SYRINGAE: A web-based application for *Pseudomonas syringae* isolate characterization

Chad Fautt[1,2,3], Estelle Couradeau[2,3] & Kevin L. Hockett[1,3]

1. Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, University Park, Pennsylvania, United States
2. Department of Ecosystem Science and Management, Pennsylvania State University, University Park, Pennsylvania, United States
3. Intercollege Graduate Degree Program in Ecology, Pennsylvania State University, University Park, Pennsylvania, United States

Abstract

The *Pseudomonas syringae* species complex (PSSC) is a diverse group of plant pathogens with a collective host range encompassing almost every food crop grown throughout the world. As a threat to global food security, rapid detection and characterization of epidemic and emerging pathogenic lineages is essential. However, phylogenetic identification and prediction of virulence is often complicated by an unclarified taxonomy and the diversity of virulence factor repertoires carried by PSSC isolates. To address these issues, we have built SYRINGAE (www.syringae.org), a web-based phylogenetic placement and functional inference pipeline for PSSC. SYRINGAE contains a comprehensive phylogeny of 2,161 quality-checked genome assemblies annotated with 120 virulence genes. From this dataset, naïve Baye classification models trained from life identification numbers (LINs) and common marker gene sequences can be used for accurate identification of isolates. SYRINGAE efficiently articulates taxonomical and functional data generated over the last several decades on PSSC and constitutes a unique tool tailored towards the rapid characterization of PSSC emerging strains of concern.

Background & Summary

The *Pseudomonas syringae* species complex (PSSC) has been co-evolving with plants since before the emergence of angiosperms (Xin, Kvitko and He, 2018), and has diversified into one of the most economically important groups of plant pathogens in the world, with a collective host range spanning almost every major food crop grown today (Baltrus, McCann and Guttman, 2017). Critically, while there are many pathogens within PSSC, there is also a wide range of virulence exhibited throughout the species complex, including non-pathogenic plant epiphytes

and clades associated with the water cycle (Morris *et al.*, 2007, 2008). The ability to discriminate between lineages within the PSSC and rapidly predict potential pathogenicity of novel lineages is crucial for preventing epidemic outbreaks (Cunty *et al.*, 2015), detecting emerging pathogenic strains (Zhao *et al.*, 2022), and untangling correlations between virulence factors carried by a pathogen, its host range, and its virulence (Preston, 2000). Although the efforts to catalog PSSC diversity and to understand the molecular determinants of virulence have yielded great insights into their ecology and behavior (Morris, Monteil and Berge, 2013), currently there is no efficient way to leverage these insights to efficiently predict the identity and pathogenicity of newly discovered PSSC strains. This is especially true for those researchers or labs that do not specialize on PSSC.

A major barrier to the characterization of PSSC strains is the inconclusive or inaccurate taxonomic identities of published genomes. By one estimate, 42% of all published PSSC genomes are misclassified at the species level, based on analysis of phylogenetic relationships described by average nucleotide identity (ANI) and multi-locus sequence analysis (MLSA) (Gomila *et al.*, 2017). As genomes deposited in databases such as GenBank often serve as reference sequences for identification of isolates found on or near diseased plants, the high rate of misclassification has a direct, negative impact on our ability to efficiently recognize pathogenic lineages. The designation of 13 phylogroups based on MLST has clarified phylogenetic relationships within PSSC (Berge *et al.*, 2014), however most published genomes aren't ascribed to a phylogroup in public databases and thus their use in classification is limited. While Berge et al. have addressed this shortcoming by providing a reference database of phylogroup type strains allowing classification based on the CTS gene, there has since been no broader effort to make the classification process more efficient. Yet another approach to

circumvent the inaccurate taxonomy at the species level while allowing for placement into clades below the species and phylogroup level is the clustering of genomes by ANI (Average Nucleotide Identity) (Vinatzer *et al.*, 2017). This approach assigns a life identification number (LIN) to each unique genome in a database, creating hierarchical clusters of genomes that largely recapitulate traditional phylogenetic clades described by the core genome, and allow for higher resolution than traditional PSSC taxonomy (Figure 1).

A second barrier to characterization of new PSSC isolates is the functional diversity exhibited throughout the species complex. Specifically, host range and virulence can vary considerably among pathogenic strains belonging to the same pathovar, while strains belonging to different pathovars can nonetheless exhibit similar host ranges. These complex patterns stem, at least in part, from the formal definition of pathovar as 'a strain or set of strains with the same or similar characteristics, differentiated at infrasubspecific level from other strains of the same species or subspecies on the basis of distinctive pathogenicity to one or more plant hosts.' which allows for broad interpretation. As such, some pathovars, such as pv. avii, have been described by their ability to cause disease on a single host (10.1023/a:1024786201793), some as having different host ranges among a small defined group of hosts (P. savastanoi pvs. savastanoi, nerii, fraxini, mandevillae and retacarpa, https://doi.org/10.1094/PHYTO-11-20-0526-R), while it has also been argued that pathogens sharing a wide common host range, regardless of a shared pathogenic potential for any single host, should also be considered as belonging to a single pathovar (Young, 2008). Given the inconsistent criteria for delineating between pathovars, and recent evidence that host ranges in PSSC strains overlap with no discernable modularity (Morris et al., 2019), some groups have called into question the validity of pathovar designations for epidemiological and disease management purposes (Lamichhane et al, 2014). Further, properly

assigning a given isolate to an appropriate pathovar requires performing host range tests that are prohibitively laborious to many labs. An alternative phylogenomic approach to predicting pathogenic potential would be beneficial, as others have demonstrated that comparative genomics can discriminate between strains known to have different host ranges (Moreno-Pérez et al., 2020) and correctly identify strains capable of infecting a given host (Almeida et al., 2022). In both of the above cases, known virulence factors, particularly those associated with the type III secretion system (T3SS), were highly correlated with known virulence patterns. Assuming T3SS effector proteins are conserved at some phylogenetic level, these results indicate that a phylogenomic signal may be present in PSSC that would be useful for assessing pathogenic potential without laborious experimental assays.

In a recent contribution we showed that we could accurately predict the presence of 77 type III effector subfamilies in PSSC isolates with an accuracy of 79-94% (REF), using only single amplicon sequence data. Here, we introduce SYRINGAE, a web application for phylogenetic classification and prediction of carried virulence factors for unknown PSSC isolates using user-submitted amplicon data generated from common PCR primer sets. To build the underlying databases of SYRINGAE, 2,161 quality-checked genomes downloaded from GenBank were screened for known virulence factors associated with the T3SS, type 3 effectors (T3E), and the Woody Host or Pseudomonas (WHOP) region associated with knot formation in woody hosts (Caballo-Ponce *et al.*, 2017). Pairwise ANI values were calculated for all genomes, and a maximum-likelihood phylogeny was constructed from the core genome of PSSC. Using the LIN clustering algorithm as a strict hierarchical classification scheme, classification models were trained and made available on the SYRINGAE website. We believe this resource will be

114 particularly useful for researchers who are new to working with PSSC isolates and do not already
115 specialize on the *P. syringae* group as a focal system.

116

117 **Methods**

118 ***reference PSSC genomes***

119 All genome assemblies classified as '*Pseudomonas syringae* group' were downloaded from the
120 GenBank via NCBI in November 2021. Genomes were checked for completeness and assembly
121 quality with BUSCO v5.3.1 using the pseudomonadales_odb10 lineage database, and genomes
122 with a BUSCO score ≥ 99 were kept for further processing (Fig 2a). A CSV file summarizing
123 each genome was created (and made searchable on the SYRINGAE website) and available at
124 xxx. Data included in this file are NCBI-submitted taxonomic data, type strain designations,
125 phylogroups as assigned in this study, LIN clusters assigned for classification purposes,
126 presence/absence of key virulence factors, and metadata found in each genome's Biosample
127 record.

128 ***Assigning phylogroups to genomes***

129 Phylogroup assignment of each genome was based on ANI shared with previously classified
130 reference strains representing Phylogroups 1a,1b,2a,2b,2c,2d,3,4,5,6,7,9,10,11, and 13 (Berge *et*
131 *al.*, 2014) (supplementary table 1). Reference strains for phylogroups 8 and 12 were not found
132 among the 2,161 genomes characterized by SYRINGAE, either because they were not
133 represented in the GenBank database or did not make it past the BUSCO quality check described
134 above.

A genome was assigned to a given phylogroup if it was the most closely related to the reference strain for that phylogroup, based on ANI. To minimize inaccurate phylogroup assignments, any genome sharing less than 95% ANI to any reference strain was left unassigned.

Assigning LIN clusters to genomes

A significant barrier to PSSC classification is unreliable and inconsistent taxonomic assignments. As such, SYRINGAE utilizes hierarchical clustering based on ANI values as an alternative to the Linnean taxonomy files typically used for Bayesian classification. Pairwise ANI between all genomes was calculated using fastANI v1.33 with default settings. Using the algorithm previously described (Vinatzer *et al.*, 2017), each genome was assigned to LIN cluster (fig 2b). To describe the algorithm briefly, a random genome was designated as belonging to group '0' at every ANI bin (e.g. assuming ANI bins of 80, 90, and 95% would give a LIN number of '0.0.0'). Each subsequent randomly selected genome was assigned a LIN number based on the genome it has the highest ANI with among genomes already assigned a LIN number. If, for example, the second genome selected had an ANI of 92% with the first genome, its LIN number would be assigned as '0.0.1', as it meets the threshold for belonging to the same group as the first genome at the 80 and 90% ANI levels but differs from the first genome at the 95% level, and so a new group '1' is created for it. All genomes were sequentially assigned LIN numbers in this way. For SYINGAE, ANI bins at 1% increments between 80-99% were used. A drawback to using LIN clustering for classification is the LIN number assigned to a given genome is highly dependent on the order of genomes selected for clustering. Thus, results obtained from classification models built with a LIN 'taxonomy' can only be interpreted when used in conjunction with a database outlining LIN number the classifier was trained on. The SYRINGAE website overcomes this limitation by using a precomputed unique set of LIN

numbers for accurate classification purposes and by conveying the results to the user in an interpretable format. To achieve that, SYRINGAE 1) cross-references the predicted cluster represented by a given LIN number with traditional taxonomy and 2) displays its position on a whole-genome phylogenetic tree.

Building Phylogenetic tree

A concatenated gene alignment based on the core genome of PSSC was generated using GTDB-TK (Chaumeil *et al.*, 2020). From this alignment, FastTree2 (Price, Dehal and Arkin, 2010) with default settings was used to construct an approximately maximum-likelihood phylogenetic tree (fig 2c).

Screening genomes for virulence factors of concern

Canonical T3SS genes from PSSC strains DC300 and B728A were used to seed a search for homologs with an 85% identity threshold using the Geneious prime 2019 software package (Dotmatics, 2022).The same method was used to find representative WHOP region gene sequences, using previously annotated genes in strain NCPPB 3335 (Caballo-Ponce *et al.*, 2017) as seeds for the homology search. T3E reference sequences were obtained from Psytec (Laflamme *et al.*, 2020). Reference sequences obtained for each gene were aligned with MAFFT (Katoh and Standley, 2013) using default settings, and alignments were used as input for a more robust identification of homologs in the entire set of 2,161 PSSC genomes using HMMER v3.3.2 (Eddy, 2020) (fig 2d). HMMER output files were manually inspected and based on the E-values associated with known virulence factors, matches with an E-value $< 10^{-20}$ were considered to be statistically significant hits. In instances where two genes were identified as more than one virulence factor (a common occurrence among closely related T3E subfamilies), the identification with the lowest E-value was chosen as the official annotation.

181 ***PSSC primer sets supported by SYRINGAE***

182 Over the last two decades, several PCR primers have been developed, often as part of MLST
183 schemes, for building evolutionary accurate phylogenies and aiding in classification of unknown
184 isolates. More recently, there has been interest in utilizing single amplicon sequences for these
185 purposes. To help in our decision of which primer sets to include in SYRINGAE, as well as to
186 investigate which provide the most value for classification using a single amplicon, we
187 conducted a short but thorough in-silico investigation of 16 commonly used primer sets (Fautt et
188 al., submitted). Briefly, we assessed successful amplification in genomes representing the full
189 diversity of the species complex as we currently know it, concordance between pairwise
190 amplicon distance and whole genome ANI, resolution of naïve Bayes classifiers trained from
191 amplicon data, as well as the potential for functional prediction based on the classification
192 results. Given the results of that study, SYRINGAE supports the following primer sets: GyrB,
193 GapA, CTS, RpoD, (Hwang *et al.*, 2005) and PGI (Yan *et al.*, 2008) (Table 1).

194 ***Training Naïve Bayes classification models***

195 For each marker gene, a classification model was trained using the feature-classifier plugin for
196 QIIME 2. Training naïve Bayes classifiers requires both a list of sequences, and an associated
197 taxonomy file for each sequence (typically in the format 'Order_Pseudomonadales; Family_
198 Pseudomonadaceae; Genus_Pseudomonas; Species_syringae;'). LIN numbers assigned to each
199 genome were used to construct a hierarchical taxonomy, with ANI bins within each LIN number
200 acting as taxa levels, and groups acting as individual taxa (e.g., a taxonomy format of '80%_0;
201 90%_0; 95%_1) (fig 2e).

202

Technical Validation

Genome records included in SYRINGAE were validated for assembly quality using BUSCO, and all genomes scoring < 99 were removed from the dataset. Accuracy of the classification models and functional predictions were investigated and published separately (Fautt et al., submitted). Beyond the T3SS, T3E, and WHOP genes, which were annotated using HMM models built for this study, all gene annotations were taken directly from the NCBI Prokaryotic Genome Annotation Pipeline.

Usage Note

Identify

The primary functionality on syringae.org is the rapid characterization of *Pseudomonas sp.* isolates from single amplicon sequences.

Input:

The query sequence(s) are untrimmed amplicon sequences generated from primer sets targeting the genes gyrB, CTS, gapA, PGI, or rpoD (table 1), in FASTA or multiFASTA format.

Outputs:

1) A list of genomes predicted to be most closely related to the unknown isolate
2) Phylogenetic classification for each query sequence is also displayed as a phylogenetic tree rooted at the most recent common ancestor of all genomes in the Syringae database predicted to be closely related to the unknown isolate. By default, the tree only shows the

most closely related genomes, but users can toggle the ANI threshold scale to lower values, to "zoom out" and display more distant relatives. This extra functionality allows the ability to visualize the predicted placement of the unknown isolate in a larger phylogenetic context. The predicted shared ANI with the unknown isolate is displayed as a radial bar chart around the perimeter of the tree (fig. 3, external ring).

3) For the currently displayed tree, the abundance of represented species, phylogroups, and pathovars are to the left of the tree. (fig. 3)
4) A summary of the virulence factors found among the most closely related of the unknown isolate genomes given the selected ANI threshold is available as a second tab in the results section. (fig. 4)

To quickly learn more about any genomes predicted to be similar to the unknown isolate, users can utilize the *Explore* and *Search* functionalities.

Explore

An interactive visualization tool for exploring the phylogenetic and genetic diversity of PSSC. Users can filter the 2,161-genome phylogenetic tree Syringae uses for visualizing classification data by taxa (phylogroup, species, and pathovar), and annotate up to six features, including multiple taxa and presence/absence of up to 3 NCBI-annotated genes (fig. 5)

Search

A search tool for quickly finding metadata for any genome in SYRINGAE's database. Search results include a list of the closest PSSC relatives in the dataset, as measured by fastANI (fig. 6).

Users can also search for any NCBI-annotated gene names or virulence factors annotated by us for Syringae. Search results include a list of all genomes in our database carrying the gene of interest, as well as each unique protein accession number associated with the gene name found within the species complex.

Figure Legends

Figure 1 Comparison of clustering within PSSC that results from a maximum likelihood phylogenetic tree and LIN assigned based on ANI. Digits from left to right in each LIN correspond to inclusion of a strain in increasingly smaller clades within the phylogeny. Figure adapted from Vinatzer et al., 2017

Figure 2 Schematic of bioinformatic pipeline used for SYRINGAE.

Figure 3 screenshot of classification results provided by Syringae.org. The tree is rooted at the most recent common ancestor of all genomes within the LIN cluster the unknown isolate has been placed into. The exterior ring shows the predicted ANI similarity between the unknown isolate and each reference genome. The interior ring shows phylogroup assigned to each reference genome. Relative abundance of phylogroups, species, and pathovars found within the unknown isolates predicted LIN cluster are shown to the left of the tree.

Figure 4 screenshot of virulence factor prediction provided by Syringae.org. The proportion of genomes within an unknown isolate's predicted LIN cluster that carry canonical type III secretion system genes, type III effectors, and WHOP genes are displayed. Green, yellow, and

red bars denote virulence factors found in >90%, >50%, and <50% of related genomes, respectively.

Figure 5 Screenshot of the Explore functionality on Syringae.org. The form at top allows for filtering of genomes to display by taxa, and annotation by up to 3 simultaneous genes and 4 taxa. Annotation rings surrounding the phylogenetic tree are based on user-selected annotations; in this case showing, from innermost ring outward, phylogroups, pathovars, and absence/presence of effector protein subfamily HopB4.

Figure 6 screenshot of search results for a PSSC genome displaying from top to bottom: metadata associated with the genome, predicted virulence factors, and most closely related genomes.

Table 1 primer sets accepted by Syringae.org for isolate characterization.

References

Almeida, R. N. D. *et al.* (2022) 'Predictive modeling of Pseudomonas syringae virulence on bean using gradient boosted decision trees', *PLOS Pathogens*. Public Library of Science, 18(7), p. e1010716. doi: 10.1371/JOURNAL.PPAT.1010716.

Baltrus, D. A., McCann, H. C. and Guttman, D. S. (2017) 'Evolution, genomics and epidemiology of Pseudomonas syringae: Challenges in Bacterial Molecular Plant Pathology', *Molecular Plant Pathology*. Blackwell Publishing Ltd, pp. 152–168. doi: 10.1111/mpp.12506.

Berge, O. *et al.* (2014) 'A User's Guide to a Data Base of the Diversity of Pseudomonas syringae and Its Application to Classifying Strains in This Phylogenetic Complex', *PLOS ONE*. Public Library of Science, 9(9), p. e105547. doi: 10.1371/JOURNAL.PONE.0105547.

Caballo-Ponce, E. *et al.* (2017) 'WHOP, a Genomic Region Associated With Woody Hosts in the Pseudomonas syringae Complex Contributes to the Virulence and Fitness of Pseudomonas savastanoi pv. savastanoi in Olive Plants', *Ornston*, 30(2), p. 113. doi: 10.1094/MPMI-11-16-0233-R.

Chaumeil, P. A. *et al.* (2020) 'GTDB-Tk: a toolkit to classify genomes with the Genome

Taxonomy Database', *Bioinformatics*. Oxford Academic, 36(6), pp. 1925–1927. doi: 10.1093/BIOINFORMATICS/BTZ848.

Cunty, A. *et al.* (2015) 'Origin of the outbreak in France of Pseudomonas syringae pv. actinidiae biovar 3, the causal agent of bacterial canker of kiwifruit, revealed by a multilocus variable-number tandem-repeat analysis', *Applied and Environmental Microbiology*. American Society for Microbiology, 81(19), pp. 6773–6789. doi: 10.1128/AEM.01688-15/SUPPL_FILE/ZAM999116610SO1.PDF.

Dotmatics (2022) *Geneious*. Available at: https://www.geneious.com/ (Accessed: 31 October 2022).

Eddy, S. R. (2020) *HMMER*. Available at: www.hmmer.org (Accessed: 31 October 2022).

Gomila, M. *et al.* (2017) 'Clarification of taxonomic status within the Pseudomonas syringae species group based on a phylogenomic analysis', *Frontiers in Microbiology*. Frontiers Media S.A., 8(DEC), p. 2422. doi: 10.3389/FMICB.2017.02422/BIBTEX.

Hwang, M. S. H. *et al.* (2005) 'Phylogenetic characterization of virulence and resistance phenotypes of Pseudomonas syringae', *Applied and Environmental Microbiology*. American Society for Microbiology, 71(9), pp. 5182–5191. doi: 10.1128/AEM.71.9.5182-5191.2005/ASSET/8595E3BB-82F4-4338-BA61-23CBE0B7B3A4/ASSETS/GRAPHIC/ZAM0090557940001.JPEG.

Katoh, K. and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*. Oxford Academic, 30(4), pp. 772–780. doi: 10.1093/MOLBEV/MST010.

Laflamme, B. *et al.* (2020) 'The pan-genome effector-triggered immunity landscape of a host-pathogen interaction', *Science (New York, N.Y.)*. Science, 367(6479), pp. 763–768. doi: 10.1126/SCIENCE.AAX4079.

Moreno-Pérez, A. *et al.* (2020) 'Host Range Determinants of Pseudomonas savastanoi Pathovars of Woody Hosts Revealed by Comparative Genomics and Cross-Pathogenicity Tests', *Frontiers in Plant Science*. Frontiers Media S.A., 11, p. 973. doi: 10.3389/FPLS.2020.00973/BIBTEX.

Morris, C. E. *et al.* (2007) 'Surprising niche for the plant pathogen Pseudomonas syringae', *Infection, Genetics and Evolution*. Elsevier, 7(1), pp. 84–92. doi: 10.1016/J.MEEGID.2006.05.002.

Morris, C. E. *et al.* (2008) 'The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle', *The ISME Journal*, 2, pp. 321–334. doi: 10.1038/ismej.2007.113.

Morris, C. E. *et al.* (2019) 'The overlapping continuum of host range among strains in the Pseudomonas syringae complex', *Phytopathology Research*. Springer Science and Business Media LLC, 1(1), p. 4. doi: 10.1186/s42483-018-0010-6.

Morris, C. E., Monteil, C. L. and Berge, O. (2013) ' The Life History of Pseudomonas syringae : Linking Agriculture to Earth System Processes ', *Annual Review of Phytopathology*. Annual Reviews, 51(1), pp. 85–104. doi: 10.1146/annurev-phyto-082712-102402.

Preston, G. M. (2000) 'Pseudomonas syringae pv. tomato: the right pathogen, of the right plant, at the right time', *Molecular Plant Pathology*. John Wiley & Sons, Ltd, 1(5), pp. 263–275. doi: 10.1046/J.1364-3703.2000.00036.X.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments', *PLOS ONE*. Public Library of Science, 5(3), p. e9490. doi: 10.1371/JOURNAL.PONE.0009490.

Vinatzer, B. A. *et al.* (2017) 'A proposal for a genome similarity-based taxonomy for plant-pathogenic bacteria that is sufficiently precise to reflect phylogeny, host range, and outbreak affiliation applied to Pseudomonas syringae sensu lato as a proof of concept', *Phytopathology*. American Phytopathological Society, 107(1), pp. 18–28. doi: 10.1094/PHYTO-07-16-0252-R.

Xin, X. F., Kvitko, B. and He, S. Y. (2018) 'Pseudomonas syringae: What it takes to be a pathogen', *Nature Reviews Microbiology*. Nature Publishing Group, 16(5), pp. 316–328. doi: 10.1038/NRMICRO.2018.17.

Yan, S. *et al.* (2008) 'Role of recombination in the evolution of the model plant pathogen Pseudomonas syringae pv. tomato DC3000, a very atypical tomato strain', *Applied and environmental microbiology*. Appl Environ Microbiol, 74(10), pp. 3171–3181. doi: 10.1128/AEM.00180-08.

Zhao, M. *et al.* (2022) 'Pseudomonas alliivorans sp. nov., a plant-pathogenic bacterium isolated from onion foliage in Georgia, USA', *Systematic and Applied Microbiology*. Urban & Fischer, 45(1), p. 126278. doi: 10.1016/J.SYAPM.2021.126278.

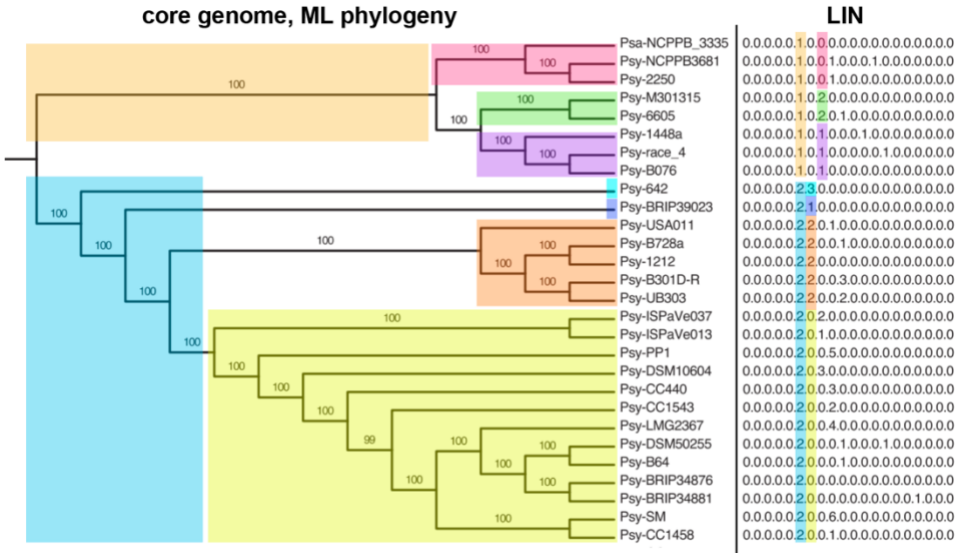


Figure 1

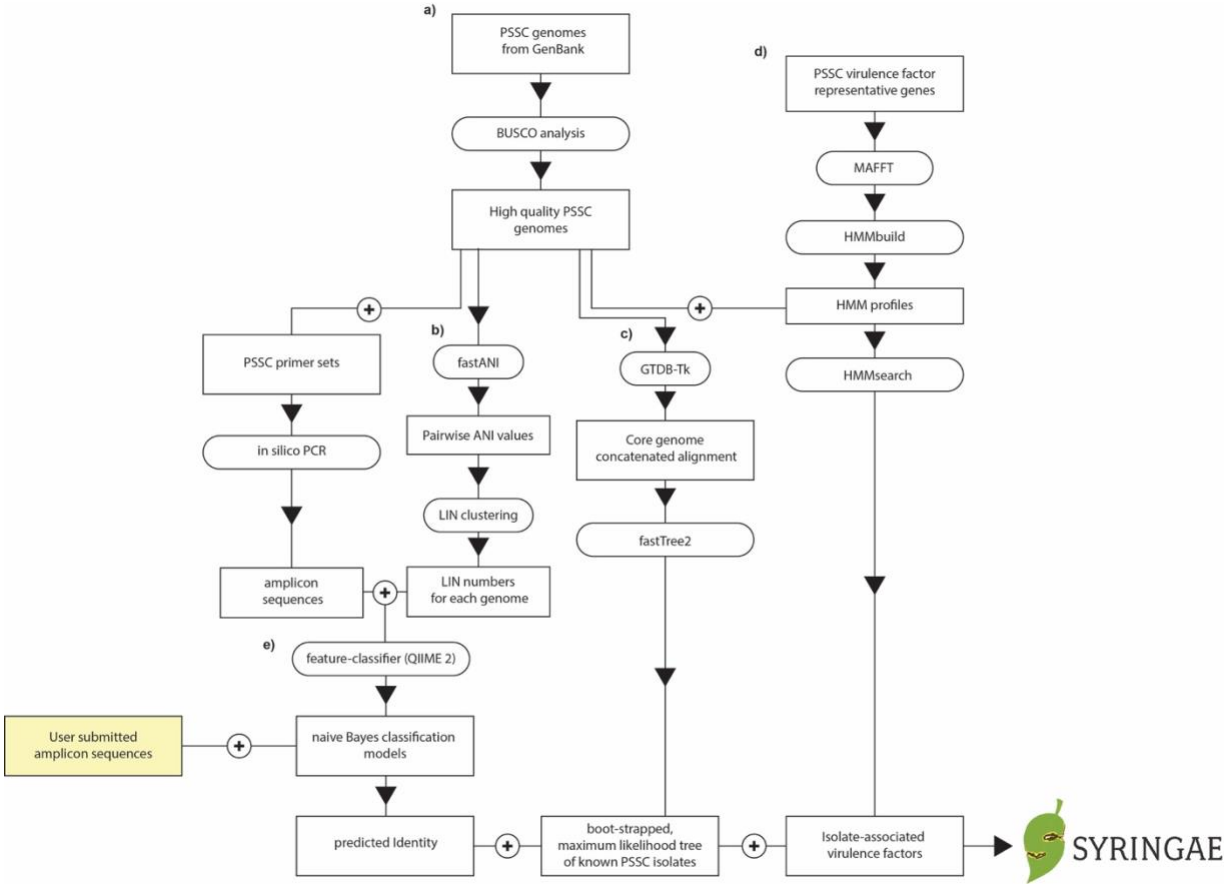
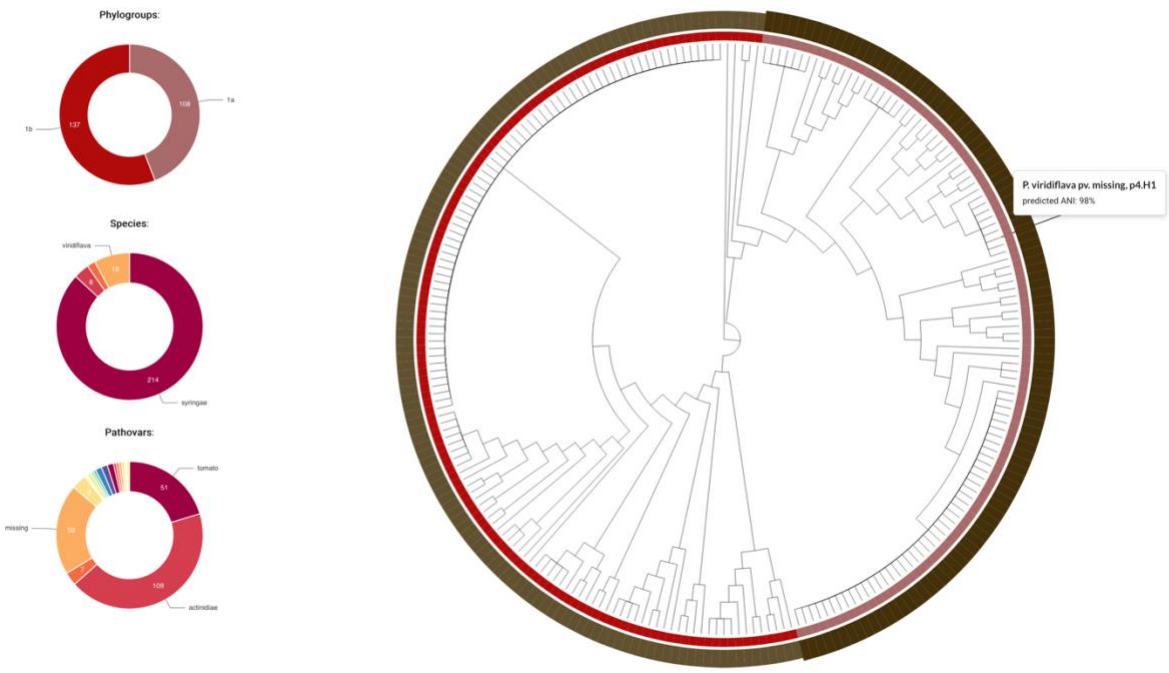


Figure 2

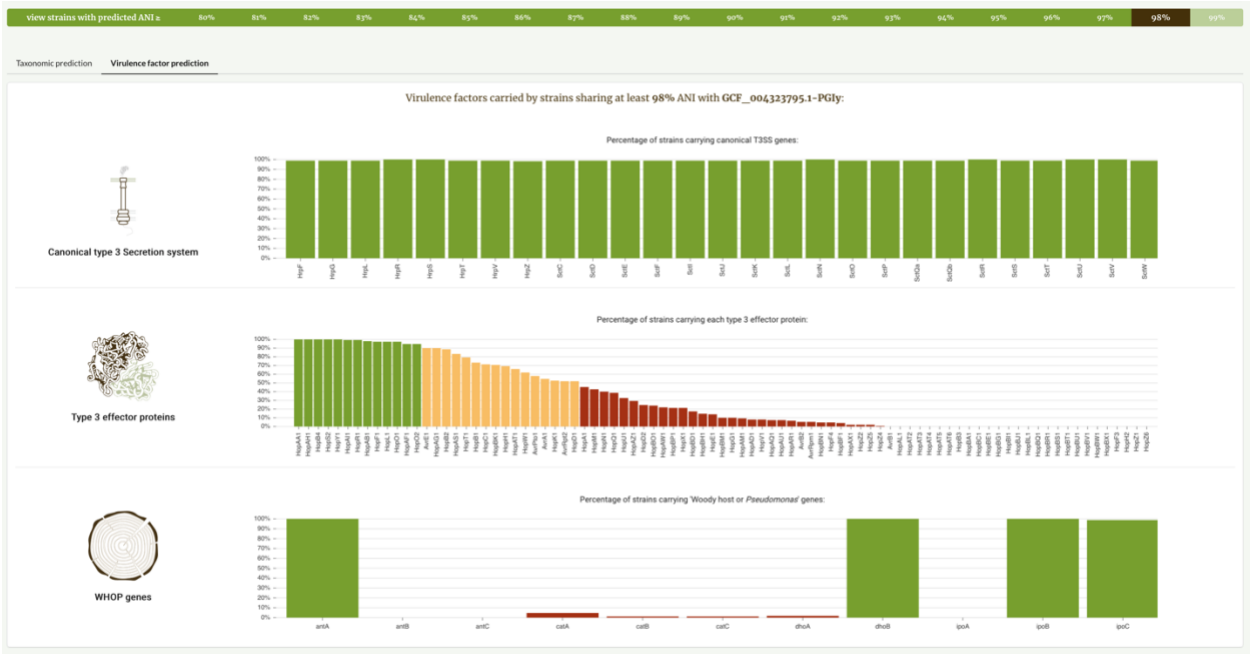
348



349
350

Figure 3

351

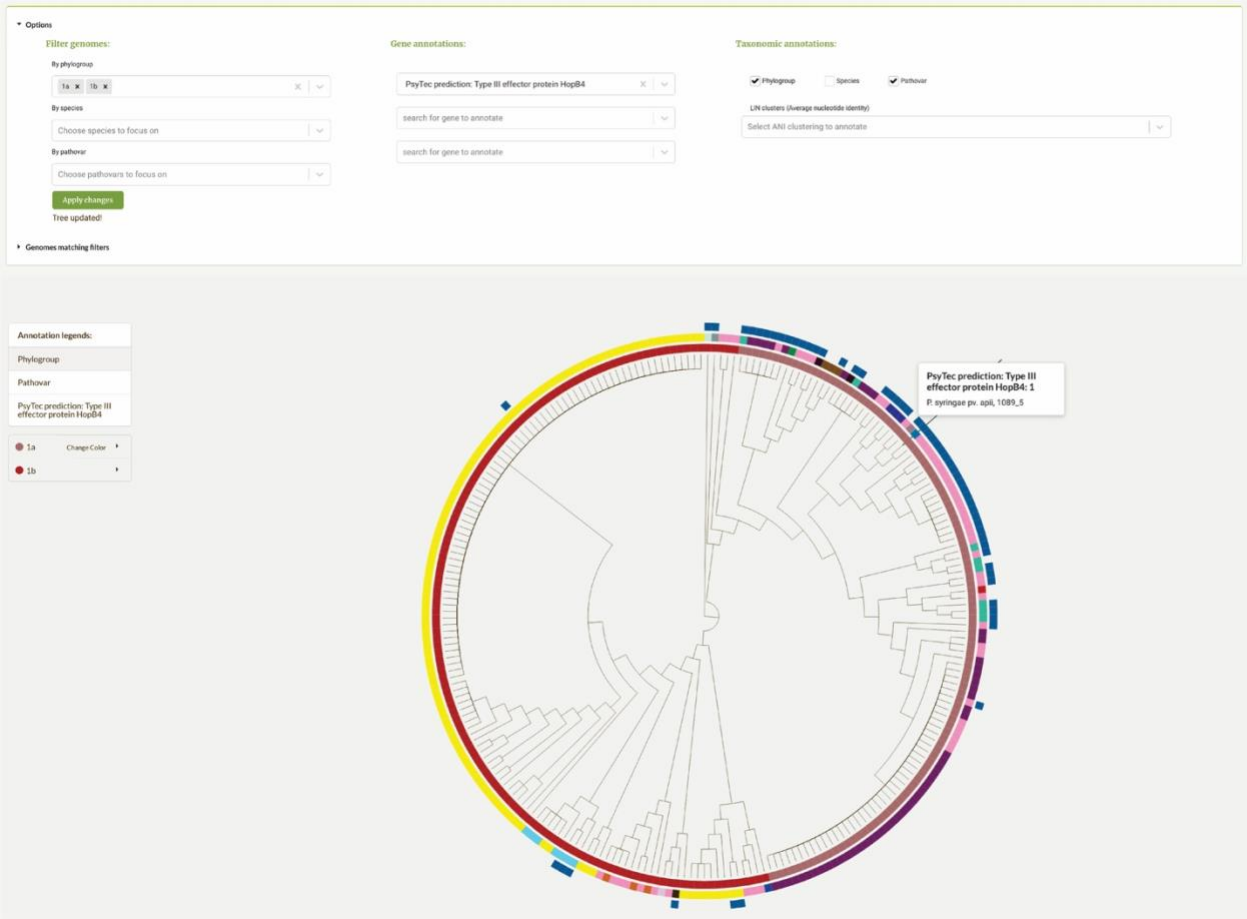


352

353

Figure 4

354



355

356 **Figure 5**

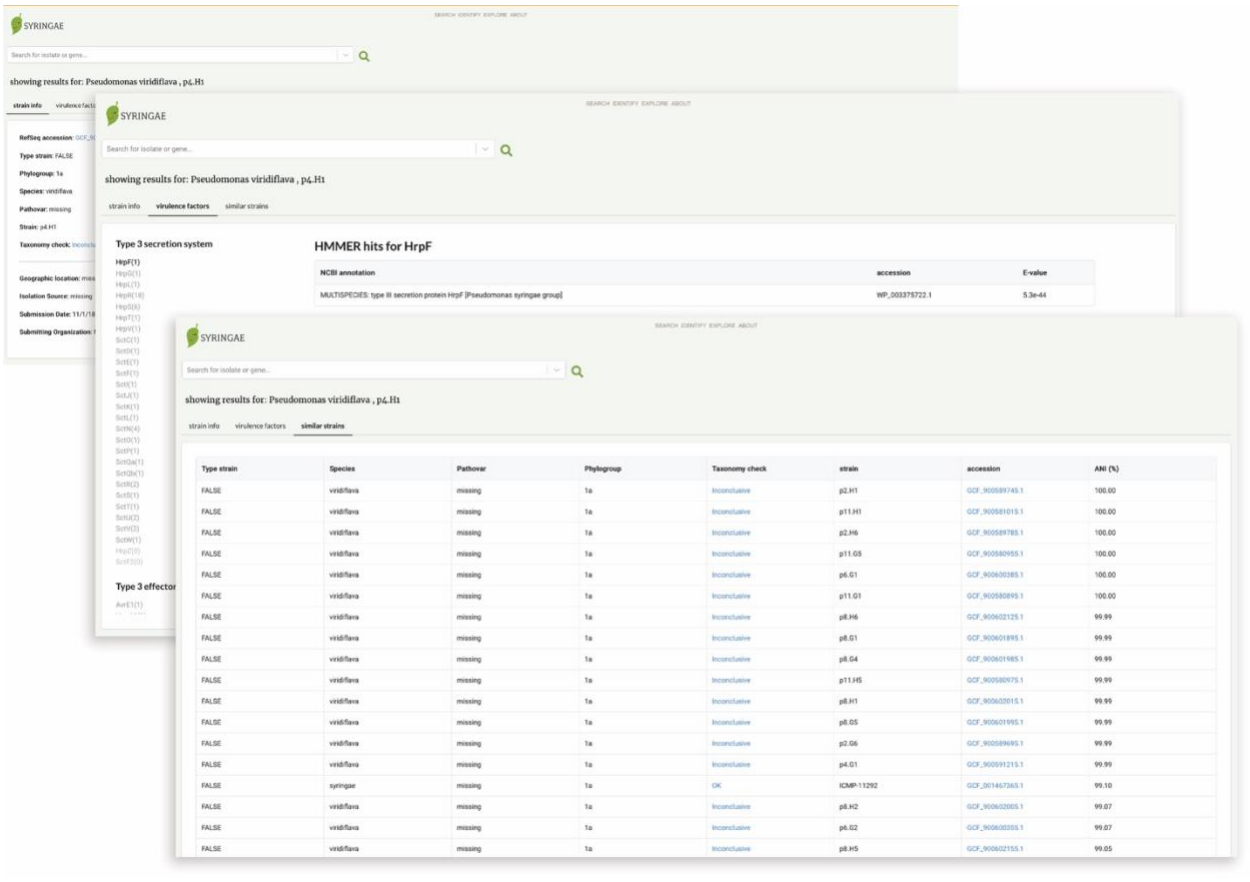


Figure 6

Target gene	Forward sequence (5'-3')	Reverse Sequence (5'-3')	primer names	Source
gapA	TCGARTGCACSGGBCTSTTCACC	GTGTGRTTGGCRTCGAARATCGA	gapA+312s/ gapA−874ps	Hwang et al., 2005
gyrB	TCBGCRGCVGARGTSATCATGAC	TTGTCYTTGGTCTGSGAGCTGAA	gyrB+271ps/ gyrB−1022ps	Hwang et al., 2005
CTS	CCTGRTCGCCAAGATGCCGAC	CGAAGATCACGGTGAACATGCTGG	gltA+513s/ gltA−1130s	Hwang et al., 2005
rpoD	GYGAAGGCGARATYGRAATCG	CCGATGTTGCCTTCCTGGATCAG	rpoD+364s/ rpoD−1222ps	Hwang et al., 2005
PGI	GCGTACTACCGYAMYCCBTC	CCACATMGGRAARATRTTYT	pgi	Yan et al., 2008

368 **Table 1**