, although the toxin component was missing.

Finally, a 31kb prophage was detected exclusively in 30 TP-Pbr isolates collected between 2012 and 2020, mainly in the Netherlands (phage_grp_46 in Supplementary Table S6). The five *P. brasiliense* isolates in this clade lacking this prophage were sampled in 2016 or later in Belarus and the Netherlands (Figure 5a,d). Prophage signature comparison shows a remnant region in FN-Pbr genomes (g_406). Therefore, an insertion in the common ancestor of TP-Pbr and FN-Pbr clades and two independent excision events, first in the FN-Pbr clade and another in the five isolates of TP-Pbr clades likely explains these results. In summary, we captured ongoing prophage and immune system dynamics in the clonal isolates based on the robust core phylogeny.

Discussion

During annual sampling of soft-rot *Pectobacteriaceae* in the Netherlands, we detected BL-causing *P. brasiliense* isolates that escaped the previously designed BL-phenotype assays (*32*). To provide a broader context for the emergence of this isolate, we applied a pangenome approach to the *Pectobacterium* genus. Compared to the previous pangenome, we observed a stable core but an almost 50% increase in the accessory genome. Prophages contributed significantly to the *Pectobacterium* genus pangenome, mostly non-core genes. Further characterization of these prophages revealed the presence of both generalist and specialist prophages of *Pectobacterium* species.

The genetic diversity of *P. brasiliense* is evident from the broad range of intra-species ANI scores (94-99%). Pathogens can evade diagnosis by natural selection of isolates that acquire mutations in assay target genes. Diagnostic assay escape is not new for soft-rot pathogens. Previously, the Dia-A primer developed to detect *D. dianthicola* failed to detect a *Dickeya* outbreak in the USA because of target gene deletion (*2*). In this and our case, the newly evolved isolates were very similar to those for which assays were developed (for *P. brasiliense*, 97.7% ANI). This highlights the challenges in designing stable phenotype diagnosis assays for soft-rot pathogens like *P. brasiliense* with high genomic diversity and an open pangenome. To address this, we propose monitoring as described in this study and periodic pangenome updates.

Bacterial species show a wide range of prophage dissemination, from 50% of the genomes in actinobacterial species harboring one or more prophage-like elements (*50*) to 99.5% of *Acinetobacter baumanii* genomes (*51*). Higher dissemination of prophages is known to drive the plasticity of bacterial genomes and contributes to pathogenicity and virulence (*51*). In our collection, including *Pectobacterium* spp. isolates from all over the world, 99.7% of the genomes had one or more prophage-like elements, implying high genome plasticity at both genus and species levels. We further showed that prophages account for $\sim$ 16% of the *Pectobacterium* genus pangenome, resulting in an open pangenome. We observed a significant difference

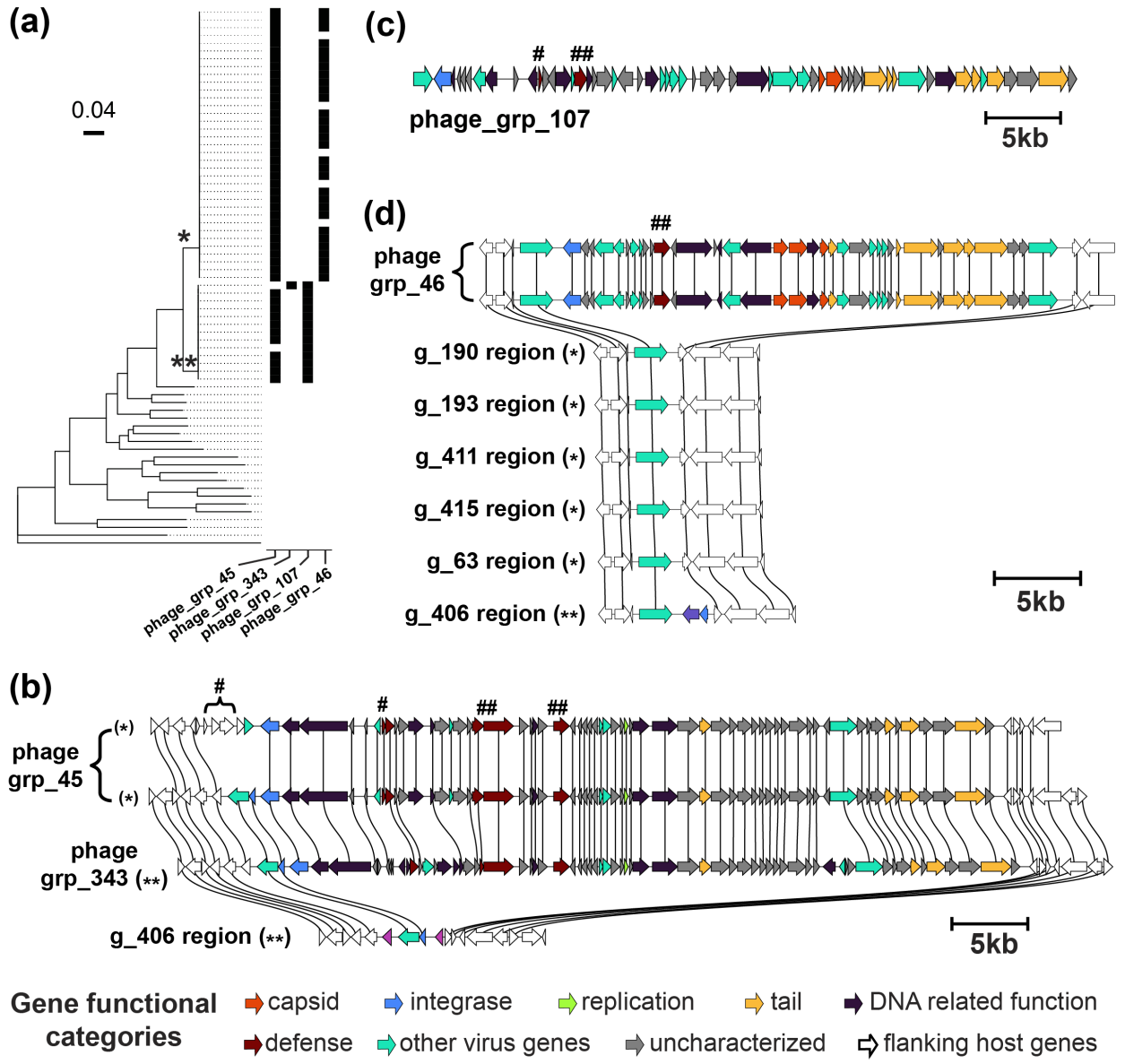


Figure 5: (a) Presence-absence (black-white) of the four prophages exclusive to BL-causing *P. brasiliense* lineage are shown next to a subtree for this clade. *: TP-Pbr clade, **: FN-Pbr clade. The remaining nodes in the tree belong to one representative *Pectobacterium* species. (b) BL-causing clade-specific prophage homology group signatures: four regions showing two representative prophages from the orthologous phage group phage_grp_45, a single prophage from genome g_400 belonging to phage_grp_343 group, and a region from genome g_406 in which the prophage is absent. (c) FN-Pbr clade-specific prophage homology group signature: a representative from the orthologous phage group phage_grp_107. (d) TP-Pbr clade-specific prophage homology group signatures comparison: two representative prophages from phage_grp_46 group and regions from the five TP-Pbr genomes and the single FN-Pbr genome in which this prophage is absent are shown. For all figures, flanking genes are colored white and genes in the same homology groups are connected with vertical lines. Bacterial defense systems are marked as #: TA-system and ##: methyltransferase. BL diagnosis is marked as *: TP-Pbr, **: FN-Pbr.

in mean genome size between the BL-causing and non-causing *P. brasiliense* isolates. Prophages explain a quarter of this observed genome size difference. Bacteria and their phage predators are locked in an arms race and it is possible that to counter these phages, prey bacteria are under selection pressure to acquire additional mobile genetic elements such as plasmids, integrative conjugating elements and insertion sequences. We did not find convincing evidence for the presence of plasmids in *Pectobacterium* species as we observed a discrepancy in plasmid prediction on high quality versus fragmented genome assemblies. It remains to be seen whether other mobile genetic elements exist and how they interplay with the prophages in *Pectobacterium* species.

As observed in *P. brasiliense* lineages from our collection, lineage-specific prophages are also detected in other pathogens, for example $\Phi RS551$ in *Ralstonia solanacearum* (*52*) and $\Phi RE2010$ in *Salmonella enterica* (*53*). Genes in such prophages are known to be transcriptionally active and up-regulated post-infection in different plant pathogens, for example in *P. brasiliense* (*54*) and *R. solanacearum* (*55*, *56*). These prophages modulate the mobility and virulence of bacterial pathogens and may provide a competitive advantage over other closely related species not bearing prophages (*29*, *54*). Similar examples of horizontally acquired mobile elements modulating virulence and quorum sensing are also reported for the *Pectobacterium* species (*29*, *57*, *58*). Moreover, annual sampling across the Netherlands enabled us to capture ongoing gain-loss dynamics of prophages, even within clonal lineages, and therefore highlights the contribution of prophages to genome plasticity. These lineage-specific prophages in *P. brasiliense* probably provide a competitive advantage over other closely related species and the active dynamics of prophages enable such pathogens to continuously radiate novel lineages as observed.

We also showed the existence of generalist prophages targeting multiple *Pectobacterium* species. Such prophage identification across species validates host recognition by the phage, an important characteristic for determining phage host range. Phage-host interaction networks often show modularity (*49*, *59*). Taxonomic separation of hosts due to structural and metabolic differences is the intrinsic driver of such modular host ranges (*45*), emphasizing the need to include the host phylogeny while analyzing phage host ranges. A pangenome approach to trace prophages shown here allows such paired study of host range and host phylogeny estimated from the core genome. These phage host range studies have important applications such as phage typing, controlling microbial community composition and phage therapy.

Biocontrol of plant pathogens using phages provides an environmentally friendly alternative to antibiotics (*60*). Although lytic phages are more suitable phage therapy candidates, studying the dynamics of lysogenic phages will enable learning of the host-detection mechanisms employed by phages in general. Recently, several studies have shown the effectiveness of this for *Pectobacterium* species and other plant pathogens (*61–67*). However, just as antibiotic resistance, the development of phage resistance in bacterial hosts (*68*) is a major bottleneck in phage therapy applications. Prophages also carry various defense systems, such as TA-systems, restriction-modification system and CRISPR (*69*, *70*), that can provide immunity to the host bacteria against both lytic and lysogenic phages (*71*). All of the BL-causing *P. brasiliense* exclusive prophages harbored at least one defense system. Therefore, a comprehensive characterization of prophage diversity and its interplay with bacterial defense systems can pave the way for smarter development of

phage therapy applications. Additionally, a broader pangenome study that includes pathogens sharing niches with *Pectobacterium* species could help to discover prophages with cross-genus range.

However, such large-scale comparative analysis to study prophage dynamics comes with several challenges, such as reference bias and prophage overestimation because of fragmented assemblies (*72*). Our approach overcomes these by combining prophage identification with the graph pangenome approach to merge fragmented prophages and correctly estimate their abundance in pangenome. We further used homology groups to identify orthologous prophages and constructed a non-redundant prophage set for *Pectobacterium* species. Thus, our approach provides a novel and efficient strategy to screen and select bacteriophage candidates, also for developing phage therapy applications.

In conclusion, our study demonstrates an important contribution of prophages in the growing *Pectobacterium* species pangenome. We identified an ongoing bacteriophage-mediated exchange of genes in the new lineage of *P. brasiliense* as well as highly conserved prophage-like elements in the *P. brasiliense* genomes. Our results thereby emphasize the importance of routine sampling and pangenome analysis in understanding the evolutionary trajectories of bacterial plant pathogens.

Methods

Sample collection and phenotyping

During the annual screening and field inspection conducted by the Dutch General Inspection Service for agricultural seeds and seed potatoes (NAK) in the Netherlands from 2018 to 2020, *Pectobacterium* spp. strains were collected from symptomatic potato plants and non-symptomatic potato tubers. Non-symptomatic potato peels were crushed in water and incubated in pectate buffer before spreading on single-layer crystal violet pectate (CVP) plates. Characteristic cavity-forming colonies were sub-cultured on the CVP and nutrient agar plates to obtain pure strains and stored at $-80°C$ in 15% glycerol with half-strength nutrient broth.

BL-causing ability of the *P. brasiliense* isolates was tested on potato cultivar Agria during 2019 to 2023 field trials. Isolates were grown on nutrient agar plates for one day at 28°C. Suspensions of $OD_{600} = 0.1$ were prepared in 10 mM phosphate buffer (pH 7.2) and diluted an extra 100×, resulting in a suspension of about 106 cfu/ml. Tubers were submerged in the solution and brought under a -0.07Pa vacuum. The vacuum was kept for 10 minutes, after which tubers were left submerged for another 15 minutes. Tubers were air-dried before planting and the ability to cause BL was scored as the development of typical plant symptoms (*14*). Isolates that induced BL phenotype in $\geq 30\%$ plants were categorized as BL-causing, in $< 5\%$ plants as non-causing and the remaining were labelled inconclusive.

Additionally, 35 *Pectobacterium* species isolates were included from the Netherlands Institute for Vectors, Invasive plants and Plant health (NIVIP) collection.

Genome sequencing, quality control and annotation

Whole genome sequencing and genome assembly were performed for 23 *P. brasiliense* isolates from the

NAK collection and 35 isolates from the NIVIP collection as described in Jonkheer et al. (*17*) and Blom et al. (*73*) respectively. Publicly available genomes from *Pectobacterium* species and related BioSample metadata were downloaded (date: 29/08/2022) using the NCBI E-utility tool (*74*). For uniform taxonomy validation of our collection, we combined in-house and NCBI type-strains to perform an independent ANI comparison and taxonomy correction. Of the 508 genomes, 26 marked as "anomalous", "excluded" or "replaced" by NCBI were excluded. We used BUSCO (v5.2.2, database: "enterobacterales_odb10") protein to evaluate genome assembly completeness and removed 21 genomes that had less than 99% "complete BUSCOs" (*75*). Finally, seven genomes were found to be duplicated and hence only one copy was kept for further analysis. The remaining 454 genomes were annotated using Prokka (v1.14.6) (*76*). InterProScan (v5.56-89.0) was used to annotate genomes with protein family and domain information from the InterPro database (*77*, *78*). EggNOG-mapper (v2.1.10) was used to annotate proteins by transferring annotations from the EggNOG database (v5.0.2) (*79*, *80*).

Pangenome construction and analysis

A *Pectobacterium* spp. pangenome was built using PanTools (v4.1.1) (*35*, *81*). Both InterProScan and EggNOG annotations were added to the PanTools pangenome database using the *add_annotations* module. To group genes into homology groups, PanTools' *optimal_grouping* command tests eight different similarity cutoffs and evaluates the grouping of single-copy BUSCO genes. The time required to optimize the grouping increases quadratically with the number of genomes. To efficiently determine the relaxation setting for grouping of genes, we used two subsampling approaches. First, genomes of the type strain for each species were used. In the second, the k-mer tree built using *kmer_classification* command was cut into 20 clusters (number of *Pectobacterium* species) and a genome was selected from each cluster at random. This second method of random subsampling was repeated 100× to generate 100 subsets of *Pectobacterium* genomes. The optimal grouping step of PanTools was performed on these subsets and the relaxation setting with lowest average F-score was selected. This resulted in a relaxation setting of four, 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 five.

Phylogenetic analysis

Core genes were used to construct a maximum likelihood phylogenetic tree using the PanTools *core_phylogeny* command (*82*). Phylogenetic trees were processed using R (v4.2.1) packages Ape (v5.7.1) (*83*, *84*) and treeio (v1.22.0) (*85*) and data was visualized with R packages ggtree (v3.6.2) (*86*) and ComplexHeatmap (v2.15.1) (*87*).

Prophage detection and quality control

Prophages were detected in genomes using the geNomad (v1.5.0) pipeline and checkV (v1.0.1) was used for virus genome completeness assessment (*38*, *88*). Host sequence regions at the prophage boundaries predicted by CheckV were trimmed to remove the contamination resulting from host genes. 139 prophages without any gene of viral origin were excluded from downstream analysis. In fragmented genome assemblies, prophages may be split over multiple contigs, leading to an over-estimation of the number of prophages.

We used homology group information of the orthologous prophages from the high quality genome assemblies to identify such fragmented prophages ($n = 145$) and merged them into 61 prophages. Merging was considered confident if the merged prophages together covered at least 80% of the closest matching orthologous prophages ($n = 50$ from 110 fragmented prophages) and the rest were excluded ($n = 11$). Finally, the shortest prophage detected with $> 90\%$ completeness was of length 5.2kb and hence prophages smaller than 5kb were also excluded, resulting in 1,369 prophages $\geq 5kb$ in length.

Orthologous prophage detection

Each putative prophage region was represented as a sequence of homology groups from the pangenome, defined as the prophage signature. Prophage similarity was quantified by evaluating the syntenic conservation of homology groups between two prophage signatures. To quantify such syntenic conservation, we applied dynamic programming to align two prophage signatures and calculate the pairwise Jaccard index. Only syntenicly matching homology groups between a pair of prophage signatures are considered for determining the size of the intersection for the Jaccard index calculation. During dynamic programming scoring, a score of 5 was used for a match and -2 for a mismatch of homology groups between two prophage signatures being compared. To account for the natural variation in prophages due to the mosaicism and evolution, a maximum gap length of two was allowed. A syntenic Jaccard index was considered significant if two prophages had the longest local subsequence with a chain length, i.e. the number of syntenicly matching homology groups, of at least five. An all-vs-all syntenic Jaccard index matrix was clustered using the complete linkage hierarchical clustering function 'hclust' in R. A conserved phage tail-like bacteriocin, CTV, was present in 428/454 *Pectobacterium* genomes and clustered together. We used CTV's intra-cluster similarity as a threshold to decide the grouping of similar prophage-like elements into clusters. The complete-linkage dendrogram was cut at height 0.66, which was just above the node height for the CTV cluster, resulting in 436 prophage clusters. Here, the complete-linkage clustering ensures that each resulting cluster includes prophages with intra-cluster similarity equal to or higher than the chosen cut-height or similarity threshold. A cluster representative backbone was selected using criteria in the following order: highest checkV completeness score, longest prophage length and highest average intra-cluster Jaccard index.

Prophage cluster visualization

A 5kb flanking region for each prophage was included to provide a broader context to understand integration at the same or different sites in the genome. An R script was written to extract the homology groups for prophages in a cluster and the data was stored in required JSON format by clustermap.js (https://github.com/gamcil/clustermap.js). This JSON data was used in clinker (*89*) to generate interactive plots to visualize orthologous prophages. Prophage genes were manually classified into broad functional categories based on the PFAM and COG description: integrase, capsid, replication, defense, DNA-related functions, phage tail-related, other virus genes, or uncharacterized.

Data and code availability

All the 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.13631140. Sequencing data can be accessed at NCBI with BioProjects PRJNA1122109 for the NIVIP collection and PRJNA1140644 for the NAK collection.

Supplementary information

Additional file 1: Supplementary_tables.xlsx

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. ID, MP: Resources. SS, TL, DR: Supervision, Funding acquisition, Project administration. LP, SS, TL, DR, AK: Writing - review and editing. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We would like to thank Harm Nijveen and the IT team of Wageningen University & Research for managing the Linux servers.

References

(*1*) Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S. V., Machado, M. A., Toth, I., Salmond, G., and Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology 13*, 614–629.

(*2*) Charkowski, A. O. (2018). The Changing Face of Bacterial Soft-Rot Diseases. *Annual Review of Phytopathology 56*, 269–288.

(*3*) Oulghazi, S., Sarfraz, S., Zaczek-Moczydłowska, M. A., Khayi, S., Ed-Dra, A., Lekbach, Y., Campbell, K., Novungayo Moleleki, L., O'Hanlon, R., and Faure, D. (2021). Pectobacterium brasiliense: Genomics, Host Range and Disease Management. *Microorganisms 9*, 106.

(*4*) Duarte, V., De Boer, S., Ward, L., and de Oliveira, A. (2004). Characterization of atypical Erwinia carotovora strains causing blackleg of potato in Brazil. *Journal of Applied Microbiology 96*, 535–545.

(*5*) Nabhan, S., De Boer, S., Maiss, E., and Wydra, K. (2012). Taxonomic relatedness between Pectobacterium carotovorum subsp. carotovorum, Pectobacterium carotovorum subsp. odoriferum and Pectobacterium carotovorum subsp. brasiliense subsp. nov. *Journal of Applied Microbiology 113*, 904–913.

(*6*) Adeolu, M., Alnajar, S., Naushad, S., and S. Gupta, R. (2016). Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology 66*, 5575–5599.

(*7*) Zhang, Y., Fan, Q., and Loria, R. (2016). A re-evaluation of the taxonomy of phytopathogenic genera Dickeya and Pectobacterium using whole-genome sequencing data. *Systematic and Applied Microbiology 39*, 252–259.

(*8*) Portier, P., Pédron, J., Taghouti, G., Fischer-Le Saux, M., Caullireau, E., Bertrand, C., Laurent, A., Chawki, K., Oulgazi, S., Moumni, M., Andrivon, D., Dutrieux, C., Faure, D., Hélias, V., and Barny, M.-A. (2019). Elevation of Pectobacterium carotovorum subsp. odoriferum to species level as Pectobacterium odoriferum sp. nov., proposal of Pectobacterium brasiliense sp. nov. and Pectobacterium actinidiae sp. nov., emended description of Pectobacterium carotovorum and description of Pectobacterium versatile sp. nov., isolated from streams and symptoms on diverse plants. *International Journal of Systematic and Evolutionary Microbiology 69*, 3207–3216.

(*9*) 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.

(*10*) 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*.

(*11*) 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.

(*12*) 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.

(*13*) 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.

(*14*) 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.

(*15*) van der Wolf, J. M. et al. In *Plant Diseases Caused by Dickeya and Pectobacterium Species*, Van Gijsegem, F., van der Wolf, J. M., and Toth, I. K., Eds.; Springer International Publishing: Cham, 2021, pp 215–261.

(*16*) Pérombelon, M. C. M. (2002). Potato diseases caused by soft rot erwinias: an overview of pathogenesis. *Plant Pathology 51*, 1–12.

(*17*) Jonkheer, E. M., Brankovics, B., Houwers, I. M., van der Wolf, J. M., Bonants, P. J. M., Vreeburg, R. A. M., Bollema, R., de Haan, J. R., Berke, L., Smit, S., de Ridder, D., and van der Lee, T. A. J. (2021). The Pectobacterium pangenome, with a focus on Pectobacterium brasiliense, shows a robust core and extensive exchange of genes from a shared gene pool. *BMC Genomics 22*, 265.

(*18*) Li, X., Ma, Y., Liang, S., Tian, Y., Yin, S., Xie, S., and Xie, H. (2018). Comparative genomics of 84 Pectobacterium genomes reveals the variations related to a pathogenic lifestyle. *BMC genomics 19*, 889.

(*19*) Zoledowska, S., Motyka-Pomagruk, A., Sledz, W., Mengoni, A., and Lojkowska, E. (2018). High genomic variability in the plant pathogenic bacterium Pectobacterium parmentieri deciphered from de novo assembled complete genomes. *BMC Genomics 19*, 751.

(*20*) Li, L., Yuan, L., Shi, Y., Xie, X., Chai, A., Wang, Q., and Li, B. (2019). Comparative genomic analysis of Pectobacterium carotovorum subsp. brasiliense SX309 provides novel insights into its genetic and phenotypic features. *BMC genomics 20*, 486.

(*21*) Zhang, Y., Chu, H., Yu, L., He, F., Gao, Y., and Tang, L. (2022). Analysis of the Taxonomy, Synteny, and Virulence Factors for Soft Rot Pathogen Pectobacterium aroidearum in Amorphophallus konjac Using Comparative Genomics. *Frontiers in Microbiology 13*, 868709.

(*22*) Camargo, A. P. et al. (2023). IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Research*, gkad964.

(*23*) 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.

(*24*) Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., and Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature Communications 1*, 147.

(*25*) Bondy-Denomy, J., and Davidson, A. R. (2014). When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *Journal of Microbiology 52*, 235–242.

(*26*) Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature Reviews Microbiology 13*, 641–650.

(*27*) Harrison, E., and Brockhurst, M. A. (2017). Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. *BioEssays 39*, 1700112.

(*28*) Taylor, V. L., Fitzpatrick, A. D., Islam, Z., and Maxwell, K. L. In *Advances in Virus Research*, Kielian, M., Mettenleiter, T. C., and Roossinck, M. J., Eds.; Academic Press: 2019; Vol. 103, pp 1–31.

(*29*) Evans, T. J., Coulthurst, S. J., Komitopoulou, E., and Salmond, G. P. C. (2010). Two mobile Pectobacterium atrosepticum prophages modulate virulence. *FEMS microbiology letters 304*, 195–202.

(*30*) Nasser, W. et al. (2014). Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. *Proceedings of the National Academy of Sciences of the United States of America 111*, E1768–E1776.

(*31*) Foxall, R. L., Means, J., Marcinkiewicz, A. L., Schillaci, C., DeRosia-Banick, K., Xu, F., Hall, J. A., Jones, S. H., Cooper, V. S., and Whistler, C. A. Inoviridae prophage and bacterial host dynamics during diversification, succession and Atlantic invasion of Pacific-native Vibrio parahaemolyticus, 2023.

(*32*) van der Lee, T. A. J., van Gent-Pelzer, M. P. E., Jonkheer, E. M., Brankovics, B., Houwers, I. M., van der Wolf, J. M., Bonants, P. J. M., van Duivenbode, I., Vreeburg, R. A. M., Nas, M., and Smit, S. (2023). An Efficient Triplex TaqMan Quantitative PCR to Detect a Blackleg-Causing Lineage of Pectobacterium brasiliense in Potato Based on a Pangenome Analysis. *Microorganisms 11*, 2080.

(*33*) Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology 57*, 81–91.

(*34*) Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., Brover, S., Schoch, C. L., Kimchi, A., and DiCuccio, M. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology 68*, 2386–2392.

(*35*) 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.

(*36*) Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology 11*, 472–477.

(*37*) Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., and Kyrpides, N. C. You can move, but you can't hide: identification of mobile genetic elements with geNomad, 2023.

(*38*) 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.

(*39*) Bobay, L.-M., Rocha, E. P., and Touchon, M. (2013). The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Molecular Biology and Evolution 30*, 737–751.

(*40*) Wylie, S. J., Adams, M., Chalam, C., Kreuze, J., López-Moya, J. J., Ohshima, K., Praveen, S., Rabenstein, F., Stenger, D., Wang, A., and Zerbini, F. M. (2017). ICTV Virus Taxonomy Profile: Potyviridae. *The Journal of General Virology 98*, 352–354.

(*41*) Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America 96*, 2192–2197.

(*42*) Hatfull, G. F., and Hendrix, R. W. (2011). Bacteriophages and their Genomes. *Current opinion in virology 1*, 298–303.

(*43*) Moura de Sousa, J. A., Pfeifer, E., Touchon, M., and Rocha, E. P. C. (2021). Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa. *Molecular Biology and Evolution 38*, 2497–2512.

(*44*) Kupczok, A., Bailey, Z. M., Refardt, D., and Wendling, C. C. (2022). Co-transfer of functionally interdependent genes contributes to genome mosaicism in lambdoid phages. *Microbial Genomics 8*, 000915.

(*45*) Holtappels, D., Alfenas-Zerbini, P., and Koskella, B. (2023). Drivers and consequences of bacteriophage host range. *FEMS Microbiology Reviews 47*, fuad038.

(*46*) Ranjani, P., Gowthami, Y., Samuel, Gnanamanickam, S., and Palani, P. (2018). Bacteriophages: A New Weapon for the Control of Bacterial Blight Disease in Rice Caused by Xanthomonas oryzae. *46*, 346–359.

(*47*) Mohammed, M., Casjens, S. R., Millard, A. D., Harrison, C., Gannon, L., and Chattaway, M. A. (2023). Genomic analysis of Anderson typing phages of Salmonella Typhimrium: towards understanding the basis of bacteria-phage interaction. *Scientific Reports 13*, 10484.

(*48*) Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., and Weitz, J. S. (2011). Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences of the United States of America 108*, E288–E297.

(*49*) Göller, P. C., Elsener, T., Lorgé, D., Radulovic, N., Bernardi, V., Naumann, A., Amri, N., Khatchatourova, E., Coutinho, F. H., Loessner, M. J., and Gómez-Sanz, E. (2021). Multi-species host range of staphylococcal phages isolated from wastewater. *Nature Communications 12*, 6965.

(*50*) 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.

(*51*) 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.

(*52*) Greenrod, S. T. E., Stoycheva, M., Elphinstone, J., and Friman, V.-P. (2022). Global diversity and distribution of prophages are lineage-specific within the Ralstonia solanacearum species complex. *BMC Genomics 23*, 689.

(*53*) Mottawea, W. et al. (2018). Salmonella enterica Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. *Frontiers in Microbiology 9*, 836.

(*54*) Bellieny-Rabelo, D., Tanui, C. K., Miguel, N., Kwenda, S., Shyntum, D. Y., and Moleleki, L. N. (2019). Transcriptome and Comparative Genomics Analyses Reveal New Functional Insights on Key Determinants of Pathogenesis and Interbacterial Competition in Pectobacterium and Dickeya spp. *Applied and Environmental Microbiology 85*, e02050–18.

(*55*) Coupat-Goutaland, B., Bernillon, D., Guidot, A., Prior, P., Nesme, X., and Bertolla, F. (2011). Ralstonia solanacearum virulence increased following large interstrain gene transfers by natural transformation. *Molecular plant-microbe interactions: MPMI 24*, 497–505.

(*56*) Jacobs, J. M., Babujee, L., Meng, F., Milling, A., and Allen, C. (2012). The in planta transcriptome of Ralstonia solanacearum: conserved physiological and virulence strategies during bacterial wilt of tomato. *mBio 3*, e00114–12.

(*57*) Vanga, B. R., Ramakrishnan, P., Butler, R. C., Toth, I. K., Ronson, C. W., Jacobs, J. M. E., and Pitman, A. R. (2015). Mobilization of horizontally acquired island 2 is induced in planta in the phytopathogen Pectobacterium atrosepticum SCRI1043 and involves the putative relaxase ECA0613 and quorum sensing. *Environmental Microbiology 17*, 4730–4744.

(*58*) Panda, P., Vanga, B. R., Lu, A., Fiers, M., Fineran, P. C., Butler, R., Armstrong, K., Ronson, C. W., and Pitman, A. R. (2016). Pectobacterium atrosepticum and Pectobacterium carotovorum Harbor Distinct, Independently Acquired Integrative and Conjugative Elements Encoding Coronafacic Acid that Enhance Virulence on Potato Stems. *Frontiers in Microbiology 7*, 397.

(*59*) Kauffman, K. M., Chang, W. K., Brown, J. M., Hussain, F. A., Yang, J., Polz, M. F., and Kelly, L. (2022). Resolving the structure of phage–bacteria interactions in the context of natural diversity. *Nature Communications 13*, 372.

(*60*) Sieiro, C., Areal-Hermida, L., Pichardo-Gallardo, Á., Almuiña-González, R., de Miguel, T., Sánchez, S., Sánchez-Pérez, Á., and Villa, T. G. (2020). A Hundred Years of Bacteriophages: Can Phages Replace Antibiotics in Agriculture and Aquaculture? *Antibiotics 9*, 493.

(*61*) 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.

(*62*) 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.

(*63*) 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.

(*64*) 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.

(*65*) 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.

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

(*67*) 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.

(*68*) Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology 8*, 317–327.

(*69*) LeRoux, M., and Laub, M. T. (2022). Toxin-Antitoxin Systems as Phage Defense Elements. *Annual Review of Microbiology 76*, 21–43.

(*70*) Samson, J. E., Magadán, A. H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology 11*, 675–687.

(*71*) Owen, S. V., Wenner, N., Dulberger, C. L., Rodwell, E. V., Bowers-Barnard, A., Quinones-Olvera, N., Rigden, D. J., Rubin, E. J., Garner, E. C., Baym, M., and Hinton, J. C. (2021). Prophages encode phage-defense systems with cognate self-immunity. *Cell Host & Microbe 29*, 1620–1633.e8.

(*72*) Morgan, T., de Rezende, R. R., Lima, T. T. M., Souza, F. d. O., and Alfenas-Zerbini, P. (2022). Genomic Analysis Unveils the Pervasiveness and Diversity of Prophages Infecting Erwinia Species. *Pathogens 12*, 44.

(*73*) Blom, N., Gorkink-Smits, P., Landman, M., Bilt, J. v. d., Vogelaar, M., Raaymakers, T., Visser, M., Pel, M., and Bergsma-Vlami, M. Ralstonia solanacearum (phylotype II) isolated from Rosa spp. in the Netherlands is closely related to phylotype II isolates from other sources in the Netherlands and is virulent on potato, 2024.

(*74*) Sayers, E., *E-utilities Quick Start*; National Center for Biotechnology Information (US): 2018.

(*75*) Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics 31*, 3210–3212.

(*76*) Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics 30*, 2068–2069.

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

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

(*79*) 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.

(*80*) 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.

(*81*) 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.

(*82*) 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.

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

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

(*85*) 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.

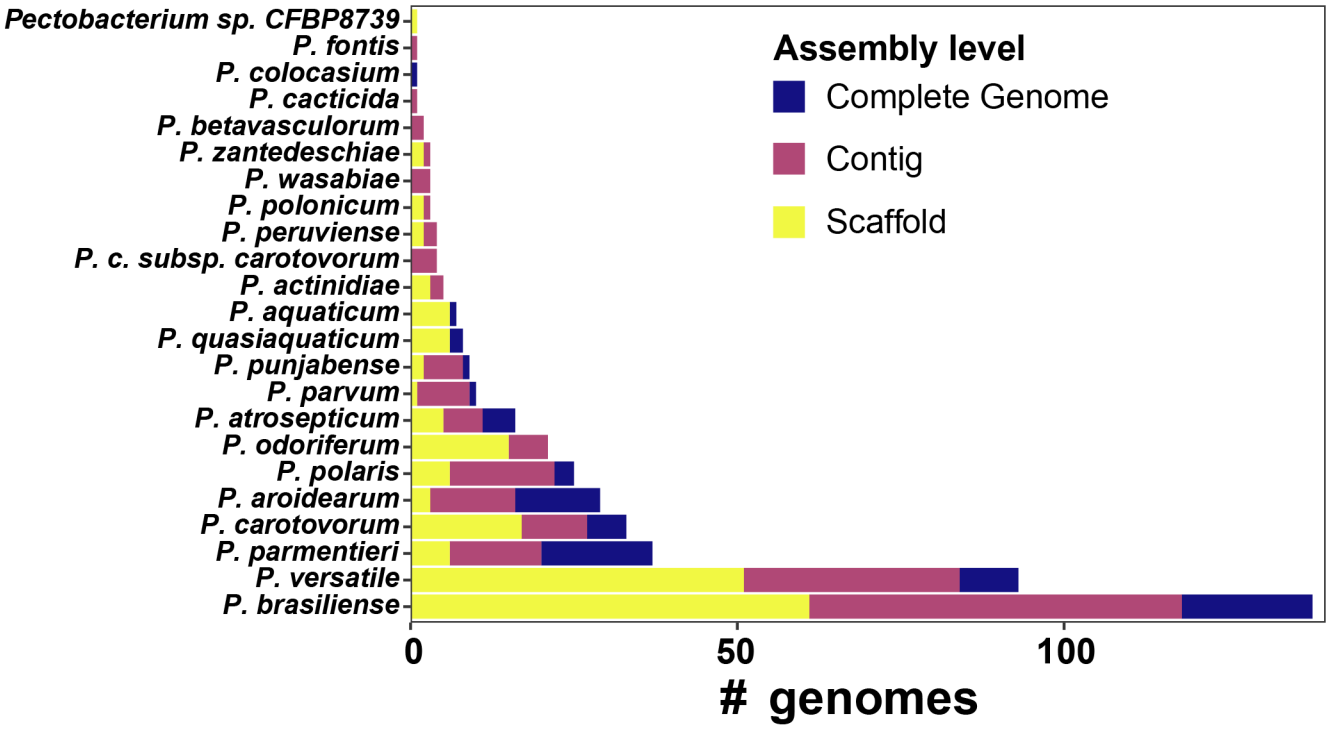
(*86*) Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T. T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta 1*, e56.

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

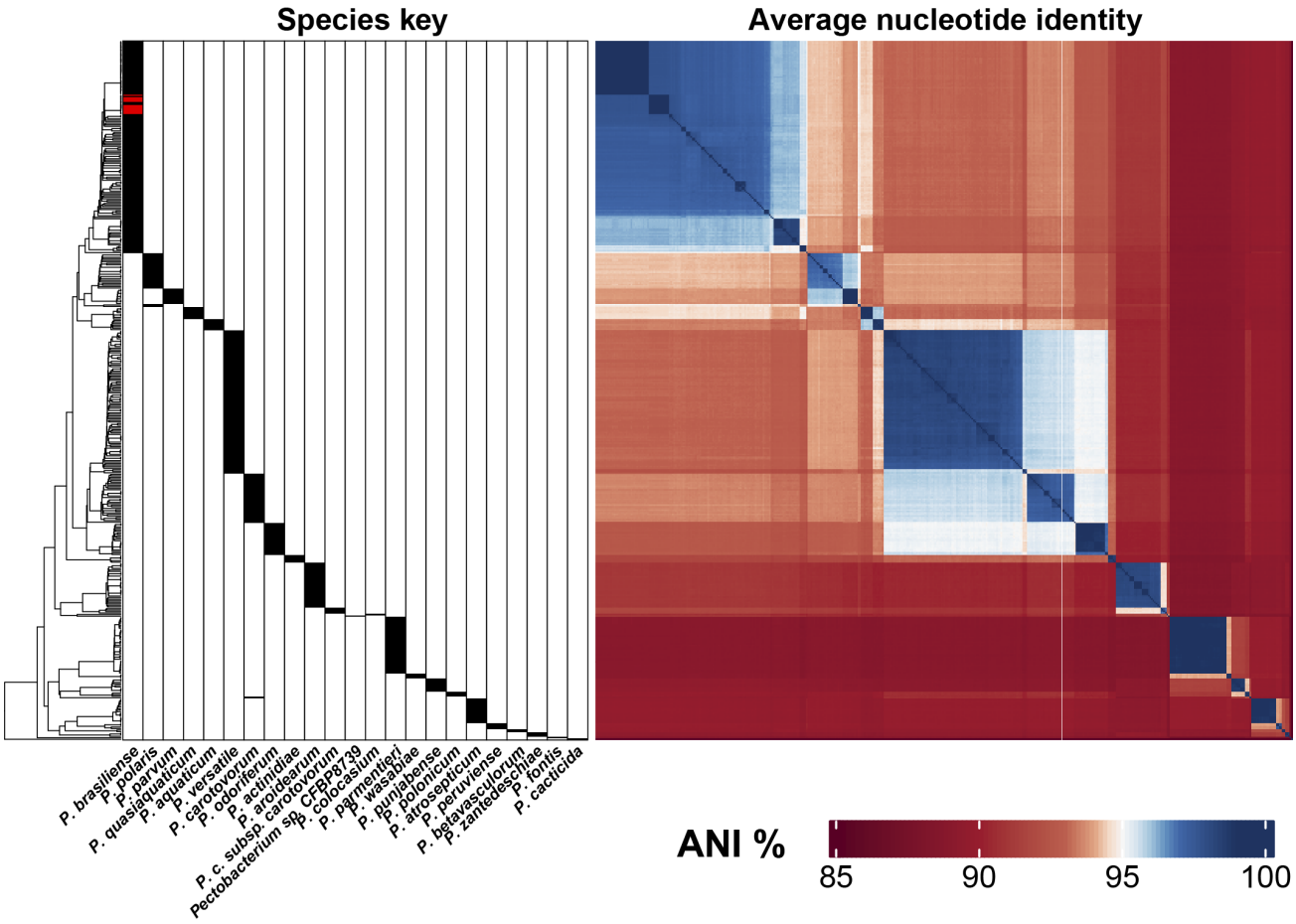
(*88*) Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., and Kyrpides, N. C. (2023). Identification of mobile genetic elements with geNomad. *Nature Biotechnology*, 1–10.

(*89*) Gilchrist, C. L. M., and Chooi, Y.-H. (2021). clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics 37*, 2473–2475.

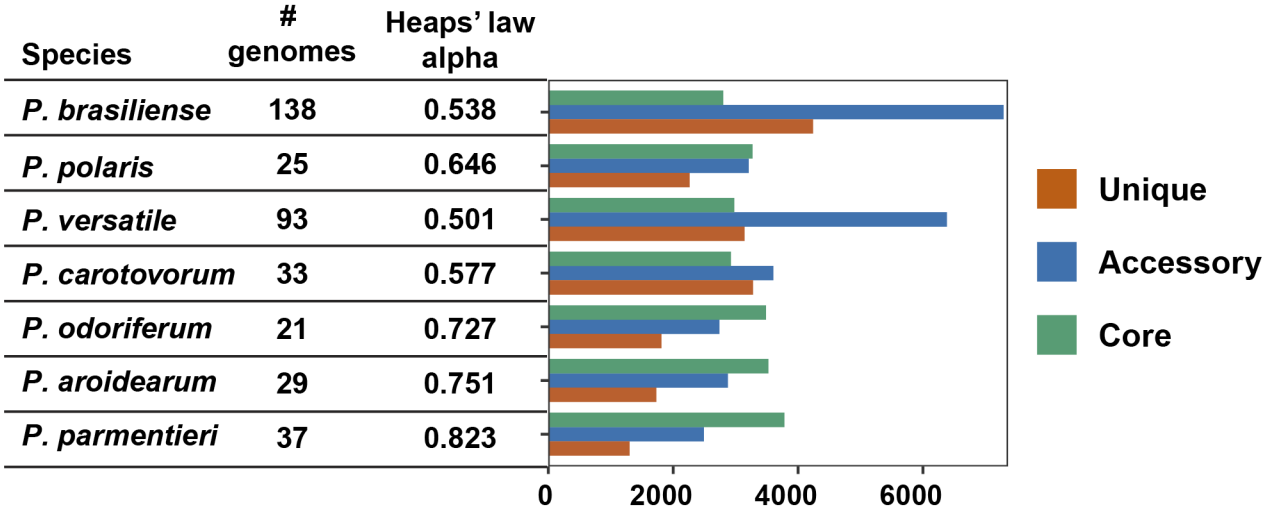
Supplementary Figures, belonging to Pardeshi et al., "Pangenomics approach to understand the prophage dynamics in the radiating lineages of P. brasiliense"



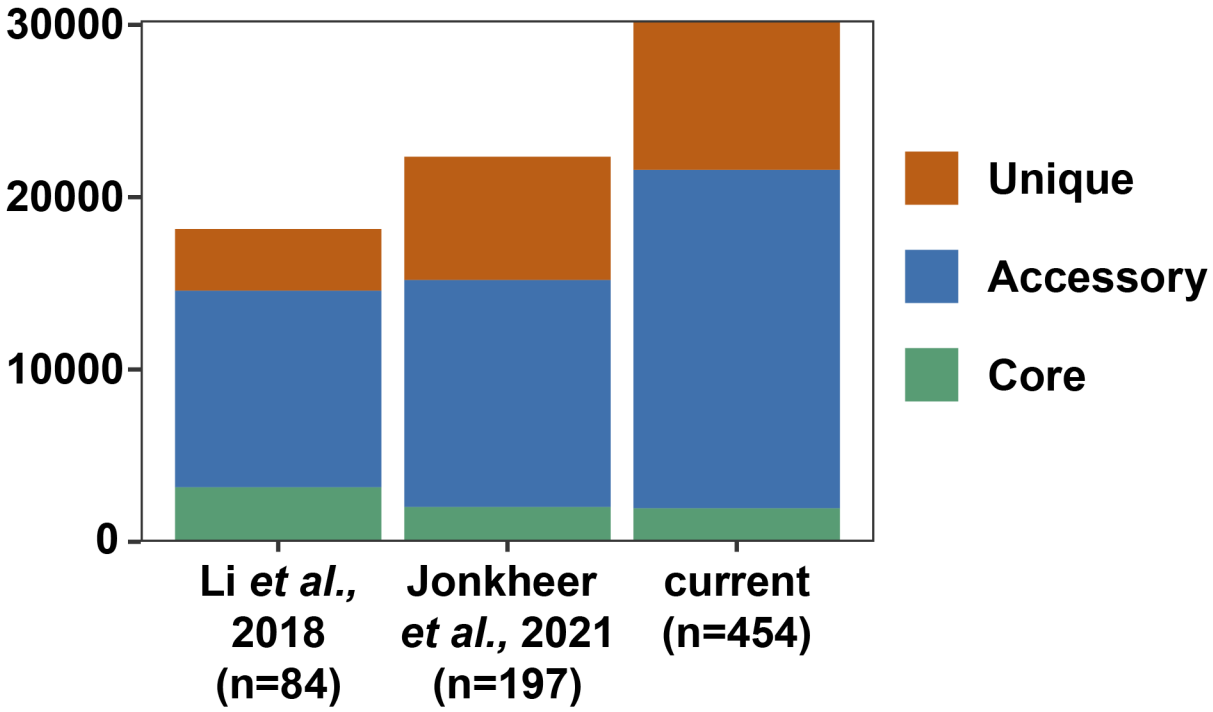
Supplementary Figure S1: Numbers of *Pectobacterium* species genomes in the pangenome.



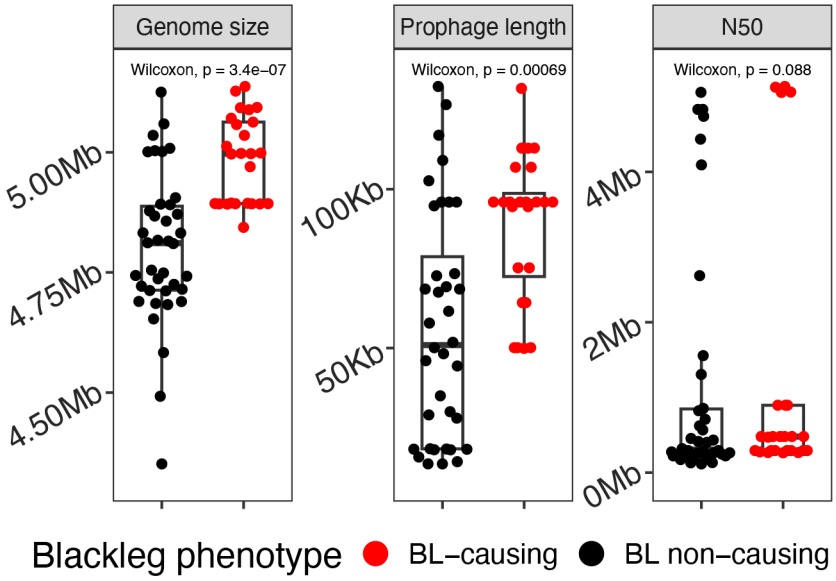
Supplementary Figure S2: ANI between all pairs of *Pectobacterium* spp. genomes: from left to right, a dendrogram based on UPGMA clustering on a (1-ANI) distance matrix; species key, where FN-Pbr isolates are marked in red; ANI heatmap. ANI heatmap color scale is centered around white color which represents 95% ANI, a commonly used species delineation cutoff.



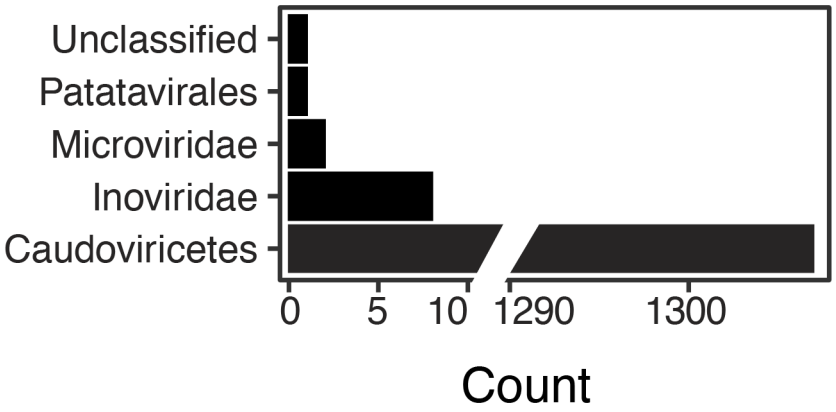
Supplementary Figure S3: Core, accessory and unique gene statistics for *Pectobacterium* species, with pangenome openness depicted by Heaps' law α value.



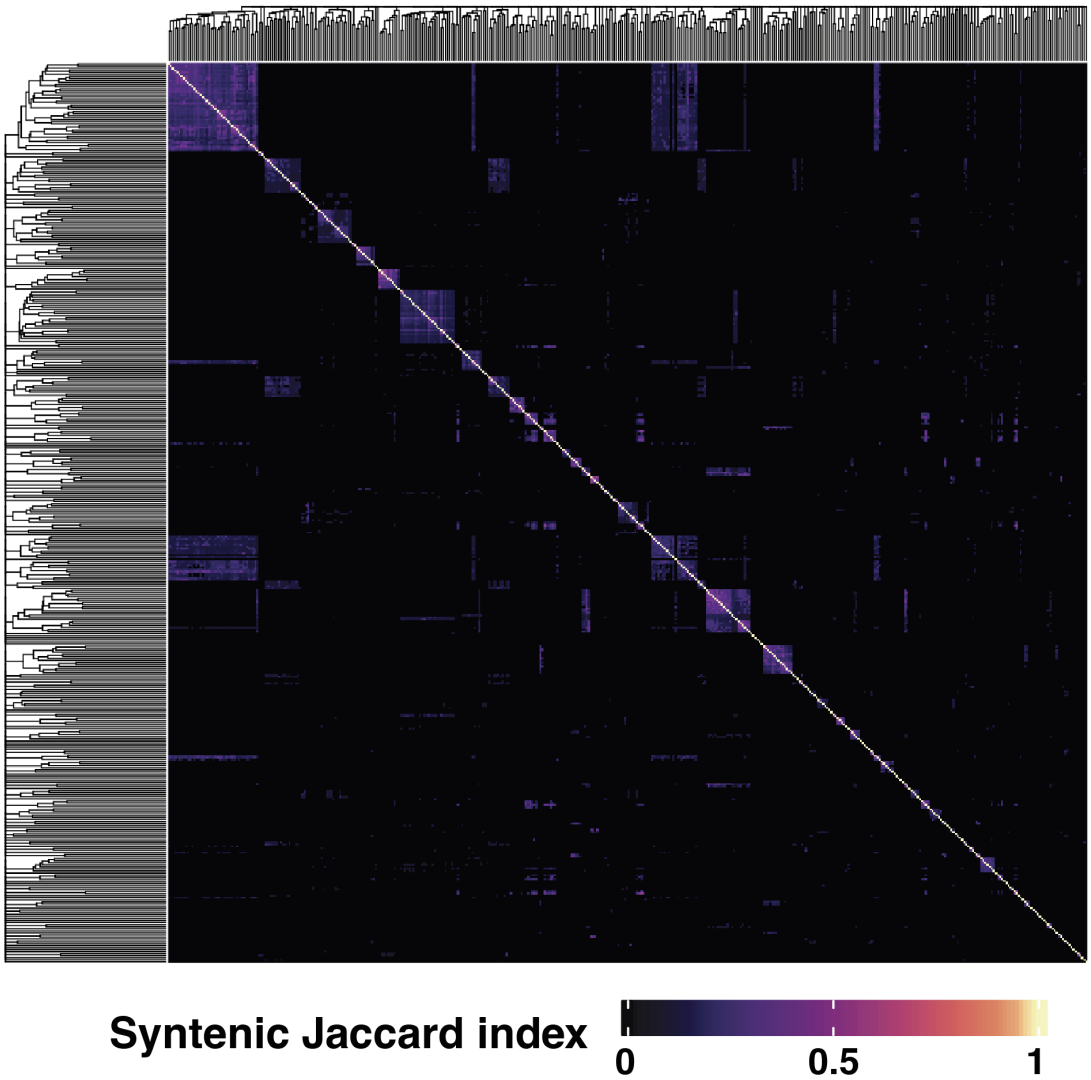
Supplementary Figure S4: Homology group counts in previous *Pectobacterium* spp. pangenomes and the current one. Number of genomes in the pangenome are mentioned in the parenthesis.



Supplementary Figure S5: Prophage length comparison between BL-causing and non-causing *P. brasiliense* isolates



Supplementary Figure S6: Prophage taxonomy



Supplementary Figure S7: Syntenic Jaccard index heatmap for 436 representative prophages