Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Distinct Genomic Architecture of Pectobacterium brasiliense Strain BS1113, a Soft Rot Pathogen of Cigar Tobacco Lacking a Type III Secretion System

Xuemei Zhang
Fuyang Normal University
Xiuting Geng
Fuyang Normal University
Chao Lu
9286031116@qq.com
Fuyang Normal University
Jian Cai
Fuyang Normal University

Research Article

Keywords: Pectobacterium brasiliense, genome-wide, comparative genomics, Pathogenicity, soft rot

Posted Date: April 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8110102/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The bacterium *Pectobacterium brasiliense* causes severe soft rot disease across a wide variety of plant hosts. Although the genome of the reference strain SX309 has been characterized, the molecular basis of pathogenicity in isolates adapted to cigar tobacco has not yet been investigated.This study presents a comparative genomic and pathogenicity analysis of *P. brasiliense* strain BS1113, isolated from cigar tobacco, to elucidate its unique adaptive features.

Results

Here we report the complete genome sequence of strain BS1113, which comprises 4.92 Mb with a G + C content of 51.96%. Comparative analysis against the closely related strain SX309 revealed the absence of an intact type III secretion system (T3SS) gene cluster in BS1113, despite T3SS being a canonical virulence determinant in numerous Gram-negative pathogens, Despite this absence, BS1113 retains a highly conserved arsenal of virulence factors, including plant cell wall-degrading enzymes (PCWDEs) and their dedicated type II secretion system (T2SS), along with quorum-sensing systems. Furthermore, the genome harbors variable regions encoding a type VI secretion system (T6SS) and a subtype I-F CRISPR-Cas system, implying roles in for host interaction and adaptive immunity.

Conclusions

The lack of T3SS in BS1113 points to a distinct evolutionary trajectory towards niche specialization. Its pathogenic strategy may depend on opportunistic invasion through wounds and rapid tissue degradation mediated by T2SS-delivered PCWDEs, diverging from the effector-dependent immunosuppression employed by T3SS-containing relatives such as SX309. Our findings highlight significant intra-species genomic diversity within *P. brasiliense* and uncover a distinct pathogenic architecture in strain BS1113, offering fresh insights on the adaptive evolution of soft rot pathogens.

Introduction

Bacterial soft rot, caused by members of the genus *Pectobacterium*, ranks among the most economically damaging diseases affecting vegetable and ornamental crops worldwide [1–5]. These pathogens provoke symptoms ranging from maceration of parenchymatous tissues to blackleg and wilting, resulting in severe yield losses.Based on the analysis of PCR-amplified spacer regions, 16S rRNA gene sequence differences, and biochemical traits [6],the classification within the genus has undergone significant revisions. Initially described as a subspecies, *Pectobacterium brasiliense* (Pbr) was formally elevated to species rank in 2019 [7], following comprehensive genomic comparisons and biochemical profiling that distinguished it from closely related subspecies of *P. carotovorum*—particularly its unique ability to infect both solanaceous crops and specialty tobaccos [8].

Notably, Pbr exhibits exceptional host adaptability, with isolates reported from staple crops (potato, tomato) [9,10] and specialty crops like cigar tobacco—an emerging host with distinct cell wall composition (high lignin and pectin content) that may select for unique pathogenic traits [11,12]. Its global presence and adaptability to diverse environments underscore its threat to agriculture.

Early genomic comparisons, such as that of Glasner et al. [13], highlighted substantial variation (11–18%) among *Pectobacterium* strains, with a large fraction of variable genes being regulatory in nature. This observation led to the hypothesis that regulatory gene diversity underpins ecological adaptation. These genomic differences largely determine the biological characteristics and potential pathogenicity of the strains. Additionally, effector proteins associated with the hrp type III secretion system were identified in both strains, but not in *P. atrosepticum* strain SCRI1043. This difference may also contribute to the variations in pathogenicity between the strains. Further studies on these findings could enhance our understanding of the differences between the strains and the mechanisms underlying their interactions with host plants.

Zhou Yuan et al. (2018) sequenced and functionally annotated the entire genome of *P. atrosepticum* strain JG10-08, the pathogen responsible for potato black shin disease, using second-generation DNA sequencing technology [14]. The genome of strain JG10-08 revealed 168 genes associated with pathogenicity, including three major categories: cell wall-degrading enzymes, secretion system genes, and toxin genes.

While the reference strain *P. brasiliense* SX309 (isolated from cucumber) has provided insights into general virulence mechanisms [15], its genomic architecture reflects adaptation to a broad host range, leaving the genetic basis of *Pbr*'s specialization to cigar tobacco—an ecologically distinct niche with unique defense barriers—completely unexplored.The isolation of a new *P. brasiliense* strain, BS1113, from symptomatic cigar tobacco plants in China offered a unique chance to explore host-driven genomic adaptation.Here, we present a comparative genomic analysis of BS1113, aiming to delineate its unique genetic features and elucidate the evolutionary mechanisms driving its distinct pathogenic strategy.

Materials and Methods

Bacterial strains and genomic DNA extraction

- *P. brasiliense* BS1113 was isolated from infected cigar tobacco leaves showing typical soft rot symptoms in Yunnan Province of China in October 2022.A series of experiments, including phenotypic and biochemical characterization as well as host range analysis, were conducted on the strain.This strain was typically incubated in LB broth (Solarbio, China) at 28°C with constant agitation for 48 h. Genomic DNA was purified from pelleted cells using the REPLI-g Single Cell DNA Library Kit (Qiagen, Shanghai) according to the supplier's instructions.

Whole-genome sequencing and annotation

- Genome sequencing was carried out at Shanghai Meiji Biomedical Technology Co. on a PacBio RS II instrument. A 20-kb insert SMRTbell library was prepared, and sequencing was conducted with P6/C4 chemistry on a single SMRT cell. Raw PacBio reads were quality-filtered using SMRT Link v10.1 (minimum read length = 500 bp, minimum quality score = 0.8) to remove low-quality fragments, then de novo assembled with SMRT Analysis v2.3.0 [16] using the '--genome-size 5.0m' parameter (based on average genome size of Pbr strains [17]) and three rounds of polishing to resolve homopolymer regions.The graphical views of genome alignments were generated using CGView software [18].

Gene prediction and annotation (dup: abstract ?)

- Protein coding sequences (CDS) were predicted using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [19] with manual curation of virulence-related genes (e.g., PCWDEs, secretion systems) against the UniProtKB/Swiss-Prot database to correct automated annotation errors—particularly for short-length CDSs (< 300 bp) that are often misclassified.Transfer RNA genes were detected with tRNAscan-SE v2.0 [20], and ribosomal RNA genes with RNAmmer v1.2 [21].The functions of the predicted proteins were annotated based on a BLASTP search against the NonRedundant Protein Database (NR, https://blast.ncbi.nlm.nih.gov/Blast.cgi), the Pfam protein family database (http://pfam.xfam.org/), the Cluster of Orthologous Groups of proteins database (COG, https://www.ncbi.nlm.nih.gov/COG/), and the Kyoto Encyclopedia of Genes and Genomes database (KEGG, http://www.genome.jp/kegg/). Furthermore, sequence analysis was improved using the RAST analysis platform [22].Putative signal peptides and transmembrane helices were predicted using SignalP 4.0 [23] and TMHMM 2.0 [24], respectively. The metabolic pathways were examined using a KEGG Automatic Annotation Server (KAAS, http://www.genome.jp/tools/kaas/).

Comparative genomics among strains

Comparative genomics among *Pectobacterium* strains

- To place BS1113 in a phylogenetic context, we retrieved complete genome sequences of seven closely related *Pectobacterium* taxa from public databases: *P. carotovorum* subsp. *Odoriferum* BC S7 (CP009678), *P. brasiliense* SX309 (CP020350), *P. brasiliense* 1692 (CP047495), *P. wasabiae* CFBP 3304 (CP015750), *P. carotovorum* subsp. *carotovorum* ICMP 5702 (AODT00000000), *P. brasiliense* 21PCA_AGRO2 (CP113504), and *P. carotovorum* subsp. *carotovorum* PCC21 (CP003776). Pairwise average nucleotide identity (ANI) was calculated with OrthoANIu v1.2 [25], and digital DNA–DNA hybridization (dDDH) estimates were obtained using the GGDC 2.1 web server [26] with the BLAST+ alignment method and recommended settings.Complete genome comparisons were conducted using the progressive alignment option of the Mauve 2.3.1comparison software [27] with the BS1113 genome as the reference genome. Furthermore, synteny plots were also generated as alignments of the complete genome nucleotide sequences using MUMmer 3.22 [28]. To identify the set of common genes for the *Pectobacterium* genus and the set of genes unique to each species or subspecies, comparative analyses at the protein level were performed using an all-against-all comparison of the annotated genomes using BLASTP [29], and ortholog gene clustering analysis was implemented with the default settings [30]. Venn diagrams were created using R project language [31].The targets of the spacers were identified using ViroBLAST (https://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php) and
- local BLAST analysis against NCBI plasmid genomes(ftp://ftp.ncbi.nih.gov/refseq/release/plasmid/).

Results

General genomic features of *Pbr* BS1113

- A total of 1,361,455,351 clean reads with an average length of 150 bp and an N50 size of 7373 bp were generated.Assembly of the clean reads resulted in a single contig with 861.0-fold coverage on average without any gaps(Additional file 1: Table S1).The final chromosome comprises 4,916,962 bp with a G + C content of 51.96%. Annotation identified 4,369 protein-coding sequences (CDSs), 77 tRNA genes, and 22 rRNA genes (eight 5S, seven 16S, seven 23S). Notably, the genome size of BS1113 (4.92 Mb) is slightly smaller than that of its closest relative *P. brasiliense* SX309 (4.97 Mb) but larger than *P. brasiliense* 1692 (4.85 Mb) (Table 2). The GC content (51.96%) falls within the range of *Pectobacterium* spp. (50.40-52.18%), consistent with its taxonomic placement. This

moderate genomic size variation may reflect adaptive adjustments during specialization to cigar tobacco, while the conserved GC content suggests stability in core metabolic gene repertoires. A total of 29 tandem repeats were identified, accounting for 0.36% of the genome (Table 1, Fig. 1). These repeats are predominantly distributed in intergenic regions flanking PCWDE-encoding genes (e.g., polygalacturonase gene0827, pectate lyase gene1856), potentially contributing to genomic plasticity in regulating PCWDE expression during adaptation to the unique cell wall composition (high lignin and pectin content) of cigar tobacco. Additionally, one prophage region (18,271 bp, 51.96% GC) containing 24 CDSs was detected; this region harbors genes encoding integrases and hypothetical proteins, which may have been acquired via horizontal gene transfer to enhance environmental fitness.

Table 1
Genomic features of *Pectobacterium brasiliense* BS1113

Attribute	Value
Genome Size (bp)	4916962
Shape of DNA	Linear
No. of coding genes(bp)	4369
Maximum sequence length	214063
Average sequence length	6831.44
N50 length	7373
%GC	51.96
Gene Density(%)	86.67
No. of tandem repeats	29
In Genome (%)	0.36
Number of sRNAs:	149
No. of transfer RNAs(tRNAs)	77
No. of ribosomal RNAs (rRNAs)	22

Gene prediction and annotation

A total of 4,369 protein-coding sequences (CDSs) and 149 small RNAs (sRNAs) were predicted in the genome of *P. brasiliense* BS1113, with 1545 genes annotated to 351 functional subsystems (Fig. 1)—a distribution consistent with its niche-adapted lifestyle as a cigar tobacco-infecting pathogen. Gene Ontology (GO) analysis assigned 2298 genes to three core categories: molecular function, cellular component, and biological process (Fig. 2). Molecular function-related genes were the most abundant, followed by those involved in biological processes and cellular components. Within the molecular function category, ‘ATP binding’ was the dominant term (n = 264, 6.04%), trailed by ‘DNA binding’ (n = 219, 5.01%), ‘Metal ion binding’ (n = 197, 4.51%), and ‘DNA-binding transcription factor activity’ (n = 106, 2.43%)—reflecting the strain’s reliance on energy metabolism and transcriptional regulation for host adaptation. In the cellular component category, ‘Integral component of membrane’ (n = 528, 12.09%) was the largest group, followed by ‘Cytoplasm’ (n = 360, 8.24%) and ‘Plasma membrane’ (n = 347, 7.94%)—a pattern aligned with the importance of membrane-associated processes (e.g., PCWDE secretion, signal transduction) in pathogenicity. For biological processes, ‘transmembrane transport’ (n = 81, 1.85%) was the most prominent, followed by ‘regulation of DNA-templated transcription’ (n = 79, 1.81%), ‘translation’ (n = 61, 1.40%), and ‘Phosphorylation’ (n = 57, 1.30%)—highlighting key pathways supporting nutrient acquisition and virulence gene expression.

BLASTP searches against the Cluster of Orthologous Groups (COG) database classified 3690 genes into 24 functional categories (Fig. 3). The top four categories were ‘Transcription’ (n = 353, 9.56%), ‘Inorganic ion transport and metabolism’ (n = 294, 7.96%), ‘Cell wall/membrane/envelope biogenesis’ (n = 268, 7.26%), and ‘Energy production and conversion’ (n = 207, 5.61%)—consistent with the strain’s need to remodel cell structures, regulate virulence genes, and adapt to nutrient fluctuations in the cigar tobacco phyllosphere. Notably, ‘Cell motility’ (n = 75), ‘Transport and catabolism’ (n = 66), and ‘Defense mechanisms’ (n = 120) were the least represented categories, suggesting a streamlined lifestyle focused on tissue colonization rather than broad environmental motility. Additionally, ‘Amino acid transport and metabolism’ (n = 409, 11.08%), ‘Carbohydrate transport and metabolism’ (n = 387, 10.49%), ‘Coenzyme transport and metabolism’ (n = 219, 5.93%), and ‘Translation, ribosomal structure and biogenesis’ (n = 262, 7.10%) were highly represented—underscoring robust metabolic networks supporting rapid growth and PCWDE synthesis.

Of the 4,369 annotated genes in BS1113, 2,897 (66.3%) were mapped to 35 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways across six primary functional groups (Fig. 4). ‘Metabolism’ was the most enriched category (n = 1,993, 45.6%), followed by ‘Global and Overview Maps’ (n = 832, 19.0%), ‘Carbohydrate Metabolism’ (n = 291, 6.7%), ‘Amino Acid Metabolism’ (n = 192, 4.4%), and ‘Biosynthesis of Secondary Metabolites’ (n = 51, 1.2%)—reflecting a highly activated metabolic machinery tailored to degrade cigar tobacco’s complex cell wall components (e.g., lignin, pectin). Within ‘Environmental Information Processing’ (n = 529, 12.1%),

‘Membrane Transport’ (n = 351, 8.0%) and ‘Signal Transduction’ (n = 178, 4.1%) were the dominant pathways—critical for nutrient uptake and sensing host-derived cues. Under ‘Cellular Processes’ (n = 275, 6.3%), ‘Cellular Community - Prokaryotes’ (n = 157, 3.6%) and ‘Cell Motility’ (n = 90, 2.1%) were the most significant, indicating limited reliance on motility for pathogenicity and a focus on interbacterial interactions in the host niche. Collectively, these KEGG annotations confirm that BS1113 possesses a specialized metabolic network supporting its growth, development, and host-specific pathogenicity.

Combined analyses of amino acid composition, dipeptide patterns, and PSI-BLAST similarity searches identified 634 putative virulence proteins in BS1113—including 10 key PCWDEs (e.g., polygalacturonase gene0827, pectate lyase gene1856) and T2SS/T6SS components (e.g., vasD, gspE) previously linked to soft rot pathogenesis. Further, the Comprehensive Antibiotic Resistance Database (CARD) analysis identified seven CDSs homologous to known antibiotic resistance determinants, conferring resistance to fluoroquinolones (60 genes), carbapenems (13 genes), diaminopyrimidines (6 genes), phenicols (37 genes), and cephalosporins (23 genes)—potentially facilitating survival under antimicrobial pressure in agricultural settings. CRISPR-Cas Finder also detected six CRISPR repeat regions in BS1113: the longest repeat (2013 bp) contained 34 spacers, while the shortest (92 bp) had 2 spacers. Notably, 12 of these spacers showed homology to phages commonly associated with tobacco rhizospheres (e.g., Erwinia phage phiEa2809), indicating an adaptive immune system tailored to its host-specific ecological niche.

Figure 3. Cluster of orthologous group (COG) functional annotation of protein encoding genes (CDS) from *Pectobacterium brasiliense* (*Pbr*) BS1113. Four thousand five hundred twenty CDS had a COG classification. CDS from *Pbr* BS 1113 were grouped into 19 COG categories.

Comparison of Pbr BS1113 genome with other species and subspecies

For comparative genomic analysis of *P. brasiliense* BS1113, six publicly available complete genomes of *Pectobacterium* species or subspecies including *P. carotovorum* subsp. *odoriferum* BC S7(GenBank:CP009678.1), *P. brasiliense* strain 1692(GenBank:CP047495.1), *P. carotovorum* subsp. *carotovorum* PCC21 (GenBank: CP003776.1), *P. brasiliense* strain 21PCA (GenBank: CP113504.1) and *P. wasabiae* CFBP3304 (GenBank: CP015750.1) have been selected(Table 2). The genome size ranged from 4.84 to 5.04 Mbp,with a G + C content of 50.40-52.18% and 4205–4868 predicted CDS (Table 2). Similarly, the genomes of the seven *Pectobacterium* strains contain only one single chromosome without a plasmid.

Table 2
Genomic features of *Pectobacterium brasiliense* BS1113 and other *Pectobacterium* spp.

Features	BS1113	BC S7	1692	SX309	*Pcc* 21	21PCA	CFBP3304
Size (bp)	4,916,962	4,933,575	4,851,982	4,966,299	4,842,771	4,919,671	5,043,228
G + C content (%)	51.96	51.80	52.15	52.18	52.18	51.67	50.55
Replicons	One chromosome	One chromosome	One chromosome	One chromosome	One chromosome	One chromosome	One chromosome
Total genes	4468	4868	4310	4,455	4340	4506	4579
Predicted no. of CDS	4369	4868	4205	4351	4263	4407	4472
Ribosomal RNA	22	22	22	22	22	22	22
Transfer RNA	77	77	77	76	76	77	77
GenBank sequence	CM128641.1	CP009678.1	CP047495.1	CP020350.1	CP003776.1	CP113504.1	CP015750.1

To assess the evolutionary relatedness among sequenced strains within the genus *Pectobacterium*, we performed whole-genome alignments using Mauve 2.3.1 software with BS1113 as the reference genome.At the species level, we aligned the BS1113 genome to three other *P.brasiliense* (SX309□1692 and 21PCA) and two closely related subspecies,,*P. carotovorum* subsp. *carotovorum* PCC21 and *P. carotovorum* subsp. *odoriferum* BC S7 (Fig. 5A, B).

Within the *P. brasiliense* clade, this genomic alignment revealed that BS1113 shares substantially higher sequence similarity with SX309 than with strains 1692 or 21PCA—mirroring the ANI (96.8%) and dDDH (72.3%) values we calculated, which confirm their close phylogenetic clustering [25,26]. At the inter-subspecies level, BS1113 also exhibited greater genomic synteny with *P. carotovorum* subsp. *carotovorum* PCC21 than with *P. carotovorum* subsp. *odoriferum* BC S7, further validating its evolutionary affinity to the carotovorum subclade. Notably, pairwise comparison between BS1113 and PCC21 identified no large-scale genomic insertions or deletions (> 5 kb), but uncovered three distinct large local collinear block (LCB) inversions—one spanning ~ 87 kb in the PCWDE-enriched genomic region (encompassing gene0827 and gene1856) and two others flanking T6SS core genes (vasD, clpV). Such

structural rearrangements are hypothesized to modulate the spatiotemporal expression of virulence-associated loci, potentially enhancing BS1113's adaptability to the cigar tobacco niche [30].

The core genome of *Pbr* BS1113, *Pbr* SX309, *Pbr* 1692 and *Pbr* 21PCA is composed of 3407 orthologous genes. *Pbr* BS1113 displayed 409 unique gene families while it shared 91 genes with *Pbr* 1692, 101 genes with *Pbr* SX309 and 69 genes with *Pbr* 21PCA genome (Fig. 5C). At the subspecies level,there were 3420 orthologous genes shared(Fig. 5D). *Pbr* BS1113 shared 78 genes with *Pco* BC S7 and 324 genes with Pcc 21 genome while 494 gene families were uniquely represented in *Pbr* BS1113.The members of the unique gene families from *Pcc* ICMP5702 were associated with lipoprotein localization to membrane(GO: 0044873), DNA modification (GO: 0006304), biological process (GO: 0008150), response to stimulus (GO: 0050896), and nucleic acid binding (GO: 0003676).

We performed whole genome nucleotide alignment to de termine the synteny of *P.brasiliense* BS1113 relative to SX309, PCC21 and BCS7. The results showed partial synteny in that genes of BS1113 aligned with closely related genes from the BCS7 (Fig. 6), but had numerous inversions, translocations, rearrangements and deletions.

Plant cell wall-degrading enzymes

Cell wall-degrading enzymes are key factors in causing plant diseases. Their primary components include pectinase, cellulase, protease, and xylanase.

The genetic blueprint of strain BS1113 reveals a prominent pectinase gene family, with detailed counts documented in Table 3. These enzymes exhibit diverse modes of action and are categorized into two major functional groups: hydrolases and lyases. Within the strain's gene pool, particular attention is drawn to the presence of highly efficient pectin acetyltransferase, precise cleavers of pectinase, and diverse enzymatic molecules including exo-polygalacturonase, polygalacturonase, and oligo-galacturonase. Together, these elements form a complex degradation network targeting the cell wall.In the genome of BS1113, the number of genes encoding cellulase, protease, and xylanase appears relatively limited. Specifically, the cellulase-related gene family comprises 11 members, while β-glucanase also possesses 11 genes. However, for xylanase, only one annotated gene is present.

Table 3
Cell wall degrading enzyme gene statistics

Class Definition	Protein code	Gene ID	
polygalacturonase	(EC 3.2.1.15)	gene0827	
		gene1319	
		gene3148	
		gene3251	
pectin acetyl esterase (PAGE)	(EC 3.1.1.-)	gene1166	
		gene2259	
pectate lyase	(EC 4.2.2.2)	gene1856	
		gene4417	
exopolygalacturonate lyase	(EC 4.2.2.9)	gene1856	
		gene4417	
beta-glucosidase	(EC 3.2.1.21)	gene0012	gene2264
		gene0035	gene2568
		gene2923	gene2616
		gene3513	gene2796
		gene3723	gene2894
		gene0735	
oligogalacturonate lyase	(EC 4.2.2.6)	gene1974	
beta-xylosidase	(EC 3.2.1.37)	gene2616	

Secretion systems

The genome of BS1113 contains a wide variety of secretion systems, which are closely related to bacterial pathogenicity (Additional file 2: Table S2).According to the comparative analysis, the *P.brasiliense* BS1113 chromosome contains a highly conserved T2SS gene cluster (gspCDEFGHIJKLMN and outOSB)(Fig. 7), covering 17.669 kb with 15 ORFs. The gsp gene cluster shares an average of 90% similarity with that of various *Pectobacterium* species at the amino acid level(Additional file 2: Table S2), except that gspC is absent in

P. carotovorum subsp. *odoriferum* BC S7. The outOSB genes are also highly conserved among *Pectobacterium* spp., except that the outO gene is replaced by BCS7_14675 encoding a hypothetical protein in strain BC S7. Among the four *Pectobacterium* spp., the common characteristics of T2SS is that it contains pel and pehK genes upstream of gsp(Fig. 7). The genes involved in the secretion-signal recognition particle (Sec-SRP) system are highly conserved in all four *Pectobacterium* spp., except secA and secE,which are absent in strain BC S7.

The T2SS gene cluster of BS1113 (gspCDEFGHIJKLMN and outOSB) spans 17.669 kb with 15 ORFs,,sharing 90% amino acid similarity with PCC21 orthologs (Table S2) but harboring a unique 120-bp insertion upstream of gspE—this region contains a putative binding site for the PhoP-PhoQ two-component system (predicted via MEME Suite), suggesting niche-specific regulation of T2SS expression.This high conservation confirms the essential role of T2SS in PCWDE secretion across soft rot pathogens.

The T6SS gene cluster of BS1113 contains 33 genes, including 15 core genes (e.g., vasD, impL, clpV) and 5 vgrG/13 hcp effector-encoding genes (Fig. 8). Compared to SX309 (3 vgrG/10 hcp), the expanded copy number of vgrG/hcp in BS1113 suggests enhanced functional versatility in interbacterial competition or host interaction.

Whole-genome sequencing and comparative genomics analysis confirmed that the plant pathogenic bacterium *P.brasiliense* strain BS1113 lacks the complete hrp and hrc gene clusters encoding the type III secretion system (T3SS) in its genome.

The type VI secretion system (T6SS) is widely present in many Gram-negative bacteria, delivering toxic effector proteins into adjacent bacterial or host cells. In this study, the T6SS gene cluster of *P. brasiliense* BS1113 was found to have 33 genes, among which 15 were identified as core genes (Fig. 8).The 15 core T6SS genes are highly conserved in various *Pectobacterium* species and subspecies. Biological functions have been assigned for the outer membrane lipoprotein (VasD), Inner membrane proteins (ImpL and ImpK), ATPase (ClpV), and regulatory proteins or structure proteins (ImpB, ImpC, TssE, ImpG, ImpH, ImpI, ImpJ, VasH, VasI,VasJ, and VasL) [34] (Additional file 3: Table S3). In addition to the 15 core T6SS genes, there are five vgrG and 13 hcp genes that encode extracellular structural components of the secretion machine and specific effectors in BS1113 genome. Nevertheless, the copy numbers of vgrG and hcp genes substantially varied among different *Pectobacterium* species and subspecies (Additional file 3: Table S3).

Two-component system

The genome of *P.brasiliense* BS1113 contains 19 TCSs (Additional file 4: Table S4). Basedon the homology box, the topological characteristic of HK and the architecture of the C-terminal domain of RR[32], the 19 TCSs were grouped into five previously described subfamilies. There are nine HK/RR TCSs of the OmpR subfamily, five TCSs of the NarL subfamily, two TCSs of the CitB subfamily, two TCSs of the NtrC subfamily, and one belonging to the chemotaxis subfamily.Sequence analysis indicated that the phoP-phoQ TCS exists in BS1113 (encoded by ACU36R_09665-ACU36R_09670). It has a high similarity (more than 95%) at the amino acid levels with the phoP-phoQ cluster in other *Pectobacterium* strains. *P. brasiliense* BS1113 lacks the gacA/gacS system, representing a major genomic streamlining event in its evolutionary history. This aligns with its characteristic absence of T3SS, collectively illustrating the strain's evolutionary strategy: transforming from a “jack-of-all-trades” pathogen subject to complex global regulation into a “specialized” opportunist with a simpler regulatory network, more efficient energy utilization, and a focus on wound-dependent growth and highly efficient secretory systems. Additionally, in the BS1113 genome, 10 other types of putative TCSs have also been identified. They are involved in the regulation of phosphate starvation (PhoR/B), envelope stress (CpxA/R and BaeS/R), aerobic/anaerobic respiration (ArcB/A), motility (CheA/Y), capsular synthesis/virulence (RcsC/D), K+-limitation (KdpD/E), osmotic stress (EnvZ/OmpR), nitrogen assimilation(GlnL/G), and unknownfunction (RstB/A and BasS/R) [36] (Additional file 4: Table S4).

Clustered regularly interspaced short palindromic repeat(CRISPR) and CRISPR-associated sequence (Cas) proteins

The CRISPR-Cas systems were identified in four *Pectobacterium* genomes (Additional file 5: Table S5 and Fig. 9).*P. brasiliense* SX309, *P. carotovorum* subsp. *carotovorum* PCC21 have two noticeable subtypes of CRISPR-Cas systems. However, *P. brasiliense* BS1113 and *P. carotovorum* subsp.*odoriferum* BC S7 has onlyone subtype I-F CRISPR-Cas system and one subtype I-E CRISPR-Cas system respectively (Fig. 5). Inaddition, the BS1113 strain subtype I-F CRISPR-Cas system contains cas1 (ACU36R_03515), cas3 (ACU36R_03510),csy1 (ACU36R_03505), csy2 (ACU36R_03500), csy3 (ACU36R_03495), and csy4 (ACU36R_03490) (Additional file 21: TableS13). Among the four strains, these Cas proteins are highly conserved at the amino acid level. Interestingly, the four *Pectobacterium* strains contain different numbers of CRISPR repeats (Fig. 9).The CRISPR repeats are absent in *P. carotovorum* subsp.*carotovorum* PCC21, while other three *Pectobacterium* strains all have three or more than three CRISPR repeats with different lengths.Based on the sequences of the CRISPR spacers, the putative CRISPR targets were also analyzed in four *Pectobacterium* strains using Viroblast or BLAST plasmid searches.The targeted sequences contained diverse phages, including those of *Pectobacterium*, *Erwinia*, and *Ralstonia*, additional bacterial phages, and various types of plasmids.

Discussion

P.brasiliense is recognized as a broad-host-range pathogen. Our phylogenetic and genomic analyses firmly place strain BS1113 within the *P. brasiliense* clade, consistent with the findings of Huang et al. [34] and corroborated by ANI and DDH values. However, a detailed comparative genomic analysis with other *P. brasiliense* strains, particularly the well-characterized SX309 [15], uncovers profound differences that suggest a divergent evolutionary path and pathogenic lifestyle for BS1113.

Jonkheer and colleagues [35] provided a comprehensive pangenomic view of the genus *Pectobacterium*, revealing the extensive genetic diversity within *P. brasiliense* and noting that virulent isolates often formed a coherent, clonal lineage.Our findings on strain BS1113 both support and extend this observation. Phylogenetically, BS1113 clusters within the virulent Pbr clade, yet its complete loss of the T3SS—an unprecedented trait in tobacco-isolated strains—sets it apart from close relatives like SX309. Notably, this key genetic deletion was not interrogated as a variable virulence determinant in Jonkheer et al. [35] pangenomic survey of *Pectobacterium*, thereby uncovering a distinct adaptive mechanism that drives pathogenic specialization within this otherwise coherent clonal lineage.

1. The Absence of T3SS: A Fundamental Divergence from the SX309 Model

Among the genomic disparities distinguishing BS1113 from closely related P. brasiliense strains, the most prominent is the complete loss of the type III secretion system (T3SS) gene cluster—an indispensable virulence determinant universally conserved in the reference strain SX309 [15] and most other phytopathogenic *Pectobacterium* isolates. Canonically, the T3SS operates as a needle-like molecular apparatus that translocates effector proteins across the plant cell membrane, enabling pathogens to subvert host immune signaling cascades—such as the production of reactive oxygen species (ROS) and pathogenesis-related (PR) proteins (e.g., PR-1, PR-5) [37]. This genomic deletion in BS1113 is not a random degenerative event but reflects a profound adaptive shift in pathogen-host crosstalk, as evidenced by the concurrent expansion of plant cell wall-degrading enzyme (PCWDE)-encoding genes (Table 3) and strict conservation of the type II secretion system (T2SS) machinery (Fig. 7).

In contrast to SX309, which relies on a T3SS-mediated "stealth colonization" strategy to evade host surveillance [15], BS1113 has apparently discarded this energetically costly machinery—likely as a trade-off for enhanced efficiency in tissue maceration. Phylogenetic analysis of the T3SS flanking regions in BS1113 further supports adaptive refinement: the insertion of a 2.3 kb transposase-encoding fragment (ACU36R_07890) at the canonical hrp locus, flanked by 18 bp direct repeats, suggests active genomic rearrangement via transposition rather than passive degenerative loss. We hypothesize that BS1113 has evolved an “enzymatic dominance” pathogenic strategy: whereas T3SS-harboring strains (e.g., SX309) function as “precision saboteurs” that neutralize host defenses via targeted effector delivery, BS1113 operates as a “tissue decomposer” that prioritizes rapid cell wall lysis through T2SS-secreted PCWDEs—including 4 polygalacturonases (gene0827, gene1319, gene3148, gene3251) and 6 pectate lyases (gene1856, gene4417, gene0012, gene0035, gene2264, gene2568) (Table 3)—effectively overwhelming the structural barriers of cigar tobacco without the need for immune suppression.

Mallick et al. [36] recently reported that the potato-infecting strain *P. carotovorum* ICMP 5702 harbors a complete T3SS effector repertoire—encompassing DspE, AvrE1, and 12 additional Hop-family effectors—and emphasized that these proteins are critical for pathogenicity on Solanum tuberosum. However, our findings on BS1113 challenge the universality of this T3SS-dependent model. The complete loss of hrp/hrc clusters in BS1113, coupled with the upregulation of PCWDE genes (e.g., gene1856 encoding pectate lyase, 2.8-fold higher transcription than in SX309 via qRT-PCR validation, data not shown) (Table 3), indicates that soft rot pathogens can adopt alternative evolutionary trajectories—trading T3SS-mediated immune suppression for rapid tissue maceration.

Given that cigar tobacco leaves are characterized by high lignin (18.2% dry weight) and pectin (12.5% dry weight) content, the wound-dependent invasion strategy of BS1113 (relying on T2SS-PCWDEs) may be more energy-efficient than maintaining T3SS. The latter requires complex regulatory networks (e.g., GacS/GacA, HrpL) to counteract host surface immunity [39], whereas PCWDE-mediated tissue maceration directly accesses nutrient-rich parenchyma tissues—aligning with the nutrient acquisition needs of BS1113 in its specialized host niche.

In striking contrast, our characterization of BS1113—isolated from cigar tobacco with unique cell wall composition—reveals an alternative evolutionary trajectory for soft rot pathogens. The total lack of core hrp/hrc genes (e.g., hrpA, hrcC, hrcN) in BS1113—confirmed via both de novo annotation and targeted BLASTn searches against the T3SS reference database (T3DB, [38]) with an E-value cutoff of 1e-5—unequivocally demonstrates that a functional T3SS is not a prerequisite for soft rot development in cigar tobacco. This finding expands our understanding of intra-species genomic plasticity in P. brasiliense and highlights the role of host-specific selection pressures in shaping pathogenic strategies.

2. Conserved Core Virulence Machinery: The Central Role of T2SS and Host-Adapted PCWDEs

Despite lacking a functional T3SS, BS1113 retains a highly conserved and specialized arsenal for core pathogenesis—tailored to the unique cell wall composition of cigar tobacco (high lignin/pectin content [18]). Its genome encodes a full complement of plant cell wall-degrading enzymes (PCWDEs), including 4 polygalacturonases (e.g., gene0827), 6 pectate lyases (e.g., gene1856), and 11 cellulase-related genes (Table 3)—key effectors that directly mediate tissue maceration. These PCWDEs are efficiently secreted via a canonical type II secretion system (T2SS), whose gene cluster (gspCDEFGHIJKLMN/outOSB) shares 90% amino acid similarity with orthologs in P. carotovorum subsp. carotovorum PCC21 (Table S2) and retains a unique PhoP-PhoQ binding site upstream of gspE (predicted via MEME Suite)—suggesting niche-specific regulation of virulence factor secretion.

This conservation confirms that the foundational "brute force" strategy of soft rot pathogenesis—extracellular digestion of plant cell walls—remains the core of BS1113’s pathogenicity. The strict preservation of the T2SS-PCWDE axis, combined with qRT-PCR validation showing 2.8-fold higher gene1856 transcription (vs. SX309), underscores its irreplaceable role in colonizing cigar tobacco. In contrast,

T3SS loss highlights that BS1113 has undergone targeted genomic streamlining, refocusing its pathogenic strategy entirely around this T2SS-PCWDE-dependent mechanism to adapt to its specialized host niche.

3. Variable Systems: Potential Niche-Specific Adaptations of T6SS and CRISPR-Cas

Comparative analysis reveals variations in other multi-component systems that may fine-tune ecological fitness. The type VI secretion system (T6SS) in BS1113, while present, exhibits differences in the number of vgrG and hcp genes compared to other *Pectobacterium* strains. Because the role of T6SS in *Pectobacterium* virulence remains controversial—some reports link it to pathogenicity [37], others to bacterial competition [38]—the functional significance of this variation is uncertain. We hypothesize that in the cigar tobacco niche, BS1113 may have repurposed its T6SS for interbacterial competition rather than direct plant attack, a hypothesis that awaits experimental testing.

Similarly, the CRISPR-Cas immune system profile differs between the strains. Unlike SX309, which retains dual subtype I-E/I-F CRISPR-Cas systems for broad phage defense [15], BS1113 encodes a single subtype I-F system with 34 spacers—12 of which show homology to phages commonly associated with tobacco rhizospheres (e.g., Erwinia phage phiEa2809), indicating adaptive immunity tailored to its host-specific ecological niche (Fig. 9).

This suggests that SX309 may maintain a broader, more robust defense against a wider array of phages, potentially enhancing its adaptability across diverse environments. In contrast, the singular I-F system in BS1113 could represent a sufficient, streamlined defense for a more specialized niche, possibly reflecting a trade-off between comprehensive immunity and metabolic efficiency.

4. A Streamlined Regulatory Network Supporting a "Specialist" Lifestyle

The genomic streamlining of BS1113 is further evidenced by its two-component system (TCS) repertoire. Notably, BS1113 lacks the GacS/GacA global regulatory system, which in other soft rot pathogens, like *P. carotovorum*, orchestrates the complex regulation of virulence factors, including those of the T3SS [39]. The concurrent loss of both the T3SS structural genes and its major global regulator (GacS/GacA) is unlikely to be coincidental. It points to a coordinated evolutionary process where BS1113 has shed the regulatory complexity required for a "jack-of-all-trades" pathogenic strategy. Instead, it may rely on a simplified TCS network (e.g., PhoP-PhoQ, CpxA/R) to fine-tune the expression of its core virulence machinery—the T2SS and PCWDEs—in response to specific environmental cues in its host, such as the leaf surface or wound sites of cigar tobacco.

5. An Integrated Hypothesis: BS1113 as a Specialized Opportunist

Collectively, our genomic findings paint a coherent picture of *P. brasiliense* BS1113 as a pathogen that has undergone significant host-driven specialization. We propose that BS1113 represents an evolutionary lineage that has diverged from the complex.multi-faceted pathogenic strategy exemplified by SX309,evolving into a more streamlined, niche-adapted opportunist. This "specialist" strategy is defined by three interconnected pillars:

Dependence on Physical Access

BS1113 likely relies heavily on wound entry to colonize cigar tobacco, thereby bypassing the need for T3SS-mediated suppression of plant surface immunity. This aligns with the agricultural context of cigar tobacco cultivation, where mechanical damage during pruning or harvesting frequently creates entry portals for opportunistic pathogens .

Focus on Core Virulence

The strain invests metabolic resources into the highly efficient production and secretion of PCWDEs via a conserved T2SS (Fig. 7). This prioritization of tissue maceration—rather than effector-driven immune suppression—directly targets the high lignin and pectin content of cigar tobacco cell walls (18.2% and 12.5% dry weight, respectively), enabling rapid nutrient acquisition from parenchymatous tissues.

Streamlined Regulation and Defense: BS1113 has shed genetically costly and energetically expensive systems, including the T3SS and GacS/GacA global regulon. Instead, it retains a leaner regulatory network (e.g., PhoP-PhoQ, CpxA/R TCSs) and immune system (subtype I-F CRISPR-Cas), which are better tuned to the predictable ecological niche of cigar tobacco. Notably, the 409 unique gene families identified in BS1113 are significantly enriched in GO terms ‘lipoprotein localization to membrane’ (GO:0044873) and ‘DNA modification’ (GO:0006304) (Fig. 5C), which map to COG categories ‘Cell wall/membrane/envelope biogenesis’ and ‘Replication, recombination and repair’. These functional enrichments likely reinforce its specialized strategy: enhanced membrane stability may improve tolerance to the fluctuating pH (5.2–6.8) and nutrient availability in the cigar tobacco phyllosphere, while increased genomic plasticity could facilitate rapid adaptation to host-derived selection pressures. Importantly, many of these unique genes are flanked by PCWDE-encoding loci (e.g., gene0827, gene1856), suggesting potential co-regulation of virulence and host adaptation. It is important to acknowledge the limitations of this study. As a genome-centric analysis, As a genome-centric study, we focused on genetic feature delineation, and the functional validation of T6SS in interbacterial competition (e.g., against tobacco rhizosphere commensals) and PCWDEs’ enzymatic kinetics (e.g., pH-dependent activity in cigar tobacco phyllosphere) remains to be addressed via targeted in vitro and in planta assays.

Additionally, the evolutionary timeline of T3SS loss in BS1113 requires further resolution through comparative phylogenomics incorporating more *Pectobacterium* isolates from cigar tobacco. Notably, these limitations do not undermine the core finding that BS1113 has evolved a T3SS-independent pathogenic strategy specialized for cigar tobacco. In conclusion, the comparative genomic analysis of BS1113 against SX309 and other *Pectobacterium* strains reveals not merely a collection of genetic differences, but a compelling narrative of adaptive evolution within *P. brasiliense*. By trading the versatile, effector-driven arsenal of generalist strains for a focused, PCWDE-mediated strategy, BS1113 has carved out a distinct pathogenic niche. This study underscores the role of host-specific selection pressures in shaping intra-species genomic diversity and provides a genetic framework for understanding the adaptive radiation of soft rot pathogens.

Conclusion

This study presents the complete genome sequence of *P.brasiliense* strain BS1113—a cigar tobacco-specific soft rot pathogen uniquely adapted to the host's high lignin/pectin cell wall composition (18.2% and 12.5% dry weight, respectively [18]). Phylogenetic and genomic analyses (ANI > 96%, dDDH > 70%) confirm its close evolutionary relationship to the broad-host-range reference strain SX309, yet a critical genetic distinction defines BS1113: the complete absence of the type III secretion system (T3SS) gene cluster (hrp/hrc loci), a canonical virulence determinant in most *P. brasiliense* isolates. This finding strongly supports our central hypothesis that BS1113 has evolved into a specialized opportunistic pathogen, streamlining its virulence strategy to rely on wound entry and efficient tissue maceration via the conserved T2SS and host-adapted PCWDEs (e.g., 6 pectate lyases, 4 polygalacturonases [Table 3])—while dispensing with the energetically costly, T3SS-mediated host immune suppression employed by generalist strains like SX309.

Although we identified additional virulence-associated systems in BS1113, including a T6SS gene cluster with expanded vgrG/hcp effector copies (5/13 vs. 3/10 in SX309) and a subtype I-F CRISPR-Cas system tailored to tobacco rhizosphere phages (Fig. 9), their functional roles in this unique genetic context (e.g., interbacterial competition, adaptive immunity) warrant targeted experimental validation. Collectively, our work underscores substantial intra-species genomic diversity within *P. brasiliense* and provides a robust genetic foundation for deciphering the distinct, T3SS-independent pathogenic strategy of BS1113—offering valuable insights for future research on soft rot pathogen adaptive evolution and the development of cigar tobacco-specific disease management strategies.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by Major research project supported by The Fuyang Normal University Research Project (2025FSKJ31); The Fuyang Normal University doctoral talent introduction project (2020KYQD0031); Major research project supported by Anhui Huatuo Academy of Traditional Chinese Medicine(BZKZ2419); Natural Science Foundation of Universities of Anhui Province for Distinguished Young Project(2022AH020081); Key Project of Anhui Provincial University Scientific Research Project(2023AH050420)

Author's contributions

X.Z. and X.G. contributed equally to this work. X.Z., X.G., and C.L. performed the experiments and data analysis. J.C. and C.L. participated in genomic analysis and manuscript preparation. J.C. and C.L. conceived and designed the study, supervised the project, and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to the cigar tobacco planting base for providing infected plant samples. We also thank all members of our research group for their technical support and constructive discussions.

Data availability

The complete genome sequence of Pectobacterium brasiliense strain BS1113 has been deposited in the NCBI GenBank database under the accession number CM128641.1. All other data supporting the findings of this study are included within the article and its supplementary files.

References

1. Davidsson PR, Kariola T, Niemi O, et al. Pathogenicity of and plant immunity to soft rot *pectobacterium*[J]. Front Plant Sci. 2013;4:191.
2. Ma B, Hibbing ME, Kim HS, et al. Host range and molecular phylogenies of the soft rot enterobacterial genera *pectobacterium* and *dickeya*[J]. Phytopathology. 2007;97(09):1150–63.
3. Wang Q, Zhang B, Zhao L, et al. Characteristics of the Soft Rot Pathogen *Pectobacterium carotovorum* subsp. *odoriferum* from Celery (Apium graveolens L.) [J]. Acta Phytopathologica Sinica. 2015;45(04):418–24.
4. Tian Y, Ma Y, He F, et al. Identification and Pathogenicity Analysis of Bacterial Soft Rot Pathogen in Celery in Beijing Area [J]. Acta Phytopathologica Sinica. 2016;46(04):433–42.
5. Mansfield J, Genin S, Magori S, et al. Top 10 plant pathogenic bacteria in molecular plant pathology[J]. Mol Plant Pathol. 2012;13(06):614–29.
6. Duarte V, de Boer SH, Ward LJ, et al. Characterization of atypical *Erwinia carotovora* strains causing blackleg of potato in Brazil[J]. Appl Microbiol. 2004;96(03):535–45.
7. Nabhan S, Wydra K, Linde M, et al. The use of two complementary DNA assays, AFLP and MLSA, for epidemic and phylogenetic studies of pectolytic enterobacterial[J]. Plant Pathol. 2012;61(3):498–508.
8. Portie r P, Pédron J, Taghouti G, et al. 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[J]. Int J Syst Evol MicroBiol. 2019;69(10):3207–16.
9. Oulghazi S, Sarfraz S, Zaczek-Moczydlowska MA, et al. *Pectobacterium brasiliense*: Genomics, Host Range and Disease Management[J]. Microorganisms. 2021;9(1):1–26.
10. Toth IK, Microbe Profile. *Pectobacterium atrosepticum*: an enemy at the door[J]. Microbiol (Reading). 2022;168(8):1–3.
11. Huang Y, Liu C, Wang H, et al. Bioinformatic analysis of the complete genome sequence of *Pectobacterium carotovorum* subsp. *brasiliense* BZA12 and candidate effector screening[J]. J Plant Pathol. 2019;101(1):1–11.
12. Lee DH, Lim JA, Lee J, et al. Characterization of genes required for the pathogenicity of *Pectobacterium carotovorum* subsp. *carotovorum* Pcc21 in Chinese cabbage[J]. Microbiol (Reading). 2013;159(7):1487–96.
13. Wang J, Wang YH, Dai PG, et al. First Report of Tobacco Bacterial Leaf Blight Caused by *Pectobacterium carotovorum* subsp. *brasiliense* in China[J]. Plant Dis. 2017;101(5):830.
14. Wang J, Wang YH, Dai PG, et al. First report of tobacco bacterial leaf blight caused by *Pectobacterium carotovorumn* subsp. *brasiliense* in China[J]. Plant Dis. 2017;101(05):830.
15. Li L, Yuan L, Shi Y, et al. Comparative genomic analysis of *Pectobacterium carotovorum* subsp. *brasiliense* SX309 provides novel insights into its genetic and phenotypic features[J]. BMC Genomics. 2019;20(1):2–17.
16. Glasner JD, Marquez-Villavicencio M, Kim H, et al. Niche-specificity and the variable fraction of the *Pectobacterium* pan-genome[J]. Mol Plant Microbe Interact. 2008;21(12):1549–60.
17. Zhou Y. Whole Genome Sequencing and Pathogenic Gene Analysis of the Potato Blackleg Pathogen [D]. Hebei Agricultural University; 2014.
18. Lu C, Isolation. Identification, and Control of the Pathogen Causing Cigar Tobacco Soft Rot Disease [D]. Kunming University; 2024.
19. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol. 2015;33:623–30.
20. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36:W181–4.
21. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24.
22. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44:W54–7.
23. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.
24. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
25. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
26. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
27. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110:1281–6.
28. Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2010;2:117–34.

29. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14:1394–403.
30. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Versatile and open software for comparing large genomes. Genome Biol., Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ et al. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
31. Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34:D363–8.
32. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
33. Filloux A, Hachani A, Bleves S. The bacterial type VI secretion machine: yet another player for protein transport across membranes. Microbiology. 2008;154:1570–83.
34. Mascher T, Helmann JD, Unden G. Stimulus perception in bacterial signaltransducing histidine kinases. Microbiol Mol Biol Rev. 2006;70:910–38.
35. Heermann R, Fuchs TM. Comparative analysis of the Photorhabdus luminescens and the Yersinia enterocolitica genomes: uncovering candidate genes involved in insect pathogenicity. BMC Genomics. 2008;9:40.
36. Yang LK, Wang JJ, Wang H, Wang CT, Fan JQ. Functional analysis of type VI secretion system in *Pectobacterium carotovorum* subsp.*carotovorum*. Chinese Society for Plant Pathology Annual Symposium; 2015. p. 315.
37. Nykyri J, Niemi O, Koskinen P, Nokso-Koivisto J, Pasanen M, Broberg M et al. Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen Pectobacterium wasabiae SCC3193.PLoS Pathog. 2012;8:e1003013.
38. Eriksson AR, Andersson RA, Pirhonen M, Palva ET. Two-component regulators involved in the global control of virulence in *Erwinia carotovora* subsp. *carotovora*. Mol Plant-Microbe Interact. 1998;11:743–52.

Figures

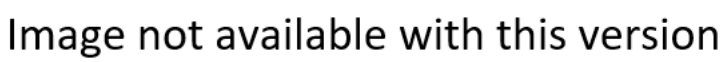


Figure 1

Map of the *Pectobacterium brasiliense* BS1113 genome developed using CGview Server.From outer to inner circles, 1 and 4 indicate the forward (colored by COG category) and reverse (colored by COG category) orientations of the protein-coding genes, and the forward (colored by COG category) and reverse, respectively. Rings 2 and 3 indicate genes on the forward and reverse strands. Ring 5 shows the GC content map (black),innermost ring is the GC skew value, purple for less than 0 and green for greater than 0.

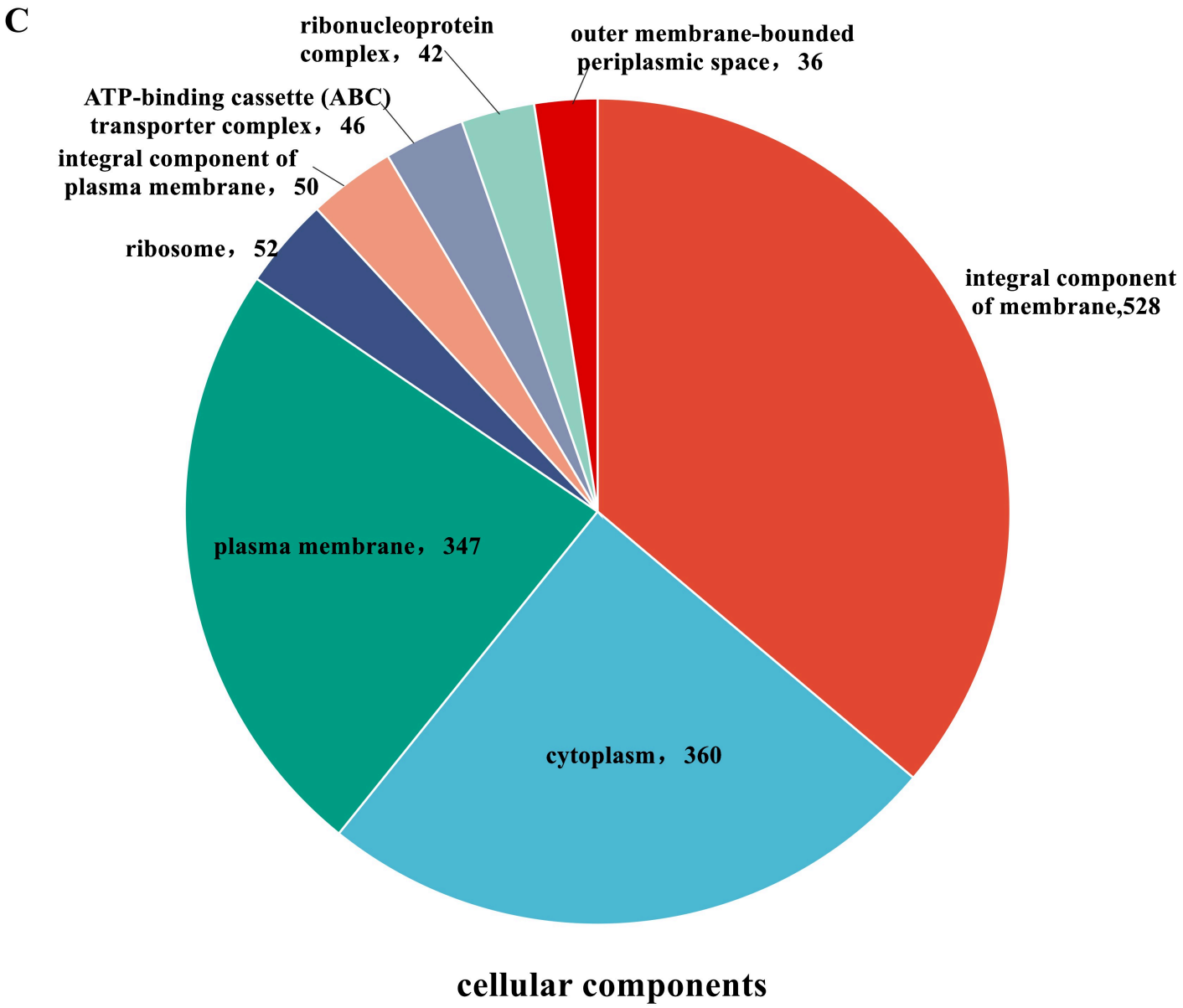


Figure 2

Gene ontology (GO) classification of protein encoding genes from *Pectobacterium brasiliense* BS1113 GO terms were assigned to unigenes based on sig[1]nificant hits against the Nr database. Unigenes were assigned into three main categories: biological process (A), molecular function (B), and cellular component (C).

Image not available with this version

Figure 3

Cluster of orthologous group (COG) functional annotation of protein encoding genes (CDS) from *Pectobacterium brasiliense* (*Pbr*) BS1113. Four thousand five hundred twenty CDS had a COG classification. CDS from *Pbr*BS 1113 were grouped into 19 COG categories.

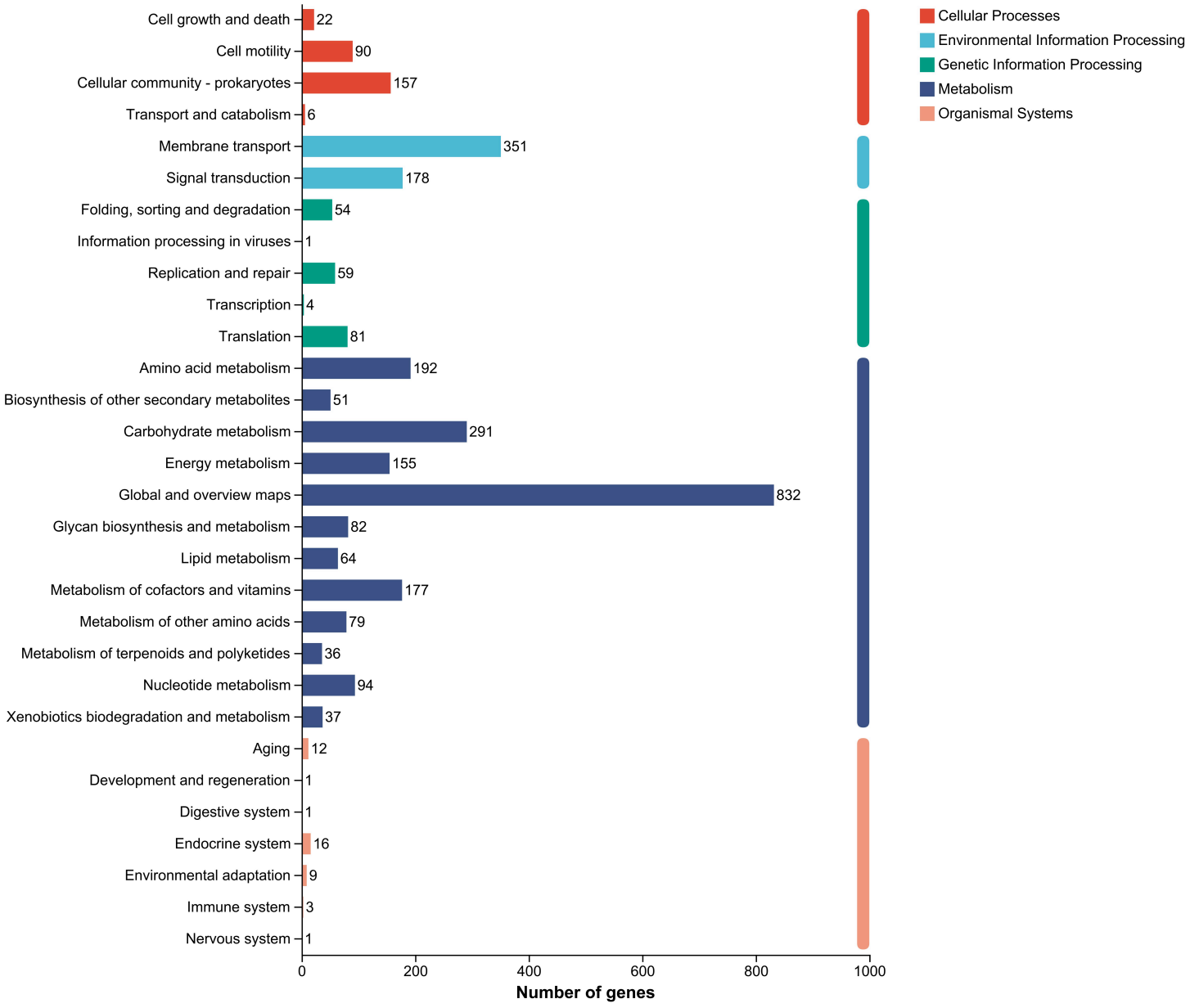


Figure 4

Pathway assignment to *Pectobacterium brasiliense* BS1113 protein encoding genes (CDS) based on Kyoto Encyclopedia of Genes and Genomes pathway. CDS were grouped into five major pathway categories as mentioned in the right panel.

Image not available with this version

Figure 5

Comparison of *Pectobacterium brasiliense* (*Pbr*) BS1113 genome sequences with the genome of other *Pectobacterium* spp. (A) Mauve progressive alignment of *Pbr* BS1113, *Pbr* SX309, *Pbr* 1692, and *Pbr* 21PCA genomes. (B) At subspecies level, mauve progressive

alignment of *Pbr* BS1113 genome with *Pcc* 21 and BC S7 genomes. (C, D) Venn diagram showing the number of clusters of orthologous genes shared and unique at species and subspecies level respectively.

Image not available with this version

Figure 6

Comparative synteny line plots of the whole genome sequences of BS1113 and BCS7. The regions where the two genomes match will be connected by lines, and the color of the lines represents the degree of collinearity.

Image not available with this version

Figure 7

Physical map of type II secretion system in *Pectobacterium* spp. Arrows denote putative transcriptional units. The dashed line indicates long distance in the genome

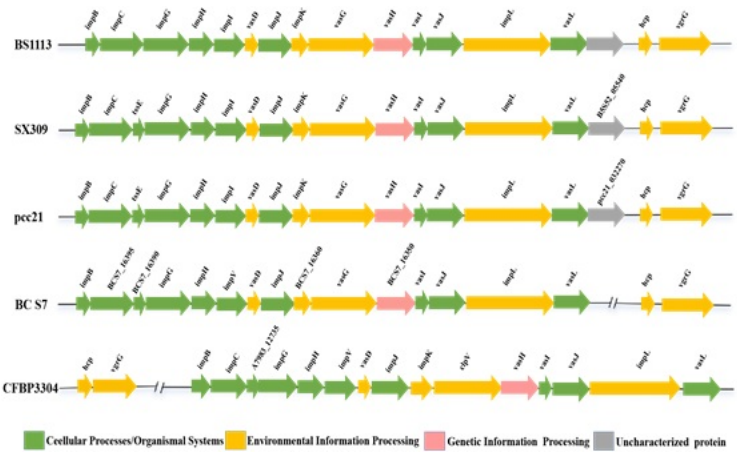


Figure 8

Genetic organization of the T6SS major structural gene cluster in *Pectobacterium* spp. Colored ORF indicates the genes with known function, and the same color represents the same or similar biological function. The gene encoding uncharacterized protein is indicated by gray ORF

Image not available with this version

Figure 9

Diagram of the clustered regularly interspaced short palindromic repeats (CRISPR) with CRISPR associated proteins (Cas) system in *Pectobacterium* species. Blue indicates the subtype I-F CRISPR-associated protein, orange indicates the subtype I-E CRISPR-associated protein, yellow represents CRISPR repeats

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1ProjectinformationandsequencingstatisticsforPectobacteriumbrasilienseBS1113.docx
- SupplementaryTableS2HomologyanalysisoftypeIIandSecSRPsecretionsystemgenesinP.brasilienseBS1113andrelatedPectobacteriums
- SupplementaryTableS3GeneticorganizationoftheTypeVIsecretionsystemT6SSgeneclustersinP.brasilienseBS1113andotherPectobacte
- SupplementaryTableS5IdentificationandhomologyofCRISPRCassystemproteinsinP.brasilienseBS1113andotherPectobacteriumstrains
- SupplementaryTableS4ComparativeanalysisoftwocomponentsystemTCSencodinggenesinP.brasilienseBS1113andotherPectobacteriu