

RESEARCH ARTICLE

Predicting Ecological Roles in the Rhizosphere Using Metabolome and Transportome Modeling

Peter E. Larsen^{1,2*}, Frank R. Collart¹, Yang Dai²

1 Argonne National Laboratory, Biosciences Division, Argonne, IL, United States of America, **2** University of Illinois at Chicago, Department of Bioengineering, Chicago, IL, United States of America

* plarsen@anl.gov



OPEN ACCESS

Citation: Larsen PE, Collart FR, Dai Y (2015) Predicting Ecological Roles in the Rhizosphere Using Metabolome and Transportome Modeling. PLoS ONE 10(9): e0132837. doi:10.1371/journal.pone.0132837

Editor: Jeffrey L Blanchard, University of Massachusetts, UNITED STATES

Received: March 18, 2015

Accepted: June 18, 2015

Published: September 2, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This contribution originates in part from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"), and the U.S. Department of Energy, Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program. This research has been funded by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357. The funders had no

Abstract

The ability to obtain complete genome sequences from bacteria in environmental samples, such as soil samples from the rhizosphere, has highlighted the microbial diversity and complexity of environmental communities. However, new algorithms to analyze genome sequence information in the context of community structure are needed to enhance our understanding of the specific ecological roles of these organisms in soil environments. We present a machine learning approach using sequenced *Pseudomonad* genomes coupled with outputs of metabolic and transportomic computational models for identifying the most predictive molecular mechanisms indicative of a *Pseudomonad*'s ecological role in the rhizosphere: a biofilm, biocontrol agent, promoter of plant growth, or plant pathogen. Computational predictions of ecological niche were highly accurate overall with models trained on transportomic model output being the most accurate (Leave One Out Validation F-scores between 0.82 and 0.89). The strongest predictive molecular mechanism features for rhizosphere ecological niche overlap with many previously reported analyses of *Pseudomonad* interactions in the rhizosphere, suggesting that this approach successfully informs a system-scale level understanding of how *Pseudomonads* sense and interact with their environments. The observation that an organism's transportome is highly predictive of its ecological niche is a novel discovery and may have implications in our understanding microbial ecology. The framework developed here can be generalized to the analysis of any bacteria across a wide range of environments and ecological niches making this approach a powerful tool for providing insights into functional predictions from bacterial genomic data.

Introduction

Terrestrial plants rarely exist simply as solitary organisms. Rather they encompass complex interacting communities of soil fungi, subsurface bacteria, and animals whose combined functions are crucial to above and below ground plant biomass [1, 2]. Some of these subsurface communities are correspondingly dependent upon their plant host. Around 20% to 40% of

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

photosynthetically-derived sugars from plants are consumed directly by the subsurface community [3, 4], making the root-associated ecosystem an important part of the terrestrial carbon cycle. These communities reside and interact in the narrow region in the soil directly influenced by plant root exudates called the rhizosphere, and within the rhizosphere, soil bacteria fill multiple ecological niches. A niche is defined here the set of specific roles by which organisms interact with the abiotic environment, the plant roots, and other microbes as they compete for available nutrients.

Efforts to understand the compositions and interrelationships of bacteria in the rhizosphere community [1, 5–7] are limited by the inability to culture these organisms in the laboratory [8]. As such, much of the information about these organisms can only be learned indirectly (i.e. inferred from genomic sequences assembled from metagenomic data sets). In this context, ecological functions of uncharacterized but genomically sequenced bacteria are frequently inferred from genomic sequence homology to characterized species [9, 10]. This approach is not without limitations, however, since high levels of sequence homology between bacteria can occur between organisms of different ecological functions [9, 11–13]. Other approaches focus on the presence of key genes linked to ecological function, such as those for the fixation of nitrogen [14–16] or for injecting toxins into a host cell [17–20]. While a few important ecological characteristics can be inferred in this fashion, many ecological functions cannot be reliably linked to small sets of specific genes. More sophisticated approaches use computational approaches such as flux balance analysis (FBA) modeling of predicted bacterial metabolomes to infer a bacterium's ecological niche [21, 22]. Although, FBA can infer a bacterium's nutritional requirements, the approach is often not predictive for a bacterium's role in its community.

To circumvent these limitations, we evaluated the utility of machine learning computational tools to infer a bacterium's ecological role from genomic data. Support Vector Machine (SVM) models were used to predict rhizosphere ecological niches using outputs from system-scale computational models for genomic, metabolomic, and transportomic features. Given a set of training examples, each marked as belonging to one of two categories, such as membership to an ecological niche, an SVM training algorithm builds a model that assigns new examples into one category or the other. SVMs are particularly powerful in their ability to avoid over-fitting. To evaluate the utility of this computational framework, we selected *Pseudomonads*, a genus of bacteria commonly found in the rhizosphere community and of particular interest to terrestrial carbon cycling. *Pseudomonads* are widely distributed and sequence data from representative organisms indicate their genomes encode a diverse spectrum transporters, enzymes and secondary metabolic activities [23]. This functional and metabolic diversity makes these bacteria highly relevant to computational modeling of metabolic and transportomic capacities.

Pseudomonads occupy a wide variety of habitats including soil and marine environments and can be plant or animal pathogens [24, 25]. For the present analysis, the selections of environmental niche labels are based upon those previously reported by Silby et al [25] as well as other investigators (Identified in Table 1). Classes of ecological niches are non-exclusionary and a single *Pseudomonad* species may be associated with any number of niches. For the present analysis, we considered four ecological niches from the rhizosphere: biocontrol, biofilm formation, plant growth promotion, and plant pathogen. Biocontrol is an ecological niche associated with *Pseudomonads* in the rhizosphere in which the bacteria protect the plant's roots from detrimental fungi, bacteria, or other pathogens [26–28]. Biofilm formation is the ability to form biofilms in any environment, but can include soils or bacteria which reside internal to the host organism [29–31]. Plant pathogenicity is the ability to cause disease in plant roots or leaves [32, 33]. Plant growth promotion is the ability to form beneficial relationships with plant roots that result in increased plant biomass [34]. For associating these

Table 1. Assigned Ecological Niche Classifications of Pseudomonad Species.

Species	# Genomes	Biocontrol	Biofilm	Plant Pathogen	Plant Growth	Reference
Aeruginosa	9	N	Y	Y	Y	(Silby, Winstanley et al. 2011)
Brassicacearum	1	Y	N	N	N	(Ortet, Barakat et al. 2011)
Denitrificans	1	Y	N	N	N	(Ainala, Somasundar et al. 2013)
Entomophila	1	Y	N	N	N	(Vodovar, Vallenet et al. 2006)
Flourescens	4	Y	Y	Y	Y	(Silby, Winstanley et al. 2011)
Fulva	1	N	N	N	N	(Renault, Deniel et al. 2007)
Mendocina	2	Y	Y	N	N	(Silby, Winstanley et al. 2011)
ND	1	N	N	Y	N	(Li, Zhao et al. 2013)
Poae	1	Y	Y	N	Y	(Muller, Zachow et al. 2013)
Protogens	2	Y	Y	Y	Y	(Jousset, Schuldes et al. 2014)
Putida	11	Y	Y	N	Y	(Silby, Winstanley et al. 2011)
Stutzeri	6	N	N	N	Y	(Silby, Winstanley et al. 2011)
Syringae	3	Y	Y	Y	N	(Silby, Winstanley et al. 2011)

doi:10.1371/journal.pone.0132837.t001

ecological niches with molecular mechanisms, four modeling output types were considered: enzyme function profiles, metabolic models, secondary metabolism models, and transporter profile (transportomic) models. Enzyme function profiles are generated using the number and distribution of the genome encoded set of specific enzyme activities as identified by assignment of an Enzyme Commission (EC) annotation number. Similarly, transportomic models are generated using the transporter functions encoded in a set of bacterial genomes. Enzyme function abundances and a set of all possible metabolic transformations performed by those functions were used to derive the metabolic models predicting the relative rates of metabolic turnover for specific metabolites, and the transportomic models, predicting the relative capacity of bacteria to transport specific ligands across cell membranes. Secondary metabolism models were derived using subset of enzymes involved in the generation of secondary metabolites (organic compounds that are not directly involved in the normal growth, development, or reproduction of an organism).

Materials and Methods

Pseudomonad Genomes

There were 43 fully sequenced and annotated Pseudomonad strains available from the NCBI (<ftp://ftp.ncbi.nih.gov/genomes/>) that are confidently associated with specific rhizosphere ecological niche classes at the time this analysis originated. The files for predicted protein sequences (.faa files in NCBI genomic sequence database) were used for all analysis strains. Pseudomonad rhizosphere ecological niche was defined as a function of Pseudomonad species and assigned based on published manuscripts (Table 1). The complete list of strains and accompanying references is available in [S1 Data](#).

Metabolomic and Transportomic Modeling

SVMs were trained on the outputs of four different computational models which are generated using annotation data derived from sequenced and annotated genomic information: enzyme function profiles, metabolomic, secondary metabolism, and transportomic. The generation of each type is described below and the relationships between data types are pictured in [Fig 1](#).

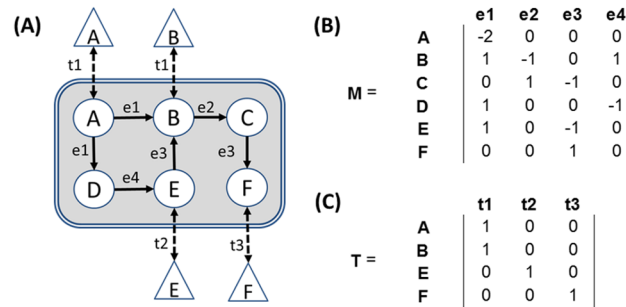


Fig 1. Metabolomic and Transportomic Modeling. In (A), a simplified metabolomic/transportomic network is featured. Triangles are extracellular compounds, circles are intracellular compounds, and double-line is cellular membrane. Dashed edges t1-t3 are transmembrane transport interactions. Solid arrows e1-e4 are directed metabolic transformations. Each enzyme or transporter annotation can be associated with one or more compounds. (B-C) represents the network in (A) transformed in matrices for use in PRMT and PRTT calculations. In (B), matrix does not consider enzymatic flux or mass balanced reactions. In (C), the transportome matrix is constructed such that a '0' indicates that a ligand is not transported and a '1' indicates that a ligand is transported by a transporter of a given annotation. For example, in the cartoon above, ligands A and B are transported by a transporter annotated with function 't1'.

doi:10.1371/journal.pone.0132837.g001

Enzyme Function Profiles

All *Pseudomonad* predicted gene models from the published genomic sequence were re-annotated for protein functions. This approach ensured that all functional assignments for the predicted proteins from genomic sequence data use uniform annotation criteria and a consistent ontology for enzyme functions and ligands.

The database of the Kyoto Encyclopedia of Genes and Genomes (KEGG) was used as the source of annotated protein sequences of metabolic enzymes and transmembrane transporter activities [35, 36]. For enzyme function annotations, Enzyme Commission (EC) annotation numbers [37] were used. A database of bacterial enzymes annotated with EC numbers and associated with specific reactions in KEGG metabolic pathways (downloaded May 16, 2011) was used for this analysis. The set of 754,066 protein sequences is annotated with 2,605 unique EC number enzyme function descriptions and the complete collection of annotated enzymes is available, in FASTA-format in [S2 Data](#). For transmembrane transporter function KEGG Orthology (KO) annotations were used [38]. The complete list of transmembrane transporter KO annotations used can be found in [S3 Data](#). There are 164,321 protein sequences, annotated with 891 unique transporter/sensor functions, and are associated with the transport of 272 unique ligands in the set of annotated transmembrane transporters and the complete FASTA-formatted set of annotated transporter proteins is available in [S4 Data](#). It is possible for a single protein sequence to be present in both the set of enzymes and the set of transmembrane transporters. Protein annotations were assigned to single best BLAST-P hit with e -values $< 1 \times 10^{-10}$ (NCBI-Blast 2.2.23+). Enzyme function profiles for *Pseudomonads* were generated as lists of all possible enzyme or transmembrane transporter annotations and the number of genes in each *Pseudomonad* for the assigned function.

Metabolomic and Secondary Metabolism Models

Predicted Relative Metabolic Turnover (PRMT) uses enzyme function profiles for quantifying the relative metabolic turnover between two metabolomes and has been described in detail elsewhere [39]. The necessary tools for performing PRMT, instructions, and demonstration data can be downloaded from www.bio.anl.gov/PRMT.html and the computational approach is briefly summarized below.

Required input for PRMT is a set of relative unique enzyme function abundances and a set of all possible metabolic transformations performed by those functions (Fig 1A). EC annotations and KEGG metabolic pathways are used for this purpose. Enzyme function abundances are provided as vectors of the log₂-transformed number of enzyme function representation in genomes of length *ec*, where *ec* is the number of enzyme function annotations in the metabolic model. The network of possible enzyme-mediated metabolic transformations is provided by a matrix *M* of size *m* by *ec*, where *m* is the total number of metabolites present in the metabolic network (Fig 1B). This matrix is the Enzyme Interaction Network (EIN) described in [39] and is generated using the PRMT script “GenerateEIN_fromECList.pl” (www.bio.anl.gov/PRMT.html).

The PRMT vector between metabolomes encoded by genomes *x*, and *y* is given by:

$$\overrightarrow{PRMT} = M(\vec{e}^x - \vec{e}^y) \quad (1)$$

The resulting vector of PRMT-scores of length *m* contains the comparison of predicted relative metabolic turnover of each metabolite in *M* for metabolome encoded by genome *x* relative to genome *y*. A positive PRMT score indicates an increased relative capacity for the synthesis of a compound in the metabolome encoded by genome *x* relative to genome *y*. A negative PRMT score indicates an increased relative capacity for the consumption of a compound in the metabolome encoded by genome *x* relative to genome *y*. PRMT scores do not indicate rates of reaction or predict quantities or concentrations of compounds in a metabolome. PRMT scores are generated using the script “CalculatePRMT_AllCols.pl” (www.bio.anl.gov/PRMT.html).

Two sets of PRMT models were generated. The first used the complete set of enzyme functions identified in the set of the 43 Pseudomonad genomes to generate the metabolomic models. For generation of secondary metabolism models, the set was restricted to the subset of enzyme activities that is present in the KEGG Biosynthesis of Secondary Metabolites pathway (KEGG map 01110). Both sets were calculated using the average enzyme function count across all Pseudomonads. In this analysis, the reference genome *y* is always calculated as the average unique enzyme function counts of all Pseudomonad genomes, as has been similarly done for normalization in previous applications of PRMT (e.g. [39–42]). Average unique enzyme function counts are calculated:

$$EFC_{AVE}^x = \frac{\sum_{t=1}^{\max t} EFC_t^x}{T}, \quad (2)$$

where EFC_{AVE}^x is the average enzyme function count for enzyme activity *x*, and EFC_t^x is the enzyme function count for activity *x* in taxa *t* of a total of *T* taxa.

Transportomic Models

Predicted Relative Transmembrane Transport (PRTT) is a system-scale metric that quantifies relative ability of organism to transport specific metabolites across the cellular membrane and is introduced here for the first time. PRTT-scores are calculated as a special case of PRMT-scores, using the same tools as PRMT, but using the pre-calculated matrix of transporter annotations and transported ligands as the EIN matrix. The transportomic matrix is available as [S5 Data](#) and for download from the PRMT website at www.bio.anl.gov/PRMT.html.

Required input for PRTT is the set of transporter function abundances and a matrix of transporter annotations and transported ligands. A selected subset of KO annotations was used for transporter annotations. Log₂-transformed representations of transmembrane transport function annotations in genomes are provided as vectors of length *ko*, where *ko* is the number transporter function annotations in the transportomic model. Also required is a transporter

ligand specificity matrix T of size l by ko , where l is the total number of ligands present in the transporter ligand specificity matrix (Fig 1C).

The PRMT score vector between transportomes encoded by genomes x , and y is given by:

$$\overrightarrow{PRTT} = T(\overrightarrow{k^x} - \overrightarrow{k^y}) \tag{3}$$

The resulting vector of PRTT-scores is of length l for the comparison of predicted relative transmembrane transport of each ligand in T for transportome encoded by genome x relative to genome y . A positive PRTT score indicates an increased relative capacity for transmembrane transport of a specific ligand in the transportome in genome x relative to genome y . A negative PRTT score indicates a decreased relative capacity of transmembrane transport of a ligand. PRTT scores do not indicate absolute rates or directionality of transmembrane transport activity. As with PRMT scores, all PRTT scores were calculated using reference genome y calculated as the average transmembrane transport function counts for all Pseudomonad genomes.

SVMs and Training Procedure

SVMs to predict Pseudomonad ecological niche were trained using subsets of calculated enzyme profiles, metabolic and secondary metabolomic model outputs, and transportomic model outputs. Enzyme function profiles (S6 Data), PRMT scores (S7 Data), secondary metabolism PRMT scores (S8 Data), or PRTT (S9 Data) scores used as features in training SVMs were non-zero in more than half of the genomes and had a standard deviation greater than 0.2 indicating features were present in most Pseudomonas genomes and there is variation in feature values.

SVMs were generated using a One Versus Rest (OVR) strategy, implemented as a set of four independent binary classifiers, and validated using a Leave One Out Validation (LOOV) scheme (Fig 2). In the OVR SVM binary classification approach, separate SVMs were generated for each ecological niche class (Biocontrol, Biofilm, Plant Pathogen, and Plant Growth Promotor), that is, Biocontrol vs non-Biocontrol, Biofilm vs. non-Biofilm, Plant Pathogen vs.

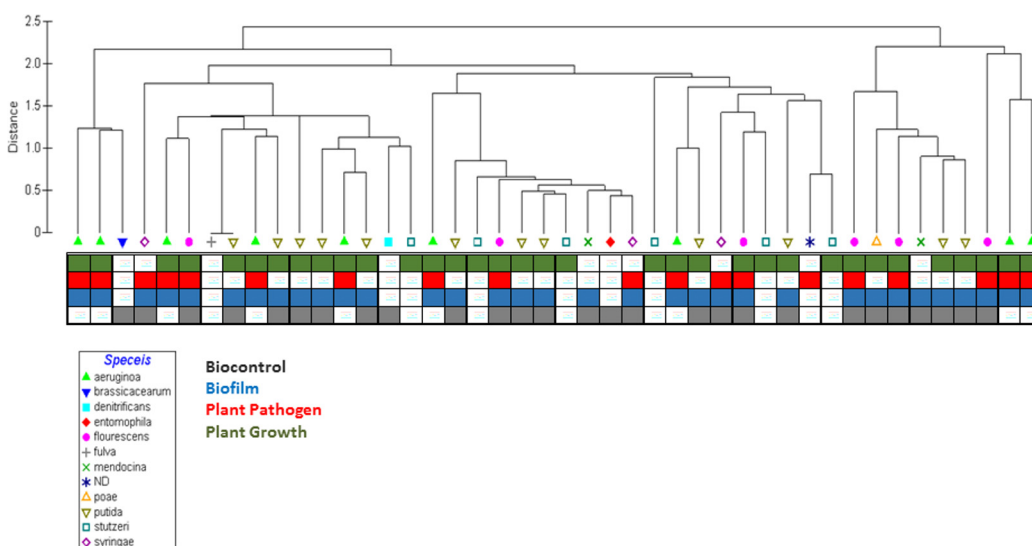


Fig 2. Clustering Pseudomonad genomes using enzyme function counts. The “Primer 6” core package and enzyme function profile data were used to generate hierarchical clusters. No obvious pattern by species or by ecological function is apparent using only enzyme function count and hierarchical clustering. Suggesting additional data and/or alternate methods are required to deduce Pseudomonad environmental niche using sequenced and annotated genomes.

doi:10.1371/journal.pone.0132837.g002

non-Plant Pathogen, and Plant Growth Promoter vs. non-Plant Growth Promoter. A LOOV scheme is a special case of a K -fold cross validation. It is most appropriate for the data in this study as the number of *Pseudomonas* is small relative to the number of possible model features and some *Pseudomonads* are represented by a very small number of examples that would go un-represented in the training sets of a K -fold cross validation. In the LOOV experimental design, a single genome is used as a validation set and the model is trained on the remaining genomes with a 10-fold cross-validation procedure and linear kernels. The selection of validation sample and training SVM is repeated until each of the 43 *Pseudomonas* genomes was used as the validation sample once. For generation of SVM, package ‘e1071’ v1.6–1 in R-project (August 29, 2013, <http://cran.r-project.org/web/packages/e1071/index.html>) was used. The outputs collected included class predictions, decision values for all training and validation samples and SVM files. A total of 16 SVM models, each with 43 LOOV, were generated: Four feature types based on computational model output types (enzyme function profiles, metabolomic model, secondary metabolism model, and transportomic model) were used to train for the prediction for each of the four ecological niche classes (biofilm formation, biocontrol agent, plant pathogen, and plant growth promoter).

A prediction confidence measurement was assigned to class predictions of validation samples. Using the SVM model for the positive examples in SVM training sets, the averages and standard deviations of distances from the hyperplane of SVM classifier were calculated. The statistical significance of the prediction of validation sample was calculated using its SVM distance x from the SVM classifier hyperplane and the standard normal distribution:

$$\text{Confidence of prediction} = 1 - \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (4)$$

where z , the normalized distance for the validation sample, is calculated as x divided by the standard deviation of the absolute values of training set positive decision value distances. A confidence of greater than or equal to 95% was considered a significant assignment of validation sample to ecological niche class. The complete set of confidence values for all validation predictions can be found in [S1 Data](#).

To quantitate the predictive power of SVMs, F-score was used. F-score is a metric that combines precision and recall of predictions and is calculated as follows:

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

where,

$$\text{Precision} = \frac{tp}{tp + fp}, \quad \text{Recall} = \frac{tp}{tp + fn} \quad (6, 7)$$

In *Precision* and *Recall*, tp is the number of true positives, fp is the number of false positives, and fn is the number of false negatives in predictions.

Results

Computational Model Overview

Enzyme function profiles were generated from the re-annotated *Pseudomonad* genomes. These profiles identified 1092 unique enzyme activities and 195 transmembrane transport annotations that were present in at least one genome. 606 of the enzyme functions were present and showed variation across *Pseudomonads* in over half of the re-annotated genomes and were

used to train Enzyme Function Profile SVMs. Metabolic (PRMT-scores) and transportomic (PRTT-scores) models for Pseudomonads were calculated using the complete enzyme function profiles. The complete metabolomic model is comprised of 6642 enzymatic transformation interactions between 3688 metabolites, of which 2143 were present and showed variation across Pseudomonads in over half of the re-annotated genomes and were used to train metabolomic SVMs. The secondary metabolism model is comprised of 1649 enzymatic transformation interactions between 1494 metabolites, of which 714 are variable across Pseudomonads and were used to train secondary metabolism SVMs. The transportomic model is predicted to transport 271 metabolites, of which 169 are predicted to be variably transported and were used to train transportomic SVMs.

Clustering by Enzyme Function Profiles Does Not Distinguish Between Ecological Niches

To determine if genomic data alone are sufficient to predict ecological niches, enzyme function profile data was used to generate hierarchical clusters using ‘Primer 6’ v6.1.10 (Primer-E Ltd., Luton, UK). Hierarchical clustering of genomic representation of enzyme functions (Fig 2) shows that Pseudomonads do not group by species or by ecological niche annotation. This inability to cluster genomes into species ecological activity groups by hierarchical clustering with these data indicates that other computational approaches are required to predict ecotype from genomic information.

SVMs Accurately Predict Rhizosphere Ecological Niche from System-Scale Model Outputs

Accuracy of SVM predictions, as quantitated by F-score, varies by type of model output used to train SVMs and by niche type (Fig 3). The transportomic model was the most predictive (i.e. highest F-score) for three out of four environmental niches: biocontrol, biofilm, and plant growth. Secondary metabolism was most predictive for the plant pathogen ecological niche. Enzyme function profile and complete metabolomic model were never the most predictive for any environmental niche. Considering the average F-score by SVM classifier type, both secondary metabolism features (t-test p-value 0.009) and transportomic model features (p-value 0.0003) are significantly more predictive than enzyme function profiles. Average F-score for

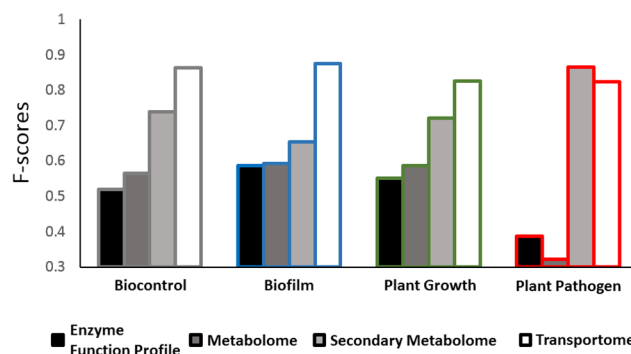


Fig 3. F-score for SVM predictions of ecological niche by input data type. Predictive capacity of SVM models is function of input type used to train model (enzyme profile, metabolomic, secondary metabolism or transportomic) and ecological niche in the rhizosphere (biocontrol, biofilm, plant growth promoter, or plant pathogen). Environmental niche classes were assigned at a confidence of > = 95%. F-scores are calculated from Leave One Out Validation (LOOV).

doi:10.1371/journal.pone.0132837.g003

SVM prediction using transportomic model features was significantly higher than average for SVM trained with metabolomic model features (p-value 0.002).

To avoid consideration of redundancy between features in the secondary metabolism and the less predictive complete metabolome feature types, only enzyme function profile, secondary metabolism model, and transportomic model are considered in the subsequent analysis of high-weight SVM features.

Highly Predictive SVM Features Provide Insights into Mechanisms of Adaptations to Ecological Niches

In SVMs, features used for training are assigned weights, proportional to their predictive capabilities with high-weight features more predictive than low weight features. We considered high-weight as more than 2 standard deviations +/- average feature weight for each SVM input type. The complete lists of high-weighted features are found in **Tables A-D** in [S2 File](#). The biological relevance of many of these features is supported by prior published observations and discussed in subsequent sections.

Highly predictive features for one ecological niche type are often present in the highly predictive features of another niche ([Fig 4](#)). For all training model output types, plant growth is the niche with the least overlap with other niches. Secondary metabolism has the largest proportion of high-weight features in common with all rhizosphere ecological

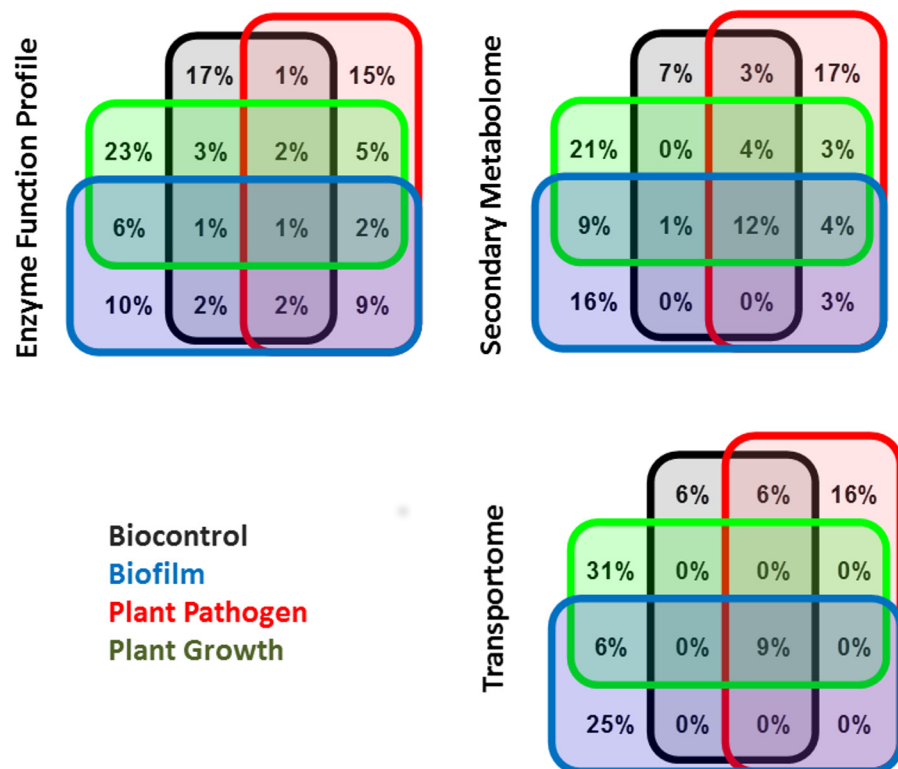


Fig 4. Venn diagram for significant features identified by SVM for each model feature type and for each ecological niche. All values in diagram are presented as percent of features out of total number of high-weight features for SVM feature type.

doi:10.1371/journal.pone.0132837.g004

niches. The transportomic model has the least overlap of high-weight features between niches with a small number of transported ligands common to all ecological niches.

Rhizosphere

High-weight features that are common to all rhizosphere ecological niches (Fig 4) are the molecular signatures for Pseudomonads that occupy the rhizosphere ecosystem, relative to those Pseudomonads that inhabit other environments (Tables A-D in S2 File). Secondary metabolism and transportomic have the largest sets of rhizosphere-specific features and many of the functions associated with these features are consistent with growth characteristics of organisms in a soil environment. Transport activities that are identified as predictive for inhabiting the rhizosphere involve carbohydrate transporters (e.g. 2-O-alpha-manosyl-D-glycerate) suggestive for osmoregulation in soils [43, 44] and 3-hydroxyphenylpropionic, one of many lignin breakdown products, which are ubiquitous in soils. We also identified the general class of cation transport as predictive for inhabiting the rhizosphere. This transporter class is possibly associated with maintaining charge balance in negatively charged soils in Pseudomonads and other soil bacteria [45, 46]. The ability of rhizosphere Pseudomonads to import lignin breakdown products is particularly relevant for the known important saprophytic capacities of Pseudomonads [47] which consume organic matter in soils. Another key metabolomic predictor is the capacity for catecholamine biosynthesis (Table C in S2 File), which are regulatory compounds found in many plants that are involved in growth and development and are regulated by stress conditions [48].

Biocontrol

Biocontrol is most predictive by its transportome (Table D in S2 File), specifically by transport of cobamide coenzyme (biochemically active forms of vitamin B12) and monosaccharides. Cobamide coenzyme is part of a vitamin biosynthesis pathway that induces resistance against pathogens and synthesis of growth factors in plant roots. Monosaccharide transport is a part of a sensor system associated with wound response and pathogen detection [49, 50]. Metabolomics (Table C in S2 File) predicts that metabolism of acetyl-D-glucosamine, a sugar that does not occur in plants or prokaryotes but is a structural polymer of fungi, is an indicator of biocontrol activity, specifically against fungal infection [51]. Additionally, intermediates in the pathway of isoniazid metabolism, which is an antimicrobial compound [52], is also identified as important to Pseudomonas biocontrol activities.

Biofilm

The most predictive metabolic activities for biofilm formation (Table C in S2 File) is the metabolism of anti-biofilm compounds protoporphyrin [53, 54] and methylglyoxal [55, 56], suggesting that a metabolic feature of these organisms involves a defense against biofilm inhibition synthesized by other competitors in the rhizosphere. Additional pathways previously implicated in biofilm formation include antranilate degradation pathways implicated in biofilm formation in *P. aeruginosa* [57] and the shikimate pathway [58, 59], which was identified as predictive for biofilm formation. Important transport functions predictive for biofilm formation (Table D in S2 File) are growth required environmental nutrients, specifically phosphorus and nitrogen. Limiting availability of both of these nutrients is an inducer of biofilm formation [60–62].

Plant Pathogen

Fatty acid biosynthesis pathways were identified as features predictive for plant pathogenicity in Pseudomonads. This computational prediction corresponds to the recently reported biological observation that lipid signaling is important for plant resistance to pathogens [63, 64]. Transport of plant sugars, such as arabinose [65, 66], and polyamines [67] are both important signals in plant stresses and defense against pathogens, and are also predictive of Pseudomonads' pathogenicity.

Plant Growth

Metabolomic input type (**Table C in S2 File**) predicts that synthesis of a number of plant signaling compounds is predictive of plant growth promotion by Pseudomonads including indole [68] and flavones eriodictyol, neringenin [69–71]. A number of transport functions were also identified (**Table D in S2 File**). C4-dicarboxylate is indicative of increased organic acid metabolism in the rhizosphere. Calcium transport is important in plant-root symbiote signaling [72, 73]. Glutathione transport is also predictive of plant growth promotion, and bacterially synthesized glutathione is previously reported as detected in the rhizosphere. Transport of a number of simple sugars (i.e. malonate, mannose, sucrose, galactose, and hexose) was found to be predictive of plant growth promotion by Pseudomonads and is suggestive of an ecological niche that is able to take advantage of exuded photosynthetic sugars present in the rhizosphere.

Discussion

Our analysis indicated that SVMs trained on outputs from enzyme function profiles, metabolic models, and transportomic models can be used to accurately predict the ecological niche of biofilms, biocontrol, plant growth promotion, or plant pathogen for a group of Pseudomonad organisms. A simple hierarchical clustering based on enzyme function profiles failed to distinguish Pseudomonads at the level of species or by rhizosphere ecological niche, suggesting a need for alternate approaches. The accuracy of SVM prediction was dependent upon model feature types used to train the SVM, with enzyme function profiles being the least predictive (F-scores between 0.44 and 0.60) and transportomic models the most predictive (F-scores between 0.82 and 0.89). Of intermediate predictive power is metabolic modeling, with secondary metabolome model features being more predictive of ecological niche than those of the complete metabolome. These results suggest that the most characteristic capability of an organism to fit into an ecological niche is not the set of enzyme functions available to it, but the system-scale mechanisms by which the bacterium senses and interacts with its environment. The most predictive model feature type for biofilm formation, plant growth promotion, and biocontrol in Pseudomonas was identified at its transportome. Our analysis revealed a novel aspect that the capability of an organism to occupy an ecological niche is most indicated by its ability to sense and manipulate its environment via its transmembrane transport capacity. The ability to model and quantitate bacterial transportomes using PRTT may provide new insights into microbial ecology and evolution of function.

Analysis of the most predictive features, i.e. those features with the highest SVM weights, for each ecotype identifies considerable overlap with prior biological knowledge, suggesting that not only are metabolomic and transportomic model features highly predictive for ecological niche, but also return results that are biologically significant. The agreement between computational predictions and previously published observations indicates that this analysis framework yields a number of potential hypotheses suitable for the design of molecular biological experiments. While transportomic feature data are found to be most predictive alone, it is very possible that a model using mixed feature data types may prove to have higher overall

predictive capabilities in future models. However, as optimizing predictive capacity, particularly on this relatively small set of Pseudomonad genomes, was not the goal of this research effort, we have elected not to present a mixed data-type SVM model here. In the context of a more general tool for accurately predicting ecological niches of uncharacterized bacteria from genomic data, such a mixed data-type SVM would be more appropriate and we are currently pursuing this goal.

Examining the most predictive features that were derived from each type of computational model data type provides a framework for a system-scale understanding of how specific molecular mechanisms in Pseudomonads contribute to their capacity to fill their varied ecological niche spaces in the rhizosphere. While Pseudomonads were considered here, the framework presented can be generalized for mining metagenomics and genomics data for new insights to bacterial functional determination and ecological niche prediction. As technology for sequencing and genome assembly continuously improves, the ability to generate completely sequenced bacterial genomes from environmental [74], clinical [75], or even single cell isolates [76] is expanding at an exponential rate. Of the 2749 completely sequenced and annotated genomes since 1995 listed in KEGG Organisms (http://www.genome.jp/kegg/catalog/org_list.html), over half have been generated since 2011. Many more thousands of draft bacterial genomes are currently in the process of completion and annotation. Yet only a relative handful of these bacteria have been characterized in the laboratory and the totality of what is known about many of these organisms is inferred from their genomic sequences. While the model generated in this study is not likely applicable to other taxa, the proposed framework can be applied to many other bacterial groupings and environmental niches. Computational approaches, such as the one that we have demonstrated here will be increasingly important to analyze and understand the role that bacteria play across all ecosystems.

Supporting Information

S1 Data. Compilation of Pseudomonas Species and Reference Data.
(XLSX)

S2 Data. FASTA-formatted file of 754,066 bacterial proteins annotated with EC enzyme functions.
(7Z)

S3 Data. List of KEGG Orthology (KO) annotations
(TXT)

S4 Data. FASTA-formatted file of 164,321 bacteria proteins annotated with KO transporter functions.
(ZIP)

S5 Data. Transporter-Ligand matrix. Matrix for use with PRMT scripts to generate PRTT-scores from tabular formatted data of transporter annotation counts in Pseudomonad annotated genomes.
(TXT)

S6 Data. All Enzyme function profiles for 43 re-annotated Pseudomonads. This file is comprised of two lists: EC enzyme annotation function counts and KO transporter function counts.
(TXT)

S7 Data. PRMT-scores for Pseudomonad metabolic models.
(TXT)

S8 Data. PRMT-scores for Pseudomonad secondary metabolic models.
(TXT)

S9 Data. PRTT-scores for Pseudomonads transportomic models.
(TXT)

S1 File. Ecological niche predictions for all Pseudomonads. There is one table in file for each feature type: **Table A**–Enzyme function profile, **Table B**–Metabolome, **Table C**–Secondary metabolome, and **Table D**–Transportome.
(XLSX)

S2 File. Lists of high-weight SVM features. Lists for each input model type and each rhizosphere ecological niche are given on separate tables: **Table A**, Enzyme Function Profile–Biocontrol, Biofilm, Plant Pathogen, and Plant Growth; **Table B**, Metabolomic–Biocontrol, Biofilm, Plant Pathogen, and Plant Growth; **Table C**, Secondary Metabolism–Biocontrol, Biofilm, Plant Pathogen, and Plant Growth; **Table D**, Transportomic–Biocontrol, Biofilm, Plant Pathogen, and Plant Growth. Each table is a model output feature type. Within each table, rows are model output features and columns are model output type. A ‘1’ indicates that a model output feature is high-weight by SVM, ‘0’ otherwise.
(XLSX)

Acknowledgments

This contribution originates in part from the “Environment Sensing and Response” Scientific Focus Area (SFA) program at Argonne National Laboratory. This research was supported by the U.S. Department of Energy, Office of Biological and Environmental Research (BER), as part of BER’s Genomic Science Program. This research has been funded by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

Author Contributions

Conceived and designed the experiments: PEL FRC YD. Performed the experiments: PEL YD. Analyzed the data: PEL FRC YD. Contributed reagents/materials/analysis tools: PEL FRC YD. Wrote the paper: PEL FRC YD.

References

1. Mendes R, Garbeva P, Raaijmakers JM. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev Sep*; 37(5):634–63. doi: [10.1111/1574-6976.12028](https://doi.org/10.1111/1574-6976.12028) PMID: [23790204](https://pubmed.ncbi.nlm.nih.gov/23790204/)
2. Newton AC, Fitt BD, Atkins SD, Walters DR, Daniell TJ. Pathogenesis, parasitism and mutualism in the trophic space of microbe-plant interactions. *Trends Microbiol Aug*; 18(8):365–73. doi: [10.1016/j.tim.2010.06.002](https://doi.org/10.1016/j.tim.2010.06.002) PMID: [20598545](https://pubmed.ncbi.nlm.nih.gov/20598545/)
3. Gamper H, Hartwig UA, Leuchtman A. Mycorrhizas improve nitrogen nutrition of *Trifolium repens* after 8 yr of selection under elevated atmospheric CO₂ partial pressure. *The New phytologist*. [Research Support, Non-U.S. Gov’t]. 2005 Aug; 167(2):531–42. PMID: [15998404](https://pubmed.ncbi.nlm.nih.gov/15998404/)

4. Drigo B, Pijl AS, Duyts H, Kielak AM, Gamper HA, Houtekamer MJ, et al. Shifting carbon flow from roots into associated microbial communities in response to elevated atmospheric CO₂. *Proceedings of the National Academy of Sciences of the United States of America*. [Research Support, Non-U.S. Gov't]. 2010 Jun 15; 107(24):10938–42. doi: [10.1073/pnas.0912421107](https://doi.org/10.1073/pnas.0912421107) PMID: [20534474](https://pubmed.ncbi.nlm.nih.gov/20534474/)
5. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* Nov; 11(11):789–99. doi: [10.1038/nrmicro3109](https://doi.org/10.1038/nrmicro3109) PMID: [24056930](https://pubmed.ncbi.nlm.nih.gov/24056930/)
6. Berg G, Smalla K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* 2009 Apr; 68(1):1–13. doi: [10.1111/j.1574-6941.2009.00654.x](https://doi.org/10.1111/j.1574-6941.2009.00654.x) PMID: [19243436](https://pubmed.ncbi.nlm.nih.gov/19243436/)
7. Knief C. Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Front Plant Sci*; 5:216. doi: [10.3389/fpls.2014.00216](https://doi.org/10.3389/fpls.2014.00216) PMID: [24904612](https://pubmed.ncbi.nlm.nih.gov/24904612/)
8. Lennon JT, Aanderud ZT, Lehmkuhl BK, Schoolmaster DR Jr. Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* Aug; 93(8):1867–79. PMID: [22928415](https://pubmed.ncbi.nlm.nih.gov/22928415/)
9. Wu X, Monchy S, Taghavi S, Zhu W, Ramos J, van der Lelie D. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* Mar; 35(2):299–323. doi: [10.1111/j.1574-6976.2010.00249.x](https://doi.org/10.1111/j.1574-6976.2010.00249.x) PMID: [20796030](https://pubmed.ncbi.nlm.nih.gov/20796030/)
10. Suen G, Goldman BS, Welch RD. Predicting Prokaryotic Ecological Niches Using Genome Sequence Analysis. *PLoS One* 2007; 2(8).
11. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 2007 Sep; 45(9):2761–4. PMID: [17626177](https://pubmed.ncbi.nlm.nih.gov/17626177/)
12. Rajendhran J, Gunasekaran P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* Feb 20; 166(2):99–110. doi: [10.1016/j.micres.2010.02.003](https://doi.org/10.1016/j.micres.2010.02.003) PMID: [20223646](https://pubmed.ncbi.nlm.nih.gov/20223646/)
13. Kunin V, Engelbrekton A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology*. [Research Support, U.S. Gov't, Non-P.H.S.]. 2010 Jan; 12(1):118–23. doi: [10.1111/j.1462-2920.2009.02051.x](https://doi.org/10.1111/j.1462-2920.2009.02051.x) PMID: [19725865](https://pubmed.ncbi.nlm.nih.gov/19725865/)
14. Bannert A, Kleineidam K, Wissing L, Mueller-Niggemann C, Vogelsang V, Welzl G, et al. Changes in diversity and functional gene abundances of microbial communities involved in nitrogen fixation, nitrification, and denitrification in a tidal wetland versus paddy soils cultivated for different time periods. *Appl Environ Microbiol* Sep; 77(17):6109–16. doi: [10.1128/AEM.01751-10](https://doi.org/10.1128/AEM.01751-10) PMID: [21764972](https://pubmed.ncbi.nlm.nih.gov/21764972/)
15. Sims A, Horton J, Gajaraj S, McIntosh S, Miles RJ, Mueller R, et al. Temporal and spatial distributions of ammonia-oxidizing archaea and bacteria and their ratio as an indicator of oligotrophic conditions in natural wetlands. *Water Res* Sep 1; 46(13):4121–9. doi: [10.1016/j.watres.2012.05.007](https://doi.org/10.1016/j.watres.2012.05.007) PMID: [22673339](https://pubmed.ncbi.nlm.nih.gov/22673339/)
16. You J, Das A, Dolan EM, Hu Z. Ammonia-oxidizing archaea involved in nitrogen removal. *Water Res* 2009 Apr; 43(7):1801–9. doi: [10.1016/j.watres.2009.01.016](https://doi.org/10.1016/j.watres.2009.01.016) PMID: [19232671](https://pubmed.ncbi.nlm.nih.gov/19232671/)
17. Kapitein N, Mogk A. Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. *Curr Opin Microbiol* Feb; 16(1):52–8. doi: [10.1016/j.mib.2012.11.009](https://doi.org/10.1016/j.mib.2012.11.009) PMID: [23290191](https://pubmed.ncbi.nlm.nih.gov/23290191/)
18. Cascales E. The type VI secretion toolkit. *EMBO Rep* 2008 Aug; 9(8):735–41. doi: [10.1038/embor.2008.131](https://doi.org/10.1038/embor.2008.131) PMID: [18617888](https://pubmed.ncbi.nlm.nih.gov/18617888/)
19. Jani AJ, Cotter PA. Type VI secretion: not just for pathogenesis anymore. *Cell Host Microbe* Jul 22; 8(1):2–6. doi: [10.1016/j.chom.2010.06.012](https://doi.org/10.1016/j.chom.2010.06.012) PMID: [20638635](https://pubmed.ncbi.nlm.nih.gov/20638635/)
20. Sels J, Mathys J, De Coninck BM, Cammue BP, De Bolle MF. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*. [Research Support, Non-U.S. Gov't/Review]. 2008 Nov; 46(11):941–50.
21. Winter G, Kromer JO. Fluxomics—connecting 'omics analysis and phenotypes. *Environ Microbiol* Jul; 15(7):1901–16. doi: [10.1111/1462-2920.12064](https://doi.org/10.1111/1462-2920.12064) PMID: [23279205](https://pubmed.ncbi.nlm.nih.gov/23279205/)
22. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol Biol*; 985:17–45. doi: [10.1007/978-1-62703-299-5_2](https://doi.org/10.1007/978-1-62703-299-5_2) PMID: [23417797](https://pubmed.ncbi.nlm.nih.gov/23417797/)
23. Gross H, Loper JE. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat Prod Rep* 2009 Nov; 26(11):1408–46. doi: [10.1039/b817075b](https://doi.org/10.1039/b817075b) PMID: [19844639](https://pubmed.ncbi.nlm.nih.gov/19844639/)
24. Clarke PH. The metabolic versatility of pseudomonads. *Antonie Van Leeuwenhoek* 1982 May; 48(2):105–30. PMID: [6808915](https://pubmed.ncbi.nlm.nih.gov/6808915/)

25. Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 2011 Jul; 35(4):652–80. doi: [10.1111/j.1574-6976.2011.00269.x](https://doi.org/10.1111/j.1574-6976.2011.00269.x) PMID: [21361996](https://pubmed.ncbi.nlm.nih.gov/21361996/)
26. Haas D, Defago G. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol* 2005 Apr; 3(4):307–19. PMID: [15759041](https://pubmed.ncbi.nlm.nih.gov/15759041/)
27. Ryan RP, Germaine K, Franks A, Ryan DJ, Dowling DN. Bacterial endophytes: recent developments and applications. *FEMS Microbiol Lett* 2008 Jan; 278(1):1–9. PMID: [18034833](https://pubmed.ncbi.nlm.nih.gov/18034833/)
28. Garbeva P, de Boer W. Inter-specific interactions between carbon-limited soil bacteria affect behavior and gene expression. *Microb Ecol* 2009 Jul; 58(1):36–46. doi: [10.1007/s00248-009-9502-3](https://doi.org/10.1007/s00248-009-9502-3) PMID: [19267150](https://pubmed.ncbi.nlm.nih.gov/19267150/)
29. Alhede M, Bjarnsholt T, Givskov M. *Pseudomonas aeruginosa* biofilms: mechanisms of immune evasion. *Adv Appl Microbiol*; 86:1–40. doi: [10.1016/B978-0-12-800262-9.00001-9](https://doi.org/10.1016/B978-0-12-800262-9.00001-9) PMID: [24377853](https://pubmed.ncbi.nlm.nih.gov/24377853/)
30. Quiles F, Humbert F. On the production of glycogen by *Pseudomonas fluorescens* during biofilm development: an in situ study by attenuated total reflection-infrared with chemometrics. *Biofouling* Jul; 30(6):709–18. doi: [10.1080/08927014.2014.915956](https://doi.org/10.1080/08927014.2014.915956) PMID: [24835847](https://pubmed.ncbi.nlm.nih.gov/24835847/)
31. Rudrappa T, Biedrzycki ML, Bais HP. Causes and consequences of plant-associated biofilms. *FEMS Microbiol Ecol* 2008 May; 64(2):153–66. doi: [10.1111/j.1574-6941.2008.00465.x](https://doi.org/10.1111/j.1574-6941.2008.00465.x) PMID: [18355294](https://pubmed.ncbi.nlm.nih.gov/18355294/)
32. Sands DC, Schroth MN, Hildebrand DC. Taxonomy of phytopathogenic pseudomonads. *J Bacteriol* 1970 Jan; 101(1):9–23. PMID: [5411761](https://pubmed.ncbi.nlm.nih.gov/5411761/)
33. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* Aug; 13(6):614–29. doi: [10.1111/j.1364-3703.2012.00804.x](https://doi.org/10.1111/j.1364-3703.2012.00804.x) PMID: [22672649](https://pubmed.ncbi.nlm.nih.gov/22672649/)
34. Frey-Klett P, Garbaye J, Tarkka M. The mycorrhiza helper bacteria revisited. *New Phytol* 2007; 176(1):22–36. PMID: [17803639](https://pubmed.ncbi.nlm.nih.gov/17803639/)
35. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*. [Research Support, Non-U.S. Gov't]. 2012 Jan; 40(Database issue):D109–14. doi: [10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988) PMID: [22080510](https://pubmed.ncbi.nlm.nih.gov/22080510/)
36. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*. [Research Support, Non-U.S. Gov't]. 1999 Jan 1; 27(1):29–34. PMID: [9847135](https://pubmed.ncbi.nlm.nih.gov/9847135/)
37. Bairoch A. The ENZYME database in 2000. *Nucleic acids research* 2000 Jan 1; 28(1):304–5. PMID: [10592255](https://pubmed.ncbi.nlm.nih.gov/10592255/)
38. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 2005 Oct 1; 21(19):3787–93. PMID: [15817693](https://pubmed.ncbi.nlm.nih.gov/15817693/)
39. Larsen PE, Collart F, Field D, Meyer F, Keegan KP, Henry CS, et al. Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microbial Informatics and Experimentation* 2011; 1(4).
40. Scott NM, Hess M, Bouskill NJ, Mason OU, Jansson JK, Gilbert JA. The microbial nitrogen cycling potential is impacted by polyaromatic hydrocarbon pollution of marine sediments. *Front Microbiol* 2014; 5:108. doi: [10.3389/fmicb.2014.00108](https://doi.org/10.3389/fmicb.2014.00108) PMID: [24723913](https://pubmed.ncbi.nlm.nih.gov/24723913/)
41. Larsen PE, Collart FR, Dai Y. Using metabolomic and transportomic modeling and machine learning to identify putative novel therapeutic targets for antibiotic resistant *Pseudomonad* infections. *Conf Proc IEEE Eng Med Biol Soc* 2014; 2014:314–7. doi: [10.1109/EMBC.2014.6943592](https://doi.org/10.1109/EMBC.2014.6943592) PMID: [25569960](https://pubmed.ncbi.nlm.nih.gov/25569960/)
42. Larsen PE, Scott N, Post AF, Field D, Knight R, Hamada Y, et al. Satellite remote sensing data can be used to model marine microbial metabolite turnover. *ISME J* 2015 Jan; 9(1):166–79. doi: [10.1038/ismej.2014.107](https://doi.org/10.1038/ismej.2014.107) PMID: [25072414](https://pubmed.ncbi.nlm.nih.gov/25072414/)
43. Miller KJ, Kennedy EP, Reinhold VN. Osmotic adaptation by gram-negative bacteria: possible role for periplasmic oligosaccharides. *Science* 1986 Jan 3; 231(4733):48–51. PMID: [3941890](https://pubmed.ncbi.nlm.nih.gov/3941890/)
44. Talaga P, Fournet B, Bohin JP. Periplasmic glucans of *Pseudomonas syringae* pv. *syringae*. *J Bacteriol* 1994 Nov; 176(21):6538–44. PMID: [7961404](https://pubmed.ncbi.nlm.nih.gov/7961404/)
45. Cornelis P. Iron uptake and metabolism in pseudomonads. *Appl Microbiol Biotechnol* May; 86(6):1637–45. doi: [10.1007/s00253-010-2550-2](https://doi.org/10.1007/s00253-010-2550-2) PMID: [20352420](https://pubmed.ncbi.nlm.nih.gov/20352420/)
46. Kraepiel AM, Bellenger JP, Wichard T, Morel FM. Multiple roles of siderophores in free-living nitrogen-fixing bacteria. *Biomaterials* 2009 Aug; 22(4):573–81. doi: [10.1007/s10534-009-9222-7](https://doi.org/10.1007/s10534-009-9222-7) PMID: [19277875](https://pubmed.ncbi.nlm.nih.gov/19277875/)
47. Ruiz-Duenas FJ, Martinez AT. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb Biotechnol* 2009 Mar; 2(2):164–77. doi: [10.1111/j.1751-7915.2008.00078.x](https://doi.org/10.1111/j.1751-7915.2008.00078.x) PMID: [21261911](https://pubmed.ncbi.nlm.nih.gov/21261911/)

48. Kulma A, Szopa J. Catecholamines are active compounds in plants. *Plant Science*2007 Mar; 172(3):433–40.
49. Lalonde S, Boles E, Hellmann H, Barker L, Patrick JW, Frommer WB, et al. The dual function of sugar carriers. Transport and sugar sensing. *Plant Cell*1999 Apr; 11(4):707–26. PMID: [10213788](#)
50. Conde C, Agasse A, Glissant D, Tavares R, Geros H, Delrot S. Pathways of glucose regulation of monosaccharide transport in grape cells. *Plant Physiol*2006 Aug; 141(4):1563–77. PMID: [16766675](#)
51. Haran S, Schickler H, Chet I. Molecular mechanisms of lytic enzymes involved in the biocontrol activity of *Trichoderma harzianum*. *Microbiology-Uk*1996 Sep; 142:2321–31.
52. Suarez J, Rangelova K, Jarzecki AA, Manzerova J, Krymov V, Zhao XB, et al. An Oxyferrous Heme/Protein-based Radical Intermediate Is Catalytically Competent in the Catalase Reaction of *Mycobacterium tuberculosis* Catalase-Peroxidase (KatG). *Journal of Biological Chemistry*2009 Mar 13; 284(11):7017–29. doi: [10.1074/jbc.M808106200](#) PMID: [19139099](#)
53. Olczak T, Maszczak-Seneczko D, Smalley JW, Olczak M. Gallium(III), cobalt(III) and copper(II) protoporphyrin IX exhibit antimicrobial activity against *Porphyromonas gingivalis* by reducing planktonic and biofilm growth and invasion of host epithelial cells. *Archives of Microbiology*2012 Aug; 194(8):719–24. doi: [10.1007/s00203-012-0804-3](#) PMID: [22447101](#)
54. Ma HY, Darmawan ET, Zhang M, Zhang L, Bryers JD. Development of a poly(ether urethane) system for the controlled release of two novel anti-biofilm agents based on gallium or zinc and its efficacy to prevent bacterial biofilm formation. *Journal of Controlled Release*2013 Dec 28; 172(3):1035–44. doi: [10.1016/j.jconrel.2013.10.005](#) PMID: [24140747](#)
55. Majtan J, Bohova J, Horniackova M, Klaudiny J, Majtan V. Anti-biofilm Effects of Honey Against Wound Pathogens *Proteus mirabilis* and *Enterobacter cloacae*. *Phytotherapy Research*2014 Jan; 28(1):69–75. doi: [10.1002/ptr.4957](#) PMID: [23494861](#)
56. Kilty SJ, Duval M, Chan FT, Ferris W, Slinger R. Methylglyoxal: (active agent of manuka honey) in vitro activity against bacterial biofilms. *International Forum of Allergy & Rhinology*2011 Sep-Oct; 1(5):348–50.
57. Costaglioli P, Barthe C, Claverol S, Brozel VS, Perrot M, Crouzet M, et al. Evidence for the involvement of the anthranilate degradation pathway in *Pseudomonas aeruginosa* biofilm formation. *Microbiologyopen* Sep; 1(3):326–39. doi: [10.1002/mbo3.33](#) PMID: [23170231](#)
58. Dorel C, Lejeune P, Rodrigue A. The Cpx system of *Escherichia coli*, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities? *Res Microbiol*2006 May; 157(4):306–14. PMID: [16487683](#)
59. Munoz-Elias EJ, Marcano J, Camilli A. Isolation of *Streptococcus pneumoniae* biofilm mutants and their characterization during nasopharyngeal colonization. *Infect Immun*2008 Nov; 76(11):5049–61. doi: [10.1128/IAI.00425-08](#) PMID: [18794289](#)
60. Bowden GH, Li YH. Nutritional influences on biofilm development. *Adv Dent Res*1997 Apr; 11(1):81–99. PMID: [9524446](#)
61. Filiatrault MJ, Tomblin G, Wagner VE, Van Alst N, Rumbaugh K, Sokol P, et al. *Pseudomonas aeruginosa* PA1006, which plays a role in molybdenum homeostasis, is required for nitrate utilization, biofilm formation, and virulence. *PLoS One*; 8(2):e55594. doi: [10.1371/journal.pone.0055594](#) PMID: [23409004](#)
62. Tielen P, Rosin N, Meyer AK, Dohnt K, Haddad I, Jansch L, et al. Regulatory and metabolic networks for the adaptation of *Pseudomonas aeruginosa* biofilms to urinary tract-like conditions. *PLoS One*; 8(8):e71845. doi: [10.1371/journal.pone.0071845](#) PMID: [23967252](#)
63. Okazaki Y, Saito K. Roles of lipids as signaling molecules and mitigators during stress response in plants. *Plant J* May 20.
64. Raffaele S, Leger A, Roby D. Very long chain fatty acid and lipid signaling in the response of plants to pathogens. *Plant Signal Behav*2009 Feb; 4(2):94–9. PMID: [19649180](#)
65. Hu X, Zhao J, DeGrado WF, Binns AN. *Agrobacterium tumefaciens* recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. *Proc Natl Acad Sci U S A* Jan 8; 110(2):678–83. doi: [10.1073/pnas.1215033110](#) PMID: [23267119](#)
66. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, et al. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci U S A*2003 Sep 2; 100(18):10181–6. PMID: [12928499](#)
67. Hussain SS, Ali M, Ahmad M, Siddique KH. Polyamines: natural and engineered abiotic and biotic stress tolerance in plants. *Biotechnol Adv* May-Jun; 29(3):300–11. doi: [10.1016/j.biotechadv.2011.01.003](#) PMID: [21241790](#)
68. Peer WA. From perception to attenuation: auxin signalling and responses. *Curr Opin Plant Biol* Oct; 16(5):561–8. doi: [10.1016/j.pbi.2013.08.003](#) PMID: [24004572](#)

69. Abdel-Lateif K, Bogusz D, Hocher V. The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and Frankia bacteria. *Plant Signal Behav* Jun; 7(6):636–41. doi: [10.4161/psb.20039](https://doi.org/10.4161/psb.20039) PMID: [22580697](https://pubmed.ncbi.nlm.nih.gov/22580697/)
70. Steinkellner S, Lenzemo V, Langer I, Schweiger P, Khaosaad T, Toussaint JP, et al. Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules* 2007 Jul; 12(7):1290–306. PMID: [17909485](https://pubmed.ncbi.nlm.nih.gov/17909485/)
71. Phillips DA, Tsai SM. Flavonoids as plant signals to rhizosphere microbes. *Mycorrhiza* 1992; 1(2):55–8.
72. Kosuta S, Hazledine S, Sun J, Miwa H, Morris RJ, Downie JA, et al. Differential and chaotic calcium signatures in the symbiosis signaling pathway of legumes. *Proc Natl Acad Sci U S A* 2008 Jul 15; 105(28):9823–8. doi: [10.1073/pnas.0803499105](https://doi.org/10.1073/pnas.0803499105) PMID: [18606999](https://pubmed.ncbi.nlm.nih.gov/18606999/)
73. Singh S, Parniske M. Activation of calcium- and calmodulin-dependent protein kinase (CCaMK), the central regulator of plant root endosymbiosis. *Curr Opin Plant Biol* Aug; 15(4):444–53. doi: [10.1016/j.pbi.2012.04.002](https://doi.org/10.1016/j.pbi.2012.04.002) PMID: [22727503](https://pubmed.ncbi.nlm.nih.gov/22727503/)
74. Thomas T, Gilbert J, Meyer F. Metagenomics—a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2012; 2(1):3. doi: [10.1186/2042-5783-2-3](https://doi.org/10.1186/2042-5783-2-3) PMID: [22587947](https://pubmed.ncbi.nlm.nih.gov/22587947/)
75. Jones ML, Ganopoulos JG, Martoni CJ, Labbe A, Prakash S. Emerging science of the human microbiome. *Gut Microbes* Jul 11; 5(4).
76. Baslan T, Hicks J. Single cell sequencing approaches for complex biological systems. *Curr Opin Genet Dev* Jul 10; 26C:59–65.