Revisiting the Genomic Structure of the Genus *Pseudomonas* with Whole Genome Data: Insights into Diversity and Host-related Genetic Determinants

Buqing Yi* and Alexander H. Dalpke

Institute of Medical Microbiology and Hygiene, Medical Faculty, Technical University Dresden, Dresden, Germany

*corresponding author:

Dr Buqing Yi, Institute of Medical Microbiology and Hygiene, Medical Faculty, Technical University Dresden, Dresden, Germany Fetscherstraße 74, 01307 Dresden, Germany, Tel.: +49 351 4586440, buqing.yi@uniklinikum-dresden.de

Abstract:

Pseudomonas spp. exhibit considerable differences in virulence, host specificity, and colonization. Many severe clinical infections are caused by *Pseudomonas* spp., and these bacteria are also responsible for many plant diseases. Interestingly, a variety of *Pseudomonas* spp. may actually promote plant growth at the same time. Understanding genome variations that generate the tremendous diversity in *Pseudomonas* biology is important in reducing and controlling the incidence of diseases, and can also help predict the pathogenic potential of newly discovered strains. In this study, we aimed to elucidate the genome-wide population structure of *Pseudomonas*. With a data set of 704 *Pseudomonas* whole genome sequences representing 186 species, we calculated a genus wide pan-genome composed of 62,202 genes. The genomic structure of the genus *Pseudomona*s was explored by hierarchical clustering based on average nucleotide identity, and also by phylogeny analysis based on concatenated core-gene alignment. Interestingly, most species that may act as opportunistic human pathogen appear closely related and fall into one clade. Further comparative functional analyses indicate that *Pseudomonas* species that only live in natural habitats lack multiple functions involved in stress response, which are predominantly present in species that may act as opportunistic human pathogen. In summary, this study provided detailed insights into the dynamics of genome diversity of *Pseudomonas,* and revealed host-related genetic determinants, which might help the development of more targeted antibiotics for the treatment of clinical *Pseudomonas* infections.

Introduction

The gram-negative genus *Pseudomonas* (phylum Proteobacteria) constitutes a highly diverse group which is widely distributed in nature, utilizing a wide range of organic compounds as energy sources (Lessie & Phibbs, 1984; Wu et al., 2011) and producing a large number of secondary metabolites (Gross & Loper, 2009).

Pseudomonas spp. are responsible for many plant diseases, but at the same time some *Pseudomonas* spp. may actually promote plant growth or protect plant from infections (Haas & Défago, 2005; Mendes et al., 2011). A variety of *Pseudomonas* spp. may act as opportunistic human pathogen. For example, *P. putida* strains are prevalent in natural habitats, including soil, water, and plant surfaces, but have also been reported as opportunistic human pathogens causing nosocomial infections (Fernández et al., 2015). *P. stutzeri* is a nonfluorescent denitrifying bacterium widely distributed in the environment, and several strains of this species have also been isolated as opportunistic pathogen from humans (Lalucat et al., 2006). Similarly, in addition to being frequently detected from many different environmental sources, *P. aeruginosa* acts as an important opportunistic human pathogen implicated in the pathogenesis of many diseases, e.g. in lung infection in cystic fibrosis (CF). In the latter disease, dominance of *P. aeruginosa* in the lung microbial community correlates with CF disease features, such as cough scores, chest radiograph scores, and loss of pulmonary function. *P. aeruginosa* mucoid conversion within lungs of CF patients is a hallmark of chronic infection and predictive of poor prognosis (Emerson et al., 2002; Konstan et al., 2007; Nixon et al., 2001)

Pseudomonas can be transmitted via several distinct routes from animals, plants or environmental sources to human hosts. To control disease incidences related to *Pseudomonas* infection and predict the pathogenic potential of newly isolated strains, it is important to understand the genome variations behind the highly diversified host options. However, most studies dealing with *Pseudomona*s spp. genomic variation have been focused on one specific human or plant pathogen, such as *P. aeruginosa* (Freschi et al., 2019; Wiehlmann et al., 2007) or *P. syringae* (Hulin et al., 2018), the genomic variation among the *Pseudomonas* species remains poorly understood. Only a few studies have so far investigated the genomic diversity of *Pseudomonas* species (e.g. (Gomila et al., 2015; Loper et al., 2012; Mulet et al., 2010; Yamamoto et al., 2000)), among which the most recent study was performed with genome sequences available till December 2012 (Gomila et al., 2015). Therefore, it appears helpful to update investigations on the genomic variation of *Pseudomonas*. In this study, we aimed to elucidate the genome-wide population structure of *Pseudomonas* with a data set of 704 *Pseudomonas* whole genome sequences representing 186 species, comprising all complete genome sequences from NCBI available by January 2020.

Results

1. Pan-genome characteristics of *Pseudomonas*

To investigate the genomic diversity of *Pseudomonas* spp., we compiled a total of 704 genomes downloaded from NCBI (see Table S1 in the supplemental material). This dataset contains genomes representing 186 species, among which 95 species have been validly published (verified through The Bacterial Diversity Metadatabase https://bacdive.dsmz.de/), and the other 91 species have not been validated yet (see Table S2 for a detailed species information). Across the entire data set, genome size varied between 3.19 and 7.76 Mb (mean, 6.32 Mb), GC content varied between 0.38 and 0.68 (mean, 0.63), the number of predicted genes ranged from 2,831 to 7,360 (mean, 5,637).

We performed pangenome analysis of the *Pseudomonas* dataset with anvi'o (version 6.1) (Eren et al., 2015) following the workflow for microbial pangenomics. We first analyzed the number of orthologous gene families that are shared by all the strains, and the results showed that only 122 genes are present in 100% *Pseudomonas* stains, which represents only 0.2% of the entire pan-genome (62,202 genes). To take sequencing and assembly errors into account, based on a widely used criterion (Page et al., 2015; Park & Andam, 2019, 2020), orthologous gene families have been classified into core genome (present in 99% ≤ strains ≤100%) and accessory genome (present in < 99% stains). We found a considerably small core genome at the genus level, composed of 1025 genes (see Table S3 for detailed information), representing 1.65 % of the entire pan-genome. It is notable that in the accessory genome, 36,131 accessory genes representing 58% of the pan-genome are only present in < 1% of the strains. The large accessory genome (98.35% of the pan-genome) indicates a high level of genomic diversity of *Pseudomonas* genomes, which is consistent with the observed biological diversity of the *Pseudomonas*, highlighting the heterogeneity of this group of bacteria.

2. Genome-wide population structure

To determine the degree of genomic relatedness and hence clarify the relationship between the *Pseudomonas* spp., we calculated the pairwise average nucleotide identity (ANI) for all possible pairs of genomes. ANI estimates the average nucleotide identity of all orthologous genes shared between any two genomes, usually ≥ 95% when comparing organisms belonging to the same species (Jain et al., 2018). This index is a robust similarity metric that has been broadly used to delineate inter- and intra-species strain relatedness. It was reported in previous *Pseudomonas* genus phylogeny analysis (Gomila et al., 2015) that ANI results show very similar topology as phylogeny analysis based on multi-locus sequence analysis (MLSA).

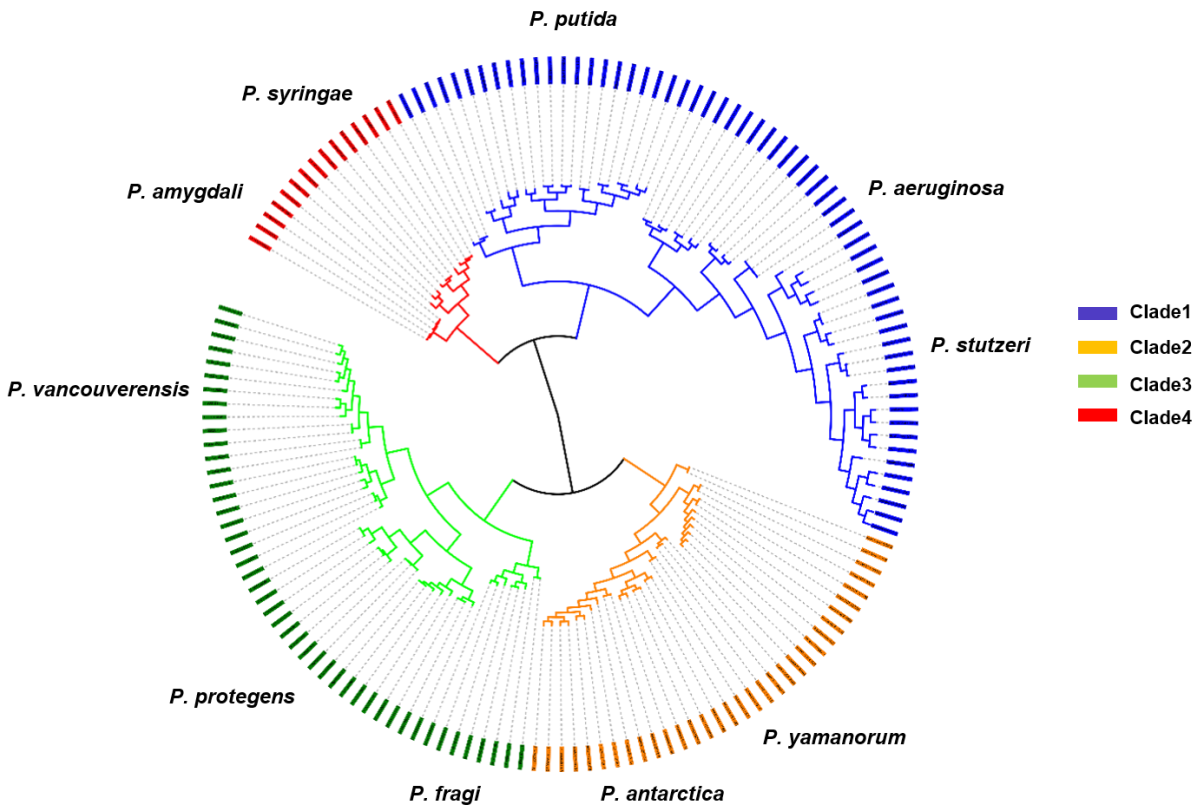

Fig. 1 Population structure of the genus *Pseudomonas* based on whole-genome comparisons (ANIb) of 155 *Pseudomonas* strains representing 95 valid *Pseudomonas* species. The calculations were performed using the Python module PYANI. The strains fall into four major Clades (Clade1-Clade4). The positions of representative species in each Clade were labelled.

We included all valid *Pseudomonas* species in our dataset for ANI analysis. To achieve a better resolution of the population structure at the genus level, for species with multiple genomes, maximal three genomes from one species were included in the ANI analysis. Thus, we have performed ANI analysis with 155 representative strains from all 95 valid species in our dataset including 25 reference or representative genomes (see Table S4 for detailed information). Based on whole-genome similarity index ANIb, *Pseudomonas* strains clearly fall into four clades. Representative species in each clade are labelled in Figure 1. Most species that may act as opportunistic human pathogen or use animals as host appear closely related and fall into one clade, such as *P. aeruginosa, P. stutzeri and P. putida*. Strains of the *P. syringae* group (note: *P. amygdali* has been verified belonging to the *P. syringae* group (Hwang et al., 2005; Parkinson et al., 2011)) do not share much features with strains of other *Pseudomonas* species, forming a one-species clade. Likewise, the population structure derived by ANIm analysis comprises four clades (Figure S1), and the species fall into each clade are almost identical between ANIb and ANIm based analyses.

We also analyzed the genomic structure of the genus *Pseudomonas* using concatenated alignment of all single-copy core genes based on pan-genome analysis. As shown in Figure 2, *Pseudomonas* phylogeny tree constructed with this method showed similar topology as ANI dendrograms (Figure 1 & Figure S1)

based on the relative position of representative species. In this phylogeny tree, all 704 genomes from both valid and not-yet validated species have been included.

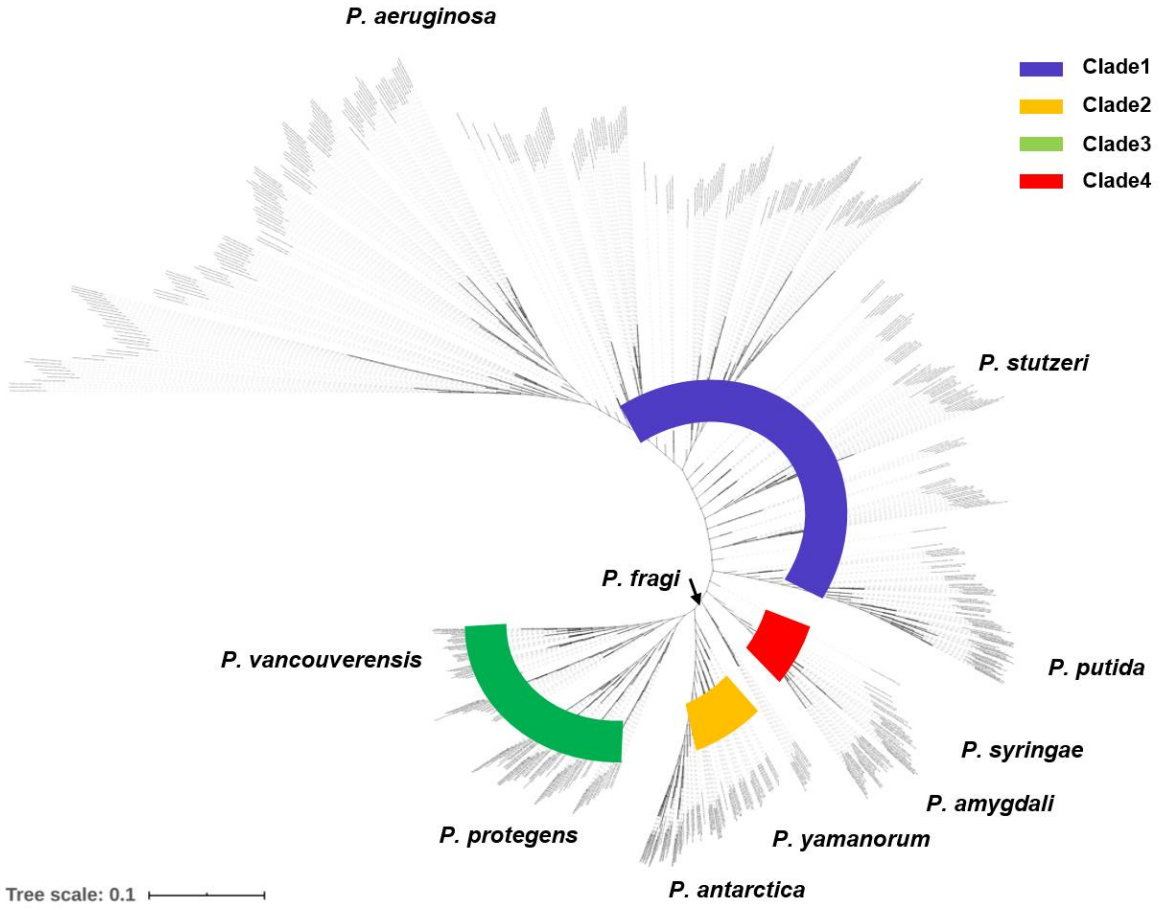

Fig. 2 Phylogeny of 704 *Pseudomonas* strains. The phylogenetic analysis was performed using the Maximum Likelihood method based on concatenated single-copy core genes. Four Clades comparable to the four major Clades formed by whole-genome comparisons (ANI) have been detected and color-labelled. The positions of representative species in each Clade were labelled.

3. **Comparative functional enrichment analysis reveals host-specific features**.

Among the *Pseudomonas* spp., a variety of species are human opportunistic pathogens causing devastating symptoms in patients suffering from immunodeficiency, cystic fibrosis, or intensive care unit patients. Meanwhile, there are many *Pseudomonas* spp. purely isolated from environmental sources ranging from soil to plant. Plants can represent reservoirs for human infections as some pathogens can survive and sometimes proliferate in the rhizosphere. As we know, many *Pseudomonas* spp. have been isolated from plants, including human opportunistic pathogen *P. aeruginosa*. It remains unclear why some species may only use plants as host, but some other species may use both humans and plants as host.

ANI based genome population structure analysis indicates that most species that may act as opportunistic human pathogen or use animals as host are closely related and fall into clade1 (Figure 1 & Figure S1). We asked whether there are any functions predominantly present in genomes from clade 1, while absent from other clades. This information might provide hints for host-related genetic determinants. We therefore conducted comparative functional analyses with the aim of finding specific niche factors. Because the numbers of genomes in each clade were greatly dissimilar, we restricted our comparative analyses to five representative genomes from each clade, including all known reference genomes or representative genomes since these genomes usually have well-documented ecological information. The detailed information of these genomes is listed in Table S5. Their hosts or environments are representative of the diversity of ecological niches of *Pseudomonas*.

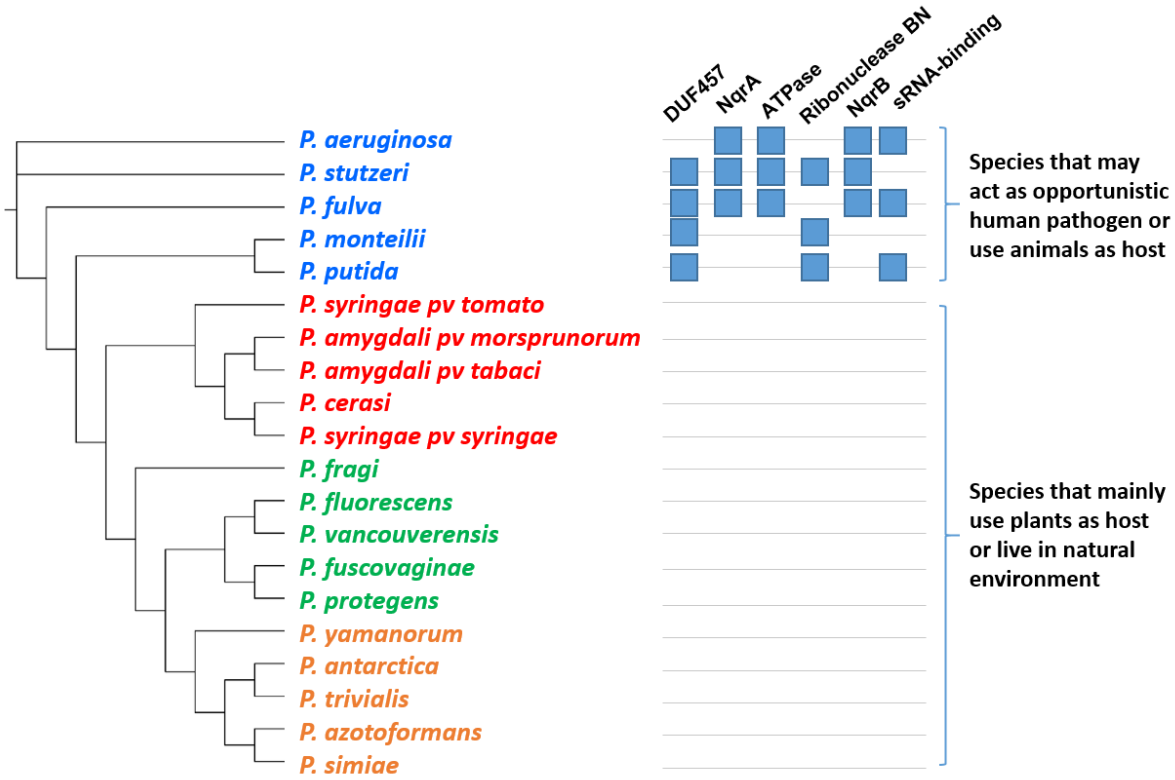

Fig. 3 Phylogeny and distribution of enriched functions for 20 representative *Pseudomonas* species. The evolutionary distances were computed using the Maximum Likelihood method based on concatenated single-copy core genes. Branch label colors are indicative of the ANI based Clade that the respective species fall into, with Clade 1 labelled as blue, Clade 2 as yellow, Clade 3 as green and Clade4 as red. Squares correspond to functions that were found to be preferably enriched in species that may act as opportunistic human pathogen or use animals as host. From left to right: Membrane-bound metal-dependent hydrolase YbcI, DUF457 family; Na+-transporting NADH:ubiquinone oxidoreductase (NQR), subunit NqrA; AAA+-type ATPase, SpoVK/Ycf46/Vps4 family; Ribonuclease BN; Na+-transporting NADH:ubiquinone oxidoreductase (NQR), subunit NqrB; sRNA-binding protein.

Comparative functional enrichment analyses showed that multiple functions are highly enriched in genomes from clade1, while predominantly absent in genomes from other clades mainly consisting of species living in natural habitats (Figure 3). These functions include: Membrane-bound metal-dependent

hydrolase YbcI, DUF457 family; AAA+-type ATPase, SpoVK/Ycf46/Vps4 family; Ribonuclease BN, tRNA processing enzyme; Na+-transporting NADH:ubiquinone oxidoreductase (NQR), subunit NqrA/NqrB/NqrC/NqrD/NqrE/NqrF; sRNA-binding protein and a few others (details are listed in Table S6). The results have been verified with the 155 genomes for ANI analysis representing 95 valid *Pseudomonas* species, and all the functions were also highly enriched in genomes from clade1 in this dataset, while predominantly missing in genomes from other clades (Table S7).

It has been shown that the major function of ribonuclease BN is to control the level of certain sRNAs (e.g. 6S RNA) that may act as important global regulators of transcription and play a critical role in bacterial stress responses (Bechhofer & Deutscher, 2019; Chen et al., 2016). Membrane-bound metal-dependent hydrolase and AAA+-type ATPase are involved in the regulation of metabolism, and have been reported to contribute to metabolic adaptation to physiological host conditions and pathogenicity (Mougous et al., 2006; Parisien et al., 2008; Vermassen et al., 2019; Wyckoff et al., 2012). Furthermore, it has been reported that small RNA binding protein might contribute to virulence controlling (Djapgne et al., 2018; Law et al., 2019; Wang et al., 2020). Interestingly, it has been recently reported that *P. aeruginosa* NADH:ubiquinone oxidoreductase (NQR) is actually a completely new form of proton pump, which plays an important role in the adaptation against autotoxicity during the infection process (Raba et al., 2018).

In general, these functions are involved in bacteria stress responses, making it easier for the bacteria to get adjusted to the host conditions. The results suggest the presence of these functions might be characteristic of *Pseudomonas* species with potential of becoming human pathogen.

4. Homogeneity analysis unveils geometrically conserved genes with diversified functions

In addition to gain-of-function or loss-of-function, diversity and evolution can also be driven by geometrically highly conserved genes, which are conserved across most genomes but have lots of room to diversify in a way that impacts their functions (Sabi & Tuller, 2015; Zou & Zhang, 2019). It would be informative to investigate what kind of conserved genes in *Pseudomonas* play an important role in driving diversity.

Therefore, with the dataset of 20 representative genomes defined for the comparative functional enrichment analysis (Table S5)**,** we have performed a pan-genome based functional homogeneity analysis to detect genes that are present in all the genomes but with a low functional homogeneity index. Functional homogeneity analysis takes aligned residues into account to quantify differences across residues in a site by evaluating the biochemical properties of different residues that may affect the functional conservation of the proteins (see methods part for more details). Figure 4 shows the pan-genome of the 20 representative genomes, with each genome visualized by one layer of circle. In the pan-genome, the clustering of the genome layers is based on presence/absence of orthologous gene

family using Euclidean distance and Ward Linkage. Consistent to other population structure analyses, the genomes from each clade were grouped together. Functional homogeneity index is visualized in one of the extra layers in dark green (layer name: Func. Homeogeneity Ind.). The value of functional homogeneity index for single-copy core genes varied between 0.642 and 0.999 (mean, 0.877).

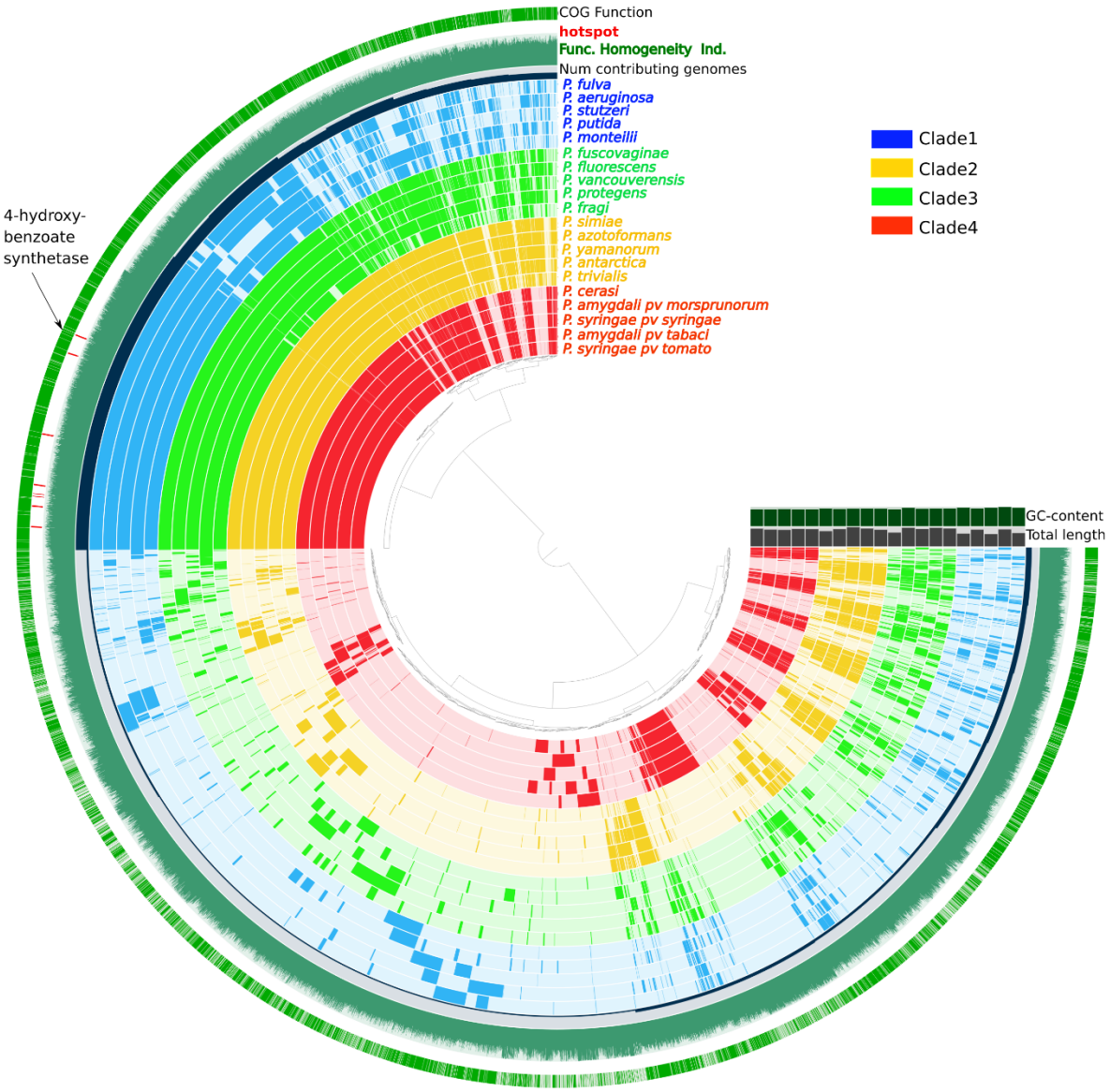

Fig. 4 Pan-genome of 20 representative *Pseudomonas* genomes. One layer represents one genome, with a few extra layers showing distinct pan-genome features. The color of each genome layer (and layer label) is indicative of the ANI based clade that the respective species fall in. Num of contributing genomes, the number of genomes each orthologous gene family identified from (range 1-20). Func. Homeogeneity Ind., functional homeogeneity index. The layer "hotspot" indicates the exact positions of geometrically conserved genes with diversified functions, and one representative gene 4-hydroxybenzoate synthetase was labelled.

We have discovered only eight genes that are geometrically conserved across all the genomes but show functional homogeneity < 0.7. These genes encode proteins including D-hexose-6-phosphate mutarotase, sulfur transfer complex TusBCD TusB component, 4-hydroxybenzoate synthetase,

Molybdopterin-guanine dinucleotide biosynthesis protein A, Flagellar hook-length control protein FliK and a few others (details are listed in Table S8). The exact positions of these genes are marked in the layer of "hotspot", with one representative gene encoding 4-hydroxybenzoate synthetase labelled in the pan-genome (Figure 4). We have further verified the results with the 155 genomes for ANI analysis representing 95 valid *Pseudomonas* species, and most of the genes have similar functional homeogeneity index in this dataset (Table S9).

Among these proteins, D-hexose-6-phosphate mutarotase, sulfur transfer complex TusBCD TusB component, 4-hydroxybenzoate synthetase and Molybdopterin-guanine dinucleotide biosynthesis protein A are involved in resource usage with broad substrate specificity, and Flagellar hook-length control protein FliK is implicated in regulating bacterial motility. These functions typically play an important role in environmental adaptation of the bacteria.

Discussion

Pseudomonas spp. are highly diverse, and responsible for many severe clinical infection cases as well as plant diseases. Despite the medical and agricultural importance, the understanding about *Pseudomonas* genome characteristics and intrageneric population structure is limited. In this study, we have explored the pan-genome characteristics and genome-wide population structure of 704 *Pseudomonas* genomes. As far as we know, this is the first study addressing the *Pseudomonas* genome varieties completely based on whole-genome sequence information.

Pan-genome analysis of the 704 *Pseudomonas* genomes has revealed a small core genome (representing 1.65% of the pan-genome) and a very large accessory genome (98.35% of the pan-genome), indicating a high level of genomic diversity. Since in this dataset there are some species that have not been validly published yet, to verify the data quality, we have also checked the core-genome size (in 99% ≤ strains ≤100%) of the 155 genomes for ANI analysis representing 95 valid *Pseudomonas* species. The core-genome of this dataset constitutes 2.39% of the pan-genome, which is only slightly bigger than the core-genome size (1.65% of the pan-genome) of the complete 704 genomes. Since these 155 genomes are from validly published species, this further confirmed the high-level of diversity of the *Pseudomonas* genomes.

Population structure analysis based on whole-genome similarity index ANI grouped the *Pseudomonas* strains in four clades, which was reinforced by *Pseudomonas* phylogeny established with concatenated alignment of single-copy core genes of 704 genomes and by relatedness of orthologous genes between 20 representative strains, as displayed in Figure 1, 2&4. Interestingly, most known opportunistic human pathogens, such as *P. aeruginosa, P. stutzeri and P. putida*, are closely related and fall into one clade, while commonly known plant pathogens of *P. syringae* strains are also found to form one distinct clade.

This population structure is congruent with those from two *Pseudomonas* phylogeny analyses in which a large number of strains representing many species were evaluated by either four concatenated genes (16S rRNA, gyrB, rpoB, and rpoD) or MLSA (Gomila et al., 2015; Mulet et al., 2010). In both studies, *Pseudomonas* species that may act as opportunistic human pathogens or plant pathogens are mostly separated from other non-pathogenic *Pseudomonas* species such as *P. fluorescens, P. fragi, and P. antarctica*. Furthermore, the population structure based on *Pseudomonas* phylogeny established with concatenated core-gene alignment is largely consistent with the population structure based on ANI. This is in line with the most recent investigation about *Pseudomonas* phylogeny (Gomila et al., 2015) reporting that ANI results show very similar topology as phylogeny analysis based on MLSA.

Moreover, in this study, we have investigated host-related functional elements. Although many *Pseudomonas* stains may live in tremendously different environments, it is still poorly understood why some *Pseudomonas* spp. may colonize human or use animals as host, while some species only live in environment or live with plants. With a combination of population structure analysis and comparative functional enrichment analysis, we have identified functions that are predominantly present in human (or animal) pathogen species, while largely lacking in species only living in natural habitats. Most of these functions have been reported to contribute to bacterial stress response. Bacterial stress responses may help bacteria survive and grow in hostile environments, and potentially trigger diseases in the hosts. Bacteria that inhabit or infect human organs such as gut or lung are exposed to various stresses, such as changes in pH, nutrient structure, the attack from the host immune system, and the presence of toxic compounds produced by other bacteria or themselves (Fang et al., 2016; Tattoli et al., 2012). The interaction between pathogenic bacteria and host immune cells may stimulate complicated responses in bacteria including rapid adjustment of gene expression (Gunderson & Seifert, 2015; Sansonetti & Di Santo, 2007). It is not surprising that functions preferably enriched in human opportunistic pathogens are mostly implicated in stress responses. These functions may provide the microbes a better chance to survive harsh challenges during infection and colonization, and the possession of these functions might be the key attribute of those *Pseudomonas* species with potential of becoming human pathogen.

From the other aspect, these functions can be suitable targets for the development of new antibiotics. Most *Pseudomonas* species that may act as human pathogens are resistant to multiple antibiotics, which makes it extremely difficult to get rid of the infections they have caused (Langdon et al., 2016). For example, controlling chronic *P. aeruginosa* airway infection in cystic fibrosis patients is still one of the most challenging tasks for cystic fibrosis treatment (Boutin et al., 2017). These functions we revealed through this study might be able to promote the development of more targeted antibiotics.

In addition to newly achieved genetic features, bacterial diversity can also be driven by geometrically conserved but functionally diversified genes. Understanding the homogeneity/heterogeneity within one orthologous gene family could yield detailed ecological or evolutionary insights with respect to the forces that shape the genomic context across closely related taxa (Andrés et al., 2007; Cao et al., 2019;

Delmont & Eren, 2018). Therefore, we have further performed functional homogeneity analysis, and detected multiple genes with diversified functions but still under structural constraints showing homogeneous sequence alignment across the genomes. The proteins encoded by these genes are typically involved in resource usage or bacterial motility, and play a critical role in environmental adaptation. This is consistent with the assumption that these genes are also important genetic elements driving diversity.

The information we achieved from this study added to our understanding about the genome varieties that generate the tremendous diversity in *Pseudomonas*, and provided insights into host-related functional determinants, which might be beneficial for the development of *Pseudomonas*-targeted antibiotics.

Materials and methods

Dataset

The analyzed data set consisted of 704 genomes of *Pseudomonas* strains with complete genome or chromosome available in January 2020 in the RefSeq database of the National Center for Biotechnology Information (NCBI)**.** Accession numbers and other detailed information about the genomes are shown in Supplementary Table S1.

Pangenome analysis

Pangenome analysis of *Pseudomonas* was done in anvi'o (version 6.1) (Eren et al., 2015) following the workflow for microbial pangenomics (Delmont & Eren, 2018; Delmont et al., 2019). Prodigal (Hyatt et al., 2010) was used to identify open reading frames, then each database was populated with profile hidden Markov models (HMMs) by comparing to a collection of single-copy genes using HMMER (Eddy, 2011). We used the Diamond software to calculate gene similarity and MCL (Van Dongen & Abreu-Goodger, 2012) for clustering under the following settings: minbit, 0.5; mcl inflation, 2; and minimum occurrence, 2.

Population structure analysis

Whole-genome comparisons by ANI (average nucleotide identity) was calculated for blast based ANI (ANIb) and MUMmer based ANI (ANIm) with PYANI (Pritchard et al., 2016). For phylogenomic analysis of *Pseudomonas* genomes, we used the sequence information of single-copy core genes present in 100% *Pseudomonas* stains. Sequence alignments of core genes were concatenated to give a single core alignment and a maximum likelihood phylogeny was then generated using the program IQ-TREE (Nguyen et al., 2015) with a time-reversible amino-acid substitution model WAG (Whelan & Goldman,

2001), four gamma categories for rate heterogeneity and 1000 bootstrap replicates. All phylogenies were visualized using the Interactive Tree of Life (Letunic & Bork, 2019).

Comparative functional enrichment analysis

All genomes were first annotated with the functional annotation source COGs (Tatusov, 2000). Then functional enrichment test was performed using a Generalized Linear Model with the logit linkage function to compute an enrichment score and p-value for each function. False Detection Rate correction to p-values to account for multiple tests was done using the R package qvalue (Delmont & Eren, 2018).

Functional homogeneity analysis

Functional homogeneity index indicates how aligned amino acid residues are conserved across genes. To perform this analysis, first, based on amino acids' biochemical properties, all amino acids are divided into seven different "conserved groups": nonpolar, aromatic, polar uncharged, both polar and nonpolar characteristics, acidic, basic, and mostly nonpolar. Next, based on how close the biochemical properties of the amino acid residues are to each other, for every pair of amino acids at the same position in the orthologous gene, a similarity score is assigned. Functional homogeneity index is then calculated with all assigned similarity scores and will reach its maximum value of 1.0 if all residues are identical. This analysis was performed using functions integrated in anvi'o (version 6.1) (Delmont & Eren, 2018; Eren et al., 2015).

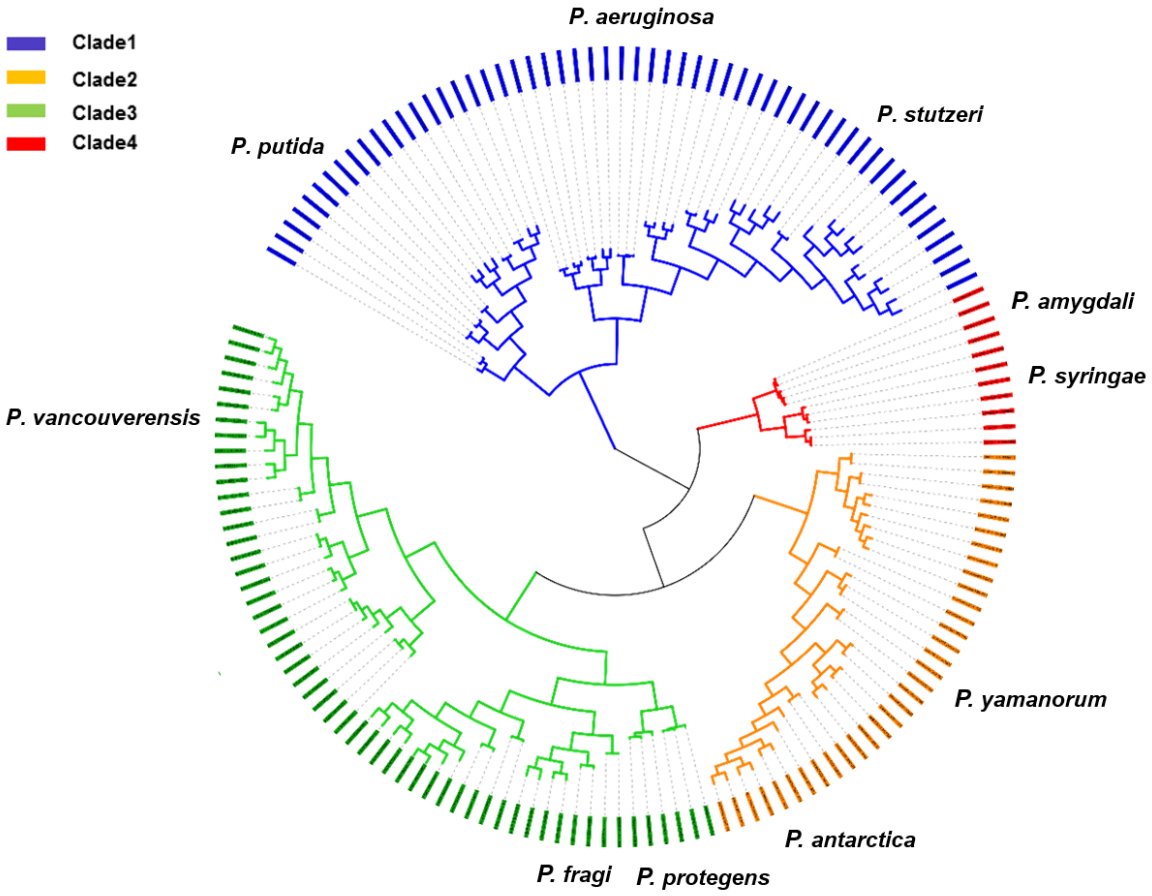

Fig. S1 Population structure of the genus *Pseudomonas* based on whole-genome comparisons (ANIm) of 155 *Pseudomonas* strains representing 95 valid *Pseudomonas* species. The calculations were performed using the Python module PYANI. The strains fall into four major Clades (Clade1-Clade4). The positions of representative species in each Clade were labelled.

References

Andrés, A. M., De Hemptinne, C., & Bertranpetit, J. (2007). Heterogeneous rate of protein evolution in serotonin genes. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msm202

Bechhofer, D. H., & Deutscher, M. P. (2019). Bacterial ribonucleases and their roles in RNA metabolism. In *Critical Reviews in Biochemistry and Molecular Biology*. https://doi.org/10.1080/10409238.2019.1651816

Boutin, S., Graeber, S. Y., Stahl, M., Dittrich, A. S., Mall, M. A., & Dalpke, A. H. (2017). Chronic but not intermittent infection with Pseudomonas aeruginosa is associated with global changes of the lung microbiome in cystic fibrosis. *European Respiratory Journal*, *50*(4). https://doi.org/10.1183/13993003.01086-2017

Cao, Y., Jiang, L., Wang, L., & Cai, Y. (2019). Evolutionary rate heterogeneity and functional divergence of orthologous genes in pyrus. *Biomolecules*. https://doi.org/10.3390/biom9090490

Chen, H., Dutta, T., & Deutscher, M. P. (2016). Growth phase-dependent variation of RNase BN/Z affects small RNAs. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M116.757450

Delmont, T. O., & Eren, E. M. (2018). Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*. https://doi.org/10.7717/peerj.4320

Delmont, T. O., Kiefl, E., Kilinc, O., Esen, O. C., Uysal, I., Rappé, M. S., Giovannoni, S., & Eren, A. M. (2019). Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *ELife*. https://doi.org/10.7554/eLife.46497

Djapgne, L., Panja, S., Brewer, L. K., Gans, J. H., Kane, M. A., Woodson, S. A., & Oglesby-Sherrouse, A. G. (2018). The Pseudomonas aeruginosa PrrF1 and PrrF2 small regulatory RNAs promote 2-alkyl-4- quinolone production through redundant regulation of the antR mRNA. *Journal of Bacteriology*. https://doi.org/10.1128/JB.00704-17

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1002195

Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B., & Gibson, R. L. (2002). Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatric Pulmonology*. https://doi.org/10.1002/ppul.10127

Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platformfor 'omics data. *PeerJ*. https://doi.org/10.7717/peerj.1319

Fang, F. C., Frawley, E. R., Tapscott, T., & Vázquez-Torres, A. (2016). Bacterial Stress Responses during Host Infection. In *Cell Host and Microbe*. https://doi.org/10.1016/j.chom.2016.07.009

Fernández, M., Porcel, M., de la Torre, J., Molina-Henares, M. A., Daddaoua, A., Llamas, M. A., Roca, A., Carriel, V., Garzón, I., Ramos, J. L., Alaminos, M., & Duque, E. (2015). Analysis of the pathogenic potential of nosocomial Pseudomonas putida strains. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2015.00871

Freschi, L., Vincent, A. T., Jeukens, J., Emond-Rheault, J. G., Kukavica-Ibrulj, I., Dupont, M. J., Charette, S. J., Boyle, B., & Levesque, R. C. (2019). The Pseudomonas aeruginosa Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evy259

Gomila, M., Peña, A., Mulet, M., Lalucat, J., & García-Valdés, E. (2015). Phylogenomics and

systematics in Pseudomonas. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2015.00214

Gross, H., & Loper, J. E. (2009). Genomics of secondary metabolite production by Pseudomonas spp. In *Natural Product Reports*. https://doi.org/10.1039/b817075b

Gunderson, C. W., & Seifert, H. S. (2015). Neisseria gonorrhoeae elicits extracellular traps in primary neutrophil culture while suppressing the oxidative burst. *MBio*. https://doi.org/10.1128/mBio.02452-14

Haas, D., & Défago, G. (2005). Biological control of soil-borne pathogens by fluorescent pseudomonads. In *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro1129

Hulin, M. T., Armitage, A. D., Vicente, J. G., Holub, E. B., Baxter, L., Bates, H. J., Mansfield, J. W., Jackson, R. W., & Harrison, R. J. (2018). Comparative genomics of Pseudomonas syringae reveals convergent gene gain and loss associated with specialization onto cherry (Prunus avium). *New Phytologist*. https://doi.org/10.1111/nph.15182

Hwang, M. S. H., Morgan, R. L., Sarkar, S. F., Wang, P. W., & Guttman, D. S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of Pseudomonas syringae. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.71.9.5182-5191.2005

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-11-119

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*. https://doi.org/10.1038/s41467-018-07641-9

Konstan, M. W., Morgan, W. J., Butler, S. M., Pasta, D. J., Craib, M. L., Silva, S. J., Stokes, D. C., Wohl, M. E. B., Wagener, J. S., Regelmann, W. E., & Johnson, C. A. (2007). Risk Factors For Rate of Decline in Forced Expiratory Volume in One Second in Children and Adolescents with Cystic Fibrosis. *Journal of Pediatrics*. https://doi.org/10.1016/j.jpeds.2007.03.006

Lalucat, J., Bennasar, A., Bosch, R., Garcia-Valdes, E., & Palleroni, N. J. (2006). Biology of Pseudomonas stutzeri. *Microbiology and Molecular Biology Reviews*. https://doi.org/10.1128/mmbr.00047-05

Langdon, A., Crook, N., & Dantas, G. (2016). The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. In *Genome Medicine*. https://doi.org/10.1186/s13073-016-0294-z

Law, C. O. K., Huang, C., Pan, Q., Lee, J., Hao, Q., Chan, T. F., Lo, N. W. S., Ang, I. L., Koon, A., Ip, M., Chan, E., & Lau, T. C. K. (2019). A Small RNA Transforms the Multidrug Resistance of Pseudomonas aeruginosa to Drug Susceptibility. *Molecular Therapy - Nucleic Acids*. https://doi.org/10.1016/j.omtn.2019.02.011

Lessie, T. G., & Phibbs, P. V. (1984). Alternative Pathways of Carbohydrate Utilization in Pseudomonads. *Annual Review of Microbiology*. https://doi.org/10.1146/annurev.mi.38.100184.002043

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz239

Loper, J. E., Hassan, K. A., Mavrodi, D. V., Davis, E. W., Lim, C. K., Shaffer, B. T., Elbourne, L. D. H., Stockwell, V. O., Hartney, S. L., Breakwell, K., Henkels, M. D., Tetu, S. G., Rangel, L. I., Kidarsa, T. A., Wilson, N. L., van de Mortel, J. E., Song, C., Blumhagen, R., Radune, D., … Paulsen, I. T.

(2012). Comparative genomics of plant-associated pseudomonas spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002784

Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., Van Der Voort, M., Schneider, J. H. M., Piceno, Y. M., DeSantis, T. Z., Andersen, G. L., Bakker, P. A. H. M., & Raaijmakers, J. M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*. https://doi.org/10.1126/science.1203980

Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordoñez, C. L., Lory, S., Walz, T., Joachimiak, A., & Mekalanos, J. J. (2006). A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science*. https://doi.org/10.1126/science.1128393

Mulet, M., Lalucat, J., & García-Valdés, E. (2010). DNA sequence-based analysis of the Pseudomonas species. *Environmental Microbiology*. https://doi.org/10.1111/j.1462-2920.2010.02181.x

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msu300

Nixon, G. M., Armstrong, D. S., Carzino, R., Carlin, J. B., Olinsky, A., Robertson, C. F., & Grimwood, K. (2001). Clinical outcome after early Pseudomonas aeruginosa infection in cystic fibrosis. *Journal of Pediatrics*. https://doi.org/10.1067/mpd.2001.112897

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv421

Parisien, A., Allain, B., Zhang, J., Mandeville, R., & Lan, C. Q. (2008). Novel alternatives to antibiotics: Bacteriophages, bacterial cell wall hydrolases, and antimicrobial peptides. In *Journal of Applied Microbiology*. https://doi.org/10.1111/j.1365-2672.2007.03498.x

Park, C. J., & Andam, C. P. (2019). Within-species genomic variation and variable patterns of recombination in the tetracycline producer Streptomyces rimosus. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2019.00552

Park, C. J., & Andam, C. P. (2020). Distinct but Intertwined Evolutionary Histories of Multiple Salmonella enterica Subspecies . *MSystems*. https://doi.org/10.1128/msystems.00515-19

Parkinson, N., Bryant, R., Bew, J., & Elphinstone, J. (2011). Rapid phylogenetic identification of members of the Pseudomonas syringae species complex using the rpoD locus. *Plant Pathology*. https://doi.org/10.1111/j.1365-3059.2010.02366.x

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. In *Analytical Methods*. https://doi.org/10.1039/c5ay02550h

Raba, D. A., Rosas-Lemus, M., Menzer, W. M., Li, C., Fang, X., Liang, P., Tuz, K., Minh, D. D. L., & Juárez, O. (2018). Characterization of the Pseudomonas aeruginosa NQR complex, a bacterial proton pump with roles in autopoisoning resistance. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.RA118.003194

Sabi, R., & Tuller, T. (2015). A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics*. https://doi.org/10.1186/1471-2164-16-S10-S5

Sansonetti, P. J., & Di Santo, J. P. (2007). Debugging how Bacteria Manipulate the Immune Response. In *Immunity*. https://doi.org/10.1016/j.immuni.2007.02.004

Tattoli, I., Sorbara, M. T., Vuckovic, D., Ling, A., Soares, F., Carneiro, L. A. M., Yang, C., Emili, A., Philpott, D. J., & Girardin, S. E. (2012). Amino acid starvation induced by invasive bacterial pathogens triggers an innate host defense program. *Cell Host and Microbe*. https://doi.org/10.1016/j.chom.2012.04.012

Tatusov, R. L. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*. https://doi.org/10.1093/nar/28.1.33

Van Dongen, S., & Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-361-5_15

Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M., & Desvaux, M. (2019). Cell wall hydrolases in bacteria: Insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. In *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2019.00331

Wang, Y. L., Sun, L. J., Xian, C. M., Kou, F. L., Zhu, Y., He, L. Y., & Sheng, X. F. (2020). Interactions between Biotite and the Mineral-Weathering Bacterium Pseudomonas azotoformans F77. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.02568-19

Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a003851

Wiehlmann, L., Wagner, G., Cramer, N., Siebert, B., Gudowius, P., Morales, G., Köhler, T., Van Delden, C., Weinel, C., Slickers, P., & Tümmler, B. (2007). Population structure of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0609213104

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., & van der Lelie, D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in Pseudomonas putida. In *FEMS Microbiology Reviews*. https://doi.org/10.1111/j.1574-6976.2010.00249.x

Wyckoff, T. J., Taylor, J. A., & Salama, N. R. (2012). Beyond growth: Novel functions for bacterial cell wall hydrolases. In *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2012.08.003

Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A., & Harayama, S. (2000). Phylogeny of the genus Pseudomonas: Intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. *Microbiology*. https://doi.org/10.1099/00221287-146-10-2385

Zou, Z., & Zhang, J. (2019). Amino acid exchangeabilities vary across the tree of life. *Science Advances*. https://doi.org/10.1126/sciadv.aax3124