

RESEARCH

Open Access



Pan-genome analysis of *Ralstonia pseudosolanacearum* associated with tobacco bacterial wilt in China

Jia Kang^{1†}, Muhammad Aamir Sohail^{1†}, Yiqian Zhang¹, Zhao Wang¹, Rubin Xu², Tom Hsiang³, Lu Zheng¹, Junbin Huang¹, Yuan Fang^{4*} and Yanyan Li^{2*}

Abstract

Tobacco bacterial wilt, caused by *Ralstonia solanacearum* species complex (RSSC), poses a significant threat to tobacco production in most major tobacco-growing regions of China. Here, 204 strains of *Ralstonia* were isolated from infected tobacco plants across 12 provinces and autonomous regions in China. Molecular identification and biovar typing of the RSSC strains revealed that all the isolates were found in phylotype I (classified as *R. pseudosolanacearum*) and biovar III. Seven sequevars were identified and sequevar 15 was the most prevalent. Pathogenicity tests indicated that 29% of the strains showing high virulence were found in the southwest tobacco-growing region of China. Notably, significant pathogenic variation was observed within strains of the same sequevar, and no clear correlation was found between the sequevar type and pathogenicity. Using Oxford Nanopore sequencing, we analyzed 103 strains of tobacco *R. pseudosolanacearum* from different geographical origins and pathotypes. The pangenome of *R. pseudosolanacearum* is comprised of 9008 non-redundant genes, divided into a core genome (36%), accessory genome (48.7%) and isolate-specific genes (15%). Core genome functions were related to oxidation–reduction reaction process and DNA transcription regulation, while the accessory genome was linked to DNA recombination, integration, and transposition. SNP analysis revealed an average of 36,740 SNP loci per strain, indicating evolutionary purification with k_a/k_s value below 1. Phylogenetic analysis divided 99 strains into 4 main groups, with sequevars correlated to specific branches, reflecting evolutionary relationships. We identified 2226 genomic islands across the strains, with each strain containing 18–25 islands, primarily related to translation regulation, transposition, and transposase activity. Analysis of virulence factors using the virulence factors database (VFDB) highlighted 1252 virulence genes within these islands, with a significant portion (36.5%) linked to effector transport systems, predominantly the type III secretion system (66.7%). Each strain averaged 60 type III effector proteins, with RipBK and RipAZ2 unique to highly pathogenic strains. This study provides a comprehensive understanding of pan-genome of *R. pseudosolanacearum* causing tobacco

[†]Jia Kang and Muhammad Aamir Sohail have contributed equally to this work.

*Correspondence:

Yuan Fang

fy0579@zjnu.cn

Yanyan Li

yanyanli0025@126.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

bacterial wilt in China, providing valuable insights into virulence variation and environmental adaptation of the pathogen.

Keywords Pangenome, *R. pseudosolanacearum*, Tobacco bacterial wilt

Introduction

Tobacco bacterial wilt, caused by *Ralstonia solanacearum* species complex (RSSC), is a soil-borne disease that can significantly reduce tobacco yield and quality. It can cause poverty among farmers, and hinder development of the tobacco industry. The pathogen persists in both soil and water environments, and retains its pathogenicity even without a host plant [1, 2]. Soil contaminated with bacteria serves as the primary infection source, with pathogens easily spreading through irrigation water, farming tools, and crop residues [3]. RSSC can invade host plants through natural or mechanical wounds [4]. Once inside, it damages xylem tissue, leading to leaf yellowing and unilateral wilting, with water-soaked stem spots deepening over time [5]. Tobacco bacterial wilt is widespread across China, affecting tobacco-growing areas in 30 provinces. The disease incidence exceeds 30%, and when combined with black shank disease, it can reach up to 75%. Continuous planting exacerbates disease rates especially in high rainfall areas with elevated field humidity, potentially leading to complete crop failure [6, 7].

Ralstonia solanacearum species complex exhibits a broad host range, capable of infecting over 450 plant species from 50 families [8]. In China, RSSC is classified into five races based on the host plant and symptom severity, with race 1 being predominant. Six biovars have been identified based on the oxidation of three disaccharides (maltose, lactose, cellobiose) and three hexols (mannitol, sorbitol, dulcitol) [9]. Most RSSC strains from China belong to biovar III, but this classification is insufficient for distinguishing strains within biovar II. Therefore, various techniques are essential to study genetic variation to explore the global population dynamics of RSSC. Fegan and Prior [10] proposed phylotypes to classify RSSC into four types, each nearly corresponding to a distinct species or subspecies. Specific primers and phylotype primers distinguish these bacterial strains through PCR amplification. This classification reflects genetic relationships and associates strain phylotypes with geographical origins: Phylotype I (Asian branch); Phylotype II (American branch); Phylotype III (African branch); Phylotype IV (Indonesian clade). A taxonomic revision of *R. solanacearum* by Safni et al. divided the four Phylotypes into three distinct species within the RSSC: (1) all phylotype II strains belong to *R. solanacearum*, (2) *R. pseudosolanacearum* includes all phylotype I and III strains, and (3) *R. syzygii* covers all phylotype IV strains [11, 12].

Based on similarity of endoglucanase genes (*egl*), evolutionary type can be further divided into sequevars. In

China, 14 sequevars of phylotype I have been reported, including variants 12, 13, 14, 15, 16, 17, 18, 34, 44, 48, 54, 55, and two unidentified sequevars [13]. Two sequevars of phylotype II, including sequevars 1 and 7 have also been reported [13]. Sequevars 13, 14, 15, 17, 34, 44, 54, and 55 can infect tobacco [14]. With the widespread application of molecular biology and high-throughput sequencing technology, the cost of obtaining high-quality bacterial genome information has decreased, facilitating research on the bacteria at genome level to uncover more information on biology and pathogenicity. However, extensive structural variation within bacterial genomes and promiscuity among bacterial strains and species can alter the standard genomic make-up for any particular species, highlighting genomic plasticity and genetic diversity among bacteria. Reliance on a single reference genome may overlook structural and genetic variation related to any particular phenotype. Thus, genetic diversity of RSSC associated with tobacco bacterial wilt in China needs to be analyzed by pangenomics.

The concept of pangenome analysis was originally introduced during a genomic study of *Streptococcus agalactiae* [15], with the definition of the entirety of genetic information within a species rather than for a particular or standard genotype. In traditional pangenomic analysis, the core genome comprises genes and gene families shared by an evolving lineage, which are essential for growth and survival, and the additional genomic information (accessory genome) aids in the understanding of species diversity and metabolism [16]. Accessory genomes consist of genes present in one or more genomes, considered dispensable but contributing to species diversity and adaptation due to their high mobility. Research on *Bacillus* pangenomic differences has emphasized horizontal gene transfer (HGT) as a driver of species adaptation to new niches during evolution [17]. A study of 1311 *Pseudomonas aeruginosa* genomes found that gradual evolution and HGT significantly contributed to antimicrobial resistance and virulence mechanisms [18]. A study of 96 prokaryotic genomes revealed that approximately 40% of the genes were associated with metabolic functions, significantly expanding biosynthesis potential and improving population adaptability through intraspecific metabolite exchange [19]. Studies have shown that accessory genomes may evolve faster than core genomes [20], leading to greater environmental selection pressure on affiliated genomes due to frequent gene transfer and acquisition of beneficial genes. A pangenomic study of the genome sequences of 131 strains of the RSSC from

various hosts revealed multiple types of mobile genetic elements, which were crucial for enhancing genomic plasticity and environmental adaptability [21]. In another genomic analysis of eleven RSSC strains, approximately 82% of metabolic pathways were conserved in core metabolic pathways [22]. However, pangenomic function of RSSC associated with tobacco bacterial wilt in China has not been explored.

In this study, a total of 204 *R. pseudosolanacearum* strains were isolated from tobacco bacterial wilt samples collected across 12 provinces and autonomous regions in China. Molecular identification and pathogenicity assays were performed on selected strains to assess sequence variation and pathogenicity differences within and between the bacterial populations. Pangenomic analysis of long-read whole genome sequencing data from 103 bacterial isolates revealed genomic diversity, enabling insights into the genomic structure, functional elements, and SNPs. This allowed further investigation into the relationships between SNP variation and population evolution. Additionally, the study explored the differences in type III effector proteins among pathogenic strains, evaluating the correlation between specific type III effector proteins and pathogenic types. These findings provide a comprehensive understanding of pangenomic characteristics of the *R. pseudosolanacearum* causing tobacco bacterial wilt in China, providing valuable insights into molecular variation, population variation, and establishing a basis for scientific strategies for disease management.

Results

Isolation and identification of *Ralstonia* strains from diseased tobacco

A total of 204 strains were isolated from main tobacco-growing regions across Southeast China (Fujian, Anhui, Zhejiang, Guangdong and Jiangxi provinces), Southwest China (Yunnan, Sichuan, Guizhou and Guangxi Zhuang Autonomous Region), the middle and upper reaches of the Yangtze River (Hubei, Hunan and Chongqing Municipality) and Huanghuai tobacco-producing region (Henan Province) (Table S1). On 2,3,5-Triphenyl-tetrazolium chloride (TTC) medium, colonies of the isolated strains showed low motility and poor light transmission, with a pink center and a milky white periphery, lacking significant protrusion (Fig. S1A). On Nutrient Agar (NA) medium, colonies were also milky white with low motility and no significant protrusion (Fig. S1B). These morphological traits were consistent with typical characteristics of RSSC from tobacco. To confirm the identification, two pairs of species-specific primers (759/760 and Rsol-flic-F/R, Table S2) of RSSC were used for PCR amplification of genomic DNA from the 204 isolates. The amplification yielded bands of 280 bp and 480 bp, respectively (Fig.

S2A, B), confirming that all 204 isolates could be identified as RSSC.

Identification of phylotype, sequevar and biovar of *Ralstonia* strains from tobacco

Multiplex PCR amplification with primers specific to four different phylotypes (Table S2) and species-specific primers 759/760 was performed on the genomic DNA of the 204 *Ralstonia* strains, yielding two bands of 144 bp and 280 bp (Fig. S2C). These results confirmed that all isolates belonged to phylotype I of the Asian origin, which are classified as *R. pseudosolanacearum*.

To analyze the sequevar of these strains, the *egl* sequences of all 204 strains were obtained after PCR amplification and sequencing of genomic DNA using the primers Endo-F and Endo-R (Table S2). A phylogenetic analysis of the *egl* gene sequences from 204 strains, along with 34 reference strains, was performed to construct a phylogenetic tree. The analysis identified seven sequence variants (sequevars), designated as 13, 14, 15, 17, 34, 44 and 54, which constituted approximately 2.5%, 5.4%, 34.3%, 27.5%, 3.9%, 26% and 0.5% of the total strains, and sequevar 15 emerged as the most prevalent (Fig. 1A).

To determine the biovar, these strains were inoculated into six biochemical media containing cellobiose, maltose, lactose, mannitol, sorbitol or dulcitol, and the utilization of these organic compounds was assessed based on oxidation activity and acid production. The results showed that a color change from blue to yellow in the biochemical medium indicated the effective utilization of all six substrates confirming that all 204 *R. pseudosolanacearum* isolates from tobacco belonged to biovar III (Fig. S3).

In sequevar analysis, seven sequevars were identified in the middle and upper reaches of the Yangtze River, which have the largest number of sequevars among different tobacco-growing regions of China. Sequevars 13 and 14 were exclusively identified in the middle and upper reaches of the Yangtze River and the southwest tobacco-growing region, while sequevar 54 was unique to the middle and upper reaches of the Yangtze River (Fig. 1B). Thus, the distribution of sequevars varied significantly among provinces and autonomous regions.

R. pseudosolanacearum isolates from tobacco exhibit high variability in virulence

Based on ecological characteristics of different tobacco-growing regions, as well as location, 63 isolates were randomly selected for pathogenicity testing. Each isolate was inoculated onto three tobacco varieties, Honghuadajinyuan (susceptible), K326 (moderately resistant) and Yanyan 97 (resistant), and the area under the disease progress curve (AUDPC) was calculated for each isolate on all three tobacco varieties based on the disease

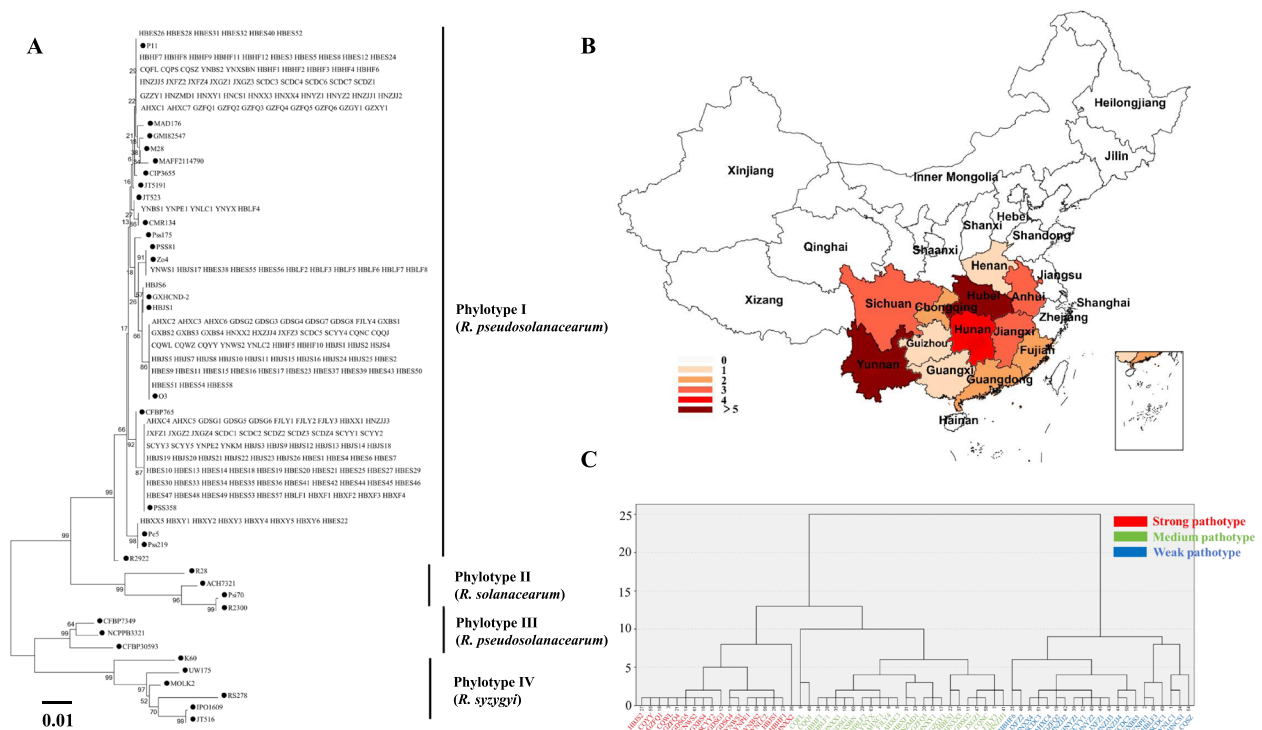


Fig. 1 Genetic analysis of RSSC isolates causing tobacco bacterial wilt in China. **A** Phylogenetic tree of the 204 isolates of RSSC based on partial *egf* gene sequences. The tree was generated using the Neighbor-Joining (NJ) method in MEGA version 10.0.5. Symbol "●" represents reference strains, and the other nodes represent the 204 isolates of RSSC from tobacco; **B** The number of sequevars in each province and autonomous region. Different colors indicate the abundance of sequevars, with darker colors signifying a greater abundance. This distribution reflects the genetic diversity of RSSC across different regions; **C** Systematic clustering of different pathotypes of RSSC isolates. In pathogenicity assays, the 63 strains of RSSC were inoculated onto three varieties of tobacco (Honghuadajinyuan, K326 and Yanyan 97). Based on disease index of the three varieties of tobacco on the 10th, 12th, 15th, 20th, 25th, and 30th days, the Area Under the Disease Progress Curve (AUDPC) was calculated. Colors indicate pathotype strength: red for strong pathotype, green for medium pathotype, and blue for weak pathotype

incidence at these time points (Table S3). The virulence of 63 *R. pseudosolanacearum* strains was then analyzed by systematic clustering based on AUDPC values (Fig. 1C). Pathogenic types were classified into three categories: strong, medium, and weak.

The weakly pathogenic type (W), with 21 strains (33.3% of the total) was isolated from tobacco-growing regions in southeast, southwest, and the middle and upper reaches of Yangtze River. The weakly pathogenic strains inoculated into Honghua Dajinyuan resulted in a disease index ranging from 41.7–100, with an average of 81.9. The AUDPC values were between 7.2–19.5, with an average of 11.7. In K326, the weakly pathogenic strain produced a disease index ranging from 20.8–88.9, with an average of 66.8, and AUDPC values between 2.4–14.2, with an average of 9.3. Finally, in Yanyan 97, the weakly pathogenic strain resulted in a disease index ranging from 0–52.8, with an average of 27.1, and the AUDPC values from 0–13.0, with an average of 3.3.

Medium pathogenic strains (M), comprising 24 strains (38.09%), were collected from southeast, southwest, Huanghuai area, and middle and upper reaches of the Yangtze River. The moderately pathogenic strain

inoculated into Honghua Dajinyuan resulted in a disease index ranging from 79.2–100, with an average of 97.2, and AUDPC values were between 13.5–21.3, with an average of 17.7. When inoculated onto K326, the moderately pathogenic strain produced a disease index of 34.7 to 100, with an average of 87.4, and the AUDPC values ranged from 11.7 to 21.3, with an average of 15.5. For Yanyan 97, disease index for the moderately pathogenic strain varied between 3.9–100, averaging 50.5, with AUDPC values ranging from 1.8–13.0, with an average of 7.0.

Strongly pathogenic type (S), totaling 18 (28.57%), also originated from these regions. The disease index for the strongly pathogenic strain inoculated into Honghua Dajinyuan ranged from 97.2–100, with an average of 99.8. The corresponding AUDPC values ranged from 18.6–22.1, with an average of 20.2. For the highly pathogenic strain inoculated into K326, the disease index ranged from 83–100, with an average of 97. The AUDPC values were between 12.5 and 21.4, with an average of 18.8. In the case of Yanyan 97, the disease index for the highly pathogenic strain ranged from 59.7–100, averaging 85.7,

and the AUDPC value spanned 10.7–19.5, with an average of 14.1.

All three pathogenic types were present in the south-east, southwest, and middle and upper reaches of the Yangtze River. Notably, 61.11% (11/18) of the strong pathogenic strains were isolated from the southwest tobacco-growing areas, indicating that highly pathogenic strains dominate these regions.

Genomic characteristics of *R. pseudosolanacearum* isolates from tobacco in China

After genome sequencing and quality control, the assembled clean data yielded the complete genome of tobacco *R. pseudosolanacearum*. The genome sizes of the 103 *R. pseudosolanacearum* strains ranged from 5,551,041 bp to 6,041,745 bp. The GC content spanned a range from 66.8% to 67.2%, with an average of 66.9%. The genome

consisted of a chromosome and a megaplasmid. The GC content of chromosome averaged 66.8%, while the megaplasmid had an average GC content of 67.1%. Coding gene counts ranged between 4,778 and 5,724 per strain. The gene-coding region accounted for an average of 85.8% of the entire genome, ranging from 84.8% to 86.2%. Average Nucleotide Identity (ANI) analysis was performed between the reference genome GMI1000 and the genome sequences of 103 tobacco *R. pseudosolanacearum* strains. An ANI above 95% served as the standard for classifying them as the same species. Our findings showed that ANI values of all 103 strains exceeded 98% when compared to GMI1000 (Fig. 2A, Table S4), as well as among themselves, confirming that all tested strains belong to the same species as GMI1000.

The genomes of 103 tobacco *R. pseudosolanacearum* strains varied in size, with the largest being HBES4

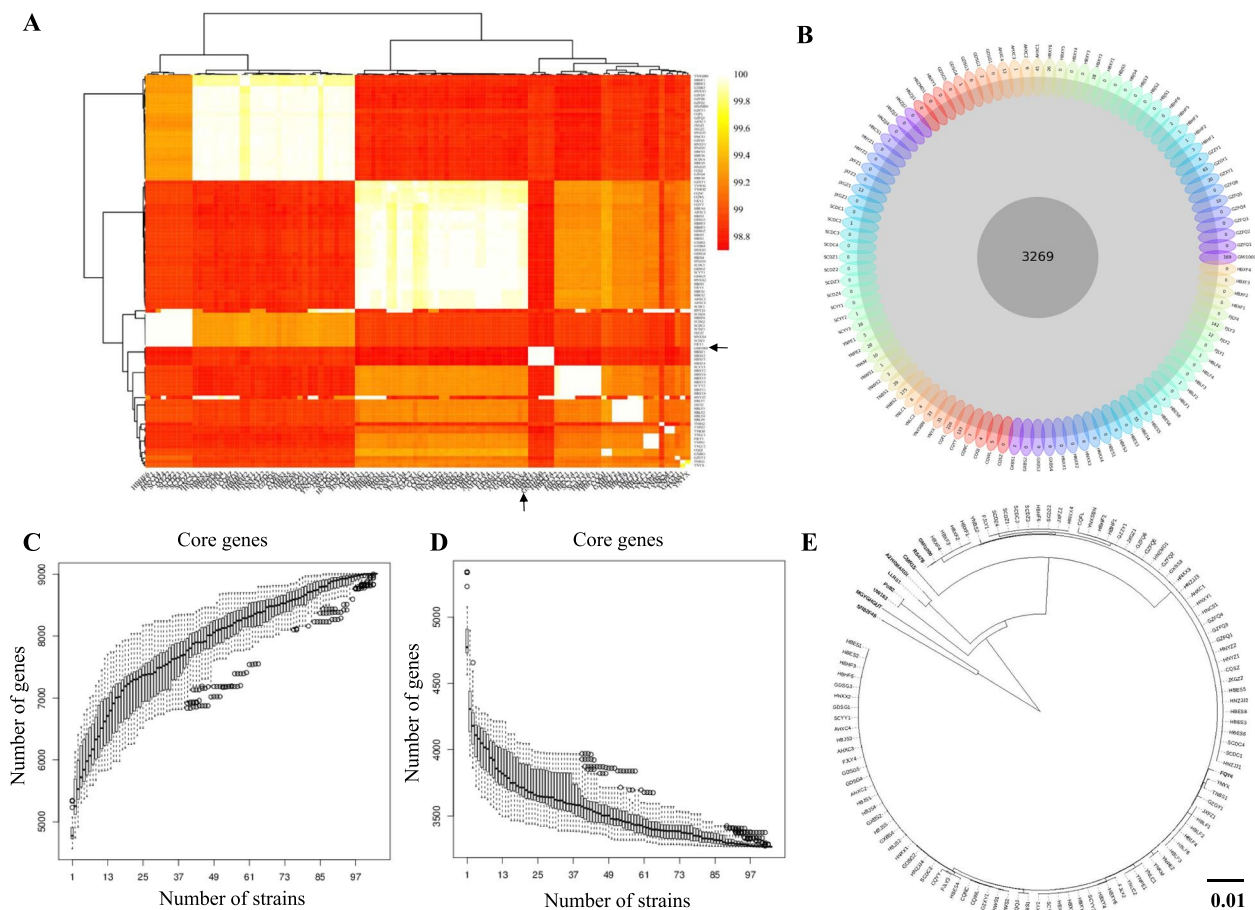


Fig. 2 Pan-genomic analysis of *R. pseudosolanacearum* isolates causing tobacco bacterial wilt in China. **A** Heat map of Average Nucleotide Identity (ANI) comparative analysis of the 103 *R. pseudosolanacearum* strains with reference strain GMI1000. Red represents the lowest average nucleotide consistency of 98.8% among the strains of *R. pseudosolanacearum*. When the color is lighter, the average nucleotide consistency among strains is higher. An arrow indicates the reference strain GMI1000; **B** Pan-genomic structure of the population of the *R. pseudosolanacearum* isolates from tobacco. The pangenome flower plot shows the core genome and unique genes for each strain, with different colors corresponding to different strains; **C** Gene accumulation curves for the pangenome; **D** Core gene solution curves for the pangenome. The calculative sizes were determined by selecting strains without replacement in random order for 1,000 times; **E** Phylogenetic tree constructed using the core genome by the Neighbor-Joining method with 5,000 bootstraps. The black bold lines represent 10 reference isolates from different species in the *R. solanacearum* species complex

(6,041,745 bp), isolated from the middle and upper reaches of the Yangtze River tobacco-growing region, and the smallest being JXFZ1 (5,551,014 bp) isolated from the Southeast tobacco-growing region. The size difference between these two genomes was 490,740 bp, accounting for over 8% of a whole genome. Even among strains isolated from the same field, significant differences in genome size were observed. Among the seven strains isolated from Enshi County, Hubei Province, the genome size of HBES4 was the largest among all isolates. The genome sizes of the remaining six strains from that county were ranked 43, 44, 47, 48, 50 and 78 among the 103, with a maximum variation of 239,391 bp between them. These results indicate that genome sizes and structures of *R. pseudosolanacearum* isolates associated with tobacco bacterial wilt are diverse in different locations of China.

Pangenomic structures of tobacco *R. pseudosolanacearum*

In Orthologous gene cluster analysis of the 103 strains, the composition of core genome, accessory genome, and strain-specific genes of tobacco *R. pseudosolanacearum* was determined (Fig. 2B). The core genome consisted of conserved genes and gene families present in all bacterial isolates (3,269 non-redundant genes), while the accessory genome contained genes shared by at least two isolates (4,385 genes), and strain-specific genes, unique to individual strains, accounted for 1,354 genes. Strain-specific genes along with the accessory genome form the dispensable genome. When combined together, there were 532,461 genes across all 103 genomes, but the pangenome consisted of 9,008 unique genes. Among these isolates, CQFL from Sichuan Province contained the most strain-specific genes (226), followed by YNLIC1 in Yunnan Province with 175 specific genes. The core genome accounted for 36.3% of the pangenome, while the dispensable genome made up 63.7%, with 48.7% attributed to the accessory genome, and 15.0% to strain-specific genes.

By fitting the pangenomic accumulation curve and core genome reduction curve of tobacco *R. pseudosolanacearum*, we found that both curves conformed to Heap's law ($n = kN^y$) pangenomic model: addition of new sequenced bacterial isolates to the pangenome resulted in a continuous increase in non-redundant gene set (Fig. 2C). When the genome of the 25th newly sequenced strain was added, the pangenome exhibited a logarithmic increase in non-redundant genes, maintaining a rapid growth. Subsequently, as more genomes were included up to all 103 strains, the growth rate of non-redundant genes slowed but continued to show an increasing trend, confirming the open nature of the pangenome.

With continuous addition of newly sequenced genomes, most initially conserved core genes were still

found to be present across all strains. As more *R. pseudosolanacearum* genomes were added, some homologous genes present in earlier genomes, transitioned from core genes to accessory genes as they were no longer found in all strains. The number of core genes decreased to 3,269 upon inclusion of all 103 strain genomes (Fig. 2D). This indicated that tobacco *R. pseudosolanacearum* strains from different locations in China might exhibit high genetic variability.

To clarify the taxonomic status of tobacco-infecting *Ralstonia* strains in China, the genomes of 10 reference strains representing different phylotypes within RSSC, along with 103 tested strains of tobacco-infecting *Ralstonia* strains were subjected to phylogenetic analyses. A core genome phylogenetic tree was constructed based on 113 single-copy core genes of *R. pseudosolanacearum*, revealing two distinct branches. The first branch was comprised of the 103 tobacco *Ralstonia* strains along with reference strains GMI1000, RS 476, and FQY_4, all classified as *R. pseudosolanacearum*. The second branch included the strains CMR15, Po82, and UW163, classified as *R. solanacearum*, along with LLRS1 (*R. syzygii*), MGYGHGUT (*R. pickettii*), and SN82F48 (*R. mannitolilytica*) (Fig. 2E).

Functional annotation of the pangenome of tobacco *R. pseudosolanacearum*

To investigate the primary functions of the core genome, accessory genome, and specific genes of *R. pseudosolanacearum*, COG functional annotation was conducted for each component of the pangenome (Fig. 3A). A total of 9,008 genes within the tobacco *R. pseudosolanacearum* pangenome were categorized into 20 COG functional groups out of a maximum 21 categories. The core genome was associated with 13 functional categories, primarily related to translation, ribosome structure and biosynthesis. In contrast, the accessory genomes were linked to 11 functional categories, and also centered around translation, ribosome structure and biosynthesis. Notably, a higher proportion of genes in the core genome were annotated for these functions compared to those in the entire accessory genome. Specific genes were classified into 20 functional categories, with predominant functions including replication, recombination, and repair, translation, functional correlation of mobile elements (prophage and transposon).

The GO function annotation of 9,008 genes in the pangenome of *R. pseudosolanacearum* was performed (Fig. 3B), and the results were classified into three categories: biological processes, cellular components, and molecular functions. The core genome was primarily associated with biological process such as transcriptional regulation, transport, and signal transduction. Additionally, cell membrane and intracellular components

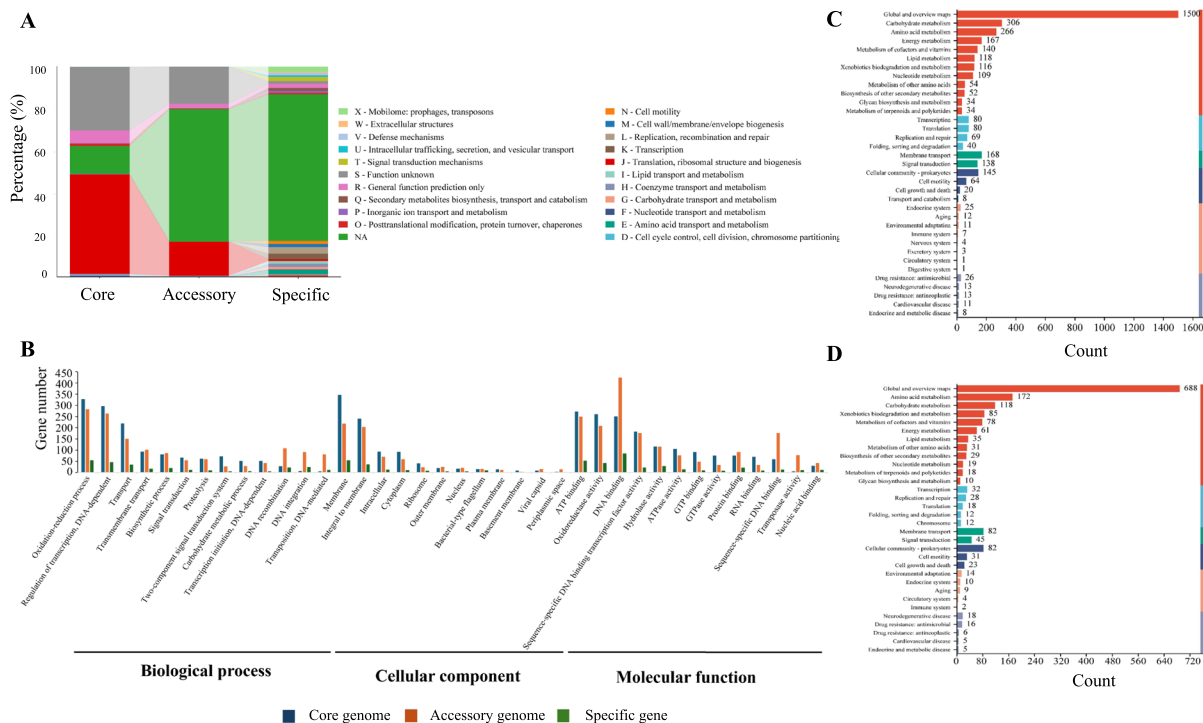


Fig. 3 Function annotation of core, accessory and specific genes in genomes of *R. pseudosolanacearum* isolates causing tobacco bacterial wilt in China. **A** COG function annotation of core, accessory, and specific genes. The Clusters of Orthologous Groups (COG) database was used to annotate the functions of these genes, providing insights into their roles in various biological processes; **B** GO function annotation of core, accessory, and specific genomes. The Gene Ontology (GO) database was used to annotate the core, accessory, and specific genomes. The annotations were categorized under Biological Processes, Cellular Component, and Molecular Functions, with a maximum of 10 terms displayed for each category; **C** KEGG pathway function annotation of core, accessory, and specific genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to annotate the core and accessory genomes. The annotations were categorized under Metabolism, Genetic Information, Environmental Information, Cellular Information, Organismal System, and Human Diseases, with a maximum of 10 pathways displayed for each category

were found to be involved. For molecular functions, ATP binding was linked to oxidoreductase activity. The accessory genomes was associated mainly with DNA recombination, integration, and transposition in biological processes. It also included functions related to the cell membrane and inner membrane. Molecular functions showed correlations with the binding of specific sequence DNA and transcription factor binding activity. The specific genes were primarily linked to the functions of redox processes and transcriptional regulation within biological processes. The functions of cell components such as membrane and intima are related. Functions related to membranes and intracellular components were also observed. Additionally, molecular functions revealed correlations with the binding of specific sequence DNA and the binding activity of specific sequence transcription factors. More genes related to DNA recombination, integration and transposition were annotated in the accessory genome, probably associated with DNA rearrangement and gene sequence transposition caused by transposons occurring more frequently in the accessory genome. This higher frequency of transposon activity

may explain the increased mutation rate in the accessory genome of tobacco *R. pseudosolanacearum*.

The pangenome was subjected to KEGG functional annotation of all 9,008 genes (Fig. 3C, D). The results showed that both the core and accessory genomes were linked to metabolism, genetic information processing, environmental signal recognition, cell components, biological systems, and human diseases. The core genome's KEGG annotation highlighted carbohydrate metabolism, amino acid metabolism, energy metabolism, lipid metabolism and secondary metabolites synthesis under "metabolism", signal transduction and membrane transport under "environmental signal processing" and cellular community under "cellular process" (Fig. 3C). Moreover, the KEGG annotation of the accessory genome focused on amino acid metabolism, carbohydrate metabolism, and xenobiotics biodegradation and metabolism within "metabolism", membrane transport within "environmental signal processing" and cell mobility and prokaryotic community within the "cellular process" (Fig. 3D).

Analysis of SNP loci of tobacco *R. pseudosolanacearum*

Using the GMI1000 genome as reference, the genomes of 103 strains were each compared to the reference genome, yielding comprehensive SNP loci data. A total of 3,645,466 SNPs were predicted, with an average of 36,740 SNPs per strain. Among these, 84.73% of SNPs were located in protein-coding genes, and 57.49% of these mutations were synonymous (Fig. 4A, B). Further analysis of SNPs frequency within coding genes revealed that most coding genes contained fewer than 50 SNPs (Fig. 4C). The Ka/Ks ratio for nearly all coding genes was less than 1, suggesting that the genome of *R. pseudosolanacearum* primarily underwent purifying selection (Fig. 4D).

Among 103 strains of *R. pseudosolanacearum*, 99 strains were selected to construct the SNP phylogenetic tree based on their genome SNPs loci (Fig. 4E). The topology of the phylogenetic tree revealed that all isolates clustered into four groups, with similar sequence variants defining each group. The red, green, purple and blue groups were primarily represented by sequevars 17, 15, 34 and 44, respectively. The genetic evolutionary relationship among *R. pseudosolanacearum* strains, as reflected by the SNPs phylogenetic tree exhibited high similarity with the sequevars identified by the *egl* gene. However, sequevars 15 and 44 were also present in a branch predominantly composed of sequevar 17, and sequevars 14,

15, and 17 were found within a branch largely formed by sequevar 44.

Identification of genomic islands (GIs) from tobacco *R. pseudosolanacearum*

Genomic islands (GIs), which are mobile genetic elements, play a crucial role in the adaptive evolution of bacterial populations and the acquisition of genes related to survival. A total of 2,226 GIs were predicted in 103 strains of *R. pseudosolanacearum*, with each strain containing an average of 22 GIs. These islands ranged in length from 3,144 bp to 59,034 bp, with an average length of 14 kb and an average GC content of 63.7%. The number of encoded genes varied among the genomic islands.

To further explore the function of the GIs in *R. pseudosolanacearum*, gene annotation was conducted using COG, GO and KEGG databases. Functional annotation of 2,226 GIs containing coding genes revealed that host-associated prophage and transposon functions accounted for the largest proportion at 17.5%, followed by transcription. Additionally, genes related to replication, and recombination and repair represented 11.3% and 8.8% of all annotated genes, respectively (Fig. 5A).

GO functional annotation of these 2,226 GIs was categorized into three groups: biological processes, cellular components, and molecular functions. Biological processes were mainly divided into DNA-mediated transposition and translation regulation. In the cellular

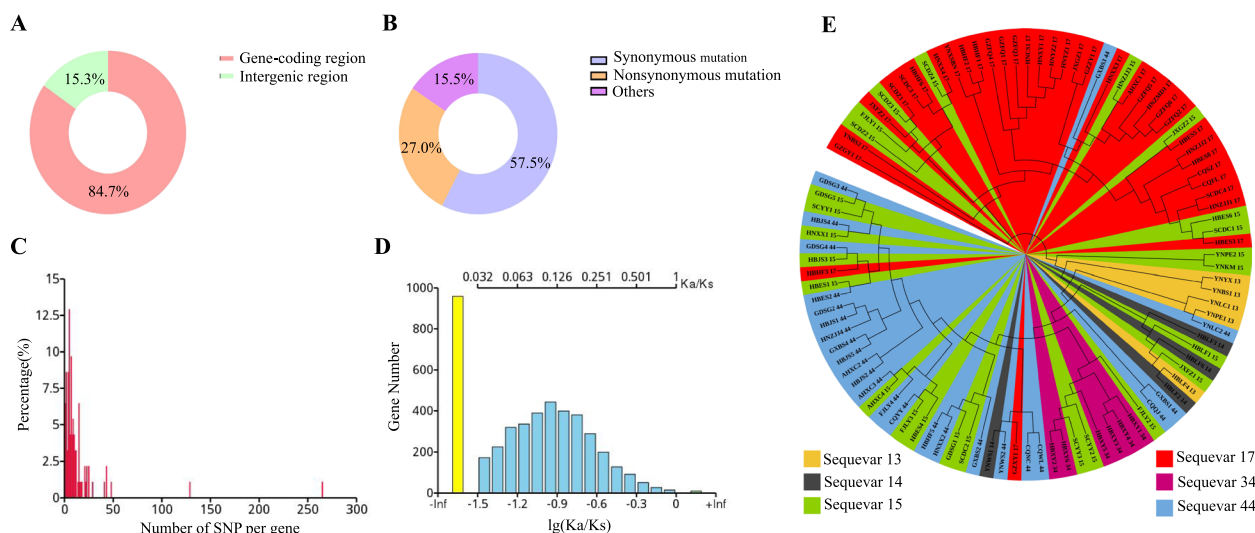


Fig. 4 SNP analysis in the genomes of *R. pseudosolanacearum* isolates from tobacco. **A** Proportion of SNPs occurring in coding and non-coding regions of the genome. This panel shows the distribution of SNPs across different genomic regions, highlighting the relative abundance of SNPs in coding versus non-coding areas; **B** The proportion of different types of SNPs in the gene coding region. This panel categorizes SNPs based on their type (e.g., synonymous, non-synonymous) within the coding regions, providing insights into the nature of genetic variation; **C** Frequency distribution of SNPs in coding genes. This panel presents the frequency of SNPs across coding genes, showing how SNPs are distributed among different genes and their potential impact on gene function; **D** Ka/Ks frequency distribution map. This panel shows the ratio of non-synonymous to synonymous substitutions (Ka/Ks) across the genome, which can indicate regions under positive selection, neutral evolution, or purifying selection; **E** Phylogenetic tree constructed based on whole-genome SNPs. The dendrogram was generated by MEGA version 10.0.5 using the Neighbor-Joining (NJ) method. Sequevars 13, 14, 15, 17, 34, and 44 are represented by yellow, grey, green, red, purple and blue, respectively

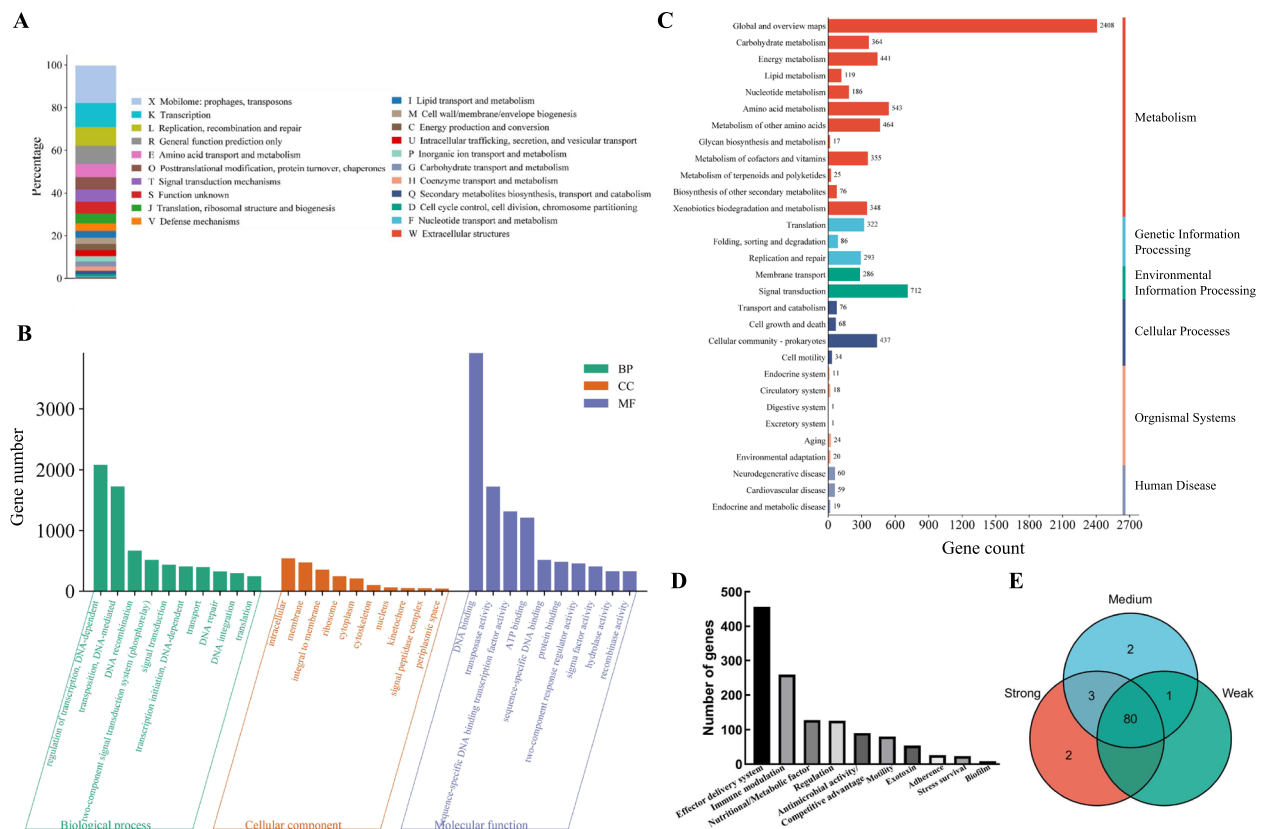


Fig. 5 Analysis of genomic islands (GIs) and type III effectors of *R. pseudosolanacearum* isolates from tobacco. **A** COG function annotation of genes belonging to 2226 GIs. The Clusters of Orthologous Groups (COG) database was used to annotate the functions of genes within the 2226 genomic islands, providing insights into their roles in various biological processes; **B** GO function annotation of genes belonging to 2226 GIs. The genes within the 2226 genomic islands were annotated based on the Gene Ontology (GO) database. The annotations are categorized under Biological Processes, Cellular Component, and Molecular Functions, with a maximum of 10 terms displayed for each category; **C** KEGG pathway function annotation of genes belonging to 2226 GIs. The annotations are categorized under six main pathways: Metabolism, Genetic Information, Environmental Information, Cellular Information, Organismal System, and Human Diseases, with a maximum number of pathways displayed for each category; **D** Function annotation of virulence factors of genes belonging to 2226 GIs. The genes within the 2226 genomic islands were annotated based on the Virulence Factor Database. This panel provides insights into the virulence factors associated with these genes, highlighting their potential roles in pathogenicity; **E** Number of type III effectors in different pathotypes. Different colors signify different pathotype strains, and overlapping types indicate type III effectors shared by different pathogenic *R. pseudosolanacearum* strains

components category, functions were associated with intracellular substances and plasma membrane; in molecular functions, genes were linked to DNA binding and transposon activity (Fig. 5B).

Functional annotation of genes carried by 2,226 GIs was conducted using the KEGG database. The results demonstrated that the GIs contained genes involved in metabolism, genetic information processing, environmental information processing, cellular processes, organic systems, and human diseases. Notably, over 50% of the genes were associated with metabolic pathways, which mainly includes energy metabolism, amino acid metabolism and secondary metabolic pathways, such as symbiotic factors and vitamins. Additionally, 712 genes were implicated in signal transduction in environmental information processing (Fig. 5C).

Bacterial GIs typically harbor virulence factors. Based on matches with virulence factors database (VFDB), 10 virulence factors were identified within the 2,226 GIs, encompassing 1,252 genes related to virulence. The number of virulence factors associated with the effector transport system reached 457, representing 36.5% of all virulence factors identified in GIs, with the majority linked to the type III secretion system (305/457) (Fig. 5D).

Analysis of type III effectors from tobacco *R. pseudosolanacearum*

Tobacco RSSC possesses a diverse range of virulence factors, and numerous studies have demonstrated the significant role of Type III effectors in the interaction between tobacco and RSSC. Prediction from the Type III effector protein database identified 90 types of type III effector

proteins across 103 strains of *R. pseudosolanacearum*, averaging 60 types per strain. This suggests that certain effector proteins are not universally present in all bacterial isolates.

The association between these specific effectors and pathogenicity among bacterial isolates was analyzed in more detail in three representative isolates. The results revealed 88 type III effector proteins from three pathogenic strains, with 80 effector proteins being common among them (Fig. 5E). The weakly pathogenic strains did not exhibit any specific Type III effector proteins. In contrast, the virulent strains exhibited unique type III secreted effector proteins, namely RipBK and RipAZ2. Additionally, RipS6 and RipG4 were identified as specific effector proteins. The highly and moderately pathogenic strains contained three other type III effector proteins (RipG3, RipBO and RS_T3E_Hyp18). These results indicate that some specific type III effectors might be the key factors involved in virulence level of *R. pseudosolanacearum* from different tobacco-planting locations in China.

Discussion

The expansion of tobacco bacterial wilt, caused by RSSC, into northern regions presents a growing concern for tobacco cultivation, particularly in the context of global warming and changes in the tobacco planting system. Our study isolated 204 *R. pseudosolanacearum* strains from 12 provinces and cities in China, with the majority originating from warm tropical low latitudes. Notably, the isolation of strains from high latitudes, such as Xinyang and Zhumadian in Henan Province, suggests an alarming trend of northward expansion, which aligns with previous reports on the changing distribution patterns of the pathogen [13, 23]. The predominant biovar of RSSC infecting tobacco in China has been type III. However, infection caused by biovars I, II, and IV has also been reported [13]. In our study, all 204 *R. pseudosolanacearum* strains isolated from tobacco were identified as biovar III with no other biovars detected. These findings confirm the dominance of biovar III in the RSSC populations affecting tobacco in China. In this study, different phylotypes of RSSC correspond to different geographical origins. It provides insight into the relationship between genetic variation and geographical distribution during genome evolution [24]. All 204 strains of RSSC were classified as phylotype I (*R. pseudosolanacearum*) through PCR primers which can distinguish biovars and phylotypes, and aligned with the phylotype reported for Japanese tobacco isolates [25]. However, reports indicate that phylotype assignments do not always align with specific geographical origins [26], which may result from the spread or mutation of RSSC through international trade.

At least 19 sequevars of RSSC have been identified in China, and sequevars 13, 14, 15, 17, 34, 44, 54, and 55 are capable of infecting flue-cured tobacco [14, 27]. The distribution of sequevar 15 was predominantly observed in low altitude regions, whereas sequevar 54 was exclusively present in high altitude areas [14]. In this study, seven distinct sequevars were identified, with sequevar 15 representing the dominant population, consistent with previous studies [14]. However, sequevar 55, which is also capable of infecting flue-cured tobacco, was not found in this study. This absence may be due to the predominance of disease samples collected in low-altitude tobacco-planting areas, coupled with a relatively small number of isolates collected from Yunnan-Guizhou Plateau region.

In the identification of pathogenicity among strains isolated from various tobacco-planting regions in China, we observed that three distinct pathogenic strains were present in the middle and upper reaches of the Yangtze River, as well as in the southwest and southeast tobacco-planting regions. This indicated a diverse distribution of different pathogenic strains within tobacco-growing regions. Among the strong pathogenic strains isolated from the southwest tobacco-planting region, they represented 61% of *R. pseudosolanacearum*, suggesting that the pathogenicity in this area may be greater than in others, such as the southeast and Huanghuai areas or those along the Yangtze river. This observation implies potential variations in the pathogenicity of *R. pseudosolanacearum* from different geographical sources.

In this study, pangenome sequencing analysis from 103 strains of *R. pseudosolanacearum* from tobacco revealed that the functions of the core genome were mainly associated with redox and transcriptional regulation, consistent with previous reports on core genome functionality [16]. Genomic variations associated with recombinant DNA and repair functions are significantly enriched in the accessory genome [28], which further illustrates the role of frequent variation in affiliated genes as a driving force for evolution within RSSC. Additionally, the open pangenome structure of this species highlights its considerable evolutionary potential and high genomic plasticity [29]. Our results confirmed the conclusion and indicated that tobacco *R. pseudosolanacearum* might have strong ability to adapt to environmental change. We conducted an analysis of the evolutionary pressure on genes of *R. pseudosolanacearum* from tobacco in China and observed that the Ka/Ks ratio of the majority of protein-coding genes in the genome was less than 1, indicating purifying selection. Previous studies have identified two homologs of the WRKY gene family in *Populus*, desert and salt-sensitive, that showed adaptive evolution under salt stress and were important for high salt tolerance [30].

This study analyzed the genetic evolution of *R. pseudosolanacearum* in China utilizing SNP loci within the

genome to construct a phylogenetic tree. In this study, the cross-distribution indicated that genetic variation within *R. pseudosolanacearum* was mainly captured through *egl* gene analysis. However, SNPs across entire bacterial genomes introduced more extensive variations, meaning that phylogenetic trees based on whole genome SNPs provided a more precise reflection of evolutionary relationships than those based on single-gene markers such as *egl*. The abundance of SNP loci in the bacterial genome facilitates a comprehensive understanding of single nucleotide mutation-induced variations and genetic distance. A phylogenetic tree based on whole-genome SNPs provides greater resolution compared to the trees derived from conserved single or multiple genes such as *egl*, *mutS*, and *gdhA*. However, factors influencing the genetic evolution of RSSC likely extend beyond single nucleotide mutations. Large insertions or deletions may also significantly impact genome size and gene expression. When the Ka/Ks ratio exceeds 1, it reflects strong positive selection acting on the gene, highlighting ongoing rapid evolution [31]. In this study, a few coding genes exhibited a Ka/Ks ratio greater than 1, which might indicate positive selection with beneficial non-synonymous mutations retained through natural selection. This supports the hypothesis that the functions of these positively selected genes may play a crucial role in bacterial survival and evolution. However, further research is needed into these specific genes and their effects on survival.

Horizontal gene transfer was found to occur frequently in the dispensable genome, making it the most dynamic and mobile part of the species genome. This process not only contributed to fluctuations in genome size but also played a crucial role in driving the evolutionary adaptability of the population. In this study, the pangenome analysis revealed a high proportion of accessory genes in *R. pseudosolanacearum*, highlighting its active interaction with the surrounding environment and potential for enhanced environmental adaptability. It is consistent with the conclusion of previous report [21]. Furthermore, the identification of specific genes in different strains suggested distinct genomic variations that may confer environmental adaptability advantages to certain strains in their particular environments.

The GIs are widely distributed in RSSC, leading to frequent gene gain and loss, which enhances the genome's adaptability and competitiveness in specific ecological niches [32]. Previous studies have identified 21 GIs in strain CQPS-1 of RSSC isolated from a continuous cropping area of tobacco [24]. In our study, we further explored the GIs present in 103 *R. pseudosolanacearum* strains, specifically focusing on the CQPS-1 strain, revealing that each strain contained between 18–25 GIs. Furthermore, we identified 10 types of virulence factors within these GIs, which may have a significant effect on

transport systems. Earlier findings suggested that bacteria frequently acquire virulence factors and drug resistance genes via GIs [33]. This indicates that pathogenic virulence factors encoded by GIs may play a critical role in disease promotion among tobacco plants.

In previous studies, the tobacco RSSC pangenome in China contained various virulence factors, including the type II secretion system, which produces extracellular enzymes such as cellulase that facilitate the infection process [34, 35]. The Type VI secretion system is also reported to be significant for pathogen colonization within the host [34]. The functionality of these secretion systems relies on the conserved protein TssM, which promotes protein transport across the membrane [36]. In this study, we identified 103 *R. pseudosolanacearum* strains that exhibited type III effector proteins, along with unique strain-specific proteins. We also identified type III effector proteins in different strains with their pathogenic types. Notably, RipBK and RipAZ2 are particularly associated with strong pathogenic strains. Importantly, these two effectors were not detected in GMI1000, CFBP2957, CMR15, Po82 and PSI07 strains, and their specific function remains to be elucidated. Moreover, RipBO has been previously referred to as Hyp16 [37], but no studies have connected it to pathogenicity. Additionally, RipS6 has been implicated in potentially reducing the virulence of *R. solanacearum* from potato showing a greater number of missense mutations compared to GMI1000 [38]. RipS6 belongs to the SKWP family, while both RipG3 and RipG4 are part of the GALA protein family, which includes F-box and leucine-rich repeats. The functions of RipBK and RipAZ2 remain unclear. While the knock out of any single effector protein does not impact the virulence of *Pseudomonas aeruginosa* [39], the simultaneous knock out all seven effectors from this family significantly reduces the pathogenicity of *Pseudomonas aquatilis*. Lastly, the role of RS_T3E_Hyp18 in pathogenesis remains unclear, despite being identified as a putative protein. While biological functions of many effectors have been ascertained, there are still some unidentified T3Es, and as well, their functions undetermined.

In conclusion, pan-genome analysis of high-quality whole-genome sequencing data from 103 *R. pseudosolanacearum* strains causing tobacco bacterial wilt in China revealed genomic variations among different strains, with detailed analysis of genome structure and functional annotation. Single nucleotide polymorphism (SNP) loci were identified in *R. pseudosolanacearum*, elucidating the relationship between SNP variations and population evolution. The study further investigated differences in genomic islands and type III effectors among strains with varying pathogenicity levels. These comprehensive investigations provide valuable insights into the pan-genome characteristics of Chinese tobacco *R.*

pseudosolanacearum populations, facilitating in-depth analysis of molecular variations and population evolution and laying a foundation for developing scientific and rational disease control strategies.

Materials and methods

Isolation of RSSC causing tobacco bacterial wilt

From 2022 to 2023, infected stems of tobacco exhibiting bacterial wilt symptoms were obtained from the South-east tobacco-growing region of China (Fujian, Anhui, Zhejiang, Guangdong and Jiangxi provinces), the South-west region (Yunnan, Sichuan, Guizhou and the Guangxi Zhuang, Autonomous region), the Yangtze River region (Hubei, Hunan and Chongqing) and the Huanghuai region (Henan province). Tobacco stem samples with typical necrotic lesions were cut to 8 cm and stored at 4 °C.

Isolation of RSSC was conducted using the serial dilution method. The surface of infected tobacco stems were cleaned with water, dried, disinfected with 75% ethanol for 30 s, and sliced vertically along the junction between diseased and healthy tissues (with sections no more than 1 cm thick). The tissues were soaked in sterile water for at least 2 h. The resulting leachate was diluted to 10^{-6} and 10^{-7} concentrations under sterile conditions. Finally, a 5–10 µL sample of the diluted solution was taken and spread on selective TTC or NA media. The plates were incubated at 28 ± 2 °C for 3–5 d.

Identification of RSSC strains from tobacco

Bacterial genomic DNA was extracted from infected tobacco stems using the TIANamp Bacteria DNA Kit (DP302, China) according to the manufacturer's protocol. The specific amplification of RSSC DNA was performed in a MyCycler™ Thermal Cycler (Bio-rad, Hercules, CA) using two specific primers, 759/760 and Rsol-flic-F/Rsol-flic-R (Table S2) [40, 41]. The PCR reaction conditions were: 95 °C for 5 min; 30 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 15 s; 72 °C for 5 min; and 4 °C for 2 min. PCR products were detected by 1% agarose gel electrophoresis.

The phylotype identification of the RSSC strains was carried out using a phylotype multiplex PCR, which involved 759/760 primers along with Nm21:1F/Nm21:2R, Nm21:2F/Nm21:2R, Nm23:AF/Nm23:2R or Nm22: InF/Nm22:2R primers (Table S2) [10]. The PCR reaction conditions and gel electrophoresis were the same as described above.

For sequevar identification, the *egl* gene sequence of the RSSC strains was amplified using Endo-F/R primers (Table S2) [10]. PCR products were sequenced by the AOKE Biological Technology Company (Wuhan, China). For reference strains, 34 previously identified sequevar (including 4 phylotypes and 27 sequevars) were obtained

from GenBank. The phylogenetic tree was constructed using MEGA10.0.5 software (<https://www.megasoftware.net/>) with the Neighbor-joining (NJ) method, and bootstrap analysis was performed based on 5000 replicates. Sequevars were defined after isolates were grouped with reference isolates with strong bootstrap support.

For biovar identification, the tested strains were inoculated into six biochemical media containing cellobiose, maltose, lactose, mannitol, sorbitol or dulcitol, and incubated at 28 ± 2 °C for 21 d. The biochemical type of RSSC was determined by assessing the oxidation of the three hexanols and three disaccharides in the tested strains, indicated by a color change from blue-green to light yellow in the medium.

Pathogenicity assays of RSSC strains

A total of 63 RSSC strains from different locations in China (Table S3) used for inoculation were grown on NA medium lacking TTC. Bacterial suspensions were prepared by washing bacterial cells from the plates with sterile water and diluted to a concentration of 1×10^8 CFU/mL ($OD_{600} = 0.1$). For each strain, bacterial suspensions were then used for inoculation onto eight 30 day-old tobacco seedlings of each cultivar, Honghuadajinyuan (susceptible), K326 (moderately resistant) and Yanyan 97 (resistant). The plants were separated into four blocks in a randomized complete block design. The experiment was repeated three times. A total of 8 ml bacterial suspensions were injected onto the root of each tobacco plant. Treatment of sterile water served as control. Following inoculation, the seedlings were grown in a greenhouse maintained at 28 ± 2 °C with 95% humidity. Disease severity was assessed at 7-d intervals up to 30 days post-inoculation (dpi), and corresponding disease index were calculated. The area under the disease progress curve (AUDPC) was calculated for all varieties, and then clustered using SPSS software with square Euclidean distance as the clustering method. Based on AUDPC values, the pathogenic types of the RSSC strains were subsequently placed into high, medium and weak.

Genome sequencing and function annotation

A total of 103 strains of tobacco *R. pseudosolanacearum* were isolated from four tobacco-growing regions, and specific strain details are provided (Table S1). Genomic DNA was extracted using the STE (Sodium Chloride-Tris-EDTA) method, and the concentrations of the DNA were assessed using the Qubit system (ThermoFisher, USA). The integrity and purity of the genomic DNA were evaluated by performing electrophoresis on a 1% agarose gel at 180 V for 20 min. Genomic DNA quality of tobacco *R. pseudosolanacearum* was checked and quantified using NanoDrop Spectrophotometer (ThermoFisher, USA).

Initially, the BluePippin automatic recovery system (Sage Science, USA) was used to isolate large DNA fragments, followed by end repair using the Oxford Nanopore Technologies SQK-LSK109 ligation kit for adapter ligation, and subsequent sequencing on the Nanopore platform (Novogene, Beijing, China). Quality control was implemented using NanoPlot software [42], applying a threshold of $Q > 7$ to ensure high-quality data.

To predict the coding genes of the sequenced genomes of *R. solanacearum*, GeneMarkS (Version 4.17) software was utilized, and the whole genome size, GC content, number of coding genes and the proportion of coding genes in the genome of each sample were determined.

Functional annotation was performed using the COG database (<https://www.ncbi.nlm.nih.gov/research/cog/>), GO database (<https://www.geneontology.org/>), KEGG database (<https://www.genome.jp/kegg/>), and PHI database (www.phi-base.org/). Predicted genes were compared using BLASTP software against the COG, GO, KEGG and VFDB (<https://www.mgc.ac.cn/VFs/>), with parameters set to e-value $\leq 10^{-5}$, similarity $\geq 40\%$, and sequence coverage $\geq 40\%$. For each gene sequence, the highest scoring values from the BLAST results were selected as the final result for the gene functional annotation. Additionally, nucleotide consistency between 103 strains of *R. solanacearum* and the reference strain GMI1000 was compared using OAT ANI (<https://www.ezbiocloud.net/tools/orthoani>) and heatmap was generated using pheatmap package in R software (v4.3.2) with Hierarchical Clustering.

Pangenome analysis of 103 strains of tobacco *R. pseudosolanacearum*

The CD-HIT (Cluster Database at High Identity with Tolerance) tool [43] was used for the genome-wide analysis of 103 tobacco *R. pseudosolanacearum* and reference GMI1000. Each XLS file generated contained a comprehensive set of clustering results, including the number of genes, core genome, and both accessory and specific genes. The clustering results for the pangenome, core genome, accessory genome, and strain-specific genes were transformed into matrix files to obtain the corresponding datasets. The expansion of the pangenome was estimated using the "micropan" package in R software based on the Heap's law model, and the attenuation parameter alpha was calculated through 5000 random number permutations. In addition, 10 strains of diverse phylotypes were selected from the RSSC. These strains, along with 103 tobacco *R. pseudosolanacearum*, were included in a pangenome analysis using CD-HIT software to derive the single-copy core gene sequence. The protein amino acid sequences of these single-copy core genes were obtained using GeneMarkS software and aligned using MUSCLE software. The resulting sequence

alignment was then imported into TreeBeST software for the generation of an evolutionary tree using the Neighbor-joining method.

The GMI1000 strain was designated as the reference strain. MUMmer alignment software was used to conduct whole genome alignment between each sample and the reference sequence to identify variant sites, followed by preliminary filtering to detect potential SNP sites. A 100-bp sequence flanking each side of the SNP site from the reference sequence was extracted, and then aligned with the assembly results using BLAT software for verification of the SNP site. SNPs from sequences with an alignment length less than 101 bp were considered unreliable and removed, and repetitive regions identified through multiple alignments also were excluded. BLAST, TRF, and Repeatmask software were used to predict repeat regions within the reference sequence, and SNPs located in these regions were filtered out to obtain reliable SNPs. The number, types, and genomic locations of identified SNPs were tallied, and natural selection pressure on protein-coding genes was assessed based on the ratio of non-synonymous mutations to synonymous mutations within individual protein-coding genes. The SNP locus matrix file from the entire genome of tobacco *R. pseudosolanacearum* strains was converted to a Fasta format file and then imported into MegaX for phylogenetic tree generation using the Neighbor-joining method based on the SNPs in the genome.

Identification of genome islands and type III effectors

The presence of genome islands in the genomes of 103 strains of tobacco *R. pseudosolanacearum* was predicted using IslandPath—DIOMB software (Version 0.2). This analysis focused on dinucleotide bias and mobility genes (such as enzymes or integrases), which are indicative of phylogenetic bias and translocation events.

The Type III secretion system (T3SS) effector proteins of 103 strains of tobacco *R. pseudosolanacearum* were predicted based on the T3Es database (<https://iant.toulouse.inra.fr/bacteria/annotation/site/prj/T3Ev3/>). Based on the pathogenicity results for the strains of tobacco *R. pseudosolanacearum*, the functions of various type III effector proteins were annotated and the specific type III effector proteins of different pathogenic strains were analyzed based on their varying levels of pathogenicity.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-025-04485-4>.

Supplementary Material 1.

Supplementary Material 2.

Authors' contributions

Conceptualization, Y.F. and Y.Y.L.; Methodology, J.K. and M.A.S.; Formal Analysis, J.K., M.A.S., Y.Q.Z. and Z.W.; Resources, R.B.X. and Y.Y. L.; Data Curation, J.K.; Writing-Original Draft Preparation, J.K., M.A.S. and Y.F.; Writing-Review & Editing, T.H., L.Z. and Y.F.; Project Administration, J.B.H. and Y.Y. L.

Funding

The research was supported by Pests and Diseases Green Prevention and Control Major Special Project (Grant no. 110202101045, LS-05).

Data availability

The clean sequence data reported in this paper are available in the Genome Sequence Archive (GSA: CRA027599) in CNCB (China National Center for Bioinformation) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The sequence data of egl gene have been deposited in the Genbank of National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) with the accession numbers PX131568—PX131771.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

²Tobacco Research Institute of Hubei Province, Wuhan, Hubei 430030, China

³Environmental Sciences, University of Guelph, Guelph, ON N1G 2W1, Canada

⁴College of Life Sciences, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

Received: 9 June 2025 / Accepted: 15 October 2025

Published online: 25 November 2025

References

- Van Nguyen T, Kröger C, Bönninghausen J, Schäfer W, Bormann J. The ATF/CREB transcription factor Atf1 is essential for full virulence, deoxynivalenol production, and stress tolerance in the cereal pathogen *Fusarium graminearum*. *Mol Plant-Microbe Interact*. 2013;26:1379–94.
- Genin S. Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum*. *New Phytol*. 2010;187:920–8.
- Tjou-Tam-Sin NNA, Van de Bilt JLI, Westenberg M, Gorkink-Smits PPMA, Landman NM, Bergsma-Vlami M. Assessing the pathogenic ability of *Ralstonia pseudosolanacearum* (*Ralstonia solanacearum* phylotype I) from ornamental Rosa spp. plants. *Front Plant Sci*. 2017;8:1895.
- Singh N, Phukan T, Sharma PL, Kabyashree K, Barman A, Kumar R, et al. An innovative root inoculation method to study *Ralstonia solanacearum* pathogenicity in tomato seedlings. *Phytopathology*. 2018;108:436–42.
- García RO, Kerns JP, Thiessen L. *Ralstonia solanacearum* species complex: a quick diagnostic guide. *Plant Health Progress*. 2019;20:7–13.
- Chen S, Qi G, Ma G, Zhao X. Biochar amendment controlled bacterial wilt through changing soil chemical properties and microbial community. *Microbiol Res*. 2020;231:126373.
- Cai Q, Zhou G, Ahmed W, Cao Y, Zhao M, Li Z, et al. Study on the relationship between bacterial wilt and rhizospheric microbial diversity of flue-cured tobacco cultivars. *Eur J Plant Pathol*. 2021;160:265–76.
- Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Guilbaud C, Fegan M, et al. *Ralstonia solanacearum* strains from Martinique (French West Indies) exhibiting a new pathogenic potential. *Appl Environ Microbiol*. 2007;73:6790–801.
- Denny TP. Plant Pathogenic *Ralstonia* species. In: Gnanamanickam, S.S. (eds) *Plant-Associated Bacteria*. Dordrecht: Springer. 2007;573–644.
- Fegan M, Prior P. How complex is the *Ralstonia solanacearum* species complex. *Bact wilt Dis*. 2005;1:449–61.
- Safni I, Cleenwerck I, De Vos P, Fegan M, Sly L, Kappler U. Polyphasic taxonomic revision of the *Ralstonia solanacearum* species complex: proposal to emend the descriptions of *Ralstonia solanacearum* and *Ralstonia syzygii* and reclassify current *R. syzygii* strains as *Ralstonia syzygii* subsp. *syzygii* subsp. nov., *R. solanacearum* phylotype IV strains as *Ralstonia syzygii* subsp. *indonesiensis* subsp. nov., banana blood disease bacterium strains as *Ralstonia syzygii* subsp. *celebesensis* subsp. nov. and *R. solanacearum* phylotype I and III strains as *Ralstonia pseudosolanacearum* sp. nov. *Int J Syst Evol Microbiol*. 2014;64:3087–103.
- Sharma K, Iruegas-Bocardo F, Abdurahman A, Alcalá-Briseño RI, Garrett KA, Goss EM, et al. *Ralstonia* strains from potato-growing regions of Kenya reveal two phylotypes and epidemic clonality of phylotype II sequevar 1 strains. *Phytopathology*. 2022;112:2072–83.
- Jiang G, Wei Z, Xu J, Chen H, Zhang Y, She X, et al. Bacterial wilt in China: history, current status, and future perspectives. *Front Plant Sci*. 2017;8:1549.
- Zhao Q, Geng M, Xia C, Lei T, Wang J, Cao C, et al. Identification, genetic diversity, and pathogenicity of *Ralstonia pseudosolanacearum* causing cigar tobacco bacterial wilt in China. *FEMS Microbiol Ecol*. 2023;99:fiad018.
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome.” *Proc Natl Acad Sci U S A*. 2005;102:13950–5.
- Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. *Curr Opin Genet Dev*. 2005;15:589–94.
- Dumas E, Christina Boritsch E, Vandenbogaert M, Rodríguez de la Vega RC, Thiberge J-M, Caro V, et al. Mycobacterial pan-genome analysis suggests important role of plasmids in the radiation of type VII secretion systems. *Genome Biol Evol*. 2016;8:387–402.
- Freschi L, Vincent AT, Jeukens J, Emond-Rheault J-G, Kukavica-Ibrulj I, Dupont M-J, et al. The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol Evol*. 2019;11:109–20.
- Goyal A. Metabolic adaptations underlying genome flexibility in prokaryotes. *PLoS Genet*. 2018;14:e1007763.
- Hurgobin B, Golicz AA, Bayer PE, Chan CK, Tirnaz S, Dolatabadian A, et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol J*. 2018;16:1265–74.
- Geng R, Cheng L, Cao C, Liu Z, Liu D, Xiao Z, et al. Comprehensive analysis reveals the genetic and pathogenic diversity of *Ralstonia solanacearum* species complex and benefits its taxonomic classification. *Front Microbiol*. 2022;13:854792.
- Baroukh C, Cottret L, Pires E, Peyraud R, Guidot A, Genin S. Insights into the metabolic specificities of pathogenic strains from the *Ralstonia solanacearum* species complex. *MSystems*. 2023;8:e00083–e123.
- Tsuchiya K. Genetic diversity of *Ralstonia solanacearum* and disease management strategy. *J Gen Plant Pathol*. 2014;80:504–9.
- Liu Q, Liu H, Gong Y, Tao Y, Jiang L, Zuo W, et al. An atypical thioredoxin imparts early resistance to sugarcane mosaic virus in maize. *Mol Plant*. 2017;10:483–97.
- Liu Y, Kanda A, Yano K, Kiba A, Hikichi Y, Aino M, et al. Molecular typing of Japanese strains of *Ralstonia solanacearum* in relation to the ability to induce a hypersensitive reaction in tobacco. *J Gen Plant Pathol*. 2009;75:369–80.
- Shutt VM, Shin G, Van Der Waals JE, Goszczynska T, Coutinho TA. Characterization of *Ralstonia* strains infecting tomato plants in South Africa. *Crop Prot*. 2018;112:56–62.
- She X, He Z, Li H. Genetic structure and phylogenetic relationships of *Ralstonia solanacearum* strains from diverse origins in Guangdong Province, China. *J Phytopathol*. 2018;166:177–86.
- Yin Z, Liu X, Qian C, Sun L, Pang S, Liu J, et al. Pan-genome analysis of *Delftia tsuruhatensis* reveals important traits concerning the genetic diversity, pathogenicity, and biotechnological properties of the species. *Microbiol Spectr*. 2022;10:e02072–e2121.
- Reis AC, Cunha MV. The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. *Microb Genom*. 2021;7:664.
- Ma J, Lu J, Xu J, Duan B, He X, Liu J. Genome-wide identification of WRKY genes in the desert poplar *Populus euphratica* and adaptive evolution of the genes in response to salt stress. *Evol Bioinforma*. 2015;11:EBO-S22067.
- Zhang Z. KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics*. 2022;20(3):536–40.

32. Leimbach A, Hacker J, Dobrindt U. *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. In: Dobrindt, U., Hacker, J., Svanborg, C. (eds) *Between Pathogenicity and Commensalism. Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer. 2013;358:3–32.
33. Bertelli C, Tilley KE, Brinkman FSL. Microbial genomic island discovery, visualization and analysis. *Brief Bioinform*. 2019;20:1685–98.
34. Lonjon F, Turner M, Henry C, Rengel D, Lohou D, van de Kerkhove Q, et al. Comparative secretome analysis of *Ralstonia solanacearum* type 3 secretion-associated mutants reveals a fine control of effector delivery, essential for bacterial pathogenicity. *Mol Cell Proteomics*. 2016;15:598–613.
35. Deslandes L, Genin S. Opening the *Ralstonia solanacearum* type III effector tool box: insights into host cell subversion mechanisms. *Curr Opin Plant Biol*. 2014;20:110–7.
36. Zoued A, Brunet YR, Durand E, Aschtgen M-S, Logger L, Douzi B, et al. Architecture and assembly of the type VI secretion system. *Biochimica et Biophysica Acta (BBA)*. 2014;1843:1664–73.
37. Peeters N, Carrère S, Anisimova M, Plener L, Cazalé A-C, Genin S. Repertoire, unified nomenclature and evolution of the type III effector gene set in the *Ralstonia solanacearum* species complex. *BMC Genomics*. 2013;14:1–19.
38. Huang M, Tan X, Song B, Wang Y, Cheng D, Wang B, et al. Comparative genomic analysis of *Ralstonia solanacearum* reveals candidate avirulence effectors in HA4-1 triggering wild potato immunity. *Front Plant Sci*. 2023;14:1075042.
39. Cunnac S, Occhialini A, Barberis P, Boucher C, Genin S. Inventory and functional analysis of the large Hrp regulon in *Ralstonia solanacearum*: identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Mol Microbiol*. 2004;53:115–28.
40. Schonfeld J, Heuer H, Van JD. Specific and sensitive detection of *Ralstonia solanacearum* in soil on the basis of PCR amplification of *flhC* fragments. *Appl Environ Microbiol*. 2003;69(12):7248–56.
41. Opina N, Tavner F, Hollway G. A novel method for development of species and strain-specific DNA probes and PCR primers for identifying *Burkholderia solanacearum* (formerly *Pseudomonas solanacearum*). *Asia-Pac J Mol Biol Biotechnol*. 1997;5(1):19–30.
42. De Coster W, D'Hert S, Schultz DT, et al. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34:2666–9.
43. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.