



Article

alfaNET: A Database of Alfalfa-Bacterial Stem Blight Protein–Protein Interactions Revealing the Molecular Features of the Disease-causing Bacteria

Raghav Kataria ¹ and Rakesh Kaundal ^{1,2,3,*}

¹ Department of Plants, Soils, and Climate, College of Agriculture and Applied Sciences, Utah State University, Logan, UT 84322, USA; raghav.kataria@usu.edu

² Bioinformatics Facility, Center for Integrated BioSystems, College of Agriculture and Applied Sciences, Utah State University, Logan, UT 84322, USA

³ Department of Computer Science, College of Science, Utah State University, Logan, UT 84322, USA

* Correspondence: rkaundal@usu.edu; Tel.: +1-(435)-797-4117; Fax: +1-(435)-797-2766

Abstract: Alfalfa has emerged as one of the most important forage crops, owing to its wide adaptation and high biomass production worldwide. In the last decade, the emergence of bacterial stem blight (caused by *Pseudomonas syringae* pv. *syringae* ALF3) in alfalfa has caused around 50% yield losses in the United States. Studies are being conducted to decipher the roles of the key genes and pathways regulating the disease, but due to the sparse knowledge about the infection mechanisms of *Pseudomonas*, the development of resistant cultivars is hampered. The database alfaNET is an attempt to assist researchers by providing comprehensive *Pseudomonas* proteome annotations, as well as a host–pathogen interactome tool, which predicts the interactions between host and pathogen based on orthology. alfaNET is a user-friendly and efficient tool and includes other features such as subcellular localization annotations of pathogen proteins, gene ontology (GO) annotations, network visualization, and effector protein prediction. Users can also browse and search the database using particular keywords or proteins with a specific length. Additionally, the BLAST search tool enables the user to perform a homology sequence search against the alfalfa and *Pseudomonas* proteomes. With the successful implementation of these attributes, alfaNET will be a beneficial resource to the research community engaged in implementing molecular strategies to mitigate the disease. alfaNET is freely available for public use at <http://bioinfo.usu.edu/alfanet/>.

Keywords: alfalfa; *Pseudomonas syringae* ALF3; bacterial stem blight; effectors; protein–protein interactions; interolog; database



Citation: Kataria, R.; Kaundal, R. alfaNET: A Database of Alfalfa-Bacterial Stem Blight Protein–Protein Interactions Revealing the Molecular Features of the Disease-causing Bacteria. *Int. J. Mol. Sci.* **2021**, *22*, 8342. <https://doi.org/10.3390/ijms22158342>

Academic Editors: Akira Ishihama and Franklin W.N. Chow

Received: 9 July 2021

Accepted: 26 July 2021

Published: 3 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Alfalfa (*Medicago sativa* L.) is one of the major forage crops in the United States and other countries around the world. The crop is widely cultivated owing to its high biomass production, role in biological nitrogen fixation and soil conservation, and high nutritional value as animal feed [1]. Besides these, various essential secondary metabolites (such as lignin, saponins, and flavonoids) are synthesized in alfalfa. These secondary metabolites have been reported to provide plants resistance against biotic and abiotic stresses [2]. An early accumulation of lignin was observed in the cell wall during the plant defense mechanism against the disease [3]. The intermittent occurrence of bacterial stem blight of alfalfa is common in the central and western US, with many recent reports in Europe, Australia, and western Iran [4]. The disease has been reported to cause a yield loss of 40 to 50% in the first crop harvest [5]. The Gram-negative bacteria, *Pseudomonas syringae*, is classified into at least sixty pathogenic strains, characterized according to their target host. The bacteria possess virulence factors such as a type III secretion system (T3SS) to infect its host and hijack the host transcription machinery [6]. *P. syringae* pv. *syringae* ALF3,

the causal organism of bacterial stem blight, infects its host through the frost mechanism because of the presence of a unique ice-nucleation protein (INP) on the outer membrane of the bacteria, which serves as the nuclei for the disease initiation [7]. The disease occurs in two successive phases: localized foliar necrosis (blight) and systemic vascular wilt. In the initial stage, the bacterial penetration forms water-soaked lesions at the frost injury sites, which further extends and produces dried bacterial exudate throughout the stem. The diseased plants are usually dwarfs, with slender stems that blacken upon maturity [8].

Various research groups in the past have elucidated the genetics behind the interaction mechanism of *P. syringae* with its plant hosts in model plants [9], which involves gene-for-gene resistance, thus triggering a cascade of defense-related signals which eventually lead to disease resistance. Such incompatible host–pathogen interactions also initiate a hypersensitive response (HR) in plants [10]. However, information on the disease resistance mechanism in alfalfa is limited. Researchers in the past have isolated bacteria with symptoms of bacterial stem blight from alfalfa and identified strain ALF3 as *P. syringae* pv. *syringae* on the basis of the 16S rDNA sequence and PCR amplification of *syrB* for toxin production [11]. Following the genome sequencing of the strain ALF3 [8], it was found to be pathogenic in various plants such as *Medicago truncatula*, pear leaves, and bean seed pods [11]. Significant differentially expressed genes and processes associated with the host defense mechanism have also been proposed in alfalfa infected with bacterial stem blight [12]. Studies are being conducted to decipher the role of key genes and pathways regulating the disease, but the sparse availability of knowledge about the infection mechanism of *P. syringae* hampers the development of resistant cultivars.

The protein–protein interactions (PPIs) between a host and pathogen play a crucial role in the infection process and the initiation of a defense response against pathogen attack. Therefore, studying the PPI network between host and pathogen proteins helps us understand the underlying infection mechanism [13]. The identification of potential alfalfa proteins targeted by *P. syringae* will further bolster the management of the disease. There are plentiful resources from which to retrieve host–pathogen PPI data for many plant species [14], but so far no open resources have been developed for the retrieval of information for alfalfa inter-species PPIs. The implementation of host–pathogen interaction databases containing various functional analysis features supports the development of novel disease-resistant cultivars [15,16]. Databases such as *Pseudomonas* Genome Database [17], which provide information about the genome sequences and annotation of different *Pseudomonas* species, are implemented. LegumeIP [18], a database of model legumes, provides comparative genomics and transcriptomics information of various legume crops. However, no tool for the prediction of host–pathogen PPI exists for alfalfa-*P. syringae* systems.

In the present study, we aim to develop an interactomic resource, alfaNET, to provide the research community with a platform to study *Pseudomonas*–alfalfa inter-species interactions. The database holds information about annotations of *P. syringae*, along with the subcellular localization information of the pathogen proteins. Furthermore, alfaNET implements a host–pathogen interactomics tool, allowing the users to exploit the predicted PPIs between alfalfa and *P. syringae*. The tool promises a user-friendly interface, providing users with an enriched network visualization framework for better understanding the host–pathogen systems. alfaNET is freely available for public use at <http://bioinfo.usu.edu/alfanet/>.

2. Results

Database Architecture and Implementation

The sitemap architecture of the alfaNET database is depicted in Figure 1. The tool is hosted on a Linux virtual node which is located in a high-performance computing (HPC) cluster. The implementation of server code is performed in PHP and functions through an Apache server. We used Bootstrap and various JavaScript plugins to develop the front-end, thus enhancing the visual appearance. For data retrieval, separate pages are available

in the database, whereby the user can query information for the required species. The user can further filter this information using keywords and download it in a tab-delimited format, which can be accessed in spreadsheet editors such as Microsoft Excel. Furthermore, the database provides users with a BLAST utility that implements alfalfa and *P. syringae* pv. *syringae* ALF3 protein sequences as BLAST databases at the backend. We also employed BlasterJS [19] for displaying the BLAST alignment visualization.

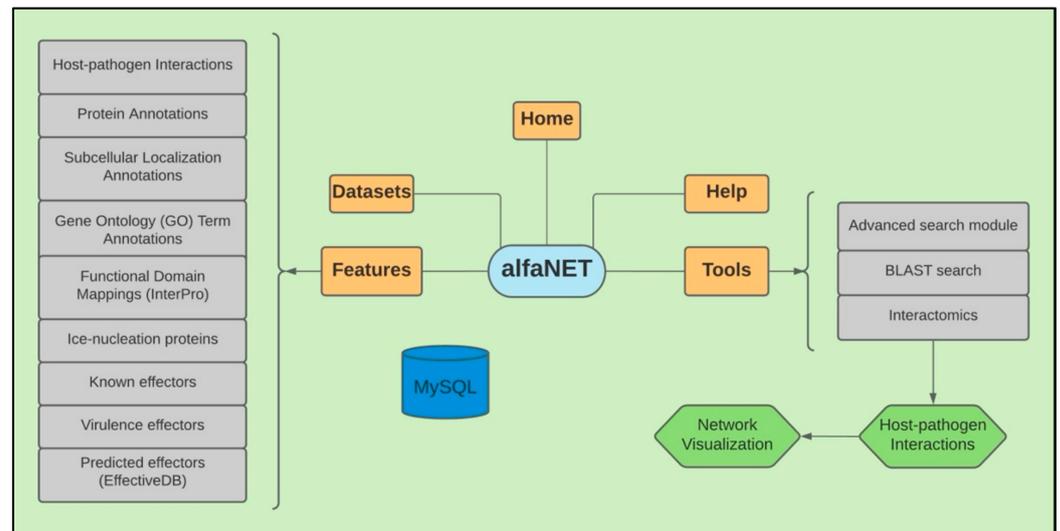


Figure 1. Sitemap architecture of the alfaNET database.

In addition to the utilities mentioned above, the major functionality of alfaNET is the host–pathogen interactome tool (Figure 2), which implements an optimized interolog approach to predict the host–pathogen interactome of the *Medicago* and *Pseudomonas* species using various in-house R scripts and SQL operations in a fraction of a second. The interface of the interactome tool is designed in such a manner that it allows users to select various protein–protein interaction databases to be used, while also providing the alignment filtering options, allowing them to filter the BLAST alignments of both the host and pathogen datasets by adjusting the parameters such as E-value, identity, (%) and coverage (%). On job submission by the user, a unique identifier (Job ID) is assigned, which the user can access to check the job status (queried, active, or completed). After the completion of the job, the results are displayed in an enhanced tabular format with advanced filtering options, enabling users to sort the result data into columns or filter using keywords. The results are available to be downloaded in different formats, such as in pdf or excel, or they can be copied in the clipboard. Additionally, the e-mail address feature in this tool notifies the user of the completion of a job. To visualize the protein interaction networks, SigmaJS [20] is implemented at the backend with various layout options, links, and buttons, thus enabling the user to export the network in JSON format. The network visualization utility also has an option for saving the network in SVG format, which is another advantage, allowing the user to generate high-quality images for publications. All the result data, including HPI predictions, features, and network files, are stored for a period of 30 days on the backend server. Access to the saved results can be further extended upon the user’s request.

Figure 2. Interface of the host-pathogen interactome tool.

3. Discussion

3.1. Features Search: A Central Resource to Retrieve *P. syringae* Annotations

alfaNET implements nine search modules to represent the result data from the host–pathogen PPI prediction analysis pipeline. From these, eight modules consist of different annotation categories (*Protein Annotations, Subcellular Localization Annotations, Gene Ontology (GO) Term Annotations, Functional Domain Mappings (InterPro), Ice-nucleation proteins, Known effectors, Virulence effectors, and Predicted effectors from EffectiveDB*) of *P. syringae* proteins. Furthermore, the module ‘*Host-pathogen interactions*’ includes the alfalfa proteins involved in the predicted interactions and those related to bacterial stem blight. The eight *P. syringae* annotation modules can be accessed using the *Pseudomonas* dataset, while the results for the interactions module can be displayed by specifying the *Medicago* dataset.

The database aims to provide comprehensive information on the *Pseudomonas* proteins to users through the ‘*advanced search*’ functionality (Figure 3). The interface of this module permits the user to specify the filtering parameters by using the keywords for the required species. As an extensive module, it incorporates the functioning of other modules, and thus searches for the given keyword in the respective backend tables to pull the particular record that matches the required criteria. Various information modules such as protein annotations, subcellular localization, and other annotations are queried in order to match with the specific keyword in question. Furthermore, the user can also obtain the protein accessions using the “*basic*” search option, which is easily accessible through all the modules of alfaNET.

alfaNET is structured in such a manner that it can perform several tasks, for example: obtaining protein annotations of *P. syringae* proteins (Figure 4a); obtaining subcellular localization of the proteins (Figure 4b); and fetching functional domain mappings (InterPro) (Figure 4c) of the pathogen proteins related to bacterial stem blight. With the availability of these interfaces, we believe that this database will be a great resource for the research community.

Advanced Search Module

You may use this form to search proteins that match selected filtering criteria. Wild cards (partial words/names) are enabled for this search. This module could be use if for example you would like to retrieve certain records of a given length or that are related with some biological process.

This module is comprehensive and will look for protein annotations, Gene Ontology terms, experiment descriptions and many others that would match to the provided keyword.

Keyword * (e.g: Psyr_2612252608, Kinase) **Species**

Protein length **Genome range** **Subcellular localization**

- -

Protein length and genome range values must be numeric, if those are not numeric the values will be ignored in the search.

*Keyword is the only field that is required to run the search, you may leave the others blank if preferred. Placeholders inside input boxes are just hints for the users to help them to know what to type inside. If the user want to use an additional filter (e.g: Protein length) must type something in those boxes to run a valid search.

Figure 3. Advanced search module of alfaNET displaying the default parameters.

(a)

Protein Annotations ⓘ

Show 25 entries Search:

Accession	Locus	Description	Length
Psyr_2612250406	Ga0057439_10042	metabolite-proton symporter	858
Psyr_2612250407	Ga0057439_10043	Mg2+ and Co2+ transporter CorB, contains DUF21, CBS pair, and CorC-HlyC domains	415
Psyr_2612250408	Ga0057439_10044	ABC-type uncharacterized transport system, permease component	270
Psyr_2612250409	Ga0057439_10045	signal recognition particle subunit FFH/SRP54 (srp54)	458

(b)

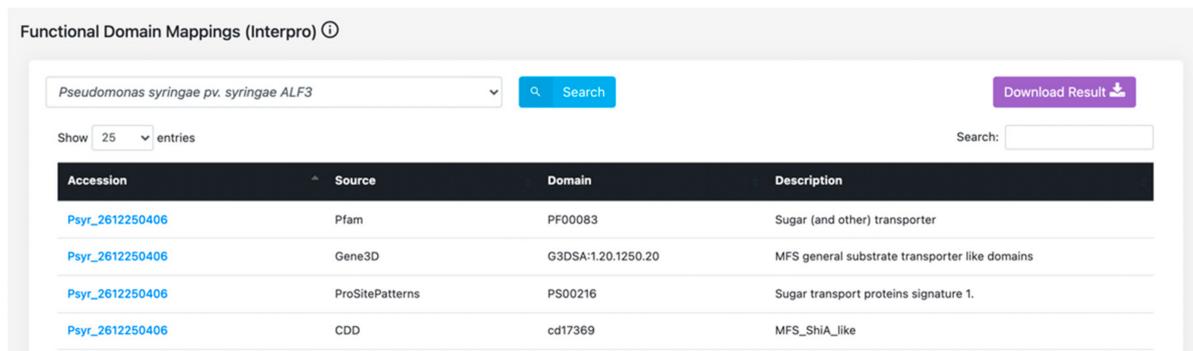
Subcellular Localization Annotations ⓘ

Show 25 entries Search:

Accession	Subcellular Localization	Score
Psyr_2612250407	Cytoplasmic Membrane	10
Psyr_2612250408	Cytoplasmic Membrane	10
Psyr_2612250409	Multi-target	-
Psyr_2612250410	Cytoplasmic	9.26

Figure 4. Cont.

(c)



Accession	Source	Domain	Description
Psyr_2612250406	Pfam	PF00083	Sugar (and other) transporter
Psyr_2612250406	Gene3D	G3DSA:1.20.1250.20	MFS general substrate transporter like domains
Psyr_2612250406	ProSitePatterns	PS00216	Sugar transport proteins signature 1.
Psyr_2612250406	CDD	cd17369	MFS_ShiA_like

Figure 4. Examples of annotation modules of alfaNET: (a) protein annotations; (b) subcellular localizations; (c) functional domain mappings.

3.2. Host-Pathogen Interactome: Towards a Better Understanding of Bacterial Stem Blight Infection Mechanisms

The study of PPIs provides an in-depth understanding of the biological functions of proteins [21]. Currently, the host–pathogen PPI prediction is primarily carried out using various experimental techniques such as yeast two-hybrid [22] and co-immunoprecipitation [23], among others. These techniques are primarily useful for the validation of few PPI pairs, but for large-scale PPI prediction such methods are expensive, time-consuming, and labor-intensive [24]. The advancements in high-throughput computational approaches have made the extensive prediction of PPIs rapid and tremendously efficient [25].

Accounting for the variations in bacterial strains, the PPI prediction for a particular pathogenic strain can contribute to the development of enhanced strain-specific treatment. Therefore, alfaNET implements an interactomics module to identify the PPIs between *P. syringae* and alfalfa that are responsible for causing bacterial blight disease. The interactomics tool allows the user to select protein–protein interaction databases, which serve as a reference for PPI prediction. The user can also define the filters for BLASTp alignments to determine the protein homologs. Furthermore, our database implements an efficient plugin, SigmaJS, for network visualization. This plugin is known for its enhanced performance and ability to illustrate large networks. The network visualization platform provides the user with detailed information about each node, such as species, degree, and description, along with the identification of hub nodes. The hub nodes are of high research interest, as these play a significant role in deciphering disease-related pathways and cellular processes [26]. Besides analyzing the network with this tool, the user can also download the resultant network files in JSON or SVG format, enabling the user to examine the networks in third-party network analyzer tools.

3.3. BLAST Server and Bulk Data Download

The BLAST search module (Figure 5a) provides homology sequence search functionality to users. In this module, we implemented the NCBI’s BLAST locally on the server with the availability of two proteomes: alfalfa and *P. syringae*. The link to the protein sequences of the host and pathogen species used in alfaNET is available on the “Datasets” page, which forwards the user to the source of the dataset. The user has the choice to query sequences against the individual proteome or both (by default) the proteomes together at the same time. The module is of great advantage to users, as either the nucleotide or amino acid sequences can be uploaded as a query and the specific BLAST program (BLASTx or BLASTp) that needs to be performed is automatically detected by the system.

(a)

BLAST search

Upload FASTA file

Drag and drop a file here or click

Or Paste your FASTA sequence (Load demo data)

Datasets

All proteomes

Advance BLAST options

Expectation (e-value) Weight matrix Number of results

Email address (Optional)

Enter email

ⓘ

(b)



Figure 5. BLAST search module: (a) BLAST search interface; (b) visualization of BLAST alignments using BlasterJS.

The “BLAST Results” page summarizes the information of BLAST alignments, which can be downloaded in various formats such as Excel or PDF, or can be copied to the clipboard. Additionally, the alignments can be visualized in an enhanced mode using the “Detailed” option (Figure 5b). This option allows the user to download the alignments in PNG or JPEG format.

3.4. Applicability of alfaNET: A Case Study on Resistant and Susceptible Plant Responses in *Medicago sativa* to Bacterial Stem Blight

In order to confirm the viability of alfaNET, we obtained the differentially expressed genes (DEGs) identified in resistant and susceptible plants of the alfalfa cultivar "Maverick", infected with *P. syringae* pv. *syringae* ALF3, causing bacterial stem blight [12]. The bioinformatic analysis in this study revealed 810 genes up-regulated in resistant plants and 577 genes down-regulated in susceptible plants 72 hours after inoculation. We identified these DEGs in our PPIs obtained using the interolog-based computational approach and found 388 down-regulated genes and 542 up-regulated genes. The identified up-regulated genes were found to encode for plant resistance (R) genes, suggesting the defense response of the resistant plants. On the other hand, the down-regulated genes were involved in various functional categories such as protein phosphorylation, response to biotic stimulus, and transmembrane transport, thus being the reason for the plant becoming susceptible. Additionally, we also identified the interacting pathogens for these differentially expressed genes. A total of 1734 *P. syringae* proteins were found interacting with 542 up-regulated alfalfa genes (67% correctly identified), involved in 173,587 interactions (Figure 6). Meanwhile, for 388 down-regulated genes (67.24% correctly identified), 1457 *P. syringae* proteins were involved in 105,897 interactions (Figure 7). The annotations of the pathogen proteins are available in different modules of alfaNET. This shows the enrichment of alfaNET, and the identified genes can serve as a valuable dataset for biologists for further validation and characterization against bacterial stem blight disease.

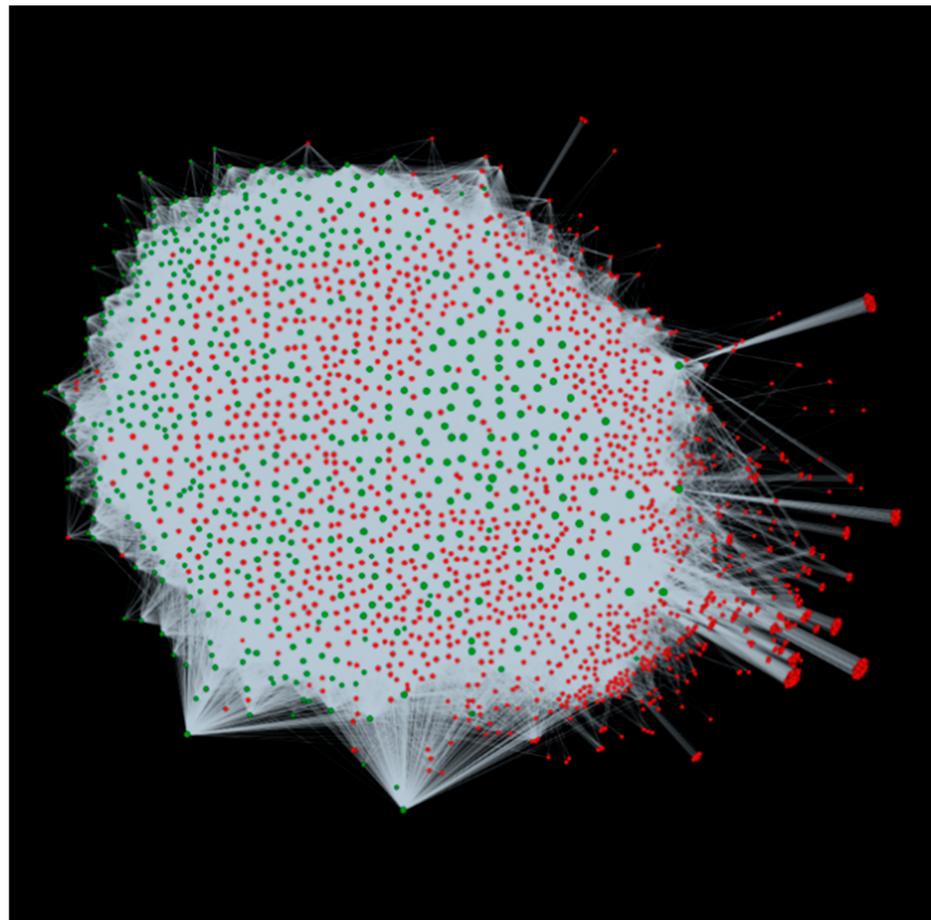


Figure 6. Visualization of the interactions for up-regulated genes of alfalfa interacting with *P. syringae*. Green nodes are host proteins and red nodes are pathogen proteins. Edges in grey represent the interactions from the interolog-based method.

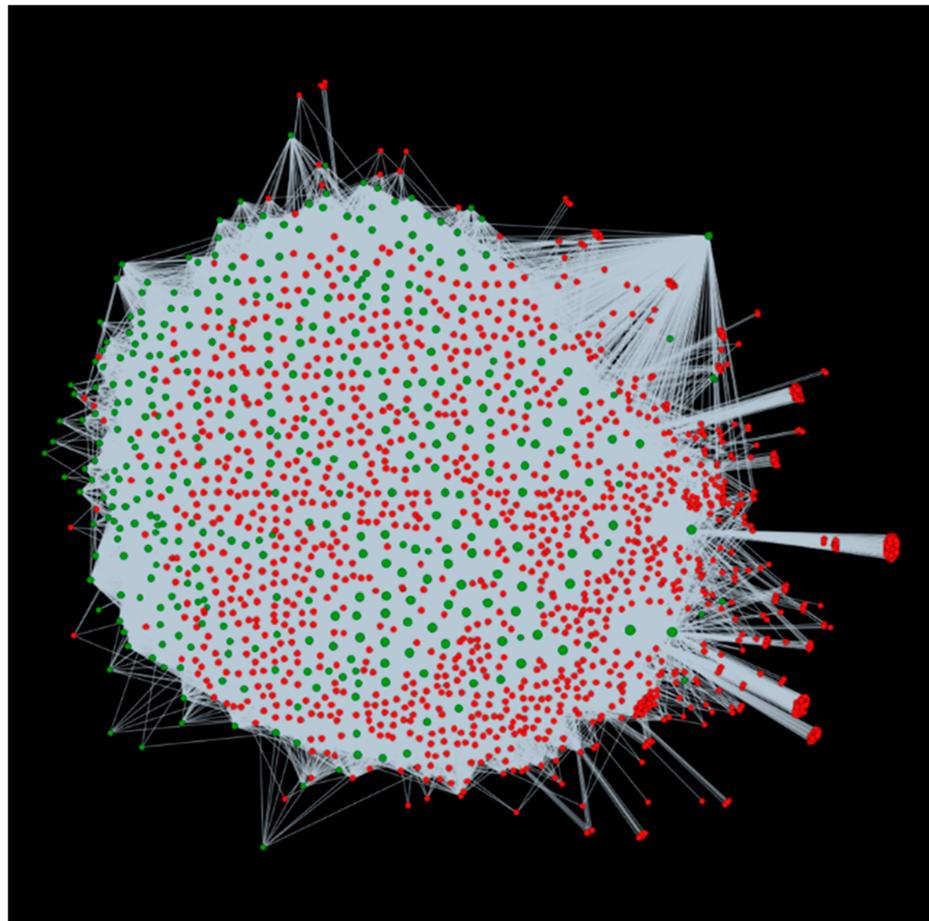


Figure 7. Visualization of the interactions for down-regulated genes of alfalfa interacting with *P. syringae*. Green nodes are host proteins and red nodes are pathogen proteins. Edges in grey represent the interactions from the interolog-based method.

To validate the interactions predicted using the case study above, the host–pathogen interaction pair “MSAD_210711-Psyr_2612254330” was further investigated. The functional analysis revealed that the host protein “MSAD_210711” is involved in the lipid catabolic process (GO:0016042), phospholipid metabolic process (GO:0006644), and glycerolipid metabolic process (GO:0046486). The enzymes involved in lipid biosynthesis are reported to play an essential role in various biotic and abiotic defense pathways. Phosphatidic acid (PA), a major component of lipid biosynthesis, is involved in abscisic acid-mediated stomatal closure during pathogen attack [27]. On the other hand, the pathogen protein “Psyr_2612254330” was found to be involved in the plant–pathogen interaction pathway (psb04626). The subcellular localization analysis showed that both the host and pathogen proteins are located in the cytoplasm, thus suggesting the site of protein–protein interaction. The above pair could be a good candidate for future experimental validations, among others.

3.5. Limitations and Future Development

The database implements multiple modules, a few of which consist of inadequate information about *P. syringae* pv. *syringae* ALF3 proteins due to the current availability of only the draft genome assembly, leading to incomplete annotation information. Predominantly, the polished genomes are used to perform molecular mechanism-related studies. Furthermore, to our knowledge, there is no previous report on the PPI prediction in the alfalfa-*P. syringae* system. In the future, we would probably enhance alfaNET by incorporating other *Pseudomonas* strains infecting alfalfa and include more experimentally

validated data. Future developments will also include the addition of domain–domain interaction databases.

4. Materials and Methods

4.1. Data Source and Processing

The proteomes of alfalfa and *Pseudomonas syringae* pv. *syringae* ALF3 were retrieved from various sources. For alfalfa, the proteome was obtained from LegumeIP (<https://plantgrn.noble.org/LegumeIP/gdp/>, accessed on 10 November 2020) [28]. In total, 87,892 protein sequences were downloaded, which were further analyzed with CD-HIT [29] at 100% identity to reduce redundancy, leading to 87,156 proteins. The proteome of *Pseudomonas syringae* pv. *syringae* ALF3 was downloaded from JGI Integrated Microbial Genomes and Microbiomes (IMG/M) (<https://img.jgi.doe.gov/>, accessed on 10 November 2020; Taxon ID: 2609460275) and UniProt proteomes (<https://www.uniprot.org/proteomes/>, accessed on 30 July 2021; UP000028706). The *Pseudomonas* protein sequences from both the sources were combined and 100% identical sequences were clustered, leading to 5527 unique protein sequences. Since the cytoplasmic proteins of pathogens are believed not to take part in host–pathogen interactions, such proteins were eliminated [30]. Subsequently, the cytoplasmic proteins were analyzed at EffectiveDB (<https://effectors.csb.univie.ac.at/>, accessed on 16 December 2020) to predict the secreted proteins, and those predicted as secreted were included in the analysis. This approach gave 4427 *P. syringae* proteins for further analysis.

To employ the homology-based interolog approach, various PPI databases of The International Molecular Exchange Consortium (IMEX) [31] were used as templates. alfaNET implements five major databases from IMEX, viz., DIP 2020 [32], IntAct v4.2.16 [33], MINT 2018 [34], BioGRID v4.2.191 [35], and HPIDB v3.0 [36]. We also implemented all the plant species from the STRING v11.0 database [37]. In addition, the ArabHPI database was also included in alfaNET. The above-mentioned databases were downloaded from the respective sources and implemented as an individual MySQL table locally.

4.2. Host-Pathogen Interactome Comparison Tool

A novel module, known as an “interactomics” tool, was developed to compare PPIs between host and pathogen. This module implements the homology-based interolog approach at the backend. The interolog approach relies on the concept of transferring the conserved interactions between the species based on the sequence homology [38]. For example, if two proteins A and B interacting in an organism contain their orthologs in A' and B' in other organisms, then A-B and A'-B' are considered to be an interolog [39]. Furthermore, the ortholog proteins obtained from the above-mentioned approach are queried against the PPI databases. If the host and pathogen proteins match with the corresponding interactions in the PPI databases, then the particular protein pair is anticipated to be interacting.

In alfaNET, we employed this method at the backend to predict the possible host–pathogen and protein–protein interactions between alfalfa and *P. syringae* proteins. In actuality, we aligned the proteomes of alfalfa and *Pseudomonas* individually against the six PPI databases using the BLASTp v2.7.1 tool with default parameters, thus obtaining six alignment files for each of the proteomes. Following the alignments, the interolog prediction is performed using our in-house R and python scripts, which call on various SQL functions to extract the data from the database tables. All these steps are executed after the submission of the job by a user. The process of ortholog match querying was accelerated by indexing the columns of the SQL tables for the PPI databases as `interactor_A` and `interactor_B`.

4.3. Protein Annotation

For *Pseudomonas* proteins, the detailed description was obtained from the *P. syringae* pv. *syringae* ALF3 annotations available at JGI IMG/M. The functional annotation, including the prediction of conserved domains and important sites, was carried out by analyzing

the *P. syringae* proteome in InterProScan [40] using the “iprlookup” and “goterms” parameters. The subcellular localization of the *Pseudomonas* proteins was obtained using PSORTb [41]. Furthermore, the effectors for bacterial proteins were predicted using EffectiveDB, a database for retrieving bacterial secreted proteins [42].

4.4. Dataset Collection for *P. syringae* Effectors

The plant pathogenic bacteria subvert the host cell machinery and suppress the immune responses by deploying effector proteins into the plant cell using the type III secretion system (T3SS) [43]. To enhance the pathogen protein annotations, we extracted the known, potential, and virulence proteins for *P. syringae* from three different sources mentioned below. In total, 655 unique effectors (known T3SS effectors, virulence effectors, and predicted effectors) were procured and implemented in the database.

4.4.1. Known T3SS Effectors

There are 50 known T3SS effector proteins reported in the literature, which were obtained from *Pseudomonas syringae* Genome Resources (<http://www.pseudomonas-syringae.org/>, accessed on 18 December 2020).

4.4.2. Virulence Effectors

We obtained 8 virulence factors for *P. syringae* by searching its orthologs against *Pseudomonas aeruginosa* PAO1 (reference) from Pseudomonas genome DB (<https://www.pseudomonas.com/>, accessed on 5 January 2021).

4.4.3. Predicted Effectors

To determine the type III secreted proteins of *P. syringae*, we extracted the pre-calculated potential effector proteins from EffectiveDB (<https://effectors.csb.univie.ac.at/>, accessed on 5 January 2021), using *P. syringae* pv. *syringae* B728a as a reference. A total of 641 proteins were obtained, which were classified as secreted proteins on the basis of their signal peptide, secretion chaperon binding site, or eukaryotic-like domains.

4.5. Dataset Collection for Ice-Nucleation Proteins

As mentioned earlier, INPs serve as initiators of bacterial stem blight. In line with this, we obtained the orthologs of the gene *Psyr1608*, which belongs to *P. syringae* pv. *syringae* B728a [44]. This gene is localized on the outer membrane of the bacteria and encompasses the ice-nucleation proteins octamer repeat. A total of 30 orthologs *Psyr1608* were procured from the Pseudomonas genome DB (<https://pseudomonas.com/>, accessed on 24 February 2021).

5. Conclusions

We developed alfaNET, an extensive framework that provides various annotation functionalities for *P. syringae* proteins. The database provides a platform namely, ‘interactomics’ tool—which predicts the protein–protein interactions of the alfalfa-*P. syringae* system by implementing an interolog-based computational approach. This tool also presents an enhanced network visualization, thus providing in-depth information on host–pathogen interactions. We believe that alfaNET will be a useful resource for the research community as well as plant pathologists/breeders, and that it will provide experimental biologists with a better understanding of the infection mechanisms of the disease.

Author Contributions: R.K. (Rakesh Kaundal) formulated and designed the research. R.K. (Raghav Kataria) analyzed the data, developed prediction models, and designed and constructed the web database. Writing—original draft preparation, R.K. (Raghav Kataria); writing—review and editing, R.K. (Rakesh Kaundal); visualization, R.K. (Raghav Kataria) and R.K. (Rakesh Kaundal); supervision, R.K. (Rakesh Kaundal); project administration, R.K. (Rakesh Kaundal); funding acquisition, R.K. (Rakesh Kaundal). Both authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Utah Agricultural Experiment Station (UAES), Utah State University, and approved as journal paper number 9503. The authors also acknowledge the support to this study from UAES seed grant # A48025 to R. Kaundal. The funding body did not play any role in the design of this study; the collection, analysis, or interpretation of data; or in the writing of this manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are thankful to all the colleagues in KAABiL lab for their beneficial inputs while creating the database, alfaNET. Special thanks to Cristian D. Loaiza for devoting his valuable time to configuring and fixing the backend scripts required for the working of the database. The authors also sincerely thank Shelby McCowan (Bioinformatics System Administrator), who helped with the designing of the database and implementing the database successfully on an HPC cluster.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fernandez, A.; Sheaffer, C.; Tautges, N.; Putnam, D.; Hunter, M. *Alfalfa, Wildlife & the Environment*, 2nd ed.; National Alfalfa and Forage Alliance: St. Paul, MN, USA, 2019; Volume 34.
2. Rafińska, K.; Pomastowski, P.; Wrona, O.; Górecki, R.; Buszewski, B. *Medicago sativa* as a source of secondary metabolites for agriculture and pharmaceutical industry. *Phytochem. Lett.* **2017**, *20*, 520–539. [[CrossRef](#)]
3. Nicholson, R.L.; Hammerschmidt, R. Phenolic Compounds and Their Role in Disease Resistance. *Annu. Rev. Phytopathol.* **1992**, *30*, 369–389. [[CrossRef](#)]
4. Harighi, B. Occurrence of alfalfa bacterial stem blight disease in Kurdistan province, Iran. *J. Phytopathol.* **2007**, *155*, 593–595. [[CrossRef](#)]
5. Proud Sponsor of Midwest Forage Association. *Forage Focus*; 2017; Technical Report.
6. Saint-Vincent, P.M.B.; Ridout, M.; Engle, N.L.; Lawrence, T.J.; Yeary, M.L.; Tschaplinski, T.J.; Newcombe, G.; Pelletier, D.A. Isolation, characterization, and pathogenicity of two *Pseudomonas syringae* pathovars from populus trichocarpa seeds. *Microorganisms* **2020**, *8*, 1137. [[CrossRef](#)]
7. Li, Q.; Yan, Q.; Chen, J.; He, Y.; Wang, J.; Zhang, H.; Yu, Z.; Li, L. Molecular characterization of an ice nucleation protein variant (InaQ) from *Pseudomonas syringae* and the analysis of its transmembrane transport activity in *Escherichia coli*. *Int. J. Biol. Sci.* **2012**, *8*, 1097–1108. [[CrossRef](#)]
8. Harrison, J.; Dornbusch, M.R.; Samac, D.; Studholme, D.J. Draft genome sequence of *Pseudomonas syringae* pv. *syringae* ALF3 isolated from alfalfa. *Genome Announc.* **2016**, *4*, 2015–2016. [[CrossRef](#)]
9. Quirino, B.F.; Bent, A.F. Deciphering host resistance and pathogen virulence: The *Arabidopsis/Pseudomonas* interaction as a model. *Mol. Plant Pathol.* **2003**, *4*, 517–530. [[CrossRef](#)] [[PubMed](#)]
10. Alfano, J.R.; Collmer, A. Bacterial pathogens in plants: Life up against the wall. *Plant Cell* **1996**, *8*, 1683–1698. [[CrossRef](#)]
11. Samac, D.; Studholme, D.J.; Ao, S. Characterization of the bacterial stem blight pathogen of alfalfa, *Pseudomonas syringae* pv. *syringae* ALF3. *Phytopathology* **2014**, *104*, 102.
12. Nemchinov, L.G.; Shao, J.; Lee, M.N.; Postnikova, O.A.; Samac, D.A. Resistant and susceptible responses in alfalfa (*Medicago sativa*) to bacterial stem blight caused by *Pseudomonas syringae* pv. *syringae*. *PLoS ONE* **2017**, *12*, e0189781. [[CrossRef](#)] [[PubMed](#)]
13. Loaiza, C.D.; Duhan, N.; Lister, M.; Kaundal, R. In silico prediction of host–pathogen protein interactions in melioidosis pathogen *Burkholderia pseudomallei* and human reveals novel virulence factors and their targets. *Brief. Bioinform.* **2020**, *22*, bbz162. [[CrossRef](#)] [[PubMed](#)]
14. Wattam, A.R.; Abraham, D.; Dalay, O.; Disz, T.L.; Driscoll, T.; Gabbard, J.L.; Gillespie, J.J.; Gough, R.; Hix, D.; Kenyon, R.; et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **2014**, *42*, 581–591. [[CrossRef](#)]
15. Memišević, V.; Kumar, K.; Zavaljevski, N.; DeShazer, D.; Wallqvist, A.; Reifman, J. DBSecSys 2.0: A database of *Burkholderia mallei* and *Burkholderia pseudomallei* secretion systems. *BMC Bioinform.* **2016**, *17*, 1–7. [[CrossRef](#)] [[PubMed](#)]
16. Yue, J.; Zhang, D.; Ban, R.; Ma, X.; Chen, D.; Li, G.; Liu, J.; Wisniewski, M.; Droby, S.; Liu, Y. PCPPI: A comprehensive database for the prediction of Penicillium-crop protein-protein interactions. *Database* **2017**, *2017*, 1–9. [[CrossRef](#)] [[PubMed](#)]
17. Winsor, G.L.; Lam, D.K.W.; Fleming, L.; Lo, R.; Whiteside, M.D.; Yu, N.Y.; Hancock, R.E.W.; Brinkman, F.S.L. *Pseudomonas* Genome Database: Improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res.* **2011**, *39*, 596–600. [[CrossRef](#)]
18. Li, J.; Dai, X.; Liu, T.; Zhao, P.X. LegumeIP: An integrative database for comparative genomics and transcriptomics of model legumes. *Nucleic Acids Res.* **2012**, *40*, 1221–1229. [[CrossRef](#)]

19. Blanco-Míguez, A.; Fdez-Riverola, F.; Sánchez, B.; Lourenço, A. BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results. *PLoS ONE* **2018**, *13*, e0205286. [[CrossRef](#)]
20. Jacomy, A. Sigmajs. 2012. Available online: <http://sigmaj.s.org/> (accessed on 1 August 2021).
21. Wei, L.; Xing, P.; Zeng, J.; Chen, J.X.; Su, R.; Guo, F. Improved prediction of protein–protein interactions using novel negative samples, features, and an ensemble classifier. *Artif. Intell. Med.* **2017**, *83*, 67–74. [[CrossRef](#)]
22. He, Y.; Wu, L.; Liu, X.; Zhang, X.; Jiang, P.; Ma, H. Yeast two-hybrid screening for proteins that interact with PFT in wheat. *Sci. Rep.* **2019**, *9*, 1–10. [[CrossRef](#)]
23. Free, R.B.; Hazelwood, L.A.; Sibley, D.R. Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectroscopy. *Curr. Protoc. Neurosci.* **2009**, *46*, 5–28. [[CrossRef](#)]
24. Ding, Z.; Kihara, D. Computational identification of protein-protein interactions in model plant proteomes. *Sci. Rep.* **2019**, *9*, 1–13. [[CrossRef](#)]
25. Yang, L.; Han, Y.; Zhang, H.; Li, W.; Dai, Y. Prediction of Protein-Protein Interactions with Local Weight-Sharing Mechanism in Deep Learning. *Biomed Res. Int.* **2020**, *2020*, 5072520. [[CrossRef](#)]
26. Saha, S.; Sengupta, K.; Chatterjee, P.; Basu, S.; Nasipuri, M. Analysis of protein targets in pathogen-host interaction in infectious diseases: A case study on *Plasmodium falciparum* and *Homo sapiens* interaction network. *Brief. Funct. Genomic.* **2018**, *17*, 441–450. [[CrossRef](#)] [[PubMed](#)]
27. Lim, G.H.; Singhal, R.; Kachroo, A.; Kachroo, P. Fatty Acid- and Lipid-Mediated Signaling in Plant Defense. *Annu. Rev. Phytopathol.* **2017**, *55*, 505–536. [[CrossRef](#)]
28. Dai, X.; Zhuang, Z.; Boschiero, C.; Dong, Y.; Zhao, P.X. LegumeIP V3: From models to crops—An integrative gene discovery platform for translational genomics in legumes. *Nucleic Acids Res.* **2020**, *49*, D1472–D1479. [[CrossRef](#)]
29. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)] [[PubMed](#)]
30. Sahu, S.S.; Weirick, T.; Kaundal, R. Predicting genome-scale Arabidopsis-*Pseudomonas syringae* interactome using domain and interolog-based approaches. *BMC Bioinform.* **2014**, *15*, S13. [[CrossRef](#)] [[PubMed](#)]
31. Orchard, S.; Kerrien, S.; Abbani, S.; Aranda, B.; Bhate, J.; Bridge, A.; Briganti, L.; Brinkman, F.S.L.; Cesareni, G.; Chatr-aryamontri, A.; et al. Exchange Consortium (IMEx). *Nat. Methods* **2012**, *9*, 345–350. [[CrossRef](#)] [[PubMed](#)]
32. Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32*, 449–451. [[CrossRef](#)]
33. Kerrien, S.; Aranda, B.; Breuza, L.; Bridge, A.; Broackes-Carter, F.; Chen, C.; Duesbury, M.; Dumousseau, M.; Feuermann, M.; Hinz, U.; et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **2012**, *40*, 841–846. [[CrossRef](#)]
34. Licata, L.; Briganti, L.; Peluso, D.; Perfetto, L.; Iannuccelli, M.; Galeota, E.; Sacco, F.; Palma, A.; Nardozza, A.P.; Santonico, E.; et al. MINT, the molecular interaction database: 2012 Update. *Nucleic Acids Res.* **2012**, *40*, 857–861. [[CrossRef](#)] [[PubMed](#)]
35. Chatr-Aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N.K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D369–D379. [[CrossRef](#)] [[PubMed](#)]
36. Ammari, M.G.; Gresham, C.R.; McCarthy, F.M.; Nanduri, B. HPIDB 2.0: A curated database for host-pathogen interactions. *Database* **2016**, *2016*, 1–9. [[CrossRef](#)] [[PubMed](#)]
37. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)]
38. Castillo-Lara, S.; Abril, J.F. PlanNET: Homology-based predicted interactome for multiple planarian transcriptomes. *Bioinformatics* **2018**, *34*, 1016–1023. [[CrossRef](#)]
39. Yu, H.; Luscombe, N.M.; Lu, H.X.; Zhu, X.; Xia, Y.; Han, J.D.J.; Bertin, N.; Chung, S.; Vidal, M.; Gerstein, M. Annotation transfer between genomes: Protein-protein interologs and protein-DNA regulogs. *Genome Res.* **2004**, *14*, 1107–1118. [[CrossRef](#)]
40. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
41. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Cenk Sahinalp, S.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. [[CrossRef](#)]
42. Jehl, M.A.; Arnold, R.; Rattei, T. Effective-A database of predicted secreted bacterial proteins. *Nucleic Acids Res.* **2011**, *39*, 591–595. [[CrossRef](#)]
43. Shames, S.R.; Finlay, B.B. Bacterial effector interplay: A new way to view effector function. *Trends Microbiol.* **2012**, *20*, 214–219. [[CrossRef](#)]
44. Feil, H.; Feil, W.S.; Chain, P.; Larimer, F.; DiBartolo, G.; Copeland, A.; Lykidis, A.; Trong, S.; Nolan, M.; Goltsman, E.; et al. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11064–11069. [[CrossRef](#)] [[PubMed](#)]