Genomic Analysis of *Curtobacterium Flaccumfaciens* Reveals the Differences Between Pathogenic and Nonpathogenic Strains

Qingde Li
China Agricultural University

Lianjun Sun (✉ sunlj@cau.edu.cn)
China Agricultural University

Research Article

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

Abstract

Purpose

Curtobacterium flaccumfaciens is a Gram-positive bacterium which has been isolated from different plants and abiotic environment. *Curtobacterium. flaccumfaciens* pv. *flaccumfaciens* (Cff) is a pathogenic bacterium that infects legume, which is causing great economic losses. At the genomic level, the metabolic and phylogenetic characteristics, and differences in pathogenicity between pathogenic and nonpathogenic *C. flaccumfaciens* strains have not been analyzed in detail.

Methods

Therefore, in order to discuss the differences in genome, phylogeny, gene function and mobile genetic elements between pathogenic and nonpathogenic strains, pangenomics and comparative genomics were used in this study to analyze 12 *C. flaccumfaciens* strains.

Result

The pangenome of *C. flaccumfaciens* is open. Phylogenetic analysis showed that there was no correlation between the phylogeny and pathogenicity of *C. flaccumfaciens*. KAAS annotation of the core genome shows that the citrate cycle was incomplete. In addition, gene islands analysis of the three pathogenicity-related genes encoding for pectate lyase, serine protease and cellulases showed that they only existed in the Cffs and LMG3645 strains. LMG3645 might be a pathogenic strain.

Conclusion

This study clearly and reliably revealed the differences between the pathogenic and nonpathogenic strains of *C. flaccumfaciens* at the genomic level, and paves the way for further research on its pathogenicity.

1. Introduction

Curtobacterium flaccumfaciens is a Gram-positive, short rod or coryneform, aerobic, and motile bacterium. It has been isolated from multiple environments such as residential carpet, ice-wedge polygon, and different plants. *C. flaccumfaciens* pv. *flaccumfaciens* (Cff) is a pathogenic bacterium that can cause bacterial wilt of common bean (*Phaseolus vulgaris*) as well as bacterial tan spot disease on soybean (Jeger et al. 2018), the former is one of the most important diseases threatening edible legume production around the globe (Chen et al. 2020b). In addition, Cff can also infect cowpea, wheat and red nightshade plants, and may survive and overwinter on solanaceous vegetable residues (Osdaghi et al. 2018). Previous studies have mainly focused on the identification of the pathogenicity of Cff in relation

to legumes (Puia et al. 2021; Valdo et al. 2016), the selection of crops in rotation system to reduce the effect of Cff (Goncalves et al. 2021; Goncalves et al. 2017), the antagonistic effect of Cff on *Bacillus* spp. (Leao et al. 2016), and the sensitivity of Cff to protein synthesis group inhibitors (Tumbarski et al. 2018).

At present, only Cff strains were found to be pathogenic in the *C. flaccumfaciens* species (Osdaghi et al. 2018). However, there is no obvious morphological difference between pathogenic and nonpathogenic *C. flaccumfaciens* strains (Osdaghi et al. 2018). Showing high genetic diversity amongst the strains, a phylogenetic tree of *C. flaccumfaciens* was constructed by multilocus sequence analysis (MLSA) based on six housekeeping genes (Goncalves et al. 2019). Nevertheless, to further reveal the differences between the pathogenic and nonpathogenic strains, systems biology approaches such as genomics is necessary. At present, genome sequencing and annotation has made great progress in understanding the molecular mechanism of plant colonization, pathogenicity and survival mechanism of plant pathogenic actinomycetes (Thapa et al. 2019). To date, whole genome sequences of 12 *C. flaccumfaciens* strains have been sequenced and published (Osdaghi et al. 2016), Three *C. flaccumfaciens* strains were analyzed by Chen et al. (Chen et al. 2020b). This involved studying the virulence factors of four phylogenetically closely related *Curtobacterium spp.* strains in relation to Cff using comparative genomics.

The accumulation of *C. flaccumfaciens* genomic data and the development of corresponding bioinformatics software can facilitate pangenomic analysis. Pangenomic analysis, which can help in understanding species diversity and the metabolic capabilities of a species (Golicz et al. 2020), includes studying the core genomes as they relate to the basic biological functions, accessory genomes, and unique genomes demonstrating the diversity of bacterial strains (Tettelin et al. 2005). Previously, there have been pangenomic studies on plant pathogenic bacteria such as *Brenneria spp.*, *Dickeyasolani* and *Pantoeaananatis* (Bakhshi Ganje et al. 2021; Golanowska et al. 2018; Stice et al. 2018).

In this study, the pangenome of 12 published *C. flaccumfaciens* genomes were analyzed. The phylogenetic relationship of *C. flaccumfaciens* was studied at the genome level, while the metabolic capabilities in the core and accessory *C. flaccumfaciens* genomes were also analyzed. Lastly, mobile genetic elements (MGEs) of *C. flaccumfaciens* were identified and analyzed. Collectively, this study analyzed the differences across genomes, phylogeny, gene function and mobile genetic elements between pathogenic and nonpathogenic strains.

2. Materials And Methods

2.1 Pangenome analysis

The 12 *C. flaccumfaciens* genomes used for the pangenome analysis were download from the NCBI database. Genome accession numbers are shown in Table 1. The pangenome and core genome analysis of *C. flaccumfaciens* were conducted via the Bacterial Pangenome Analysis (BPGA) pipeline (Chaudhari et al. 2016). Usearch is the clustering tool with a cutoff of 50% sequence identity. Using the gene

presence–absence binary matrix (pan-matrix) obtained from BPGA as the input data, the pangenome and core genome profiles were evaluated with PanGP (Zhao et al. 2014).

To construct the phylogenetic tree of the 12 *C. flaccumfaciens* strains, the concatenated protein amino acid sequences in the core genome of each strain obtained from the BPGA pipeline analysis were included in the ClustalW multiple sequence alignment step (Larkin et al. 2007). The alignment file was imported into MEGA X in order to construct a maximum-likelihood phylogenetic tree (Kumar et al. 2018).

2.2 Functional annotation of the pangenome

Protein-coding genes were assigned to Cluster of Orthologous groups (COG) according to their functions by eggNOG-mapper (Huerta-Cepas et al. 2019). KEGG Automatic Annotation Server (KAAS) Pipeline was used to perform (KEGG) Orthology (KO) analysis (Moriya et al. 2007). In addition, *C. flaccumfaciens* mobile genetic elements (MGEs), which including genome islands and prophage sequences, were predicted. IslandViewer 4 was used to predict the Genome islands (GIs) through three methods: SIGI-HMM, IslandPath-DIMOB, and IslandPick (Bertelli et al. 2017). PHASTER was used to annotate the prophage sequences where the instant score was > 100(Arndt et al. 2016).

3. Results And Discussion

3.1 Pangenome of *C. flaccumfaciens*

The detailed information of the published genomes of *C. flaccumfaciens* is shown in Table 1. The genome size of the 12 *C. flaccumfaciens* strains ranged from 3.68 Mb to 3.94 Mb. The average number of protein coding genes in each genome was 3,536. A total of 6,577 gene clusters were identified via the pangenome analysis. Of these, the 1,893 genes found in all 12 genomes represented the core *C. flaccumfaciens* genome (Fig. 1). Six housekeeping genes (atpD, dnaK, gyrB, ppK, recA, and rpoB), which had been used to construct a phylogenetic tree of *C. flaccumfaciens* were all identified in the core genome (Goncalves et al. 2019). A total of 4,296 genes were found in more than two, but not all 12 genomes, these constituted the accessory genomes (Fig. 1). In addition, and representing the strain-specific genes, 2281 genes were uniquely found in individual genomes. The strain-specific gene in each genome varied from seven to 596 (Fig. 2A). Strain JUb65 and WW7 had the largest number of strain-specific genes (596 and 534, respectively).

The changes in pangenome size and core gene number is shown in Fig. 2B, with the relationship between the pangenome size (y) and the genome number (x) represented by the equation $y = \alpha x^\beta + \sigma$. When $\beta = 0.45$, the pan genome is considered to be open. Under this condition, the magnitude of the pangenome tends to be enlarged with an increased number bacterial genomes. That is to say, the species are capable of obtaining homologous and nonhomologous genes from other gene pools. This serves an important evolutionary function. This is in contrast to what is observed for the *Dickeyasolani* plant pathogen, which presents a nearly closed pangenome structure (Motyka-Pomagruk et al. 2020).

3.2 Phylogenetic analysis

Plant, residential carpet and ice-wedge polygon *C. flaccumfaciens* isolates were obtained, and a phylogenetic tree constructed based on the core *C. flaccumfaciens* genome. The 12 strains were divided into two clades (Fig. 3). In the first clade, strains were all isolated from plants. Forming a single cluster and demonstrating similar degrees of evolution, both CFBP3418 and LMG 3645 were isolated from *Phaseolus vulgaris*, and their evolutionary degrees are consistent. In addition, given that they had the largest number of strain-specific genes, the highest degree of evolution was found between the JUb65 and WW7 strains. Among the pathogenic strains, only Cff1037 was grouped into the second clade. Hence, in this phylogenetic tree, it was found that there was no formed a monophyletic cluster among the three pathogenic strains, and there was no correlation between phylogeny and pathogenicity of *C. flaccumfaciens*, which was consistent with previous studies (Goncalves et al. 2019; Osdaghi, Taghavi, Hamzehzarghani, et al. 2018). Similarly, no correlation between *Pantoeaananatis* pathogenicity and genetic diversity was seen (Stice et al. 2018).

3.3 COG functional annotation of the *C. flaccumfaciens* pangenome

Following eggNOG mapper analysis (Huerta-Cepas et al. 2019), the genes in the *C. flaccumfaciens* pangenome represented 96.1%, 90.9% and 84.7% of core, accessory, and strain-specific genomes, respectively. The genes of each group were assessed according to COG categories. COG functional annotation is classified into four groups: cell physiological processes and signals (COG categories: D, M, N, O, T, U, V), information storage and processing (COG categories: J, K, L), metabolism (COG categories: C, E, F, G, H, I, P, Q), and limited characterization (COG categories: R, S).Function unknown (S) was the largest category into which annotated genes were assigned and accounted for 18.0%, 22.8%, and 23.8% of core, accessory, and strain-specific genes respectively. Since the core genome typically encodes proteins responsible for basic housekeeping functions (Loper et al. 2012), most core genes were involved in the housekeeping processes. Among core genes annotated to the metabolism group, Amino acid transport and metabolism (E) (11.55%) and Carbohydrate transport and metabolism(G) (10.28%) were the largest two categories; while Transcription (K, 8.17%) and Translation, and ribosomal structure and biogenesis (J, 7.83%) were the two most commonly observed among the information storage and processing groups (Fig. 4). Among the accessory genes, the largest category was K (15.47%), followed by G (10.56%), E (10.13%), and Cell wall/membrane/envelope biogenesis (M) (7.03%). In the two most prevalent groups within the strain-specific genes, the K and G categories occupied 14.68% and 9.67%, respectively. Replication, recombination and repair (L) accounted for 5.84% of the strain-specific genomes, and 2.26% of the accessory genomes. No core genes were annotated to group L.

In addition, accessory genes uniquely found only in the three pathogenic Cff strains were not seen. However, while 42 COG annotated genes were identified in three Cff strains, they were absent in other eight strains. Four of the COGs were related to serine protease, while others were found to relate to cellulase and pectate lyase (Table S1). It is worth noting that, except for Cff, these 42 COGs only existed in strain LMG3645.

3.4 Metabolic analysis of the *C. flaccumfaciens* pangenome

The main metabolic pathways of the *C. flaccumfaciens* core, accessory and strain-specific genomes were represented by KO numbers. In the core genome, the basal metabolic, glycolysis/glyconeogenesis metabolic and pentose phosphate pathway are complete. However, the citrate cycle (TCA cycle) was found to be missing the gene encoding 2-oxoglutarate dehydrogenase (EC:1.2.4.2) in all 12 strains. KAAS annotation of the core genome showed that *C. flaccumfaciens* can utilize a variety of carbon sources. The metabolic pathways of sucrose, starch, maltose, glucose, fructose, cellotriose, cellobiose, trehalose, mannose, xylose, and xylitol were complete. The metabolic pathway of sulfur source of *C. flaccumfaciens* is complete, but an enzyme (cysJ, EC: 1.8.1.2) with only WW7 strain missing. In metabolism involving nitrogen sources, only ammonia and formamide could be metabolized. Lastly, it was found that the core *C. flaccumfaciens* genome can produce ethanol.

3.5 Prediction of MGEs

Mobile genetic elements (MGEs) are segments that mediate the movement of DNA within genomes or between bacterial cells, which are not commonly found among chromosomal housekeeping genes. Known to contribute to the pathogenicity, environmental adaptation, and intraspecific diversity of *C. flaccumfaciens* (Frost et al. 2005), the MGEs associated with genomic islands (GIs) across the 12 *C. flaccumfaciens* genomes were investigated. The number of predicted GIs in the *C. flaccumfaciens* genome ranged from 10 (strain CFBP3418) to 27 (strainS5.26), indicating the widespread existence of MGEs in *C. flaccumfaciens*. The size of total GIs in each strain ranged from 157kb (strain CFBP3418) to 414kb (strainS5.26). In each genome, the genes of GIs, range from 320 to 854, most genes identified in the GIs encode hypothetical protein. Interestingly, the size of the total GIs of the three pathogenic strains and LMG3645 are the smallest among the 12 strains, while the nonpathogenic strains were all larger than that of the pathogenic strains, which may indicate that the insertion of foreign genes is less prevalent in the evolution of pathogenic strains.

The disease symptoms caused by Cff include water stress and leaf wilt. In the case of severe infection, even if the environmental conditions are suitable, the whole plant will wilt and die (Harveson et al. 2008). Cff reduces grain yield by colonizing xylem vessels, subsequently impeding the translocation of water and nutrients to the superior plant parts (Jeger et al. 2018). In this process, because plant pathogens must break through the barrier of plant cell wall, the existence of enzymes that degrade cell wall is very important for pathogenicity. Therefore, three genes which encoding pectate lyase, serine protease and cellulases, were inferred as virulence-associated (Chen et al. 2020; Osdaghi et al. 2018; Goncalves et al. 2019). And the presence of the three genes in 12 *C. flaccumfaciens* strains is shown in Fig. 5. Pectate lyase is an important enzyme that breaks down and attacks xylem vessels. Pectate lyase (protein id: QFS80865.3) was only found in the three Cff strains and LMG3645 strain. At present, while the role of serine proteinase (protein ids: QIH95653.1, QIH95654.1, and QFS80891.2) in the pathogenicity of Cff has not been elucidated, four, four, 10, and 10 serine proteases were found in the GIs of P900, CFBP3418, Cff1037, and LMG3645, respectively. Serine protease mainly involved in the breakdown of peptide bonds to amino acids required for nutritional purposes or to degrade proteins in the plant cell wall, allowing the bacterial translocation or overcoming plant chemical defenses (Dow et al. 1990). This indicates that

serine proteases may have a significant effect on the pathogenicity of *C. flaccumfaciens*. In addition, found in the glycosyl hydrolase family, cellulase plays a key role in degrading cell wall and infecting xylem. Enzymes associated with the glycosylhydrolase family (protein id: QFS80892.1) were found in the GIs of P990, CFBP3418, Cff1037, LMG3645 and S5.26.

In the analysis of 12 *C. flaccumfaciens* strains, these Gram-positive plant pathogens often produced lyases that can cause diseases on host plants. This included the production of pectate lyase, serine protease and cellulases (Gartemann et al. 2008), typically through genes found in the GIs of LMG3645 and Cff. Previously, while the pathogenicity of *C. michiganensis* was shown to be attributed to a genomic island (Jung et al. 2014). Besides, LMG3645 has not been reported to be pathogenic, but the presence of pectate lyase, serine protein and cellulases related genes were found in the GIs of LMG3645, which also confirms the analysis of COG (Table S1). As the LMG3645 was also additionally isolated from *Phaseolus vulgaris*, we can infer that LMG3645 may also be pathogenic. Further experiment on the pathogenic effect of *C. flaccumfaciens* LMG3645 on *P. vulgaris* could be carried out. In addition, the online service PHASTER was used to identify and annotate the protophage sequences in the genomes of 12 strains of *C. flaccumfaciens*. A total of five incomplete prophage sequences were found in the WW7, VKM Ac-1386, S5.26 strains.

4. Conclusions

Cff is an emerging plant pathogen that is spreading rapidly worldwide and occurs in many bean-producing countries. Systems biology approaches, such as genomics, has been used to further reveal its pathogenicity. For 12 strains of *C. flaccumfaciens*, the analyses used in this study showed that strains were associated with an open pangenome, which meant that the size of new genes and pan genome would increase with the addition of new genome sequence. Phylogenetic analysis showed that there was no correlation between phylogeny and pathogenicity of *C. flaccumfaciens*. KAAS annotation of the core genome shows that the basal metabolic pathway, glycolysis/glyconeogenesis metabolic pathway and pentose phosphate pathway are complete while the citrate cycle (TCA cycle) was incomplete, missing the gene encoding 2-oxoglutarate dehydrogenase (EC:1.2.4.2) in all the 12 *C. flaccumfaciens* strains. GIs analysis showed that genes encoding pectate lyase, serine protein and cellulases, which may be related to the pathogenicity of *C. flaccumfaciens*, were found in the GIs of three known Cff pathogenic strains, and based on the analysis of COG and GIs, it was concluded that LMG3645 might also be a pathogenic strain. However, bacteriophages are not abundant in *C. flaccumfaciens*. This study deepens our understanding about the pathogenicity of *C. flaccumfaciens* at the genomic level, and may encourage further exploration thereof.

Declarations

Ethics approval and consent to participate:

Not applicable. This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication:

Not applicable.

Availability of data and materials:

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests:

The authors declare to have no conflicts of interest.

Funding:

Authors' contributions:

Conceptualization, Q.L.; genomic analysis, Q.L.; writing—original draft preparation, Q.L.; writing—review and editing, L.S.; supervision, L.S.; project administration, L.S.; funding acquisition, L.S. All authors have read and agreed to the published version of the manuscript.

Acknowledgements:

Authors' information (optional):

Not Applicable.

References

1. Arndt, David, Jason R. Grant, Ana Marcu, Tanvir Sajed, Allison Pon, Yongjie Liang, David S. Wishart (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research 44: W16-W21.

2. Bakhshi Ganje, Meysam, John Mackay, Mogens Nicolaisen, Masoud Shams-Bakhsh (2021) Comparative genomics, pangenome, and phylogenomic analyses of *Brenneria spp.*, and delineation of *Brenneria izadpanahii sp. nov.* Phytopathology 111: 78-95.

3. Bertelli, Claire, Matthew R. Laird, Kelly P. Williams, Britney Y. Lau, Gemma Hoad, Geoffrey L. Winsor, Fiona S. L., Brinkman Grp Simon Fraser Univ Res Comp (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Research 45: W30-W35.

4. Chaudhari, Narendrakumar M., Vinod Kumar, Gupta Chitra Dutta (2016) BPGA- an ultra-fast pan-genome analysis pipeline. Scientific reports 6.

5. Chen, G., M. Khojasteh, A. Taheri-Dehkordi, S. M. Taghavi, T. Rahimi, E. Osdaghi (2020) Complete genome sequencing provides novel insight into the virulence repertories and phylogenetic position of dry beans pathogen *Curtobacterium flaccumfaciens pv. flaccumfaciens*. Phytopathology: PHYTO06200243R.

6. Dow, J. M., B. R. Clarke, D. E. Milligan, J. L. Tang, M. J. Daniels (1990) Extracellular proteases from xanthomonas-campestris pv campestris, the black rot pathogen. Applied and Environmental Microbiology 56: 2994-2998.

7. Frost, L. S., R. Leplae, A. O. Summers, A. Toussaint (2005) Mobile genetic elements: The agents of open source evolution. Nature Reviews Microbiology 3: 722-732.

8. Gartemann, Karl-Heinz, Birte Abt, Thomas Bekel, Annette Burger, Jutta Engemann, Monika Fluegel, Lars Gaigalat, Alexander Goesmann, Ines Graefen, Jorn Kalinowski, Olaf Kaup, Oliver Kirchner, Lutz Krause, Burkhard Linke, Alice McHardy, Folker Meyer, Sandra Pohle, Christian Rueckert, Susanne Schneiker, Eva-Maria Zellermann, Alfred Puehler, Rudolf Eichenlaub, Olaf Kaiser, Daniela Bartels (2008) The genome sequence of the tomato-pathogenic actinomycete Clavibacter michiganensis subsp michiganensis NCPPB382 reveals a large island involved in pathogenicity. Journal of Bacteriology 190: 2138-2149.

9. Golanowska, Malgorzata, Marta Potrykus, Agata Motyka-Pomagruk, Michal Kabza, Giovanni Bacci, Marco Galardini, Marco Bazzicalupo, Izabela Makalowska, Kornelia Smalla, Alessio Mengoni, Nicole Hugouvieux-Cotte-Pattat, Ewa Lojkowska (2018) Comparison of Highly and Weakly Virulent Dickeya solani Strains, With a View on the Pangenome and Panregulon of This Species. Frontiers in Microbiology 9.

10. Golicz, Agnieszka A., Philipp E. Bayer, Prem L. Bhalla, Jacqueline Batley, David Edwards (2020) Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. Trends in Genetics 36: 132-145.

11. Goncalves, R. M., M. I. Balbi-Pena, J. M. Soman, A. C. Maringoni, G. Taghouti, M. Fischer-Le, Saux P. Portier (2019) Genetic diversity of *Curtobacterium flaccumfaciens* revealed by multilocus sequence analysis. European Journal of Plant Pathology 154: 189-202.

12. Goncalves, Ricardo M., Carlos A. Schipanski, Lincom Koguishi, Jose M. Soman, Renate K. Sakate, Tadeu A. F. Silva, Junior Antonio C. Maringoni (2017) Alternative hosts of *Curtobacterium flaccumfaciens pv. flaccumfaciens*, causal agent of bean bacterial wilt. European Journal of Plant Pathology 148: 357-365.

13. Goncalves, Ricardo Marcelo, Tadeu Antonio Fernandes da Silva Junior, Jose Marcelo Soman, Joao Cesar da Silva, Antonio Carlos Maringoni (2021) Effect of crop rotation on common bean cultivars

against bacterial wilt caused by *Curtobacterium flaccumfaciens pv. flaccumfaciens*. European Journal of Plant Pathology.

14. Harveson, R. M., A. K. Vidaver (2008) A new color variant of the dry bean bacterial wilt pathogen (*Curtobacterium flaccumfaciens pv. flaccumfaciens*) found in Western Nebraska. Plant Health Progress: PHP-2008-0815-01-BR.

15. Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernandez-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, Ivica Letunic, Thomas Rattei, Lars J. Jensen, Christian von Mering, Peer Bork (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47: D309-D314.

16. Jeger, Michael, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Jean-Claude Gregoire, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Bjoern Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van der Werf, Jonathan West, Stephan Winter, Stefania Tegli, Gabor Hollo, David (2018) Pest categorisation of *Curtobacterium flaccumfaciens pv. flaccumfaciens*. Efsa Journal 16.

17. Jung, W. J., F. Mabood, A. Souleimanov, L. G. Whyte, T. D. Niederberger, D. L. Smith (2014) Antibacterial activity of antagonistic bacterium Bacillus subtilis DJM-51 against phytopathogenic Clavibacter michiganense subsp michiganense ATCC 7429 in vitro. Microbial Pathogenesis 77: 13-16.

18. Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, Koichiro Tamura (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35: 1547-1549.

19. Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins (2007) Clustal W and clustal X version 2.0. Bioinformatics 23: 2947-2948.

20. Leao, E. U., J. C. da Silva, F. R. Medeiros, G. S. S. R. Macedo, G. C. Adorian, A. C. Maringoni (2016) In vitro potential of Bacillus spp. in the control of *Curtobacterium flaccumfaciens pv. flaccumfaciens* in common bean. Summa Phytopathologica 42: 360-362.

21. Loper, Joyce E., Karl A. Hassan, Dmitri V. Mavrodi, Edward W. Davis, II, Chee Kent Lim, Brenda T. Shaffer, Liam D. H. Elbourne, Virginia O. Stockwell, Sierra L. Hartney, Katy Breakwell, Marcella D. Henkels, Sasha G. Tetu, Lorena I. Rangel, Teresa A. Kidarsa, Neil L. Wilson, Judith E. van de Mortel, Chunxu Song, Rachel Blumhagen, Diana Radune, Jessica B. Hostetler, Lauren M. Brinkac, A. Scott Durkin, Daniel A. Kluepfel, W. Patrick Wechter, Anne J. Anderson, Young Cheol Kim, Leland S. Pierson, III, Elizabeth A. Pierson, Steven E. Lindow, Donald Y. Kobayashi, Jos M. Raaijmakers, David M. Weller, Linda S. Thomashow, Andrew E. Allen, Ian T. Paulsen (2012) Comparative genomics of plant-associated *Pseudomonas spp.*: Insights into Diversity and Inheritance of Traits Involved in Multitrophic Interactions. Plos Genetics 8.

22. Moriya, Yuki, Masumi Itoh, Shujiro Okuda, Akiyasu C. Yoshizawa, Minoru Kanehisa (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35: W182-W185.

23. Motyka-Pomagruk, Agata, Sabina Zoledowska, Agnieszka Emilia Misztak, Wojciech Sledz, Alessio Mengoni, Ewa Lojkowska (2020) Comparative genomics and pangenome-oriented studies reveal high homogeneity of the agronomically relevant enterobacterial plant pathogen *Dickeya solani*. Bmc Genomics 21.

24. Osdaghi, E., S. M. Taghavi, S. Calamai, C. Biancalani, M. Cerboneschi, S. Tegli, R. M. Harveson (2018) Phenotypic and Molecular-Phylogenetic Analysis Provide Novel Insights into the Diversity of *Curtobacterium flaccumfaciens*. Phytopathology 108: 1154-1164.

25. Osdaghi, E., S. M. Taghavi, H. Hamzehzarghani, A. Fazliarab, R. M. Harveson, S. Tegli, J. R. Lamichhane (2018) Epiphytic *Curtobacterium flaccumfaciens* strains isolated from symptomless solanaceous vegetables are pathogenic on leguminous but not on solanaceous plants. Plant Pathology 67: 388-398.

26. Osdaghi, Ebrahim, S. Mohsen Taghavi, Habiballah Hamzehzarghani, Amal Fazliarab, Robert M. Harveson, Jay Ram Lamichhane (2016) Occurrence and characterization of a new red-pigmented variant of *Curtobacterium flaccumfaciens*, the causal agent of bacterial wilt of edible dry beans in Iran. European Journal of Plant Pathology 146: 129-145.

27. Puia, Jacqueline Dalbelo, Merilin Gabrieli Dal Ben Ferreira, Adriano Thibes Hoshino, Leandro Camargo Borsato, Marcelo Giovanetti Canteri, Sandra Cristina Vigo (2021) Occurrence of *Curtobacterium flaccumfaciens pv. flaccumfaciens* in the state of Parana and its pathogenicity in beans. European Journal of Plant Pathology.

28. Stice, Shaun P., Spencer D. Stumpf, Ron D. Gitaitis, Brian H. Kvitko, Bhabesh Dutta (2018) Pantoea ananatis Genetic Diversity Analysis Reveals Limited Genomic Diversity as Well as Accessory Genes Correlated with Onion Pathogenicity. Frontiers in Microbiology 9.

29. Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. M. Y. Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. W. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. B. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, C. M. Fraser (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America 102: 13950-13955.

30. Thapa, S. P., E. W. Davis, Q. Lyu, A. J. Weisberg, D. M. Stevens, C. R. Clarke, G. Coaker, J. H. Chang (2019) The evolution, ecology, and mechanisms of infection by gram-positive, plant-associated bacteria in J. E. Leach, S. E. Lindow (eds.), Annual Review of Phytopathology 57.

31. Tumbarski, Y., V. Georgiev, R. Nikolova, A. Pavlov (2018) Isolation, identification and antibiotic susceptibility of *Curtobacterium flaccumfaciens* strain PM_YT from Sea daffodil (*Pancratium maritimum L.*) shoot cultures. Journal of Microbiology, Biotechnology and Food Sciences 7: 623-627.

32. Valdo, S. C. D., A. Wendland, L. G. Araujo, L. C. Melo, H. S. Pereira, P. G. Melo, L. C. Faria (2016) Differential interactions between *Curtobacterium flaccumfaciens pv. flaccumfaciens* and common bean. Genetics and molecular research : GMR 15.

33. Zhao, Yongbing, Xinmiao Jia, Junhui Yang, Yunchao Ling, Zhang Zhang, Jun Yu, Jiayan Wu, Jingfa Xiao (2014) PanGP: A tool for quickly analyzing bacterial pan-genome profile. Bioinformatics 30: 1297-1299.

Tables

Table 1
Genome information of 12 C. flaccumfaciens isolates used in this study

Strain#	Source	Size (Mb)	GC%	Genes	Protein	Assembly
P990	Capsicum sp.: dry beans	3.93	70.37	3740	3625	GCA_009363215.3
Cff1037	Phaseolus vulgaris	3.78	70.95	3585	3500	GCA_013348005.1
CFBP3418	Phaseolus vulgaris	3.83	71.00	3627	3570	GCA_004103915.1
LMG 3645	Phaseolus vulgaris	3.83	71.00	3668	3579	GCA_013359815.1
VKM Ac-1386	Agrostis sp.	3.90	70.90	3762	3687	GCA_015351095.1
VKM Ac-1795	Agrostis sp.	3.94	70.70	3772	3683	GCA_015351035.1
208	Arthrocereusglaziovii: stalk	3.83	70.92	3642	3579	GCA_014156405.1
MEB126	Arabidopsis thaliana: leaf	3.68	70.90	3520	3441	GCA_000834535.1
UCD-AKU	Residential carpet	3.71	70.80	3556	3434	GCA_000349565.1
S5.26	Ice-wedge polygon	3.81	70.70	3637	3588	GCA_006438835.1
JUb65	not applicable	3.80	71.50	3633	3514	GCA_004361695.1
WW7	not applicable	3.49	71.40	3345	3237	GCA_011759505.1

Table 2
GIs distributions of the 12 C. flaccumfaciens

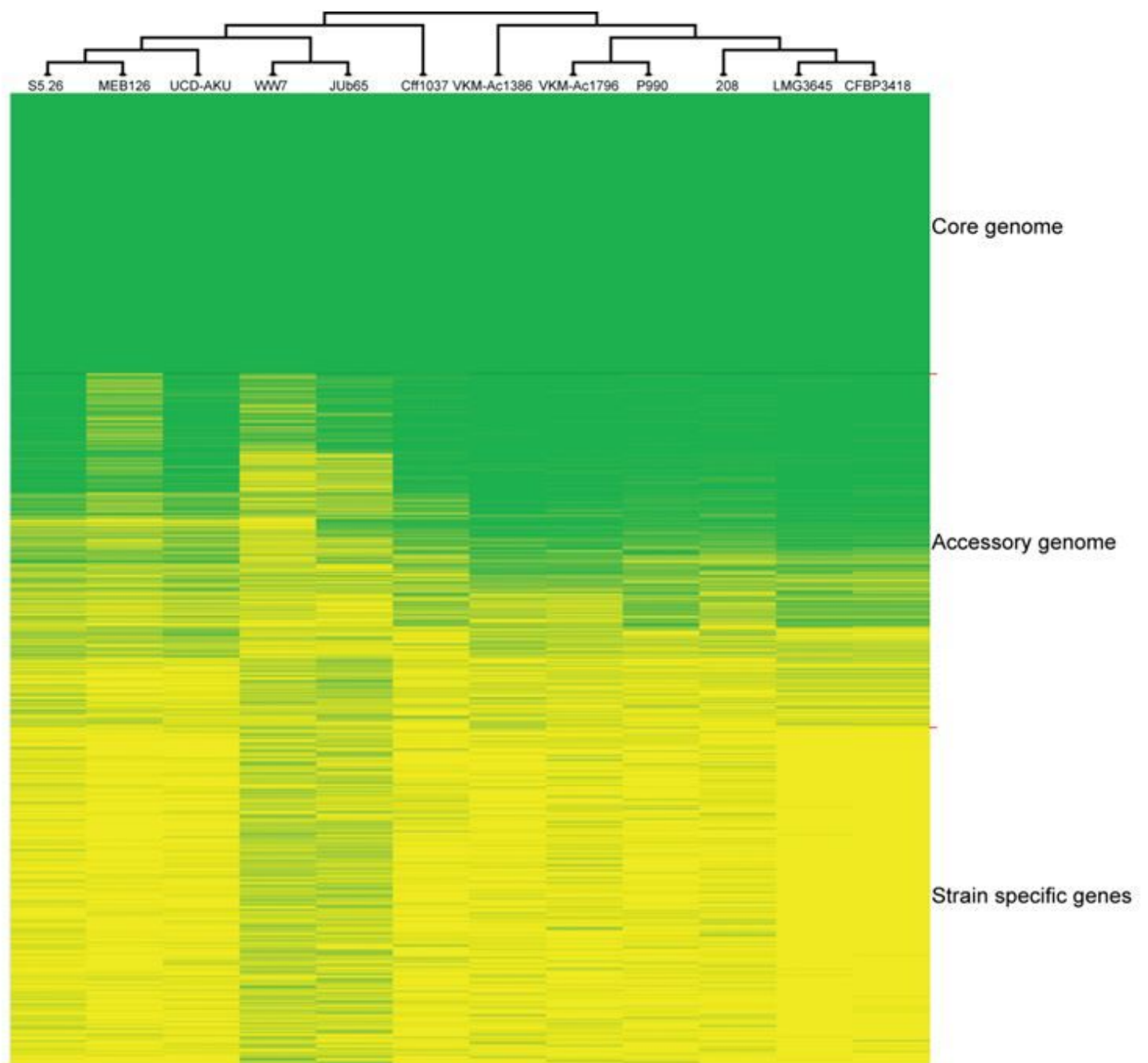| Strain#	Number of GIs	Total proteins	Total length(bp)	Genome size (Mb)	GIs size/Genome size
CFBP3418	10	320	156926	3.82758	4.10%
P990	15	344	167152	3.93170	4.25%
Cff1037	19	396	185856	3.78144	4.91%
LMG 3645	12	392	183651	3.82763	4.80%
UCD-AKU	18	362	193074	3.70532	5.21%
WW7	15	560	226576	3.48645	6.50%
JUb65	18	596	250787	3.79968	6.60%
208	16	518	259828	3.83102	6.78%
MEB126	15	388	266872	3.68442	7.24%
VKM Ac-1386	18	590	275586	3.89885	7.07%
VKM Ac-1795	17	648	347778	3.94493	8.82%
S5.26	27	854	413821	3.81209	10.86%

Figures

Figure 1

The presence of each gene in 12 strains of C. flaccumfaciens. The phylogenetic trees of the 12 C. flaccumfaciens are illustrated at the top of this figure. In the pangenome, the presence of each gene across the 12 C. flaccumfaciens strains is indicated by green, and the absence thereof is indicated by yellow. Each column represents the genome of one strain, while rows represent one gene. The row where all genes are represented in green indicate the core genes, the rows where accessory genes are located have at least two strains whose positions are green, while the row where the strain-specific genes are located have only one strain whose relative positions are green.
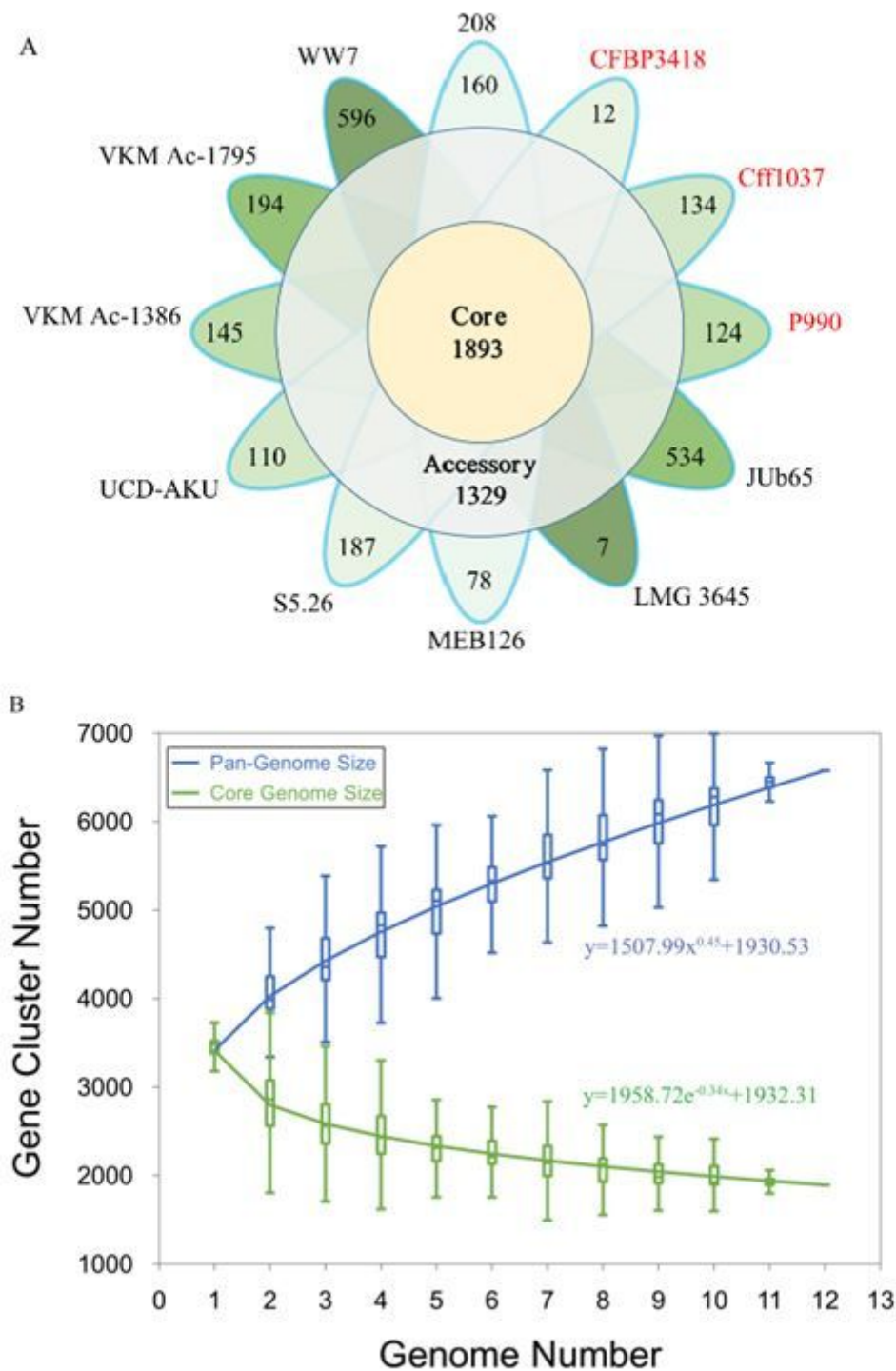
Figure 2

Pangenome of C. flaccumfaciens. (A) Core and accessory genome sizes, and the number of unique genes of each C. flaccumfaciens strain, the names of pathogenic strains are marked in red while nonpathogenic strains in black. (B) Development of pan- and core genome sizes when the genome number of C. flaccumfaciens strains varied from one to12. The strains corresponding to 1-12 on the abscissa axis were as follows: 208, CFBP3418, Cff1037, JUb65, LMG 3645, MEB126, P990, S5.26, UCD-AKU, VKM Ac-1386,

VKM Ac-1795, WW7. The cumulative curve (in blue) supports an open pangenome. Based on Heaps' law, the formula y =αxβ + σ (where y denotes the number of genes of the pangenome; x denotes the analyzed genome number; And α, β and σ are fitting parameters) can be used to calculate the pan-genome size. When 0 < β< 1, the pan-genome is open, Whenβ> 1, the pan-genomic is closed. y =αeβx + σ (where Y denotes the number of genes of the core genome; x denotes the number of analyzed genomes; And α, β and σ are fitting parameters), the size of the core genome can be determined according to the number of pan-genomes.
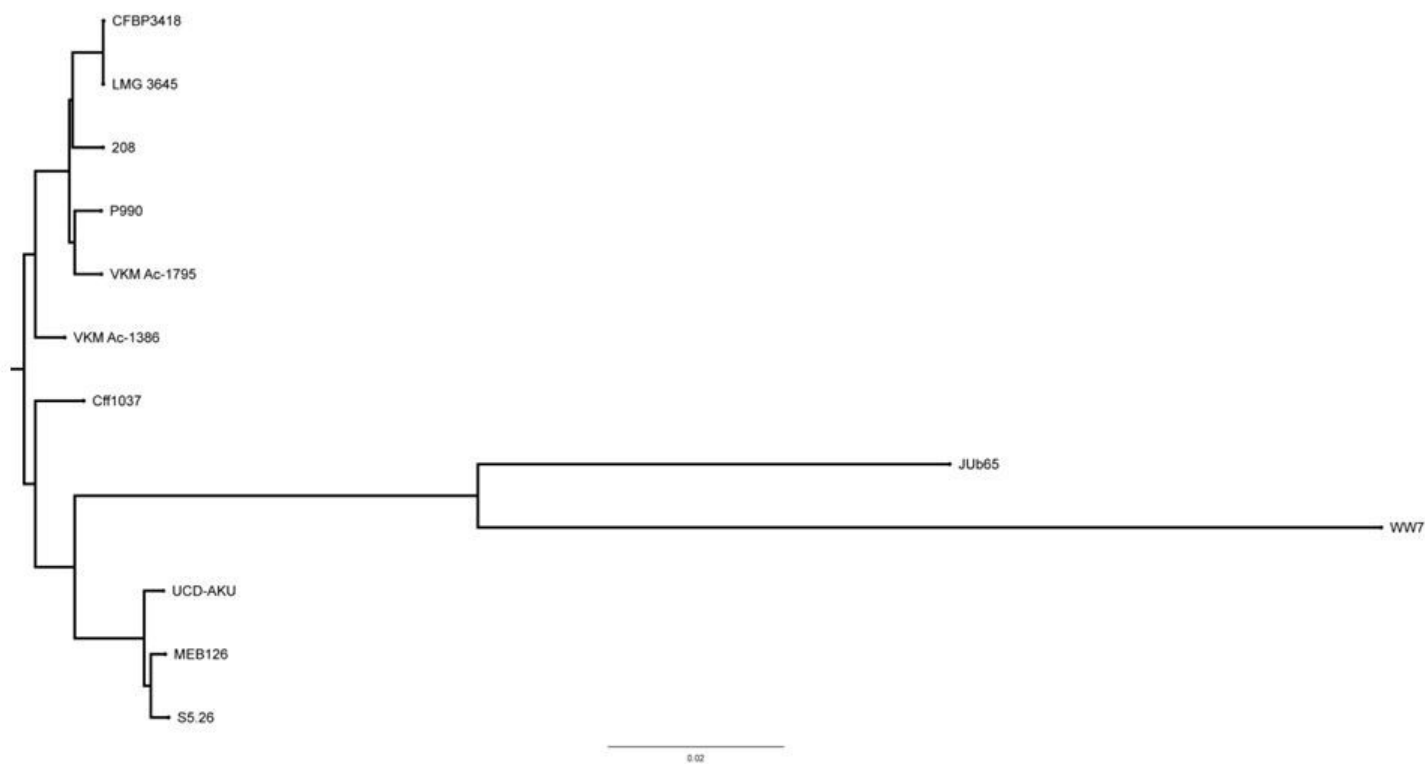

CFBP3418
LMG 3645
208
P990
VKM Ac-1795
VKM Ac-1386
Cff1037
JUb65
WW7
UCD-AKU
MEB126
S5.26

0.02

Figure 3

Phylogenetic trees of 12 C. flaccumfaciens strains based on the concatenated amino acid sequences of the core genomes.
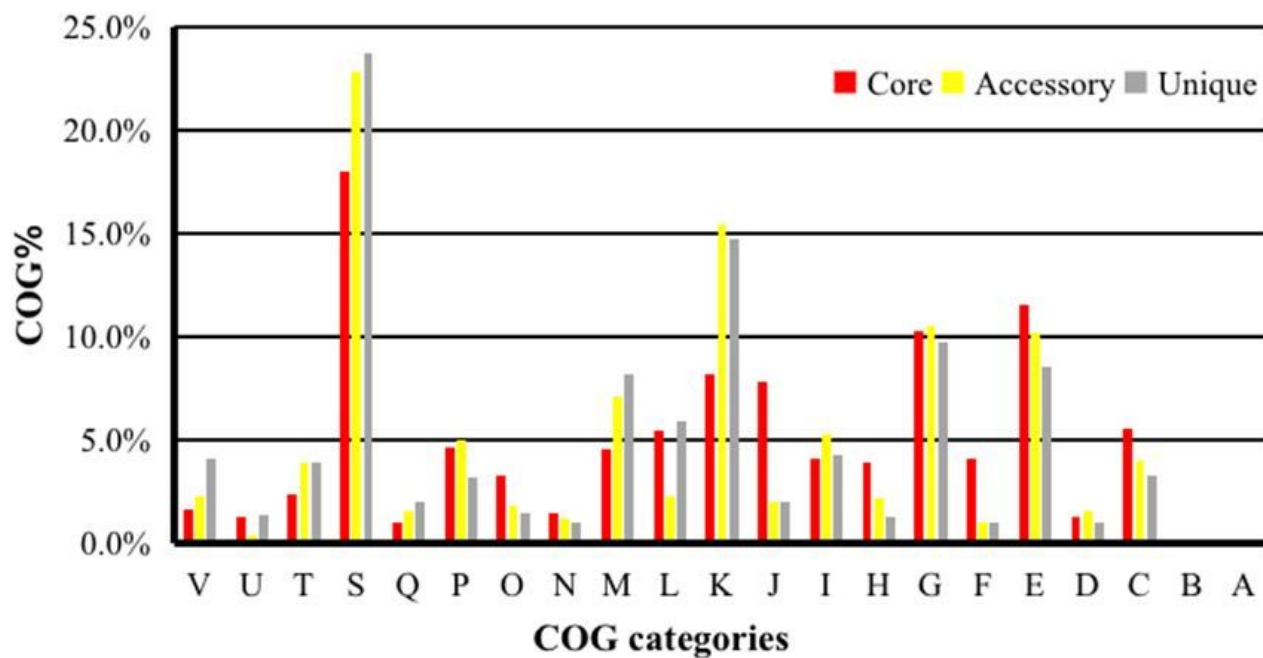
Figure 4

Distribution of COG categories between the core genome, accessory genome, and strain specific gene of C. flaccumfaciens strains. (A) RNA processing and modification; (B) chromatin structure and dynamics; (C) energy production and conversion; (D) cell cycle control, cell division, chromosome partitioning; (E) amino acid transport and metabolism; (F) nucleotide transport and metabolism; (G) carbohydrate transport and metabolism; (H) coenzyme transport and metabolism; (I) lipid transport and metabolism; (J) translation, ribosomal structure, and biogenesis; (K) transcription; (L) replication, recombination, and repair; (M) cell wall/membrane/envelope biogenesis; (N) cell motility; (O) posttranslational modification, protein turnover, chaperones; (P) inorganic ion transport and metabolism; (Q) secondary metabolite biosynthesis, transport, and catabolism; (S) function unknown; (T) signal transduction mechanisms; (U) intracellular trafficking, secretion, and vesicular transport; (V) defense mechanisms; and (W) extracellular structures.
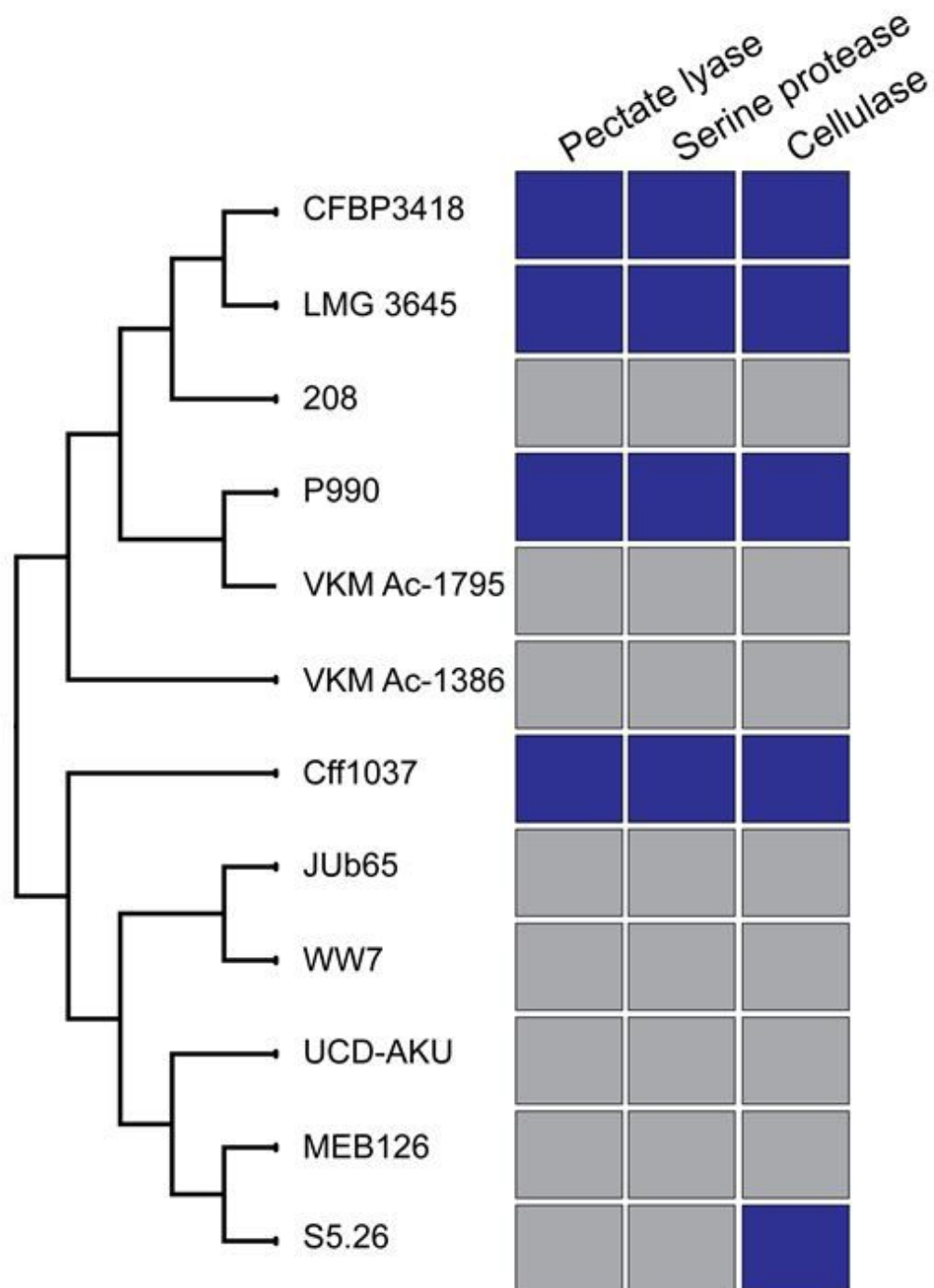
Figure 5

The existence of three enzymes (pectate lyase, serine protease and cellulases) in 12 strains of C. flaccumfaciens. The presence of the enzyme in strain is indicated by blue, while the absence is indicated by grey.

Supplementary Files

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

- TableS1.docx