Phylogeny, pangenomics, and predicted functional diversity of maize rhizosphere *Pseudomonas*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Jessica Lynn Sutter

aus Silver Spring/Vereinigte Staaten von Amerika

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	22.09.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Prof. Dr. Ruth E. Ley
2. Berichterstatterin:	Prof. Dr. Nadine Ziemert

To my Mom, my Grandma, my Brother, and of course my whole family for their endless support and encouragement

and to my beautiful Maggie:
Well, a good dog on the ground's worth three in the saddle
No matter where you're from
Been many good dog was a friend to a man
But you were the greatest one

Acknowledgements

I first want to give my sincerest thanks to Ruth Ley for this opportunity to grow as a scientist, back at Cornell and now here at the MPI for Biology. After ten years she can finally scrape this crusty old barnacle from her hull. Next I want to thank Nick Youngblut and Tony Walters for all their instruction, suggestions, and excellent advice despite my seeming inability to take it sometimes. I also want to thank my committee members past and present: Detlef Weigel, Nadine Zeimert, Talia Karasov, Dan Buckley, Gregory Martin, and of course Eugene Madsen.

I'm eternally grateful for all the help, encouragement, and camaraderie from the members of the Department of Microbiome Sciences here in Tübingen: Daphne Welter, Zach Henesler, Jacobo de la Cuesta for without whom my figures would be a great deal uglier, Tanja Schön for translating my abstract and otherwise being a very kind and helpful person, Andrea Borbón and Leonardo Moreno for being wonderful friends and colleagues, Liam Fitzstevens, Albane Ruaud, Guillermo Luque, Yihua Liu, Michael Bell, Mirabeau Mbong Ngwese, Xiaoying Liu for all those excellent discussions of phylogeny, Hagay Enav and his big scientist dad energy, Kelsey Huus, Taichi Suzuki, Claudia Mirretta Barone, Jillian Waters, Sara Clasen, Boram Seo, Victor Schmidt, Carolin Kolmender, and and all the technicians and staff who so patiently endured my flask after flask of stinky *Pseudomonas* cultures– Athina Iliopoulou, Katharina Vernali, Silke Dauser, Sophie Maisch, Iris Holdermann, Ursula Schach, and a hearty "good luck" to all the new technicians & students that started recently of whom I haven't met in person. I also want to give a massive shoutout to James Marsh for spiking my bloomer into his PacBio run and saving my pangenome, and of course Karin Klein for helping me navigate not only the institute but Germany as a whole.

Rewinding even further back I want to thank everyone from Cornell who supported and inspired me to even attempt a PhD at all: Julia Goodrich, Angela Poole, Sara di Rienzi, Shao-Pei Chou, Elizabeth Johnson, Wei Zhang, Sha Li, Omry Koren, Joe Usack, Catherine Spirito, Clara Cho, Noah Clark, Tim DeMarsh for being a great coworker and for caring for Jermagisty in their final months, Beth Bell, Qiaojuan Shi, and Zhao Jin for setting me up with a project she was lucky to escape.

Here at the institute I couldn't have accomplished anything without the assistance of so many people. At the Genome Center: Julia Hildebrandt, Katrin Fritschi, and Christa Lanz who trained me on the HiSeq and graciously didn't murder me when I forgot to lower the sipper handle that one time. All those at the Tübingen International PhD Program: Jeanette Müller, Sibylle Patheiger, George Deffner, and Susan Jones especially for helping me over the finish line. All the administrators that kept ahead of what I needed to be here at all: Christian Neff, Sarah Kellner, and Brigitte Walderich. Without Birgit Moldovan MPQueer would never have been open to our campus– she was a profoundly kind and sincere person. Her passing was a terrible loss.

From Department 6 many individuals assisted me over the past 6 years: Derek Lundberg, Or Shalev-Scriptchak, Alejandra Duque, Haim Ashkenazy, Ilja Bezrukov, Bridgit Waithaka, Moisés Expósito Alonso, Dino Jolic and his ungodly ability to eat ice cream, Sergio Latorre, Grey Monroe and his absolutely spot-on Bernie Sanders impression, Manuela Neumann, Julian Regalado, Clemens Weiss, and Fernando Rabanal. My deepest appreciation to Rebecca Schwab for making the garden a reality, despite my near complete inability to grow anything besides a sunflower.

I also want to thank Ziduan Han for being a generous worm steward, Athina Gavriilidou patiently helping wrest my data from SQL, Melanie Kirch for her sympathetic ear, and Marek Kučka for his magic beads. To Honour McCann I want to give my sincerest appreciation for her mentorship regardless of how brief, to Daria Evseeva for motivating us all to get outside, and to Andrea Sajuthi for being a good sport when Maggie dropped by unannounced. Speaking of Maggie, I have to commend Joachim Sieler for taking the time to win her over and allowing her to be the official IT bretzel inspector– and of course Andre Noll who always gave me more space in my project folder immediately after knocking me down a peg (as I rightly deserved).

Lastly, I want to acknowledge those German and German-adjacent folks who have been wonderful friends and neighbors: again, Michael Werner and Talia Karasov, Stacey Heaver, and of course Florian and Christiane Emmerich. I must also express my appreciation for Kavita Venkataramani, Isabella Casini, Effie Symeonidi, Claudi Friedemann, Akanksha Mishra, Brandon Seah, Aditi Singh, Minakshi Singh, and Agnes Henschen for all being lovely people to spend time with. Also to Christian Kubica and Alexander Ringwald for inciting my love of miniature painting, and to Erica McGale, Barbara Safaric, JD Rolfes, and Frauke Logermann for all their work on MPQueer. Then there's Wiebke Mollik, my unicorn of a roommate who managed to survive not only me but my pets as well. I've been incredibly fortunate.

Table of Contents

Abstract

Plants, like all multicellular organisms on this planet, exist as a holobiont. Their underground tissues are often anchored in soil, which in itself is a very microbe-rich habitat. From this vast reservoir of microbial diversity the plant acquires an essential community of bacteria, archaea, fungi, and microscopic eukaryotes that comprises their root microbiome, or rhizosphere. This relationship has existed for hundreds of millions of years; it involves a complex food web with the host as a source of nutrients and the microbes as both competitors for resources and intermediaries between the host and the environment. Essential interactions are often modulated by abiotic factors such as moisture, temperature, pH, and inorganic mineral availability. The symbioses formed run the gamut from mutualistic to parasitic, with the vast majority of microbial species existing in a commensal relationship with the plant.

To better understand the genetic mechanism by which the plant selects for certain bacterial taxa a study was conducted to profile these communities in maize, and to determine which genera or species of bacteria appeared to be heritable. 16S rRNA sequencing of several thousand maize rhizospheres revealed a possible heritability of some taxa and a striking period of increased abundance of the genera *Pseudomonas*. This phenomena could not be explained as contamination or as a sequencing artifact. At genus level it was impossible to relate phylogeny or function to this abundance phenotype due to *Pseudomonas* being highly diverse and containing thousands of individual species, which range from beneficial to pathogenic.

The aims of this work are first to identify these blooming *Pseudomonas* at species level, next to ascertain their phylogenetic relationships, then to perform *in silico* predictions based on gene content to infer the functional potential of these taxa. Furthermore, to use metagenomic sequencing to assemble the genomes of the

Pseudomonas strains endemic to these rhizospheres, as well as profile other phyla that are present within this environment and are known to interact with *Pseudomonas.*

By whole-genome shotgun sequencing of each rhizosphere, species level resolution was achieved for all bacterial and eukaryotic taxa. This revealed the *Pseudomonas* population consisted of 394 species with a high degree of phylogenetic relatedness. The most abundant of these were *P. brassicacearum, P. frederiksbergensis E,* and *P. silesiensis,* with the proportions of each varying significantly by field. A threshold for biological relevance in *Pseudomonas* relative abundance was set and representative genomes for each species were used to construct a phylogeny based on single copy core genes. With this, monophyletic groups were identified along with subgroups named based on literature type strains. These groups were categorized as belonging to either the fluorescens lineage or non-fluorescens lineage, with the former dominating the rhizosphere in both number of species and their combined relative abundance.

A pan-genome analysis of these *Pseudomonas* species revealed that beta diversity metrics such as Bray-Curtis dissimilarity can distinguish phylogenetic group membership based on whole genomes. Phenotype predictions indicated that fluorescens lineage groups are more likely to utilize certain carbohydrates such as arabinose, trehalose, and maltose, while high abundance species of this lineage appear to be enriched in genes related to nitrogen reduction. Antimicrobial and metal resistance gene content appears to be highly conserved across all phylogenetic groups, with virulence gene content being the most discriminatory between both groups and subgroups. Unlike the resistome, secondary metabolite/biosynthetic gene cluster content was highly unique between individual species, with the vast majority of these gene clusters appearing in fewer than ten genomes. The scant number that were consistent between groups were largely β-lactone antibiotics and bacteriocins.

These results indicate both core and accessory genes can be utilized in discriminating between fluorescens lineage *Pseudomonas* independent of taxonomy. The resistome is highly conserved between groups but secondary metabolites are largely unique to individual species. This suggests rhizosphere *Pseudomonas* may be under pressure to maintain consistent defenses involving specific antibiotic and multidrug efflux operons, while concurrently evolving or acquiring novel offensive secondary metabolites to maintain competitiveness between closely related species.

Zusammenfassung

Wie alle mehrzelligen Organismen auf diesem Planeten, sind Pflanzen Holobionten. Meistens sind ihre unterirdischen Gewebe im Boden, ein mikrobenreicher Lebensraum bereits an sich, verankert. Aus diesem immensen Reservoir mikrobieller Vielfalt erwirbt die Pflanze eine wesentliche Gemeinschaft von Bakterien, Archaeen, Pilzen und mikroskopischen Eukaryoten, die ihr Wurzelmikrobiom, auch Rhizosphäre genannt, bilden. Diese Form der Lebensgemeinschaft existiert bereits seit Hunderten von Millionen Jahren; sie umfasst ein komplexes Nahrungsnetz mit dem Wirt als Nährstoffquelle und den Mikroben als Konkurrenten um Ressourcen und Vermittler zwischen dem Wirt und der Umwelt. Die wesentlichen Interaktionen werden häufig durch abiotische Faktoren wie Feuchtigkeit, Temperatur, pH-Wert und die Verfügbarkeit anorganischer Mineralien beeinflusst. Diese Symbioseformen reichen von Mutualismus bis Parasitismus, wobei die überwiegende Mehrheit der Mikrobenarten in einer kommensalen Beziehung mit der Pflanze lebt.

Um den genetischen Mechanismus der Auswahl bestimmter Bakterientaxa durch die Pflanze näher zu untersuchen, wurde eine Studie wurde eine Studie durchgeführt, um ein Profil dieser Gemeinschaften in Mais zu erstellen und festzustellen, welche Gattungen oder Arten von Bakterien vererbbar zu sein scheinen. Anhand der 16s rRNA Sequenzierung mehrerer tausend Maisrhizosphären zeigte sich, dass einige Taxa vererbbar zu sein schienen, interessanterweise gab es eine auffällige zeitlich begrenzte Zunahme der Gattung Pseudomonas-Abundanz in fast allen Proben. Bei diesm Phänomen handelte es sich weder um eine Kontamination der Proben noch um Sequenzierungsartefakte. Diese Abundanz konnte auf der Gattungsebene der Mikroben nicht mit ihrer Phylogenie oder ihrerer Funktionen in Verbindung gebracht werden, da Pseudomonas äußerst divers und artenreich ist, wobei die Arten von Nützlingen bis Pathogenen reichen.

Die Ziele dieser Arbeit umfasst die Identifizierung die Arten dieser blühenden Pseudomonas, die Ermittlung ihrer phylogenetischen Verwandtschaft und die Durchführung von In-silico-Vorhersagen ihres funktionellen Potenzials auf der Grundlage ihres jeweiligen Geninhalts. Darüber hinaus soll mittels metagenomischer Sequenzierung die Genome der Rhizosphären-endemischen Pseudomonas-Stämme zusammengesetzt werden und andere Phyla zu charackterisieren, die in dieser Umgebung vorkommen und von denen bekannt ist, dass sie mit Pseudomonas interagieren.

Durch die Shotgun-Sequenzierung des Metagenomes jeder Rhizosphäre wurde eine Auflösung auf Artniveau für alle bakteriellen und eukaryotischen Taxa erreicht. Dabei zeigte sich, dass die Pseudomonas-Population tatsächlich aus 394 verschiedenen Arten mit einem hohen Grad an phylogenetischer Verwandtschaft bestand. Die am häufigsten vorkommenden Arten waren *P. brassicacearum, P. frederiksbergensis E* und *P. silesiensis*, wobei der Anteil der einzelnen Arten je nach Feld erheblich variierte. Ein

Schwellenwert für die relative Häufigkeit von Pseudomonas wurde für die biologische Relevanz festgelegt, und aus repräsentativen Genomen jeder Art wurde eine Phylogenie auf der Grundlage von Kerngenen mit einer einfachen Kopie erstellt. Auf diese Weise wurden monophyletische Gruppen und Untergruppen identifiziert, benannt auf Grundlage Literatur beschriebener Stämme. Kategorisiert wurden diese Gruppen als Fluorescens-Linie oder Nicht-Fluorescens-Linie, wobei erstere die Rhizosphäre sowohl in Bezug auf die Anzahl der Arten als auch deren kombinierte relative Häufigkeit dominierte.
Eine Pan-Genomanalyse dieser Pseudomonas-Arten ergab, dass phylogenetische Gruppenzugehörigkeit auf der Grundlage ganzer Genome mittels Beta-Diversitätsmetriken, wie die Bray-Curtis-Dissimilarität, unterschieden werden können. Phänotypische Vorhersagen deuten darauf hin, dass Gruppen der Fluorescens-Linie gewisse Kohlenhydrate wie Arabinose, Trehalose und Maltose bevorzugt verwerten, während häufig vorkommende Arten dieser Linie offenbar vermehrt Gene für die Stickstoffreduktion aufweisen. Der Gehalt an Metall- und antimikrobiellen Resistenzgenen war über alle phylogenetischen Gruppen hinweg konserviert, wohingegen der Gehalt an Virulenzgenen den größten Unterschied zwischen beiden Linien und Untergruppen ausmachte. Im Gegensatz zum Resistom war der Gehalt an Sekundärstoffwechsel-/Biosynthese-Geneclustern bei den einzelnen Arten sehr unterschiedlich, wobei die überwiegende Mehrheit dieser Gencluster in weniger als zehn Genomen vorkam. Bei der geringen Anzahl, die zwischen den Gruppen übereinstimmte, handelte es sich größtenteils um Beta-Lakton-Antibiotika und Bakteriozine.

Diese Ergebnisse zeigen, dass sowohl Kern- als auch akzessorische Gene zur Unterscheidung zwischen Pseudomonas der Fluorescens-Linie unabhängig von der Taxonomie verwendet werden können. Das Resistom ist zwischen den Gruppen sehr konserviert, wobei die Sekundärmetaboliten weitgehend einzigartig für die einzelnen Arten sind. Dies lässt vermuten, dass Pseudomonas in der Rhizosphäre unter dem Druck stehen, eine konsistente Abwehr aufrechtzuerhalten, die spezifische Antibiotika- und Multidrug-Efflux-Operons umfasst, während sie gleichzeitig neue offensive Sekundärmetaboliten entwickeln oder erwerben, um die Wettbewerbsfähigkeit zwischen eng verwandten Arten zu erhalten.

Contributions

James Marsh prepared and generated the long reads for the isolate URIL14HWK12:I7. William (Tony) Walters designed a custom script for the generalized linear mixed model. Daphne K. Welter, Zachariah Henseler, and Jacobo de la Cuesta generated custom scripts to automate some data processing and figure generation. Zhao Jin and Jason Peiffer collected and isolated the 2010 *Pseudomonas* strains from Urbana IL and Lansing NY.

List of figures

Main figures

Supplemental figures

List of tables

Main tables

Supplemental tables

Abbreviations

AMR - antimicrobial resistance
B73 - inbred accession of stiff stalk *Zea mays* subsp. *mays* var. B73
BacMet2 - Antibacterial Biocide and Metal Resistance Genes Database
BGC - biosynthetic gene cluster
CARD - The Comprehensive Antibiotics Resistance Database
GC - nucleotides guanine and cytosine
GCF - gene cluster family
gDNA - genomic DNA
GTDB - Genome Taxonomy Database
HGT - horizontal gene transfer
IL14H - inbred accession of sweet *Zea mays*
LPS - lipopolysaccharide
MAG - metagenomically-assembled genome
MAMP - microbe-associated molecular pattern
MCL - markov cluster
MHB - mycorrhizal helper bacteria
ML - maximum likelihood method
MLSA - multilocus sequence analysis
MO17 - inbred accession of non-stiff stalk *Zea mays*
NAGGN - N-acetylglutaminylglutamine amide
NF - non-fluorescens lineage *Pseudomonas*
NGS - next generation sequencing
NRP - nonribosomal peptide
PAMP - pathogen-associated molecular patterns
PCoA - principal coordinate analysis
PC - principal coordinate
PGPR - plant growth promoting rhizobacteria
Resistome - the sum of all antibiotic resistance gene content
RiPP - ribosomally synthesized and post-translationally modified peptides
SCG - single copy gene
SG1 - fluorescens lineage supergroup 1 *Pseudomonas*
SG2 - fluorescens lineage supergroup 2 *Pseudomonas*
SG3 - fluorescens lineage supergroup 3 *Pseudomonas*
T4P - type IV pili
T6SS - type VI secretion system
tRNA - transfer RNA
V1 - first principal coordinate (X-axis)
V2 - second principal coordinate (Y-axis)
VFDB - Virulence Factor Database

Introduction

Similar to the manner in which higher-order vertebrates have complex communities of bacteria and archaea inside their gut and on the exterior surface of their bodies, plants have co-evolved intimate associations with their own microbiota both above and below ground. As with the study of human microbiota, the attempt to determine which of these microbes are heritable and to what degree has driven many experiments using diverse plant hosts including arabidopsis and sorghum (Deng *et al.*, 2021). To address this question during the summer of 2010 an ambitious field study was conducted using 27 inbred lines of maize planted in two Midwest field sites, plus three additional locations across the Finger Lakes region of Upstate New York, to assess 16S rRNA gene diversity within the rhizosphere (Peiffer *et al.*, 2013; Walters *et al.*, 2018).

This longitudinal study which totaled 4,866 samples was informative in many regards; foremost, it revealed that the variation in relative abundance among the majority of "core" operational taxonomic units (OTUs) shared by all samples was under modest but significant influence by the plant genotype. This study also demonstrated the strong effect of plant age amid the inherent local diversity of microbes supplied at each geographic location. Overall, rhizosphere microbiome richness (alpha, or within sample taxonomic diversity) varied between maize types. These phenotypic categories of maize included stiff stalk, non stiff stalk, tropical, sweet, popcorn, and mixed. When comparing the taxonomic diversity between samples, the field effect resulted in strong discrimination between samples using both weighted (lineage abundance) and unweighted (taxa shared) UniFrac distances.

Among the many observations of this study, one conspicuous result was the predominance of Proteobacteria within these rhizospheres. This is not unusual, as this phyla is well understood to be highly abundant within many agriculturally relevant plant cultivars including maize, tobacco, cucumber, cotton, rice, and wheat (García-Salamanca *et al.*, 2013; Tian and Gao, 2014; Breidenbach, Pump and Dumont, 2015; Saleem, Law and Moe, 2016; Rossmann *et al.*, 2020; Shi *et al.*, 2020). However, the 16S rRNA gene sequencing for the 2010 maize study indicated that, beginning at week 8 after planting, the relative abundance of the Proteobacteria genus *Pseudomonas* increased within all fields from ~3% to ~45% for the duration of the season. At the time this event was referred to as the "*Pseudomonas* bloom", and of the 16S rRNA gene sequences classified as *Pseudomonas,* more than 90% of those comprised just three OTUs.

The "core" OTUs shared by 100% of all rhizosphere samples were found to be exclusively Proteobacteria; and unsurprisingly among them was the genus *Pseudomonas.* The influence of host genetics on the variance seen in OTU abundances across each sample was calculated, and 143 OTUs were identified as significantly heritable. Unlike the core OTUs, the heritable OTUs were taxonomically diverse and included Alpha-, Beta-, and Gammaproteobacteria, Actinobacteria, Verrucomicrobia, Bacteroidetes, Planctomycetes, Fimicutes, Chloroflexi, Acidobacteria,

Gemmatimonadetes, and the Archaea Nitrososphaera. Despite the prevalence of *Pseudomonas* and its ubiquity across every rhizosphere, aside from the family Nevskiaceae, the only other Gammaproteobacteria OTU found to be heritable was *Pseudomonas viridiflava.*

An attempt was made five years later in 2015 to capture this blooming effect using a selection of maize genotypes at one of the same field plots as before. Despite careful adherence to the sampling protocol, in addition to improved metadata collection concerning temperature, precipitation, soil moisture, and plant health, *Pseudomonas* did not achieve the high relative abundance observed in 2010. However, rhizosphere samples were retained for isolate culturing and bulk soil was collected later that year for *in vitro* experimentation. The soil was used to perform a substrate-induced growth assay where glucose, fructose, glutamic acid, and salicylic acid were applied individually to rested bulk soil microcosms to stimulate *Pseudomonas* growth. After 12, 24, and 48 hour incubations the microcosm soils were harvested for metagenomic sequencing and *Pseudomonas* isolation.

The previous finding that *Pseudomonas* is core to the maize rhizosphere microbiome, and these taxa maintain high abundance despite field and host genotype raises questions that are central to the aims of this thesis. The first dimension that will be explored is the species-level diversity and phylogeny of these rhizosphere *Pseudomonas*, and to use *in silico* methods for gene content comparisons between phylogenetically distinct groups of this genera. The second dimension will be to use the rhizosphere metagenomes to profile other taxa, and to generate assemblies of high-abundance *Pseudomonas* to expand the pangenomes of the “blooming” species. Combining phylogeny, phenotype, secondary metabolite and antimicrobial gene content, and virulence potential will help to contextualize the bloom phenomenon of 2010; or at the very least provide some insight on broad-scale gene content variation to differentiate *Pseudomonas* lineages within the rhizosphere.

Chapter 1
Background

1.1. Members of the genus *Pseudomonas* are pervasive colonizers of the plant microbiome

1.1.1. Anatomy of the plant microbiome

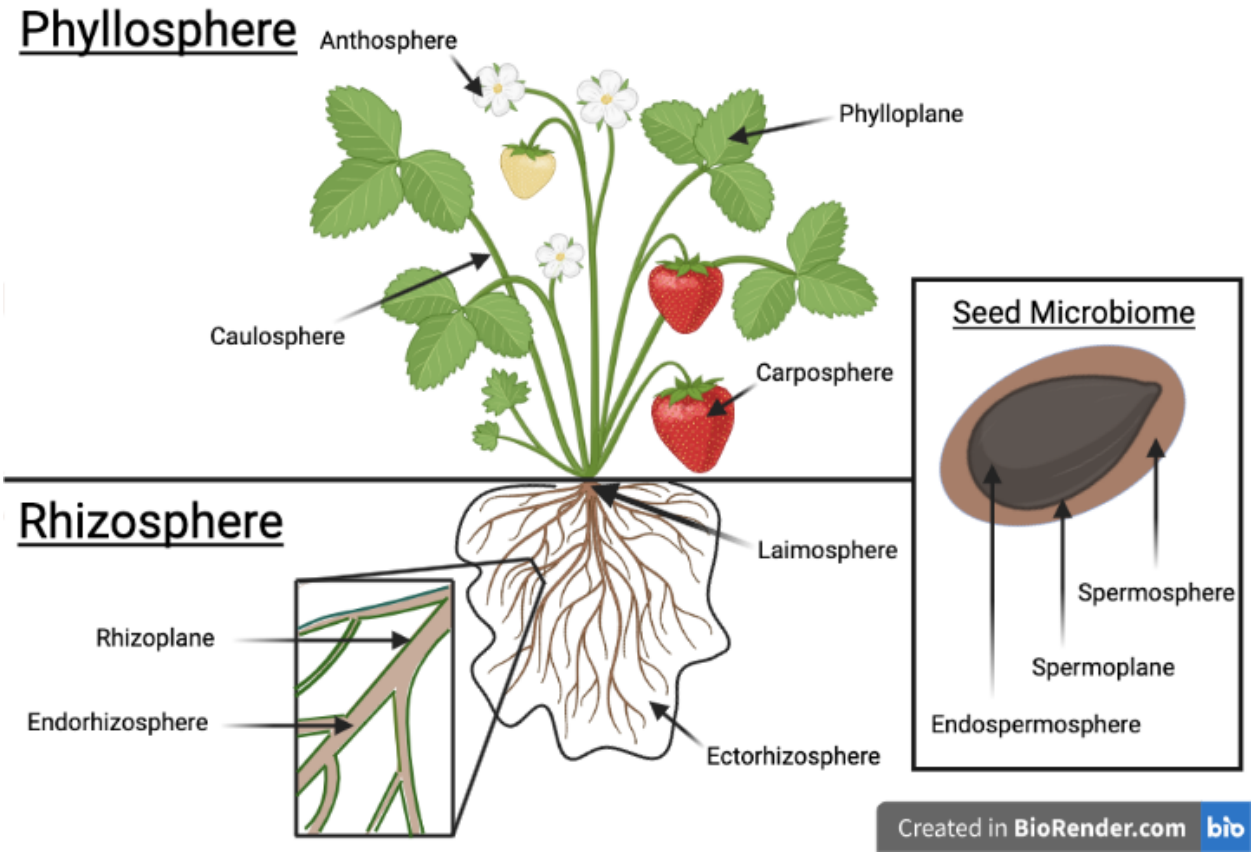


Figure 1 - General structure of the plant microbiome

For millions of years plants have co-evolved with microbes, and are colonized on their surfaces and within their tissues by diverse communities of bacteria. To truly understand the plant as a whole organism, it is essential to consider their microbiome and how it affects the processes we wish to interpret. Regarding the plant as a "holobiont" helps to contextualize the microbiome when comparing plant phenotypes (Vandenkoornhuyse *et al.*, 2015). The plant is subject to not only to immune interactions with the microbes that can range from beneficial to detrimental, but the microbes that exhibit mutualism and antagonism between each other, and both factors affect the structure of the microbial community. Persisting symbioses within the "holobiome" can

potentially influence the diversification, of both the eukaryotic as well as the prokaryotic members, within a volatile environment (Guerrero, Margulis and Berlanga, 2013).

The plant microbiome consists of numerous compartments made of different tissues and changes over the lifespan of the plant (Nelson, 2018). The first compartment is the phyllosphere, which comprises the microbes present on the aerial portions. This includes the caulosphere (stem), the phylloplane (leaves), the anthosphere (flowers), and the carposphere (fruit). The second compartment is the rhizosphere, which comprises the microbes found on the below-ground portions of the plant. This includes the laimosphere (the portion of the stem between the seed and the soil surface), the rhizoplane (the zone containing root rhizodeposits, exudates, and lysates), and the endosphere (interior of the root). The third compartment is the most transient yet foundational to the plant, the seed microbiome. This includes microbes that exist as epiphytes (on the surface of the seed), endophytes (within the interior of the seed) and within the spermosphere (the zone around the seed under the influence of seed exudates before the radicle emerges to form the nascent rhizoplane). Each of these compartments have different qualities that influence the microbes that occupy them, and these microbes are essential to mediating how the plant responds to immune and abiotic challenges.

The phyllosphere is a complex microbial habitat; this is due to the variety of plant tissues, varying degrees of exposure to the plant immune system, as well as physiochemical conditions such as UV radiation, temperature, and desiccation. Depending on individual species physiology, the leaves of the plant's phylloplane can represent a massive amount of microbiome surface area. When considering the upper and lower portions of the leaf, the calculated combined global leaf surface area may exceed 1,017,260,200 km^2 which is double the land surface area of Earth (Lindow and Brandl, 2003; Vorholt, 2012). Inoculum for the phyllosphere originates from both the air in a stochastic manner (Maignien *et al.*, 2014) and the soil which acts as a bacterial reservoir (Whipps *et al.*, 2008; Zarraonaindia *et al.*, 2015). Insect pollinators may also contribute to the transmission of microbes between the anthospheres of different plant species (Vannette, 2020), and predation by phytophagous insects such as caterpillars (Lilley *et al.*, 2006).

The phyllosphere is considered a low-nutrient environment where surface availability of simple carbohydrates are spatially diffuse and localized only in certain parts of the leaf (Leveau and Lindow, 2001). The age of the leaf affects the integrity of its cuticle, the waxy and hydrophobic layer that covers the leaf and acts as a diffusion barrier; its erosion causes an increase in wettability and thus increases the epiphytic bacterial populations (Whipps *et al.*, 2008). Overall, the phyllosphere's propensity for cycles of evaporation and rewetting, along with fluctuating temperatures and exposure to UV radiation makes it a challenging habitat for a bacteria to adapt to (Hirano Susan S. and Upper Christen D., 2000). It's no surprise that endophytic bacteria have found routes to penetrate the cuticle and enter the protected interior of the leaf, typically by way of the

stomata or wounds to the leaf surface (Frank, Saldierna Guzmán and Shay, 2017). However, despite the appeal of inhabiting the interior, only 1% of the phylloplane microbiota are considered endophytic and preferentially colonize the apoplast of the leaf (T. Chen *et al.*, 2020). The majority of microbes preferentially colonize the surface of the leaf and form aggregates of cells that can be phylogenetically diverse or clonal, and often form biofilms as a means of survival (Brandl and Mandrell, 2002; Tecon and Leveau, 2012).

In summary, the phyllosphere consists of many different compartments, with the largest by volume being the phylloplane (leaves). The phylloplane is low in labile carbon as well as nitrogen and phosphorus and is subject to drastic shifts in abiotic factors; despite this, the phyllosphere maintains high species richness (Dastogeer *et al.*, 2020). In experiments that swapped the legume *Medicago trunctula* between different soil types, the amplicon sequencing revealed that the phyllosphere community restructured each time to more closely resemble the rhizoplane microbiota (Tkacz *et al.*, 2020). These results suggest that the soil acts as the largest and most consistent reservoir for phyllosphere inocula.

Soil is among one of the most diverse microbiomes on earth; a single gram of soil may contain up to 10^{10} bacterial cells (Gans, Wolinsky and Dunbar, 2005; Raynaud and Nunan, 2014). Taxonomic diversity varies depending on land usage; forest soils tend to have higher richness at phylum level whereas agricultural soils have fewer represented phyla but very high species diversity (Roesch *et al.*, 2007). High taxonomic richness holds a strong positive correlation with plant biomass in greenhouse studies (Q.-L. Chen *et al.*, 2020). Comparing that to the fact air may contain only 10^6 cells/m^3 (Tignat-Perrier *et al.*, 2019), it should not come as a shock that the composition of the rhizosphere has a strong influence on all compartments of the plant microbiome.

It behooves us however to start from the beginning, with the seed; the seed originates from the flower, develops within a fruit, and therefore informs the seed microbiome inocula of angiosperms (Rodríguez *et al.*, 2018). Depending on the species and lifestyle of the plant, the fruit may pass through the digestive tract of an animal, be spread by air or water, or be handled by humans prior to its contact with the soil for germination. There has been increased interest in exploring how much of an influence the seed microbiome has on the establishment of the rhizosphere and if the phylogeny of the plant has any effect on what microbes get selected for within the endospermosphere (Nelson, 2018). Studies investigating the influence of host genetics on rhizosphere community structure grew *Zea mays* in sterile sand vs standard garden soil and observed the dominant taxa of the sterile sand group were also dominant within the soil rhizospheres, suggested that some taxa (specifically Proteobactera and Bacteroides) may originate with the seed and persist as the plant develops (Johnston-Monje *et al.*, 2016; Johnston-Monje, Gutiérrez and Lopez-Lavalle, 2021).

There is also evidence that endospermosphere microbiota may be vertically or horizontally transmitted (Shade, Jacques and Barret, 2017; Shahzad *et al.*, 2018) and select for endophytes that have beneficial functions within the rhizosphere such as phosphate solubilization and nitrogen fixation in *Zea mays* (maize) and *Pachycereus pringlei* (giant cardon cactus) (Puente, Li and Bashan, 2009; Johnston-Monje and Raizada, 2011). The authors of that study also observed an overrepresentation of some taxa within species of *Zea* ancestral to *Zea mays* producing a fruitcase, or exocarp; and these taxa were nearly absent from the seeds of *Z. mays*. This suggests the anatomical changes brought about by domestication may alter the colonization of seed microbiomes. Human microbiomes have been described as “an ecosystem on a leash” where complex interactions between all of its constituents (bacteria, eukaryotic, fungal, and archaeal) are on a “leash”, or constrained by the host via immune responses or the modulation of what substrates are available (Foster *et al.*, 2017). Domestication can, in effect, attach another leash; the modification of a host through artificial selection for traits beneficial to humans puts additional constraints on the host’s capacity to interact with its own microbiome (Soldan *et al.*, 2021).

This principle applies just as strongly to the rhizosphere of plants. As the embryo within the seed germinates, the radicle protrudes into the soil where it develops into the primary root. The microbes of the seed spermosphere are highly influenced by the type of soil and the resident microbial community; studies of fluorescent *Pseudomonas* demonstrated significant differences between the taxa found within the rhizospheres of flax (*Linum usitatissimum*) and tomato (*Solanum lycopersicum*) versus the bulk soil, and significant differences between the plants grown in different geographic locations with differing soil types (Latour *et al.*, 1996). A follow-up to this study further explored the influence of soil type on synthetic communities of fluorescent *Pseudomonas* using sterilized soils and demonstrated that abiotic parameters such as clay composition affect the carrying capacity of these strains; however, they observed no consistency in community structure between similar conditions which may suggest biotic factors such as soil microbial community structure play an important part in defining what is competitive in this environment (Latour *et al.*, 1999). These results were recapitulated by a contemporary study with maize and the Betaproteobacteria *Burkholderia cepacia*, where soil type had the most influence on the genetic diversity of rhizosphere-isolated *B. cepacia* when compared to maize cultivar or root compartment (Dalmastri *et al.*, 1999). All in all, despite the seed carrying its own unique microbiome that has been cultivated by selective factors, the soil in which the seed grows sets the foundation for both the rhizosphere and the phyllosphere.

Once the seed germinates the root is in intimate contact with an immense number and diversity of microbes; fungal, bacterial, archaeal, and eukaryotic. The plant has established mechanisms for the recruitment and management of the rhizosphere; first by affecting the structure and biochemistry of the soil, including modulating soil pH and

oxygen concentrations, altering soil structure, and mobilizing iron (Marschner, Treeby and Römheld, 1989; Lundberg *et al.*, 2012). Plants are sessile and therefore unable to physically avoid predation or infection by phytopathogenic microbes; these organisms have multiple ways to cause illness; biotrophic pathogens behave as parasites and siphon off nutrients from the plant without killing it whereas necrotrophic pathogens prefer to secrete toxins and feast on the nutrients released from dead cells. Hemibiotrophic pathogens exhibit a dual strategy, consisting first of a biotrophic phase then eventually switching to a necrotrophic phase upon specific regulatory signals (Horbach *et al.*, 2011; Koeck, Hardham and Dodds, 2011; Kemen and Jones, 2012).

The plant also manages colonizing microbes by way of immune surveillance using pattern-recognition receptors (PRRs). This type of innate immunity responds to pathogen-associated molecular patterns (PAMPs) and microbe-associated molecular patterns (MAMPs) that induces what is known as PAMP or pattern or pathogen-triggered immunity (PTI) to fend off the infection. Defense mechanisms can include the closing of leaf stomata, restricting nutrient flow to the apoplast, secretion of antimicrobial peptides, and dramatically increasing reactive oxygen species (ROS burst) to kill the invader (Bigeard, Colcombet and Hirt, 2015).

In *Arabidopsis thaliana* studies have revealed different categories of PRRs that specialize in recognizing different types of microbial patterns. One such PRR in *Arabidopsis* are the intracellular nucleotide binding site leucine-rich repeat (NBS-LRR) pathogen-resistance proteins, which are similar in structure to mammalian c-type lectin receptors (CLRs) and are sensitive to pathogen-encoded virulence factors and microbial effector proteins that may enter plant cells by bacterial type III secretion systems (Ausubel, 2005; DeYoung and Innes, 2006).

Plants also have transmembrane PRRs that are poised to identify PAMPs and MAMPs such as bacterial flagellin, elongation factor Tu, peptidoglycan, lipopolysaccharides (LPS), fungal chitin, and oomycete beta-glucans that elicit immune responses (Newman *et al.*, 2013). A well-studied example of one such PRR in *Arabidopsis* is the LRR-receptor kinase FLAGELLIN-SENSING-2 (FLS2); it binds to flg22, a highly conserved domain within bacterial flagellin where the D_0 meets the D_1 domain. In mammals the transmembrane Toll-like receptor 5 (TLR5) serves a similar function, flagellin recognition, but is keyed to the amino acids within the D_1 domain (Zipfel and Felix, 2005). Essentially, plants are vigilant to both conserved and generalized microbial products with their external PRRs, such as FLS2; and they are also capable of discerning species-specific virulence factors and pathogenic effectors that enter the plant cell using NBS-LRR pathogen-resistance proteins (Jones and Dangl, 2006). Recent work in *Arabidopsis* with rhizosphere *Pseudomonas* has demonstrated the ability for some commensal taxa to increase colonization of lateral roots by suppressing flg22 signaling by reducing the pH of the soil with gluconic acid (Yu *et al.*, 2019). In the absence of a

disease phenotype this behavior is considered largely beneficial, as the resident microbes may outcompete a pathogen.

Once a pathogen has been detected, this typically induces the hypersensitive response (HR) which triggers apoptosis of the affected cells. However, PRR response also causes the upregulation of pathogenesis-related proteins (PRs) and antimicrobial secondary metabolites such as Cys-rich peptides (CRPs); in *Arabidopsis*, CRPs may account for up to 3% of all proteins expressed (Stintzi *et al.*, 1993; Tam *et al.*, 2015). Early studies of HR in tobacco plants challenged with the tobacco mosaic virus (TMV) gave evidence for increased resistance to TMV in adjacent tissues of infected plants. This observation led to the discovery of the mechanisms behind plant systemic acquired resistance (SAR), a process modulated by salicylic acid signaling (Dempsey, Shah and Klessig, 1999; Durrant and Dong, 2004).

The process of SAR is typically triggered by a pathogen, yet the vast majority of rhizosphere microbes are not pathogenic, or at least not capable of pathogenesis in a colonized rhizosphere (Wang *et al.*, 2021). However, many taxa of bacteria produce flagellins, LPS, and other MAMPs that are detectable by the host. Explorations into this phenomena have lead to the discovery of an immune response pathway that exists parallel to SAR, termed rhizobacteria-mediated induced systemic resistance (ISR), that relies on microbial products often detected by PRRs but does not trigger salicylic acid signaling (van Loon, Bakker and Pieterse, 1998; Meziane *et al.*, 2005). It is worth noting that some bacterial MAMPs may induce systemic resistance in the plant against a pathogen of a different phyla, or even kingdom (Audenaert *et al.*, 2002; De Vleesschauwer *et al.*, 2008; Omoboye *et al.*, 2019). Notably, however, some strains of rhizosphere *Pseudomonas* are known to increase susceptibility to pathogenic phyllosphere *Pseudomonas,* yet these same rhizosphere strains are capable of inducing systemic resistance to herbivory by insects. This effect appears to be influenced by increasing jasmonic acid (JA) signaling while suppressing salicylic acid (SA) signaling (Haney *et al.*, 2018), and this depression of SA signaling is not observed in other closely related *Pseudomonas.*

The rhizosphere generally refers to the below-ground portion of the plant microbiome, but specifically it is the zone around the root that is under the influence of plant exudates. Plants dedicate a substantial amount of energy to root exudation; in trees and grassland biomes, may range between 30-50% of the plant's photosynthates are released as root exudates. In cereals that use the C3 photosynthetic cycle such as wheat and barley, root exudates may range from 20-30% of photosynthates (Kuzyakov, Domanski and Others, 2000). Early studies of photosynthate exudation in roots measured under 0.4%, but that modest allocation of carbon remained highly stimulating to soil microbes (Rovira, 1969).

The act of releasing carbon and other compounds into the soil is referred to as rhizodeposition. This can be broadly categorized based on whether the exudate was

secreted through an active, metabolically costly process, or if the exudate was passively released through a basal process (Jones, Hodge and Kuzyakov, 2004). Examples of passive rhizodeposition include the release of amino acids, sugars, and nucleotides from the lysis of dying root cells, and the translocation of gasses such as ethylene and CO_2 from the roots to the soil (Lynch and Whipps, 1990).

Rhizodermal cells, or the epidermal cells of the root, are capable of secreting a complex selection of compounds via both rhizodeposition processes. These can include the passive leaking of water-soluble exudates like simple sugars, organic acids, amino acids, nucleosides, purines, and phytohormones (Lynch and Whipps, 1990). Rhizodermal cells may also actively secrete secondary metabolites such as phytosiderophores for the uptake of iron and other essential metal cofactors, phenolics, flavonoids and terpenoids, in addition to the aforementioned antimicrobial CRPs (Dakora and Phillips, 2002). The rate and type of exudates released are affected by the morphological region of the root; the root cap at the tip of the root protects the root apical meristem, and beyond the meristem is the differentiation and elongation zones, followed by the maturation zone where the lateral roots will eventually emerge (Burstrom, 1953).

For a long time it was thought that the mucilage produced to lubricate the root cap as the root penetrates the soil was merely a source of high molecular weight insoluble polysaccharides; any other exudates detected near the root apical meristem were from cells being lysed by the friction of the growing root (Voeller, Ledbetter and Porter, 1964). It was therefore hypothesized that the area around the root cap was an attractive region for rhizosphere microbes due to the guaranteed polysaccharides and passively-released cytoplasmic leakage. However, results from scanning electron microscopy of field grown wheat roots revealed that even with high microbial adherence along the root surface, the root cap was surprisingly devoid of colonization (Foster *et al.*, 1983). This result was recapitulated later even with studies that inoculated roots with bacteria known to be strong colonizers, such as *Pseudomonas* (Gamalero *et al.*, 2005), and in one experiment it was observed that the removal of the root cap allowed *Pseudomonas fluorescens* to colonize the root apical meristem directly (Humphris *et al.*, 2005). Clearly, some manner of host mechanism was hard at work to protect the vulnerable site of elongation from infection despite this area being especially rich in exudates.

In mammals some immune cells known as neutrophils produce histone-linked extracellular DNA (“exDNA”), or neutrophil extracellular traps (“NETs”), that capture pathogens and attract macrophages that go on to utilize the neutrophil-derived antimicrobial peptides enhance killing. (Papayannopoulos, 2018; Monteith *et al.*, 2021). Although plants lack specialized immune cells such as neutrophils and macrophages, they are capable of producing exDNA at the root cap to serve a similar function (Hawes *et al.*, 2011). In addition to the mucilage and exDNA produced at the root cap, specialized cells called border cells are also released into this elongating region. Border cells originate from the rhizodermis adjacent to the root cap and detach; they are not dead, but

remain metabolically active and will divide if placed in culture. However, *in planta,* border cells exist to deliver a host-specific cocktail of exudates, mucilage, and antimicrobial peptides into the rhizosphere before succumbing to apoptosis (Hawes *et al.*, 2011).

The exact profile of border cell exudates appears to be tailored by the host to attract commensal and beneficial microbes to the rhizosphere. The rate at which the plant releases border cells is highly influenced by a combination of host genotype and environmental signals, specifically free water and plant age (Ponce *et al.*, 2005; Odell *et al.*, 2008; Somasundaram, Fukuzono and Iijima, 2008). Border cells retain PRR activity and will secrete mucilage in response to exposure to bacteria, fungi, or in some cases heavy metals thus forming aggregates with other border cells hundreds of cells wide (Wen *et al.*, 2009; Hawes *et al.*, 2011). In maize, border cells can remain metabolically active for more than a week, although this timespan may vary by host (Vermeer and McCully, 1982). Together, the root cap and border cells exist synergistically by stimulating the production of mucilage, exDNA, and secondary metabolites to that act as signaling molecules to detect various biotic and abiotic factors within the rhizosphere that trigger host metabolic and transcriptional changes, especially in hosts that form nitrogen-fixing symbioses with species of *Rhizobium* (Maxwell and Phillips, 1990; Baluška *et al.*, 1996; Hawes *et al.*, 2011).

As the plant is capable of exerting some degree of influence over what sources of carbon are more abundant relative to the surrounding bulk soils as well as the entrapment and killing of some microbes in the vicinity of the root tip, are there specific microbes the plant is attempting to recruit, if any? What microbial taxa appears to be successful at rhizosphere colonization regardless of host? Studies in *Arabidopsis* demonstrated that the ectorhizosphere and endorhizosphere are strongly influenced by soil type, as expected; however the endorhizosphere of plants grown in different soil types were enriched in *Actinobacteria* and *Proteobactera* (Lundberg *et al.*, 2012). It has been hypothesized that the rhizosphere exhibits a two-step selection model for microbiota differentiation; edaphic factors provide the basal soil microbiome, then rhizodeposition (including exudation, border cells, exDNA, and root structure) restrict the growth of some taxa while cultivating others. This reduction in diversity is further evident once the microbes contact the rhizoplane and cross into the endorhizosphere (Bulgarelli, Schlaeppi and Spaepen, 2013). In both wild and agricultural settings, ectorhizospheres have lower taxonomic diversity than bulk soil but can be 10-100 times higher in bacterial density (Berendsen, Pieterse and Bakker, 2012).

Each compartment of the rhizosphere (ectorhizosphere, rhizoplane, and endorhizosphere) represent different niches with varying access to carbon and host immune influence. Outcomes from ectorhizosphere microbial processes can benefit the host through nitrogen fixation, phosphorus solubilization, antagonism or outcompeting pathogens, and improve abiotic stress tolerance such as under drought conditions (Mendes *et al.*, 2011; Naylor and Coleman-Derr, 2017).

The microbes that colonize the endorhizosphere without pathogenesis, either as commensals or symbionts, originate from the ectorhizosphere, the rhizoplane and from the endospermosphere. Obligate endophytic bacteria originate from the spermorhizosphere and require the host to survive, whereas facultative endophytic bacteria are ectorhizosphere bacteria that opportunistically colonize the interior of the root under specific conditions. In the case of facultative endophytic bacteria, some are passive colonizers that are ectorhizosphere or soil residents that take advantage of available nutrients in the proximity of root tissues, and can be cleared by triggering ISR pathways. Facultative or obligate endophytes that possess genetic machinery to subvert plant immune responses or to produce signaling molecules that aid in colonization tissues or within the apoplast are considered “competent” endophytic rhizobacteria (Hardoim, van Overbeek and van Elsas, 2008).

These microbes colonize the cortex of the root; in the case of arbuscular mycorrhizal fungi, they form arbuscles that penetrate into the root cells as part of their symbiosis (Parniske, 2008a). Ectomycorrhizal fungi colonize between the rhizodermal and cortex cells. Bacteria can enter the rhizodermal tissue at cracks between cells or in wounds caused by nematode grazing. The edge of an emerging lateral root is a vulnerable area where bacteria may penetrate into the xylem and phloem where they translocate to tissues in the phyllosphere (Reinhold-Hurek and Hurek, 1998; Liu *et al.*, 2017).

The rhizoplane is essentially the “gate” to the endorhizosphere; the plant, though exudation can induce changes in soil structure and biochemistry is the first level of exclusion, and the bacteria that have a strong chemotactic attraction to the source of exudates are more likely to make contact with rhizodermal cells (Edwards *et al.*, 2015). There is evidence that translocation from the rhizoplane to the endorhizosphere requires more than simple attachment, and may involve the formation of biofilms or direct selection (or exclusion) by the host (Walker *et al.*, 2004; Compant, Clément and Sessitsch, 2010; Edwards *et al.*, 2015).

In summary, the plant microbiome is a complex habitat that is primarily distinguished between tissues that are aerial or underground. The phyllosphere consists of the surface and interior of leaves, stems, flowers and fruits. Microbes that inhabit these regions generally originate from the air and soil. Unlike the leaves, plant roots are exposed to a very high density of soil microbes and this has driven dynamic physiological and immune responses within the plant to manage their microbiota. The ectorhizosphere is the zone of soil around the root mass that is under the influence of root exudates and border cells, and these mechanisms serve to provide carbon substrates that attract and enrich for commensal and symbiotic microbiota, whereas the release of mucilage, exDNA, and various types of antimicrobial peptides discourage the growth of pathogenic microbes and deterring colonization of vulnerable tissues. Some microbes, whether pathogenic or symbiotic, are attracted to the source of these exudates and colonize the

rhizoplane. The plant immune system is vigilant for PAMPs and will initiate ISR or SAR to defend against pathogenic infection. Bacteria that colonize the rhizoplane may gain entrance to the endorhizosphere through nematode grazing wounds, gaps between rhizodermal cells, or at sites of lateral root emergence. Once inside the endorhizosphere facultative endophytes take advantage of the sheltered and nutrient rich environment without significant signaling with the host. Obligate endophytes require the host to survive and often form symbioses that are predicated on the mutual exchange of nutrients or secondary metabolites. Both facultative and obligate endophytes have the capacity to enter the xylem and be transported up to the phyllosphere. There they may potentially colonize reproductive compartments such as the anthrosphere or carposphere and be vertically transmitted to the endospermosphere within the resulting seed.

1.1.2. The convoluted history of *Pseudomonas*

The rhizosphere wouldn't exist without the plant; but as demonstrated in the above section, the plant is under enormous pressure to reign in the microbes that it finds itself surrounded by. Each species, even each individual plant, maintains a unique rhizosphere (excluding gnotobiotic and experimental synthetic communities) that is influenced primarily by the soil inocula, abiotic factors, and host genotype. There are microbial taxa that have become highly adapted to the rhizosphere niche; some in a broad sense as generalists while others may have formed symbioses or parasitic relationships with the host or other host-associated members of the rhizosphere. To highlight each prominent family, genera, or species of rhizosphere generalist is very much beyond the scope of this monograph. However, the bloom of *Pseudomonas* observed during the 2010 maize rhizosphere study captured a very peculiar event in time that, with a more thorough investigation of species diversity, may provide new insight into *Pseudomonas* functional diversity at species level.

Pseudomonas is a nearly inescapable environmental inhabitant and a tremendous amount of work spanning has been done to characterize it. The first attempt to describe this aerobic, gram negative, polar-flagellated, rod-shaped bacterium was in 1894 (Migula, 1895) at the Karlsruhe Institute; shortly after the type species was established as *Pseudomonas pyocyanea*, later renamed to *Pseudomonas aeruginosa.* As early as the 1920's it was observed that members of *Pseudomonas* had very diverse metabolic capabilities, being able to consume and degrade a variety of organic compounds. Some strains were able to decompose otherwise toxic compounds such as some aromatic hydrocarbons. Even botanists contemporaneous to Lorenz Hilter, the plant scientist and director of the Royal Agriculture-Botanical Institute in Munich who initially defined the rhizosphere as "soil influenced by roots" in relation to the nitrogen fixation activity of root nodules (HILTNER and L, 1904), identified *Pseudomonas syringae* as the causative agent of disease across multiple plant species (Smith, 1904).

As the field of bacteriology progressed during the early and mid portion of the 20th century species of bacteria were continually being assigned to the genus *Pseudomonas;* however, at this time bacterial taxonomy was entirely based on observations of morphology, physiology, and biochemical assays (Lysenko, 1961; Schleifer, 2009). Archaea wasn't identified as a separate kingdom of microbial life until 1977, and that was preceded by the development of molecular sequencing techniques (Woese and Fox, 1977). Even without the ability to analyze genotype, by the early 1960's there were an estimated 800 species names assigned to the genus *Pseudomonas (Palleroni, 2010)*. Not a surprise at all considering how incredibly common it is in the environment, and how amenable it is to grow in culture. Eventually it was acknowledged the state *Pseudomonas* taxonomy was becoming "chaotic"; some microbiologists even expressed a desire for more consistent and uniform methods of connecting nomenclature to bacterial phenotype (Van Niel and Stanier, 1962).

It wasn't until the development of DNA/DNA hybridization (DDH) that comparisons between strains could be made independent of an observable phenotype, and this thankfully came along shortly after the discourse on standardizing bacterial nomenclature with the foundational work exploring taxonomy in enteric bacteria using *Escherichia coli* and *Salmonella typhimurium* (Brenner and Cowie, 1968; Brenner, 1973). This chromatographic technique involved taking labeled single-stranded DNA (ssDNA) from a reference bacteria and combining it with unlabeled single-stranded DNA of another bacteria and measuring the degree of reassociation, or hybridization, that occurs between the two. Highly similar sequences will bind while the dissimilar unbound ssDNA can be eluted and measured as a fraction of the whole (Bernardi, 1965). After a few decades many species assigned *Pseudomonas sensu stricto* were eventually reassigned to other genera after DDH analysis, typically to another class of *Proteobacteria (Kersters et al., 1996)*. It was also observed that some species of *Pseudomonas* had varying degrees of DDH between strains and subspecies; specifically those of *P. aeruginosa* being consistently more similar to each other whereas strains within *P. fluorescens* and *P. syringae* were much less similar to each other, to the point of likely being different species (Silby *et al.*, 2011).

As mentioned above, improvements in molecular sequencing techniques opened a new dimension to bacterial phylogenetics and taxonomy. The pioneering work of Carl Woese on distinguishing Archaea from Bacteria using comparisons of the highly conserved 16S ribosomal RNA (rRNA) gene was immediately utilized to better characterize *Pseudomonas* phylogenetic relationships. Earlier work on rRNA/DNA hybridization in *Pseudomonas* observed five broad phylogenetic groups based on rRNA homology, and their phylogenetic distances from each other strongly suggested reassigning at least one or more to a new genus, or family (Palleroni *et al.*, 1973). Carl Woese himself was especially critical of *Pseudomonas* nomenclature and how it permitted such phenotypically diverse isolates to be classified within this genus. The

results of the rRNA hybridization experiments was a clear indicator that the classical definition of *Pseudomonas* was not a proper phylogenetic unit and an inappropriate means to define the genera. He then took a variety of isolates classified as *Pseudomonas* and compared their 16S rRNA sequences by oligonucleotide cataloging. The results recapitulated the 5 groups observed by Palleroni and colleagues, but scrutinizing only rRNA revealed an even more stark phylogenetic divergence between these "Pseudomonads" (Woese, Blanz and Hahn, 1984). This experiment was clearly a flex; he absolutely expected the outcome he observed, as per the last sentence of the introduction: "It would seem that the initial choice of the name *pseudo-monad* (i.e. a false unit) was an especially wise one". The final statement of the discussion was especially prescient:

> "*What is becoming clear with genetic (i.e., sequence) characterizations of bacteria is that most classical phenotypic characteristics are generally poor a priori indicators of phylogenetic relationships. The phylogenetic relationships, which hold the answers to questions of bacterial evolution, must be determined by methods that measure differences in molecular sequences - the best of these being those that actually determine sequence itself.*"

As Woese predicted, the study of bacterial evolution was destined to become a predominantly bioinformatic affair once oligonucleotide sequencing technologies increased in accuracy while also decreasing substantially in cost. The first DNA polymerase-based sequencing method was developed by Frederick Sanger in 1977 using electrophoresis and chain-termination to determine the order by which labeled deoxynucleotides are added to a ssDNA template (Sanger, Nicklen and Coulson, 1977). This allowed for the precise and accurate determination of DNA sequences under 700 base pairs (bp), but this method is limited to 96 lanes of one sample per lane. Ten years later Edward Hyman devised a method without electrophoresis that measured the release of pyrophosphate (PPi) during DNA polymerization (Hyman, 1988), with this assay being refined with continuous monitoring by Pål Nyrén a year later (Nyrén, 1987). Concurrent development of solid phase sequencing methods that allowed for automation of the pyrosequencing technique by fixing the DNA template on streptavidin beads (Hultman *et al.*, 1989) which lead to its commercialization as a platform with the final method proposed by Mostafa Ronaghi and colleagues (Ronaghi *et al.*, 1996). Pyrosequencing was the first of the next generation sequencing technologies to enable higher throughput with multiplexing samples but was accurate for sequences <500 bp.

Also around this time other research groups were exploring other methods of sequencing-by-synthesis (SBS). While pyrosequencing relied on measuring the release of PPi during DNA polymerization, which entailed cycling through dNTPs one at a time

and measuring the intensity of light emitted when a nucleotide is successfully incorporated, Bruno Canard and colleagues developed a method that utilized reversibly tagged fluorescent dNTPs (3'-RTa-dNTPs) that acted as chain terminators (Canard and Sarfati, 1994). The tagged dNTPs were added to solid phase ssDNA and when the appropriate 3'-RTa-dNTP was incorporated the unreacted components were washed away, thus allowing the fluorescent tag to be removed and eluted off as an indicator of which dNTP was incorporated. Since the latter method requires fewer reagents to produce sequences for the same short read distance, it quickly became the main competitor to pyrosequencing. In 1998 researchers from Cambridge University committed to optimizing the 3'-RTa-dNTP method to found the company Solexa (later Illumina), while the pyrosequencing platform came to market with 454 Life Sciences in the year 2000. Both companies would begin the second era of next generation sequencing; during this time there was a tremendous push by the Human Genome Project (HGP) to develop technology that will allow for the complete sequencing of the human genome. The dream of personalized medicine was the $1,000 draft human genome and the promise it held, which led to immense pressure between competing technologies to develop comprehensive platforms to sequence at the lowest cost.

Naturally, at this point in the discussion of DNA sequencing technologies, what does this have to do with the history of *Pseudomonas*? One could posit the real winners in the race to produce cheap human genomes are the microbiologists who, after this period in the mid 2000's, reaped the benefits of short read sequencing to produce data hand over fist. As bioinformatic methods caught up to the quantity of reads being produced, the immediate focus was on multi-locus sequence typing (MLST); primers targeting specific highly-conserved genes allowed for the amplification and phylogenetic comparisons between bacterial isolates in culture. MLST became a powerful tool in discriminating pathogenic bacteria in a clinical setting (Maiden *et al.*, 1998; Maiden, 2006). While not the only biochemical method for comparing strains, beginning in 2002 and until about 2011, MLST became the most common bacterial typing method by publication. In 2012 however it was surpassed by whole genome sequencing (WGS) (Pérez-Losada *et al.*, 2013).

Again, why does this matter to those interested in *Pseudomonas*? As Palleroni and Woese had been belaboring for years, *Pseudomonas* was rife with species identified by classical methods as *sensu stricto* when their genotype would place them well outside the genus. As of January 2022, there are 10,819 genomes assigned *Pseudomonas* within the Genome Taxonomy Database (GTDB) and 27,899 genomes uploaded as *Pseudomonas* within the National Center for Biotechnology Information (NCBI). Within NCBI is GenBank (established 1983) and the Reference Sequence Database (RefSeq, established 1999), with the latter being a highly curated public database of non-redundant nucleotide and protein sequences. Notably, GTDB utilizes the genomes available from RefSeq and clusters genomes based on average nucleotide identity (ANI), resulting in

taxonomic classifications by ANI thresholds of aligned fractions greater than 150 kb (Parks *et al.*, 2020). The disparity between the number of genomes submitted to NCBI assigned to *Pseudomonas* versus roughly half that when clustered by ANI is evidence that even with genomic tools, what is or is not a *Pseudomonas* can vary depending on the database used.

Also, within GTDB those approximately 10k *Pseudomonas* genomes are highly multiphyletic, and contain 16 subgroups. Since the discovery of the genus in the late 1800's there have been several relevant species that are highly overrepresented within culture collections and publicly available databases; some strains of *Pseudomonas aeruginosa* are common opportunistic pathogens of plants, animals, even nematodes, and therefore appear very frequently in clinical settings; it is of significant concern to immunocompromised individuals and those with cystic fibrosis. The Pseudomonas Genome Database (PGD) (Winsor *et al.*, 2016) has (as of January 2022) 4,955 curated strains of *P. aeruginosa* listed, GTDB suggests monophyletic subgroups with the largest containing 5,211 genomes and the other 41, and NCBI 15,506. Another highly represented species is *Pseudomonas viridiflava*, which is a common pathogen in many species of plants. It is poorly represented within the PGD with 18 strains, GTDB again suggests three separate groups with the largest containing 1,360 strains while the other two have 2 strains each. Unsurprisingly, NCBI lists 3,073 *P. viridiflava* strains. Does this mean half of the total *Pseudomonas* genomes uploaded to NCBI are not the correct genus? Unlikely, as many genomes may not pass quality control thresholds but *Pseudomonas* was observed as a genera with commonly reassigned sequences during the construction of GTDB (Parks *et al.*, 2020). Taken as a whole, it is clear that *Pseudomonas* is a very diverse organism that is difficult to taxonomically resolve without taking genotype into account.

1.1.3. *Pseudomonas* facilitates complex inter-phyla interactions

Members of the genus *Pseudomonas* exhibit a vast repertoire of metabolic potential that allow it to thrive wherever there is available carbon. Individual taxa are more common in specific environments such as the soil, water, and within both the rhizosphere and phyllosphere. Soils are the predominant reservoir for the *Pseudomonas* that colonize both plant biomes; many of the *Pseudomonas* species that inhabit soils are not competitive within the rhizosphere. Niche formation within these habitats is a complicated process which involves considering both host-microbe and microbe-microbe interactions, in addition to the abiotic factors that affect both (Kroll, Agler and Kemen, 2017). This hints at a possible driver of the vast diversity within *Pseudomonas;* for every host there is opportunity, and this attraction creates competition between not only disparate taxa but also closely related species. These ancient circumstances are carried along by qualities

specific to bacteria, such as their capacity for horizontal gene transfer (HGT) and rapid mutation (Spiers, Buckling and Rainey, 2000).

The size of *Pseudomonas* genome accurately reflects its behavior as a jack-of-all-trades type generalist with the low end being ~4.8 kilobases up to ~7 kb at the high end. Genome size is typically positively correlated with the frequency of HGT (Nakamura *et al.*, 2004) which makes sense considering the close proximity that microbes experience when colonizing a host. Species with smaller genomes and proportionally fewer genes are less likely to be competitive within the ectorhizosphere and rhizoplane, but may be competitive with the endorhizosphere if the mechanisms for host symbiosis allow for limited competition with other microbes. Larger genome sizes suggest an accumulation of, and selection for, genes relating to host colonization (Barret, Morrissey and O'Gara, 2011).

Competency, or the ability to successfully colonize a rhizosphere, only describes the presence of a *Pseudomonas* and not the specific behavior once established. As mentioned numerous times above, *Pseudomonas* are capable of a variety of lifestyles. The raw diversity of this genus mandates that unless there is evidence of a particular phenotype most rhizosphere competent *Pseudomonas* are likely commensal. However, as discussed already there are species highly characterized which are pathogenic to agriculturally significant hosts. Examples include the aforementioned *P. viridiflava* (bacterial blight) but other notable plant pathogens include *P. syringae* (bacterial speck), *P. corrugata* (pith necrosis), *P. marginalis* (root browning), *P. salomonii* (brown blotch), *P. cichorii* (leaf blight), and *P. aeruginosa* (soft rot). There are also species of rhizosphere-competent *Pseudomonas* that become opportunistically pathogenic in atypical hosts, such as fish in the case of *P. anguilliseptica* (red spot disease) in cold conditions when the fish becomes immunocompromised (Contessi *et al.*, 2006; Magi *et al.*, 2009). As a general rule any organism would do well to avoid *Pseudomonas* when immunocompromised.

However not all *Pseudomonas* are closet hemibiotrophic pathogens; many species have evolved to engage with the host in beneficial ways, and this has been observed for as long as microbiologists have studied plants. In this context these microbes are referred to as plant growth-promoting rhizobacteria (PGPR) and often are identified among the genera *Azoarcus*, *Azospirillum*, *Azotobacter*, *Arthrobacter*, *Bacillus*, *Clostridium*, *Enterobacter*, *Gluconacetobacter, Serratia,* and of course *Pseudomonas.* A non-exhaustive list of PGPR activities include: the solubilization of phosphorus, preferentially colonizing and outcompeting pathogens within the rhizoplane, or the production of phytohormones that suppress plant immune responses in situations of abiotic stress, or stimulate ISR to increase host resistance to pathogens (Benizri, Baudoin and Guckert, 2001; Somers, Vanderleyden and Srinivasan, 2004; Hayat *et al.*, 2010).

The production of phytohormones by rhizobacteria was observed during the study of "bacterial fertilizers" and how their application to soils and seeds affected plant yield

(Mishustin and Naumova, 1962), and later work to disentangle the influence of bacteria-derived phytohormones on plant morphology suggested they have a strong influence on root biomass in a dose-dependent manner (Müller, Deigele and Ziegler, 1989). This phytostimulation effect by PGPR can alter root system architecture (RSA) leading to changes in the spatial distribution of primary and lateral roots compared to plants grown axenically, or in the absence of rhizobacteria. Many species of *Pseudomonas* have been found to produce phytohormones; the highly diverse *Pseudomonas fluorescens* have many plant growth-promoting species that produce the auxin indole-3-acetic acid (IAA) and cytokinin zeatin. Both IAA and zeatin stimulate root hair formation, with high concentrations of IAA also increasing lateral root emergence. Cytokinins increase cell proliferation in meristematic tissues and increase shoot to root ratio (Vacheron *et al.*, 2013). Another example of PGPR-mediated RSA modification includes the ability of some rhizoplane and endophytic *P. fluorescens* and *P. putida* to induce ISR which results in the accumulation of lignin in rhizodermal cells, leading to increased resistance to invasion by pathogenic oomycetes (Benhamou, Bélanger and Paulitz, 1996; Iavicoli *et al.*, 2003).

Fungi are also influential members of the rhizosphere, and unsurprisingly exhibit just as diverse relationships with the host. However, despite the threat that some species pose as pathogens, plants have the benefit of recruiting types of plant-beneficial fungi that can greatly increase nutrient availability in challenging environments. The most commonly observed relationship is that of arbuscular mycorrhizae (AM), which consists exclusively of the phylum *Glomeromycota (Schüßler, Schwarzott and Walker, 2001)*. This taxa is present in the roots of over 70% land plants, and fossil and molecular evidence of the emergence of ancestral AM fungi from 353-462 Myr ago suggests they may have been instrumental in land colonization (Simon *et al.*, 1993; Remy *et al.*, 1994). Species of this taxa are obligate endosymbionts that form unique branching structures called arbuscules inside cortex cells; these allow phosphorus and nitrogen collected by the fungal hyphae to enter the plant in exchange for carbon, essentially acting as an extension of the root system (Parniske, 2008b).

Rhizosphere mutualisms between plants and fungi can also involve the creation of extracellular networks of hyphae within root tissues between cells and without the formation of arbuscules, known as ectomycorrhizae (EcM). Unlike AM fungi, multiple phyla contain EcM taxa, such as *Basidiomycota, Ascomycota,* and *Zygomycota.* Again, unlike AM fungi only 2% of plant species (predominantly temperate and boreal trees) are known to form this type of reciprocal parasitism, and whereas AM fungi reproduce asexually all EcM fungi reproduce sexually and some produce commercially significant fruiting bodies, e.g. truffles (Smith and Read, 2010; Tedersoo, May and Smith, 2010).

As the relationship between fungi and plants were evolving, it was always in the presence of bacteria. Although today we can't know exactly what functions rhizosphere microbes served for early plants during the Silurian period 443-420 Myr ago, there is increasing evidence that diverse rhizobacteria, including several species of

Pseudomonas, may play a role in facilitating interactions between plants and mycorrhizal fungi. Early studies of AM fungi/bacteria co-culture determined that under nitrogen-deficient conditions the fungi could not infect the root without the presence of *Pseudomonas (Mosse, 1962)*. Species of *Glomus* were stimulated to infect the roots of *Zea mays* (corn) by *Pseudomonas putida,* and the *Pseudomonas* behaved synergistically with the *Glomus* to proliferate within the endorhizosphere. However, more experimentation was needed to determine the molecular drivers of this effect (Gryndler and Vosátka, 1996).

The term “Mycorrhizal helper bacteria” (MHB) was coined to describe the bacteria that appeared beneficial to the establishment of the fungus-plant mutualism (Garbaye, 1994) but this only described the phenomenon and not the mechanism; notably the bacteria appeared to show specificity to the fungi, but not the plant. Studies of the EcM basidiomycete fungi *Laccaria bicolor* that infects the roots of *Pseudotsuga menziesii* (Douglas Fir) demonstrated when a strain of MHB *Pseudomonas fluorescens* cultured from the *L. bicolor* sporocarp (the fruiting body) was applied to trees in both a greenhouse and a nursery setting it stimulated *L. bicolor* growth but steadily decreased until completely undetectable after 19 weeks. Interestingly, it was not culturable from the rhizoplane or endorhizosphere, or even from the resulting sporocarps (Frey-Klett, Pierrat and Garbaye, 1997). Later studies of fluorescent *Pseudomonas* reported an increase in both root and fungal biomass when co-inoculating Australian Acacias (*Acacia holosericea*) with the Basidiomycete fungus *Pisolithus alba (Founoune et al., 2002)*. With the implementation of NGS at lower cost, more groups have been exploring plant-fungi-bacteria mutualism and the effect of Pseudomonas spp. on spore germination and facilitating the establishment of the AMR-plant symbiosis (Giovannini *et al.*, 2020).

The apparent time limit on the viability of the *Pseudomonas* lead to the authors suggesting three working hypotheses for the “helper effect”: that MHB facilitate recognition between the plant and EcM, that the MHB stimulate the root to accept the EcM, or that the MHB directly stimulates the growth of the EcM prior to contact with the root. The authors contended that, based on their observations, the last of these hypotheses was the most supported and therefore plant growth promotion may be an indirect effect of the interaction between the MHB and the EcM. Other groups have observed *Pseudomonas* among the culturable endobacteria found in multiple EcM morphotypes associated with *Pinus sylvestris* (Scots Pine), and some of these endobacteria preferentially utilized fungal sugars over plant sugars (Izumi *et al.*, 2006). Taken together, it is clear that *Pseudomonas*, as with many diverse taxa of rhizobacteria, are influential in not only the host response, but in mediating interactions between the host and other members of the rhizosphere.

1.1.4. *Pseudomonas* has a massive pangenome

In 2005 the term “pan-genome” was coined to describe the variations found among pathogenic isolates of *Streptococcus agalactiae (Tettelin et al., 2005)*. The *Pseudomonas* pangenome, or the sum of all genes present in all known strains within the genus, is enormous and grows with each strain discovered. There are three essential categories to describe any pangenome and these include “core”, “accessory”, and “singleton” genes. Core genes are present in generally 90-100% of strains analyzed, whereas the accessory (or shell) genes are present in multiple strains, and singleton genes are unique to a single strain. Concerning *Pseudomonas,* there are likely tens of thousands of genes for each species; for example, there are over 30,497 core genes in one study of nine *Pseudomonas putida* genomes with an additional 200-1000 unique accessory genes per genome (Udaondo *et al.*, 2016). Another pangenome study of *Pseudomonas aeruginosa* with 181 strains observed that only 15% of the 16,820 genes were found in 100% of strains (Mosquera-Rendón *et al.*, 2016). When the number of strains or genomes included in a pangenome analysis increases the number of core genes is reduced proportionally, especially in species with large genomes or generalist lifestyles that favor the acquisition of horizontally-transferred genes (Popa and Dagan, 2011). The acquisition of new genes using the main three mechanisms of genetic transfer are transformation, transduction, and conjugation; the success of the first mechanism often depends on the native competency of a bacteria to uptake and integrate exogenous DNA into itself, the second requires susceptibility to a lysogenic bacteriophage, and the last entails physical contact between two cells to facilitate the transfer of DNA such as plasmids or to initiate homologous recombination (Ochman, Lawrence and Groisman, 2000).

As discussed at length above, the process of disentangling prokaryotic taxonomy and phylogeny is difficult at the best of times. Yet with pangenomics it is just as essential to have a clear understanding of what constitutes a “species”. Phylogenetic approaches to delineating species from each other typically expect monophyletic groups (Mallet, 2008). However, as again mentioned above, multilocus sequence alignments indicate *Pseudomonas* is not monophyletic but contains species that group into multiple clades that share a common ancestor. Discerning genus membership is relatively straightforward by ANI comparisons but prokaryotic speciation is complicated by adaptive HGT in addition to population effects such as genome and gene-specific selective sweeps (Wiedenbeck and Cohan, 2011; Bendall *et al.*, 2016). Some argue that monophyly as a threshold to define species will not be universally applicable when each individual loci being compared may have potentially polyphyletic inheritance . However, as history has shown taxonomic classification by shared phenotypes isn’t the answer either; in an effort to reconcile phylogeny with taxonomy some argue that species definitions should reflect *exclusivity*, or phylogenetically related groups that are more closely related to each other on the whole when compared to others (Velasco, 2009).

This concept has been applied to pangenome analyses of 701 species of *Streptomyces* and 1,586 species of *Bacillus*. In this study it was observed that, even with high rates of HGT, vertically inherited genes (core, or single copy genes) contribute a stronger phylogenetic signal. However, this study also gave evidence that exclusivity exists on a continuum that makes it impossible to use phylogeny to objectively rank as one species vs another. There has yet to be discovered a non-phylogenetic method to assign completely unique traits to only one species out of all others, and therefore the act of defining a species is essentially fluid based on what criteria is being examined. The concept of exclusivity defines species as a group of strains that are comparatively more similar to each other on a whole-genome level, not as a concrete taxonomic rank based simply on single-copy gene inheritance (Wright and Baum, 2018).

Consistent definitions of what constitutes a species informs how one measures parameters relevant to population genetics, and thus how inferences are made regarding how selection and drift affects organisms with relatively small genomes and variable effective population sizes (N_e), i.e. bacteria. Unlike in higher eukaryotes where large genomes tend to have accumulated many non-coding sequences over evolutionary time, prokaryote genomes generally are only inflated with “useful” accessory genes due to the need for balance between overall genome size and the pressure to “streamline” energy expenditure through purifying selection of pseudogenes (Sela, Wolf and Koonin, 2016). When comparing regulatory networks of gene transcription factors (TFs) between eukaryotes and prokaryotes, unlike in eukaryotes were a single gene could be regulated by multiple redundant transcription factors that may be distantly located from the start site, prokaryotic genes are subject to fewer redundant TFs that in turn reduce the overall need for long intergenic regions (Molina and van Nimwegen, 2008).

Several other factors potentially contribute to the diversity and size of different bacterial pangenomes. The drift-barrier model postulates that the larger the N_e of a species, the less genetic drift and thus less pressure to delete genes with neutral mutations. Also, larger population sizes favor contact and therefore HGT between phylogenetically similar organisms. As N_e shrinks, drift increases and genes with only modest fitness benefits will eventually be purged (Lynch *et al.*, 2016). This model suggests that species with a larger N_e will maintain larger pangenomes due to the retention of accessory genes (why can’t I cite this book for fucks sake)

Early attempts to quantify bacterial pangenomes were often focused on clinically relevant, culturable isolates of pathogenic strains. The formation of biofilms to enhance virulence is known to facilitate HGT among related strains and species (Molin and Tolker-Nielsen, 2003) and host pressure incentivizes conjugation and transformation which drives an increase in the accessory genome of the resident species. This effect is referred to as the distributed-genome hypothesis where the accessory pangenome of a species is substantially larger than within any individual genome, which acts as a pool of diversity that benefits the population as whole (Hiller *et al.*, 2007; Ehrlich *et al.*, 2010). However,

this hypothesis was developed to help explain how some bacteria use HGT to increase pathogenesis; although initially applicable to biofilm-producing bacteria, the distributed-genome model is considered relatively universal to all taxa with high N_e (Lapierre and Gogarten, 2009).

In nature there are genera that are widely distributed and have the ability to occupy multiple niches. Intuitively, it tracks that these large populations would be genomically diverse and the question arose of how well the distributed-genome model would predict pangenome sizes of taxa with distinct subpopulations and varying phenotypes. Predicting the scope of pangenomes requires accounting for the frequency of gene gain, loss, gene flow, and rates of drift that are hard to estimate when the taxa is large and diverse. Another method for performing quantitative predictions of pangenomes involves creating a “true” organismal tree using single-copy core genes, observing the pangenome of a subset of a group (genus or species), and using this infinitely many genes (IMG) model, extrapolating gene frequencies based on population size (Baumdicker, Hess and Pfaffelhuber, 2012).

The IMG model utilizes a standard neutral model of evolution that assumes no inherent fitness advantage to any gene, that all members of a population are equally fit to reproduce, and that new genes are introduced only once. In essence this model treats core genomes as essential for survival (i.e. permanent) and all accessory genes as dispensable; however, in nature there are traits originating from the accessory genome that confer situational fitness benefits. For example, studies of soil ecotypes observed enrichment of phosphatases and xylose utilization genes within a population of rhizosphere *Pseudomonas koreensis* versus those found in the nearby bulk soil (Lopes *et al.*, 2018). Despite the evidence for positive selection on individual genes within a specific population, the initial study postulating the IMG model used 22 geographically diverse *Prochlorococcus* genomes and observed that only a small fraction (~1%) of all genes in the pangenome had significant different frequencies (Baumdicker, Hess and Pfaffelhuber, 2012).

A study conducted with the express purpose of validating the IMG model using 172 complete *Bacillus* genomes suggests that this method is generally robust when genes are classified in one of three ways: the essential class which are found in every genome, the slow class of accessory genes that persist but can be slowly lost as genomes diverge, and a fast class that represent genes that are rapidly appear and are rapidly lost. Recently acquired genes (whether through HGT, mutation, lysogenic infection, or conjugation) within any specific genome are almost entirely fast class and are often singletons. Slow class genes are vertically inherited with varying frequency within the population, implying that the trait they influence is beneficial in at least one circumstance. Under the IMG model the projected distribution of fast, slow, and essential genes within any given pangenome is calculated based on the phylogenetic tree of the genomes used. This model considers pangenomes inherently open, or always capable of new genes

appearing either de novo or through HGT from outside populations (Collins and Higgs, 2012). There's no guarantee any specific accessory gene will persist, and strains exist in nature that have yet to be observed. Yet there exists a massive range of niches, lifestyles, competency, genome size, among other properties, between all prokaryotes. Therefore any method that attempts to predict pangenome size or gene frequency must take phylogeny, ecology, and sample size into account.

1.2. *In silico* analysis of microbial diversity within rhizospheres and soils

1.2.1.16S rRNA amplicons vs shotgun metagenomics for taxonomic profiling

One of the main challenges in functional profiling of soils and rhizospheres aside from the inconsistencies inherent in *Pseudomonas* taxonomy as mentioned above, is the cost of high-throughput sequencing of metagenomes. Historically 16S rRNA sequencing has been the most robust method of profiling bacterial communities as the Earth Microbiome Project optimized an Illumina protocol for targeting the V4 region of the variable 16S SSU rRNA for high-throughput processing of environmental samples (Thompson *et al.*, 2017). Methodological consistency between groups, datasets, and habitats is beneficial for reproducibility, but there are many considerations to account for when analyzing the sequence data produced in any experiment. For rhizospheres, moisture and texture can dramatically affect the amount of soil adhering to roots during sample collection and this can alter relative abundance profiles of the microbial taxa detected. Methods for genomic DNA extraction can introduce biases caused by the efficiency of cell lysis (mechanical vs chemical), as well as any number of other factors.

Overall however, the biggest considerations for 16S rRNA profile analysis is the limitation of using a partial sequence of one, albeit highly conserved, gene to infer taxonomy. The bioinformatic method PICRUSt was developed to infer metagenomes from 16S rRNA sequencing using reference genomes to predict community function; this software is on a second iteration with increased protein orthology database coverage (Douglas *et al.*, 2020). Inferring species or strain level taxonomy using 16S rRNA is difficult if not impossible for genera of bacteria that have a high level of recent speciation, such as *Pseudomonas*. Specifically, the V4 region of one strain of *Pseudomonas* can be identical to potentially dozens of other *Pseudomonas* species.

Shotgun metagenomics uses non-targeted sequencing to generate libraries containing whole genomes of all the microbes present. Under ideal circumstances the workflow of a metagenomic analysis begins with sample collection, followed by genomic DNA extraction. This is the step where the process diverges between long read and short read sequencing methods, especially for long read sequencing where the desired gDNA

size may need to reach >10K bp. Regardless of the sequencing length used the data generated typically, after quality control preprocessing, undergoes several profiling methods: read-based taxonomic profiling, assembly-based profiling, or read-based metabolic profiling (Quince *et al.*, 2017). Read-based taxonomic profiling relies on mapping the metagenomic sequences to a database of phylogenetic marker genes and provides an output of detected taxa. Assembly-based profiling first assembles the metagenomic reads into contigs then proceeds with taxonomic and functional profiling using databases of marker genes or protein sequences using annotated contigs. Lastly, read-based metagenomic profiling maps reads directly to annotated genes from a database with the intention of predicting metabolic pathways. In the latter case, unlike with assembly-based profiling that may assign taxonomy to each assembled contig, mapping reads to genes can provide insight to the metabolic capacity of the whole microbial community, as opposed to individual taxa. As will all of these methods, the outcome is a taxonomic profile and a collection of contigs and/or genes to infer a functional profile for each sample. This is a very broad overview of metagenomic processing; there are numerous techniques that enable the target enrichment of specific taxa of interest, or to isolate rare community members.

1.2.2. The benefits & limitations of metagenomes for bacterial data mining

The appeal of shotgun metagenomics over single locus amplicon sequencing to taxonomically profile complex communities has only increased over time as sequencing costs have decreased, along with the advancement of bioinformatic methods of phylogenetic marker classification. The method chosen to profile microbial communities at species level in a culture-independent manner will depend on factors such as the complexity of the community being studied, the inherent diversity of the taxa of interest, and accessibility of the computational resources necessary to process the metagenomic reads. The more complex the sample in diversity, or varied in individual taxa relative abundances, requires adequate sequencing depth (number of reads per metagenomic sequence). To profile low-abundance taxa often high sequencing depth is required, especially to capture the virome or to profile specific genes (Zaheer *et al.*, 2018).

A major limitation of shotgun metagenomic analysis of complex communities is the inability to link the presence of specific taxa or genes to ecologically relevant functional profiles. In order to connect these aspects, the WGS data should supplement experimental phenotypic evidence such as stable isotope probing (SIP), synthetic communities, comparisons between treatment groups against controls, or other methods of hypothesis testing (Kalyuzhnaya *et al.*, 2008). Stable isotope probing is a powerful technique that utilizes incorporating stable-isotope-labeled carbon, nitrogen, oxygen, sulfur, or hydrogen into microbial food webs and tracking their assimilation by specific taxa by isopycnic centrifugation and mass spectrometry (Neufeld *et al.*, 2007). Labeled

carbon is especially useful for investigating nutrient cycling in soils; a recent study tracked 1,286 bacterial taxa capable of assimilating labeled carbon in agricultural soils over 48 days and observed low phylogenetic conservation among bacteria. The conclusion was that carbon cycling within soils is not well predicted by bacterial phylogeny but predicted life-history; they hypothesize this may help explain the geographic variability observed among taxa in similar habitats (Barnett *et al.*, 2021).

1.2.3. Common methods for investigating microbial diversity within metagenomes

There are several popular methods for profiling metagenomes depending on what computational resources are available and whether taxonomic profiles, functional profiles, or both, are desired. For taxonomic classification of metagenomic reads there is Krakentools, which is a suite of programs designed to classify either to a database (Kraken2) or a user-specified set of genomes (KrakenUniq), and a means to extrapolate the relative abundance of these taxa for each metagenome (Bracken) (Lu *et al.*, 2017; Breitwieser, Baker and Salzberg, 2018; Wood, Lu and Langmead, 2019). Other options include MetaPhlAn3 which uses a pre-built multi-kingdom reference database to assign taxonomy to metagenic reads and estimating organismal relative abundance (Beghini *et al.*, 2021), and MCP to accurately profile human fecal metagenomes (Parks *et al.*, 2021). While MCP uses GTDB taxonomy to classify reads and Krakentools allows the use of either GTDB or other pre-built databases, the popular microbiome analysis tool MEGAN uses NCBI taxonomy for read classification plus a functional classifier and DNA-to-protein alignment tool (Huson *et al.*, 2016). These are just a few of the most common tools, as they have been externally validated and are actively being maintained (as of this monograph).

Unlike taxonomic profiling which only gives insight to community structure, there are methods of assembling whole genomes from metagenomic reads. An increasing number of publicly available genomes are assembled from metagenomic reads. These metagenomically-assembled genomes (MAGs) have become a useful tool for investigating functional capacity and uncovering novel, sometimes unculturable, taxa within diverse environments. To assemble MAGs the metagenomic reads can either be mapped to a database of reference genomes, or the reads can be assembled de novo. In the former case a de novo assembly can be attempted on the unmapped reads.

An enormous study that sampled human microbiomes from multiple geographic locations (western and non-western) generated 154,723 MAGs and uncovered thousands of new species, many of which were observed in non-western populations, and improved the overall ability of human gut metagenomic reads to be mapped to a known species by over 87% (Pasolli *et al.*, 2019). Even more species were derived when 50,000 additional MAGs were generated from the human oral microbiome; of the species-level bins over 64% had no publicly available reference genome (Zhu *et al.*, 2021).

Soils are more taxonomically diverse than the human gut, and assembling MAGs from these habitats poses certain challenges. One of the earliest attempts to succeed in reconstructing genomes at strain level from metagenomes involved shotgun sequencing an acidophilic biofilm comprising a low-complexity community with little evidence of HGT between taxa (Tyson *et al.*, 2004). The challenge comes from the variability of relative abundance between taxa where the most highly abundant species will represent the lion's share of the metagenome reads, and the less abundant taxa would be more difficult to assemble unless the sequencing depth is increased. Another consideration is the degree of similarity between conserved regions of bacterial genomes at higher phylogenetic levels; without high enough coverage the assembly tool may struggle to distinguish between regions of high similarity within genomes of closely related species or genera. Lastly, soil and rhizosphere communities may contain species with high strain diversity, i.e. strain mixtures, which is especially computationally difficult to correctly bin contigs to genomes when the difference between them is <0.01 ANI (Nurk *et al.*, 2017). There is evidence that MAGs may under-represent both the core genome and accessory genome within any given genome, and common bioinformatic methods of determining contamination between taxa may underestimate how many genes are the result of binning errors (Meziti *et al.*, 2021).

As discussed above, assembling high quality genomes from metagenomes with highly variable relative abundances and diverse strain mixtures can be difficult. While the criteria for quality in isolate genomes is typically >97% completeness and <5% contamination, inclusion criteria for MAGs may depend largely on the downstream analysis they're contributing to. Unless the MAGs is destined to become a representative strain for a new species, MAGs with low contamination but also lower completion may be useful when examining the species pangeome within a metagenome. If the experiment does not require taxonomic resolution or whole genomes then assembling genes de novo and assigning taxonomy to the assembled proteins may be more helpful in uncovering community-level functional diversity. This method can be computationally intensive but has demonstrated an increased sensitivity compared to assembling contigs for MAGs (Steinegger, Mirdita and Söding, 2019).

Chapter 2

Secondary metabolite diversity between fluorescens lineage *Pseudomonas* suggest diverging strategies to survive intragenus competition within the rhizosphere

2.1 Aims

The first objective in this effort to characterize species-level diversity and pangenome content of these rhizosphere *Pseudomonas* was to generate high quality metagenomes from a selection of the remaining root samples. Next, using bulk soils collected from one field different labile carbon sources were applied to stimulate *Pseudomonas* growth *in vitro.* These enriched soils were used to culture *Pseudomonas* along with a selection of rhizospheres from 2015.

For any species detected via taxonomic profiling that lacked a representative isolate from this study, the publicly available reference genome was obtained to generate a phylogeny by MLSA of shared housekeeping genes. The goal with the phylogeny was to create a robust classification mechanism independent of taxonomy, instead to group these species by monophyletic clade with the reasoning that phenotypic similarity will be reflected in phylogeny.

To determine if these groupings are indicative of broad scale gene-content similarity when including the whole genome, a pangenome of all the *Pseudomonas* detected within the 2010 and 2015 rhizospheres was created. If these monophyletic groups were in fact measurably different from each other, then gene content in regard to predicted phenotypes, antibiotic resistance, and secondary metabolite production could potentially serve as context for the high abundance phenotype observed in 2010.

2.2 Methods

Rhizosphere sampling

Maize planting and sample collection were conducted as described in (Peiffer *et al.*, 2013) for the 2010 rhizosphere samples. The 2015 samples were planted and collected as per Walters and colleagues (Walters *et al.*, 2018). Briefly, in 2010, 27 inbred cultivars representing a genotypically diverse panel of maize were planted in a randomized block design at several locations across upstate New York at Musgrave Research Farm in Aurora at coordinates: 42°43'25.0"N 76°39'25.5"W; Ketola Organic Research Farm in Ithaca at 42.471968598071406, -76.43819497351113; Willet Dairy, Lansing at 42.6238111793627, -76.58472637350556 (supplemental table 1). For weeks 1-15 & 20 for Aurora and Lansing, and 2-15 & 20 for Ithaca, plants were chosen at random to avoid border effects. Each rhizosphere was sampled destructively by removal with a spade and collecting root with adherent soil roughly 5 cm from the base of the stem. The rhizosphere samples were transported on ice and stored at -80 ˚C until further processing. In 2015, a subset of genotypes planted in 2010 (B73, MO17, IL14H, CML277 & HP301) were planted and collected as per Peiffer et al. (Peiffer *et al.*, 2013) at the Aurora location. Sampling of the rhizosphere and bulk soil was performed weekly between weeks 4 to 12. Rhizospheres from cultivars B73, MO17, and IL14H grown at Aurora, Ithaca, and Lansing

in 2010, plus those grown in Aurora in 2015 were selected for WGS metagenomic analysis.

Substrate-induced respiration assay

Micro-organisms were extracted from each rhizosphere soil sample by washing 0.1 g of soils in 5 mL sterile phosphate buffered saline (PBS) with 10 % glycerol for 1 h with gentle rocking at room temperature. Then I isolated single Pseudomonas colonies by plating 100 µL of the wash liquid onto *Pseudomonas* Isolation Agar (BD Diagnostic Systems, Franklin Lakes, NJ) using a disposable inoculating loop.

In December 2015, 1.5 kg of silt loam soil (10 cm depth) was collected from the Musgrave Field Station field site in Aurora, New York (coordinates: 42°43'25.0"N 76°39'25.5"W). Two treatment groups were prepared, with one containing all carbons ^{13}C-labeled to permit stable isotope probing (DNA-SIP). The soil was prepared as per (Pepe-Ranney *et al.*, 2016). Briefly, the soil was dried shortly at 25 ˚C to permit sieving (2 mm), homogenized, with 10 g distributed to each 250 ml flask and sealed. To deplete the residual carbon within the soil headspace CO2 was measured every two days via GC-MS to determine the point at which microbial respiration plateaued to ensure residual labile carbon was exhausted, 17 days total. A single labile carbon substrate, either glucose, fructose, glutamic acid, or salicylic acid was suspended in murashige & skoog basal salt medium to attain a CN ratio of 1:13 with uniform moisture content. Each substrate was amended to individual microcosms at the concentration of 4 mg substrate g−1 d.w. soil. Microbial respiration was measured from sampling headspace CO2 via GCMS with no flushing between samplings at 12, 24, and 48 hours. For each timepoint a single microcosm of each substrate treatment was collected, the soil frozen at -80 °C and later processed for metagenomic sequencing as per the maize study.

Pseudomonas isolations

From each rhizosphere soil sample, 0.1 grams of soils were washed in 5 mL sterile phosphate buffered saline with 10% glycerol for 1 hour with gentle rocking at room temperature. 100 µL of the wash liquid was plated onto *Pseudomonas* Isolation Agar (BD Diagnostic Systems, Franklin Lakes, NJ) using a disposable inoculating loop. Individual colonies were picked and incubated in 1.2 ml LB broth overnight in 96 deep-well format. The resulting genomic DNA was purified with Mag-Bind beads as above and quantified using Picogreen (Invitrogen P7589).

Glycerol stocks were prepared by adding 200 ul of 50% glycerol to a 200 ul aliquot of the *Pseudomonas* culture for storage at -80 °C. The remaining volume of *Pseudomonas* culture segued directly into a DNA extraction with , followed by the Nextera LITE library preparation (Karasov *et al.*, 2018). These genomes were sequenced using the Illumina Hiseq3000 platform, 150 bp paired end. Selected isolates for long-read

sequencing were cultured at a volume of 10ml in LB overnight and high molecular weight DNA was obtained using the MagAttract HMW DNA kit (Qiagen 67563)

Rhizosphere gDNA extraction, NGS library preparation, and metagenome sequencing

Rhizosphere DNA extraction and purification was performed in 96 well plates using the MagAttract PowerSoil DNA Kit (Qiagen 27100-4-EP) with approximately 250 mg of root and attached soil. The genomic DNA was purified by 1:1 ratio of Mag-Bind TotalPure NGS magnetic beads and quantified via Picogreen. The short-read Illumina WGS libraries were prepared using the Illumina Nextera protocol (Caruccio, 2011) modified to accommodate smaller volumes. Size selection (450-650 bp range) was performed with a BluePippin (Sage Science, Beverly, MA) using a 1.5% agarose cassette with internal marker R2. The Illumina libraries were sequenced using 150 PE HiSeq 3000 with an average sequencing depth of 40x.

Pseudomonas isolates were prepared for long-read sequencing using the Qiagen MagAttract HMW DNA Kit (Qiagen 67563) and were size-selected using the BluePippin (0.75% agarose high-pass cassette and S1 marker) to obtain genomic DNA > 20 kb in size. The size-selected HMW DNA was prepared for long-read sequencing on the Oxford Nanopore MinION using the NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (E7180S) and the Oxford Nanopore Ligation Sequencing Kit (SQK-LSK109). Libraries were multiplexed in groups of 12 using the Native Barcoding Expansion 1-12 & 13-24 kits (EXP-NBD104 & EXP-NBD114).

Sequence processing and quality control

Raw metagenomic and genomic sequences were initially processed using bbtools/bbduk (Bushnell, 2014) and Skewer v0.2.2 (Jiang *et al.*, 2014) to remove adapters and filter reads based on PHRED score. Maize sequences were removed from the filtered reads using bbtools/bbmap by mapping reads to the complete B73 cultivar genome (B73 RefGen_v4). The quality of these trimmed and filtered reads were assessed with fastQC v0.11.7 (Andrews, 2010) and multiQC v1.5a (Ewels *et al.*, 2016).

Isolate genome assembly

The processed genomic reads were used to assemble draft genomes for each *Pseudomonas* isolates. Isolate genomes were assembled with SPAdes v3.14.1 (Bankevich *et al.*, 2012) for short reads. The Guppy algorithm v1.5.3 (Wick, Judd and Holt, 2019) was used to basecall the MinION long reads, and Unicycler v0.4.5 (Wick *et al.*, 2017) was used to perform a hybrid assembly with both the short and long reads. CheckM v1.2.0 (Parks *et al.*, 2015) was used to determine genome contamination, completeness, GC content, and genome size. Genomes with less than 97% completeness and greater than 5% contamination were excluded from downstream

analysis. To estimate read coverage, Nonpareil v3.3.1 (Rodriguez-R *et al.*, 2018) was used for the metagenomic sequences and Bandage v0.8.1 (Wick *et al.*, 2015) was used for assessing individual genome graph assembly quality.

Metagenome taxonomic profiling

Reads were classified using Kraken2 v2.1.2 (Wood, Lu and Langmead, 2019), and a Bayesian re-estimation of the species-level abundance of each sample was then performed using Bracken v2.2 (Lu *et al.*, 2017). The reads were mapped to the GTDB-r95 (Parks *et al.*, 2020) database for taxonomic identification.

The *Pseudomonas* species with a relative abundance above 0.0001% were included in the downstream analysis and phylogeny. For species with an isolate on hand derived from the original study, the 2015 season, or the microcosms, these were used as the representatives for the phylogeny. For all other species the GTDB representative species genome was obtained from RefSeq (O'Leary *et al.*, 2016), with taxonomy and corresponding RefSeq number included in supplemental table 4.

Quality control for each publicly available genome was conducted in the same manner as the isolate genomes (CheckM >97% completeness, <5% contamination). KrakenUniq v0.6 (Breitwieser, Baker and Salzberg, 2018) was used to measure the precise relative abundance of these specific strains within each metagenome. The command line parameters were as follows: build: --kmer-len 31 --minimizer-len 15, map: --hll-precision 12. The percent abundance of the *Pseudomonas* genomes with fewer than 1000 kmers per sample was converted to zero to compensate for the lack of sensitivity of KrakenUniq below that threshold.

Inference of Pseudomonas phylogeny

A phylogenetic tree of the abundance-filtered *Pseudomonas* was created using Anvi'o interactive interface v7 (Eren *et al.*, 2021) phylogenomic workflow (--hmm-source Bacteria_71), using Fastree v2.1.10 (Price, Dehal and Arkin, 2010) approximate maximum likelihood method to infer the phylogeny based on the alignment of 71 single copy core genes present in 90% of the genomes. The tree was rooted using *Shewanella oneidensis MR-1* (GCF_000146165.2) as the outgroup. Pruning of tree tips was achieved using ETE3 toolkit v3.1.2 (Huerta-Cepas, Serra and Bork, 2016).

Pangenome analysis of Pseudomonas strains

All *Pseudomonas* included in the phylogeny were included in the pangenome. The Anvi'o pangenome workflow was used, with a minbit heuristic of 0.5 and an MCL inflation of 6 (van Dongen and Abreu-Goodger, 2012; Eren *et al.*, 2021). Each amino acid sequence was queried via NCBI blast. Partial gene calls were included. The full pangenome was used to ordinate the Jaccard distances for each genome (R package vegan v2.5-7, ggplot v3.3.5)

Genome annotation & gene content profiling

The genomes above the relative abundance threshold were annotated using Prokka v1.14.5 (Seemann, 2014). Culture-independent phenotyping with Traitar3 v1.1.2 (Weimann *et al.*, 2016) and biosynthetic gene cluster annotation using Abricate v1.0.1 (Seemann, 2015). Antibiotic and secondary metabolite biosynthetic gene clusters (BGCs) were annotated with Antismash v5.1.2 (Blin *et al.*, 2019). These BGCs were clustered using BiG-SLiCE v1.1.1 (Kautsar *et al.*, 2021) with the euclidean distance threshold for each gene cluster family set at 300.

2.3 Results

2.3.1. Temporal dynamics of *Pseudomonas* within the rhizosphere & microcosms

To characterize the diversity of *Pseudomonas* species within the 2010 & 2015 maize rhizospheres, and the response of the 2015 bulk soil response to the addition of labile carbons substrates constitutive of maize root exudates, whole shotgun metagenomes were profiled from biological replicates in storage remaining from a previous study. The 2010 and 2015 field studies included three maize genotypes: B73, IL14H, and MO17, of the stiff stalk, sweet, and non-stiff stalk heterotic groups, respectively. In 2010 these rhizospheres were sampled at 4, 6, 8, 11, and 14 weeks post emergence as they were grown in three fields in NY: Aurora, Ithaca, and Lansing. By contrast, in 2015 only Aurora was planted and the maize rhizospheres were sampled at weeks 4-5, 8, and 12. On completion of the 2015 growing season, soils were collected from Aurora to perform a substrate-induced growth assay to measure *Pseudomonas* response to carbon sources found in maize root exudates. Soils were treated with either glucose, fructose, glutamic acid, or salicylic acid, and metagenomes were produced at 12, 24, and 48 hours post treatment.

This approach was chosen specifically to capture the species-level diversity within *Pseudomonas* and to utilize GTDB as a means of consistent bacterial taxonomy. The rhizospheres obtained from 2010 were collected during a previous 16S rRNA gene profiling study and therefore constrained to unused biological replicates remaining for the time points selected. The 2015 field study was conducted as a follow-up to the 2010 field study to determine if the relative abundance of *Pseudomonas* observed was typical for maize grown in that location. The rationale for applying root exudate-analogous substrates was to determine if *Pseudomonas* growth could be stimulated in the absence of maize.

Whole shotgun metagenomes were taxonomically classified via GTDB; the number of K-mers identified were used to calculate the relative abundance of rhizosphere bacteria and archaea first at class level, then species. Each dataset is presented

separately with the 2010 field season with the *Pseudomonas* bloom, followed by the 2015 field season follow up (figure 2), then the substrate microcosm experiment using 2015 bulk soil (figure 3)

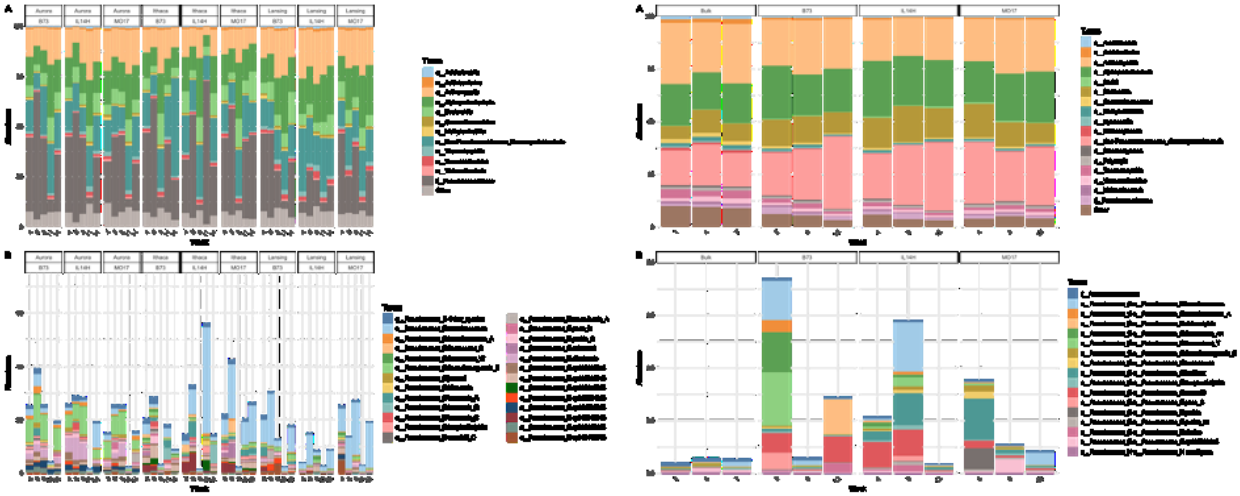


Figure 2 - Temporal dynamics of the genus Pseudomonas in the 2010 maize rhizospheres for each field sampled by week

A) Relative abundance barplot of all GTDB-detected bacterial taxa observed at class level during the 2010 field study, with Pseudomonadaceae at family level and all other phyla collapsed by class. The annotations above the bars are grouped by field, with B73, IL14H, and MO17 representing each maize genotype. The X-axis is in order of ascending collection week. B) Relative abundance of the genus Pseudomonas from panel A collapsed by class level, with the most highly abundant species (>0.2% mean abundance) shown individually. C) The relative abundance of the GTDB-detected bacterial taxa observed during the 2015 field study, again collapsed at class level. Only Aurora is represented in this study, with grouping by maize genotype and bulk soil. The X-axis is again in order of ascending collection week. D) Relative abundance of the genus Pseudomonas from panel C collapsed by GTDB group, with the most highly abundant species (>0.01% mean abundance) shown individually.

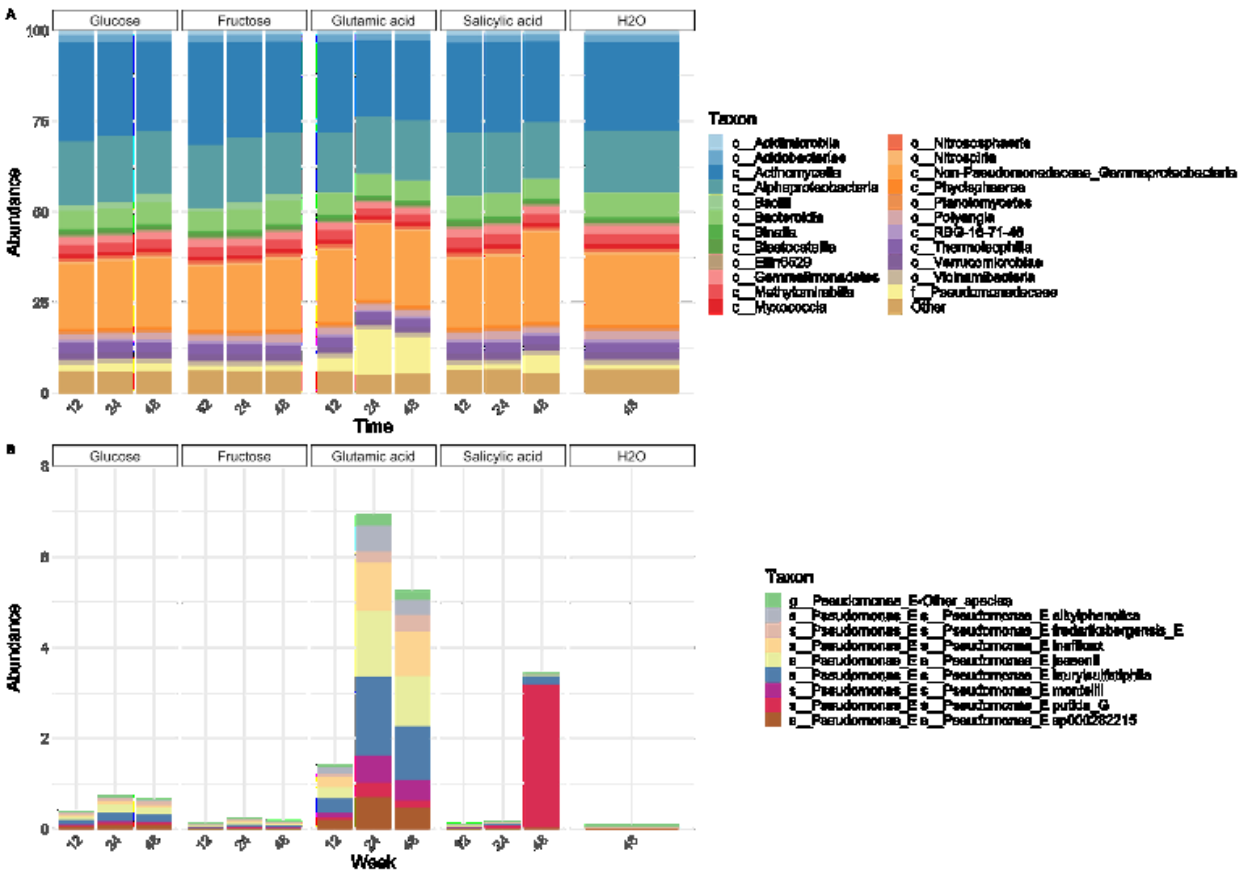


Figure 3 - Taxonomic distribution of labile carbon-treated soil microcosms by substrate and incubation time

A) Relative abundance barplot of all GTDB-detected bacterial taxa observed at class level within each microcosm with the carbon substrate addition annotated above and the incubation time on the X-axis. The H2O control was measured only after 48 hours. B) Relative abundance of the genus Pseudomonas from panel A collapsed by GTDB group, with the most highly abundant species (>0.05% mean abundance) shown individually.

To determine the abundance of individual *Pseudomonas* species within the maize rhizospheres of 2010, each metagenome (n=45) was taxonomically profiled. The percentage of reads remaining unmapped varied between 19% and 57%, (median 48%). For each metagenome within the 2010 field study, Gammaproteobacteria dominated, with other abundant classes including Alphaproteobacteia, Actinomycetia, and Bacteroidia. Of the Gammaproteobacteia, the genus *Pseudomonas* was the most highly abundant. Comparing the taxonomic profiles between fields, a distinct reduction of *Pseudomonas* was observed relative to other non-Gammaproteobacteria across each maize genotype at week 11 in Aurora and Lansing, and at week 8 in Ithaca (figure 2). This drop in *Pseudomonas* coincided with a relative increase in Alphaproteobacteria, and in some cases Actinomycetia. Across all fields, the species richness did not vary significantly by week. However, the Shannon Index demonstrated significant differences in species

evenness between weeks 4 and 6, and weeks 6 and 8 (figure 18). The relative abundance of *Pseudomonas* species varied in composition by each field, with the most highly abundant species at Aurora consisting of *P*. *brassicacearum, P*. *brassicacearum A, P*. *frederiksbergensis E, P*. *silesiensis, P*. *sp000282215, P. sp004786035,* and *P. sp900187425.* The most highly abundant species in Ithaca consisted of *P*. *brassicacearum, P*. *fluorescens C, P*. *koreensis A, P*. *sp000282315,* and *P*. *sp900005815.* The most highly abundant species for Lansing consisted of *P*. *brassicacearum, P*. *frederiksbergensis E, P*. *sp000282315, P. sp002754355,* and P. *sp900187605.*

To determine the abundance of individual *Pseudomonas* species within the maize rhizospheres of 2015, each metagenome (n=12) were also taxonomically profiled. The percentage of unmapped reads varied between 52% and 67% (median 62%). In this season Pseudomonadaceae increased only slightly over the bulk soil, with no pattern of increasing or decreasing abundance by collection week. For example, *Pseudomonas* was most abundant within the maize genotype B73 at week 5, while *Pseudomonas* was most abundant for genotype IL14H at week 8. The proportion of Pseudomonadaceae did not increase or decrease relative to other Gammaproteobacteria, nor other bacterial classes including Actinomycetia, Alphaproteobacteria, or Bacteroidia. The taxa most abundant in the bulk soil compared to the rhizospheres included Acidomicrobiia, Acidobacteriae, Gemmatimonadetes, Methylomirabilia, and Thermoleophilia. Overall the relative abundance of *Pseudomonas* for each time point was not consistent, and the species composition was not consistent between maize genotypes. The most abundant *Pseudomonas* species for the genotype B73 included *P. brassicacearum, P*. *frederiksbergensis E, P. chlororaphis, P*. *fluorescens AA, P. fluorescens T, P*. *monteilii,* and *P*. *poae B.* The most abundant species for the genotype IL14H included *P*. *brassicacearum, P*. *fluorescens T, P*. *inefficax,* P. *monteilii, P*. *poae B,* and *P*. *simiae.* The most abundant species for the genotype MO17 included *P. brassicacearum, P*. *frederiksbergensis E, P*. *hunanensis, P*. *inefficax, P*. *monteilii, P*. *putida,* and *P*. *sp000282215.*

To determine whether simple carbon amendments analogous to common maize root exudates added to soil could increase the relative abundance of *Pseudomonas*, bulk soil was collected at the Aurora field for a substrate-induced respiration assay. Sources of labile carbon were selected per (Kraffczyk, Trolldenier and Beringer, 1984) and included glutamic acid, glucose, salicylic acid, and fructose. Each microcosm (n=15) received one carbon source. The concentration of carbon dioxide in the headspace of each microcosm was cumulatively measured and while the basal salt control did not cause an increase in CO2 production, the glucose, fructose, and glutamic acid exhibited a logistic increase until saturation at 24 hours; whereas salicylic acid lagged by 12 hours (supplemental figure 1). The soils were collected for DNA extraction and metagenomic

sequencing at incubation intervals of 12, 24, and 48 hours with the water+basal salt negative control being collected at 48 hours.

The taxonomic profile of the microcosms was attained in the same manner as the field study metagenomes. When compared to simply rewetting the soil, adding glutamic acid increased the relative abundance of Pseudomonadacae with a peak at 12 hours. Glucose resulted in a modest increase, whereas salicylic acid yielded an increase in abundance only at 48 hours. Pseudomonadaceae did not respond significantly to the addition of fructose (figure 3A). In general, the magnitude of the increases observed was similar to that of the 2015 field study.

The species that attained a relative abundance >0.05% within the microcosms were more similar to the most abundant *Pseudomonas* of the 2015 rhizospheres compared to the 2010 rhizospheres, as would be expected due to the soils being collected that same year. The glutamic acid experienced a robust increase in abundance from specifically *P. alkylphenolica*, *P. frederiksbergensis E*, *P. inefficax*, *P. jessenii*, P. *laurylsulphatiphila*, *P. monteilii, P. putida G,* and *P. sp000282215*. The glucose microcosms exhibited a much lower increase in *Pseudomonas* relative abundance, but a relatively uniform increase in these same species. The salicylic acid microcosms essentially exhibited no response until the 48-hour time point, which was dominated exclusively by P. *putida G* (figure 3B). This is to be expected, considering the type strain for this species of *Pseudomonas* was found to use acetylsalicylic acid as a carbon source (Parales *et al.*, 2017).

A discerning eye would immediately consider that overabundance of *Pseudomonas* within the 2010 season maize rhizospheres as potentially a contamination issue, specifically due to its ubiquity within the environment and how the genus *Pseudomonas* is often identified as a contaminant of DNA extraction kit reagents (Salter *et al.*, 2014; Sheik *et al.*, 2018). Blank extractions were not particularly illuminating due to the high-throughput method of DNA extraction in 96-well plate format and its overall propensity for well-to-well contamination. All care was taken to avoid this, but the use of a pipetting robot does not eliminate this possibility. Despite bulk soil samples being unavailable for metagenomic sequencing, the 16S rRNA sequencing of the biological replicates processed during the Walters et al. study indicated the bulk soil collected and processed in parallel did not demonstrate an increase in *Pseudomonas* as was observed with rhizospheres.

When examining the abundances observed using 16S rRNA sequencing, it is worth remembering that species within the genus *Pseudomonas* often have between 1-7 copies of the 16S rRNA gene; however, attempting to normalize taxonomic relative abundance to 16S rRNA gene copy number is difficult, especially for such a diverse taxa as *Pseudomonas* (Louca, Doebeli and Parfrey, 2018). The method to taxonomically profile the metagenomes used a kmer approach which did not contain such a bias. Although the exact proportions of the *Pseudomonas* relative abundance per rhizosphere

did not mirror the 16S rRNA profiles, their overall abundance relative to all other taxa remains disproportionately high.

Bulk soils collected in the winter of 2015 at Aurora were used to determine if the species of highly abundant rhizosphere *Pseudomonas* would respond to the addition of labile carbon added directly to soils. The metagenomes of these microcosms of individual carbon substrates including glucose, fructose, glutamic acid, and salicylic acid were profiled after fixed incubation periods. *Pseudomonas* increased in relative abundance in response to glutamic acid, salicylic acid, and glucose with the most abundant species being among the most abundant for the rhizospheres of the Aurora that year.

In summation, the metagenomes of the maize rhizospheres of the 2010 and 2015 maize seasons were taxonomically profiled to determine the abundance of *Pseudomonas* relative to all other detected taxa and to achieve resolution at species level. For the 2010 season, *Pseudomonas* was the most highly abundant genera across all time points excluding week 11 in Aurora and Lansing, and week 8 in Ithaca. The distribution and relative abundance of individual species varied by field, where Ithaca and Lansing were dominated by P. *brassicacearum*, whereas the distribution of highly abundant *Pseudomonas* was more diverse within Aurora. Revisiting Aurora in 2015 the overall abundance of *Pseudomonas* was an order of magnitude lower than in 2010, yet P. *brassicacearum* was still one of the most highly abundant taxa. However, several species that were not above the threshold of 0.0001% relatively abundant in 2010 were detected above threshold for 2015, and among these include P. *inefficax* and P. *monteilii.*

2.3.2. Phylogeny of abundant maize rhizosphere *Pseudomonas*

The number of Pseudomonas species detected via Kracken2, approximately 500 per rhizosphere regardless of year, required a cutoff threshold going forward with the computational limits of later analyzing the Pseudomonas pangenome. Only Pseudomonas species with a relative abundance above the cutoff (0.0001%) were used to construct a phylogenetic tree. Both maize seasons, in addition to the microcosms, were used with the Alteromonadales Gammaproteobacteria *Shewanella oneidensis* MR-1 as the outgroup.

This “main tree” consisted of 389 Pseudomonas species, with each being assigned a “group” based on the local alignments of single-copy genes present in at least 90% of all genomes via the Anvi’o v7 bacteria 71 database (Eren *et al.*, 2021). The genomes used were obtained by downloading the Genome Taxonomy Database (GTDB) species representative from NCBI RefSeq or Genbank (O’Leary *et al.*, 2016); with the exception of those species within our culture collection which were isolated directly from maize rhizospheres or soils during either of the seasons observerved.

Other studies of comparative Pseudomonas phylogenetics organized the genus into main paraphyletic groups and subgroups based on multilocus sequence analysis

(MLSA) with those of the fluorescens lineage being the most populous (Gomila *et al.*, 2015; Garrido-Sanz *et al.*, 2016; Hesse *et al.*, 2018). The taxonomic composition of the maize rhizospheres included in this study were grouped in accordance with the literature, and this structure informed the phylogenetic group assignments of Pseudomonas taxa with unresolved or inconsistent nomenclature according to GTDB.

The phylogeny of these abundant maize rhizosphere *Pseudomonas* enables the comparison of how different species were distributed across each year and maize genotype, which is beneficial for assessing broad trends in abundance between these variables. The genus *Pseudomonas* contains a massive accessory pangenome (Koehorst *et al.*, 2016; Winsor *et al.*, 2016) which makes discrimination between closely related taxa difficult. Using a multi locus sequence analysis of 71 single-copy genes (Eren *et al.*, 2021) as opposed to 16S rRNA or a limited selection of housekeeping genes enabled the most accurate tree within the bioinformatic resources available. The use of GTDB as the primary taxonomic classification method provided an essential means to compare between closely related taxa via a highly curated database.

The genomes of *Pseudomonas* species that were above the minimum relative abundance threshold for the 2010, 2015 seasons and the microcosms were used to create a phylogenetic tree. Annotated to this phylogeny were basic genome qualities such as size in base pairs, GC content,and the number of contigs in the assembly (figure 4). The 10% most abundant *Pseudomonas* were highlighted to enable comparisons between taxa with varying magnitudes of abundance (figure 6). The full phylogeny with 2010 *Pseudomonas* abundances by field was too large to visualize in a publication but was still useful to get a sense for the scale of the bloom (figure 5).

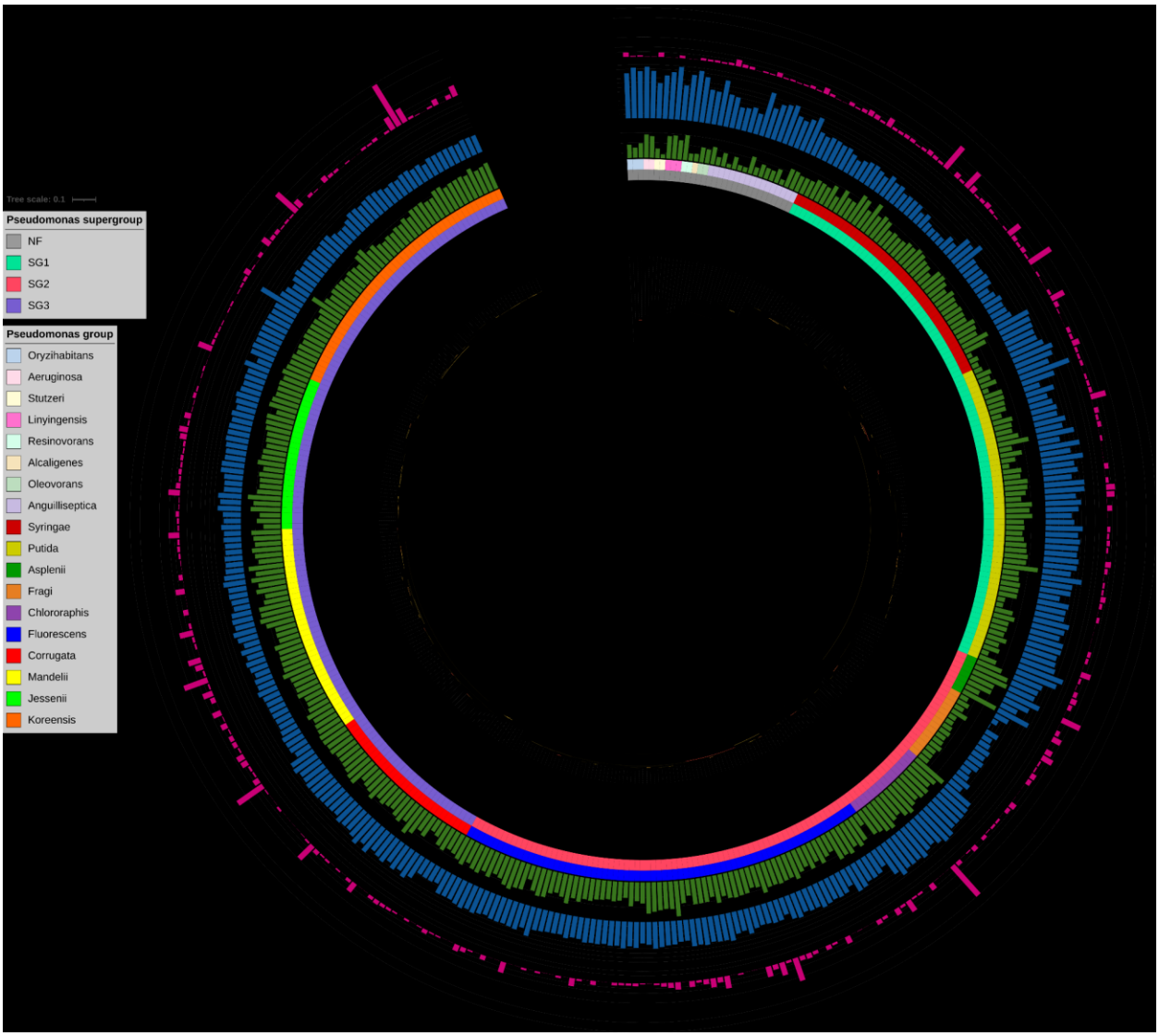


Figure 4 - 2010 and 2015 Pseudomonas phylogeny of 71 conserved single copy genes for each genotype, with annotations for Fluorescens supergroup, group, genome size (scale begins at 4 Mb) represented by the green bars, GC content (scale begins at 55%) represented by the blue bars, and the number of contigs per each genome assembly represented by the red bars.

Figure 5 - Combined 2010-2015 Pseudomonas phylogeny of species above the relative abundance per metagenome threshold, with red indicating Aurora, blue Ithaca, and green Lansing.

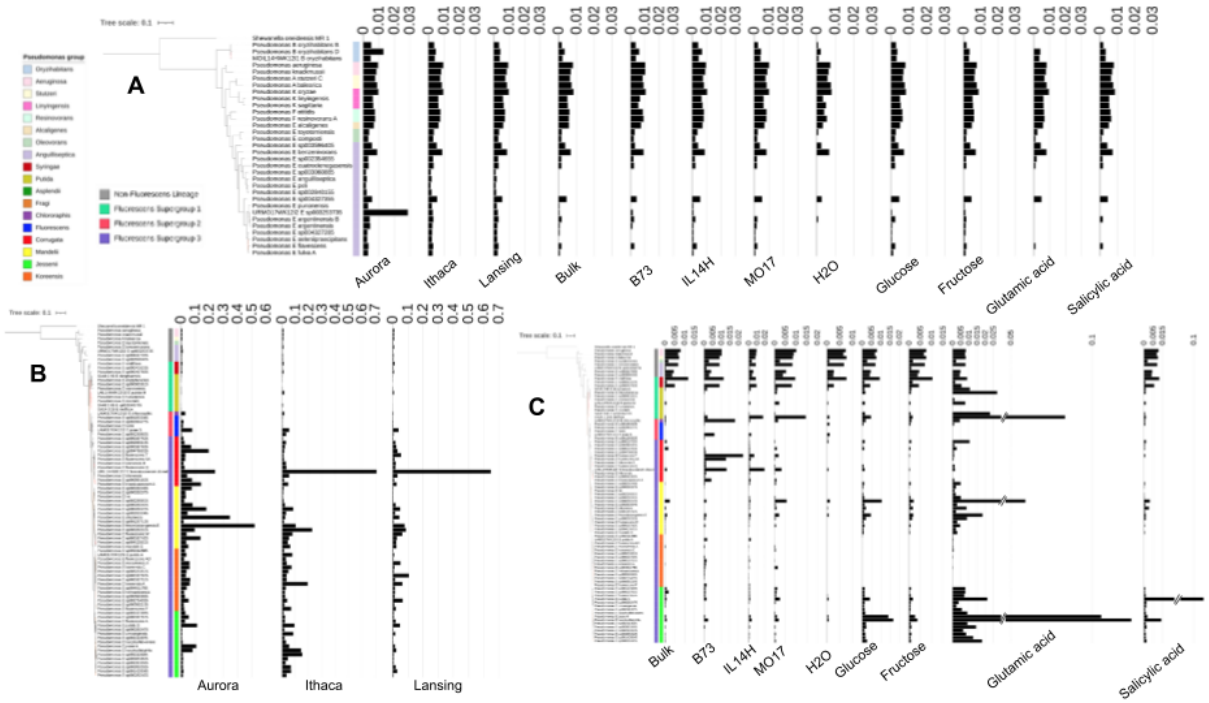


Figure 6 - Phylogeny of the top 10% most abundant Pseudomonas using 71 conserved single copy genes for seasons 2010, 2015, and the microcosms
A) Pseudomonas phylogeny of 71 conserved single copy genes for the non-fluorescens lineage within each rhizosphere and microcosm, with the 1+log relative abundance for each species. Colored bars represent each Pseudomona group. Tree scale indicates substitutions per site using the Jones-Taylor-Thorton ML model via Fasttree 2.1. Bootstrap values are indicated by branch color, with the minimum (0) as red, mid (0.5) as yellow, and maximum (1) as black. B) The Pseudomonas phylogeny includes the fluorescens lineage with the 1+log relative abundances for the 2010 field rhizospheres. The inner bar represents Pseudomonas supergroup and the outer bar the group. C) The same phylogeny as panel B with the 2015 rhizospheres and substrate microcosm 1+log relative abundance.

To explore the genomic and phylogenetic architecture of these maize rhizosphere *Pseudomonas*, all species present above a relative abundance threshold were used to generate a master phylogeny. A total of 394 *Pseudomonas* species were included, spanning the 2010 and 2015 metagenomes, and also the microcosms. These genomes were selected from the GTDB as representative genomes and acquired from the NCBI, or selected from the microcosm isolate culture collection.

This phylogeny includes numerous taxonomically uncharacterized species within NCBI and spans multiple *Pseudomonas* species groups. To better understand the relationship between these species and any potential phenotypic or evolutionary similarities they might possess, each genome was assigned placement to a phylogenetic group based on the work of (Hesse *et al.*, 2018) and (Garrido-Sanz *et al.*, 2016) whom

assessed Pseudomonas phylogeny using conserved housekeeping genes. The phylogeny of the 394 *Pseudomonas* revealed that these species fell into 18 of the established *Pseudomonas* groups, with the vast majority being of the fluorescens lineage. Basal to the outgroup were the non-Fluorescens species, comprised of Oryzihabitans, Aeruginosa, Stutzeri, Linyingensis, Resinovorans, Alcaligenes, Oleovorans, and Anguilliseptica. Of these basal *Pseudomonas*, there were only a small number of species per group, with the exception of Anguilliseptica (17 species). These groups are comparably the most phylogenetically divergent from each other, and the few species contained in each group tend to be more diverse.

The remainder of the phylogeny consisted of fluorescens lineage *Pseudomonas* species, classified into three main “supergroups”. The first supergroup consists of the groups Syringae and Putida (SG1) with a total of 104 species, the second Aspendii, Fragi, Chlororaphis, and Fluorescens (SG2) with 112 species, and the third with Corrugata, Mandelii, Jessenii, and Koreensis (SG3) with 146 species. The number of species contained within each fluorescens lineage group varies from 7 to 76, from Aspendii to Fluorescens respectively.

When the number of species is restricted to the top 10% most abundant taxa for the 2010 Aurora & 2015 Aurora fields and microcosms, the topology of the resulting tree remains largely the same as the full tree (Figure 6B). However, groups Corrugata and Fluorescens exchange positions with each other, as do Jessenii and Koreensis. When the mean abundances of these *Pseudomonas* are applied to each tip per each maize genotype, it is observed that there are substantial differences between not only the rhizosphere collection year, but between the rhizospheres and microcosms.

Despite the total abundance of *Pseudomonas* being an order of magnitude higher during 2010, several species above the abundance threshold remain proportionally abundant in 2015 also. Notably, the non-Fluorescens taxa share a similar relative abundance between both 2010 and 2015. Although bulk soil metagenomes were not available for the 2010 season, the taxa that are present in the bulk soils for 2015 were similar in abundance, as also seen with the water controls of the microcosms. This suggests the basal taxa are primarily soil *Pseudomonas* that persist regardless of year or host status. The *Pseudomonas* taxa that appear to be highly abundant within the rhizosphere between years are largely members of the Corrugata group, such as P. *brassicacearum* and P. *brassicacearum* A, P. *fluorescens AA* and P. *fluorescens T*. Adjacent to the Corrugata group is the one highly abundant species of the Chlororaphis group, P. *chlororaphis* shared between years, along with P. *poae B* from the Fluorescens group. However, the overall trend between the *Pseudomonas* profiles of these two field studies appears to be the “supergroup divide”. The 2010 rhizospheres exhibit the highest abundance within supergroup 3, while the 2015 rhizospheres are predominantly supergroup 1, specifically the Putida group.

The microcosms unsurprisingly reflect a distribution more in line with the 2015 rhizospheres, which is likely due to the close proximity of their sample collection. The relative abundance of *Pseudomonas* remains under 0.5% for each individual taxa, but there are some observable differences between the simple carbohydrates of the glucose and fructose versus the glutamic acid microcosms. The former there was a modest but uniform increase from within the Putida group, with the most dramatic increases relative to bulk occurring within the Mandelii and Jessenii groups. Several taxa within the glutamic acid microcosms experienced a more robust response to this carbon source, specifically P. *alkylphenolica* and P. *inefficax* from the Putida group, P. *sp000282215* from the Mandelii group, and P. *laurylsulfatiphila* from the Jessenii group. Taken as a whole, the 2015 rhizospheres, along with the microcosms, experienced a modest but discernible increase in Putida and some Jessenii group *Pseudomonas* taxa, whereas the 2010 rhizospheres observed a dramatic and highly disproportionate increase in Corrugata and Mandelii group *Pseudomonas* taxa. The *Pseudomonas* species most highly abundant and pervasive between the rhizosphere metagenomes was P. *brassicacearum*, yet that species did not respond as strongly to the addition of labile carbon directly to the soil.

The choice to constrain the genomes of this phylogeny to only those above the 0.0001% relative abundance threshold for each year biases the tree heavily toward fluorescens lineage *Pseudomonas*, while excluding other closely related species that either were either under the threshold or were simply not present in these particular fields in Upstate New York at the time. The aim of this study was not to comment on *Pseudomonas* taxonomy as a whole as with Hesse et al., but to report what *Pseudomonas* were present within the context of the 2010 bloom.

When comparing the phylogeny of these rhizosphere *Pseudomonas* with each individual genome's size and GC content, the Fragi phylogenetic group has consistently a smaller genome relative to other fluorescens lineage *Pseudomonas.* Among all groups, GC content varies; Aeruginosa lineage *Pseudomonas* appear to have proportionally higher GC content compared to fluorescens lineage taxa with the exception of the Putida and Chlororaphis groups.

Among the 10% most abundant *Pseudomonas*, the Aeruginosa lineage groups are consistently observed across each year and within the microcosms. They are proportionally lower than the fluorescens lineage taxa and appear to be ubiquitous with stable abundance. The groups that represented the most highly abundant species for the 2010 season were Corrugata, Mandelii, Jessenii, and Koreensis, whereas for 2015 it was primarily Corrugata with a few species from Putida and Syringae, and one species from Fluorescens. The groups that proliferated within the glucose microcoms were Putida, Mandelii, and especially Koreensis. This was also the case for glutamic acid but with a higher magnitude of abundance.

In summation, mapping the abundances of *Pseudomonas* to each condition (season, genotype, or microcosm) suggested a phylogenetic signal for abundance within

the fluorescens lineage. However, field is a strong determinant of the local inoculant for maize (Walters *et al.*, 2018) and this is seen by the varying taxonomic profiles between fields in 2010. Among the microcosms the most consistent group to increase in relative abundance were the Putida, Mandelii, and Koreensis groups. This appears to be an intermediate response between what was observed within the rhizospheres in 2010 vs 2015.

2.3.3. Gene content ordination of *Pseudomonas* by group

To confirm and expand upon the phylogenetic groups identified via tree topology, the pangenome of these *Pseudomonas* genomes was clustered via Jaccard distance for gene cluster presence/absence. Total gene cluster families per genome were produced using the Anvi'o pangenome pipeline (Eren *et al.*, 2021). Genomes included for the pangenome analysis were all *Pseudomonas* with relative abundances >0.0001% for each field for each year (Aurora, Ithaca, Lansing for 2010 and Aurora 2015).

This method enables comparisons based on whole-genome content, including accessory genes which vary between individuals within closely related species. Mapping the number of gene cluster families (GCFs) per genome onto the *Pseudomonas* cladogram with 2010 abundances is to determine qualitative patterns in gene content.

The cladogram of each *Pseudomonas* above the abundance threshold is annotated with the total pangenome GCFs for each genome, overlaid with the 2010 relative abundances (figure 7). Jaccard distances for each genome, based on GCF presence-absence, were compared between groups, supergroups, and abundance phenotypes by year (figure 8).

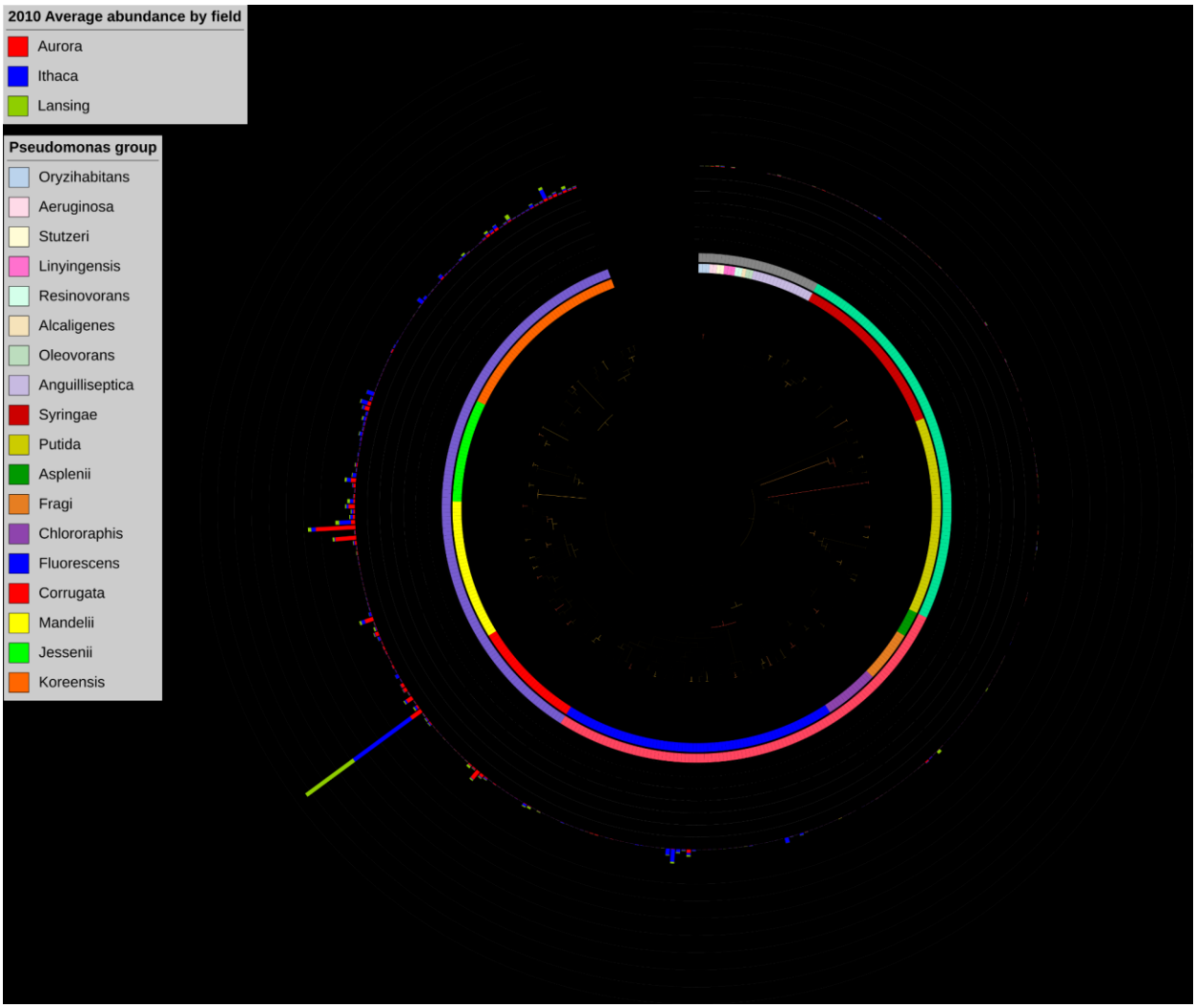


Figure 7 - Pseudomonas cladogram with pangenome gene cluster families (GCFs) versus 2010 genome relative abundance. The black bar is the number of GCFs per genome and the colored bars indicate proportional abundances for each field during the 2010 season.

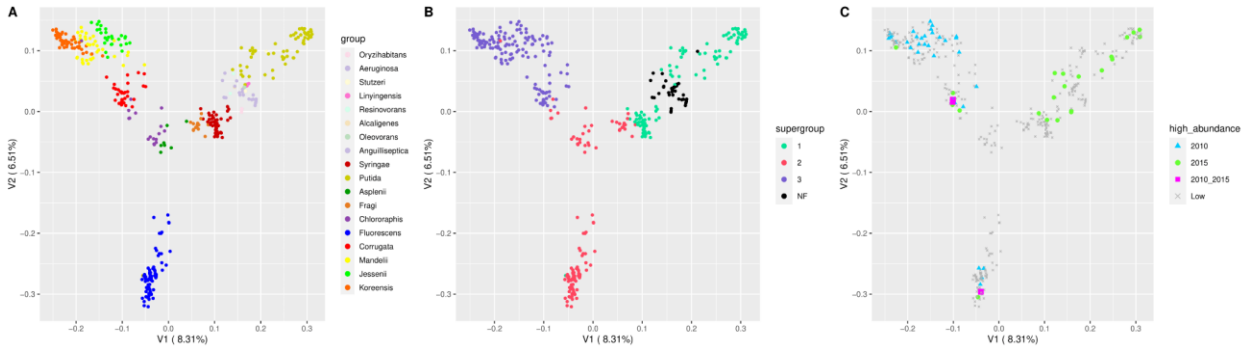


Figure 8 - PCoA of Jaccard distances for each genome based on gene cluster presence/absence

All panels: Genomes are ordinated top two dimensions, with percent of variation

explained on each axis. A) All genomes ordinated by dimension V1 (x axis) vs dimension V2 (y axis) colored by group. B) The same plot as previously colored by fluorescens lineage supergroup. C) The same plot as previous, colored by genome abundance with cyan indicating the genome being within the top ten percent relative to all genomes.

To compare the overall gene content of each species and how it relates to the phylogenetic groupings presented herein, a pan-genome analysis was conducted with the 365 *Pseudomonas* species above the abundance threshold in the 2010 metagenomes. Shared gene clusters were identified and a presence/absence matrix was used to calculate the Jaccard distances for each genome, and the resulting principal coordinate analysis (PCoA) ordination was colored by *Pseudomonas* group (figure 8A). The contribution of the first (V1) and second (V2) principal coordinates are 8.14% and 6.7% respectively, yet there is distinct clustering based on group. The non-fluorescens lineage *Pseudomonas* overlap largely with *Pseudomonas* supergroup 1, specifically the Putida group. The Syringae group was distinct with little to no overlap with either the non-Fluorescens or the Putida group.

Among the fluorescens lineage groups, supergroup 2 was unique in regard to the Fluorescens group clustering independent of the other members of this supergroup. The other groups appeared to bridge supergroup 1 to 3; the Chlororaphis group exhibiting the most diffuse clustering, with the closest proximity to both the Corrugata and Asplenii groups. Within supergroup 3, the Corrugata group was largely separate, while the Jessenii, Mandelii, and Koreensis groups experienced substantial overlap. Concerning the Fluorescens, not only did it cluster independently, but it also included several “promiscuous” genomes that cluster with other groups. P. *poae* and P. *paralactis* clustered within supergroup 1; P. *sp900187645* within supergroup 2 but adjacent to the Chlororaphis or Asplenii groups; and P. *mandelii B* clustered within the Mandelii group adjacent to Koreensis. None of these genomes appeared to be chimeric based on their GTDB assignment or CheckM contamination; however, the assembly of P. *mandelii B* had 251 contigs, whereas the other three each comprised <57 contigs. The only other positional outlier was P. *parafulva* B, which is a Putida group species that clustered within the Fluorescens group.

Taking the same ordination and coloring by supergroup, aside from a single outlier, the fluorescens lineage groups retained clustering by supergroup (figure 8B). Supergroup 1 and the non-fluorescens lineage were the most similar based on gene cluster presence/absence, which directly reflects the inferred phylogenetic relationship. Supergroup 2 seemed to be positioned between the Fluorescens group and SG1 and SG3, with that separation being driven primarily by the V2 principal component. When the top 10% most abundant *Pseudomonas* species for 2010 are mapped onto this ordination, these taxa appeared to cluster predominantly with supergroup 3 (figure 8C). There were some highly abundant species that clustered with supergroup 2, but only among the Fluorescens group.

These data represent exclusively the comparisons made between each other, and it is heavily biased toward fluorescens lineage genomes. The clustering of each group and supergroup is influenced by the number of genomes but also the parameters used to produce the pangeome. This analysis used an MCL inflation parameter of 6, which affects the granularity of clustering for each gene cluster family. This setting was chosen to balance the number of genomes vs discriminating between gene clusters of closely related species. The clustering by group and supergroup indicate the MCL inflation is adequate to discriminate at the broad phylogenetic levels. However, the Fluoresens lineage group Putida clusters very diffusely in comparison. Although this appears to be a consequence of phylogenetic diversity within the group as evidenced by the cladogram, especially within the *P. putida_LTKBVUJE* clade. There are two genomes that cluster distantly from their own group with very topographically divergent clades, and in the absence of instances of undetected contamination from closely related organisms it may indicate substantial HGT between those strains. Confirmation of this could include species level pangenome ordinations and genome alignments to assess if the shared sequences are strain or species specific.

In summary, the comparison of each phylogenetic group by pangenome GCFs shared suggests there are significant gene content distinctions between groups and supergroups. At the supergroup level, the non-fluorescens lineage species are more similar to each other and to Fluorescens supergroup 1. Notably, the distinction between each Fluorescens supergroup is slightly more nebulous compared to each group within; specifically with SG1 and SG2. The group Fluorescens clusters distinct from all groups, while group Putida also clusters distantly. The most derived groups, Koreensis, Jessenii, and Mandelii, cluster more closely than with group Corrugata. The genomes that appear to cluster outside their group may indicate occurrences of HGT, or genome contamination undetected by conventional QC methods. High abundance genomes for 2010 fall exclusively within SG2 and SG3, while 2015 encompasses all supergroups.

2.3.4. *Pseudomonas* phenotype predictions by group and abundance

To determine if the phylogenetic groupings are predictive of phenotype, in silico analysis of each genome was performed. Despite the magnitude of the *Pseudomonas* accessory and singleton pangenome, when dimensionally reduced there is evidence of phylogenetic signal at a whole genome level. To determine the core and group-specific phenotypes present the program Traitar was used (Weimann *et al.*, 2016). To account for the uneven number of genomes per group, the number of genomes with the phenotype were divided by the total number of genomes to give a proportion for that group. To investigate the phenotypic differences between *Pseudomonas* of varying relative abundances, the phenotypes present or absent in the ten percent most abundant and least abundant genomes were compared regardless of group placement.

There are known phenotypes that define the genus *Pseudomonas* (gram

negative, aerobic, flagellated, rod morphology, non-spore forming) but there are numerous accessory-genome related phenotypes specific to individual strains. These potential phenotypes that may suggest functional variations between phylogenetic groups.

To determine if similarities exist within groups that extend to the supergroup level, a heat map of the predicted phenotypes per group demonstrates a broad overview of pervasive traits (figure 9). At the individual genome level, high and low abundance *Pseudomonas* for the Aurora field heatmaps of phenotype presence-absence helps illustrate phylogenetic differences (figure 10).

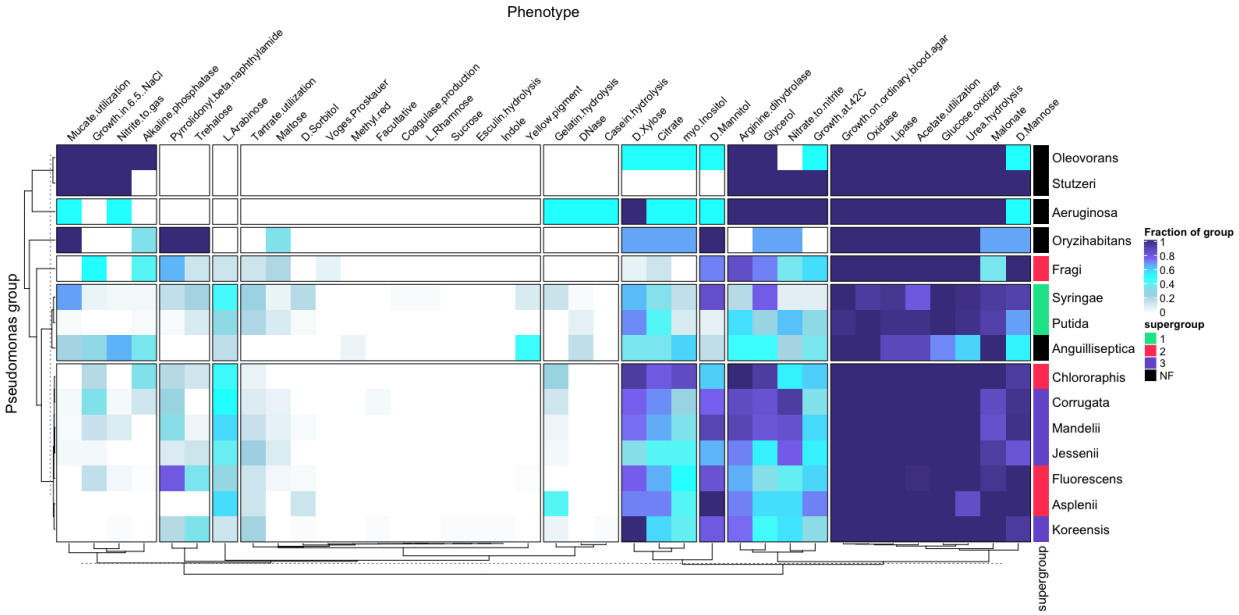
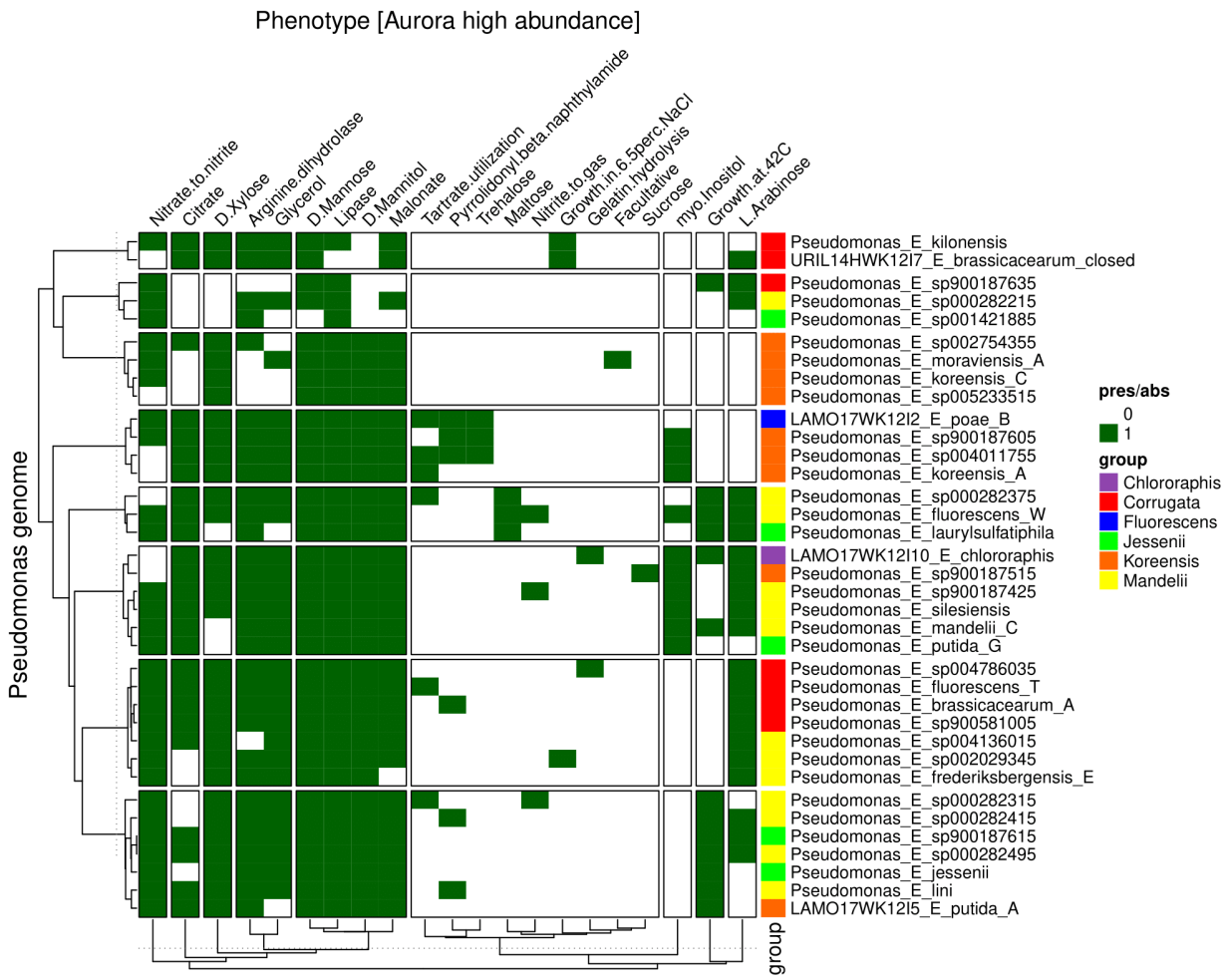


Figure 9 - Heatmap of phenotype traits shared between each groups with core traits excluded

The color of each trait indicates the fraction of each group where that trait is present, with black indicating that trait was detected in all genomes of that group. The groups are clustered via Spearman correlation and traits by k-means. Fluorescens supergroup is annotated at right, with “NF” indicating non-Fluorescens groups.


Phenotype [Aurora high abundance]
Nitrate.to.nitrite
Citrate
D.Xylose
Arginine.dihydrolase
Glycerol
D.Mannose
Lipase
D.Mannitol
Malonate
Tartrate.utilization
Pyrrolidonyl.beta.naphthylamide
Trehalose
Maltose
Nitrite.to.gas
Growth.in.6.5perc.NaCl
Gelatin.hydrolysis
Facultative
Sucrose
myo.Inositol
Growth.at.42C
L.Arabinose
Pseudomonas genome
Pseudomonas_E_kilonensis
URIL14HWK12I7_E_brassicacearum_closed
Pseudomonas_E_sp900187635
Pseudomonas_E_sp000282215
Pseudomonas_E_sp001421885
Pseudomonas_E_sp002754355
Pseudomonas_E_moraviensis_A
Pseudomonas_E_koreensis_C
Pseudomonas_E_sp005233515
LAMO17WK12I2_E_poae_B
Pseudomonas_E_sp900187605
Pseudomonas_E_sp004011755
Pseudomonas_E_koreensis_A
Pseudomonas_E_sp000282375
Pseudomonas_E_fluorescens_W
Pseudomonas_E_laurylsulfatiphila
LAMO17WK12I10_E_chlororaphis
Pseudomonas_E_sp900187515
Pseudomonas_E_sp900187425
Pseudomonas_E_silesiensis
Pseudomonas_E_mandelii_C
Pseudomonas_E_putida_G
Pseudomonas_E_sp004786035
Pseudomonas_E_fluorescens_T
Pseudomonas_E_brassicacearum_A
Pseudomonas_E_sp900581005
Pseudomonas_E_sp004136015
Pseudomonas_E_sp002029345
Pseudomonas_E_frederiksbergensis_E
Pseudomonas_E_sp000282315
Pseudomonas_E_sp000282415
Pseudomonas_E_sp900187615
Pseudomonas_E_sp000282495
Pseudomonas_E_jessenii
Pseudomonas_E_lini
LAMO17WK12I5_E_putida_A
group
pres/abs
0
1
group
Chlororaphis
Corrugata
Fluorescens
Jessenii
Koreensis
Mandelii

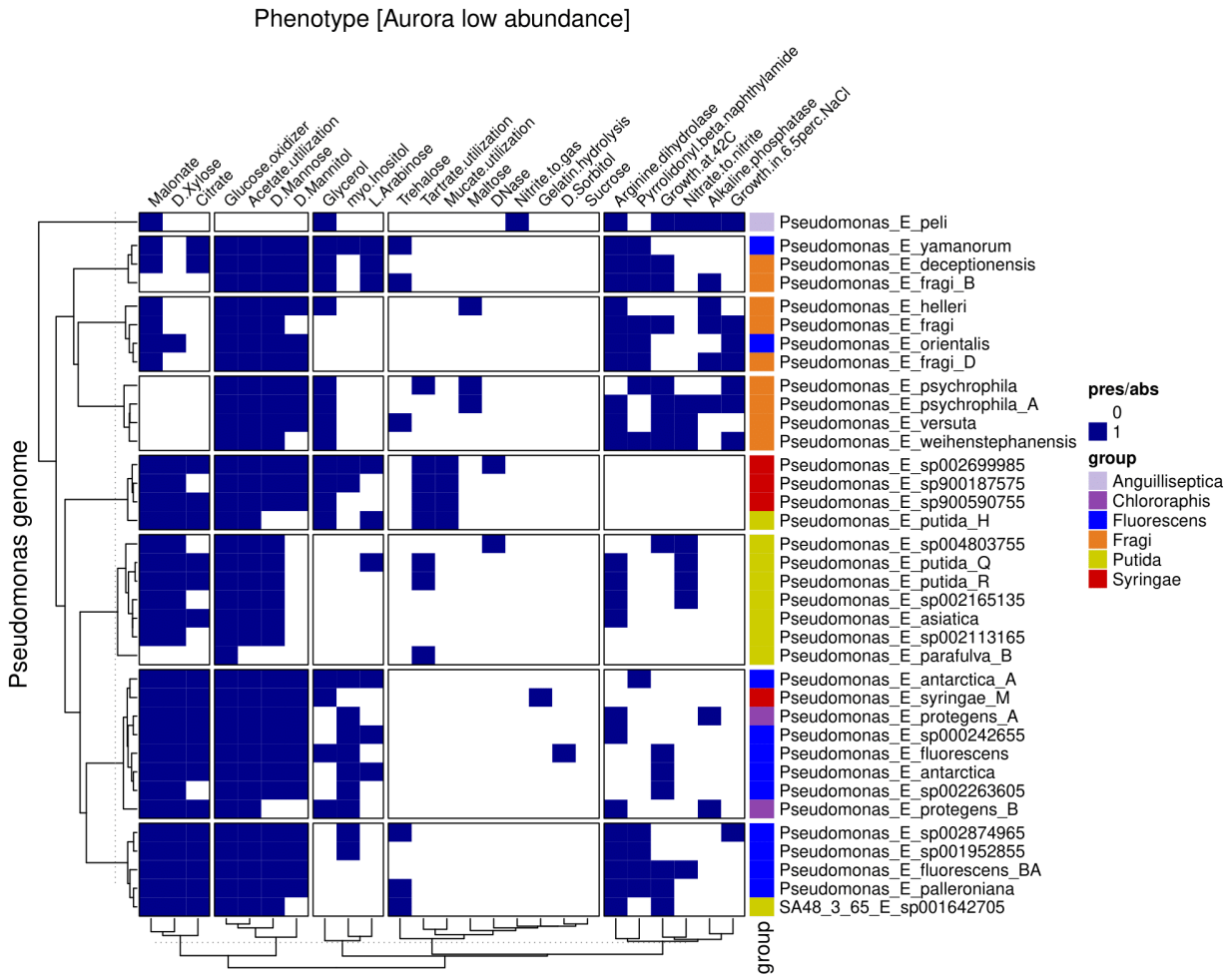


Figure 10 - Heatmap of phenotype traits by Pseudomonas genome relative abundance with core phenotypes excluded

A) Phenotype presence/absence heatmap of the top 10% **most** abundant genomes from the Aurora field, with groups are clustered via Spearman correlation and traits by k-means. B) The top 10% **least** abundant genomes from the Aurora field, again with groups are clustered via Spearman correlation and traits by k-means.

To assess the potential phenotypes present among the groups of *Pseudomonas*, each genome was examined for genomic classifiers used to predict various microbial traits. To control for the unequal numbers of individual species per group, each trait was summed and divided by the number of species in that group to determine the fraction of that group possessing that predicted trait (Figure 9). There were eleven phenotypes shared by all genomes: motile, growth on MacConkey agar, growth in KCN, gram negative, susceptible to colistin and polymyxin, cellobiose metabolism, catalase positive, bile susceptible, aerobic, and morphologically bacillus or coccobacillus. After excluding these core traits and clustering the rest via Spearman correlation, the non-Fluorescens

groups were independent of each other with the exception of the Anguilliseptica group that clustered with SG1 and the Fragi group. These placements were less instructive due to these groups having the fewest number of species represented, save the Anguilliseptica group, which had 17 species. The two groups within SG1 cluster together, with the Syringae group possessing slightly more diversity regarding the uncommon traits such as gelatin hydrolysis, d-sorbitol metabolism, and muciate utilization. Aside from the Fragi group, SG2 and SG33 clustered together, thus indicating a similar repertoire of phenotypes comparatively. *Pseudomonas* from the groups Chlororaphis, Corrugata, and Mandelii were highly predicted to convert nitrate to nitrite, and can utilize arginine as well as glycerol as carbon sources. Half of these groups were predicted to grow in 6.5N NaCl. Of all the SG2 and SG3 taxa, the Corrugata group had the most species predicted to convert nitrate to nitrite, whereas the Stutzeri, Oleovorans, and Anguilliseptica groups were predicted to convert nitrite to gas. *Pseudomonas* of SG2 and SG3 are also highly predicted to utilize sugars such as myo-inositol, D-mannitol, D-xylose, L-arabinose, and the carboxylic acids citrate and malonate (supplemental table 2).

When comparing the presence/absence of traits for only the 10% most abundant species within the 2010 Aurora field, traits associated with metabolism of sugars and carboxylic acids were highly prevalent, as well as the conversion of nitrate to nitrite (Figure 10). The 10% least abundant species for the 2010 Aurora field share much of the same capacity to metabolize sugars, but fewer species were predicted to metabolize arginine (Figure 10).

These results were derived exclusively with the program Traitar and this presumes this computational method is free of false positives or negatives. To mitigate this, only phenotypes detected with both models (phypat + PGL) were counted; the authors recommend using this consensus to avoid false positives, and with the validation dataset they observed microaccuracies of 87.5-87.9%. The differences in the number of genomes between non-fluorescens lineage and fluorescens lineage groups may suggest more consistent phenotypes within groups with a small number of genomes than may be representative in nature. Furthermore, the presence-absence heatmaps for each abundance criteria are especially sensitive to false negatives by Traitar, which may suggest the absence of shared pathways that potentially define a group such as nitrate reduction.

In summary, despite the sensitivity considerations when using Traitar, distinctions in predicted phenotypes were observable at supergroup level. At the group level, there was substantial variation in the proportion of each phenotype present in any given group. Core phenotypes consisted of aerobic respiration, motility, gram-negativity, bacillus/coccobacillus morphology, bile-susceptibility, catalase-positive, growth on cellobiose, growth on KCN, growth on MacConkey agar, and colistin-polymyxin susceptibility. Phenotypes likely to be core but due to possible false negatives or incomplete genome assemblies and absent in one or two genomes include growth on

ordinary blood agar, oxidase-positive (as expected for *Pseudomonas*), and possibly lipase and acetate utilization. Overall, the profile of phenotypes reflected similarity within supergroups more so than per group. When comparing the phenotypes between high and low abundance *Pseudomonas,* there is not a strong descrimination other than high abundance taxa having a more diverse repertoire of sugar metabolism phenotypes and a slightly higher prevalence of nitrate reduction to gas.

2.3.5. Antimicrobial resistance and virulence gene content ordination of all *Pseudomonas* genomes

Another aspect of the *Pseudomonas* accessory pangenome that may inform potential functional differences between phylogenetic groups include antimicrobial resistance (AMR) and virulence gene profiles. Again, in silico mining of AMR gene content using the program ABRicate (Seemann, 2015) which screened each genome using multiple AMR databases (Chen *et al.*, 2005; Pal *et al.*, 2014; Alcock *et al.*, 2020) was used to compare the rate of annotation of genes across each group and supergroup.

This method aggregates the results of several AMR-specific databases, allowing for broad-scale comparison between groups and supergroups. Examining the total number of AMR genes between groups versus their respective number of GCFs will help determine if the proportional differences in AMR genes are reflective of whole genome content. Comparing each genome based on AMR presence-absence will orient the groups based on the specific genes they possess, and will indicate broad similarities or differences between groups.

First, the total number of pangenome GCFs per genome for each group provides context to the proportion of genes mapped to each database of ABRIcate (figure 11). The Jaccard distance for each genome by group and supergroup suggests distinct AMR profiles, with the number of annotations to the BacMet2 database driving the signal of the V1 principal coordinate (figure 12).

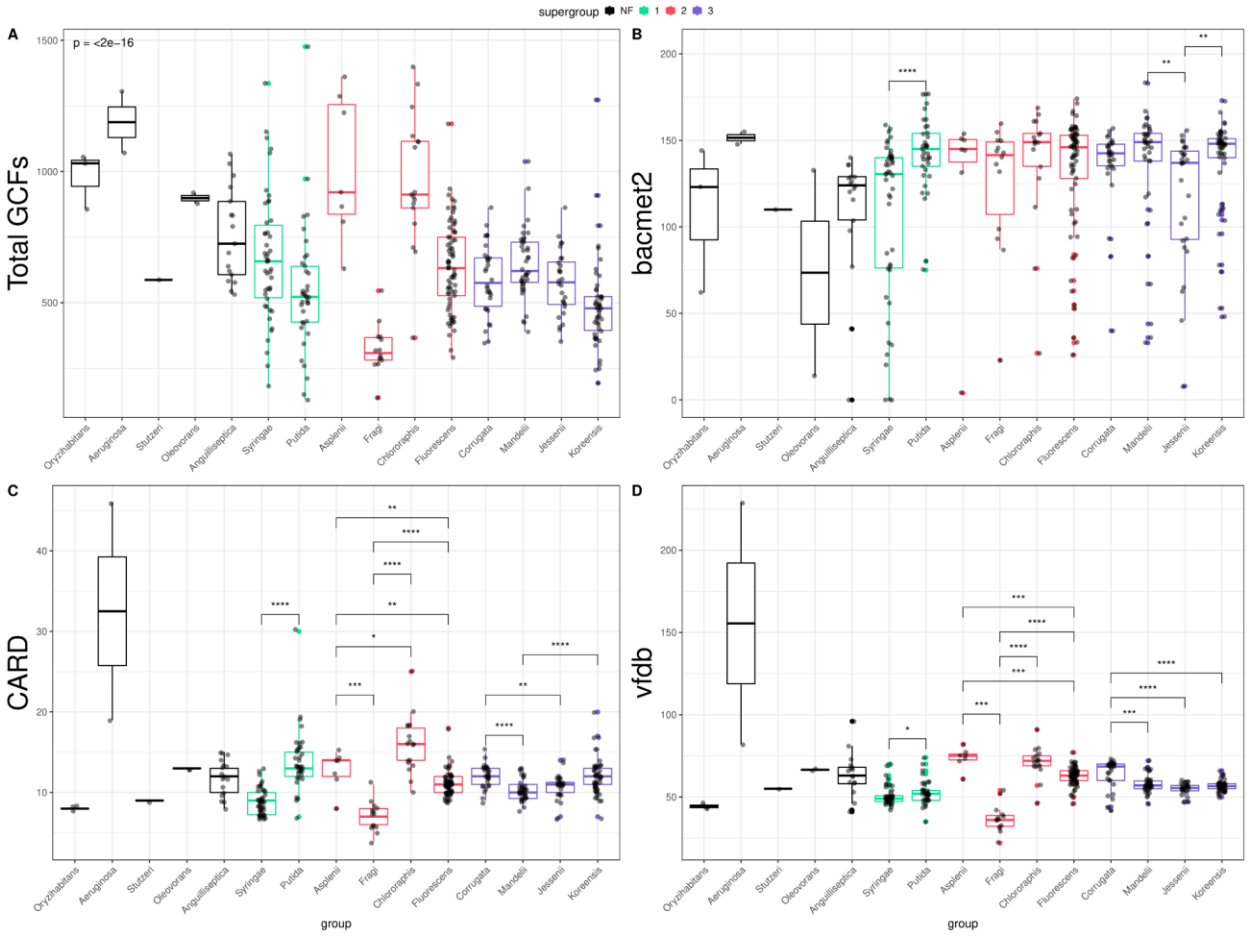


Figure 11 - Plotting the number of sequence hits to antimicrobial resistance and virulence databases per genome by phylogenetic group

A) Total pangenome GCFs for each *Pseudomonas* group colored by supergroup, with Bonferroni-corrected Kruskal-Wallis test. B) The number of sequences per genome mapping with >80% coverage to entries within the BacMet2 antibacterial biocide and metal resistance genes database by group and colored by supergroup. P-value signifiers within supergroup only. C) Sequences per genome mapping with the same coverage to the Comprehensive Antibiotic Resistance Database (CARD). D) Sequences per genome mapping with the same coverage to the virulence factor database (VFDB). P-value significance = *0.05, **0.01, ***0.001, ****0.0001

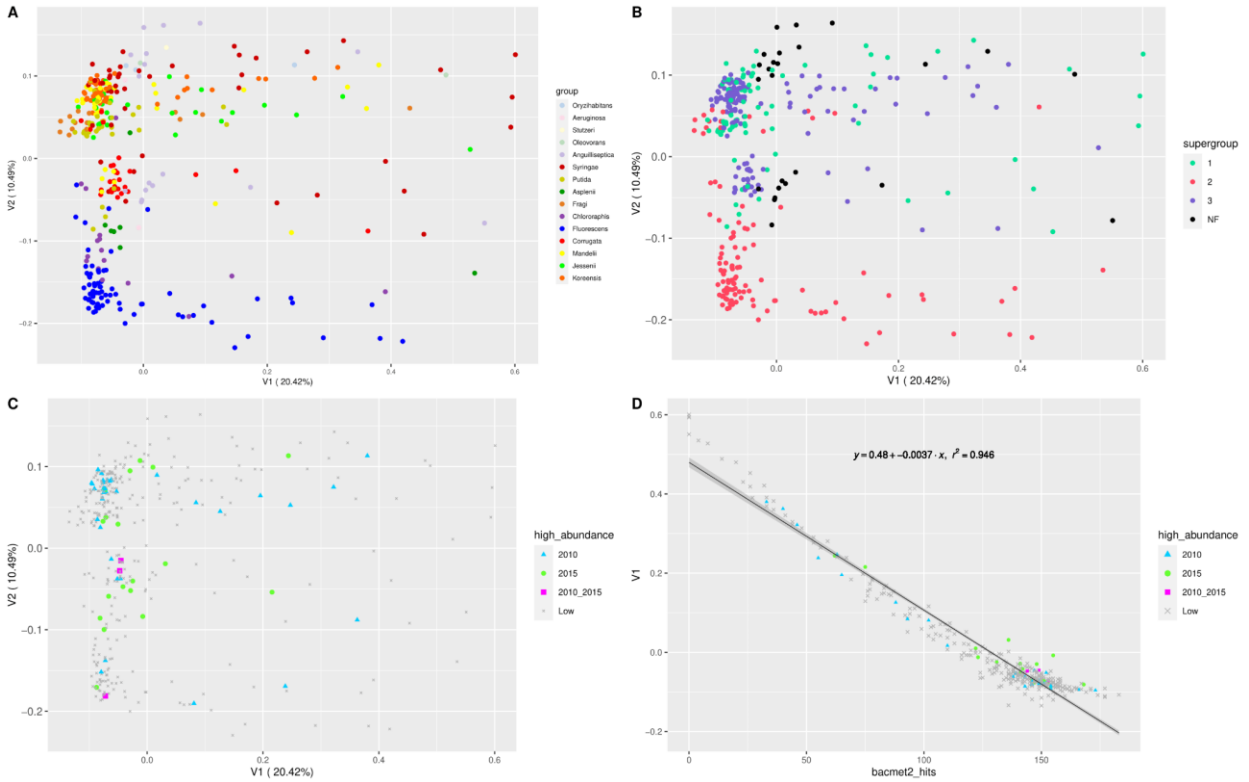


Figure 12 - PCoA of Jaccard distances for the presences/absence of antimicrobial resistance and virulence (AMR) gene content by Pseudomonas genome
A) Ordination of the first two principal components for each Pseudomonas genome colored by group. B) The same plot as panel A but colored by supergroup. C) The same plot as A, but colored by abundance status each year. D) A plot of BacMet2 database hits per genome by the V1 principal coordinate for the PCoA of the other three panels, also colored by abundance status.

Comparing each group to their respective number of genes mapping to each database first entailed orienting these data to the total number of GCFs per genome. A Kruskal-Wallis non-parametric test of each group produced a Bonferroni-corrected p-value of $2e^{-16}$, indicating a strongly significant difference between gene cluster content between the phylogenetic groups. Next, the number of genes mapped to the BacMet2 database for each group suggested that for SG1, despite the mean GCF content of group Syringae being higher than Putida, group Putida had significantly more AMR genes mapping to BacMet2. The same was observed for group Koreensis; this group had comparatively lower GCF content within SG3 but had significantly better representation in BacMet2 compared to group Jessenii. Overall, SG2 and SG3 were very similar in the number of AMR genes mapped to BacMet2.

Notably, when making these comparisons using the Comprehensive Antibiotic Resistance Database (CARD) database, the intra-group differences in mapping was highly significant for S1 and SG2, with SG3 slightly less so. Again, group Putida mapped

higher than Syringae. For SG2, the number of antibiotic resistance genes mapped from CARD to group Fragi reflected the lower number of GCFs per genome; the number of genes mapped in SG2 tracked with GCFs content overall. However that was not the case for SG3, where again group Koreensis mapped the highest number to CARD despite having the lowest mean GCF content. Lastly, the number of virulence genes mapping to each genome was similar to CARD, except SG1 was less disparate. Each group within SG2 was highly variable in their mean number of genes mapped, and SG3 was dominated by group Corrugata but otherwise relatively consistent.

To assess each genome for antimicrobial resistance and virulence (AMR) potential, genomes were screened to identify genes present in several databases including NCBI, the Comprehensive Antibiotic Resistance Database (CARD), The Virulence Factor Database (VFDB), and the Antibacterial Biocide & Metal Resistance Database (BacMet2). The full table of database hits and gene presence/absence identifies specific genes and their representation within each phylogenetic group (see data and code availability). To make broad comparisons between the AMR profiles across groups, the Jaccard distances were calculated from the presence of AMR markers from each database and plotted via a PCoA (figure 12A). As with the pan-genome PCoA, there was general clustering by group. However, all the clusters have “tails”, or genomes that extend horizontally along the V1 principal coordinate away from the main cluster. These taxa had disproportionately fewer hits to BacMet2 despite having comparable genome quality. When ordinating the jaccard indices for only the hits to the VFDB core dataset (experimentally validated virulence factors) the V1 tail artifact disappears and groups cluster more discreetly (supplementary figure 7).

Notably, as with the pan-genome PCoA, the Fluorescens group clustered the furthest distance from the others on the V2 principal coordinate. However, the Fluorescens and Chlororaphis groups were closer in proximity. Some groups, such as Anguilliseptica, split into two clusters along the V2 principal coordinate. The Corrugata group taxa clustered mainly between the other groups, with exceptions; P. *sp900581005* was highly divergent on the V1 principal coordinate, as it had 40 hits with BacMet2 and 70 hits with VFDB as compared to P. *brassicacearum*, with 153 hits to BacMet2 and 68 hits to VFDB. The majority of the Corrugata group had >120 hits to BacMet2 and >60 hits to VFDB. However, there are six taxa that were divergent on the V2 principal coordinate, and they were positioned with the Putida, Mandelii, Jessenii, and Putida groups. There are several taxa from the Putida group that also diverged on the V2 principal coordinate to be positioned with the Asplenii and Chlororaphis groups, and several taxa within the Mandelli group that were within the Corrugata group cluster. These were not the majority however. Also worth mentioning is the Putida group species P. *parafulva_B* that clustered with the Fluorescens group in the pan-genome PCoA were no longer outliers for this comparison.

When the AMR ordinations were colored by supergroups, it demonstrated general clustering trends but made it clear that each supergroup had a much less defined boundary compared to their whole-genome content (figure 12B). The non-fluorescens lineage groups Oryzihabitans and Oleovorans were clustered via the V2 principal coordinate. The Aeruginosa group was an outlier due to only two species being represented, but with P. *aeruginosa* being so thoroughly characterized, it was much better annotated compared to the other species in that group. As mentioned above, the Anguilliseptica group had two distinct clusters: one adjacent to the groups Corruata and Syringae, respectively. Among the fluorescens lineage *Pseudomonas*, SG1 clustered the most diffusely. It did not have multiple discrete clusters, but a broad overlap with SG3 and the non-Fluorescens groups. Supergroup 2 was composed of two clusters, one of which is the Fluorescens, Asplenii, and Chlororaphis groups, while the other was exclusively the Fragi group. The Fragi group overlapped heavily with SG1 and SG3, specifically the Putida, Mandelii, and Jessenii groups. Supergroup 3 also exhibited two clusters, and as with SG2, the main cluster was composed of the groups Mandelii, Jessenii, and Koreensis. The majority of the Corrugata group formed a separate cluster between the rest and SG2.

As the total number of AMR genes are constrained by the number of GCFs within that genome, the comparisons made between the AMR proportions for each group may not correlate with true functional diversity. Also, this method of comparing “hits” to databases assumes no mapping errors for any individual genome. This may in fact be an issue when observing that there are genomes from groups Anguilliseptica and Syringae that have zero genes mapped to BacMet2. A brief scan of their amino acid annotations imply this may have been erroneous but no clear malfunction of the program was evident. Also, it is worth noting that AMR genes are subject to HGT and are likely to be promiscuous between closely related taxa.

Antimicrobial resistance and virulence gene content was profiled for each genome for the purpose of supergroup and group level comparisons. The number of AMR genes mapping to each database reflected the over comprehensiveness of each database, with BacMet2 providing the majority of hits. When comparing the total number of genes mapping to each database the intra-group diversity of SG1 was consistently the highest, with group Putida being significantly higher than Syringae in both antimicrobial resistance genes and virulence factors. The intra-group variation of SG2 was significant for only a selection of antimicrobial resistance genes (CARD) but very significant for virulence factors. Only SG3 stayed relatively consistent in AMR gene content per group.

When ordinating each genome by presence-absence of all AMR genes, clustering patterns vary between “tight”, “diffuse", and “split”. Phylogenetic groups that clustered tightly were Asplenii and Koreensis. Groups that clustered less tightly but formed more of a “comet” shape (a pronounced V1 “tail”) were Asplenii, Chlororaphis, Fluorescens, and Jessenii. Some groups, however, formed multiple distinct clusters. These included

Anguilliseptica, Syringae, Putida, Corrugata, and Mandelii. At the supergroup level, SG1 was diffuse and split, but overall clustered with SG3. The non-Fluorescens groups also clustered diffuse and split with proximity to SG1 and SG3 along V2 but distinct on V1. Interestingly, SG2 was split but the majority of genomes clustered distinct from the NF, SG1 or SG3. Groups Asplenii, Fluorescens, and Chlororaphis clustered together but Fragi with its comparatively smaller genome clustered with Putida and Jessenii.

In summary, the overall AMR gene content profiles of each phylogenetic group suggest that GCF content and phylogenetic distance may influence the number of AMR genes acquired, but the specific combination of antimicrobial genes or virulence factors are highly redundant between closely related taxa. Also, the groups with split clustering may suggest diverging exposure to specific antimicrobial compounds from more distantly related taxa. Lastly, the Group Fluorescens is consistently divergent when compared to NF, SG1, and SG3.

2.3.6. *Pseudomonas* secondary metabolite profiles by group and lineage

To distinguish *Pseudomonas* groups or supergroups by secondary metabolite profile at the pangenome level, each genome was examined for possible secondary metabolite gene clusters *in silico*. First, putative secondary metabolite genes were predicted and annotated for each genome. Next, these genes were compared between all genomes and clustered based on gene position and sequence homology. Finally, the resulting GCFs and their predicted metabolite class were visualized using a Bray-Curtis dissimilary PCoA due to genomes containing multiple GCFs of the same metabolite class. Plots were examined for differences between phylogenetic group, supergroup, and abundance phenotype by year.

Pseudomonas contains a large accessory genome, and the repertoire of secondary metabolites for any individual species or strain can vary dramatically. These products, by definition, are not essential to growth or reproduction and therefore evidence of GCFs shared between groups or supergroups may indicate potential functional distinctions in lifestyle.

Comparing the secondary metabolite profiles of each group and supergroup necessitates first determining if there is any correlation between total gene cluster content/genome size and the number of biosynthetic gene clusters (BGCs) per genome. Also, to demonstrate that total contigs per genome assembly is not artificially inflating the number of BGCs as a result of fragmented clusters on the edge of contigs (figure 13). Once it is established that the number of BGCs per genome is not being driven by poor assemblies, the fractional abundance of each class of BGC within each *Pseudomonas* group and supergroup can be compared to assess consistency between these phylogenetic levels (figure 14). Finally, Bray-Curtis dissimilarities between individual genome BGC profiles allow for visualizing BGC content overlap between groups and

supergroups, and to target the highly abundant genomes as described above (figure 15).

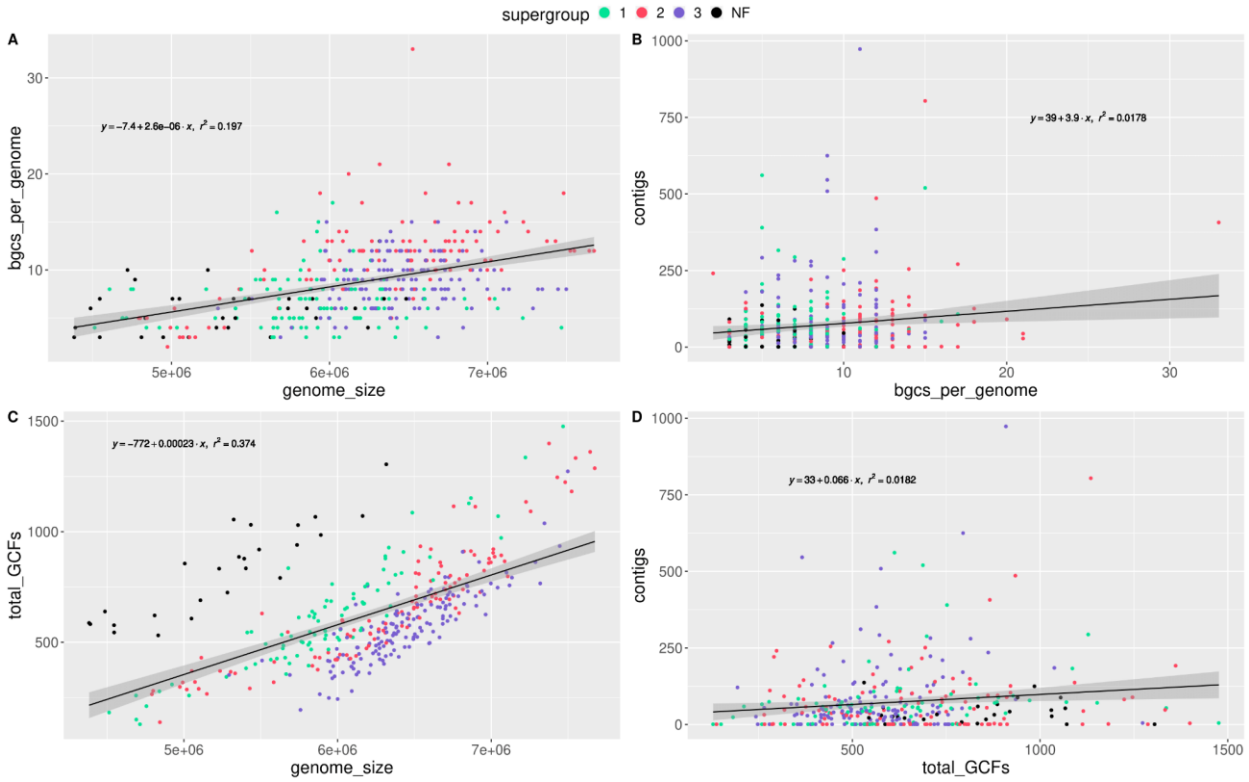


Figure 13 - Biosynthetic gene clusters (BGCs) and gene cluster families (GCFs) versus genome size and number of contigs for each *Pseudomonas* supergroup.
A) Total secondary metabolite BGCs per genome plotted against genome size, with a linear regression (formula for all panels 'y ~ x'). B) Contigs vs BGCs per genome, with the same formula. C) Total pangenome GCFs for each genome vs genome size in base pairs. D) The number of contigs for each genome vs the total pangenome GCFs.

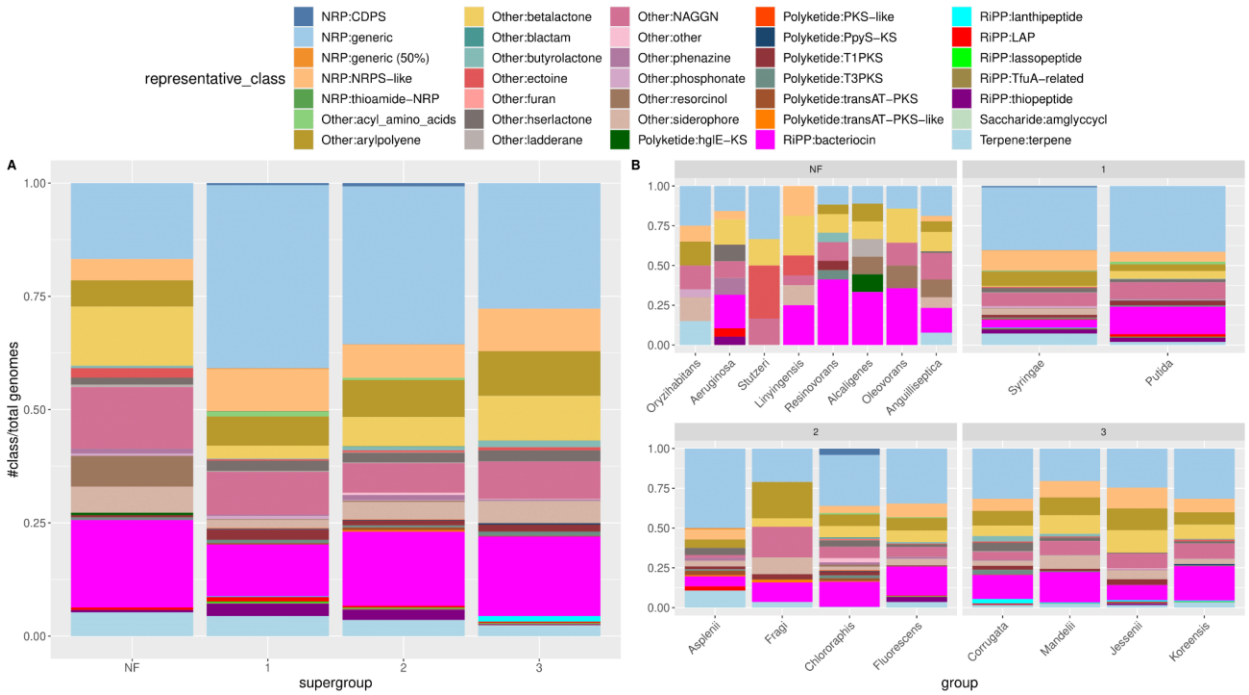


Figure 14 - Proportions of secondary metabolite representative classes present at the supergroup and group level by number of genomes.
A) The total number of each type of representative class present in each supergroup, divided by the number of genomes in that supergroup. B) The total number of each representative class divided by the number of genomes in each group.

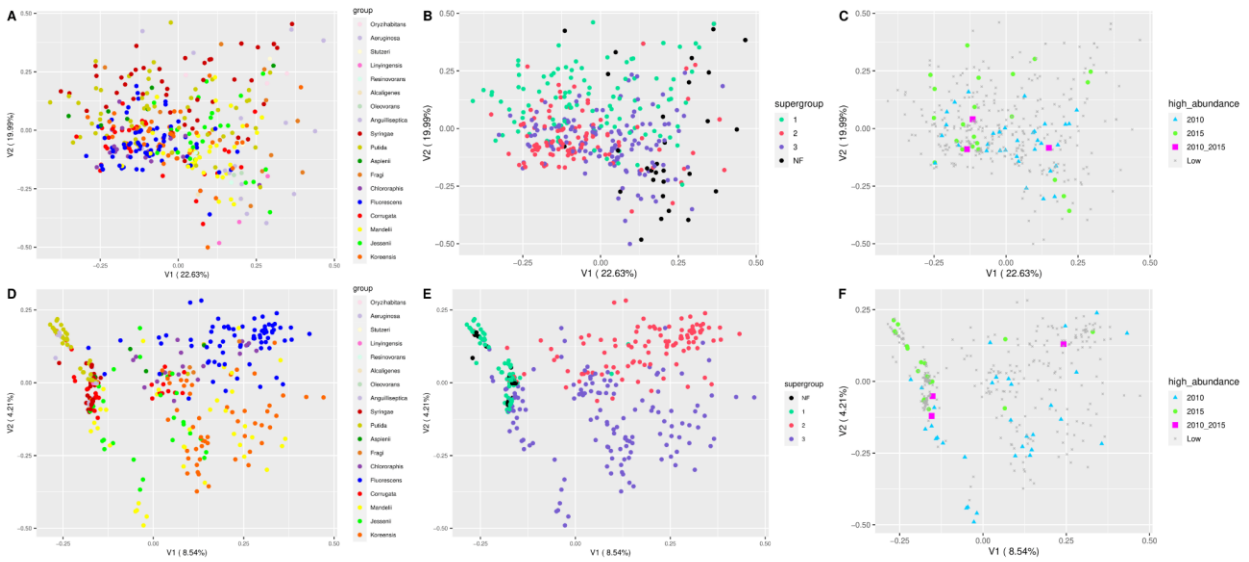


Figure 15 - PCoA of Bray-Curtis dissimilarity of secondary metabolite profiles by group and supergroup compared by GCF annotation or GCF sequence.
A) Ordination by annotation for each genome, colored by group. B) The same plot colored by supergroup with NF representing non-fluorescens lineage genomes. C) The same plot

with each color representing a genome within the top ten percent for each year. D) Ordination by GCF sequence for genome, colored by group. E) as previous, colored by supergroup. F) as previous, colored by high abundance per year.

The first comparisons made between the number of BGCs per genome versus genome size indicate that as genomes increase in the number of total base pairs, the more BGCs per genome are found (figure 13A). There is a very modest correlation in the number of contigs per genome versus BGC content. However, the linear regression of contigs versus BGC content per genome is much weaker (R^2=0.0178) and thus it is unlikely that fragmented gene clusters are artificially driving up the number of BGCs found in each genome (figure 13B). When these same comparisons are made instead with pangenome GCFs, a similar trend is observed. Notably, the correlation between total GCFs and genome size is fairly strong (R^2=0.374) with the non-fluorescens lineage genomes displaying proportionally more GCFs per genome than nearly all Fluorescens *Pseudomonas* (figure 13C). When plotting only the non-fluorescens lineage genomes, the correlation reaches an R^2=0.796, whereas plotting only the fluorescens lineage genomes the correlation is R^2=0.633 (figure not provided). Again, when comparing contigs per genome versus total GFCs per genome the correlation is weak (R^2=0.0182), indicating that assembly quality does not strongly affect the number of gene clusters detected (figure 13D).

Next, fractional abundances of each representative BGC class per group were determined for each group. Overall, no specific BGC classes are present in 100% of genomes. At supergroup level, the most consistently observed and abundant BGC classes include NAGGN, a dipeptide N-acetylglutaminylglutamine amide involved in osmotic regulation (Sagot *et al.*, 2010), aryl polyene pigments likely responsible for protection from reactive oxygen species (Schöner *et al.*, 2016), beta-lactone antibiotics (Robinson, Christenson and Wackett, 2019), siderophores (Cornelis and Matthijs, 2007), bacteriocins (Simons, Alhanout and Duval, 2020), and terpenes (Yamada *et al.*, 2015) (figure 14). The most abundant class of BGC across supergroups are generic non-ribosomal peptides, which are diverse bioactive compounds that are synthesized by large modular synthetases and often act as antimicrobials (Strieker, Tanović and Marahiel, 2010). At the group level, as expected, the non-Fluorescens and SG1 groups have substantial variation in the presence and proportion of each BGC class as reflected by the inherent phylogenetic diversity within each of those supergroups as compared to SG2 or especially SG3. There are BGC classes found in a small number of groups, such as the cytoprotectant ectoine within groups Stutzeri and Linyingensis (Hermann *et al.*, 2020), linear azol(in)e-containing peptide antibiotics for Putida, Asplenii, Corrugata, and Koreensis (Cox, Doroghazi and Mitchell, 2015), tRNA-dependent cyclopeptide antibiotics in Syringae an Chlororaphis (Gondry *et al.*, 2009), and with lanthipeptide antibiotics found predominantly within SG3 (Repka *et al.*, 2017). Each are low abundance within their groups, with only a fraction of genomes possessing these BGCs.

To determine if broad classes of secondary metabolites (annotations of each BGC) were predictive of supergroup or group status, Bray-Curtis dissimilarities were calculated between the number of each class of BGC per genome. At group level, there is only slight clustering and there is heavy overlap. At supergroup level there is slightly better discrimination between NF and the fluorescens lineage *Pseudomonas*, which is more easily observed when each group is ordinated separately (figure 15A, 15B, 15C, supplemental figure 2). Counts of shared BGC classes by group and supergroup indicate many classes are shared by both Fluorescens and NF lineages, while of the fluorescens lineage SG1 has 54% shared classes vs 25% and 33% for SG2 and SG3, respectively (supplemental figure 4).

However, when ordinating the BGCs by gene orientation within each GCF instead of by representative class, the Bray-Curtis PCoA shows greater cluster discrimination based on phylogenetic group, supergroup, and abundance phenotype as well (figure 15D, 15E, 15F). Specifically, this method of comparing cluster structure similarity as opposed to only the number of gene clusters per secondary metabolite class, indicates a greater degree of similarity between all non-Fluorescens genomes including Anguilliseptica. Notably also, the Bray-Curtis dissimilarities are much higher for SG1 when comparing annotations, but comparing GCF structure suggests a high degree of similarity. In the case of SG2, groups Asplenii and Fragi become more similar, while Chlororaphis and Fluorescens change comparatively little. For SG3, the GCF structure comparisons for group Corrugata caused two tight clusters to form, whereas group Mandelii clustered more distantly. Groups Jessenii and Koreensis clustered similarly in both cases (supplemental figure 3).

As discussed previously, many of the NF groups had less than 4 genomes; therefore making any assertions based on ordination is risky when so little is being compared. When quantifying BGC types per group, even with normalization the NF groups still suffer from an inherent imbalance. Also, each class of secondary metabolite contains a diverse array of specific compounds with varying structures and biological functions, and comparisons at this level can’t reflect function without closer investigation of individual BGC cluster sequences.

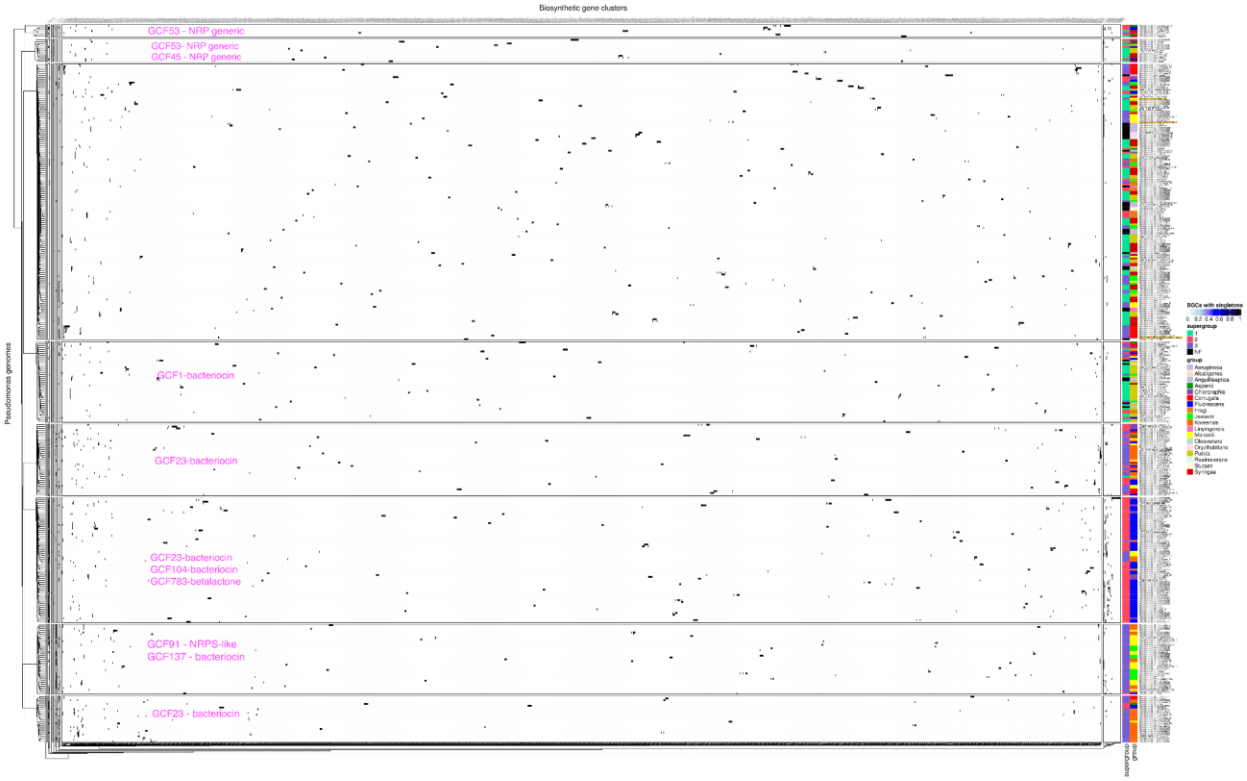
In summary, secondary metabolite profiles for each genome are influenced by the size of the genome and therefore the number of pangenome GCFs, while the number of contigs/assembly quality per genomes has a very small effect on either the number of BGCs or GFCs. Proportions of each BGC class can differentiate between groups, but the difference is driven by BGC classes found in relatively few taxa out of the group. The most consistently observed secondary metabolite classes within these *Pseudomonas* are aryl polyenes, NAGGN, siderophores, bacteriocins, and terpenes. Comparing the differences in the number of BGC types per genome is less informative than comparing BGC structure. Lastly, BGC profiles are not predictive of the abundance phenotypes observed during 2010 or 2015.

2.3.7. BGC and AMR profiles by unsupervised clustering

To determine which individual genes, gene clusters, and BGC types can be attributed to specific groups or supergroups, heatmaps of BGC GCFs and AMR genes of all genomes were produced and clustered using k-means. First the presence/absence of all BGC gene cluster families were visualized (figure 16A), then only the accessory BGCs without the singleton gene clusters (figure 16B), then the total AMR gene presence/absence for all genomes (figure 17).

Each heatmap contains hundreds of gene clusters or genes on the x-axis (columns) and therefore it is not possible to identify individual genes. The overall information being conveyed is the pattern of gene clustering and the distribution of *Pseudomonas* group and supergroup membership.

A.



B.

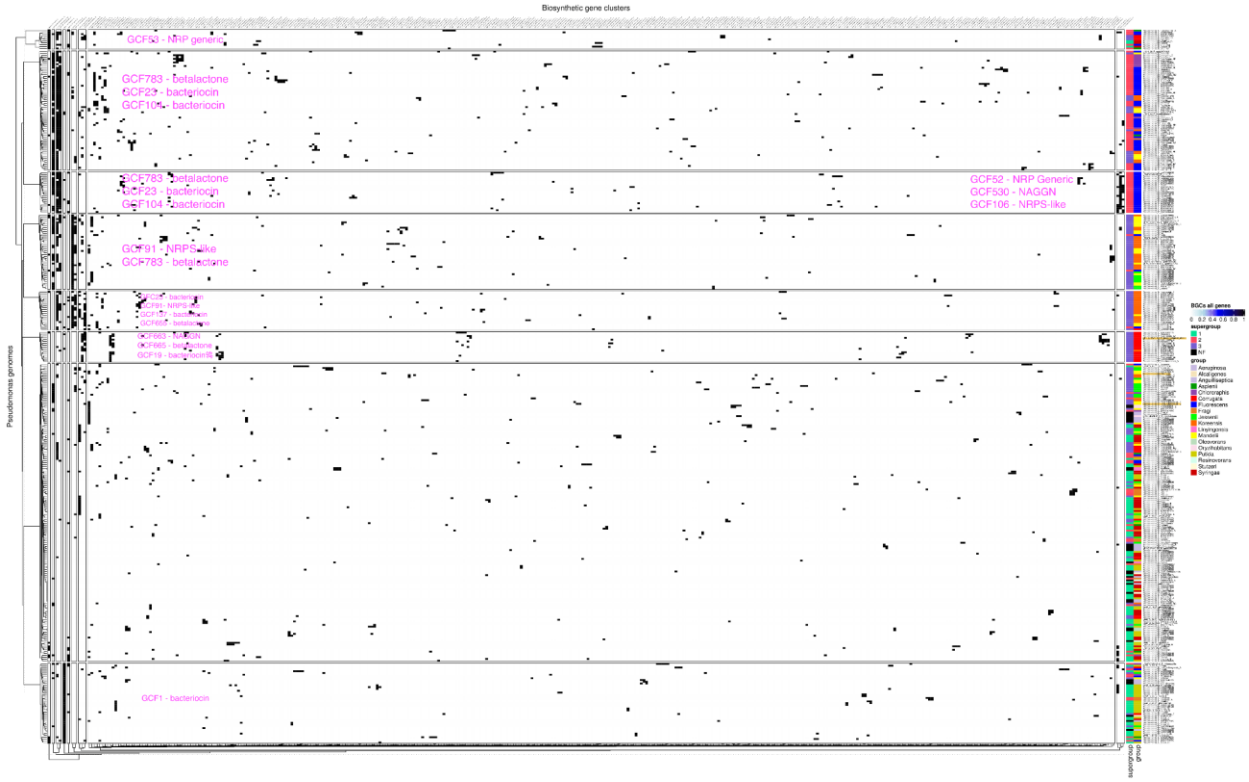


Figure 16 - BGC gene content heatmaps by genome with group and supergroup.
A) All BGC gene cluster families, clustered by k-means with supergroup and group annotated for each genome. Noted in pink are the most abundant GCFs with their annotation. Genomes highlighted in yellow are the three most highly abundant taxa for 2010. B) BGCs found in >1 genomes (no singleton GCFs) clustered by k-means with annotations and most abundant GCFs by cluster.

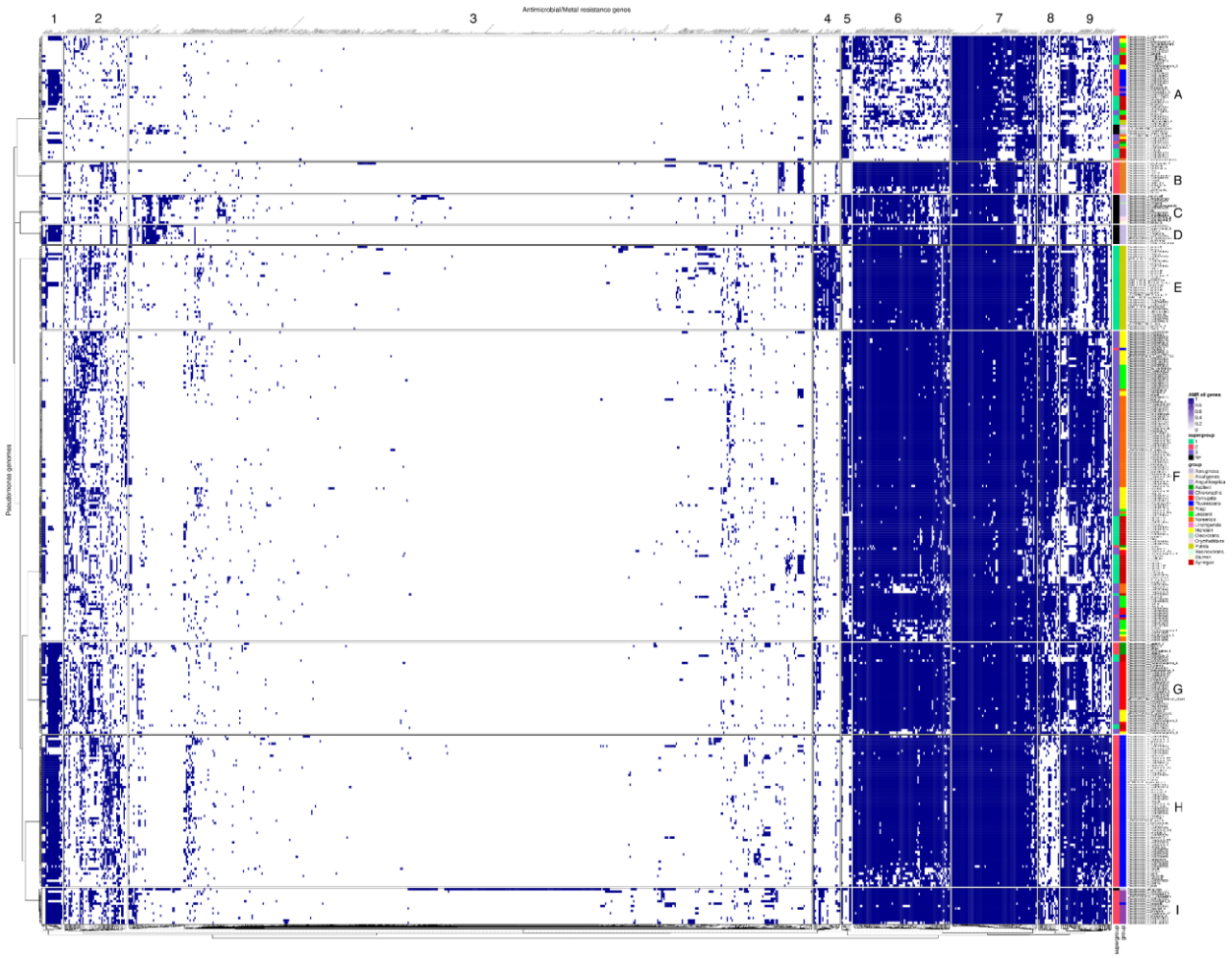


Figure 17 - AMR gene content heatmap by genome with group and supergroup. All AMR genes clustered by k-means with annotations. Pink axes designations to facilitate identifying clusters of genes by group.

The BGC presence/absence profile of all *Pseudomonas* genomes clearly indicates a great degree of diversity in both the number of BGC gene clusters and the number of unique gene clusters across all groups (figure 16A) . When regarding the entirety of the BGC profile there are only a few GCFs that are found in the majority of any given group, and there is very little consistency between the designated phylogenetic groups and supergroups. The clustering of GCFs by k-means indicates the strongest phylogenetic signal exists within the group Fluorescens, and some degree of similarity between supergroups 3 but not specifically among groups of that category. There are a large number of genomes across the non-Fluorescens and supergroup 1 lineages that completely lack any defining GCF. The total number of BGC GCFs detected within this *Pseudomonas* pangenome is 1753, with 1357 of those GCFs containing a gene cluster found in only one *Pseudomonas* genome. Of these singleton GCFs, 584 were annotated as NRP:generic, 134 as arylpolyene, 83 as NRP:NRPS-like, 74 as terpene, 59 as beta-lactone, and 58 as bacteriocin. The vast majority of singleton GCFs are complete

nonribosomal peptide synthetase gene modules with an unknown product. However, the most commonly shared GCF is annotated as a bacteriocin, found in 149 of the 394 genomes (37.8%). The next most common GCF is a beta-lactone within 119 genomes (30.2%), then another bacteriocin within 101 genomes (25.6%).

When singleton BGC GCFs are excluded the degree of k-means clustering by group and supergroup improves slightly for fluorescens lineage SG2 and SG3, with the blooming *P. brassicacearum* clustering separately with the majority of the group Corrugata (figure 16B). As with the previous figure, the first cluster of genomes remains predominantly pathogenic genomes with *P. corrugata, P. viridiflava_C, P. fuscovaginae_A, P. mediterranea, P. cichorii, and P. cichorii_B,* but with the exception of *P. antarctica* and *P. batumici* which are in the literature as environmental isolates with no currently known host association. Excluding the singleton GCFs does nothing to improve clustering for the non-fluorescens lineage or SG1 *Pseudomonas*.

Compared to the BGC profiles, the AMR genes exhibit a much larger degree of overlap between genomes with the groups Fluorescens, Syringae, Putida, and Fragi demonstrating a high degree of similarity in AMR gene content (figure 17). There is a Hcp1 secretion island (cluster 1) that encodes a T6SS and it is predominantly found in group Fluorescens and Corrugata, but also in a selection of Anguilliseptica and sparsely in Syringae. The most conspicuous cluster unique to any group was within the SG1 grouping for Putida (cluster 4E), containing genes for multidrug efflux (mdfA, smdB, acrA, ttgABCR, mexAN, and mepC). Unlike SG1 and SG2, SG3 doesn't cluster by phylogenetic group, only broadly by supergroup. The majority of Corrugata and Koreensis cluster as groups but Mandelii and Jessenii are distributed across multiple clusters. There is a type 4 pili operon (cluster 5) that is found sporadically in all three supergroups and predominantly in NF, Syringae, Mandelii, Koreensis, and Corrugata. This T4P is completely absent in Fragi, Chlororaphis, and Fluorescens. The AMR genes core to each group were mostly alginate biosynthesis, arsenic resistance, flagellin, copper resistance, iron transport, LPS biosynthesis, Mex-family drug efflux, and pyoverdin (supplemental table 3).

To provide an alternative to visualizing phylogenetic differences in BGC and AMR profiles that allows for identifying specific genes or gene clusters, heatmaps of total BGC GCFs, GCFs excluding singleton GCFs, and AMR genes were created. Sequence modularity of individual secondary metabolite GCFs are highly diverse, with the most common representative class of BGC as NRP:generic, followed by bacteriocins. Antibiotics such as beta-lactones, polyketides, and virulence factors such as siderophores were highly represented. Most BGCs were found in less than 10 genomes, and the NRP:generic GCF53 was enriched in genomes known to be pathogenic. What little phylogenetic signal there was for BGC content suggests fluorescens lineage *Pseudomonas* are enriched in bacteriocins. In contrast, antimicrobial resistance and virulence gene content are more strongly influenced by phylogeny; the genomes within

groups Fluorescens, Fragi, Syringae, and Putida clustered tightly with only a minority located apart. Multidrug efflux pump operons, metal tolerance, alginate biosynthesis, and flagellin genes were highly represented and present among multiple phylogenetic groups as expected of the genus *Pseudomonas,* while the presence or absence of type VI secretion systems, type IV pili, and pyoverdin operons were specific to only a selection of phylogenetic groups (Corrugata, Fluorescens, Chlororaphis, and some Anguilliseptica).

In summary, all *Pseudomonas* included in this pangenome contain a large degree of sequence diversity in their secondary metabolites which is congruent with the difficulty in their bray-curtis PCoA cluster discrimination by cluster annotation, despite their relatively conserved pattern of antibiotic resistance. fluorescens lineage *Pseudomonas* maintain clades of both pathogenic and commensal species highly enriched in virulence factors specifically relating to motility, host-microbe, and microbe-microbe interactions regardless of the 2010 abundance phenotype.

Chapter 3

Pangenomics of high abundance *Pseudomonas* and Eukaryotic diversity within *Pseudomonas*-enriched maize rhizospheres

3.1 Aims

To explore the taxonomic diversity inherent in the maize rhizosphere metagenomes, alpha diversity metrics of taxonomic richness and evenness were used to compare between maize genotype, week, and field. Bray-Curtis dissimilarity was used to compare each bacterial community between rhizospheres also for genotype, week, and field. Next, genomes of *Pseudomonas* will be assembled from reference sequences and *de novo* to capture high abundance species to conduct species-level pangenome analyses. Phylogenetic group pangenomes will be compared for lineage-specific gene content and evidence for niche specialization. Lastly, the metagenomes will be profiled for eukaryotic taxa to examine the diversity of organisms that potentially influence microbe-host interactions.

3.2 Methods

Measuring alpha and beta diversity

The full, unfiltered Bracken taxonomic profile was used when analyzing the alpha and beta diversity metrics using the R package vegan v2.5-7 (Oksanen *et al.*, 2020) and phyloseq v1.30.0 (McMurdie and Holmes, 2013).

Assembling genomes from metagenomes

To improve metagenomic binning, centrifuge was used to taxonomically profile the most abundant microbial reads and download the representative genomes from NCBI. During metagenome assembly, these genomes were used for reference-based assemblies with Metacompass, and the remaining unmapped sequences were assembled *de novo*. Metagenome binning was performed on a per-sample basis utilizing both maxbin2 and metabat2, with differential coverage binning being split between 29 randomly selected samples to reduce computational load. The resulting MAGS were filtered by CheckM quality for downstream analysis. The *Pseudomonas* isolates were filtered to only include genomes with a checkM completeness of >97%, contamination of <5%, while the MAGs were filtered at >80% complete and <5% contamination.

Pangenome analysis

The pan-genome analysis was performed using Anvi'o interactive interface (v7 "Hope") (Eren *et al.*, 2021) to determine the core and accessory genome of all the rhizosphere-associated *Pseudomonas*, selected high abundance species (P. *brassicacearum,* P. *frederiksbergensis E,* and P. *silesiensis* . Core is defined as a gene cluster present in all genomes, accessory as a gene cluster present at least two but no more than n-1 genomes, and singleton as a gene cluster present in only one genome. The MCL inflation was set to 4 for the whole genus, and pangenomes were generated for each taxa at species level with an MCL inflation of 6 to improve granularity and therefore to better discriminate paralogous genes. A custom python script was created to parse the orthologous gene clusters identified in Anvi'o into a presence-absence table as required for downstream analysis (see data and code availability link).

Differential abundance vs climate

Climate data was acquired via the REST API at oikolab (*OikoLab Weather API*, no date) which uses the ERA5 climate reanalysis (*ERA5 Climate reanalysis*, no date). The coordinates of the each field were as follows: Aurora at latitude: 42.7341897, longitude: -76.6570054, Ithaca at latitude: 42.471842, longitude: -76.4403944, and Lansing at latitude: 42.637274, longitude: -76.5993753. The model used: week + soil_temp_3day + precip_3day + (1|field) where week is the time point sampled, soil_temp_3day was the three day average of the soil_temperature_level_1 in degC for the day of sampling and the two days prior, precip_3day is the three day average of rain in mm/hour for the day of and three days prior, with field being a fixed effect using the R package glmmTMB v1.1.2.3 (Brooks *et al.*, 2017).

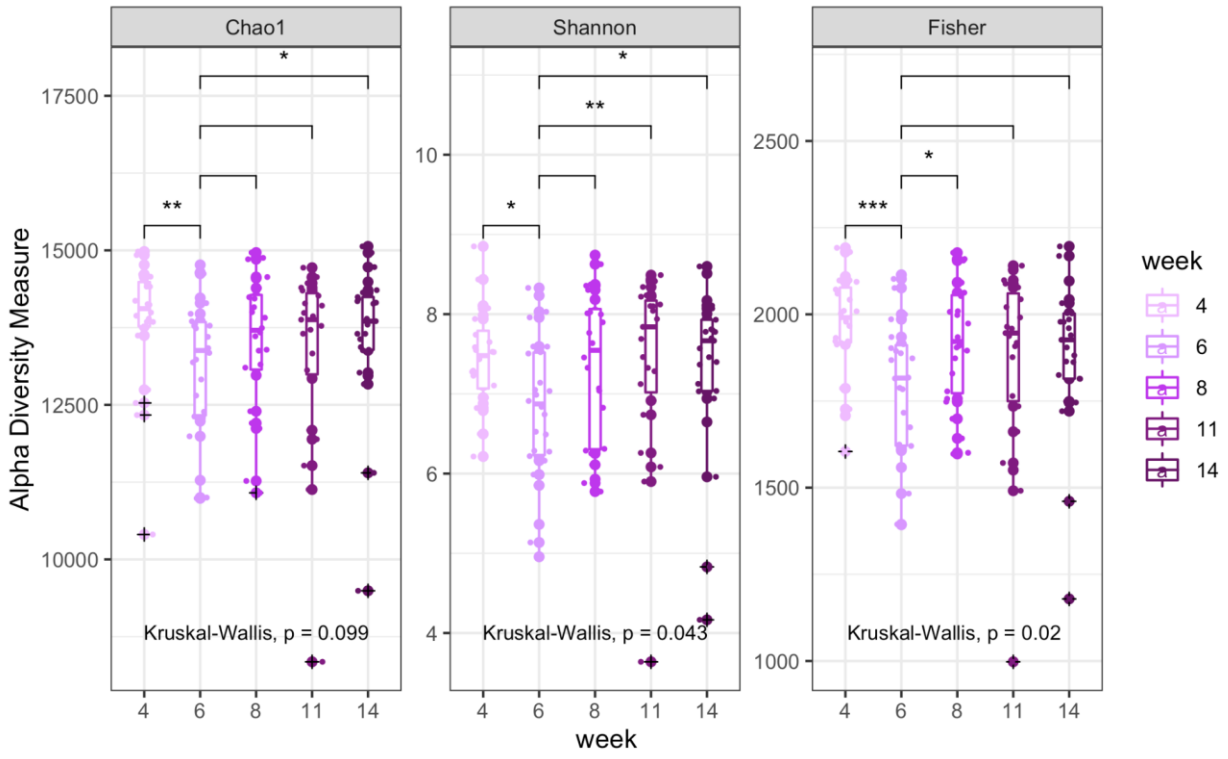
Eukaryotic species profiling

To estimate eukaryotic taxon abundances within each rhizosphere metagenome, the combined biological replicates with maize sequences filtered and subsampled to 5M reads were used to run EukDetect v1.0 (Lind and Pollard, 2020) with *min_read_len: 80*

3.3 Results

3.3.1. Maize rhizosphere alpha and beta diversity

Alpha diversity, all bacterial species detected within each individual rhizosphere metagenome, for the 2010 field season was determined by week, maize genotype, and field. For each condition, three alpha diversity metrics were measured: Chao1 (non-parametric estimation of classes within a population skewed towards rare taxa), Shannon diversity index (estimates of species diversity which accounts for the number and abundance of taxa), and Fisher's alpha (α calculated based on the number of taxa observed vs sample size) and their statistical comparisons (figure 18). Beta diversity, specifically comparing the diversity of each metagenome to each other, for the rhizospheres was measured using Bray-Curtis dissimilarity (quantification of compositional dissimilarity between samples) by field then by week for the 2010 field season (figure 19).



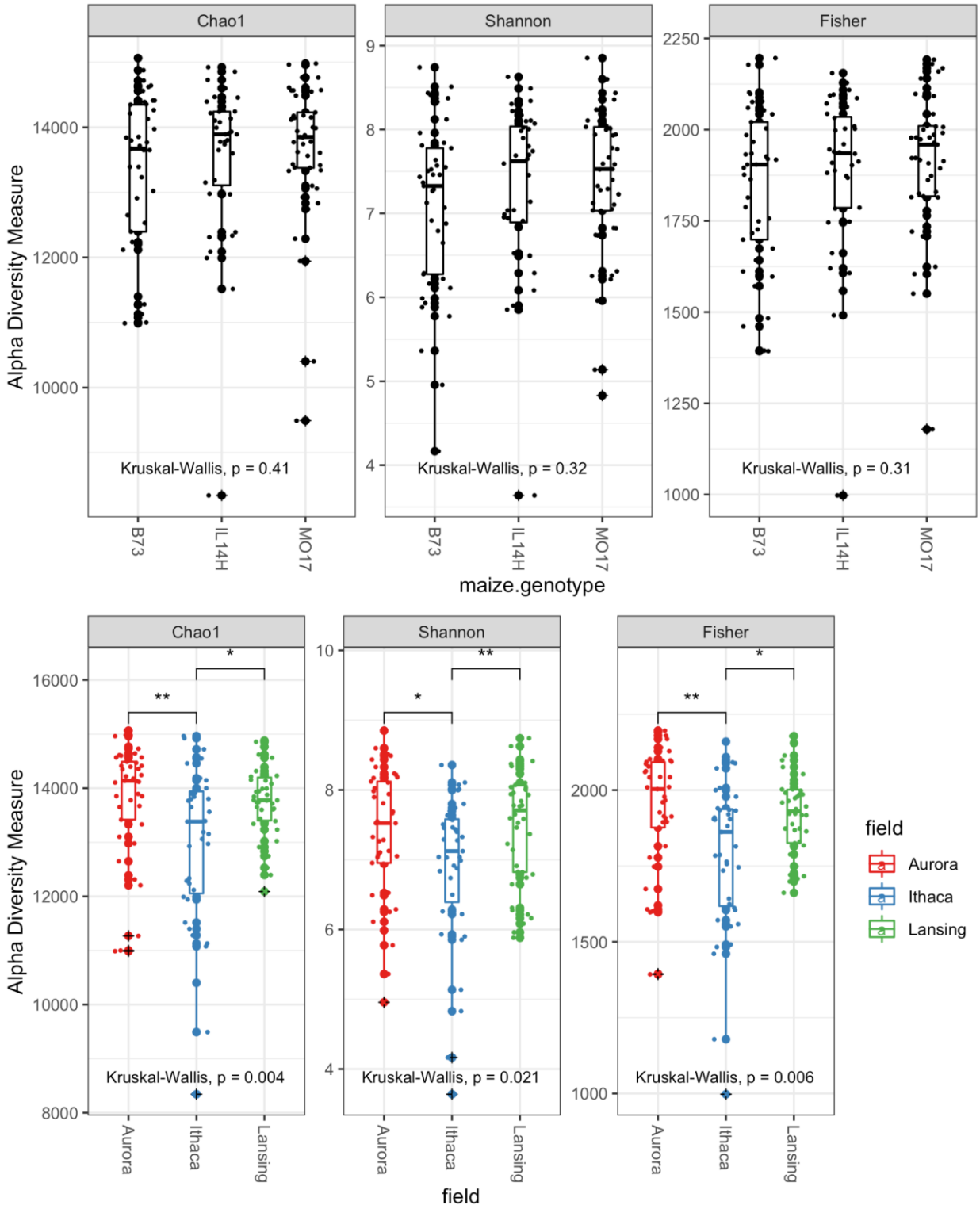


Figure 18 - Chao1, Shannon, and Fisher Alpha diversity metrics by week (A), maize genotype (B), and field (C) for the 2010 season.
Outliers denoted by a ring and cross, and "*" denotes a Wilcoxon rank-sum pairwise p <= 0.05, "**" p <= 0.01, "***" p <= 0.001

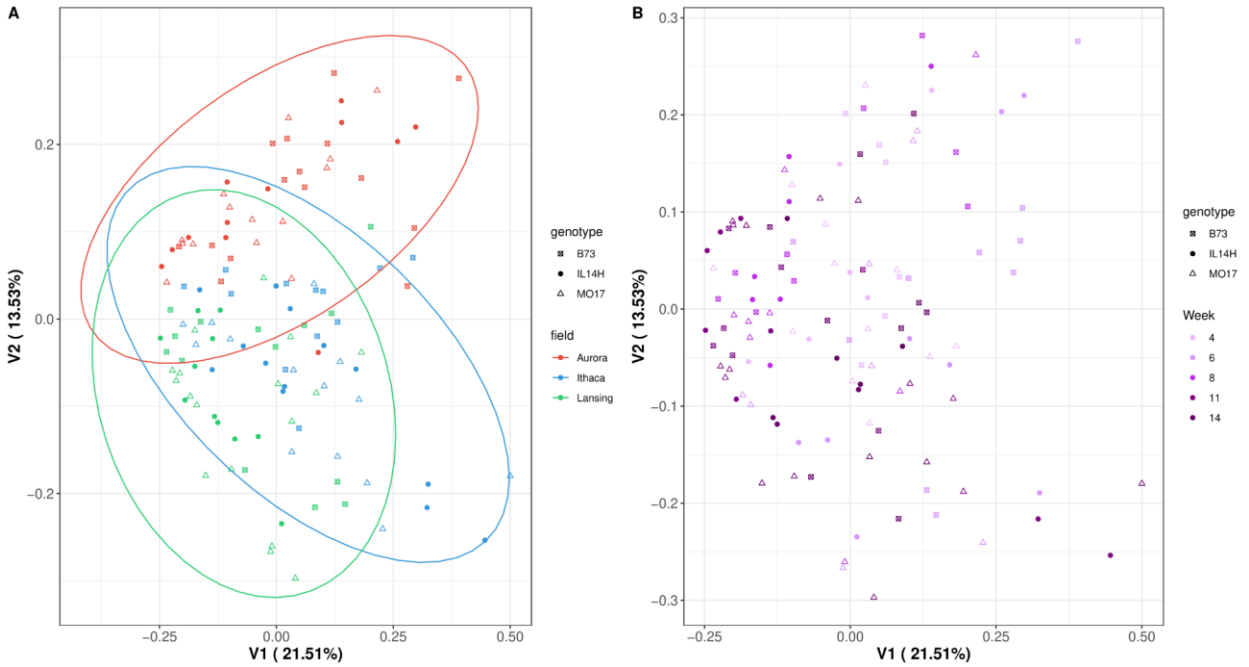


Figure 19 - Beta diversity comparisons between maize rhizosphere metagenomes

A) Bray Curtis dissimilarity PCoA by field, with density histograms for each principal coordinate. Each point is distinguished by the maize genotype the sample originated, square = B73, circle = IL14H, and triangle = MO17. Ellipses assume a multivariate normal distribution. B) The same Bray Curtis PCoA with the samples colored by time point.

The measurements of alpha diversity for each week indicate the strongest difference in both richness (Chao1, Fisher) and evenness (Shannon) was observed from week 4 to week 6, where it decreased (figure 18A). Overall evenness between all genotypes and fields increased the most from week 6 to week 11(figure 18B). Alpha diversity measurements did not indicate a significant difference between maize genotypes using the non-parametric Kruskal-Wallis test, however this is counter to the differences observed by field which indicated Ithaca as being consistently lower in both richness and evenness compared to Aurora and Lansing (figure 18C).

The beta diversity measurements suggest moderate clustering by field along the V2 principal coordinate (figure 19A). However, week and genotype do not appear to influence the beta diversity of the rhizosphere metagenomes. The total contribution of each principal coordinate is 21.51% and 13.53% for V1 and V2 respectively. Inferred measurements of the average soil temperature and precipitation three days prior to sampling were not statistically significant when comparing them to *Pseudomonas* abundance, however it is worth noting that overall 2015 was slightly cooler and wetter than 2010, and due to the geographic proximity between Aurora, Ithaca, and Lansing there was only a nominal variation in precipitation and soil temperature (supplemental figure 8, supplementary table 8). However, when using a mixed-effect model with field as a fixed effect, the only taxa found to be differentially abundant with temperature were a

selection of Fluorescens species. Their relative abundance was negatively correlated with temperature (supplemental table 5).

A serious consideration when measuring diversity within the rhizosphere is the risk of bulk soil carryover during the DNA extraction process. This could potentially inflate richness due to the higher on average diversity within soils as opposed to the rhizosphere. Contaminating bulk soil microbes may also wash out signal in conditions where the differences, say between the maize genotypes, would be weak to begin with.

The bacterial and archaeal diversity within each maize rhizosphere metagenome is affected most strongly by field where both richness and evenness is highest for Aurora and lowest for Ithaca. When comparing individual diversity by week, the rhizospheres collected at week 4 have the highest taxonomic richness but both richness and evenness drops significantly at week 6 where it recovers thereafter. No difference in individual diversity is observed when comparing each maize genotype. Comparing differences in bacterial and archaeal community composition between rhizospheres indicates a strong field effect, but no observable influence by maize genotype or week.

3.3.2. Taxonomic diversity of the metagenomically-assembed genomes

The metagenomes for each rhizosphere of the 2010 season contained a high abundance of *Pseudomonas* sequences, especially *P. brassicacearum* from Ithaca and Lansing. To better capture the strain diversity within these maize rhizospheres, I used bioinformatic methods to recover *Pseudomonas* genomes from each metagenome from the 2010 season where it was most abundant. This is accomplished by aligning the metagenome sequences to reference genomes then performing de-novo assembly on the remaining unaligned sequences. The metagenomically assembled genomes (MAGs) were assessed for completeness (>70%) and contamination (<5%) before inclusion in downstream analyses, and assigned taxonomy using GTDB R89.

MAG_id	classification	fastani_reference	closest_placement_ani	closest_placement_af	aa_percent	Completeness	Contamination	Strain heterogeneity
B73_A_8_2_metabat2_high_PE.007.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__;s__	N/A	N/A	N/A	77.48	77.12	1.75	80
IL14H_L_4_1_metabat2_low_PE.003.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Palsa-955;s__	N/A	75.99	0.07	89.03	91.38	1.3	16.67
MO17_L_6_2_metabat2_high_PE.006.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Palsa-955;s__	N/A	75.91	0.07	87.74	88.92	4.11	33.33
MO17_A_11_1_metabat2_low_PE.001.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Cyclobacteriaceae;g__Ohtaekwangia;s__	N/A	89.19	0.83	65.36	79.8	2.58	7.69
B73_L_6_1_metabat2_high_PE.006.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Dyadobacter;s__	N/A	85.16	0.78	70.71	75.75	1.24	0
B73_A_11_1_metabat2_high_PE.004.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Weeksellaceae;g__Chryseobacterium;s__	N/A	93.36	0.85	70.48	86.87	0.87	25
MO17_L_6_3_metabat2_high_PE.005.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Weeksellaceae;g__Chryseobacterium;s__	N/A	83.18	0.76	59.23	74.92	1.59	0
B73_L_14_1_metabat2_high_PE.008.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	N/A	N/A	76.73	83.41	4.49	66.67
B73_L_14_3_metabat2_high_PE.010.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	92.6	0.87	96.05	98.47	0.96	0
B73_L_14_4_metabat2_high_PE.005.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	93.17	0.78	98.12	96.65	1.7	40
IL14H_A_14_1_metabat2_high_PE.002.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	93.52	0.8	96.27	97.85	0.55	25
IL14H_I_14_3_metabat2_low_PE.006.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	88.48	0.71	98.39	97.13	0.66	0
IL14H_L_14_1_metabat2_low_PE.002.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	93.02	0.81	96.96	96.01	0.74	0
IL14H_L_6_2_metabat2_high_PE.005.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__	N/A	92.27	0.87	81.39	87.14	2.87	42.86
B73_L_14_3_maxbin2_high_prob.005.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__Pedobacter ginsenosidimutans	GCF_001442625.1	98.32	0.9	94.68	93.39	1.78	0
MO17_I_14_2_metabat2_low_PE.005.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__Pedobacter ginsenosidimutans	GCF_001442625.1	98.67	0.93	67.72	80.29	2.63	71.43
MO17_I_14_3_metabat2_low_PE.001.contigs	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter;s__Pedobacter ginsenosidimutans	GCF_001442625.1	98.46	0.89	98.69	97.61	3.85	80
MO17_A_6_1_metabat2_low_PE.004.contigs	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Gemmatimonadales;f__GWC2-71-9;g__;s__	N/A	N/A	N/A	64.84	71.4	2.3	0
IL14H_L_6_2_metabat2_low_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Asticcacaulis;s__	N/A	85	0.83	77.54	81.48	3.5	33.33
B73_L_11_3_metabat2_low_PE.002.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__;s__	N/A	N/A	N/A	69.8	80.29	0.22	100
IL14H_I_8_2_metabat2_high_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__;s__	N/A	N/A	N/A	69.9	74.5	0.27	0
MO17_I_11_1_metabat2_low_PE.002.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium;s__	N/A	N/A	N/A	84.37	88.43	1.82	35.71
MO17_I_8_1_metabat2_low_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium;s__	N/A	N/A	N/A	89.31	94.15	2.07	50
B73_A_6_4_metabat2_high_PE.007.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingopyxis;s__	N/A	88.33	0.67	72.64	78.7	2.37	30.77
IL14H_L_6_2_metabat2_low_PE.005.contigs	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingopyxis;s__Sphingopyxis sp001427085	GCF_001427085.1	96.76	0.91	70.65	79.03	2.76	50
IL14H_I_14_3_metabat2_high_PE.007.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Duganella;s__Duganella zoogloeoides	GCF_000383895.1	95.07	0.88	84.21	97.39	2.24	37.5
MO17_A_11_1_metabat2_low_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Pelomonas;s__	N/A	91.46	0.78	96.61	98.84	1.52	16.67
IL14H_I_11_2_metabat2_low_PE.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia;s__Ralstonia sp000955795	GCF_000955795.1	97.84	0.91	76.96	77.32	2.43	85.71
IL14H_I_11_1_metabat2_low_PE.009.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Variovorax;s__Variovorax paradoxus	GCF_001591365.1	98.47	0.96	82.78	73.24	0.92	0
IL14H_I_14_1_metabat2_low_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Variovorax;s__Variovorax paradoxus	GCF_001591365.1	98.51	0.96	94.98	97.34	0.25	0
MO17_I_11_1_metabat2_high_PE.010.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Variovorax;s__Variovorax paradoxus	GCF_001591365.1	98.77	0.94	98.57	100	0.23	100
MO17_I_11_2_metabat2_high_PE.005.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Variovorax;s__Variovorax paradoxus	GCF_001591365.1	98.52	0.95	98.55	99.53	0	0
MO17_I_11_3_metabat2_low_PE.009.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Variovorax;s__Variovorax paradoxus	GCF_001591365.1	97.87	0.97	93	89.55	0.56	33.33
IL14H_I_11_1_metabat2_high_PE.011.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Enterobacter;s__Enterobacter ludwigii	GCF_001750725.1	98.91	0.95	65.71	76.9	1.48	8.7
B73_A_11_1_metabat2_high_PE.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Pantoea;s__Pantoea dispersa	GCF_004655555.2	98.04	0.91	98.33	93.54	1.99	50
B73_I_14_1_metabat2_low_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Rahnella;s__Rahnella aquatilis_B	GCF_000255535.1	99.33	0.99	98.06	100	0.11	100
B73_I_4_3_metabat2_high_PE.005.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Rahnella;s__Rahnella aquatilis_B	GCF_000255535.1	99.49	0.99	98.13	99.84	0	0
B73_L_14_2_metabat2_low_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Rahnella;s__Rahnella aquatilis_B	GCF_000255535.1	99.38	0.98	95.81	94.47	0.94	0
IL14H_L_6_1_metabat2_high_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Rahnella;s__Rahnella aquatilis_B	GCF_000255535.1	99.56	0.99	97.32	97.85	0.49	0
IL14H_L_4_1_maxbin2_low_prob.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Rahnella;s__Rahnella aquatilis_B	GCF_000255535.1	99.23	0.96	97.7	100	1.25	20
IL14H_I_6_3_maxbin2_low_prob.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Serratia;s__Serratia liquefaciens	GCF_000422085.1	98.57	0.93	93.67	98.55	0.89	0
IL14H_L_6_1_maxbin2_low_prob.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Serratia;s__Serratia liquefaciens	GCF_000422085.1	98.63	0.95	95.2	98.1	2.07	0
IL14H_L_6_3_metabat2_low_PE.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Serratia;s__Serratia liquefaciens	GCF_000422085.1	98.97	1	93.15	99.02	0.45	0
IL14H_I_11_1_metabat2_high_PE.005.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Serratia;s__Serratia plymuthica	GCF_001590725.1	97.81	0.95	90.97	96.41	0.44	0
MO17_L_6_3_metabat2_low_PE.007.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Enterobacteriaceae;g__Serratia;s__Serratia plymuthica	GCF_001590925.1	97.72	0.93	92.18	96.35	1	66.67
B73_L_6_3_metabat2_high_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Cellvibrionaceae;g__Cellvibrio;s__	N/A	N/A	N/A	74.19	80.35	0.65	42.86
B73_A_4_1_metabat2_low_PE.005.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	97.34	1	92.52	90.73	0.36	33.33
B73_A_4_2_metabat2_low_PE.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	97.14	0.99	93.29	85.67	0.44	0
B73_L_14_1_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.42	0.98	97.32	99.12	4.71	93.48
B73_L_14_4_metabat2_high_PE.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.7	1	98.23	97.98	0.21	0
B73_L_6_1_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.61	0.98	97.7	100	2.42	73.81
B73_L_6_2_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.71	0.98	98.06	100	3.62	80.95
IL14H_I_11_1_maxbin2_low_prob.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.61	0.98	98.23	100	1.33	88.24
IL14H_I_6_1_maxbin2_high_prob.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.41	0.99	93.37	98.67	1.53	66.67
IL14H_I_14_1_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.68	0.98	93.53	90.31	1.64	4
IL14H_L_6_2_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.74	1	90.71	82.24	0.1	0
IL14H_L_6_3_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.68	0.98	98.23	100	2.69	68.18
MO17_I_11_1_maxbin2_high_prob.010.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.71	0.99	98.19	100	0.56	55.56
MO17_I_11_2_metabat2_high_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.57	0.99	95.1	99.18	1.76	73.33
MO17_I_11_3_metabat2_high_PE.014.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.52	0.99	93.29	96.99	1.19	42.86
MO17_I_14_3_maxbin2_low_prob.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.7	0.99	98.19	100	0.1	0
MO17_I_14_3_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.66	0.98	97.98	100	2.78	65.62
MO17_I_4_2_metabat2_low_PE.007.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.65	1	92.86	96.58	1.07	20
MO17_L_14_1_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.76	1	89.88	75.39	0.4	83.33
MO17_L_14_2_maxbin2_low_prob.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.75	0.99	98.23	100	0.5	30
MO17_L_14_3_maxbin2_low_prob.001.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.73	0.99	98.23	100	0.1	0
MO17_L_4_3_metabat2_high_PE.006.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.74	0.99	98.23	96.88	0.1	0
MO17_L_6_1_metabat2_high_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.71	1	98.23	100	0.18	0
MO17_L_6_3_maxbin2_low_prob.002.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.48	0.99	81.83	75.28	2.2	61.9
MO17_L_8_1_metabat2_high_PE.004.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.7	0.99	98.23	99.69	0.18	0
MO17_L_8_2_L_metabat2_low_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_C	GCF_000194805.1	99.77	0.99	98.23	100	0.1	0
IL14H_A_8_1_metabat2_low_PE.006.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E silesiensis	GCF_001661075.1	N/A	N/A	95.65	87.77	0.48	28.57
B73_I_14__maxbin2_high_prob.001	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	96.17	0.98	96.01	93.07	1.5	94.44
B73_I_4__metabat2_high_PE.003	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	96.81	0.92	91.73	85.58	4.86	81.13
IL14H_I_14__metabat2_high_PE.009	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	95.62	0.96	91.98	85.76	1.03	70
IL14H_I_8__metabat2_high_PE.003	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	N/A	N/A	95.77	85.77	0.08	0
MO17_I_11__maxbin2_high_prob.005	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	95.93	0.97	91.19	92.15	3.96	90.57
MO17_I_14__metabat2_high_PE.012	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	95.29	0.93	93.81	91.91	1.92	73.91
MO17_L_4__maxbin2_low_prob.009	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315	GCF_000282315.2	N/A	N/A	69.48	88.37	2.13	51.43
IL14H_I_8_2_metabat2_high_PE.003.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Rhodanobacteraceae;g__Luteibacter;s__	N/A	82.29	0.67	66.77	76.22	1.99	54.55
MO17_I_11_1_metabat2_low_PE.011.contigs	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter;s__Lysobacter enzymogenes	GCF_900106525.1	98.08	0.97	80.38	85.55	2.37	13.33

Table 1 - Metagenomically-assembled genomes from the 2010 field season maize rhizosphere metagenomes with CheckM completeness >70% and contamination <5%. s__Pseudomonas_E brassicacearum_C is the outdated assignment for s__Pseudomonas_E brassicacearum (GTDB R89-R95).

Of the total MAGs passing QC from the 2010 field season rhizosphere metagenome, 40% (33/81) of the total high-quality MAGs were of three *Pseudomonas* species: *P. brassicacearum, P. silesiensis,* and *P. sp000282315.* Of those, 25 MAGs were *P. brassicacearum,* 7 *P. sp000282315,* and a single *P. silesiensis.* Overall the MAGs that passed QC were composed mainly of Proteobacteria (63/81), and Bacteroidia (17/81). Only one taxa of Bacteroidia*, Pedobacter ginsenosidimutans*, was identified at species level. The closest placement ANI for the remaining taxa ranged from 93-75%. As reflected by the relative abundances of Enterobacteriaceae within each metagenome, multiple MAGs were obtained for *Rahnella aquatilis B, Serratia liquefaciens,* and *Serratia plymuthica.* Among the Burkholderia taxa, MAGs were also obtained for *Duganella zoogloeoides, Ralstonia sp000955795,* and multiple *Variovorax paradoxus.*

The performance of the metagenome assemblers (metacompass and metaspades) was hindered by the high proportion of closely related *Pseudomonas* within each of the 2010 field season metagenomes. Despite the abundance of reads, a great deal of those taxa were 95-99% ANI, thus challenging to distinguish which accessory

genome reads were associated with which *Pseudomonas* species or strain. This was evidenced by the low number of total *Pseudomonas* MAGs, but also many of those MAGs were determined >5% contaminated. Even MAGs passing QC exhibited high strain heterogeneity.

In the case of high abundance coupled with high strain diversity, the total number of *Pseudomonas* MAGs generated was limited but still produced 25 high quality *P. brassicacearum* genomes. Other high quality MAGs of the genera *Pedobacter, Variovorax, Rahnella, Serratia, Duganella,* and *Lysobacter* were produced, along with several novel *Sphingobium* genera and species.

3.3.3. Phylogenetic group and high abundance *Pseudomonas* pangenomes

To better characterize the most highly abundant *Pseudomonas* species of the 2010 season, I collected all the strains available to me though NCBI, isolates from the culture collection, and MAGs produced from the rhizosphere metagenomes and performed a pangenome analysis (table 2, figure 20). In addition, I performed analysis of each phylogenetic group to determine the proportions of genes and GCs assigned to core, accessory, and singleton pangenomes (table 3, figure 21). Core genes exhibiting functional amino acid substitutions were annotated for each group pangenome to investigate which highly conserved genes are potentially undergoing diversification (table 4).

species_pangenomes	total_genomes	total_genes	total_uniq_genes	total_GCs	total_unique_GCs	percent_paralogs
silesiensis_core		20892	19242	3482	3207	7.897759908
silesiensis_accessory		15822	14196	2637	2366	10.27682973
silesiensis_singletons		12558	12186	2093	2031	2.962255136
silesiensis_pangenome	6	49272	45624	8212	7604	7.403799318
frederiksbergensis_core		33887	32207	4841	4601	4.957653377
frederiksbergensis_accessory		5950	5649	850	807	5.058823529
frederiksbergensis_singletons		9667	9597	1381	1371	0.724112962
frederiksbergensis_pangenome	7	49504	47453	7072	6779	4.143099548
brassicacerum_core		160018	108452	4211	2854	32.22512467
brassicacerum_accessory		94886	75582	2497	1989	20.3444133
brassicacerum_singletons		71516	70794	1882	1863	1.009564293
brassicacerum_pangenome	38	326420	254828	8590	6706	21.93247963
sp000282315_core		30087	24633	3343	2737	18.12743045
sp000282315_accessory		27252	24777	3028	2753	9.081902246
sp000282315_singletons		12933	12627	1437	1403	2.366040362
sp000282315_pangenome	9	70272	62037	7808	6893	11.71875

Table 2 - Counts for genes, gene clusters, and percentage of paralogous genes for each classification within each species pangenome. The core genome is defined by genes present in all genomes (total), accessory by genes present in a minimum of 2 but less than the total minus 1, and singletons by genes present in only one genome. Unique genes are counted by only calling the representative gene for each cluster, whereas the total is all genes for all clusters. Paralogous genes represent gene duplications within each genome.

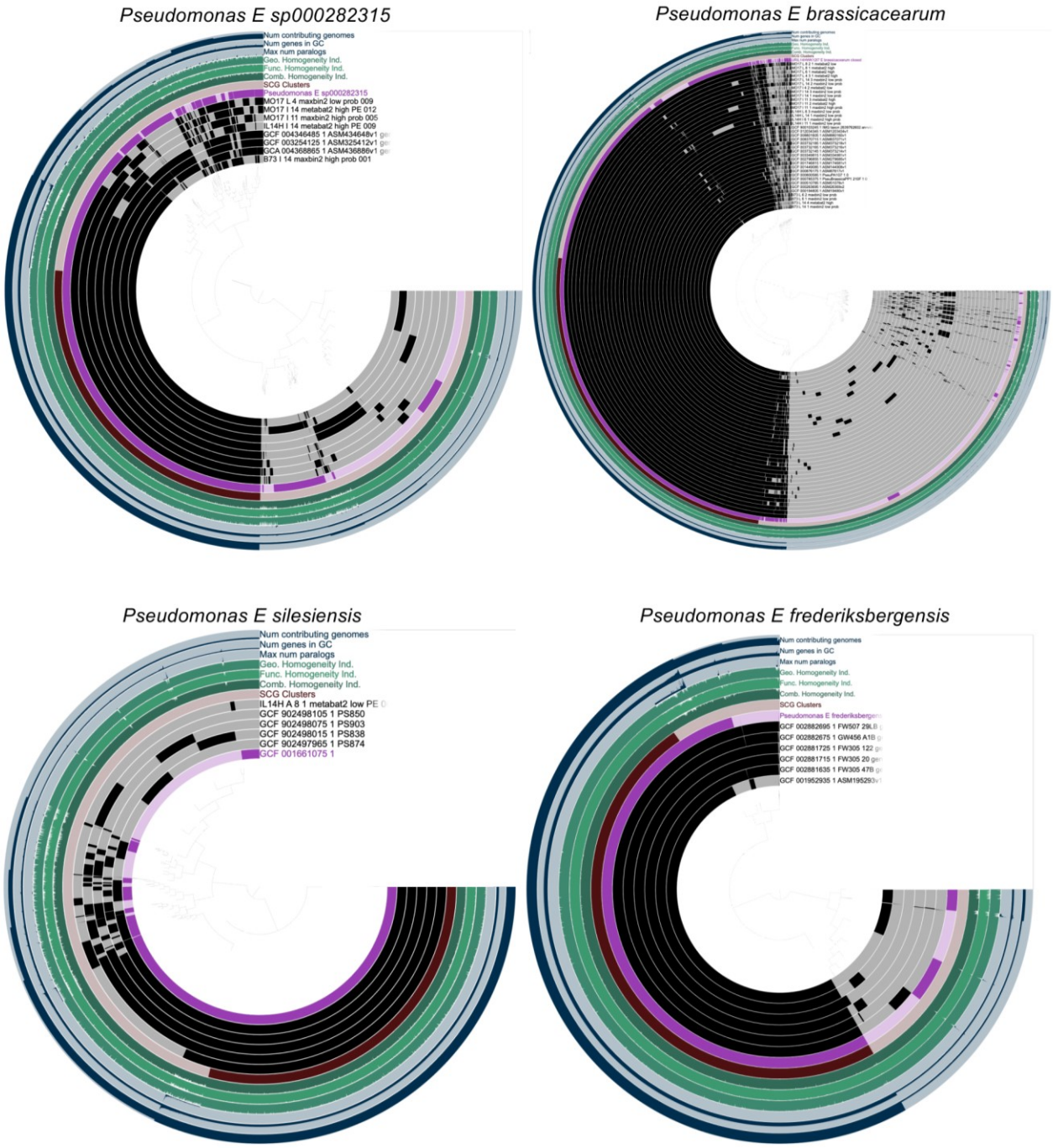


Figure 20 - Selected pangenomes of high abundance *Pseudomonas* species.
Each species contains all known isolates binned to that taxa by GTDB R95, plus MAGs obtained from the 2010 field season maize rhizospheres with >70% completion and <5% contamination, ordered by gene presence/absence with Euclidean distances and Ward linkage. The strain highlighted in purple is the representative genome for that species within the genus phylogeny. The term “SGC Clusters” refers to single-copy genes. The terms “Geo. Homogeneity”, “Func. Homogeneity”, and “Comb. Homogeneity” refer to the

geometric homogeneity index (gap/residue distribution between genomes within each GC) and the functional homogeneity index (biochemical similarity in amino acid residues between each genome within each GC). Genomes with the prefix “IL14H”, “MO17”, or “B73” are maize rhizosphere MAGs.

supergroup	group_pangenome	total_genomes	total_genes	total_unique_genes	total_GCs	total_unique_GCs	percent_paralogs
NF	Anguilliseptica_core		32334	27234	1902	1602	15.77287066
NF	Anguilliseptica_accessory		107848	94299	6344	5547	12.5630517
NF	Anguilliseptica_singletons		132804	128214	7812	7542	3.456221198
NF	Anguilliseptica_pangenome	17	272986	249747	16058	14691	8.512890771
1	Syringae_core		67177	45477	2167	1467	32.30272266
1	Syringae_accessory		315983	262074	10193	8454	17.06072795
1	Syringae_singletons		323640	315704	10440	10184	2.45210728
1	Syringae_pangenome	31	706800	623255	22800	20105	11.82017544
1	Putida_core		81363	66896	2199	1808	17.78080946
1	Putida_accessory		364339	305583	9847	8259	16.12673911
1	Putida_singletons		436711	428238	11803	11574	1.940184699
1	Putida_pangenome	37	882413	800717	23849	21641	9.258249822
2	Asplenii_core		19761	17101	2823	2443	13.46085724
2	Asplenii_accessory		29589	26110	4227	3730	11.75774781
2	Asplenii_singletons		35287	34223	5041	4889	3.015274747
2	Asplenii_pangenome	7	84637	77434	12091	11062	8.510462327
2	Fragi_core		37562	32032	2683	2288	14.72232575
2	Fragi_accessory		55356	47782	3954	3413	13.68234699
2	Fragi_singletons		65156	63728	4654	4552	2.191663086
2	Fragi_pangenome	14	158074	143542	11291	10253	9.193162696
2	Chlororaphis_core		48263	40443	2839	2379	16.20288834
2	Chlororaphis_accessory		115124	99076	6772	5828	13.93975192
2	Chlororaphis_singletons		135762	131733	7986	7749	2.967693464
2	Chlororaphis_pangenome	17	299149	271252	17597	15956	9.325453202
2	Fluorescens_core		174592	101618	2816	1639	41.796875
2	Fluorescens_accessory		753672	594084	12156	9582	21.17472853
2	Fluorescens_singletons		803520	782936	12960	12628	2.561728395
2	Fluorescens_pangenome	62	1731784	1478638	27932	23849	14.61764285
3	Corrugata_core		77400	56220	2580	1874	27.36434109
3	Corrugata_accessory		246030	207060	8201	6902	15.83953176
3	Corrugata_singletons		261180	254580	8706	8486	2.526992878
3	Corrugata_pangenome	30	584610	517860	19487	17262	11.41786832
3	Mandelii_core		110884	74252	2918	1954	33.03632625
3	Mandelii_accessory		366130	294576	9635	7752	19.5433316
3	Mandelii_singletons		401394	392730	10563	10335	2.158477705
3	Mandelii_pangenome	38	878408	761558	23116	20041	13.30247448
3	Jessenii_core		85764	66388	3063	2371	22.59222984
3	Jessenii_accessory		202496	164724	7232	5883	18.65320796
3	Jessenii_singletons		222908	218064	7961	7788	2.173093832
3	Jessenii_pangenome	28	511168	449176	18256	16042	12.12751972
3	Koreensis_core		157250	120200	3145	2404	23.56120827
3	Koreensis_accessory		369000	308150	7380	6163	16.49051491
3	Koreensis_singletons		484800	473350	9696	9467	2.36179868
3	Koreensis_pangenome	50	1011050	901700	20221	18034	10.81548885

Table 3 - Counts for genes, gene clusters, and percentage of paralogous genes for each classification within each phylogenetic group with >5 genomes. The core genome is defined by genes present in all genomes (total), accessory by genes present in a minimum of 2 but less than the total minus 1, and singletons by genes present in only one genome. Unique genes are counted by only calling the representative gene for each cluster, whereas the total is all genes for all clusters. Paralogous genes represent gene duplications within each genome.

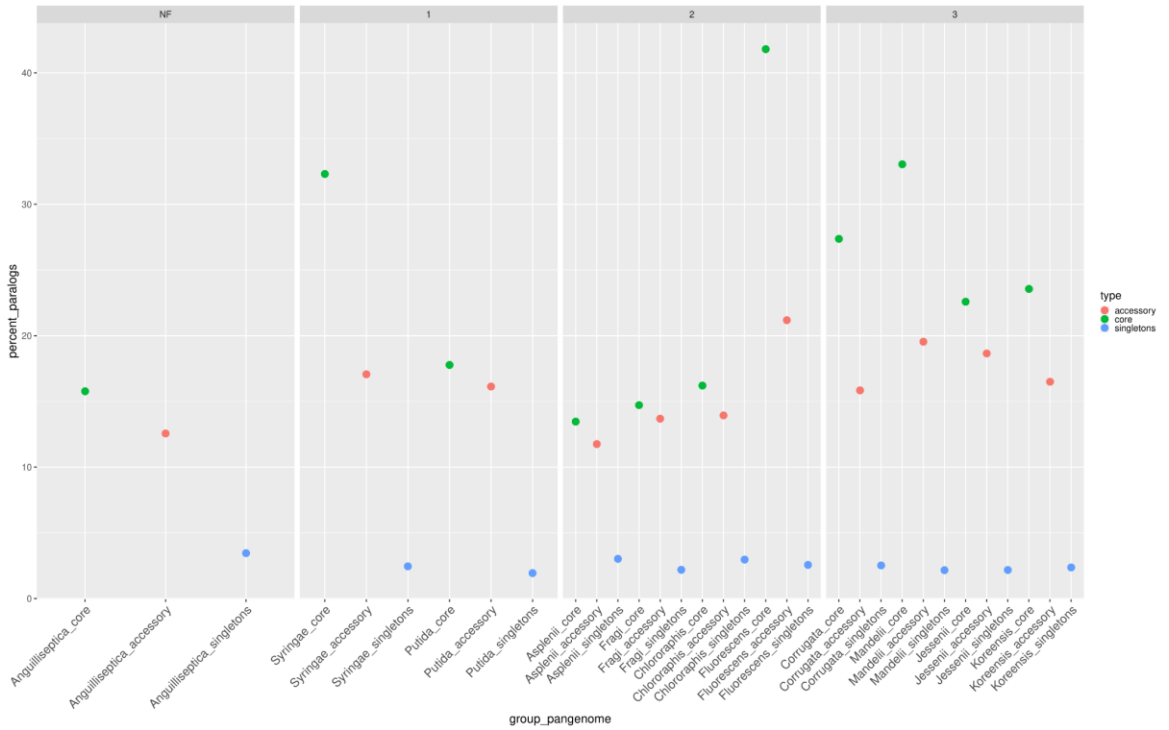


Figure 21 - Paralogous gene percentage by group pangenome compartment. The order of each phylogenetic group on the x-axis reflects the phylogenetic tree order, with supergroup above. Only groups with >5 genomes included.

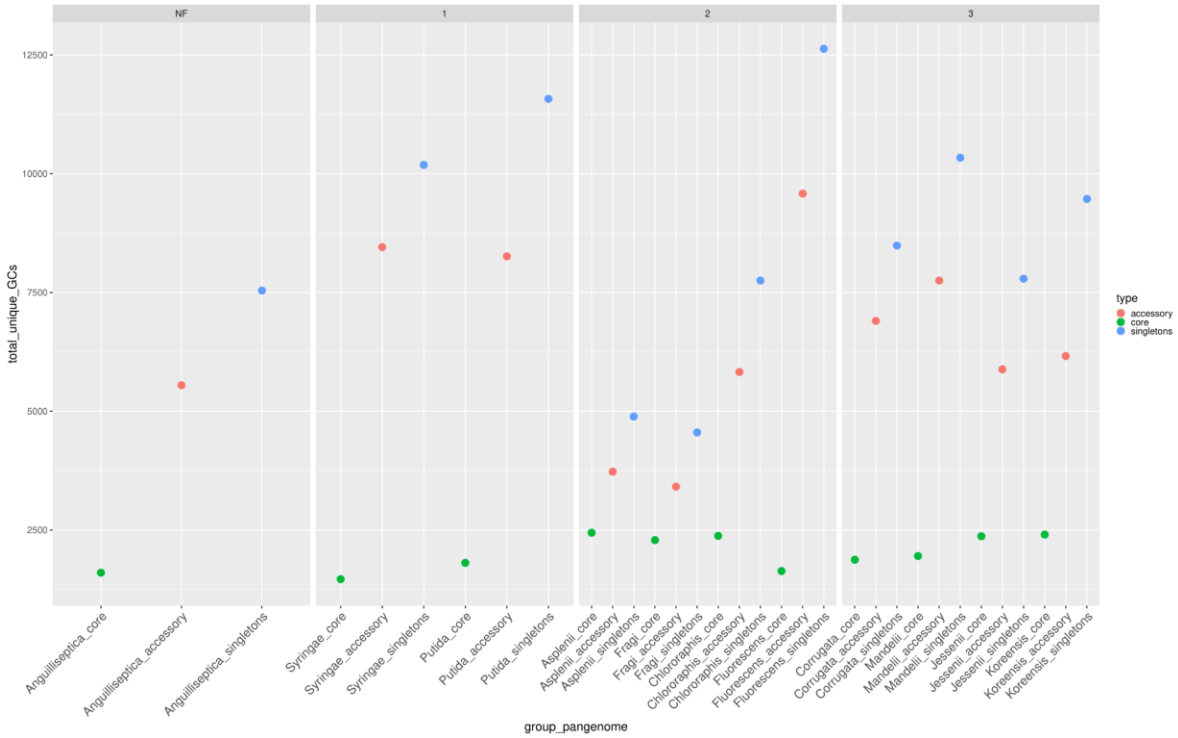


Figure 22 - Total unique gene clusters for each group pangenome compartment. The order of each phylogenetic group on the x-axis reflects the phylogenetic tree order, with

supergroup above. Only groups with >5 genomes included.

Group	KO	Definition
Anguilliseptica	K03807	ampE; AmpE protein
Anguilliseptica	K00537	arsC; arsenate reductase (glutaredoxin) [EC:1.20.4.1]
Anguilliseptica	K06598	chpC; chemosensory pili system protein ChpC
Anguilliseptica	K00768	E2.4.2.21, cobU, cobT; nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase [EC:2.4.2.21]
Anguilliseptica	K05787	hupA; DNA-binding protein HU-alpha
Anguilliseptica	K03530	hupB; DNA-binding protein HU-beta
Anguilliseptica	K11719	lptC; lipopolysaccharide export system protein LptC
Anguilliseptica	K04754	mlaA, vacJ; phospholipid-binding lipoprotein MlaA
Anguilliseptica	K00992	murU; N-acetyl-alpha-D-muramate 1-phosphate uridylyltransferase [EC:2.7.7.99]
Anguilliseptica	K00275	pdxH, PNPO; pyridoxamine 5'-phosphate oxidase [EC:1.4.3.5]
Anguilliseptica	K00564	rsmC; 16S rRNA (guanine1207-N2)-methyltransferase [EC:2.1.1.172]
Anguilliseptica	K03672	trxC; thioredoxin 2 [EC:1.8.1.8]
Anguilliseptica	K01011	TST, MPST, sseA; thiosulfate/3-mercaptopyruvate sulfurtransferase [EC:2.8.1.1 2.8.1.2]
Anguilliseptica	K04085	tusA, sirA; tRNA 2-thiouridine synthesizing protein A [EC:2.8.1.-]
Anguilliseptica	K11179	tusE, dsrC; tRNA 2-thiouridine synthesizing protein E [EC:2.8.1.-]
Anguilliseptica	K25422	yceF; 7-methyl-GTP pyrophosphatase [EC:3.6.1.-]
Asplenii	K10024	aotQ; arginine/ornithine transport system permease protein
Asplenii	K02189	cbiG; cobalt-precorrin 5A hydrolase [EC:3.7.1.12]
Asplenii	K04127	cefD; isopenicillin-N epimerase [EC:5.1.1.17]
Asplenii	K01792	E5.1.3.15; glucose-6-phosphate 1-epimerase [EC:5.1.3.15]
Asplenii	K11903	hcp; type VI secretion system secreted protein Hcp
Asplenii	K22549	lhpC; D-hydroxyproline dehydrogenase subunit alpha [EC:1.5.99.-]
Asplenii	K08967	mtnD, mtnZ, ADI1; 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase [EC:1.13.11.53 1.13.11.54]
Asplenii	K23124	pxpC; 5-oxoprolinase (ATP-hydrolysing) subunit C [EC:3.5.2.9]
Asplenii	K03546	sbcC, rad50; DNA repair protein SbcC/Rad50
Asplenii	K11065	tpx; thioredoxin-dependent peroxiredoxin [EC:1.11.1.24]
Asplenii	K03671	trxA; thioredoxin 1
Chlororaphis	K02655	pilE; type IV pilus assembly protein PilE
Fragi	K01719	hemD, UROS; uroporphyrinogen-III synthase [EC:4.2.1.75]
Fragi	K07336	K07336; PKHD-type hydroxylase [EC:1.14.11.-]
Fragi	K07215	pigA, hemO; heme oxygenase (biliverdin-IX-beta and delta-forming) [EC:1.14.99.58]
Fragi	K03546	sbcC, rad50; DNA repair protein SbcC/Rad50
Fragi	K07236	tusC, dsrF; tRNA 2-thiouridine synthesizing protein C
Jessenii	K03612	rnfG; H+/Na+-translocating ferredoxin:NAD+ oxidoreductase subunit G
Putida	K02399	flgN; flagellar biosynthesis protein FlgN
Putida	K02838	frr, MRRF, RRF; ribosome recycling factor
Putida	K07740	rsd; regulator of sigma D
Syringae	K12368	dppA; dipeptide transport system substrate-binding protein
Syringae	K01057	PGLS, pgl, devB; 6-phosphogluconolactonase [EC:3.1.1.31]
Syringae	K03745	slyX; SlyX protein

Table 4 - High confidence annotations of core genes with high geometric (>.99) but low functional (<0.85) homogeneity.

The four *Pseudomonas* used for species-level pangenomes were P. *brassicacearum,* P. *frederiksbergensis E,* P. *silesiensis,* and P. *sp000282315.* The P. *brassicacearum* was the species with the majority of high-quality MAGs, whereas only one P. *silesiensis* MAG passed quality filtering, and P. *frederiksbergensis E* had no MAGs that passed quality thresholds. The only other *Pseudomonas* species with quality MAGs was P. *sp000282315.* The pangenome for each of these taxa included the MAGs and all strains assigned within GTDB R95 and R202. The pangenome of P. *brassicacearum* had the most strains, with 18 isolates and 20 MAGs. The strain “URIL14HWK12:I7 E brassicacearum closed” was recultured from the original glycerol stock used to produce the draft genome that is publically available, and was assembled using PacBio long read sequencing to close the genome.

Each species pangenome was defined as all genes present in all genomes (n), or core as genes present in all genomes, accessory as genes present in at least two but no

more than n-1 genomes, and singleton genes that are present in only one genome. The difference in unique genes (defined by a representative gene within each cluster) versus total genes was used to calculate the number of paralogs within each compartment of the pangenome (table 2). For each species except P. *frederiksbergensis E* the majority of paralogs were among core genes, followed by accessory genes. Singleton genes are rarely duplicated. Overall, the pangenome of P. *brassicacearum* with 38 genomes contained 21.9% paralogous genes, followed by P. *sp000282315* with 11.7%.

Concerning each species, P. *frederiksbergensis E* exhibits the largest core genome at 68% of the pangenome, with 20% accessory and 12% being singletons. The percentages for P. *brassicacearum* are 42%, 30%, and 28% for the core, accessory, and singleton pangenomes respectively. The core and accessory pangenomes of P. *sp000282315* were the same at 40%, with 20% for singletons. For P. *silesiensis,* the core was closer to P. *frederiksbergensis E* at 57%, but the accessory pangenome was relatively modest at 17%, followed by 26% for singletons.

The pangenome of P. *brassicacearum* contains 20 MAGs, and based on the alignments of figure 20, the MAGs contain disproportionately fewer singleton genes than the isolate-derived genomes. However, the MAGs did not appear to dramatically affect the number of core genes or single-copy genes. Based on the euclidean distances of each genome the publicly-available isolates cluster together, with the exception of URIL14WK12I7. Notably, the MAGs cluster not by field but by maize genotype. Despite the high abundance of *Pseudomonas* from the Aurora field location, no P. *brassicacearum* MAGs produced from those metagenomes passed quality thresholds.

The pangenome of P. *sp000282315* contains five MAGs and four isolates. The four publicly-available genomes are most similar to each other, including the representative genome. The MAGs for this species contained relatively few singletons, but more than with P. *brassicacearum* MAGs. The core pangenome for P. *silesiensis* is likely to be truncated by incompleteness in the one MAG present (checkM completeness 87%). Unsurprisingly, the pangenome of P. *frederiksbergensis E* is predominantly core due to the limited number of genomes and their origin as isolates, not MAGs.

Expanding the pangenomes to the phylogenetic group level, as expected the core genome of each is less than at species level. Excluding the groups with <5 species, the core genomes of the groups with the least number of genomes (Asplenii and Fragi) were both 23%. Overall, the proportions of accessory and singleton genomes vary between 30-40%. Comparing the proportions of paralogous genes per group pangenome, figure 21 indicates that while the singleton genes are rarely duplicated, the core contains the most. The most notable difference is between the core and accessory genomes, where the groups with the most variance in percent paralogs between core and accessory are Syringae, Fluorescens, Corrugata, and Mandelii. The total number of GCs by pangenome compartment per group is largely reflective of the number of genomes in each of the groups, but it is worth noting the degree of variance between the compartments is not

directly proportional to the number of genomes (figure 22). Specifically, even though Syringae has half the number of genomes Fluorescens, it has fewer GC in the core pangenome, yet has 10,184 unique singleton GCs whereas group Fluorescens has 12,628. Group Putida is also very similar, albeit with a larger core genome. Groups within supergroup 1 (as just observed) are consistent, as is supergroup 3. Supergroup 2 however is likely hindered by the underrepresentation of groups Asplenii, Fragi, and Chlororaphis.

To explore potential functional distinctions between the core pangenomes of each group with specific emphasis on genes with high geometric homogeneity and low functional homogeneity (table 4). The purpose of this selection was to identify core genes that vary in sequence between species within each group to avoid housekeeping genes and focus on those with potentially competitive morphological or biochemical phenotypes. Gene annotations and KO numbers were determined and the results are presented on table 5. Notably, the method used to annotate these genes was unsuccessful for most of the genes, with only 44.3% (data not shown) being assigned a KEGG orthology number. The groups with the most annotations are Anguilliseptica and Asplenii, followed by Fragi, Syringae, and Putida.

Due to the massive overrepresentation of closely-related *Pseudomonas* within all of the 2010 rhizosphere metagenomes, assembling *Pseudomonas* genomes from these data proved difficult. Despite combined read counts of >25M for all metagenomes the number of reference-based and *de novo* MAGs passing quality thresholds were less than 10%. As evident above the majority of these were composed of *P. brassicacearum* as would be expected, but even with the high combined abundance of other taxa at Aurora in 2010 the *in silico* methods of assembly struggled to produce more than 1-5 other *Pseudomonas* species. Furthermore, all of the MAGs regardless of quality appear to lack singleton genes; of course this to be expected considering how much of the singleton genome of any organism is potentially the result of HGT and may contain genes with a GC content incongruent with the majority of the genome. The lack of singletons is not ideal for pangenome analyses due to the expansive pangenome of *Pseudomonas* as a whole, and it should be taken into consideration when mining for BGCs or other potentially novel phenotypes. A serious limitation of this species-level pangenome analysis is the dearth of representation for the non-fluorescens lineage, and the inclusion of *P. aeruginosa* which is complete and exhaustively annotated compared to the remainder of the genomes.

The pangenomes of a selection of high-abundance species for the 2010 field season were expanded using MAGs derived from the maize-rhizosphere metagenomes. These included numerous taxa including three species of high abundance *Pseudomonas,* which were used in conjunction with all strains of that species publicly available, to construct pangenomes. Despite the computational limitation in differentiating between sequences of closely-related species, the pangenome of *P. brassicacearum* has been

expanded from 17 genomes to 38. This species, as well as *P. silesiensis, P. frederiksbergensis E, and P. sp000282315* contain expansive accessory and singleton pangenomes, and core genomes with between 4.9-32.2% paralogous gene content. Pangenomes at the species phylogenetic group varied substantially in the proportion of paralogous gene content between their core, accessory, and singleton pangenome. When assessing functional diversity between core genes with non-synonymous mutations very few of these could be successfully annotated; of those, the groups Anguilliseptica, Asplenii, Fragi, and Syringae represent predominantly non-Fluorescens and SG1 and SG2 (sans Fluorescens).

3.3.4. Eukaryotic profiling of rhizosphere metagenomes

To characterize microbial eukaryotic diversity within each maize rhizosphere, their metagenomes were profiled by aligning sequences to curated eukaryotic marker genes. The phylum Acomycota is the most diverse, and is highly represented within the maize rhizosphere metagenomes (figure 23). The other highly abundant fungal phyla include the Basidiomycetes and Mucoromycetes (figure 24BC). Within the kingdom Animalia, the only abundant phyla is the Nematoda (figure 24A).

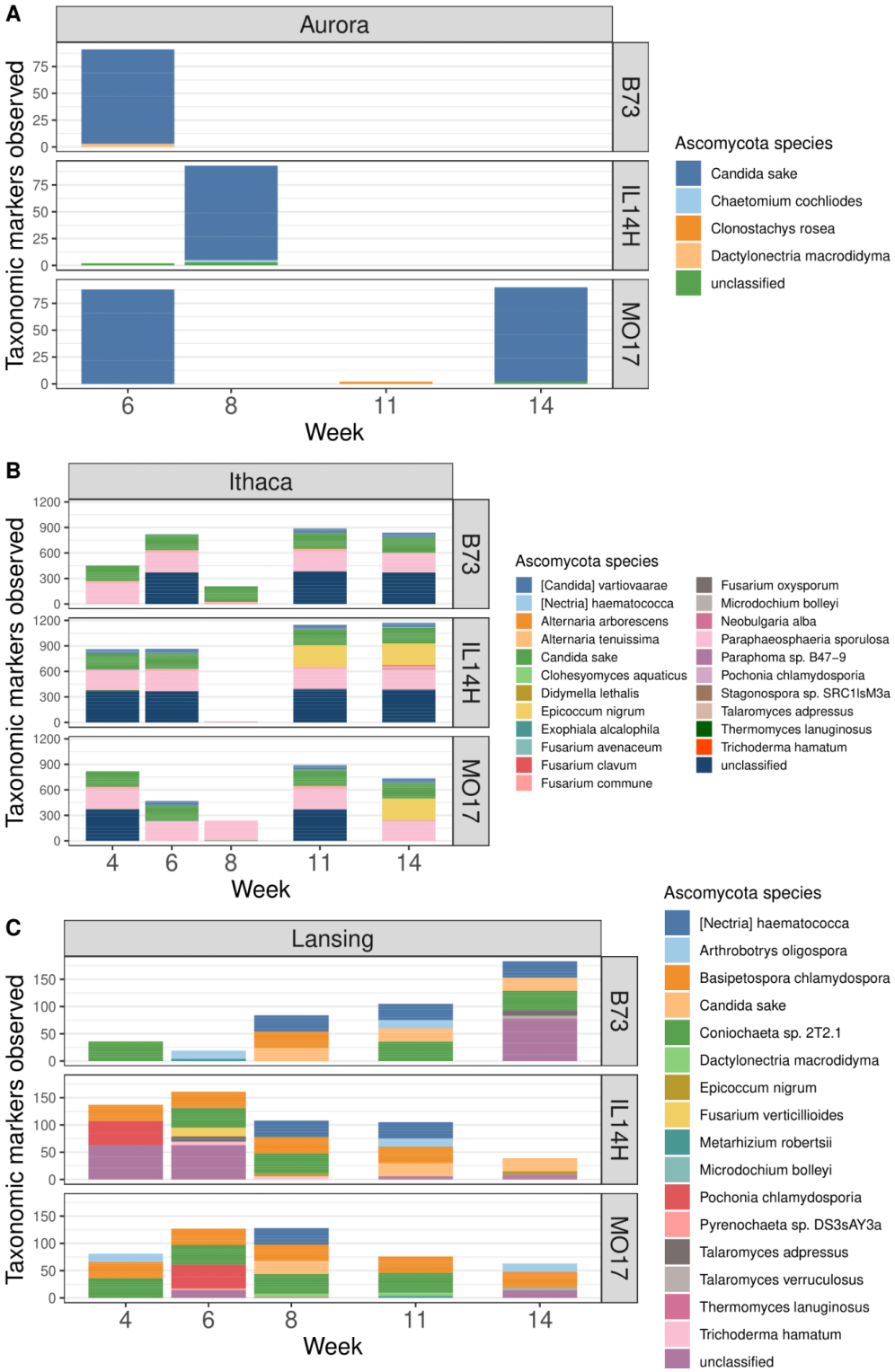


Figure 23 - Observed taxonomic marker counts of Ascomycota species for each field and maize genotype by week.

A) Taxonomic marker counts for Ascomycota species for the Aurora field. B) For Ithaca. C) For Lansing.

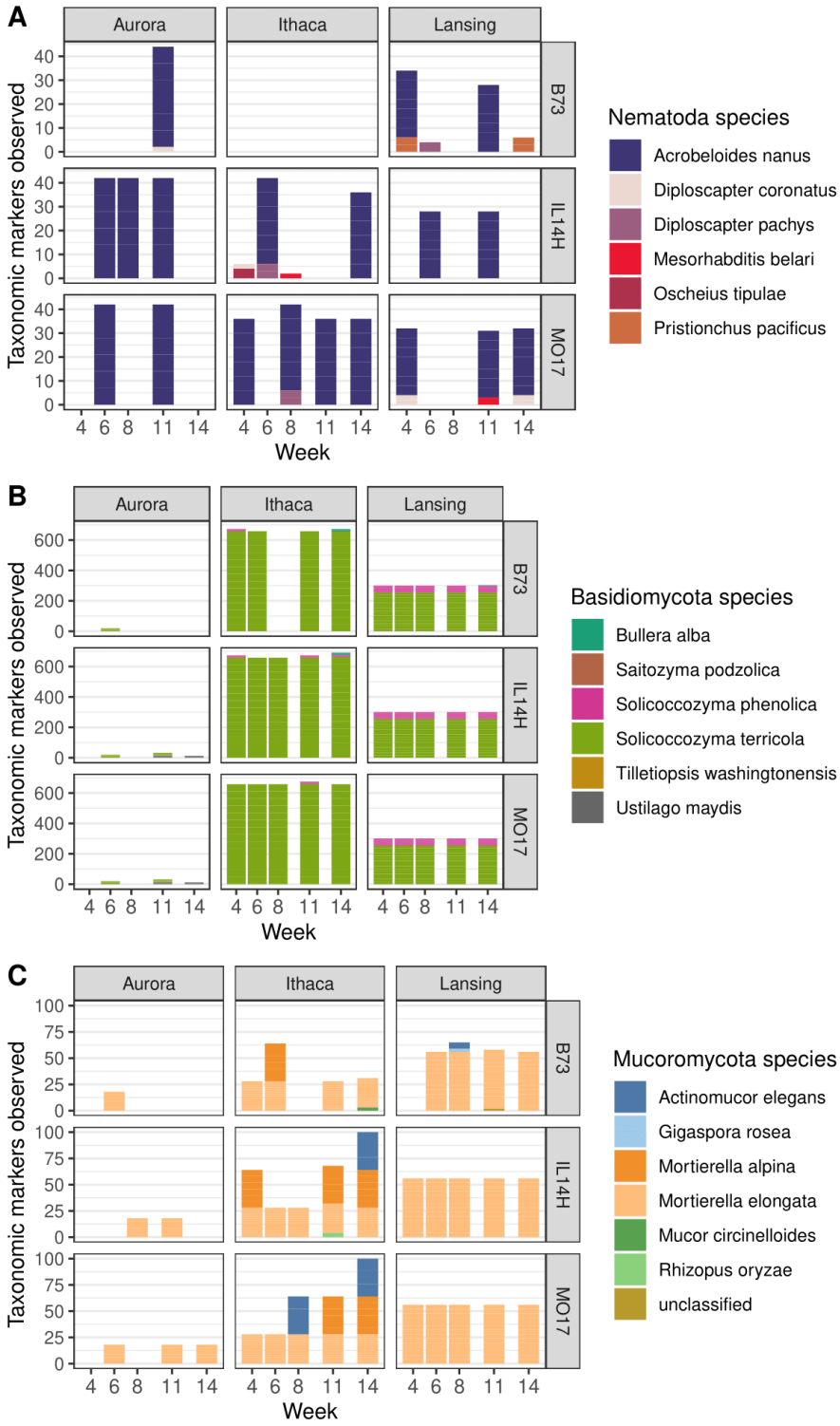


Figure 24 - Observed taxonomic marker counts of low diversity eukaryotic species by week for each field and maize genotype.

A) Taxonomic marker counts for species of the phylum Nematoda. B) Taxonomic marker counts for species of the phylum Basidiomycota. C) Taxonomic marker counts for species of the phylum Mucoromycota.

The reads assigned to eukaryotic taxa were profiled at species level for each field and collection week. The septate fungal phylum Ascomycota was the most diverse and highest in detected taxonomic markers when comparing across all fields. However, Ascomycete abundance and species varied dramatically by individual field. For Aurora, both abundance and diversity are low with only *Candida sake* being the dominant taxa out of four species. This yeast was detected within maize genotypes B73 and IL14H only at weeks six and 8, respectively. For MO17 *C. sake* was detected at weeks 6 and 14, and in all of these cases the number of markers observed was similar (fig 9A).

For the Ithaca field, 22 Ascomycete species were detected with a similar marker distribution between all three maize genotypes (figure 23). The most pervasive Ascomycetes at this location were highly abundant compared to Aurora or Lansing, with individual marker totals >250 of *Candida sake, Paraphaeosphaeria sporulosa,* and *Epicoccum nigrum.* This location had a large proportion of unclassified Ascomycota markers, and *E. nigrum* was observed during the last two weeks. The ther 19 species detected were only trace compared to the aforementioned taxa.

The diversity of Ascomycetes within the Lansing field were greater than Ithaca but the total number of taxonomic markers detected were lower, only slightly greater than the load of Aurora. Unlike Aurora where essentially all reads were attributed to *C. sake*, the 50-170 taxonomic markers found within Lansing were distributed between 16 species again with *C. sake* being prevalent beginning at week 8. The most abundant Ascomycetes in the Lansing field were *Arthrobotrys oligospora, Coniochaeta sp. 2T2.1, Pochonia chlamydosporia, Basipetospora chlamydospora,* and *Fusarium verticillioides* at week 6 for maize genotype IL14H (figure 23C). For maize genotype B73 the last week contained over 50% unclassified Ascomycota and IL14H contained >30% unclassified markers for weeks four and six.

Other eukaryotic phyla observed within these maize rhizosphere metagenomes which were discerned at species level include Nematoda, Basidiomycota, and Murcoromycota. The diversity of these taxa was considerably lower compared to Ascomycota with only one or two dominant species per phyla. For Nematoda, the total number of markers observed per species was the lowest of all the detected eukaryotic phyla and of those nearly all were of *Acrobeloides nanus.* Across each field there were trace Nematoda markers but each were observed at low abundance and were field-specific (figure 24A). The filamentous fungal Basidiomycetes were largely absent within the Aurora field but most abundant at Ithaca. The dominant Basidiomycete species was *Solicoccozyma terricola,* and the total markers observed were stable from week to week for Ithaca and Lansing (figure 24B). Lastly, the fungal phylum Mucoromycota was detected at relatively low levels in each field, again with Aurora being the field with the least Mucoromycete taxonomic markers detected. There were six species total with only three being dominant; these include *Actinomucor elegans, Mortierella alpina,* and *Mortierella elongata.* At Lansing *M. elongata* was highly dominant and stable across each

collection week. For Ithaca, all three abundant species were proportionally similar but *A. elegans* and *M. alpina* fluctuated by week and maize genotype (figure 24C).

There are several considerations when interpreting the abundances of these profile data. First, the extraction method used was not designed specifically to liberate eukaryotic DNA. There is evidence in the literature that extraction efficiency of fungi is reduced in soils with high organic content potentially due to the mechanical protection afforded to mycelia inside particle aggregates (Feinstein, Sul and Blackwood, 2009). This method also cannot distinguish between epiphytes or endophytes; there are genera of Mucoromycota which are mycorrhizal or root-associated commensal fungi that are likely to be underrepresented. Another caveat to these data is the fact that Eukdetect (Lind and Pollard, 2020), despite a commendable effort to create robust *in silico* detection of eukaryotic taxa within metagenomes which is resistant to contaminating bacterial sequences, was validated on a simulated community and uses a marker database limited to characterized eukaryotic genomes and transcriptomes. Fungal diversity in soil is enormous and therefore this method likely underrepresents true fungal abundance.

In summary, the most abundant eukaryotic phyla consisted of Ascomycota and Basidiomycota, where the former was highly diverse by field whereas the latter was predominantly one species across each field. The Aurora location was consistently low in fungal markers, but higher than Ithaca or Lansing in Nematode abundance for the weeks that they were detected. The highest number of fungal reads observed was Ithaca, followed by Lansing. The only species of Ascomycota that remained consistently abundant between all locations was *C. sake*.

Chapter 4
General conclusion and outlook

As expected, field has the strongest effect on maize rhizosphere *Pseudomonas* diversity

Between all metagenomes, evenness and species richness only matters for field and for week, but not genotype. Field contributes the most to the differences in beta diversity; week and genotype do not appear to have an effect. The proximity of these fields geographically may contribute to much of the overlap in taxa present, as abiotic factors that affect microbial abundance by week would have been subject to similar weather patterns of temperature and precipitation. Overall the rhizosphere effect may not have been noticeable if the developmental stages of these cultivars were too similar; the signal between fields could have been improved with better metadata collection.

The *Pseudomonas* bloom of 2010 is composed of numerous closely related species of the fluorescens lineage and was independent of maize genotype and was not recapitulated *in vitro*

The number, quantity, and diversity of *Pseudomonas* observed in 2010 were an anomaly, and the field metadata collected at the time of the study were not sufficient to determine a cause. The limited replication and microcosm respiration assay were informative in that nearly all the species present in 2010 at Aurora were still detected, just with abundances 1-2 orders of magnitude lower. The substrate-enriched *Pseudomonas* isolates were also very informative; the hypothesis that adding root exudate analogs (sugars, organic acids, SA) would stimulate the blooming taxa to grow was, for the most part, incorrect. Despite *P. brassicacearum* being one of the top five most abundant taxa for Aurora in both 2010 and 2015, none of the substrates included in the microcosm respiration assay stimulated it to grow. The non-Fluorescens and *Pseudomonas viridiflava* (group Syringae) taxa within SG1 with the highest abundance for that supergroup may have been bulk soil contamination based on the water controls of the microcosms and the bulk soils of 2015, specifically *P. viridiflava, P. graminis B,* and *P. sp000497835.* The two monosaccharides (glucose and fructose) mostly stimulated SG1 and SG3 taxa with the exception of group Chlororaphis, while glutamic acid experienced a similar pattern with larger magnitude, suggesting the amino group makes a difference despite the nitrogen supplied in the basal salt medium.

Not surprisingly, this combined with a failure to culture it directly from the rhizosphere samples at 2015 week 12 made it very clear the only way to possibly have attained an isolate of the blooming strain *P. brassicacearum* was to culture directly from an Ithaca or Lansing rhizosphere. Those samples have been entirely consumed so unfortunately the only way to capture it would be to return to the Ketola Research Farm. It would probably be easy enough to culture because it persisted well after 5 years at Aurora, but if the farm transitioned away from maize that might affect the dominant populations of rhizosphere-associated microbes.

Overall the lack of information provided from the fields on climate and abiotic measurements, along with the knowledge that the rows were side-dressed with fertilizer prior to tasseling but without the exact date, created a serious void where a linear mixed model would be. Even then, recording the developmental milestones for the individual maize genotypes would have been extremely helpful due to the influence it has on exudation profiles. Including multiple genotypes of maize was more of a confounding factor than anything; the number of rhizospheres available for each was not statistically powerful enough to perform a robust longitudinal analysis. Regardless, the metagenomic profiles of the 2010 rhizospheres were incredibly informative as to why the taxonomic resolution of the 16S rRNA profiles were so limited. The three “big” OTUs encompassed more than 340 species of fluorescens lineage *Pseudomonas*, while the disparity in measured weekly abundances between the 16S rRNA and metagenomes belies the inherent difficulty in comparing the same method performed years apart by different people.

Alignments of highly conserved single copy genes exhibit monophyletic groupings of species; however, overall taxonomic classifications for individual species within those groups can be highly paraphyletic

Over 500 species of *Pseudomonas* were detected using Bracken+GTDB, and the choice to establish the "biological relevance" threshold at 0.0001% in retrospect was too generous causing the tree being difficult to parse. However, once the analysis pivoted to phylogeny, having more taxa meant a more robust tree with the limitation being purely computational. The placement of each phylogenetic group was first informed by the literature using well established type species as the namesake for the group, then tracing back the nodes to form a monophyletic group that includes all taxonomically resolved species known in that group. This is not especially novel, if other than to give phylogenetic context to the 153 taxa still taxonomically unresolved. For instance, out of the 62 species within the group Fluorescens, 31 are listed under their RefSeq number in GTDB. It is unfortunate that bulk soil samples were not recovered from the 2010 season, because judging from the 2015 metagenomes the "rare" taxa detected for the basal groups of the tree are likely bulk soil carryover. Notably however, several of these taxa are rhizosphere associated, but for other host species (ie. *P. oryzihabitans* and rice).

The fluorescens lineage taxa were the most highly represented, but easily partitioned based on phylogeny; groups Syringae and Putida of SG1 were the most basal and the most divergent based on branch length (Putida with 0.01 substitutions per base). At the other end of the tree were groups Corrugata, Mandelii, Jessenii, and Koreensis which were the most derived and least phylogenetically distant from each other. The groups Asplenii, Fagi, and Chlororaphis within SG1 were the most phylogenetically distant from each other outside of SG1, with the group Fluorescens containing the most genomes of any group and with an average of 0.004 substitutions per base (with Koreensis being 0.003). The Fluorescens group taxa were the most numerous group with the detection threshold set, but outside of Ithaca they were consistently low abundance. *P. fluorescens* as the type species for that group highlights the problem of paraphyly with certain species of *Pseudomonas*. Even when using standardized taxonomy by protein phylogeny, there are many species that have GTDB subclades which are present in multiple phylogenetic groups, such as *Pseudomonas fluorescens* (species present in Chlororaphis and at least one in every SG3 group), *Pseudomonas putida* (species present in Jessenii and Koreensis), *Pseudomonas syringae (P. syringae E* present in Mandelii)*, Pseudomonas Chlororaphis (P. chlororaphis* present in Corrugata)*, Pseudomonas frederiksbergensis* (species present in Mandelii and Corrugata), *Pseudomonas brassicacearum (P. brassicacearum* present in Mandelii), *Pseudomonas mandelii (P. mandelii B* present in Fluorescens), & *Pseudomonas fulva* which is the only taxa to cross lineages to both Putida and Anguilliseptica.

Quite notably, the highest abundance species were not known for pathogenicity but were closely related to pathogens, as the type species of group Corrugata causes

pith necrosis in tomato. However, some strains of *P. brassicacearum* have shown evidence of pathogenicity on tomato while being growth promoting on other species (Belimov *et al.*, 2007; Gislason and de Kievit, 2020), so it's likely that the lifestyles of individual species within this group may depend heavily on the host and potentially other environmental factors such as nutrient availability or competition between other microbes.

Overall, 43.8% of all fluorescens lineage species are not taxonomically resolved within GTDB compared to the 17.9% of the non-fluorescens lineage species. This, combined with the paraphyly of highly represented species, suggests taxonomy is particularly unhelpful when discussing *Pseudomonas* phylogeny or attempting to predict phenotypes. Poor assemblies or incomplete genomes may be a contributing factor in these incongruencies, and if a pangenome were attempted on a paraphyletic species it would absolutely behoove the researcher to closely scrutinize this.

Pangenome gene presence/absence is predictive of both group and supergroup; group Fluorescens being the most dissimilar to all other groups, including within its own supergroup

When comparing each genome on the basis of pangenome gene cluster presence/absence, ordinating them in two dimensions provides a much clearer measure of similarity between groups and supergroups. If gene cluster content may suggest potential functional similarities between genomes, the most obvious indicator is the overall similarity between group Fluorescens and all other groups. Notably, SG1 forms a distinct cluster with the non-Fluorescens groups nesting within it, and the SG3 groups Mandelii, Jessenii, and Koreensis clustering as would be predicted by the phylogenetic tree. However, SG2 acts almost as a bridge between these supergroup clusters with groups Chlororaphis and Asplenii clustering by Corrugata in SG3 and Fragi clustering next to Syringae in SG1. Again, group Fluorescens clusters independently on both principal coordinates.

Lastly, when the highly abundant taxa are mapped onto this pangenome ordination, the distribution clearly indicates that successful taxa seem to be phylogenetically similar at the supergroup level, with 2010 being almost exclusively SG3 while 2015 being nearly all SG1. Highly abundant taxa within the group Fluorescens spanned both years, but were less common.

In silico* prediction of broad metabolic phenotypes suggest a moderate distinction between non-Fluorescens and fluorescens lineage *Pseudomonas

As is hinted at by the pangenome PCoA, the group Anguilliseptica is phenotypically more similar to SG1, while SG2 and SG3 cluster similarly. The very notable exception is group Fragi which clusters independently but between Oryzihabitans and Syringae, which is congruent with the clustering pattern observed in the pangenome PCoA. The abundance phenotype suggests only subtle variations between the high and low groups

and this may be related to the rhizosphere and how it is likely enriched in host-associated taxa. Low abundance species are possibly just inactive bulk soil residents. These results are sensitive to genome assembly quality and likely underestimates what traits are actually present; even with “good” genomes there may be unrecognized homologs that Traitar interprets as absent. Between the high and low abundance phenotype comparisons, denitrification and amino acid metabolism are the defining characteristics.

Antibiotic resistance and metal tolerance are less predictive of group and supergroup membership than virulence factors

When ordinating the presence/absence of AMR genes per genome, the first two principal coordinates were suggestive that the “tail” of x-axis (V1) was potentially an artifact of the method used. Specifically, the imbalance of hits per database was concerning, and when simply plotting the number of hits to each database against the genome’s V1 position, it was clear that the number of hits to BacMet2 was the sole contributor to the first principal coordinate. It is not clear why some genomes lack annotations within that database considering there is no relationship between number of contigs, CheckM completion, or genome size that would indicate a lack of AMR genes in those assemblies.

What follows is that the other database hits are likely the drivers of clustering along the y-axis (V2). The effect isn’t as strong as with the x-axis and BacMet2 (see supplementary figure 6) but when comparing the number of hits to the other databases (VFDB, CARD) on this axis the signal is mostly attributed to VFDB (r^2 value = 0.347). It’s not nearly as concrete an association as the former, but it is telling that virulence factor genes weigh more heavily in the distinction between groups than hits to the antibiotic resistance-specific databases. When ordinating just the VFDB hits there is clear clustering by group that again places Fluorescens separate from SG1 and SG3 groups, opposite NF on the V2 principal coordinate. Notably, very much unlike the whole pangenome ordination where the NF groups were nestled in between Syringae and Putida, in this case the SG3 groups were sandwiched between them with the exception of Corrugata. The VFDB includes a wide range of functions ranging from secretion systems, effector delivery, biofilms, flagellins, exotoxins, adherence, and other genes relating to virulence in plants and animals; despite the type species of Aeruginosa, Anguilliseptica, Syringae and Corrugata being known pathogens, this level of dimension reduction can’t predict virulence. However, it does clearly indicate virulence gene content contributes to the differences seen at a broad phylogenetic scale in *Pseudomonas.*

Overall, comparing the ordination of AMR vs pangenome genes suggests that total pangenome GCF content is poor at predicting the comparative number of metal resistance, antibiotic resistance, and virulence genes per *Pseudomonas* genome. Group Chlororaphis and Asplenii have the most virulence genes of the fluorescens lineage despite all the pathogenic species included in groups Syringae and Corrugata. Many

species of otherwise commensal or PGPR *Pseudomonas* contain virulence factors which may suggest the qualities of PGPR *Pseudomonas,* such as fungal/nematode antagonism and disease suppression, might be host dependent. There are host species susceptible to *P. brassicacearum* or other Corrugata group species (Yang *et al.*, 2020) and therefore defining whether a species is pathogenic or beneficial requires a more comprehensive view. It is relatively clear that AMR content does not correlate to group relative abundance within the 2010 rhizospheres.

Biosynthetic gene content is distinct by supergroup only when compared by cluster sequence and not predicted function, except when they're bacteriocins

Secondary metabolites produced by individual *Pseudomonas* genomes are highly diverse in sequence and structure, but these differences are hard to parse when only comparing by the representative class of metabolite. Each class of secondary metabolite contains many different molecules with different sequences and gene cluster organization. When comparing by BGC class, either by proportions of each or by ordinating by each genomes' Bray-Curtis dissimilarity, BGC annotations at this broad of a scale are not sufficient to discriminate between groups or supergroups. That being said, BGC content irrespective of group indicates *Pseudomonas* genomes contain numerous genes encoding nonribosomal peptide synthetases, arylpolyenes, NAGGNs, siderophores, polyketide antibiotics, and ribosomally synthesized and post-translationally modified peptides such as bacteriocins.

Discrimination between groups and supergroups improves greatly when ordinating the Bray-Curtis dissimilarity of each genome by gene cluster content. This method is based on the presence of shared GCFs, and demonstrates clearly that non-Fluorescens and SG1 group genomes are most similar to each other in terms of BGC content. Notably, unlike with the pangenome GCF plot, all of group Corrugata and many genomes within Jessenii and Mandelii also cluster with the NF and SG1 genomes. Overall, at supergroup level the differences in BGC content suggest that while there is some overlap between NF, SG1, and SG3, yet again group Fluorescens clusters independently from all other supergroups.

Examining individual AMR and BGC gene content highlights the disparity in antibiotic vs antibiotic resistance strategies

The vast majority of BGC gene clusters are structurally unique (i.e. singleton) to individual genomes and there is very little overlap even within groups. However, what few BGCs that differentiate between groups are bacteriocins, beta-lactone antibiotics, and 'generic' non-ribosomal peptides. There appears to be a generic non-ribosomal peptide synthetase cluster (GCF53) that is highly represented in pathogenic taxa independent of phylogenetic group, including but not limited to: *P. fuscovaginae A*, *P. viridiflava, P. cichorii & cichorii B, P. tremae, P. agarici,* and *P. floridensis.* Overall however, the general

lack of clustering by anything other than SG2 (Fluorescens) and SG3 (Koreensis) by only a handful of antimicrobial BGCs at the annotation level indicate a vast well of diversity in secondary metabolites that vary immensely between *Pseudomonas* species.

By contrast, the AMR profiles indicate a high degree of overlap in metal resistance and drug efflux gene content between groups. There are significantly fewer singleton genes per genome and k-means clustering indicates a stronger phylogenetic signal by group. The first PC of the AMR ordination was heavily influenced by the number of BacMet2 classifications, and the "V1 tail" genomes are defined by the k-means cluster 6A of figure 17. This cluster is missing what would otherwise be core or high number accessory AMR genes, yet there is no correlation between these genomes and the number of contigs, ambiguous bases, or CheckM completeness. All genes within that cluster are from BacMet2 and are a conglomeration of metal resistance, drug efflux, and the two-component response regulator irlRS which may be related to eukaryotic cell infiltration (Jones, DeShazer and Woods, 1997). The question of why these genes are absent could still be an assembly or completeness issue, especially considering how seemingly random gaps are. However, there is some clustering by group within those genomes so there may be the possibility that the operons associated with these genes are in the process of being lost, or perhaps the other direction where changes in functional homogeneity cause them not to align to the representative sequence in BacMet2.

Overall, the potential AMR phenotypes that appear to distinguish between groups are T6SS, T4P, and the Putida-specific multidrug efflux cluster; the core AMR genes are focused on alginate biosynthesis, metal resistance, and drug efflux (defense) while groups Corrugata, Chlororaphis, and Fluorescens possess the T6SS and may be more overtly antagonistic to other *Pseudomonas* taxa. The BGC profile may corroborate this, especially concerning the blooming *P. brassicacearum* and group Corrugata as a whole with their T6SS, T4P, bacteriocin GCF19. The presence of a complete T4P may facilitate exogenous DNA uptake (HGT) in biofilms and is present in syringae (Ellison *et al.*, 2018; Craig, Forest and Maier, 2019) and is highly represented in SG3 and in groups Syringae, Asplenii, and Anguilliseptica but absent in all other SG2 groups Fluorescens, Fragi, and Chlororaphis. However, group Fluorescens contains multiple shared bacteriocin GCFs which implies antagonism with other taxa. All groups of these *Pseudomonas* seem to rely on antimicrobial innovation to gain the edge required to overcome the well established drug efflux mechanisms present in the other taxa in close proximity.

The complete genome of the blooming *P. brassicacearum* ultimately eluded capture, despite both *in vitro* and bioinformatic attempts to do so

The ideal method for investigating how the phylogeny of rhizosphere *Pseudomonas* influences secondary metabolite gene cluster profiles and antimicrobial resistance gene presence/absence would have been to use MAGs from each rhizosphere instead of proxies for each species from GTDB.

In the case of the blooming *P. brassicacearum* (URIL14HWK12:I7) genome, this strain was not the GTDB representative strain nor a strain derived from any of the field locations examined. It was selected due to being cultured directly from the rhizosphere of IL14H maize and our possession of the only glycerol stock of this strain for reculturing and genome closure. The labile carbon substrate microcosms were a rudimentary effort to stimulate the growth of these blooming *Pseudomonas,* but as the data concedes, adding single carbon sources to the soil was not sufficient to capture *P. brassicacearum.*

This could have been due to any number of confounding factors, including but not limited to: the bulk soil of december 2015 being depleted in this taxa (I should have collected bulk soil from the Ithaca or Lansing fields if my intent was to culture the most abundant *Pseudomonas*), the substrates were not the preferred carbon source for that specific strain, and the obvious concern that *P. brassicacearum* has a symbiotic relationship within the rhizosphere and therefore would be at a disadvantage in the absence of a plant host. The other possibility is simply another bacterial taxa “jackpotting” on the substrate and outcompeting *Pseudomonas* during early incubation.

The high-quality MAGs combined with the publicly available strains of *P. brassicacearum* provided a pangenome that is most useful for determining what is unique to *P. brassicacearum* as a species, and what accessory functions are common. Singleton genes were largely absent from the MAGs and this is likely to greatly reduce the number of BGCs detected since the phylogenetic profiles of each group indicate the majority of secondary metabolite gene cluster families are largely unique to individual genomes. Despite the lack of BGC sensitivity in MAGs, what is shared is likely to be vertically inherited. Antimicrobial gene content is less affected by the lack of singletons, but the completeness of individual operons plus any other *in silico* method of trait prediction is going to be affected by the thresholds at which MAGs are considered “complete”. For pangenomes I chose 80%, but that is arbitrary and highly dependent on CheckM.

Accessory and singleton gene contamination may exist and go undetected by CheckM within publicly available genomes

Using CheckM as the “gold standard” for genome quality is widely accepted but not perfect; the initial pangenome created with URIL14HWK12:I7 indicated >700kb of singleton genes. When these singleton genes were mapped to contigs and these contigs were taxonomically resolved using BLAST, many aligned completely to *Rhodothermus marinus SG0.5JP17-172,* an aerobic, thermophilic, halophilic, gram-negative marine hotspring bacterium of the phylum Bacteroidota.

The likelihood that this was a contaminant during the initial culturing of this *Pseudomonas* at Cornell University, or the reculturing ten years later at the MPI on selective media containing triclosan, is astronomically low considering *R. marinus* was not a strain being studied at either location. Remarkably none of these genes were single-

copy, Rhodothermota-specific housekeeping markers and that's why the nearly megabase-worth of contamination went undetected by CheckM.

Ultimately it did not affect this species placement within the *Pseudomonas* phylogeny but its inclusion in the BGC, AMR, and pangenome analyses required completely resequencing the strain and re-analyzing these data. Since the draft genome completely lacked 16S rRNA genes, PacBio was selected to maximize quality vs effort for this attempt. Thankfully the long reads alone were sufficient to assemble a single contig with five 16S rRNA copies, plus zero reads aligning to *Rhodothermus*.

Speculating as to the origin of these contaminating reads, both genomes were sequenced at the U.S. Department of Energy Joint Genome Institute (JGI) with their respective projects being released four months apart (*R. marinius* 2013-02-12 vs *P. sp. URIL14HWK12:I7* 2013-06-16). The lack of any *R. marinus* housekeeping genes leads me to believe the contamination may have occurred during sequencing through slight cross-contamination of gDNA, because the likelihood of an accidental overlap of indexing primers would cause the whole isolate genome to be demultiplexed with the *Pseudomonas.*

Paralogous gene content between species and group pangenomes suggest intense resource competition between *Pseudomonas,* along with pressure to maintain redundancy in existing metabolism while leaving potential to innovate

Once that *P. brassicacearum* isolate was demonstrably free of non-*Pseudomonas* sequences, the pangenomes of the top three most abundant *Pseudomonas* could be confidently analyzed. The most abundant *Pseudomonas* were generally closely related; *P. brassicacearum* is a member of group Corruagata while *P. frederiksbergensis E* and *P. silesiensis* are group Mandelii, and these two groups are adjacent within SG3. The additional pangenome of species *P. sp000282315* was prepared to utilize the five high-quality MAGs produced. This taxa was also the third-most abundant *Pseudomonas* in Ithaca after *P. brassicacearum* and *P. sp900005815* (group Fluorescens), with the latter unfortunately not obtaining any MAGs or isolates.

The number of genomes per species pangenome varies depending on how many strains are publicly available, with most containing 6-9 genomes and *P. brassicacearum* possessing 38 (including the MAGs mentioned above). It is well understood that *Pseudomonas* has expansive accessory and singleton pangenomes, therefore pangenomes with fewer strains are guaranteed to represent only a fraction of the total gene diversity present within that species. When comparing the gene content between these species, the distinction was made between total genes and gene clusters vs "unique" genes and gene clusters, with only the representative of each being counted for the latter. The difference between the total gene/GC and number of representative/unique examples were interpreted as duplicate, or paralogous, genes.

When comparing paralogous gene content between each pangenome compartment the core and accessory pangenomes contained the most, with *P. brassicacearum* and *P. sp000282315* having more duplicate genes within their core, and *P. silesiensis* and *P. frederiksbergensis E* having slightly more duplicate genes within their accessory pangenome. The singleton pangenome was the least, with <3% containing duplicate genes. This amount of redundancy in gene content was notable for two main reasons: free-living bacteria are generally under nutritional constraints that select against maintaining neutral genes that inflate genome size; yet, gene duplications are associated with genome diversification and speciation. The average coding density for the *Pseudomonas* examined within this work is 88%, and this suggests these duplicate genes have not become pseudogenes. What is impossible to know without examining all gene clusters individually is whether these duplicate genes were inherited vertically or whether it was the result of HGT from a closely-related species.

Speculating on the purpose of gene duplication within the core pangenome of *Pseudomonas,* it has been suggested both spontaneous and HGT-mediated gene duplications were capable of increasing phenotypic novelty while minimizing pleiotropic cost so long as the genes themselves were highly redundant within the genome (Toll-Riera *et al.*, 2016). Duplications of the nucleoside hydrolase gene *nuh* and the adenine deaminase PA0148 in *P. aeruginosa* have been observed to increase fitness when growing on adenosine (Toussaint *et al.*, 2017). The underlying premise being that *Pseudomonas* must strike a balance between maintaining that well of genetic potential to increase metabolic flexibility in order to survive in both rhizospheres and in soils, yet not pass the threshold where the energetic costs or the accumulation of pleiotropic effects would ultimately decrease fitness when conditions eventually change.

tRNA duplication may reflect changes in codon usage in response to HGT

The singleton and accessory pangenome of *Pseudomonas* are the compartments that possess numerous genes acquired by HGT, and it has been observed in other Gammaproteobacteria (pathogenic *E. coli/Shigella*) that multi-copy tRNA genes can be introduced during HGT of selfish genetic elements from other taxa to compensate for the change in codon usage (McDonald *et al.*, 2015). *Pseudomonas* exhibit their own “core” and “accessory” multicopy tRNA genes, and within the rhizospheres of this study, the average number of tRNA genes per species of the fluorescens lineage taxa was 64.8 while the non-fluorescens lineage was 60.4; however, the total number of tRNAs was negatively correlated with the number of contigs for each assembly therefore the predominance of draft genomes is likely underrepresenting the true count. However, when only complete genomes are examined then the average for fluorescens lineage species jumps to 70 tRNAs per genome while non-Fluorescens genome tRNA averages are only 66 per genome.

Experiments in *P. fluorescens* that sequentially deleted accessory tRNAs observed little to no reduction in fitness within a selection of strains, indicating redundancy in their products. However, in strains that did suffer a fitness defect that were used as a founder population in a directed evolution experiment, sequencing revealed large (>500kb) genome duplications that contained additional copies of the tRNA genes that were originally knocked out (Khomarbaghi, 2019). The compensatory mechanism of these duplications might be worthwhile to explore to determine what cellular processes were impaired; specifically whether it was a defect in global protein synthesis or the dosage of a specific product, and if that affected competitive fitness. Overall, when comparing paralog content between each phylogenetic group (excluding those with less than five genomes), as with the species the singleton pangenomes contained the fewest duplicate genes while the core always contained the most. Notably, groups Syringae, Fluorescens, Corrugata, and Mandeli exhibited the largest core paralog content.

Mutualism and parasitism is difficult to predict based on phylogeny within the fluorescens lineage *Pseudomonas*

There is evidence in the literature that larger genomes (>6Mb) are indicative of a lifestyle that prioritizes flexibility in substrate utilization yet doesn't penalize slower growth to accommodate the larger repertoire of metabolism genes (Konstantinidis and Tiedje, 2004). The average genome size of the fluorescens lineage *Pseudomonas* included in this work is 6.2Mb with SG3 being 6.4Mb, while the non-fluorescens lineage *Pseudomonas* contains an average genome size of 5.2Mb. The hypothesis that genome size is predictive of ecological strategy extends to both eukaryotes and bacteria, with reduced genomes being linked to host dependence in vertically-transmitted symbionts (McCutcheon and Moran, 2011; Fisher *et al.*, 2017).

Numerous well-studied examples exist of obligately parasitic *Pseudomonas,* although instances of *Pseudomonas* taxa adopting obligate mutualisms to either plants or animals exist but are often studied in depth only when agricultural benefit can be obtained. There are certainly taxa that have formed mutualisms with plants as evidenced by growth-promoting phenotypes, but these taxa are capable of surviving in soils independent of the host. Studies of the gut microbes of insects have found notable metabolic symbioses. Two examples include the sap-feeding hemlock wooly adelgid *Adelges tsugae* were found to harbor a gut microbe identified as *Candiatus Pseudomonas adelgestsugas* (von Dohlen *et al.*, 2013, 2017), and the coffee berry borer *Hypothenemus hampei* harbors strains of *P. fulva* that are capable of using caffeine as sole carbon source thus detoxifying the diet of these beetles (Ceja-Navarro *et al.*, 2015). Caffeine demethylase is also found in *P. putida* (of which *P. fulva* is a member of that phylogenetic group) but only *P. fulva* appeared capable of persisting in the gut of *H. hampei.*

The discovery of a vertically-inherited obligate mutualism within *Pseudomonas* would be a fascinating avenue to explore; in the context of this work, and within the

rhizosphere, that *Pseudomonas* is at heart a promiscuous R-strategist that expends a substantial amount of energy vying with other soil copiotrophs for transiently abundant root exudates. However, in the case of temperate climes and annual hosts, with feast comes famine and otherwise highly rhizosphere-adapted *Pseudomonas* must cope with long periods of low nutrition while being poised to utilize substrates as they become available. The data accumulated in service of this monograph appears to corroborate that statement, in addition to suggesting that microbe-microbe competition between *Pseudomonas* are potentially driving speciation within the more derived fluorescens lineage taxa.

Rhizosphere fungi are taxonomically diverse and vary by field, whereas nematoda were dominated by one bacteriovorus species

The last topic necessary to discuss is an often overlooked one; *Pseudomonas* and their relationship with other soil and rhizosphere eukaryotes aside from the plant. Studies of the rhizosphere frame the plant as the “host”; however just as copiotrophic prokaryotes thrive in this habitat other organisms such as metazoa, fungi, protozoa, and oomycetes are drawn by root exudates and this forms a complex food web (Mao *et al.*, 2014). As discussed in the rhizosphere eukaryote results subsection, the extraction method for the metagenomes is likely to have affected the total markers observed for each taxa. However, other reports of rhizosphere eukaryotes in selected landraces of maize observed the relative abundance of fungal reads is lower in rhizospheres compared to the bulk soil, and the fungal taxa enriched in the rhizospheres included the Ascomycete families Mortierellaceae and Trichocomaceae (Matus-Acuña, Caballero-Flores and Martínez-Romero, 2021). This study corroborated these findings, and although the authors could not identify the nematode taxa detected at species level their maize was enriched in bacteria-feeding genera. This would include the predominant nematode found in the 2010 rhizospheres, *A. nanus*.

Pseudomonas are often preyed upon by bacterivorous nematodes and thus utilize secondary metabolites such as hydrogen cyanide and 2,4-DAPG to repel or kill nematodes (Neidig *et al.*, 2011; Burlinson *et al.*, 2013). The strain used to represent *P. brassicacearum,* URIL14HWK12:I7, contains the 2,4-DAPG biosynthetic operon and was the highest in abundance in Ithaca and Lansing; when comparing the total markers for *A. nanus* between fields, no markers were detected within the maize genotype B73 in Ithaca. However, the other genotypes provided similar abundances to Aurora. Without confirming whether or not URIL14HWK12:I7 is directly antagonistic to *A. nanus* it can’t be proven to be the reason that nematode is absent in that condition, but at least it being present on the other maize genotypes indicates the difference is not the result of the specific eukaryotic diversity of that field.

Constraints, limitations, and where to go from here

One major limitation of this analysis is that HGT was never formally measured. It is known that *Pseudomonas*, specifically pathogenic environmental *P. aeruginosa*, use HGT to evolve in novel hosts (Qiu, Kulasekara and Lory, 2009), although there is evidence that HGT overall is constrained by fitness costs between independently acquired mobile genetic elements (San Millan *et al.*, 2015). The lack of phylogenetic signal in biosynthetic gene content suggests HGT may play a role, and the fact that the rhizosphere creates an environment of high bacterial density means more opportunity for exogenous DNA uptake or plasmid conjugation.

Furthermore, the initial motivation for this thesis was to characterize the blooming *Pseudomonas* to determine if their gene content gave any indication suggestive of a PGPR lifestyle. Beyond the obvious fact that strains of *P. brassicacearum* are considered beneficial due to possessing 2,4-DAPG and generally being an excellent rhizosphere colonizer, there is nothing particularly remarkable about the strain of *P. brassicacearum* used for this analysis. It is never an outlier when comparing pangenome content, BGCs, AMR genes, or phenotype; if anything it is perfectly average when compared with other Corrugata group species. That being said, of course not every gene of this species is perfectly annotated and functionally characterized. If *P. aeruginosa* is any indication there is probably a substantial reservoir of function that is waiting to be discovered.

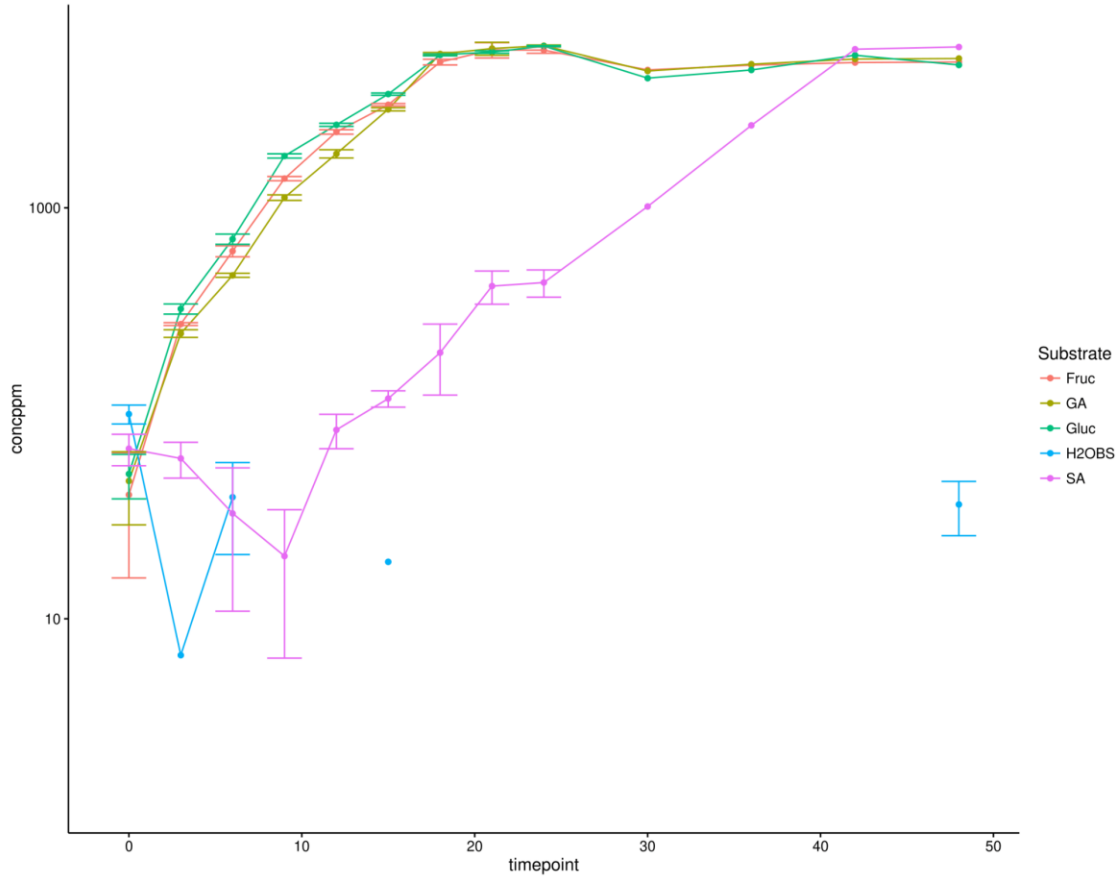
Ultimately the bloom was likely an artifact of a climatic event that occurred the summer of 2010, and the lack of environmental metadata was a seriously unfortunate oversight that should have been avoided. If the study were to be repeated as was in 2015, taking weekly measurements of precipitation, soil moisture, air and soil temperature, along with pictures of collected plants and maps of their locations would be essential. Another consideration to perhaps improve consistency and resolution of the sequence data would be to only sample each rhizosphere within a specific range of soil moisture to avoid introducing an effect where soil adherence to the roots varies by week. Also, the DNA extraction method used for this study was not ideal in that no distinction could be made between the rhizosphere compartments.

What can be gleaned from this phylogenetic exploration into rhizosphere *Pseudomonas* is less about the bloom and more about the immense diversity within this habitat and how the functional potential of these species vary in relation to each other. One useful aspect of this analysis is the placement of taxonomically unresolved *Pseudomonas* into the broader phylogeny, and to anyone wishing to study a specific phenotype may now have many additional candidates among these strains.

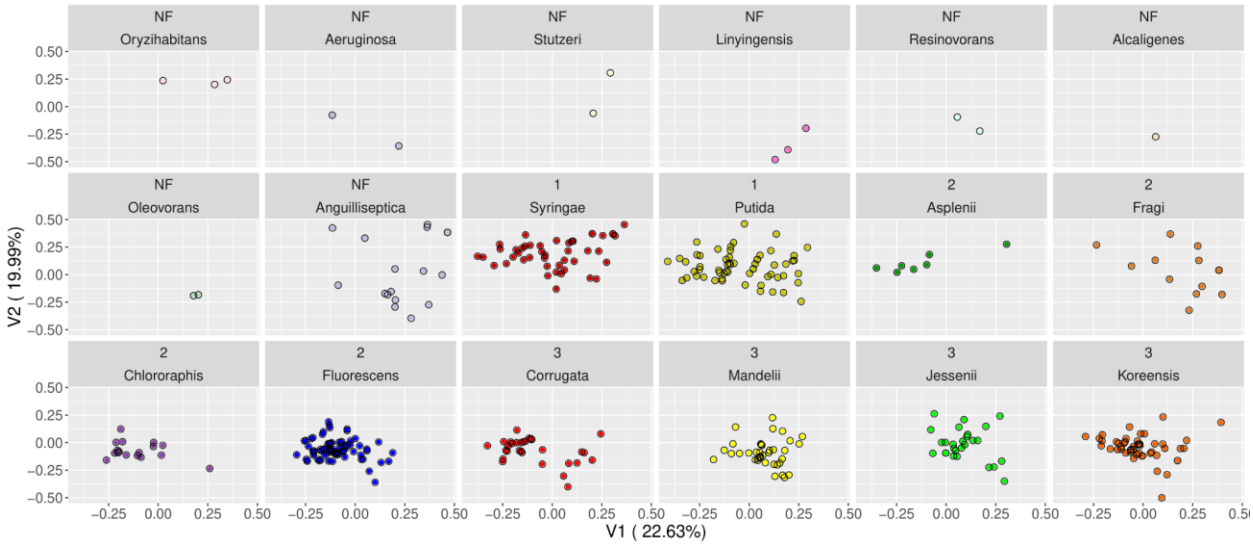
Glossary

16S rRNA - 16S small subunit ribosomal RNA
Accessory genome - genes present in more than one genome but less than all genomes
Bulk soil - soil not under the influence of the ectorhizosphere
Contig - a consensus sequence of DNA representing a portion of a genome
Core genome - genes present in all genomes
Pangenome - all genes within all genomes
Singleton genome - genes present in only one genome

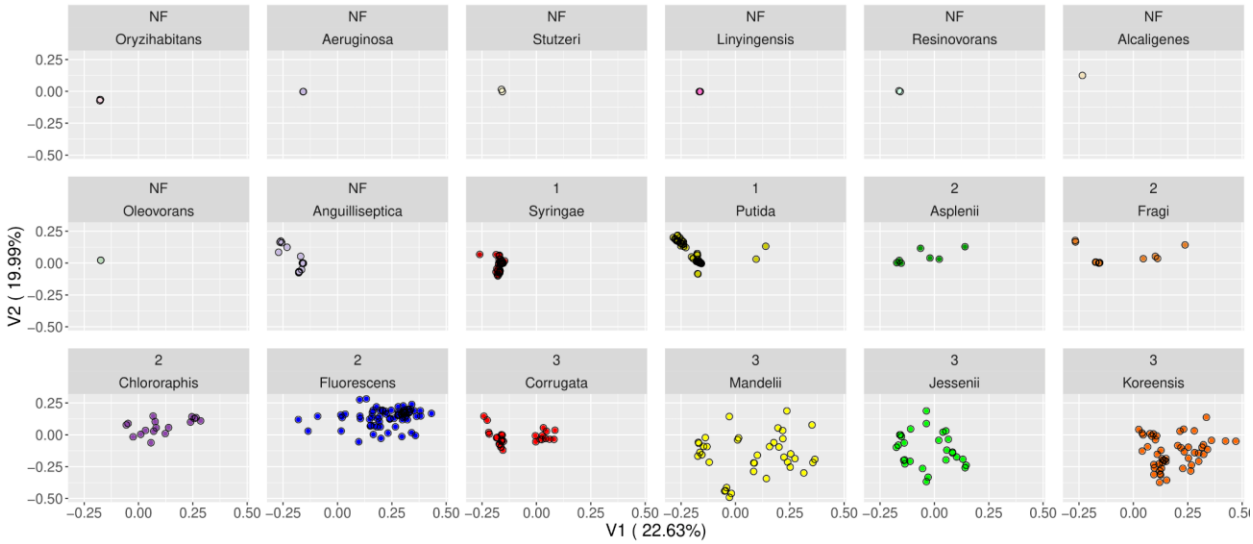
Supplemental figures



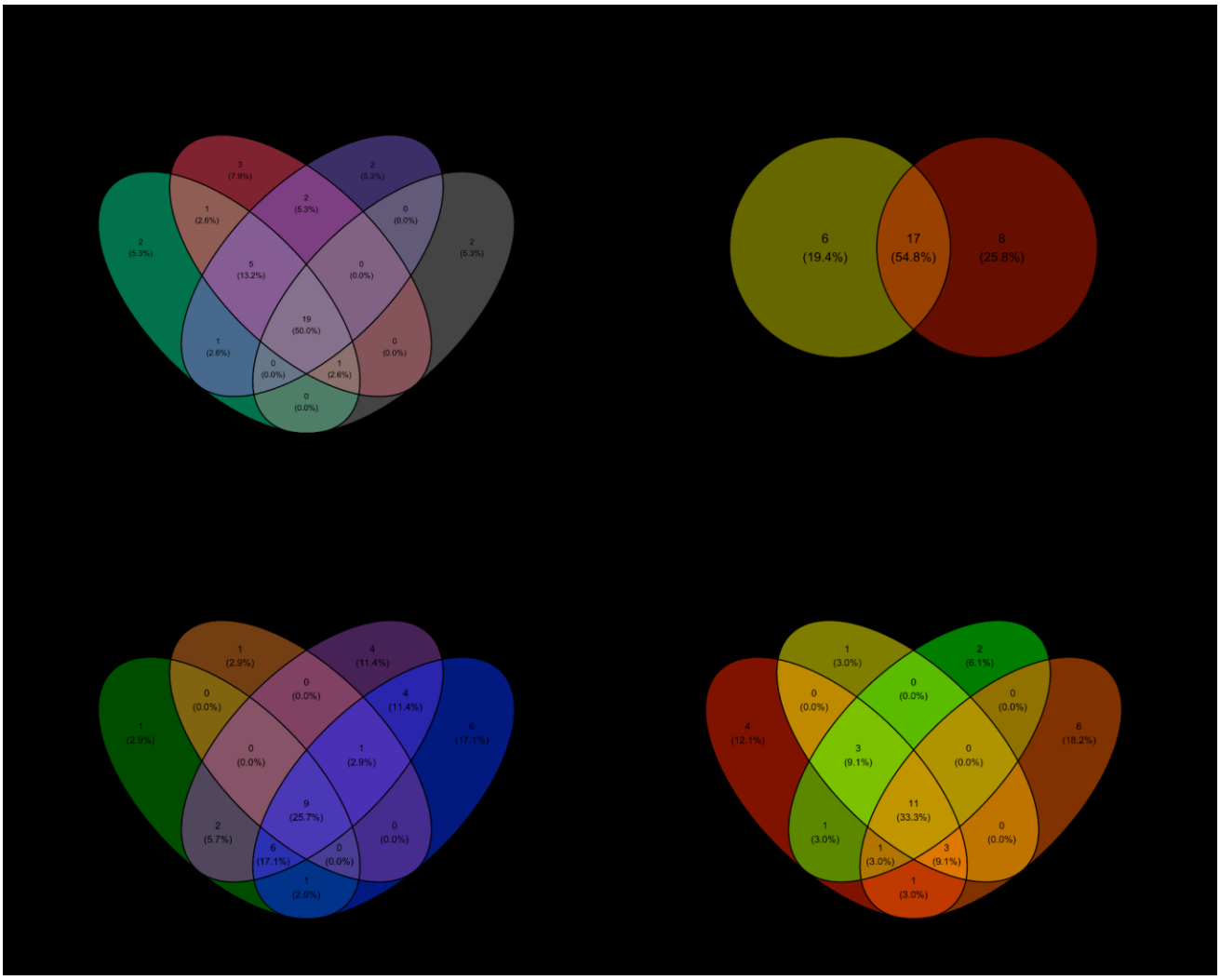
Supplemental figure 1 - Substrate-treated microcosm headspace CO2 in ppm over time. Sampling was cumulative (headspace was not flushed between collections) at 3 hour intervals from 0-24h, then every 6 hours from 24-48h. The water+basal salt control was sampled at 0, 3, 6, 12, 48h.



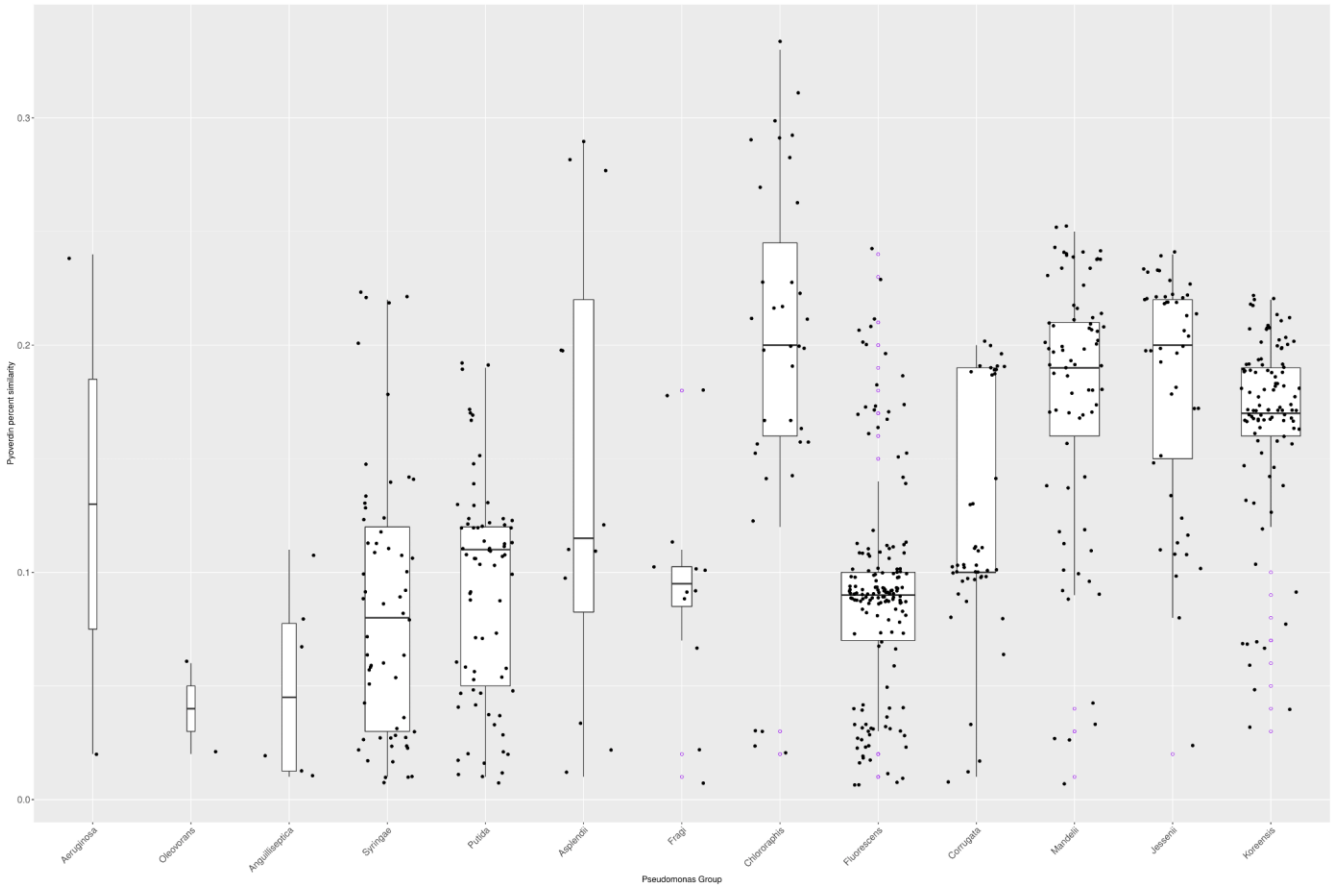
Supplemental figure 2 - PCoA of the Bray-Curtis dissimilarity clusters of BGC annotations separated by *Pseudomonas* group



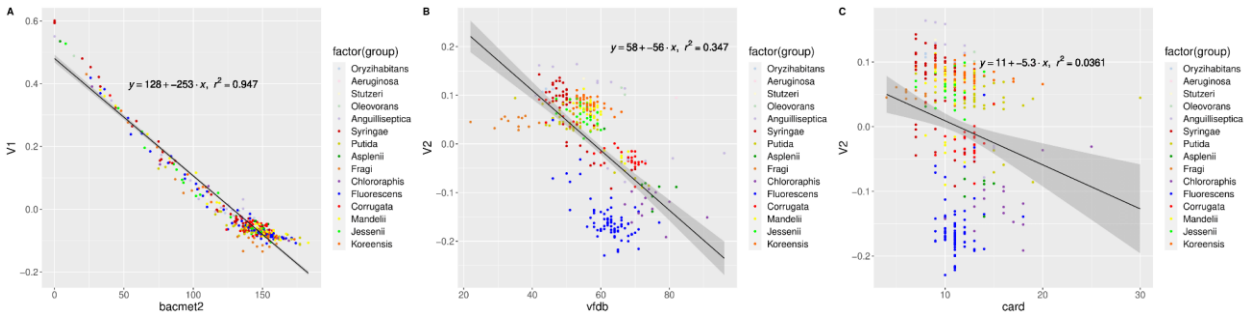
Supplemental figure 3 - PCoA of the Bray-Curtis dissimilarity clusters of biosynthetic GCFs separated by *Pseudomonas* group



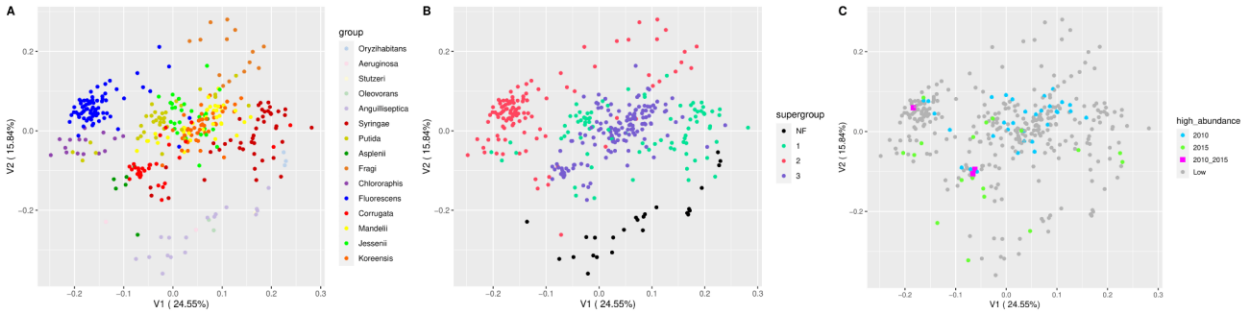
Supplemental figure 4 - The number of BGC representative classes shared between supergroups and fluorescens lineage groups.
A) Number of shared BGC classes between all supergroups. B) SG1 only. C) SG2 only. D) SG3 only.



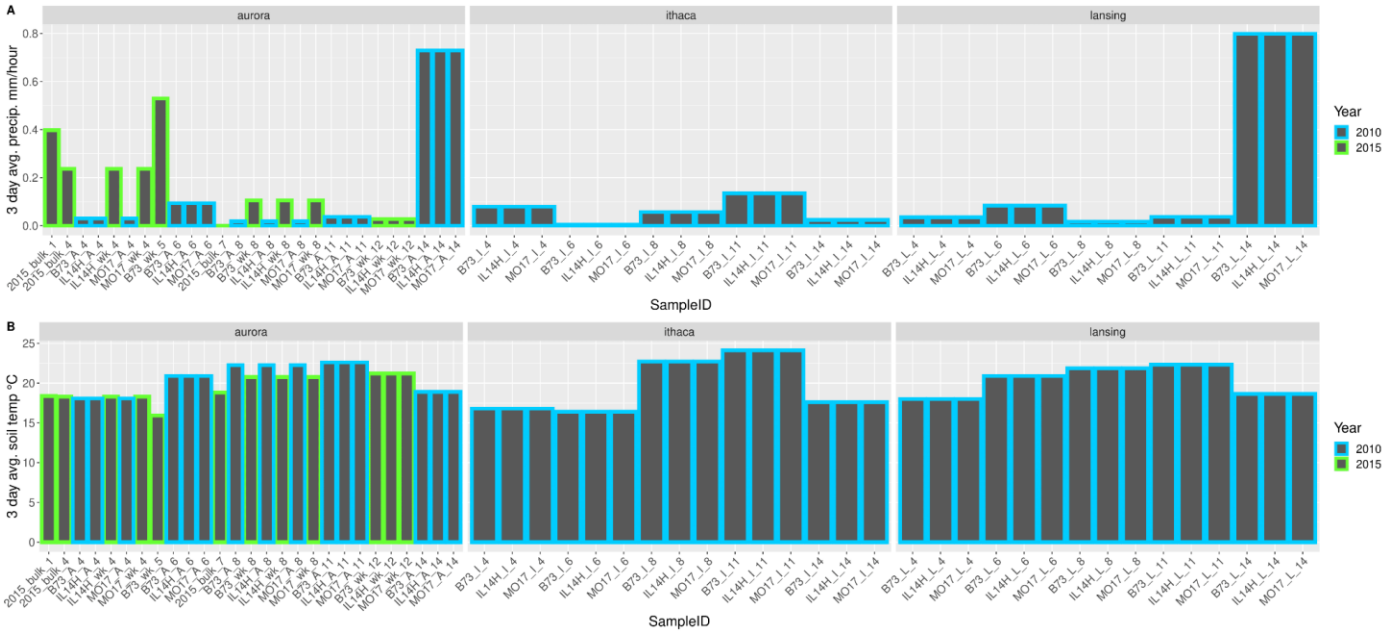
Supplemental figure 5 - Pyoverdin gene sequence similarity by *Pseudomonas* group. Each genome is represented by a point, with the pyoverdin percent similarity representing what percentage of genes the pyoverdin cluster contained significant hits to BLAST.



Supplemental figure 6 - Influence of each AMR database on the Jaccard distance principal coordinates by *Pseudomonas* group with *P. aeruginosa* removed
A) the number of BacMet2 hits by V1. B) virulence factor database hits by V2. C) antibiotic resistance database by V2.



Supplemental figure 7 - Jaccard distance PCoA for all experimentally validated genes within VFDB colored by group, supergroup, and abundance per year.

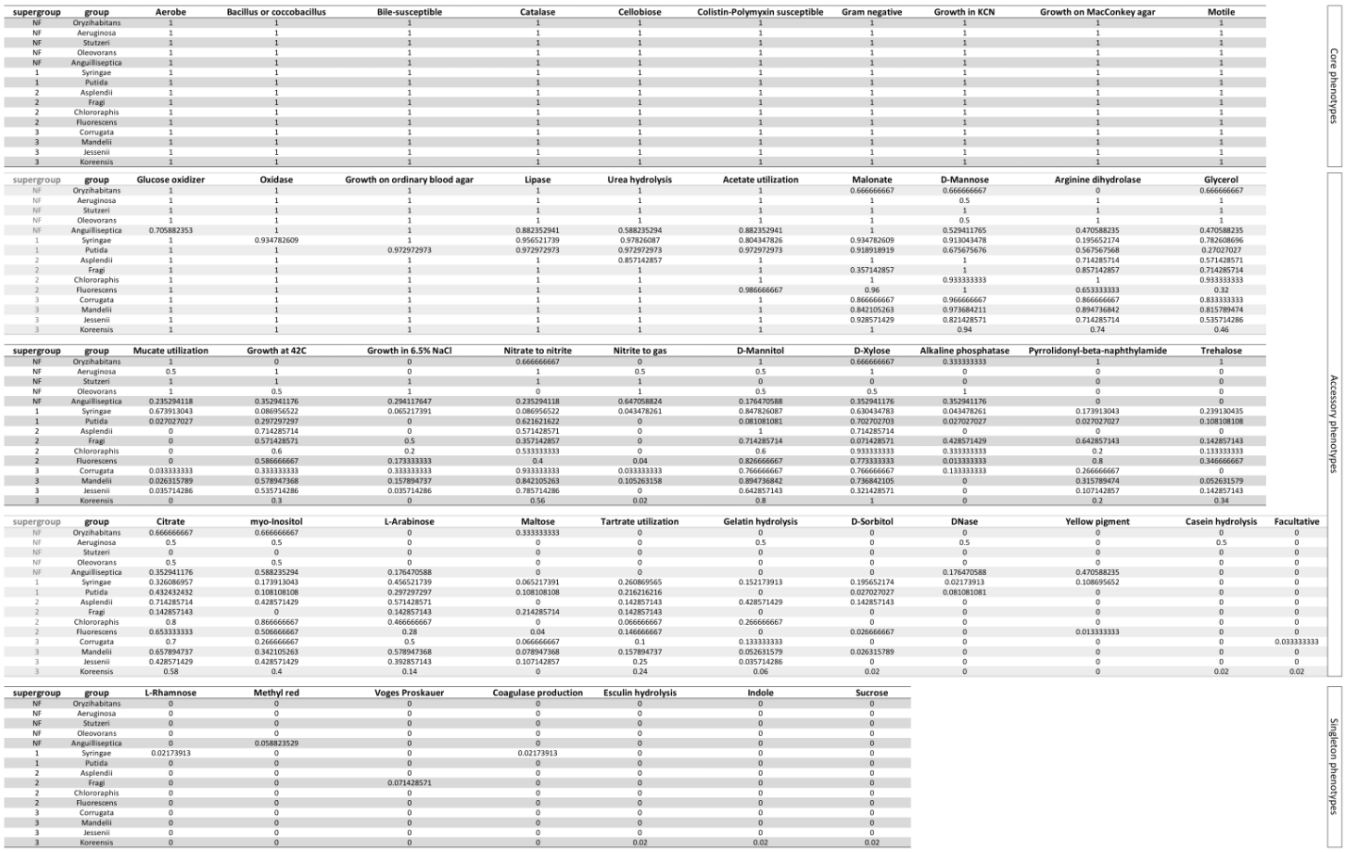


Supplemental figure 8 - Average precipitation and soil temperature for three days prior to each rhizosphere sampling, by field with the combined years in Aurora.

Supplemental tables

SampleID	maize_genotype	field	sampling_week	rhizosphere	year
2015_bulk_1	bulk	Aurora	1	NA	2015
2015_bulk_4	bulk	Aurora	4	NA	2015
2015_bulk_7	bulk	Aurora	7	NA	2015
B73_wk_12	B73	Aurora	12	rhizo	2015
B73_wk_5	B73	Aurora	5	rhizo	2015
B73_wk_8	B73	Aurora	8	rhizo	2015
IL14H_wk_12	IL14H	Aurora	12	rhizo	2015
IL14H_wk_4	IL14H	Aurora	4	rhizo	2015
IL14H_wk_8	IL14H	Aurora	8	rhizo	2015
MO17_wk_12	MO17	Aurora	12	rhizo	2015
MO17_wk_4	MO17	Aurora	4	rhizo	2015
MO17_wk_8	MO17	Aurora	8	rhizo	2015
B73_A_11	B73	Aurora	11	rhizo	2010
B73_A_14	B73	Aurora	14	rhizo	2010
B73_A_4	B73	Aurora	4	rhizo	2010
B73_A_6	B73	Aurora	6	rhizo	2010
B73_A_8	B73	Aurora	8	rhizo	2010
B73_I_11	B73	Ithaca	11	rhizo	2010
B73_I_14	B73	Ithaca	14	rhizo	2010
B73_I_4	B73	Ithaca	4	rhizo	2010
B73_I_6	B73	Ithaca	6	rhizo	2010
B73_I_8	B73	Ithaca	8	rhizo	2010
B73_L_11	B73	Lansing	11	rhizo	2010
B73_L_14	B73	Lansing	14	rhizo	2010
B73_L_4	B73	Lansing	4	rhizo	2010
B73_L_6	B73	Lansing	6	rhizo	2010
B73_L_8	B73	Lansing	8	rhizo	2010
IL14H_A_11	IL14H	Aurora	11	rhizo	2010
IL14H_A_14	IL14H	Aurora	14	rhizo	2010
IL14H_A_4	IL14H	Aurora	4	rhizo	2010
IL14H_A_6	IL14H	Aurora	6	rhizo	2010
IL14H_A_8	IL14H	Aurora	8	rhizo	2010
IL14H_I_11	IL14H	Ithaca	11	rhizo	2010
IL14H_I_14	IL14H	Ithaca	14	rhizo	2010
IL14H_I_4	IL14H	Ithaca	4	rhizo	2010
IL14H_I_6	IL14H	Ithaca	6	rhizo	2010
IL14H_I_8	IL14H	Ithaca	8	rhizo	2010
IL14H_L_11	IL14H	Lansing	11	rhizo	2010
IL14H_L_14	IL14H	Lansing	14	rhizo	2010
IL14H_L_4	IL14H	Lansing	4	rhizo	2010
IL14H_L_6	IL14H	Lansing	6	rhizo	2010
IL14H_L_8	IL14H	Lansing	8	rhizo	2010
MO17_A_11	MO17	Aurora	11	rhizo	2010
MO17_A_14	MO17	Aurora	14	rhizo	2010
MO17_A_4	MO17	Aurora	4	rhizo	2010
MO17_A_6	MO17	Aurora	6	rhizo	2010
MO17_A_8	MO17	Aurora	8	rhizo	2010
MO17_I_11	MO17	Ithaca	11	rhizo	2010
MO17_I_14	MO17	Ithaca	14	rhizo	2010
MO17_I_4	MO17	Ithaca	4	rhizo	2010
MO17_I_6	MO17	Ithaca	6	rhizo	2010
MO17_I_8	MO17	Ithaca	8	rhizo	2010
MO17_L_11	MO17	Lansing	11	rhizo	2010
MO17_L_14	MO17	Lansing	14	rhizo	2010
MO17_L_4	MO17	Lansing	4	rhizo	2010
MO17_L_6	MO17	Lansing	6	rhizo	2010
MO17_L_8	MO17	Lansing	8	rhizo	2010

Supplemental table 1 - Metagenome identities by maize genotype, location, collection week, rhizosphere influence, and year



Core phenotypes

supergroup	group	Aerobe	Bacillus or coccobacillus	Bile-susceptible	Catalase	Cellobiose	Colistin-Polymyxin susceptible	Gram negative	Growth in KCN	Growth on MacConkey agar	Motile
NF	Oryzihabitans	1	1	1	1	1	1	1	1	1	1
NF	Aeruginosa	1	1	1	1	1	1	1	1	1	1
NF	Stutzeri	1	1	1	1	1	1	1	1	1	1
NF	Oleovorans	1	1	1	1	1	1	1	1	1	1
NF	Anguilliseptica	1	1	1	1	1	1	1	1	1	1
1	Syringae	1	1	1	1	1	1	1	1	1	1
1	Putida	1	1	1	1	1	1	1	1	1	1
2	Asplendii	1	1	1	1	1	1	1	1	1	1
2	Fragi	1	1	1	1	1	1	1	1	1	1
2	Chlororaphis	1	1	1	1	1	1	1	1	1	1
2	Fluorescens	1	1	1	1	1	1	1	1	1	1
3	Corrugata	1	1	1	1	1	1	1	1	1	1
3	Mandelii	1	1	1	1	1	1	1	1	1	1
3	Jessenii	1	1	1	1	1	1	1	1	1	1
3	Koreensis	1	1	1	1	1	1	1	1	1	1

Accessory phenotypes

supergroup	group	Glucose oxidizer	Oxidase	Growth on ordinary blood agar	Lipase	Urea hydrolysis	Acetate utilization	Malonate	D-Mannose	Arginine dihydrolase	Glycerol
NF	Oryzihabitans	1	1	1	1	1	1	0.666666667	0.666666667	0	0.666666667
NF	Aeruginosa	1	1	1	1	1	1	1	0.5	1	1
NF	Stutzeri	1	1	1	1	1	1	1	1	1	1
NF	Oleovorans	1	1	1	1	1	1	1	0.5	1	1
NF	Anguilliseptica	0.705882353	1	1	0.882352941	0.588235294	0.882352941	1	0.529411765	0.470588235	0.470588235
1	Syringae	1	0.934782609	1	0.956521739	0.97826087	0.804347826	0.934782609	0.913043478	0.195652174	0.782608696
1	Putida	1	1	0.972972973	0.972972973	0.972972973	0.972972973	0.918918919	0.675675676	0.567567568	0.27027027
2	Asplendii	1	1	1	1	0.857142857	1	1	1	0.714285714	0.571428571
2	Fragi	1	1	1	1	1	1	0.357142857	1	0.857142857	0.714285714
2	Chlororaphis	1	1	1	1	1	1	1	0.933333333	1	0.933333333
2	Fluorescens	1	1	1	1	1	0.986666667	0.96	1	0.653333333	0.32
3	Corrugata	1	1	1	1	1	1	0.866666667	0.966666667	0.866666667	0.833333333
3	Mandelii	1	1	1	1	1	1	0.842105263	0.973684211	0.894736842	0.815789474
3	Jessenii	1	1	1	1	1	1	0.928571429	0.821428571	0.714285714	0.535714286
3	Koreensis	1	1	1	1	1	1	1	0.94	0.74	0.46

supergroup	group	Mucate utilization	Growth at 42C	Growth in 6.5% NaCl	Nitrate to nitrite	Nitrite to gas	D-Mannitol	D-Xylose	Alkaline phosphatase	Pyrrolidonyl-beta-naphthylamide	Trehalose
NF	Oryzihabitans	1	0	0	0.666666667	0	1	0.666666667	0.333333333	1	1
NF	Aeruginosa	0.5	1	0	1	0.5	0.5	1	0	0	0
NF	Stutzeri	1	1	1	1	1	0	0	0	0	0
NF	Oleovorans	1	0.5	1	0	1	0.5	0.5	1	0	0
NF	Anguilliseptica	0.235294118	0.352941176	0.294117647	0.235294118	0.647058824	0.176470588	0.352941176	0.352941176	0	0
1	Syringae	0.673913043	0.086956522	0.065217391	0.086956522	0.043478261	0.847826087	0.630434783	0.043478261	0.173913043	0.239130435
1	Putida	0.027027027	0.297297297	0	0.621621622	0	0.081081081	0.702702703	0.027027027	0.027027027	0.108108108
2	Asplendii	0	0.714285714	0	0.571428571	0	1	0.714285714	0	0	0
2	Fragi	0	0.571428571	0.5	0.357142857	0	0.714285714	0.071428571	0.428571429	0.642857143	0.142857143
2	Chlororaphis	0	0.6	0.2	0.533333333	0	0.6	0.933333333	0.333333333	0.2	0.133333333
2	Fluorescens	0	0.586666667	0.173333333	0.4	0.04	0.826666667	0.773333333	0.013333333	0.8	0.346666667
3	Corrugata	0.033333333	0.333333333	0.333333333	0.933333333	0.033333333	0.766666667	0.766666667	0.133333333	0.266666667	0
3	Mandelii	0.026315789	0.578947368	0.157894737	0.842105263	0.105263158	0.894736842	0.736842105	0	0.315789474	0.052631579
3	Jessenii	0.035714286	0.535714286	0.035714286	0.785714286	0	0.642857143	0.321428571	0	0.107142857	0.142857143
3	Koreensis	0	0.3	0	0.56	0.02	0.8	1	0	0.2	0.34

Singleton phenotypes

supergroup	group	Citrate	myo-Inositol	L-Arabinose	Maltose	Tartrate utilization	Gelatin hydrolysis	D-Sorbitol	DNase	Yellow pigment	Casein hydrolysis	Facultative
NF	Oryzihabitans	0.666666667	0.666666667	0	0.333333333	0	0	0	0	0	0	0
NF	Aeruginosa	0.5	0.5	0	0	0	0.5	0	0.5	0	0.5	0
NF	Stutzeri	0	0	0	0	0	0	0	0	0	0	0
NF	Oleovorans	0.5	0.5	0	0	0	0	0	0	0	0	0
NF	Anguilliseptica	0.352941176	0.588235294	0.176470588	0	0	0	0	0.176470588	0.470588235	0	0
1	Syringae	0.326086957	0.173913043	0.456521739	0.065217391	0.260869565	0.152173913	0.195652174	0.021739130	0.108695652	0	0
1	Putida	0.432432432	0.108108108	0.297297297	0.108108108	0.216216216	0	0.027027027	0.081081081	0	0	0
2	Asplendii	0.714285714	0.428571429	0.571428571	0	0.142857143	0.428571429	0.142857143	0	0	0	0
2	Fragi	0.142857143	0	0.142857143	0.214285714	0.142857143	0	0	0	0	0	0
2	Chlororaphis	0.8	0.866666667	0.466666667	0	0.066666667	0.266666667	0	0	0	0	0
2	Fluorescens	0.653333333	0.506666667	0.28	0.04	0.146666667	0	0.026666667	0	0.013333333	0	0
3	Corrugata	0.7	0.266666667	0.5	0.066666667	0.1	0.133333333	0	0	0	0	0.033333333
3	Mandelii	0.657894737	0.342105263	0.578947368	0.078947368	0.157894737	0.052631579	0.026315789	0	0	0	0
3	Jessenii	0.428571429	0.428571429	0.392857143	0.107142857	0.25	0.035714286	0	0	0	0	0
3	Koreensis	0.58	0.4	0.14	0	0.24	0.06	0.02	0	0	0.02	0.02

supergroup	group	L-Rhamnose	Methyl red	Voges Proskauer	Coagulase production	Esculin hydrolysis	Indole	Sucrose
NF	Oryzihabitans	0	0	0	0	0	0	0
NF	Aeruginosa	0	0	0	0	0	0	0
NF	Stutzeri	0	0	0	0	0	0	0
NF	Oleovorans	0	0	0	0	0	0	0
NF	Anguilliseptica	0	0.058823529	0	0	0	0	0
1	Syringae	0.02173913	0	0	0.02173913	0	0	0
1	Putida	0	0	0	0	0	0	0
2	Asplendii	0	0	0	0	0	0	0
2	Fragi	0	0	0.071428571	0	0	0	0
2	Chlororaphis	0	0	0	0	0	0	0
2	Fluorescens	0	0	0	0	0	0	0
3	Corrugata	0	0	0	0	0	0	0
3	Mandelii	0	0	0	0	0	0	0
3	Jessenii	0	0	0	0	0	0	0
3	Koreensis	0	0	0	0	0.02	0.02	0.02

Supplemental table 2 - Phenotype composition per phylogenetic group. Each column represents the fraction of genomes per group with that phenotype detected. Core phenotypes are present in all genomes in all groups. Accessory phenotypes are present in more than two but less than all groups. Singleton phenotypes are present in only one group.

all genomes	AMR oryzihabitans	AMR anguilliseptica	AMR syringae	AMR putida	AMR asplenii	AMR fragi	AMR chlororaphis	AMR fluor	AMR corrugata	AMR mandelii	AMR jessenii	AMR koreensis
adeL_BAC0508	acn_BAC0003	actR_BAC0565	algB	acn_BAC0003	algC	adeL_BAC0508	adeL_BAC0508	adeH_BAC0016	acn_BAC0003	adeH_BAC0016	adeL_BAC0508	acn_BAC0003
algB	adeL_BAC0508	actS_BAC0566	algA	actP/yjcG_BAC0564	algR	algB	algB	adeL_BAC0508	actP/yjcG_BAC0564	adeL_BAC0508	algB	actP/yjcG_BAC0564
algC	algA	adeL_BAC0508	algB	actR_BAC0565	algU	algU	algA	algB	actR_BAC0565	algB	algA	adeH_BAC0016
algU	algB	algB	algC	actS_BAC0566	algW	arnA	algB	algA	adeH_BAC0016	algA	algB	adeL_BAC0508
algW	algC	algC	algD	adeH_BAC0016	arnA	ARNA	algC	algB	adeL_BAC0508	algB	algC	algB
baeR_BAC0039	algR	algR	algI	adeL_BAC0508	ARNA	baeR_BAC0039	algD	algC	algB	algC	algD	algA
cheW_2	algU	algU	algR	algB	cheW_2	bfrA_BAC0048	algF	algD	algC	algD	algG	algB
CPXAR	algZ	algW	algU	algC	cheY	cadR_BAC0058	algG	algG	algR	algI	algI	algC
fabL/ygaA_BAC0158	arnA	bcr_BAC0041	algW	algU	corD_BAC0644	cheW_2	algI	algI	algU	algR	algR	algD
fbpC_BAC0162	ARNA	cheW_2	cheW_2	algW	EMHC	corD_BAC0644	algR	algU	algW	algU	algU	algI
fecE_BAC0164	arsB_BAC0575	corR_BAC0089	cheY	arsC_BAC0583	evgA_BAC0154	corR_BAC0089	algU	algW	arsT_BAC0714	algW	algW	algR
fleN	baeR_BAC0039	CPXAR	CPXAR	arsT_BAC0714	fbpC_BAC0162	CPXAR	algW	arnA	baeR_BAC0039	baeR_BAC0039	baeR_BAC0039	algU
fleQ	bfrA_BAC0048	cpxR_BAC0533	EMHC	baeR_BAC0039	fleN	cpxR_BAC0533	arnA	ARNA	bcr_BAC0041	cheW_2	cheW_2	algW
flgC	cheW_2	emhR_BAC0145	fleN	bcr_BAC0041	fleQ	EMHC	ARNA	arsT_BAC0714	bfrA_BAC0048	CPXAR	CPXAR	cheY
flgG	cheY	fabL/ygaA_BAC0158	fleQ	bfrA_BAC0048	fleR	evgA_BAC0154	baeR_BAC0039	baeR_BAC0039	cadR_BAC0058	EMHC	CPXAR	baeR_BAC0039
flgH	corD_BAC0644	fbpC_BAC0162	flgC	cadR_BAC0058	fleS	fabL/ygaA_BAC0158	baeR_BAC0596	bcr_BAC0041	cepA_BAC0061	fabL/ygaA_BAC0158	cpxR_BAC0533	bfrA_BAC0048
flgI	corR_BAC0089	fleN	flgG	cheW_2	flgB	fbpC_BAC0162	bfrA_BAC0048	cheW_2	cheW_2	fbpB_BAC0161	cueA_BAC0102	cheW_2
flhA	CPXAR	fleQ	flgH	cheY	flgC	fecD_BAC0163	cheW_2	CPXAR	cheY	fbpC_BAC0162	EMHC	cheY
fliA	cusR/ylcA_BAC0111	flgC	flgI	copR_BAC0753	flgH	fecE_BAC0164	cheY	cpxR_BAC0533	chrR_BAC0538	fecD_BAC0163	emrBsm_BAC0500	chrR_BAC0538
fliE	dsbC_BAC0138	flgF	flhA	copS_BAC0637	flgI	fleN	chrR_BAC0538	EMHC	copR_BAC0083	fecE_BAC0164	evgA_BAC0154	cinA_BAC0067
fliF	emrAsm_BAC0499	flgG	fliA	corB_BAC0643	flhA	fleQ	corD_BAC0644	evgA_BAC0154	copR_BAC0719	fleN	fabL/ygaA_BAC0158	copR_BAC0083
fliG	emrBsm_BAC0500	flgI	fliG	corC_BAC0088	fliA	flgC	corR_BAC0089	fabL/ygaA_BAC0158	copR_BAC0753	fleQ	fbpB_BAC0161	copR_BAC0719
fliI	emrCsm_BAC0501	flhA	fliI	corD_BAC0644	fliE	flgG	CPXAR	fbpC_BAC0162	copS_BAC0637	flgB	fbpC_BAC0162	copS_BAC0637
fliM	EMRE	fliA	fliM	corR_BAC0089	fliF	flgH	cpxR_BAC0533	fecD_BAC0163	corB_BAC0643	flgC	fecD_BAC0163	corR_BAC0089
fliN	fabL/ygaA_BAC0158	fliE	fliN	CPXAR	fliG	flgI	czcR_BAC0125	fecE_BAC0164	corC_BAC0088	flgG	fecE_BAC0164	CPXAR
fliP	fecD_BAC0163	fliF	fliP	cpxR_BAC0533	fliI	flhA	dpsA_BAC0135	fleN	corD_BAC0644	flgH	fleN	cpxR_BAC0533
MexB	fecE_BAC0164	fliG	MexB	cueA_BAC0102	fliM	fliA	EMHC	fleQ	corR_BAC0089	flgI	fleQ	EMHC
MEXE	fleN	fliI	MEXE	cutE/lnt_BAC0115	fliN	fliF	evgA_BAC0154	fleR	CPXAR	flhA	fleR	emrAsm_BAC0499
MexF	fleQ	fliM	MexF	dsbC_BAC0138	fliP	fliG	fabL/ygaA_BAC0158	flgB	cpxR_BAC0533	fliA	flgB	emrBsm_BAC0500
MEXF	flgB	fliN	MEXF	emrAsm_BAC0499	fliQ	fliI	fabV_BAC0159	flgC	cueA_BAC0102	fliE	flgC	emrCsm_BAC0501
motA	flgC	fliP	motA	emrBsm_BAC0500	fliR	fliM	fbpB_BAC0161	flgG	cutE/lnt_BAC0115	fliF	flgG	evgA_BAC0154
motC	flgG	fliQ	motC	evgA_BAC0154	MexB	fliP	fbpC_BAC0162	flgH	dsbC_BAC0138	fliG	flgH	fabL_BAC0156
pilG	flgH	MEXE	OpmH	fabL/ygaA_BAC0158	MEXE	fliQ	fecD_BAC0163	flgI	emhA_BAC0142	fliI	flgI	fabL/ygaA_BAC0158
pilH	flgI	MexE	OPMH	fbpA_BAC0160	MexE	fpvA_BAC0169	fecE_BAC0164	flhA	emhB_BAC0143	fliM	flhA	fbpA_BAC0160
pilJ	flhA	MexF	pilG	fbpB_BAC0161	motA	galE_BAC0755	fleN	fliA	EMHC	fliN	fliA	fbpB_BAC0161
CpxR	fliA	MEXF	pilH	fbpC_BAC0162	motB	MexB	fleQ	fliE	emhC_BAC0144	fliP	fliE	fbpC_BAC0162
pstB_BAC0316	fliE	mexK	pilJ	fecD_BAC0163	motC	MEXE	fleR	fliF	emhR_BAC0145	fliQ	fliF	fecD_BAC0163
waaF	fliF	MEXK	pilM	fecE_BAC0164	mucD	MexF	fleS	fliG	emrAsm_BAC0499	galE_BAC0755	fliG	fecE_BAC0164
wtpC_BAC0604	fliG	mexT_BAC0244	pilO	fleN	MuxB	MEXF	flgB	fliI	emrBsm_BAC0500	mdeA_BAC0207	fliI	fleN
zraR/hydH_BAC0467	fliI	motA	pilR	fleQ	MUXB	mexJ_BAC0241	flgC	fliM	emrCsm_BAC0501	mdtB/yegN_BAC0211	fliM	fleQ
	fliM	motC	pilT	flgB	OpmH	mexK	flgG	fliN	evgA_BAC0154	MexB	fliN	fleS
	fliN	MuxB	CpxR	flgC	OPMH	MEXK	flgH	fliP	evgS_BAC0155	MEXE	fliP	flgC
	fliP	MUXB	tapW	flgG	ostA/lptD_BAC0296	mexK_BAC0242	flgI	fliQ	fabL/ygaA_BAC0158	MexE	fliQ	flgG
	fliQ	nikA_BAC0270	TriC	flgH	pilB	modB_BAC0599	flhA	fliR	fbpB_BAC0161	MexF	fliR	flgH
	fliR	nikB_BAC0271	TTGB	flgI	pilG	motA	fliA	fptA_BAC0168	fbpC_BAC0162	MEXF	fpvA_BAC0169	flgI
	fptA_BAC0168	nikD_BAC0273	waaF	flhA	pilH	nikA_BAC0270	fliE	fpvA_BAC0169	fecD_BAC0163	mexJ_BAC0241	galE_BAC0755	flhA
	fpvA_BAC0169	nikE_BAC0274	waaG	fliA	pilJ	nikB_BAC0271	fliF	mbtH_like	fecE_BAC0164	mexK	mdeA_BAC0207	fliA
	galE_BAC0755	OpmH	xcpA/pilD	fliE	pilM	nikD_BAC0273	fliG	mdeA_BAC0207	fleN	MEXK	merA_BAC0651	fliF
	glpF_BAC0181	OPMH	zraR/hydH_BAC0467	fliF	pilO	nikE_BAC0274	fliI	MexB	fleQ	mexK_BAC0242	MexB	fliG
	hupE_BAC0483	OprN		fliG	pilR	norM/pmpM_BAC0284	fliM	MEXE	fleR	mexT_BAC0244	MEXE	fliI
	marR_BAC0561	OPRN		fliI	pilT	oscA_BAC0358	fliN	MexE	fleS	mexW	MexE	fliM
	MexB	oxyRkp_BAC0536		fliM	tapW	ostA/lptD_BAC0296	fliP	MexF	flgB	motA	MexF	fliN
	MEXB	pilB		fliN	TTGA	oxyRkp_BAC0536	fliQ	MEXF	flgC	motB	MEXF	fliP
	MEXE	pilG		fliP	TTGB	pbrA_BAC0298	fliR	mexK	flgG	motC	mexK	fliR
	MexE	pilH		fliQ	waaA	pilG	fptA_BAC0168	MEXK	flgH	oprN_BAC0562	MEXK	fpvA_BAC0169
	mexE_BAC0238	pilI		fpvA_BAC0169	waaF	pilH	fpvA_BAC0169	mexT_BAC0244	flgI	pilB	mexT_BAC0244	galE_BAC0755
	MexF	pilJ		G2alt_BAC0490	waaP	pilJ	G2alt_BAC0490	modB_BAC0599	flhA	pilG	mexW	iri5_BAC0198
	MEXF	pilM		galE_BAC0755	xcpA/pilD	pilM	glpF_BAC0181	motA	fliA	pilH	motA	mbtH_like
	mexF_BAC0239	pilR		glpF_BAC0181		pvdS	iri1R_BAC0197	motC	fliE	pilJ	motB	mdeA_BAC0207
	mexJ_BAC0241	pilT		hmrR_BAC0190		recG_BAC0356	iri5_BAC0198	mucD	fliF	pilM	motC	mdtB/yegN_BAC0211
	mexK_BAC0242	pilU		iri1R_BAC0197		smdA_BAC0353	mbtH_like	mucP	fliG	pilO	oprN_BAC0562	MexB
	mexT_BAC0244	CpxR		mbtH_like		smfY_BAC0363	mdeA_BAC0207	nikB_BAC0271	fliI	pilR	oxyRkp_BAC0536	mexC_BAC0236
	mntP/yebN_BAC0252	pstB_BAC0316		mdeA_BAC0207		smrA_BAC0364	merA_BAC0651	nikD_BAC0273	fliM	pilT	pilG	MEXE
	motA	waaF		mdtB/yegN_BAC0211		terA_BAC0386	MexB	nikE_BAC0274	fliN	pmrA_BAC0487	pilH	MexE
	motC	waaG		MexB		terB_BAC0387	mexC_BAC0236	OpmH	fliP	CpxR	pilJ	MexF
	mucD	zraR/hydH_BAC0467		MEXE		terC_BAC0388	MEXE	OPMH	fliQ	pstB_BAC0316	pilT	MEXF
	nikE_BAC0274			MexE		terE_BAC0390	MexE	oxyRkp_BAC0536	G2alt_BAC0490	pvdS	pmrA_BAC0487	mexK
	norM/pmpM_BAC0284			MexF		TTGB	MexF	pilG	hmrR_BAC0190	TriC	CpxR	MEXK
	OprN			MEXF		waaF	MEXF	pilH	ibpA_BAC0194	TTGA	pstB_BAC0316	mexT_BAC0244
	OPRN			mexJ_BAC0241		wtpC_BAC0604	mexK	pilJ	iri1R_BAC0197	TTGB	smdA_BAC0353	mexW
	oprN_BAC0562			mexK_BAC0463		znuA/yebL_BAC0463	MEXK	CpxR	mdeA_BAC0207	waaF	TriC	modC_BAC0611
	ostA/lptD_BAC0296			MEXK		znuB/yebI_BAC0464	mexT_BAC0244	pstB_BAC0316	mdtB/yegN_BAC0211	waaG	TTGA	motA
	oxyRkp_BAC0536			mexK_BAC0242		znuC/yebM_BAC0465	modE_BAC0608	pvdA	mdtC/yegO_BAC0212	waaP	TTGB	motB
	phoB_BAC0530			mexT_BAC0244		zraR/hydH_BAC0467	motA	pvdH	MexB	wtpC_BAC0604	waaF	motC
	phoR_BAC0531			modB_BAC0599		zur/yjbK_BAC0470	motB	pvdL	MEXE	ydeP_BAC0438	waaG	mucD
	pilB			modC_BAC0611			motC	pvdM	MexE	zraR/hydH_BAC0467	waaP	nikB_BAC0271
	pilG			motA			mucD	pvdS	MexF		wtpC_BAC0604	nikD_BAC0273
	pilH			motC			MuxB	TriC	MEXF		zraR/hydH_BAC0467	nikE_BAC0274
	pilJ			mucD			MUXB	TTGB	mexJ_BAC0241			OpmH
	pilM			nikA_BAC0270			nikA_BAC0270	waaF	mexK			OPMH
	pilO			nikB_BAC0271			nikB_BAC0271	waaG	MEXK			oprN_BAC0562
	pilR			nikD_BAC0273			nikC_BAC0272	wtpC_BAC0604	mexK_BAC0242			oxyRkp_BAC0536
	pilT			nikE_BAC0274			nikD_BAC0273	zraR/hydH_BAC0467	mexT_BAC0244			pilB
	pilU			norM/pmpM_BAC0284			nikE_BAC0274		mexW			pilG
	pmpM_BAC0313			nreB_BAC0285			norM/pmpM_BAC0284		mexW_BAC0246			pilH
	pmrA_BAC0487			OpmH			OpmH		modE_BAC0608			pilJ
	CpxR			OPMH			OPMH		motA			pilM
	Pseudomonas_aeru@nosa_emrE			oprN_BAC0562			OprM		motB			pilR
	pstA_BAC0315			oscA_BAC0358			OprN		motC			pilT
	pstB_BAC0316			ostA/lptD_BAC0296			OPRN		MuxB			pmpM_BAC0313
	pstC_BAC0317			oxyRkp_BAC0536			oprN_BAC0562		MUXB			pmrA_BAC0487
	pstS_BAC0318			PA0320_BAC0297			ostA/lptD_BAC0296		nikA_BAC0270			CpxR
	recG_BAC0356			pbrA_BAC0298			oxyRkp_BAC0536		nikB_BAC0271			pstB_BAC0316
	sitA_BAC0349			pilG			phoB_BAC0530		nikC_BAC0272			pvdA
	sitB_BAC0350			pilH			pilG		nikD_BAC0273			pvdH
	sitC_BAC0351			pilJ			pilH		nikE_BAC0274			pvdL
	sitD_BAC0352			pitA_BAC0312			pilJ		norM/pmpM_BAC0284			pvdM
	ssmE_BAC0377			pmrA_BAC0487			pitA_BAC0312		nreB_BAC0285			pvdS
	sugE_BAC0378			CpxR			PmpM		OpmH			smdA_BAC0353
	terC_BAC0388			pstB_BAC0316			PMPM		OPMH			smrA_BAC0364
	tolCsm_BAC0507			pvdS			CpxR		oscA_BAC0358			sugE_BAC0378
	TriC			recG_BAC0356			pstB_BAC0316		ostA/lptD_BAC0296			TriC
	tupC_BAC0607			rpoS_BAC0335			pvdA		oxyRkp_BAC0536			TTGA
	waaC			ruvB_BAC0293			pvdH		pbrA_BAC0298			TTGB
	waaF			ruvB_BAC0355			pvdL		pbrT_BAC0302			vcaM_BAC0417
	waaG			smdA_BAC0353			pvdM		phoB_BAC0530			waaF
	waaP			smrA_BAC0364			pvdS		phoR_BAC0531			waaG
	wtpC_BAC0604			sodB_BAC0706			ruvB_BAC0355		pilG			waaP
	ydeP_BAC0438			ssmE_BAC0377			smdA_BAC0353		pilH			wtpC_BAC0604
	yqjH_BAC0452			sugE_BAC0378			smfY_BAC0363		pilJ			zraR/hydH_BAC0467
	zitB/ybgR_BAC0459			TriC			smrA_BAC0364		pilT			
	znuA/yebL_BAC0463			TTGA			sodB_BAC0706		pitA_BAC0312			
	znuB/yebI_BAC0464			ttgA_BAC0404			soxR_BAC0370		pmpM_BAC0313			
	znuC/yebM_BAC0465			TTGB			TriC		pmrA_BAC0487			
	zraR/hydH_BAC0467			TTGC			TTGA		CpxR			
				TTGR			TTGB		pstB_BAC0316			
				ttgR_BAC0413			vpoC_BAC0516		recG_BAC0356			
				waaF			waaA		rpoS_BAC0335			
				waaG			waaF		ruvB_BAC0355			
				waaP			waaG		smdA_BAC0353			
				wtpC_BAC0604			waaP		smrA_BAC0364			
				ydeP_BAC0438			wtpC_BAC0604		sugE_BAC0378			
				znuA/yebL_BAC0463			xcpA/pilD		TriC			
				znuB/yebI_BAC0464			ydeP_BAC0438		TTGA			
				znuC/yebM_BAC0465			yqjH_BAC0452		TTGB			
				zraR/hydH_BAC0467			znuB/yebI_BAC0464		waaF			
							zraR/hydH_BAC0467		waaG			
									waaP			
									wtpC_BAC0604			
									ydeP_BAC0438			
									znuA/yebL_BAC0463			
									znuB/yebI_BAC0464			
									znuC/yebM_BAC0465			
									zraR/hydH_BAC0467			

Supplemental table 3 - Core AMR genes within select phylogenetic groups, >3 genomes

gtdb_genome_representative	gtdb_taxonomy : d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;	gtdb_genome_representative2	gtdb_taxonomy : d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;2
GCF_000818015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_A;s__Pseudomonas_A balearica	GCF_000282275.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282275
GCF_000263395.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_A;s__Pseudomonas_A stutzeri_C	GCF_000282315.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282315
GCF_001650425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_B;s__Pseudomonas_B oryzihabitans_B	GCF_000282375.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282375
GCF_014522265.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_B;s__Pseudomonas_B oryzihabitans_D	GCF_000282415.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282415
GCF_900100795.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E abietaniphila	GCF_000282455.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282455
GCF_900109755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E agarici	GCF_000282475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282475
GCF_000467105.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E alcaligenes	GCF_000282495.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282495
GCF_000746525.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E alkylphenolica	GCF_000282515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282515
GCF_004025535.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E alkylphenolica_A	GCF_000346225.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000346225
GCF_900105355.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E anguilliseptica	GCF_000346755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000346755
GCF_900624995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E antarctica	GCF_000416235.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000416235
GCF_001647715.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E antarctica_A	GCF_000497835.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000497835
GCF_900113905.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E argentinensis	GCF_000512695.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000512695
GCF_001839655.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E argentinensis_B	GCF_000612585.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000612585
GCF_900103875.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E arsenicoxydans	GCF_000631985.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000631985
GCF_006438915.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E arsenicoxydans_A	GCF_000633255.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000633255
GCF_009932335.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E asiatica	GCF_000633395.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000633395
GCF_900105475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E asplenii	GCF_000690905.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000690905
GCF_004011535.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E asplenii_A	GCF_014596565.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000801235
GCF_900143095.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E asturiensis	GCF_000829415.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000829415
GCF_900103345.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E azotoformans	GCF_000931465.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000931465
GCF_002813455.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E baetica	GCF_000935045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000935045
GCF_000820515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E batumici	GCF_000952175.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000952175
GCF_900100495.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E benzenivorans	GCF_000955815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000955815
GCF_002934685.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E bohemica	GCF_001269775.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001269775
GCF_000585995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_A	GCF_001269805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001269805
GCF_003732265.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_D	GCF_001269815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001269815
GCF_003732225.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E brassicacearum_E	GCF_001269925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001269925
GCF_000503215.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E canadensis	GCF_001297015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001297015
GCF_002158995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E caspiana	GCF_001297125.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001297125
GCF_001983175.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E cedrina	GCF_001308855.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001308855
GCF_900074915.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E cerasi	GCF_001421885.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001421885
GCF_001921865.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E chlororaphis_D	GCF_003627215.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001422615
GCF_001023535.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E chlororaphis_E	GCF_001427125.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001427125
GCF_003850345.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E chlororaphis_F	GCF_001429045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001429045
GCF_003731765.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E chlororaphis_G	GCF_003850565.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001705835
GCF_900104015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E cichorii	GCF_001952855.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001952855
GCF_003700275.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E cichorii_B	GCF_001984065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp001984065
GCF_900105555.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E coleopterorum	GCF_002000165.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002000165
GCF_900115475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E composti	GCF_002018875.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002018875
GCF_900103225.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E congelans	GCF_006094435.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002029345
GCF_001269905.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E corrugata	GCF_002080045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002080045
GCF_900110925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E cuatrocienegasensis	GCF_002112385.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002112385
GCF_900106095.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E deceptionensis	GCF_004006335.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002112885
GCF_001411475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E endophytica	GCF_002113025.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113025
GCF_900102035.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E extremaustralis	GCF_002113045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113045
GCF_900100535.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E flavescens	GCF_002113125.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113125
GCF_002087235.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E floridensis	GCF_002113165.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113165
GCF_900215245.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens	GCF_002113295.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113295
GCF_000802965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_A	GCF_002113375.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113375
GCF_001307275.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AA	GCF_002113595.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002113595
GCF_002022275.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AJ	GCF_002165135.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002165135
GCF_002865505.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AK	GCF_002204795.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002204795
GCF_002837185.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AM	GB_GCA_004000515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002255865
GCF_001708445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AN	GCF_009707325.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002263605
GCF_000817895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AO	GCF_002277815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002277815
GCF_013407925.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AP	GCF_002286815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002286815
GCF_000217955.3	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_AQ	GCF_002303925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002303925
GCF_000293885.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_B	GB_GCA_002338065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002338065
GCF_003626995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BA	GB_GCA_002354655.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002354655
GCF_003851495.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BB	GCF_002356535.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002356535
GCF_003732725.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BC	GCF_013522705.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002383725
GCF_005938625.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BD	GB_GCA_002386445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002386445
GCF_003732425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BE	GB_GCA_002419965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002419965
GCF_003352875.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BF	GCF_002699985.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002699985
GCF_900636825.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_BG	GCF_002753995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002753995
GCF_002022265.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_C	GB_GCA_002754355.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002754355
GCF_001307155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_E	GCF_002814235.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002814235
GCF_013363935.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_F	GCF_002836515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002836515
GB_GCA_902825215.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_G	GB_GCA_002840155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002840155
GCF_000292795.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_H	GCF_002843585.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002843585
GCF_002417665.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_J	GCF_002843605.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002843605
GCF_001679645.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_K	GCF_002874965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002874965
GCF_001648775.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_M	GCF_002878485.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002878485
GCF_001020875.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_N	GCF_002901475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002901475
GCF_000967965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_O	GCF_002979555.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002979555
GCF_001623525.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_Q	GCF_002980155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp002980155
GCF_000281895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_S	GCF_003612935.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003013355
GCF_000346775.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_T	GCF_003852315.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003014915
GCF_000968015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_W	GCF_003024385.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003024385
GCF_000730425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fluorescens_X	GCF_003033885.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003033885
GCF_900105835.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fragi	GCF_003050925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003050925
GCF_002269505.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fragi_D	GCF_003052515.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003052515
GCF_000802155.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_A	GCF_003053605.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003053605
GCF_001874645.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_B	GCF_003053805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003053805
GCF_002967995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_D	GCF_003060885.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003060885
GCF_900105495.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_E	GCF_003061005.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003061005
GCF_003732705.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_F	GCF_003097075.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003097075
GCF_003732665.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E frederiksbergensis_G	GCF_003105155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003105155
GCF_013408425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fulva_A	GCF_003151075.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003151075
GCF_000364705.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E fuscovaginae_A	GCF_003205205.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003205205
GCF_001983165.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E gessardii	GCF_003208475.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003208475
GCF_900111735.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E graminis	GCF_003331055.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003331055
GCF_001705435.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E graminis_B	GB_GCA_003428805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003428805
GCF_004364335.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E graminis_D	GCF_003596405.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003596405
GCF_900105485.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E granadensis	GCF_003605735.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003605735
GCF_007858185.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E grimontii	GCF_003851685.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003851685
GCF_001043025.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E helleri	GCF_003852235.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003852235
GCF_000783395.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E helmanticensis	GCF_003852405.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003852405
GCF_003671955.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E hunanensis	GCF_003935425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp003935425
GCF_002236115.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E jessenii	GCF_004011755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004011755
GCF_004682055.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E kairouanensis	GCF_004124275.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004124275
GCF_001269885.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E kilonensis	GCF_004136015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004136015
GCF_009601685.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E kilonensis_B	GCF_004214895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004214895
GCF_002003425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E koreensis_A	GCF_004310295.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004310295
GB_GCA_003049825.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E koreensis_B	GCF_004353925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004353925
GCF_001605965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E koreensis_C	GCF_004364435.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004364435
GCF_003999415.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E koreensis_E	GCF_004519405.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004519405
GCF_003352185.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E kribbensis	GCF_004786035.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004786035
GCF_001439845.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lactis	GCF_004803755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp004803755
GCF_002934665.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E laurylsulfatiphila	GCF_005233515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp005233515
GCF_002906155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E laurylsulfativorans	GCF_006227205.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp006227205
GCF_001439685.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E libanensis	GCF_900005815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900005815
GCF_001042905.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lini	GCF_900101695.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900101695
GCF_001042985.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lundensis	GCF_900107395.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900107395
GCF_002269125.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lundensis_A	GCF_900108875.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900108875
GCF_002563895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lurida	GCF_900109995.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900109995
GCF_000759445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E lutea	GCF_900110765.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900110765
GCF_900106065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mandelii	GCF_900113505.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900113505
GCF_002251635.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mandelii_B	GCF_900156465.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900156465
GCF_006438925.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mandelii_C	GCF_900187425.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187425
GCF_004364415.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mandelii_D	GCF_900187445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187445
GCF_900105325.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E marginalis	GCF_900187505.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187505
GCF_000774145.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mediterranea	GCF_900187515.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187515
GCF_002091715.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E migulae	GCF_900187565.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187565
GCF_900105115.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mohnii	GCF_900187575.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187575
GCF_000730605.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E monteilii	GCF_900187605.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187605
GCF_001534745.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E monteilii_A	GCF_900187615.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187615
GCF_900105805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E moraviensis_A	GCF_009834545.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187635
GCF_000621225.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mosselii	GCF_900187645.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900187645
GCF_900106045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E mucidolens	GB_GCA_900475325.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900475325
GCF_004682045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E nabeulensis	GCF_900580675.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900580675
GCF_001439815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E orientalis	GCF_900580865.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900580865
GCF_002934065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E orientalis_A	GCF_900581005.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900581005
GCF_003031675.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E palleroniana	GCF_900582195.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900582195
GCF_002021815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E parafulva_B	GCF_900583165.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900583165
GCF_001439735.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E paralactis	GCF_013201545.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900585815
GCF_900099645.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E peli	GCF_900590755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900590755
GCF_001439785.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E poae	GCF_900591205.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900591205
GCF_900105155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E prosekii	GB_GCA_900596015.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900596015
GCF_000397205.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E protegens	GCF_900601905.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900601905
GCF_003363755.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E protegens_A	GCF_900602065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp900602065
GCF_003732485.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E protegens_B	GCF_900105675.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E synxantha
GCF_001043005.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E psychrophila	GCF_000263715.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E synxantha_A
GCF_000282975.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E psychrophila_A	GCF_000507185.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae
GCF_900142655.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E punonensis	GCF_001698815.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae_A
GCF_000412675.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida	GCF_001238485.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae_E
GCF_001636055.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_B	GCF_000452705.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae_F
GCF_000019445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_E	GCF_000737245.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae_J
GCF_001976065.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_G	GCF_009176725.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E syringae_M
GCF_002356095.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_H	GCF_900475285.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E taetrolens
GCF_002025705.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_J	GCF_001269655.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E thivervalensis
GCF_002094795.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_K	GCF_002813445.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E tolaasii
GCF_001941965.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_L	GCF_900115695.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E toyotomiensis
GCF_013393325.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_P	GCF_001401155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E tremae
GCF_000498395.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_Q	GCF_001439805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E trivialis
GCF_000708715.2	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_R	GCF_001186335.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E trivialis_B
GCF_003228315.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_S	GCF_002236105.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E umsongensis
GCF_002741075.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_T	GCF_004348895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E vancouverensis
GCF_004357765.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_U	GCF_001439695.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E veronii
GCF_005080685.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E putida_V	GCF_001294575.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E versuta
GCF_900536025.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E reidholzensis	GCF_001642795.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E viridiflava
GCF_001945365.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E reinekei	GCF_002723575.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E viridiflava_B
GCF_000761155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E rhizosphaerae	GCF_003702045.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E viridiflava_C
GCF_007858255.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E rhodesiae	GCF_000425805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E vranovensis
GCF_900107155.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E salomonii	GCF_001043055.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E weihenstephanensis
GCF_900105185.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E saponiphila	GCF_900105735.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E yamanorum
GCF_900102335.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E seleniipraecipitans	GCF_900111835.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_F;s__Pseudomonas_F otitidis
GCF_001661075.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E silesiensis	GCF_000412695.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_F;s__Pseudomonas_F resinovorans_A
GCF_900111895.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E simiae	GCF_900109175.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_K;s__Pseudomonas_K linyingensis
GCF_012935715.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000242655	GCF_900104805.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_K;s__Pseudomonas_K oryzae
GCF_000263855.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000263855	GCF_900115715.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_K;s__Pseudomonas_K sagittaria
GCF_000282195.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282195	GCF_001457615.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Pseudomonas aeruginosa
GCF_000282215.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pseudomonas_E sp000282215	GCF_000689415.1	o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__Pseudomonas knackmussii

Supplemental table 4 - RefSeq numbers for all publically available *Pseudomonas* genomes

MOIL14HWK12I1_B_oryzihabitans				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	0.07260787	0.01193557	6.08332017	1.18E-09
cond.week	0.00176449	0.00051874	3.40147372	**0.00067024**
cond.soil_temp_3day	0.00089119	0.00063169	1.41080048	0.15830345
cond.precip_3day	-0.0083079	0.00743194	-1.1178608	0.26362646
Pseudomonas_E_fluorescens				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	-1.3519327	0.23832606	-5.672618	1.41E-08
cond.week	-0.0139357	0.01035811	-1.3453938	0.17849811
cond.soil_temp_3day	-0.0427344	0.01261337	-3.3880198	**0.00070399**
cond.precip_3day	0.03798887	0.14839891	0.25599155	0.79795736
Pseudomonas_E_marginalis				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	54.07828	15.1087769	3.5792626	0.00034457
cond.week	-1.744737	0.65665627	-2.657002	0.0078839
cond.soil_temp_3day	-3.5319141	0.79962992	-4.4169359	**1.00E-05**
cond.precip_3day	5.38534967	9.40780889	0.572434	0.56702799
Pseudomonas_E_sp000497835				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	-7.330206	0.69827256	-10.497629	8.86E-26
cond.week	0.10468919	0.03034826	3.44959471	**0.00056143**
cond.soil_temp_3day	-0.0043634	0.03695598	-0.1180706	0.90601168
cond.precip_3day	-0.2765168	0.43479461	-0.6359711	0.52479523
Pseudomonas_E_sp002263605				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	-0.8056306	0.33515665	-2.4037434	0.01622816
cond.week	-0.0223557	0.01456655	-1.5347306	0.12485002
cond.soil_temp_3day	-0.0707706	0.01773812	-3.989745	**6.61E-05**
cond.precip_3day	0.04128346	0.20869259	0.1978195	0.84318629
Pseudomonas_E_trivialis_B				
Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	**p.value**
cond.(Intercept)	-1.057023	0.27557992	-3.8356314	0.00012524
cond.week	-0.0272843	0.01197723	-2.2780141	0.02272574
cond.soil_temp_3day	-0.0488894	0.01458503	-3.3520296	**0.00080221**
cond.precip_3day	0.0416087	0.17159584	0.24248083	0.80840761

Supplemental table 5 - Statistically significant pairwise comparisons between abiotic factors and *Pseudomonas* relative abundance using a linear mixed-effect model. P-value adjusted to 0.001. Light blue indicates taxa is a member of group Oryzihabitans, dark blue Fluorescens, and dark red Syringae.

Data and code availability

Data and code will be made publicly available upon manuscript publication.

Bibliography

Alcock, B.P. *et al.* (2020) 'CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database', *Nucleic acids research*, 48(D1), pp. D517–D525.

Andrews, S. (2010) *FastQC: A quality control tool for high throughput sequence data*. Available at: https://github.com/s-andrews/FastQC.

Audenaert, K. *et al.* (2002) 'Induction of systemic resistance to Botrytis cinerea in tomato by Pseudomonas aeruginosa 7NSK2: role of salicylic acid, pyochelin, and pyocyanin', *Molecular plant-microbe interactions: MPMI*, 15(11), pp. 1147–1156.

Ausubel, F.M. (2005) 'Are innate immune signaling pathways in plants and animals conserved?', *Nature immunology*, 6(10), pp. 973–979.

Baluška, F. *et al.* (1996) 'Root cap mucilage and extracellular calcium as modulators of cellular growth in postmitotic growth zones of the maize root apex', *Botanica acta: Berichte der Deutschen Botanischen Gesellschaft = journal of the German Botanical Society*, 109(1), pp. 25–34.

Bankevich, A. *et al.* (2012) 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing', *Journal of computational biology: a journal of computational molecular cell biology*, 19(5), pp. 455–477.

Barnett, S.E. *et al.* (2021) 'Multisubstrate DNA stable isotope probing reveals guild structure of bacteria that mediate soil carbon cycling', *Proceedings of the National Academy of Sciences of the United States of America*, 118(47). doi:10.1073/pnas.2115292118.

Barret, M., Morrissey, J.P. and O'Gara, F. (2011) 'Functional genomics analysis of plant growth-promoting rhizobacterial traits involved in rhizosphere competence', *Biology and fertility of soils*, 47(7), p. 729.

Baumdicker, F., Hess, W.R. and Pfaffelhuber, P. (2012) 'The infinitely many genes model for the distributed genome of bacteria', *Genome biology and evolution*, 4(4), pp. 443–456.

Beghini, F. *et al.* (2021) 'Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3', *eLife*, 10. doi:10.7554/eLife.65088.

Belimov, A.A. *et al.* (2007) 'Pseudomonas brassicacearum strain Am3 containing 1-aminocyclopropane-1-carboxylate deaminase can show both pathogenic and growth-promoting properties in its interaction with tomato', *Journal of experimental botany*, 58(6), pp. 1485–1495.

Bendall, M.L. *et al.* (2016) 'Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations', *The ISME journal*, 10(7), pp. 1589–1601.

Benhamou, N., Bélanger, R.R. and Paulitz, T.C. (1996) 'Pre-inoculation of Ri T-DNA-transformed pea roots with Pseudomonas fluorescens inhibits colonization by Pythium ultimum Trow: an ultrastructural and cytochemical study', *Planta*, 199(1), pp. 105–117.

Benizri, E., Baudoin, E. and Guckert, A. (2001) 'Root Colonization by Inoculated Plant Growth-Promoting Rhizobacteria', *Biocontrol science and technology*, 11(5), pp. 557–574.

Berendsen, R.L., Pieterse, C.M.J. and Bakker, P.A.H.M. (2012) 'The rhizosphere microbiome and plant health', *Trends in plant science*, 17(8), pp. 478–486.

Bernardi, G. (1965) 'Chromatography of nucleic acids on hydroxyapatite', *Nature*, 206(4986), pp. 779–783.

Bigeard, J., Colcombet, J. and Hirt, H. (2015) 'Signaling mechanisms in pattern-triggered immunity (PTI)', *Molecular plant*, 8(4), pp. 521–539.

Blin, K. *et al.* (2019) 'antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline', *Nucleic acids research*, 47(W1), pp. W81–W87.

Brandl, M.T. and Mandrell, R.E. (2002) 'Fitness of Salmonella enterica serovar Thompson in the cilantro phyllosphere', *Applied and environmental microbiology*, 68(7), pp. 3614–3621.

Breidenbach, B., Pump, J. and Dumont, M.G. (2015) 'Microbial Community Structure in the Rhizosphere of Rice Plants', *Frontiers in microbiology*, 6, p. 1537.

Breitwieser, F.P., Baker, D.N. and Salzberg, S.L. (2018) 'KrakenUniq: confident and fast metagenomics classification using unique k-mer counts', *Genome biology*, 19(1), p. 198.

Brenner, D.J. (1973) 'Deoxyribonucleic acid reassociation in the taxonomy of Enteric bacteria', *International journal of systematic bacteriology*, 23(4), pp. 298–307.

Brenner, D.J. and Cowie, D.B. (1968) 'Thermal stability of Escherichia coli-Salmonella typhimurium deoxyribocleic acid duplexes', *Journal of bacteriology*, 95(6), pp. 2258–2262.

Brooks, M.E. *et al.* (2017) 'Modeling zero-inflated count data with glmmTMB', *bioRxiv*. doi:10.1101/132753.

Bulgarelli, D., Schlaeppi, K. and Spaepen, S. (2013) 'Structure and Functions of the Bacterial Microbiota of Plants', in Merchant, S.S. (ed.) *ANNUAL REVIEW OF PLANT BIOLOGY, VOL 64*, pp. 807–838.

Burlinson, P. *et al.* (2013) 'Pseudomonas fluorescens NZI7 repels grazing by C. elegans, a natural predator', *The ISME journal*, 7(6), pp. 1126–1138.

Burstrom, H. (1953) 'Physiology of Root Growth', *Annual review of plant physiology*, 4(1),

pp. 237–252.

Bushnell, B. (2014) *BBMap: short read aligner, and other bioinformatic tools*. Available at: https://sourceforge.net/projects/bbmap/.

Canard, B. and Sarfati, R.S. (1994) 'DNA polymerase fluorescent substrates with reversible 3'-tags', *Gene*, 148(1), pp. 1–6.

Caruccio, N. (2011) 'Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition', *Methods in molecular biology* , 733, pp. 241–255.

Ceja-Navarro, J.A. *et al.* (2015) 'Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee', *Nature communications*, 6, p. 7618.

Chen, L. *et al.* (2005) 'VFDB: a reference database for bacterial virulence factors', *Nucleic acids research*, 33(Database issue), pp. D325–8.

Chen, Q.-L. *et al.* (2020) 'Soil bacterial taxonomic diversity is critical to maintaining the plant productivity', *Environment international*, 140, p. 105766.

Chen, T. *et al.* (2020) 'A plant genetic network for preventing dysbiosis in the phyllosphere', *Nature*, 580(7805), pp. 653–657.

Collins, R.E. and Higgs, P.G. (2012) 'Testing the Infinitely Many Genes Model for the Evolution of the Bacterial Core Genome and Pangenome', *Molecular biology and evolution*, 29(11), pp. 3413–3425.

Compant, S., Clément, C. and Sessitsch, A. (2010) 'Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization', *Soil biology & biochemistry*, 42(5), pp. 669–678.

Contessi, B. *et al.* (2006) 'Evaluation of immunological parameters in farmed gilthead sea bream, Sparus aurata L., before and during outbreaks of "winter syndrome"', *Journal of fish diseases*, 29(11), pp. 683–690.

Cornelis, P. and Matthijs, S. (2007) 'Pseudomonas Siderophores and their Biological Significance', in Varma, A. and Chincholkar, S.B. (eds) *Microbial Siderophores*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 193–203.

Cox, C.L., Doroghazi, J.R. and Mitchell, D.A. (2015) 'The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles', *BMC genomics*, 16, p. 778.

Craig, L., Forest, K.T. and Maier, B. (2019) 'Type IV pili: dynamics, biophysics and functional consequences', *Nature reviews. Microbiology*, 17(7), pp. 429–440.

Dakora, F.D. and Phillips, D.A. (2002) 'Root exudates as mediators of mineral acquisition in low-nutrient environments', *Plant and soil*, 245(1), pp. 201–213.

Dalmastri, C. *et al.* (1999) ‘Soil Type and Maize Cultivar Affect the Genetic Diversity of Maize Root-Associated Burkholderia cepacia Populations’, *Microbial ecology*, 38(3), pp. 273–284.

Dastogeer, K.M.G. *et al.* (2020) ‘Plant microbiome–an account of the factors that shape community composition and diversity’, *Current Plant Biology*, 23, p. 100161.

Dempsey, D.A., Shah, J. and Klessig, D.F. (1999) ‘Salicylic Acid and Disease Resistance in Plants’, *Critical reviews in plant sciences*, 18(4), pp. 547–575.

Deng, S. *et al.* (2021) ‘Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome’, *The ISME journal*, 15(11), pp. 3181–3194.

De Vleesschauwer, D. *et al.* (2008) ‘Pseudomonas fluorescens WCS374r-induced systemic resistance in rice against Magnaporthe oryzae is based on pseudobactin-mediated priming for a salicylic acid-repressible multifaceted defense response’, *Plant physiology*, 148(4), pp. 1996–2012.

DeYoung, B.J. and Innes, R.W. (2006) ‘Plant NBS-LRR proteins in pathogen sensing and host defense’, *Nature immunology*, 7(12), pp. 1243–1249.

von Dohlen, C.D. *et al.* (2013) ‘Diversity of proteobacterial endosymbionts in hemlock woolly adelgid (Adelges tsugae) (Hemiptera: Adelgidae) from its native and introduced range’, *Environmental microbiology*, 15(7), pp. 2043–2062.

von Dohlen, C.D. *et al.* (2017) ‘Dynamic Acquisition and Loss of Dual-Obligate Symbionts in the Plant-Sap-Feeding Adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea)’, *Frontiers in microbiology*, 8, p. 1037.

van Dongen, S. and Abreu-Goodger, C. (2012) ‘Using MCL to extract clusters from networks’, *Methods in molecular biology* , 804, pp. 281–295.

Douglas, G.M. *et al.* (2020) ‘PICRUSt2 for prediction of metagenome functions’, *Nature biotechnology*, 38(6), pp. 685–688.

Durrant, W.E. and Dong, X. (2004) ‘Systemic acquired resistance’, *Annual review of phytopathology*, 42, pp. 185–209.

Edwards, J. *et al.* (2015) ‘Structure, variation, and assembly of the root-associated microbiomes of rice’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(8), pp. E911–20.

Ehrlich, G.D. *et al.* (2010) ‘The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes’, *FEMS immunology and medical microbiology*, 59(3), pp. 269–279.

Ellison, C.K. *et al.* (2018) ‘Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in Vibrio cholerae’, *Nature microbiology*, 3(7), pp.

773–780.

ERA5 Climate reanalysis (no date) *European Centre for Medium-range Weather Forecasts*. Available at: https://www.ecmwf.int/en/research/climate-reanalysis (Accessed: 2020).

Eren, A.M. *et al.* (2021) 'Community-led, integrated, reproducible multi-omics with anvi'o', *Nature microbiology*, 6(1), pp. 3–6.

Ewels, P. *et al.* (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics* , 32(19), pp. 3047–3048.

Feinstein, L.M., Sul, W.J. and Blackwood, C.B. (2009) 'Assessment of bias associated with incomplete extraction of microbial DNA from soil', *Applied and environmental microbiology*, 75(16), pp. 5428–5433.

Fisher, R.M. *et al.* (2017) 'The evolution of host-symbiont dependence', *Nature communications*, 8, p. 15973.

Foster, K.R. *et al.* (2017) 'The evolution of the host microbiome as an ecosystem on a leash', *Nature*, 548(7665), pp. 43–51.

Foster, R.C. *et al.* (1983) *Ultrastructure of the root-soil interface*. American Phytopathological Society.

Founoune, H. *et al.* (2002) 'Mycorrhiza Helper Bacteria Stimulate Ectomycorrhizal Symbiosis of Acacia holosericea with Pisolithus alba', *The New phytologist*, 153(1), pp. 81–89.

Frank, A.C., Saldierna Guzmán, J.P. and Shay, J.E. (2017) 'Transmission of Bacterial Endophytes', *Microorganisms*, 5(4). doi:10.3390/microorganisms5040070.

Frey-Klett, P., Pierrat, J.C. and Garbaye, J. (1997) 'Location and Survival of Mycorrhiza Helper Pseudomonas fluorescens during Establishment of Ectomycorrhizal Symbiosis between Laccaria bicolor and Douglas Fir', *Applied and environmental microbiology*, 63(1), pp. 139–144.

Gamalero, E. *et al.* (2005) 'Colonization of tomato root seedling by Pseudomonas fluorescens 92 rkG5: spatio-temporal dynamics, localization, organization, viability, and culturability', *Microbial ecology*, 50(2), pp. 289–297.

Gans, J., Wolinsky, M. and Dunbar, J. (2005) 'Computational improvements reveal great bacterial diversity and high metal toxicity in soil', *Science*, 309(5739), pp. 1387–1390.

Garbaye, J. (1994) 'Mycorrhization helper bacteria: a new dimension to the mycorrhizal symbiosis', *Acta botanica Gallica: bulletin de la Societe botanique de France*, 141(4), pp. 517–521.

García-Salamanca, A. *et al.* (2013) 'Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil', *Microbial biotechnology*, 6(1), pp. 36–44.

Garrido-Sanz, D. *et al.* (2016) 'Genomic and Genetic Diversity within the Pseudomonas fluorescens Complex', *PloS one*, 11(2), p. e0150183.

Giovannini, L. *et al.* (2020) 'Arbuscular Mycorrhizal Fungi and Associated Microbiota as Plant Biostimulants: Research Strategies for the Selection of the Best Performing Inocula', *Agronomy*, 10(1), p. 106.

Gislason, A.S. and de Kievit, T.R. (2020) 'Friend or foe? Exploring the fine line between Pseudomonas brassicacearum and phytopathogens', *Journal of medical microbiology*, 69(3), pp. 347–360.

Gomila, M. *et al.* (2015) 'Phylogenomics and systematics in Pseudomonas', *Frontiers in microbiology*, 6, p. 214.

Gondry, M. *et al.* (2009) 'Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes', *Nature chemical biology*, 5(6), pp. 414–420.

Gryndler, M. and Vosátka, M. (1996) 'The response of Glomus fistulosum–maize mycorrhiza to treatments with culture fractions from Pseudomonas putida', *Mycorrhiza*, 6(3), pp. 207–211.

Guerrero, R., Margulis, L. and Berlanga, M. (2013) 'Symbiogenesis: the holobiont as a unit of evolution', *International microbiology: the official journal of the Spanish Society for Microbiology*, 16(3), pp. 133–143.

Haney, C.H. *et al.* (2018) 'Rhizosphere-associated Pseudomonas induce systemic resistance to herbivores at the cost of susceptibility to bacterial pathogens', *Molecular ecology*, 27(8), pp. 1833–1847.

Hardoim, P.R., van Overbeek, L.S. and van Elsas, J.D. (2008) 'Properties of bacterial endophytes and their proposed role in plant growth', *Trends in microbiology*, 16(10), pp. 463–471.

Hawes, M.C. *et al.* (2011) 'Extracellular DNA: the tip of root defenses?', *Plant science: an international journal of experimental plant biology*, 180(6), pp. 741–745.

Hayat, R. *et al.* (2010) 'Soil beneficial bacteria and their role in plant growth promotion: a review', *Annals of microbiology*, 60(4), pp. 579–598.

Hermann, L. *et al.* (2020) 'The ups and downs of ectoine: structural enzymology of a major microbial stress protectant and versatile nutrient', *Biological chemistry*, 401(12), pp. 1443–1468.

Hesse, C. *et al.* (2018) 'Genome-based evolutionary history of Pseudomonas spp', *Environmental microbiology*, 20(6), pp. 2142–2159.

Hiller, N.L. *et al.* (2007) 'Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome', *Journal of bacteriology*, 189(22), pp. 8186–8195.

HILTNER and L (1904) 'Uber nevere Erfahrungen und Probleme auf dem Gebiet der Boden Bakteriologie und unter besonderer Beurchsichtigung der Grundungung und Broche', *Arbeit. Deut. Landw. Ges. Berlin*, 98, pp. 59–78.

Hirano Susan S. and Upper Christen D. (2000) 'Bacteria in the Leaf Ecosystem with Emphasis onPseudomonas syringae—a Pathogen, Ice Nucleus, and Epiphyte', *Microbiology and molecular biology reviews: MMBR*, 64(3), pp. 624–653.

Horbach, R. *et al.* (2011) 'When and how to kill a plant cell: infection strategies of plant pathogenic fungi', *Journal of plant physiology*, 168(1), pp. 51–62.

Huerta-Cepas, J., Serra, F. and Bork, P. (2016) 'ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data', *Molecular biology and evolution*, 33(6), pp. 1635–1638.

Hultman, T. *et al.* (1989) 'Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support', *Nucleic acids research*, 17(13), pp. 4937–4946.

Humphris, S.N. *et al.* (2005) 'Root cap influences root colonisation by Pseudomonas fluorescens SBW25 on maize', *FEMS microbiology ecology*, 54(1), pp. 123–130.

Huson, D.H. *et al.* (2016) 'MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data', *PLoS computational biology*, 12(6), p. e1004957.

Hyman, E.D. (1988) 'A new method of sequencing DNA', *Analytical biochemistry*, 174(2), pp. 423–436.

Iavicoli, A. *et al.* (2003) 'Induced systemic resistance in Arabidopsis thaliana in response to root inoculation with Pseudomonas fluorescens CHA0', *Molecular plant-microbe interactions: MPMI*, 16(10), pp. 851–858.

Izumi, H. *et al.* (2006) 'Endobacteria in some ectomycorrhiza of Scots pine (Pinus sylvestris)', *FEMS microbiology ecology*, 56(1), pp. 34–43.

Jiang, H. *et al.* (2014) 'Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads', *BMC bioinformatics*, 15, p. 182.

Johnston-Monje, D. *et al.* (2016) 'Bacterial populations in juvenile maize rhizospheres originate from both seed and soil', *Plant and soil*, 405(1-2), pp. 337–355.

Johnston-Monje, D., Gutiérrez, J.P. and Lopez-Lavalle, L.A.B. (2021) 'Seed-Transmitted Bacteria and Fungi Dominate Juvenile Plant Microbiomes', *Frontiers in microbiology*, 12, p. 737616.

Johnston-Monje, D. and Raizada, M.N. (2011) 'Conservation and diversity of seed associated endophytes in Zea across boundaries of evolution, ethnography and ecology', *PloS one*, 6(6), p. e20396.

Jones, A.L., DeShazer, D. and Woods, D.E. (1997) 'Identification and characterization of a two-component regulatory system involved in invasion of eukaryotic cells and heavy-metal resistance in Burkholderia pseudomallei', *Infection and immunity*, 65(12), pp. 4972–4977.

Jones, D.L., Hodge, A. and Kuzyakov, Y. (2004) 'Plant and mycorrhizal regulation of rhizodeposition', *The New phytologist*, 163(3), pp. 459–480.

Jones, J.D.G. and Dangl, J.L. (2006) 'The plant immune system', *Nature*, 444(7117), pp. 323–329.

Kalyuzhnaya, M.G. *et al.* (2008) 'High-resolution metagenomics targets specific functional types in complex microbial communities', *Nature biotechnology*, 26(9), pp. 1029–1034.

Karasov, T.L. *et al.* (2018) 'Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales', *Cell host & microbe*, 24(1), pp. 168–179.e4.

Kautsar, S.A. *et al.* (2021) 'BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters', *GigaScience*, 10(1). doi:10.1093/gigascience/giaa154.

Kemen, E. and Jones, J.D.G. (2012) 'Obligate biotroph parasitism: can we link genomes to lifestyles?', *Trends in plant science*, 17(8), pp. 448–457.

Kersters, K. *et al.* (1996) 'Recent Changes in the Classification of the Pseudomonads: an Overview, Systematic and Applied Microbiology', *Systematic and Applied Microbiology*, 19(4), pp. 465–477.

Khomarbaghi, Z. (2019) *The function and origin of multi-copy tRNA genes in Pseudomonas fluorescens*. Christian-Albrechts-Universität zu Kiel. Available at: https://macau.uni-kiel.de/receive/macau_mods_00000443.

Koeck, M., Hardham, A.R. and Dodds, P.N. (2011) 'The role of effectors of biotrophic and hemibiotrophic fungi in infection', *Cellular microbiology*, 13(12), pp. 1849–1857.

Koehorst, J.J. *et al.* (2016) 'Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data', *Scientific reports*, 6, p. 38699.

Konstantinidis, K.T. and Tiedje, J.M. (2004) 'Trends between gene content and genome size in prokaryotic species with larger genomes', *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), pp. 3160–3165.

Kraffczyk, I., Trolldenier, G. and Beringer, H. (1984) 'Soluble root exudates of maize: Influence of potassium supply and rhizosphere microorganisms', *Soil biology &*

biochemistry, 16(4), pp. 315–322.

Kroll, S., Agler, M.T. and Kemen, E. (2017) 'Genomic dissection of host-microbe and microbe-microbe interactions for advanced plant breeding', *Current opinion in plant biology*, 36, pp. 71–78.

Kuzyakov, Y., Domanski, G. and Others (2000) 'Carbon input by plants into the soil. Review', *Journal of Plant Nutrition and Soil Science*, 163(4), pp. 421–431.

Lapierre, P. and Gogarten, J.P. (2009) 'Estimating the size of the bacterial pan-genome', *Trends in genetics: TIG*, 25(3), pp. 107–110.

Latour, X. *et al.* (1996) 'The composition of fluorescent pseudomonad populations associated with roots is influenced by plant and soil type', *Applied and environmental microbiology*, 62(7), pp. 2449–2456.

Latour, X. *et al.* (1999) 'The establishment of an introduced community of fluorescent pseudomonads in the soil and in the rhizosphere is affected by the soil type', *FEMS microbiology ecology*, 30(2), pp. 163–170.

Leveau, J.H. and Lindow, S.E. (2001) 'Appetite of an epiphyte: quantitative monitoring of bacterial sugar consumption in the phyllosphere', *Proceedings of the National Academy of Sciences of the United States of America*, 98(6), pp. 3446–3453.

Lilley, A.K. *et al.* (2006) 'The dispersal and establishment of pseudomonad populations in the phyllosphere of sugar beet by phytophagous caterpillars', *FEMS microbiology ecology*, 24(2), pp. 151–157.

Lind, A.L. and Pollard, K.S. (2020) 'Accurate and sensitive detection of microbial eukaryotes from metagenomic shotgun sequencing data', *bioRxiv*. doi:10.1101/2020.07.22.216580.

Lindow, S.E. and Brandl, M.T. (2003) 'Microbiology of the phyllosphere', *Applied and environmental microbiology*, 69(4), pp. 1875–1883.

Liu, H. *et al.* (2017) 'Inner Plant Values: Diversity, Colonization and Benefits from Endophytic Bacteria', *Frontiers in microbiology*, 8, p. 2552.

van Loon, L.C., Bakker, P.A. and Pieterse, C.M. (1998) 'Systemic resistance induced by rhizosphere bacteria', *Annual review of phytopathology*, 36, pp. 453–483.

Lopes, L.D. *et al.* (2018) 'Genome variations between rhizosphere and bulk soil ecotypes of a Pseudomonas koreensis population', *Environmental microbiology*, 20(12), pp. 4401–4414.

Louca, S., Doebeli, M. and Parfrey, L.W. (2018) 'Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem', *Microbiome*, 6(1), p. 41.

Lu, J. *et al.* (2017) ‘Bracken: estimating species abundance in metagenomics data’, *PeerJ Computer Science*, 3, p. e104.

Lundberg, D.S. *et al.* (2012) ‘Defining the core Arabidopsis thaliana root microbiome’, *Nature*, 488(7409), pp. 86–90.

Lynch, J.M. and Whipps, J.M. (1990) ‘Substrate flow in the rhizosphere’, *Plant and soil*, 129(1), pp. 1–10.

Lynch, M. *et al.* (2016) ‘Genetic drift, selection and the evolution of the mutation rate’, *Nature reviews. Genetics*, 17(11), pp. 704–714.

Lysenko, O. (1961) ‘Pseudomonas--an attempt at a general classification’, *Journal of general microbiology*, 25, pp. 379–408.

Magi, G.E. *et al.* (2009) ‘Experimental Pseudomonas anguilliseptica infection in turbot Psetta maxima (L.): a histopathological and immunohistochemical study’, *European journal of histochemistry: EJH*, 53(2), p. e9.

Maiden, M.C. *et al.* (1998) ‘Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms’, *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), pp. 3140–3145.

Maiden, M.C.J. (2006) ‘Multilocus sequence typing of bacteria’, *Annual review of microbiology*, 60, pp. 561–588.

Maignien, L. *et al.* (2014) ‘Ecological succession and stochastic variation in the assembly of Arabidopsis thaliana phyllosphere communities’, *mBio*, 5(1), pp. e00682–13.

Mallet, J. (2008) ‘Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1506), pp. 2971–2986.

Mao, Y. *et al.* (2014) ‘Enrichment of specific bacterial and eukaryotic microbes in the rhizosphere of switchgrass (Panicum virgatum L.) through root exudates’, *Environmental microbiology reports*, 6(3), pp. 293–306.

Marschner, H., Treeby, M. and Römheld, V. (1989) ‘Role of root- induced changes in the rhizosphere for iron acquisition in higher plants’, *Zeitschrift fuer Pflanzenernaehrung und Bodenkunde*, 152(2), pp. 197–204.

Matus-Acuña, V., Caballero-Flores, G. and Martínez-Romero, E. (2021) ‘The influence of maize genotype on the rhizosphere eukaryotic community’, *FEMS microbiology ecology*, 97(6). doi:10.1093/femsec/fiab066.

Maxwell, C.A. and Phillips, D.A. (1990) ‘Concurrent Synthesis and Release of nod-Gene-Inducing Flavonoids from Alfalfa Roots’, *Plant physiology*, 93(4), pp. 1552–1558.

McCutcheon, J.P. and Moran, N.A. (2011) ‘Extreme genome reduction in symbiotic bacteria’, *Nature reviews. Microbiology*, 10(1), pp. 13–26.

McDonald, M.J. *et al.* (2015) ‘The evolutionary dynamics of tRNA-gene copy number and codon-use in E. coli’, *BMC evolutionary biology*, 15, p. 163.

McMurdie, P.J. and Holmes, S. (2013) ‘phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data’, *PloS one*, 8(4), p. e61217.

Mendes, R. *et al.* (2011) ‘Deciphering the rhizosphere microbiome for disease-suppressive bacteria’, *Science*, 332(6033), pp. 1097–1100.

Meziane, H. *et al.* (2005) ‘Determinants of Pseudomonas putida WCS358 involved in inducing systemic resistance in plants’, *Molecular plant pathology*, 6(2), pp. 177–185.

Meziti, A. *et al.* (2021) ‘The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample’, *Applied and environmental microbiology*, 87(6). doi:10.1128/AEM.02593-20.

Migula, W. (1895) *Über ein neues System der Bakterien*.

Mishustin, E.N. and Naumova, A.N. (1962) ‘Bacterial fertilizers, their effectiveness and mode of action’, *Mikrobiologiya*, 31(3), pp. 543–555.

Molina, N. and van Nimwegen, E. (2008) ‘Universal patterns of purifying selection at noncoding positions in bacteria’, *Genome research*, 18(1), pp. 148–160.

Molin, S. and Tolker-Nielsen, T. (2003) ‘Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure’, *Current opinion in biotechnology*, 14(3), pp. 255–261.

Monteith, A.J. *et al.* (2021) ‘Neutrophil extracellular traps enhance macrophage killing of bacterial pathogens’, *Science advances*, 7(37), p. eabj2101.

Mosquera-Rendón, J. *et al.* (2016) ‘Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species’, *BMC genomics*, 17, p. 45.

Mosse, B. (1962) ‘The establishment of vesicular-arbuscular mycorrhiza under aseptic conditions’, *Journal of general microbiology*, 27, pp. 509–520.

Müller, M., Deigele, C. and Ziegler, H. (1989) ‘Hormonal interactions in the rhizosphere of maize (Zea mays L.) and their effects on plant development’, *Zeitschrift fuer Pflanzenernaehrung und Bodenkunde*, 152(2), pp. 247–254.

Nakamura, Y. *et al.* (2004) ‘Biased biological functions of horizontally transferred genes in prokaryotic genomes’, *Nature genetics*, 36(7), pp. 760–766.

Naylor, D. and Coleman-Derr, D. (2017) ‘Drought Stress and Root-Associated Bacterial Communities’, *Frontiers in plant science*, 8, p. 2223.

Neidig, N. *et al.* (2011) ‘Secondary metabolites of Pseudomonas fluorescens CHA0 drive complex non-trophic interactions with bacterivorous nematodes’, *Microbial ecology*, 61(4), pp. 853–859.

Nelson, E.B. (2018) ‘The seed microbiome: Origins, interactions, and impacts’, *Plant and soil*, 422(1), pp. 7–34.

Neufeld, J.D. *et al.* (2007) ‘Methodological considerations for the use of stable isotope probing in microbial ecology’, *Microbial ecology*, 53(3), pp. 435–442.

Newman, M.-A. *et al.* (2013) ‘MAMP (microbe-associated molecular pattern) triggered immunity in plants’, *Frontiers in plant science*, 4, p. 139.

Nurk, S. *et al.* (2017) ‘metaSPAdes: a new versatile metagenomic assembler’, *Genome research*, 27(5), pp. 824–834.

Nyrén, P. (1987) ‘Enzymatic method for continuous monitoring of DNA polymerase activity’, *Analytical biochemistry*, 167(2), pp. 235–238.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) ‘Lateral gene transfer and the nature of bacterial innovation’, *Nature*, 405(6784), pp. 299–304.

Odell, R.E. *et al.* (2008) ‘Stage-dependent border cell and carbon flow from roots to rhizosphere’, *American journal of botany*, 95(4), pp. 441–446.

OikoLab Weather API (no date). Available at: https://oikolab.com (Accessed: 2020).

Oksanen, J. *et al.* (2020) ‘vegan: Community Ecology Package’. Available at: https://CRAN.R-project.org/package=vegan.

O’Leary, N.A. *et al.* (2016) ‘Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation’, *Nucleic acids research*, 44(D1), pp. D733–45.

Omoboye, O.O. *et al.* (2019) ‘Pseudomonas Cyclic Lipopeptides Suppress the Rice Blast Fungus Magnaporthe oryzae by Induced Resistance and Direct Antagonism’, *Frontiers in plant science*, 10, p. 901.

Pal, C. *et al.* (2014) ‘BacMet: antibacterial biocide and metal resistance genes database’, *Nucleic acids research*, 42(Database issue), pp. D737–43.

Palleroni, N.J. *et al.* (1973) ‘Nucleic acid homologies in the genus Pseudomonas’, *International journal of systematic bacteriology*, 23(4), pp. 333–339.

Palleroni, N.J. (2010) ‘The Pseudomonas story’, *Environmental microbiology*, 12(6), pp.

1377–1383.

Papayannopoulos, V. (2018) ‘Neutrophil extracellular traps in immunity and disease’, *Nature reviews. Immunology*, 18(2), pp. 134–147.

Parales, R.E. *et al.* (2017) ‘Genome Sequence of Pseudomonas putida Strain ASAD, an Acetylsalicylic Acid-Degrading Bacterium’, *Genome announcements*, 5(41). doi:10.1128/genomeA.01169-17.

Parks, D.H. *et al.* (2015) ‘CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes’, *Genome research*, 25(7), pp. 1043–1055.

Parks, D.H. *et al.* (2020) ‘A complete domain-to-species taxonomy for Bacteria and Archaea’, *Nature biotechnology*, 38(9), pp. 1079–1086.

Parks, D.H. *et al.* (2021) ‘Evaluation of the Microba Community Profiler for Taxonomic Profiling of Metagenomic Datasets From the Human Gut Microbiome’, *Frontiers in microbiology*, 12, p. 643682.

Parniske, M. (2008a) ‘Arbuscular mycorrhiza: the mother of plant root endosymbioses’, *Nature reviews. Microbiology*, 6(10), pp. 763–775.

Parniske, M. (2008b) ‘Arbuscular mycorrhiza: the mother of plant root endosymbioses’, *Nature reviews. Microbiology*, 6(10), pp. 763–775.

Pasolli, E. *et al.* (2019) ‘Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle’, *Cell*, 176(3), pp. 649–662.e20.

Peiffer, J.A. *et al.* (2013) ‘Diversity and heritability of the maize rhizosphere microbiome under field conditions’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), pp. 6548–6553.

Pepe-Ranney, C. *et al.* (2016) ‘Unearthing the Ecology of Soil Microorganisms Using a High Resolution DNA-SIP Approach to Explore Cellulose and Xylose Metabolism in Soil’, *Frontiers in microbiology*, 7, p. 703.

Pérez-Losada, M. *et al.* (2013) ‘Pathogen typing in the genomics era: MLST and the future of molecular epidemiology’, *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 16, pp. 38–53.

Ponce, G. *et al.* (2005) ‘Auxin and ethylene interactions control mitotic activity of the quiescent centre, root cap size, and pattern of cap cell differentiation in maize’, *Plant, cell & environment*, 28(6), pp. 719–732.

Popa, O. and Dagan, T. (2011) ‘Trends and barriers to lateral gene transfer in prokaryotes’, *Current opinion in microbiology*, 14(5), pp. 615–623.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PloS one*, 5(3), p. e9490.

Puente, M.E., Li, C.Y. and Bashan, Y. (2009) 'Endophytic bacteria in cacti seeds can improve the development of cactus seedlings', *Environmental and experimental botany*, 66(3), pp. 402–408.

Qiu, X., Kulasekara, B.R. and Lory, S. (2009) 'Role of Horizontal Gene Transfer in the Evolution of Pseudomonas aeruginosa Virulence', *Genome dynamics*, 6, pp. 126–139.

Quince, C. *et al.* (2017) 'Shotgun metagenomics, from sampling to analysis', *Nature biotechnology*, 35(9), pp. 833–844.

Raynaud, X. and Nunan, N. (2014) 'Spatial ecology of bacteria at the microscale in soil', *PloS one*, 9(1), p. e87217.

Reinhold-Hurek, B. and Hurek, T. (1998) 'Life in grasses: diazotrophic endophytes', *Trends in microbiology*, 6(4), pp. 139–144.

Remy, W. *et al.* (1994) 'Four hundred-million-year-old vesicular arbuscular mycorrhizae', *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), pp. 11841–11843.

Repka, L.M. *et al.* (2017) 'Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes', *Chemical reviews*, 117(8), pp. 5457–5520.

Robinson, S.L., Christenson, J.K. and Wackett, L.P. (2019) 'Biosynthesis and chemical diversity of β-lactone natural products', *Natural product reports*, 36(3), pp. 458–475.

Rodríguez, C.E. *et al.* (2018) 'Commentary: seed bacterial inhabitants and their routes of colonization', *Plant and soil*, 422(1), pp. 129–134.

Rodriguez-R, L.M. *et al.* (2018) 'Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity', *mSystems*, 3(3). doi:10.1128/mSystems.00039-18.

Roesch, L.F.W. *et al.* (2007) 'Pyrosequencing enumerates and contrasts soil microbial diversity', *The ISME journal*, 1(4), pp. 283–290.

Ronaghi, M. *et al.* (1996) 'Real-time DNA sequencing using detection of pyrophosphate release', *Analytical biochemistry*, 242(1), pp. 84–89.

Rossmann, M. *et al.* (2020) 'Multitrophic interactions in the rhizosphere microbiome of wheat: from bacteria and fungi to protists', *FEMS microbiology ecology*, 96(4). doi:10.1093/femsec/fiaa032.

Rovira, A.D. (1969) 'PLANT ROOT EXUDATES', *The Botanical review; interpreting botanical progress*, 35(1), p. 35–&.

Sagot, B. *et al.* (2010) 'Osmotically induced synthesis of the dipeptide N-acetylglutaminylglutamine amide is mediated by a new pathway conserved among bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 107(28), pp. 12652–12657.

Saleem, M., Law, A.D. and Moe, L.A. (2016) 'Nicotiana Roots Recruit Rare Rhizosphere Taxa as Major Root-Inhabiting Microbes', *Microbial ecology*, 71(2), pp. 469–472.

Salter, S.J. *et al.* (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC biology*, 12(1), pp. 1–12.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proceedings of the National Academy of Sciences*, 74(12), pp. 5463–5467.

San Millan, A. *et al.* (2015) 'Interactions between horizontally acquired genes create a fitness cost in Pseudomonas aeruginosa', *Nature communications*, 6, p. 6845.

Schleifer, K.H. (2009) 'Classification of Bacteria and Archaea: past, present and future', *Systematic and applied microbiology*, 32(8), pp. 533–542.

Schöner, T.A. *et al.* (2016) 'Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids', *Chembiochem: a European journal of chemical biology*, 17(3), pp. 247–253.

Schüßler, A., Schwarzott, D. and Walker, C. (2001) 'A new fungal phylum, the Glomeromycota: phylogeny and evolution', *Mycological research*, 105(12), pp. 1413–1421.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics* , 30(14), pp. 2068–2069.

Seemann, T. (2015) *Abricate: Mass screening of contigs for antimicrobial and virulence genes*. Github. Available at: https://github.com/tseemann/abricate (Accessed: 6 September 2021).

Sela, I., Wolf, Y.I. and Koonin, E.V. (2016) 'Theory of prokaryotic genome evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), pp. 11399–11407.

Shade, A., Jacques, M.-A. and Barret, M. (2017) 'Ecological patterns of seed microbiome diversity, transmission, and assembly', *Current opinion in microbiology*, 37, pp. 15–22.

Shahzad, R. *et al.* (2018) 'What Is There in Seeds? Vertically Transmitted Endophytic Resources for Sustainable Improvement in Plant Growth', *Frontiers in plant science*, 9, p. 24.

Sheik, C.S. *et al.* (2018) 'Identification and Removal of Contaminant Sequences From Ribosomal Gene Databases: Lessons From the Census of Deep Life', *Frontiers in*

microbiology, 9, p. 840.

Shi, Y. *et al.* (2020) 'Diversity and space-time dynamics of the bacterial communities in cotton (Gossypium hirsutum) rhizosphere soil', *Canadian journal of microbiology*, 66(3), pp. 228–242.

Silby, M.W. *et al.* (2011) 'Pseudomonas genomes: diverse and adaptable', *FEMS microbiology reviews*, 35(4), pp. 652–680.

Simon, L. *et al.* (1993) 'Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants', *Nature*, 363(6424), pp. 67–69.

Simons, A., Alhanout, K. and Duval, R.E. (2020) 'Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria', *Microorganisms*, 8(5). doi:10.3390/microorganisms8050639.

Smith, E.F. (1904) 'Bacterial leaf spot diseases', *Science*, 19, pp. 417–418.

Smith, S.E. and Read, D.J. (2010) *Mycorrhizal Symbiosis*. Academic Press.

Soldan, R. *et al.* (2021) 'The effect of plant domestication on host control of the microbiota', *Communications biology*, 4(1), p. 936.

Somasundaram, S., Fukuzono, S. and Iijima, M. (2008) 'Dynamics of Root Border Cells in Rhizosphere Soil of Zea mays L.: Crushed Cells during Root Penetration, Survival in Soil, and Long Term Soil Compaction Effect', *Plant production science*, 11(4), pp. 440–446.

Somers, E., Vanderleyden, J. and Srinivasan, M. (2004) 'Rhizosphere bacterial signalling: a love parade beneath our feet', *Critical reviews in microbiology*, 30(4), pp. 205–240.

Spiers, A.J., Buckling, A. and Rainey, P.B. (2000) 'The causes of Pseudomonas diversity', *Microbiology*, 146 (Pt 10), pp. 2345–2350.

Steinegger, M., Mirdita, M. and Söding, J. (2019) 'Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold', *Nature methods*, 16(7), pp. 603–606.

Stintzi, A. *et al.* (1993) 'Plant "pathogenesis-related" proteins and their role in defense against pathogens', *Biochimie*, 75(8), pp. 687–706.

Strieker, M., Tanović, A. and Marahiel, M.A. (2010) 'Nonribosomal peptide synthetases: structures and dynamics', *Current opinion in structural biology*, 20(2), pp. 234–240.

Tam, J.P. *et al.* (2015) 'Antimicrobial Peptides from Plants', *Pharmaceuticals* , 8(4), pp. 711–757.

Tecon, R. and Leveau, J.H.J. (2012) 'The mechanics of bacterial cluster formation on

plant leaf surfaces as revealed by bioreporter technology', *Environmental microbiology*, 14(5), pp. 1325–1332.

Tedersoo, L., May, T.W. and Smith, M.E. (2010) 'Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages', *Mycorrhiza*, 20(4), pp. 217–263.

Tettelin, H. *et al.* (2005) 'Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome"', *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), pp. 13950–13955.

Thompson, L.R. *et al.* (2017) 'A communal catalogue reveals Earth's multiscale microbial diversity', *Nature*, 551(7681), pp. 457–463.

Tian, Y. and Gao, L. (2014) 'Bacterial diversity in the rhizosphere of cucumbers grown in soils covering a wide range of cucumber cropping histories and environmental conditions', *Microbial ecology*, 68(4), pp. 794–806.

Tignat-Perrier, R. *et al.* (2019) 'Global airborne microbial communities controlled by surrounding landscapes and wind conditions', *Scientific reports*, 9(1), p. 14441.

Tkacz, A. *et al.* (2020) 'Influence of Plant Fraction, Soil, and Plant Species on Microbiota: a Multikingdom Comparison', *mBio*, 11(1). doi:10.1128/mBio.02785-19.

Toll-Riera, M. *et al.* (2016) 'The Genomic Basis of Evolutionary Innovation in Pseudomonas aeruginosa', *PLoS genetics*, 12(5), p. e1006005.

Toussaint, J.-P. *et al.* (2017) 'Gene Duplication in Pseudomonas aeruginosa Improves Growth on Adenosine', *Journal of bacteriology*, 199(21). doi:10.1128/JB.00261-17.

Tyson, G.W. *et al.* (2004) 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature*, 428(6978), pp. 37–43.

Udaondo, Z. *et al.* (2016) 'Analysis of the core genome and pangenome of Pseudomonas putida', *Environmental microbiology*, 18(10), pp. 3268–3283.

Vacheron, J. *et al.* (2013) 'Plant growth-promoting rhizobacteria and root system functioning', *Frontiers in plant science*, 4, p. 356.

Vandenkoornhuyse, P. *et al.* (2015) 'The importance of the microbiome of the plant holobiont', *The New phytologist*, 206(4), pp. 1196–1206.

Vannette, R.L. (2020) 'The Floral Microbiome: Plant, Pollinator, and Microbial Perspectives', *Annual review of ecology, evolution, and systematics*, 51(1), pp. 363–386.

Van Niel, C.B. and Stanier, R.Y. (1962) 'The Concept of a Bacterium', *Archiv für Mikrobiologie*, 42, pp. 17–35.

Velasco, J.D. (2009) ‘When monophyly is not enough: exclusivity as the key to defining a phylogenetic species concept’, *Biology & philosophy*, 24(4), pp. 473–486.

Vermeer, J. and McCully, M.E. (1982) ‘The rhizosphere in Zea: new insight into its structure and development’, *Planta*, 156(1), pp. 45–61.

Voeller, B.R., Ledbetter, M.C. and Porter, K.R. (1964) ‘CHAPTER 4 - The Plant Cell: Aspects of Its Form and Function’, in Brachet, J. and Mirsky, A.E. (eds) *The Cell*. Academic Press, pp. 245–312.

Vorholt, J.A. (2012) ‘Microbial life in the phyllosphere’, *Nature reviews. Microbiology*, 10(12), pp. 828–840.

Walker, T.S. *et al.* (2004) ‘Pseudomonas aeruginosa-plant root interactions. Pathogenicity, biofilm formation, and root exudation’, *Plant physiology*, 134(1), pp. 320–331.

Walters, W.A. *et al.* (2018) ‘Large-scale replicated field study of maize rhizosphere identifies heritable microbes’, *Proceedings of the National Academy of Sciences of the United States of America*, 115(28), pp. 7368–7373.

Wang, N.R. *et al.* (2021) ‘Commensal Pseudomonas fluorescens protect Arabidopsis from closely-related Pseudomonas pathogens in a colonization-dependent manner’, *bioRxiv*. doi:10.1101/2021.09.02.458786.

Weimann, A. *et al.* (2016) ‘From Genomes to Phenotypes: Traitar, the Microbial Trait Analyzer’, *mSystems*, 1(6). doi:10.1128/mSystems.00101-16.

Wen, F. *et al.* (2009) ‘Extracellular DNA is required for root tip resistance to fungal infection’, *Plant physiology*, 151(2), pp. 820–829.

Whipps, J.M. *et al.* (2008) ‘Phyllosphere microbiology with special reference to diversity and plant genotype’, *Journal of applied microbiology*, 105(6), pp. 1744–1755.

Wick, R.R. *et al.* (2015) ‘Bandage: interactive visualization of de novo genome assemblies’, *Bioinformatics* , 31(20), pp. 3350–3352.

Wick, R.R. *et al.* (2017) ‘Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads’, *PLoS computational biology*, 13(6), p. e1005595.

Wick, R.R., Judd, L.M. and Holt, K.E. (2019) ‘Performance of neural network basecalling tools for Oxford Nanopore sequencing’, *Genome biology*, 20(1), p. 129.

Wiedenbeck, J. and Cohan, F.M. (2011) ‘Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches’, *FEMS microbiology reviews*, 35(5), pp. 957–976.

Winsor, G.L. *et al.* (2016) ‘Enhanced annotations and features for comparing thousands

of Pseudomonas genomes in the Pseudomonas genome database', *Nucleic acids research*, 44(D1), pp. D646–53.

Woese, C.R., Blanz, P. and Hahn, C.M. (1984) 'What isn't a Pseudomonad: The Importance of Nomenclature in Bacterial Classification', *Systematic and applied microbiology*, 5(2), pp. 179–195.

Woese, C.R. and Fox, G.E. (1977) 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms', *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), pp. 5088–5090.

Wood, D.E., Lu, J. and Langmead, B. (2019) 'Improved metagenomic analysis with Kraken 2', *Genome biology*, 20(1), p. 257.

Wright, E.S. and Baum, D.A. (2018) 'Exclusivity offers a sound yet practical species criterion for bacteria despite abundant gene flow', *BMC genomics*, 19(1), p. 724.

Yamada, Y. *et al.* (2015) 'Terpene synthases are widely distributed in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 112(3), pp. 857–862.

Yang, M. *et al.* (2020) 'Exploring the Pathogenicity of Pseudomonas brassicacearum Q8r1-96 and Other Strains of the Pseudomonas fluorescens Complex on Tomato', *Plant disease*, 104(4), pp. 1026–1031.

Yu, K. *et al.* (2019) 'Rhizosphere-Associated Pseudomonas Suppress Local Root Immune Responses by Gluconic Acid-Mediated Lowering of Environmental pH', *Current biology: CB*, 29(22), pp. 3913–3920.e4.

Zaheer, R. *et al.* (2018) 'Impact of sequencing depth on the characterization of the microbiome and resistome', *Scientific reports*, 8(1), p. 5890.

Zarraonaindia, I. *et al.* (2015) 'The soil microbiome influences grapevine-associated microbiota', *mBio*, 6(2). doi:10.1128/mBio.02527-14.

Zhu, J. *et al.* (2021) 'Over 50,000 Metagenomically Assembled Draft Genomes for the Human Oral Microbiome Reveal New Taxa', *Genomics, proteomics & bioinformatics* [Preprint]. doi:10.1016/j.gpb.2021.05.001.

Zipfel, C. and Felix, G. (2005) 'Plants and animals: a different taste for microbes?', *Current opinion in plant biology*, 8(4), pp. 353–360.