Chapter V

Comparative Genomics of a Novel Human-Adapted Subclade of *P. fluorescens* Species Complex Bacteria That Contain a Yop/Ysc-like Type- III Secretion System

BACKGROUND

Pseudomonas fluorescens is traditionally considered a microbe of plant, soil and rhizosphere habitats (1-3). In a previous study, we reported the first sequenced strains of *P. fluorescens* bacteria from human clinical samples (4). These clinical *P. fluorescens* strains were phylogenetically classified within subclade III of the species complex, alongside strains isolated from environmental samples. Here we report fifteen additional *P. fluorescens* strains isolated from human clinical samples. These fifteen clinical strains fall into three separate subclades within the *P. fluorescens* species complex; previously reported subclades I and II and a novel subclade IV. The clinical *P. fluorescens* strains in subclades I, II and III show few genomic differences from the environmentally-derived strains within their corresponding subclades (4). In contrast, the novel subclade IV only contains newly sequenced, clinically-isolated *P. fluorescens* strains that are genomically and phenotypically divergent from all other members of the *P. fluorescens* species complex.

RESULTS AND DISCUSSION

Collection of *P. fluorescens* isolates.

The newly sequenced clinical strains of *P. fluorescens* were isolated over an eleven-year period from respiratory samples collected across the United States. The fifteen clinical strains of *P. fluorescens* were isolated between February 1999 and April 2010 from nine different hospitals across the United States, each from a separate patient suffering from a respiratory illness (**Figure V.1**). Fourteen were isolated from patients with cystic fibrosis (CF), and of these, thirteen were isolated form CF sputum and one from a CF throat swab. The one isolate that wasn't taken from a CF patient was isolated form a non-CF tracheal aspirate. Since the isolates were collected from separate patients across the United States during more than a decade long period, it is highly unlikely that the genomes presented here represent multiple isolations of the same bacteria strain.

Phenotypic and growth characteristics

Phylogenetic analysis using a MLSA approach divides the fifteen newly sequenced clinical strains into three subclades (I, II and IV) (**Figure III.1**), and we compared the phenotypic and growth characteristics between of the clinical strains from different subclades.

All the newly sequenced clinical strains share known physical characteristics of *P. fluorescens*-species, such as being motile, gram-negative, coccobacillus, lactose-fermentation negative, citrate growth positive, and fluorescent under ultra violet (UV) light (**Figure V.2**). Of the growth substrates tested, the clinical strains were only able to utilize citrate for growth, with the exception of AU12597, which was also able to utilize arabinose (**Figure V.2).** Other members of the *P. fluorescens* species complex, such as subclade III strains AU6026 and SBW256 and subclade I strain Pf-5, are able to utilize arabinose (4). The newly sequenced clinical strains share similar biochemical characteristics as previously sequenced *P. fluorescens* strains.

While members of subclade I, II and IV are similar in the majority of phenotypic properties analyzed, they differ in their ability to grow at temperatures above room temperature (RT; ~22°C). We previously reported that clinical strains in subclade III are able to grow at RT and 32°C and two strains (AU11518 and AU14440) are also able to grow at 37°C (4). While the newly sequenced clinical strains in subclade I and II reported here are unable to grow above RT (~27°C), the clinical strains in subclade IV grow readily at RT, 32°C and 37°C (**Figure V.2**). The temperature gradient found in the respiratory tract is not uniform and ranges from the inspired air in the nares (near or at room temperature) to that found at the core body temperature in the alveoli (near or at 37°C) (5). Based on the *in vitro* temperature permissive range of the newly sequenced clinical strains, those within subclade I and II could not persist in the higher temperature environment of the lower airways while strains within the newly formed subclade IV could persist at temperatures along the entire temperature gradient of the respiratory tract.

Genome assembly

Whole genomic DNA was isolated from all fifteen clinical *P. fluorescens* strains and prepared for paired-end read sequencing on the Illumina HiSeq platform (see materials and methods). Outputted sequences were de novo assembled into contigs and scaffolds using DNAstar's SeqMan NGen vs 12.0.0.22. The strains were divided into subclades based on a phylogenic approach described in the **Chapter III**. Genomic coverage ranged between 47X (AU5633) and 80X (AU12597). **Figure V.3** displays assembly statistics of each of the newly sequenced strains.

Average Nucleotide Identity

In addition to the MLSA-based approach presented in **Figure III.3**, we also calculated the shared average nucleotide identity (ANI) to investigate the phylogenetic relationship between strains. (**Appendix Figure V.1**). A similar approach was utilized to calculate the phylogenetic relationship of clinical subclade III *P. fluorescens* strains (4). ANI infers taxonomic relationship by calculating the percentage of shared nucleotide between two bacterial genomes. The intraspecies ANI cut-off is typically 95% (6). Newly sequenced clinical strain AU5633 shares the highest percent ANI to previously sequenced *P. fluorescens* subclade II strains R124 (91.72%) and Pf0-1 (87.61%) and newly sequenced strain AU11114 (87.11%) (**Appendix Figure V.1**). The previously sequenced strains that shared highest ANI to AU11114 are also R124 and Pf0-1 (**Appendix Figure V.1**). While these numbers are below the official cut-off for species delineation, they are within the range of shared ANI that we reported between members of subclade III (4). Based on shared ANI, strain AU11114 is more closely related to strains in subclade II and thus we will refer to AU11114 as a subclade II strain in all further analysis. The newly sequenced clinical strains AU11706, AU13852 and AU20219 all share their highest percent ANI with the previously sequenced subclade I strains *P. protegens* Pf-5 and CHA0 (89.8% and higher). For strain AU11706, shared ANI between other subclade I strains is 89.8 – 90.1%, while for strains AU13852 and AU20219 the range is between 97.03 and 99.18%. This agrees with the MSLA tree, which places AU11706 on a separate branch slightly away from the other members of subclade I (**Figure III.3**). For AU5633, AU11114, AU11706, AU13852 and AU20219, ANI analysis agrees with the phylogenetic placement of these strains in subclade I and II.

The shared ANI between the strains in the newly formed subclade IV is higher than that of any other subclade within the *P. fluorescens* species complex. All members of the newly formed subclade IV share an ANI of 97.83% or more (**Appendix Figure V.1**). Subclade IV strains AU2390 and AU12597 show the lowest range of shared ANI between other members of

subclade IV (97.83 – 98.18% and 98.18 – 98.42%, respectively), but this percent ANI is still well above the intraspecies cut-off of 95%. The shared percent ANI between subclade IV strains AU7350, AU10414, AU11122, AU11164, AU11235 and AU11336 is at least 99.99% while between AU1044 and AU12644 it is above 99% (**Appendix Figure V.1**). These strains also group very tightly together on the MLSA phylogenetic tree; AU10414, AU11122, AU11164 and AU11235 are all located on a single node, with AU7350 being a single branch away from this node (**Figure III.3**). These eight subclade IV strains with shared ANI of 99% and above were isolated from seven different treatment centers across the United States over an eight year time span (**Figure V.1, Figure V.4**).

AU1044 was isolated on February 21st, 1999; AU7350 was isolated on May 18th, 2004 from a treatment center in Hartford, CT; AU10414 was on December 15th, 2005 from a treatment center in Boston, MA; AU11122 was isolated on May 4th, 2006 in Palo Alto, CA; AU11164 on May 9th, 2006 in St. Louis, MO; AU11235 on May 18th, 2006 in Ann Arbor, MI and AU11336 on June 6th, 2006 from Cincinnati, OH; and AU12644 was isolated December 19th, 2005 (**Figure V.4).** The large temporal and physical range from which the subclade IV strains were isolated makes it unlikely that they represent an outbreak of a single *P. fluorescens* strain. The only other members of the *P. fluorescens* species complex that have a shared ANI near that of the subclade IV strains are the newly sequenced subclade I strains AU13852 and AU20219, which have a shared ANI of 99%. However, the shared ANI between AU13852 and AU20219 with other strains within subclade I is below 97% (**Appendix Figure V.1**). Overall, the shared percent ANI across members of the subclade IV is the highest of any subclade within the *P. fluorescens* species complex, which agrees with the tight clustering of these strains within the MLSA-based phylogenetic tree (**Figures III.1** and **III.3**).

Genomic Features

Genome size, percent G+C content and # of coding sequences within each newly sequenced genome are more similar between members of the same subclade than between members of different subclades (**Figure V.5**). Due to the draft nature of the genomes and difficulty of sequencing and assembling across the 16S rRNA gene, the final number of resolved 16S rRNA copies in each strain varied between one and seven (**Figure V.5**). A similar issue with resolving 16S rRNA gene copies in draft genomes was seen in **Chapter IV** (4). The number of coding sequences within subclade IV strains ranges between 4309 and 4441. Within subclade II the range is 5448 and 5722 and within subclade I the range is 6125 and 6574. The number of coding sequences within each subclade correlates with genomic size. Subclade IV strains have the smallest average genomic size (4.89, range 4.78 – 5.4 Mbp), subclade I strains have the largest average genomic size (7.0, range 6.68 – 7.27 Mbp) and the average genomic size of subclade II strains fall in between subclade IV and I (6.32 Mbp, range 6.08 – 6.6 Mbp) (**Figure V.5**). A more detailed analysis of genomic size is shown in **Figure V.6.** While the newly sequenced clinical strains in subclades I and II have genomes of similar size to that seen in previously sequenced *P. fluorescens* strains (**Appendix Figure V.1** and **Figure V.6**), the genome size of subclade IV strains is significantly different from any member of the *P. fluorescens* species complex (**Figure V.6** and **Appendix Figure V.2**). If the one subclade IV outlier is removed from the analysis (AU2390 with genome size of 5.4 Mbp), the average genome size of subclade IV clinical strains is nearly 1.5 Mbp smaller than the next smallest *P. fluorescens* species complex genome (**Figure I**). The genome size of the subclade IV strains is statistically different than the genome size of all the other members of the *P. fluorescens* species complex (**Appendix Figure V.2**). One of the marked signs of adaptation to a human host environment is gene loss (7). Bacteria that thrive in diverse environmental habitats typically have larger genomes than bacteria that are confined to the habitat of a single, mammalian host (8-13). The clinically-derived strains of subclade IV have genomes between 1 and 2.5 Mbp smaller than other members of the *P. fluorescens* species complex (**Figure V.6** and **Appendix**

Figure V.2) suggesting that the subclade IV strains have lost genomic information in response to adaptation to a human host environment.

The percent G+C content within subclade IV genomes is also significantly different then the percent G+C content found within any other *P. fluorescens* subclade or *Pseudomonas* species group (**Figure V.7** and **Appendix Figure V.3**). The average GC% in subclade IV genomes is 58.84% (range 58.3 – 58.8%); in subclade II it is 60.18% (range 58.8 – 61%) and in subclade I it is 63.74% (range 62.8 – 63.4%). As shown in **Figure V.7** and **Appendix Figure V.3**, the GC% of strains within subclade IV is significantly different than the GC% of strains within subclades I, II and III. A reduction in genomic G+C content is also associated with adaptation to a human-host only lifestyle (14). The reduced size and G+C content of subclade IV strain genomes could both be explained by a shift from an environmentally-based habitat to that of a human host.

GC Islands

To further investigate G+C related genomic differences between subclade IV strains and other members of the *P. fluorescens* species complex, we performed GC content scans across the newly sequenced genomes in subclades I, II and IV. Additionally, the web-based software GC-Profile (http://tubic.tju.edu.cn/GC-Profile/) was used for the GC scans (15). This tool takes inputted genomes and segments them based on parameters provided by the user. The GC content along each segment is determined. Areas where the GC content varies dramatically from the genome average are visualized as 'dips' in the GC scan graphs (**Figure V.8**) and are referred to as GC-Islands (GCI). The GCIs within each genome are numbered as GCI1, GCI2, etc in **Figure V.8.** The nucleotide sequence from each GC was then used to screen NBCI's microbial nucleotide database using the megablast tool (results displayed in **Appendix Figure V.4)**. The number of GCIs in the newly sequenced clinical strains ranges from zero in AU5633

(subclade II, **Figure V.8**) and five in AU11122 (subclade IV, **Figure V.8**). The clinical strains in subclade IV contain more GCIs (average of 2 GCI/genome) than found in subclade I and II (**Figure V.8** and **Lb**) (average of 1 GCI/genome).

The blast results of the GCIs from newly sequenced subclades I, II and IV strains reveal that some of these GC Islands represent cases of recent horizontal gene transfer (**Appendix Figure V.4**). The GCI1 and GCI2 from AU1044 retrieve best matches to non-*Pseudomonas spp.* bacteria (top hits *Xanthophyllomyces dendrorhous* and *Enterobacter cloacae,* respectively). The top hit of GCI1 from AU11122 is an *Enterobacteria* phage phiX174. The top hit from GCI1 from AU12597 is similar to that of AU1044, *Xanthophyllomyces dendrorhous*, as are the GCI1s from AU11706 and AU13852. The hit to *Xanthophyllomyces dendrorhous* is particularly interesting, as this is a yeast most widely studied for its production of the carotenoid pigment astaxanthin (16)**.** Both the GCIs from AU11114 return non-*Pseudomonas* results; GCI1 returns *Streptococcus p.* VT 162 as a top hit and GCI2 returns only two hits, *Pseudomonas sp.* VLB120 and *P. putida* W619. While the nucleotide sequences of all the other GCIs return top hits within the *Pseudomonas* genus, a few return hits from outside the *P. fluorescens*-species complex and may represent interspecies horizontal gene transfer. GCI1 from AU10414, GCI2 from AU11122, GCI2 from AU11164, GCI1 from AU11336 and GCI2 from AU12644 all return nearly identical top hits that include *P. chlororaphis* strain UFB2, *Pseudomonas cichorii* JCB1 and multiple *P. syringae* strains (**Appendix Figure V.4**). The nucleotide sequence of GCI3 form AU11122 returns only two hits, *Pseudomonas cichorii* JCB1 and a *Cupriavidus metallidurans* CH34 megaplasmid. *Pseudomonas cichorii* is pathogen of soybean, lettuce and tomato plants (17). *Cupriavidus metallidurans* CH34 is β-proteobacterium that grows well in low concentrations of heavy metals and the determinants of heavy metal resistance are found on its megaplasmids (18). The megaplasmids of *Cupriavidus metallidurans* CH34 contain multiple genomic islands, one that is similar to the PAGI-2C similar to the one found in *P. aeruginosa* clone C strains

isolated from CF patients (19). Blastn of the GCI1 from AU11164 retrieves *Pseudomonas cichorii* JCB1, *P. entomophila* L48 and *P. putida* HB3267 as top matches. GC Island analysis reveals genomic regions within the newly sequenced *P. fluorescens* strains that likely represent regions of horizontal gene transfer from bacteria both within and without the *Pseudomonas* genus. Multiple GCIs within members of subclade IV are genomically similar to the bacteria *Pseudomonas cichorii* JCB1 while no GCIs from subclade I and II clinical strains show high genomic similarity to this bacteria. Horizontal gene transfer is one of the primary methods by which bacteria evolve to thrive in new and changing environments. The high number of GC Islands within the subclade IV strains suggests that horizontal gene transfer occurs frequently within these bacteria and represent evolutionary methods by which they could adapt to a human host.

Pan, accessory, soft core and core genomes

We next calculated the pan, accessory, and core genomes of the subclade IV strains in order to investigate the number of protein-encoding genes shared within members of this subclade. The pan, accessory and core genomes were calculated by measuring the number of clusters of orthologous groups (COGs) of proteins shared between subclade IV genomes. COGs represent sets of orthologous proteins that are found in at least three evolutionarily distant species and represent proteins that have evolved from a single ancestral gene (20, 21). The core genome contains COGs found in 100% of the genomes analyzed, the pan genome contains all the COGs found across all considered genomes and the accessory genome contains all the COGs within the pan genome that are not also part of the core. The pan genome and its compartments were calculated using the COGtriangle in the program GET_HOMOLGOUES (22, 23). The pan, accessory and core genomes of the subclade IV genomes are shown in **Figure V.9.**

Major differences exist between the pan, accessory, soft core and core genomes of the subclade IV strains and the corresponding genome compartments of other members of the *P. fluorescens* species complex. The average number of COGs per subclade IV genome is 4442, with a range of 4309 to 4925 COGs. This is significantly smaller than what we previously reported for subclade III strains, where the average number of COGs per genome was 5592, with a range of 5332 – 6123 (**Chapter IV)**(4). While the pan genome of eleven clinical and environmental subclade III genomes is 11,795 COGs, the pan genome of the ten clinical subclade IV strains is nearly half that, at only 6046 COGs (4). This is significantly smaller than the pan genomes found in subclade I and II strains (**Figure VII.1**). Subclade I, which consist of three clinical strains and two environmental strains (**Figure VII.1**), has a pan genome of 8666 COGs (**Figure VII.1**). Subclade II, which contains two clinical strains and four environmental strains, has a pan genome of 10095 COGs. Interestingly, while there is a large difference in pan genome size between the subclades, there is not a large difference in core genome size different *P. fluorescens* species complex subclades: subclade I has a core genome of 4802 COGs, subclade II has a core genome of 3505 COGs, subclade III has a core genome of 3612 COGs and subclade IV has a core genome of 3692 COGs (**Figures V.9** and **VII.1**). Therefore, the large difference in pan genome size between the subclades is not a result of difference in core genome size but differences in accessory genome size. Subclade I strains has an accessory genome of 3864 COGs, subclade II strains have an accessory genome of 6590 COGS and subclade III strains have an accessory genome of 8183 COGS. In sharp contrast, the subclade IV strains only have an accessory genome of 2354 COGs. The size of the accessory genome in subclades I, II and III is highly correlated with the number of genomes considered, a trend that is also seen when the whole *P. fluorescens* species complex and all fully sequenced *Pseudomonas spp.* genomes deposited on NCBI's database are considered (**Figure VII.1**). Pan, accessory and core genome analysis reveals that subclade IV strains share

a significantly smaller pan and accessory genome than the strains in any of the other *P. fluorescens* subclades.

To further visualize genome-wide differences between the subclade IV strains and other members of the *Pseudomonas* genus, we utilized a principal component analysis (PCA) to cluster the strains based on homology between shared gene clusters within their corresponding genomes (24). In the PCA displayed in **Figure V.10**, the analyzed genomes form clusters on the multidimensional space that relates with their taxonomic clusters within the MLSA-based phylogenetic tree (**Figure III.3**). Strains within subclade I, II and III cluster near each other on the PCA space (**Figure V.10),** suggesting they share a high number of homologous gene clusters. Subclade IV strains, however, form a completely separate cluster away from the other *P. fluorescens* subclades (**Figure V.10**). Subclade IV strains AU7350, AU10414, AU11122, AU11164, AU11235 and AU11336 cluster directly on top of each other on the PCA space, suggesting they share nearly identical gene clusters. Subclade IV strains AU12644, AU2390, AU1044 and AU12597 form a second, very tight cluster (**Figure V.10**) right next to the other subclade IV cluster. The tight clustering of the subclade IV strains on the PCA reveals that they are highly homologous to each other in gene content, while the distant spacing between the subclade IV cluster and the other subclades suggests they are highly divergent in gene content from other members of the *P. fluorescens* species complex.

The *pfi*T/I2 Region

We next compared the nucleotide and amino acid sequences of the *pfi*T/I2 region across the *P. fluorescens* subclades. *pfi*T/I2 is a region found in *P. fluorescens*-species complex bacteria that encodes for a novel superantigen and is associated with various human autoimmune and enteric diseases (25). I2 was discovered through a selective screen for peptides associated with Crohn's Disease (CD) lesional tissue but absent from CD healthy

tissue (26). The 'I2' peptide was later identified as originating from a *P. fluorescens* bacteria and genome walking identified a 208-amino acid open reading frame between residues 49 and 148 of the I2 protein. This open-reading frame is termed *pfi*T (27). Within the *pfi*T/I2 literature, the names of these two slightly different expressed proteins are often used interchangeably in the literature. Here, we refer to *pfi*T as the expressed 208-amino acid open reading frame within the I2 region. To compare the *pfi*T region within the newly sequenced clinical strains, we screened each genome with the published *pfi*T amino acid sequence using a local blastp search (26). The top results from each genome were then aligned using the published *pfi*T sequence as a reference (**Figure V.11**). While all members of subclade I and II were found to contain a *pfi*T region, this region was absent in every member of subclade IV. We previously published that every strain within subclade III also contains a *pfi*T region (**Chapter IV**) (4). The *pfi*T region in subclade I and II strains differs from the published *pfi*T region at multiple positions along the peptide. Certain amino acids were altered in every strain analyzed (histidine to glutamine at position four) while certain amino acid alterations are specific to a particular subclade. The serine at position 90 in the reference *pfi*T peptide is changed to a glutamate in subclade I strains and an alanine in subclade II strains. Overall, subclade I and subclade II strains are more divergent at the amino acid level along the *pfi*T region than that of subclade III strains (4). Based on the analysis of the amino acid sequence along the *pfi*T region, the original *pfi*T amino acid sequence was likely derived from a *P. fluorescens* bacteria more genomically related to members of subclade III. Of the *P. fluorescens* species complex strains analyzed thus far, only strains within subclade IV are missing the *pfi*T region entirely.

We also compared the nucleotide sequence of the I2 region among subclade I, II and IV strains. Using the published I2 primer set we searched each genome for the nucleotide region that corresponds to the I2 peptide and aligned these nucleotide sequences to look for differences from the consensus (27) (**Figure V.12**). The I2 primers were unable to align to any

region of the subclade IV genomes. The I2 nucleotide sequences of members of *P. fluorescens* subclades I, II and III, as well as representative *P. aeruginosa* strains, are displayed in **Figure V.12** with only differences from the consensus sequence colored. Unlike what was seen in the *pfi*T amino acid comparison, the majority of nucleotide differences along the I2 sequence are not subclade specific. An exception is the guanine to thymine switch at position 32 that is specific to subclade II strains. While the *pfi*T and I2 regions are present in all *P. fluorescens* subclade I, II and III strains analyzed, they are absent in subclade IV strains.

Secretion Systems

Gene clusters involved in the formation of secretion systems in *Pseudomonas spp.* strains were annotated and compared across strains in subclade I, II and IV. A similar analysis was published on subclade III strains (**Chapter IV**) (4). No major differences were seen between the subclades in genes involved in type-II, type-IV, type-VI secretion systems (T2SS, T4SS, T6SS) or the widespread colonization island secretion system. The T2SS is the general secretion pathway commonly found among gamma-proteobacteria and the *Pseudomonas* genus (28-30). The Tad locus, or widespread colonization island (WCI), is a secretion system involved in bacterial biofilm formation and colonization of hosts (31). The type IV secretion system (T4SS) forms a pili/fimbria on the surface of bacteria that is important for adhesion, protease secretion and motility (32). The type IV secretion system (T6SS) is the first secretion system discovered in bacteria that is involved in bacteria-bacteria competition (33-36). Every strain analyzed here contained as similar assortment of gene involved in the expression of the T2SS, T4SS, T6SS and the WCI secretion systems.

While there were no large differences between the subclades in the composition of gene homologues involved in T2SS, WCI, T4SS and T6SS, the composition of gene homologues involved in type III secretion varied dramatically between strains of subclade IV and strains of

the other three *P. fluorescens* subclades. Type-III secretion systems are macromolecular, needle-like structures bacteria use to inject proteins, called effectors, into target host cells (37). There are multiple T3SS families and they differ in genetic makeup, the effector cells they deliver and the physical structure of the needle complex. The Hrp1 type III secretion system (T3SS), so named because it can trigger a hypersensitive response in plants, is the most prominent T3SS found across members of the *P. fluorescens* species complex (2) (38) (39-41). The needle-complex of the Hrp1 T3SS is longer than needle complexes of T3SSs in other families, which is likely an adaptation to the thick cellular wall that of plant host cells (42). Previous reports have found the genetic components for a Hrp1 T3SS in *P. fluorescens* environmental strains from subclade III (SBW25, BG33R, A506, SS101) and subclade II (Q8r1-96 and Q2-87). Subclade II strain Q2-87 also possessed a second T3SS gene cluster that is related to the Inv/Mxi/Spa-family of T3SSs (2). Other members of subclade II, Pf0-1 and the newly sequenced clinical subclade II strains AU5633 and AU11114, all lack a Hrp1 T3SS (2). Both clinical and environmental strains in subclade I are also missing the Hrp1 T3SS related gene cluster (2). The only Hrp1 gene homologue found in subclade IV strains is HopPmaJ, an effector protein. However, the genes involved in producing the Hrp1 needle complex, *hrpABDGJPQTV*, are completely absent in the subclade IV strains. We previously reported that clinical subclade III strains contain a similar assortment of Hrp1-related gene homologues to environmental subclade III strains (**Chapter IV**) (4). The Hrp1 T3SS is mostly confined to members of *P. fluorescens* subclade III, with the exception of two members of subclade II, Q8r1-96 and Q2-87.

While subclade IV strains lack the Hrp1 T3SS related gene cluster, every subclade IV strain contain a Yop/Ysc-like T3SS gene cluster (**Appendix Figure V.5**). The Ysc/Yop-family was first described in *Yersinia* and delivers Yop (*Yersinia* outer protein) effectors into human host cells (3, 43, 44). The Ysc/Yop-family is found in multiple known human pathogens, such as

Bordetella (Bsc), *Rhizobium*, *Chlamydia* and *Pseudomonas aeruginosa*. In *P. aeruginosa* the genes involved in the Ysc/Yop-family T3SS are given the Psc-prefix to indicate they are from a *Pseudomonas* bacteria. In *P. aeruginosa* the Yop/Ysc-like T3SS is an important virulence factor. While not required for infection, animal models of acute pneumonia, keratitis, bacteremia, peritonitis, burn infections, and gut-derived sepsis have shown that a functional T3SS enhances disease severity (45-53). Human infections with T3SS-positive *P. aeruginosa* strains are highly associated with increased mortality (46, 54-57). Subclade IV strains contain 34 genes annotated as belonging to the Yop/Ysc-like T3SS. This is a number similar to that found in *P. aeruginosa* strains, with the exception being PA7, which lacks a T3SS (**Figure V.13**) (45). The number of Yop/Ysc-like gene homologues found in subclade IV strains is statistically different than the number found in any other *P. fluorescens* subclade but is statistically the same as those found in *P. aeruginosa* strains (**Figure V.13, Appendix Figure V.6**). When the presence or absence of each gene homologue is compared between *P. fluorescens* subclade IV strains and the *P. aeruginosa* strains that contain a Yop/Ysc -like T3SS, *pscE* is the only gene homologue not found in subclade IV strains (**Appendix Figure V.5**). PscE is a chaperone protein for PscF, a structural subunit of the Yop/Ysc-like T3SS needle complex (45, 58). While PscE is important for targeting PscF to the secretion apparatus, it is possible that another chaperone protein is serving this function in the subclade IV strains (58, 59). Another possibility is that the subclade IV strains contain an equivalent gene to *pscE* whose homology is too low to be properly assigned by the annotation software, as has been the case with T3SS genes in other bacterial strains (60). Even considering the absence of a homologous *pscE* gene, the initial annotation of the Yop/Ysc -like T3SS in subclade IV strains reveals a gene cluster that is nearly identical to the one found in *P. aeruginosa.*

ExoU and SpcU

In addition to the absence of a homolog to *pscE*, the subclade IV and the *P. aeruginosa* strains analyzed in this study also differ in the assortment of Yop/Ysc -like associated effector proteins found in each strain. *P. aeruginosa* strains are known to produce four main effector proteins, ExoS, ExoT, ExoU and ExoY (45). Each effector protein involved in Yop/Ysc-like type-III secretion requires a chaperone protein to properly transverse the T3SS needle complex (45). SpcS is the chaperone protein for the effectors ExoS and ExoT; SpcU is the effector protein for ExoU and the effector protein for ExoY is currently unknown (45). Each effector protein has a slightly different effect on host cells. Both ExoS and ExoT are bifunctional molecules that have GTPase-activating (GAP) activity and ADP-ribosyltransferase activities that transiently disrupts the actin cytoskeleton of the host cell, leading to cell rounding and decreased phagocytosis of the invading bacteria (45) (61-63). ExoY is an adenylate cyclase that, when secreted into a host cell, can lead to disruption of the actin cytoskeleton, inhibition of bacterial uptake and increased endothelial permeability (45, 64-67). ExoU contains a phospholipase A (PLA2) activity that, once inside the host cell, induces rapid loss of cell membrane integrity and host cell necrosis (68, 69) (48) (47). Of the four main effector proteins, ExoU is considered the most pathogenic; ExoU is only expressed by ~30% of clinical *P. aeruginosa* strains but is expressed by 90% of clinical *P. aeruginosa* strains that cause acute disease (45, 70, 71). *P. aeruginosa* PAO1 expresses ExoS, ExoT and the corresponding chaperone SpcS; *P. aeruginosa* PA14 expresses ExoT, ExoU and SpcU; *P. aeruginosa* 2192 expresses ExoY; and *P. aeruginosa* LESB58 expresses ExoT and SpcS. Of these four main effector/chaperone combinations involved in Yop/Ysc-like T3SS, subclade IV strains contain gene homologues to ExoU and the chaperone SpcU only. Annotation of the newly sequenced clinical *P. fluorescens* genomes revealed homologues to ExoU and SpcU in each of the subclade IV strains.

To further verify the *in silico* gene annotation, we created a primer set that specifically targets the ExoU genetic sequence in the subclade IV strains and verified the existence of an

ExoU gene in each subclade IV strain through PCR (**Appendix Figure V.7** and **Figure V.14**). The ExoU-specific primers were developed by aligning the ExoU genetic sequence from each subclade IV along with the ExoU sequence from select *P. aeruginosa* strains (**Appendix Figure V.7**). The ExoU-specific primers successfully amplified the ExoU gene fragment from each subclade IV strain and did not cross react with any region in the other clinical *P. fluorescens* strains, environmental *P. fluorescens* strains, clinical *P. aeruginosa* strains or reference *P. aeruginosa* strain PAO1 (**Figure V.14**).

To compare the predicted amino acid sequence of the ExoU homologs from subclade IV strains to ExoU proteins from other *Pseudomonas spp.* strains and other bacterial species, we created a phylogenetic tree inferred on the similarity of the predicted amino acid sequence from various ExoU gene homologues (**Figure V.15**). Proteins homologous to ExoU are found in many bacterial species, including some *Pseudomonas spp.* strains that do not have a corresponding Yop/Ysc-like T3SS gene cluster. Based on the predicted amino acid sequence of their ExoU gene homologs, the subclade IV strains group tightly together on a branch where the nearest next relatives are *P. aeruginosa* strains PA14, MH27, PA99 and PA103 (**Figure V.15**). A representative ExoU amino acid sequence from AU11122 was used to screen NCBI's non-redundant GenBank CDs translations (**Appendix Figure V.8**). The top ten results are presented in **Appendix Figure V.8a**, the top ten results from the *P. fluorescens* species complex are presented in **Appendix Figure V.8b** and **Appendix Figure V.8c** shows the results of a protein to protein alignment against the reference strains included in the tree in **Figure V.15.** The ExoU protein from AU11122 has highest sequence identity and query coverage (62% and 97% respectively) with the ExoU protein from *P. aeruginosa* strains (**Appendix Figure V.8a**). In comparison, the highest match of the ExoU protein from AU11122 to another ExoU protein from the *P. fluorescens* species complex is 48% sequence identity and 90% query coverage (**Appendix Figure V.8b**). Direct protein-to-protein alignment also reveals that the ExoU protein

from AU11122 is most similar at the amino acid level to ExoU proteins from *P. aeruginosa* (**Appendix Figure V.8c**). Analysis of the predicted amino acid sequence of ExoU from subclade IV strains reveals that the highest homology is to ExoU proteins from *P. aeruginosa* strains.

The functionality of ExoU within the host environment requires its corresponding chaperone, SpcU, and all the subclade IV strains also have a homologue to this Yop/Ysc-like T3SS gene. To investigate the similarity of the SpcU homolog in subclade IV strains to SpcU from *P. aeruginosa,* we created an alignment of the predicted protein sequence of the SpcU homolog from the subclade IV strains alongside the amino acid sequence of SpcU from select *P. aeruginosa* strains and a *Photorhabdus asymbiotica* strain (**Figure V.16**). In this alignment, only amino acids that differ from the consensus sequence are labeled and colored. The predicted amino acid SpcU sequence is highly consistent across subclade IV strains, with only strain AU2390 showing any differences in amino acid composition (**Figure V.16**). A similar tree analysis and blastp analysis that was performed on the predicted amino acid sequence of subclade IV ExoU homologues was also performed for subclade IV SpcU homologues (**Figure V.15** and **Appendix Figure V.9**). While ExoU homologues are found in other members of *P. fluorescens*-species complex, SpcU homologues are only found in the newly sequenced *P. fluorescens* subclade IV strains and *P. aeruginosa* strains. A blastp screen using the predicted amino acid sequence of the SpcU homologue form AU11122 returned matches to multiple *P. aeruginosa* strains, two unknown *Pseudomonas spp.*, *Photorhabdus asymbiotica* and two *aeromonas* strains (**Appendix Figure V.9**). Nucleotide and amino acid analysis of the ExoU and SpcU homologues from subclade IV strains reveal that these two predicted Yop/Ysc-like T3SS genes are most similar to the ExoU and SpcU homologs in *P. aeruginosa* strain.

Based on the annotated homologues, we built a model of the predicted Yop/Ysc-like T3SS found in the newly sequenced subclade IV strains (**Figure V.17**). This model is built

based on the known functionality of the gene homologues of this T3SS family in *P. aeruginosa* strains (45). As the image shows, the subclade IV strains are predicted to have all the genes necessary to produce a functional Yop/Ysc-like needle complex and homologues to the effector/chaperone combo ExoU and SpcU. This is the first time that a Yop/Ysc-like T3SS has been reported in any *P. fluorescens*-species complex strain and, in fact, any *Pseudomonas spp*. other than *P. aeruginosa.*

CONCLUSIONS

Members of the newly formed subclade IV within the *P. fluorescens* species complex are highly genetically divergent from strains within the other subclades. The newly sequenced subclade IV strains have genomes that are between 1 and 2.5 Mbps smaller than other members of the *P. fluorescens* species complex (**Figure V.6** and **Appendix Figure V.2**). In addition to a reduction in genome size, the genomes of the subclade IV strains have lower overall G+C content, a feature associated with gene loss during evolution to a mammalian host (14). Multi-locus sequence analysis placed the newly sequenced subclade IV strains on a phylogenetic branch that does not include any *P. fluorescens* strains isolated from environmental habitats (**Chapter III, Figure III.3**). Average nucleotide identity between the subclade IV strains is above 97%, with eight of the ten subclade IV strains sharing 99% or a higher ANI. This is in stark contrast to the average shared ANI between members of the other subclades; where the average shared ANI between members of subclade I is 88.45; members of subclade II is 86.18 and members of subclade III it is 89.6. While the core genome of strains within subclade IV is similar in size to the core genomes of the other *P. fluorescens* subclades, the subclade IV accessory genome is considerably smaller (**Figure VII.1**), with fewer than 200 genes being unique to any one subclade IV strain (**Figure V.9)**. Unconstrained ordination based on shared gene clusters between strains placed the subclade IV strains in a cluster completely separate from the other *P. fluorescens* subclades (**Figure V.10**). The large geographical (Palo

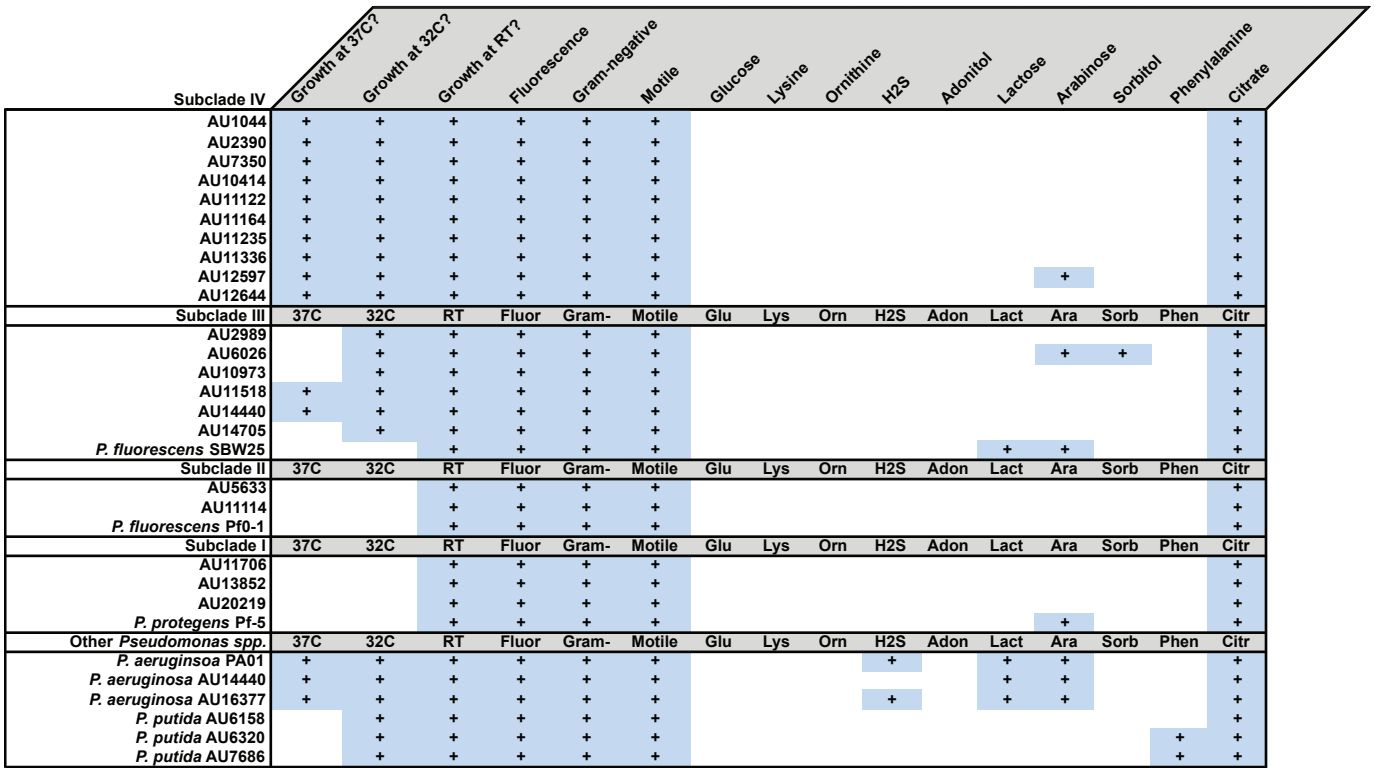
Alto, California to Boston, Massachusetts) and temporal range (February 21st, 1999 to June 6th, 2006) of where and when they were isolated makes it unlikely that the subclade IV strains represent an outbreak within the CF patient population (**Figure V.4**). Rather, members of the newly-formed fourth subclade appear to represent a group of *P. fluorescens* strains that have genomically adapted to a human host habitat. While additional analysis on the unique genetic, biochemical and physical characteristics of these strains is needed, the evidence provided here strongly suggests that the ten clinical isolates grouped into subclade IV are the first snapshot of the evolution towards a new species of human-adapted *Pseudomonas* bacteria.

Figure V.1: Date, location and source of *P. fluorescens* strains.

Isolate ID	Isolation Date	Isolation Location	Isolation Source
AU1044	2/21/99	Columbia, MO	CF Sputum
AU2390	10/2/00	Boston, MA	CF Sputum
AU5633	4/30/03	St. Louis, MO	CF Throat Swab
AU7350	5/18/04	Hartford, CT	CF Sputum
AU10414	12/15/05	Boston, MA	CF Sputum
AU11114	4/28/06	Omaha, NE	CF Sputum
AU11122	5/4/06	Palo Alto, CA	CF Sputum
AU11164	5/9/06	St. Louis, MO	CF Sputum
AU11235	5/18/06	Ann Arbor, MI	CF Sputum
AU11336	6/6/06	Cincinnati, OH	CF Sputum
AU11706	8/9/06	Hartford, CT	CF Sputum
AU12597	12/25/06	Hartford, CT	CF Sputum
AU12644	12/19/06	Hartford, CT	CF Sputum
AU13852	7/23/07	Hartford, CT	CF Sputum
AU20219	4/20/10	Austin, TX	Non-CF Tracheal Aspirate

Figure V.1: Date, location and source of *P. fluorescens* strains.
The newly sequenced strains were isolated over an elevan-year period, between February 1999 and April 2010. Isolation occurred at nine different hospitals across America. Each strain was isolated from a 2010. Isolation occurred at nine different hospitals across America. Each strain was isolated from a separate patient, fourteen isolated from individuals with cystic fibrosis and one from a non-cystic fibrosis patient. The majority (thirteen strains) were isolated from sputum; one strain was isolated from a throat swab and one was isolated from a tracheal aspirate.

Figure V.2: Growth properties of clinical and reference *Pseudomonas* strains.



Subclade IV	Growth at 37C?	Growth at 32C?	Growth at RT?	Fluorescence	Gram-negative	Motile	Glucose	Lysine	Ornithine	H2S	Adonitol	Lactose	Arabinose	Sorbitol	Phenylalanine	Citrate
AU1044	+	+	+	+	+	+										+
AU2390	+	+	+	+	+	+										+
AU7350	+	+	+	+	+	+										+
AU10414	+	+	+	+	+	+										+
AU11122	+	+	+	+	+	+										+
AU11164	+	+	+	+	+	+										+
AU11235	+	+	+	+	+	+										+
AU11336	+	+	+	+	+	+										+
AU12597	+	+	+	+	+	+							+			+
AU12644	+	+	+	+	+	+										+
Subclade III	**37C**	**32C**	**RT**	**Fluor**	**Gram-**	**Motile**	**Glu**	**Lys**	**Orn**	**H2S**	**Adon**	**Lact**	**Ara**	**Sorb**	**Phen**	**Citr**
AU2989		+	+	+	+	+										+
AU6026		+	+	+	+	+							+	+		+
AU10973		+	+	+	+	+										+
AU11518	+	+	+	+	+	+										+
AU14440	+	+	+	+	+	+										+
AU14705		+	+	+	+	+										+
P. fluorescens SBW25			+	+	+	+						+	+			+
Subclade II	**37C**	**32C**	**RT**	**Fluor**	**Gram-**	**Motile**	**Glu**	**Lys**	**Orn**	**H2S**	**Adon**	**Lact**	**Ara**	**Sorb**	**Phen**	**Citr**
AU5633			+	+	+	+										+
AU11114			+	+	+	+										+
P. fluorescens Pf0-1			+	+	+	+										+
Subclade I	**37C**	**32C**	**RT**	**Fluor**	**Gram-**	**Motile**	**Glu**	**Lys**	**Orn**	**H2S**	**Adon**	**Lact**	**Ara**	**Sorb**	**Phen**	**Citr**
AU11706			+	+	+	+										+
AU13852			+	+	+	+										+
AU20219			+	+	+	+										+
P. protegens Pf-5			+	+	+	+							+			+
Other *Pseudomonas spp.*	**37C**	**32C**	**RT**	**Fluor**	**Gram-**	**Motile**	**Glu**	**Lys**	**Orn**	**H2S**	**Adon**	**Lact**	**Ara**	**Sorb**	**Phen**	**Citr**
P. aeruginsoa PA01	+	+	+	+	+	+				+		+	+			+
P. aeruginosa AU14440	+	+	+	+	+	+						+	+			+
P. aeruginosa AU16377	+	+	+	+	+	+				+		+	+			+
P. putida AU6158		+	+	+	+	+										+
P. putida AU6320		+	+	+	+	+									+	+
P. putida AU7686		+	+	+	+	+									+	+

Figure V.2: Growth properties of clinical and reference *Pseudomonas* strains.
Plus sign represents a positive result to the growth quality listed above. Growth at various temperatures performed aerobically in LB broth. Fluorescence tested on solid LB agar under UV light. Both gram-negative and motility were assessed under a standard light microscope. Utilization of various substrates was performed using the BBL Enterotube II growth assay system.

Figure V.3: Assembly statistics of clinical strains of *P. fluorescens* sequenced in this study.

		Subclade IV							
		AU1044	**AU2390**	**AU7350**	**AU10414**	**AU11122**	**AU11164**	**AU11235**	**AU11336**
Assembly Totals	# of Contigs	59	106	60	75	84	89	98	91
	Contigs >2K	19	73	30	23	26	27	23	28
	Assembled Sequences	3828761	2924945	3446930	3789204	3125039	3809383	3859620	3918024
	Unassembled Sequences	243151	117257	197326	169552	239017	169767	188200	164566
	Sequences not assembled due to complete trimming	9451	7814	7160	8821	7827	7854	9175	9075
	Sequences removed due to small contig size	158681	53298	121474	99319	147432	87016	99233	84892
	All Sequences	4071912	3042202	3644256	3958756	3364056	3979150	4047820	4082590
	Contig N50	443 kbases	181 kbases	357 kbases	296 kbases	283 kbases	374 kbases	291 kbases	254 kbases
	Average Coverage	74	52	69	76	62	76	77	79
Average Totals	Sequences per Contig	64894	27593	57448	50522	37202	42802	39383	43055
Average Lengths	Contigs	81211	51177	80835	64680	57932	54679	49742	53483
	Assembled Sequences	97	97	97	98	97	97	97	98
	Unassembled Sequences	82	74	77	81	82	80	80	80
	All Sequences	96	96	96	91	96	97	97	97
Average Quality	Assembled Sequences	33	33	33	33	33	33	33	33
	Unassembled Sequences	30	28	30	29	30	29	29	29
	All Sequences	33	33	33	33	33	33	33	33
Assembled Pair Statistics	Read Pairs	2035956	1521101	1822128	1979378	1682028	1989575	2023910	2041295
	Assembled Pairs	1880674	1434080	1696270	1868195	1537078	1875195	1899943	1930808
	Pairs Consistent Within a Contig	1876532	1427998	1689161	1861798	1530451	1867777	1893081	1923280
	Pairs Inconsistent Within a Contig	136	109	367	173	215	177	317	208

		Subclade IV		Subclade II		Subclade I		
		AU12597	AU12644	AU5633	AU11114	AU11706	AU13852	AU20219
Assembly Totals	# of Contigs	61	76	69	95	165	70	123
	Contigs >2K	21	47	41	45	89	33	66
	Assembled Sequences	4019860	3784046	3322127	3374130	4044023	3628817	3535384
	Unassembled Sequences	137148	180488	92379	178788	151355	123193	166982
	Sequences not assembled due to complete trimming	7947	10325	7781	8185	7300	6677	8769
	Sequences removed due to small contig size	61284	107375	29100	109232	66439	55268	75301
	All Sequences	4157008	3964534	3414506	3552918	4195378	3752010	3702366
	Contig N50	590 kbases	219 kbases	369 kbases	253 kbases	173 kbases	493 kbases	272 kbases
	Average Coverage	80	76	49	47	65	58	57
Average Totals	Sequences per Contig	65899	47899	48146	35517	24509	51840	28742
Average Lengths	Contigs	80639	61735	88251	65583	44145	95750	59356
	Assembled Sequences	97	97	97	97	97	97	97
	Unassembled Sequences	73	80	66	80	77	74	77
	All Sequences	96	97	96	96	96	96	96
Average Quality	Assembled Sequences	33	33	33	33	33	33	33
	Unassembled Sequences	28	29	26	29	29	28	29
	All Sequences	33	33	33	33	33	33	33
Assembled Pair Statistics	Read Pairs	2078504	1982267	1707253	1776459	2097689	1876005	1851183
	Assembled Pairs	1977383	1861226	1633711	1658381	1988520	1786775	1734867
	Pairs Consistent Within a Contig	1972380	1854612	1630174	1654436	1980928	1783169	1727783
	Pairs Inconsistent Within a Contig	112	173	78	131	262	194	129

Figure V.3: Assembly statistics of clinical strains of *P. fluorescens* sequenced in this study.
Clinical *P. fluorescens* strains were de novo assembled using the DNAstar SeqMan NGen software package.

Figure V.4: Subclade IV isolation across the United States.

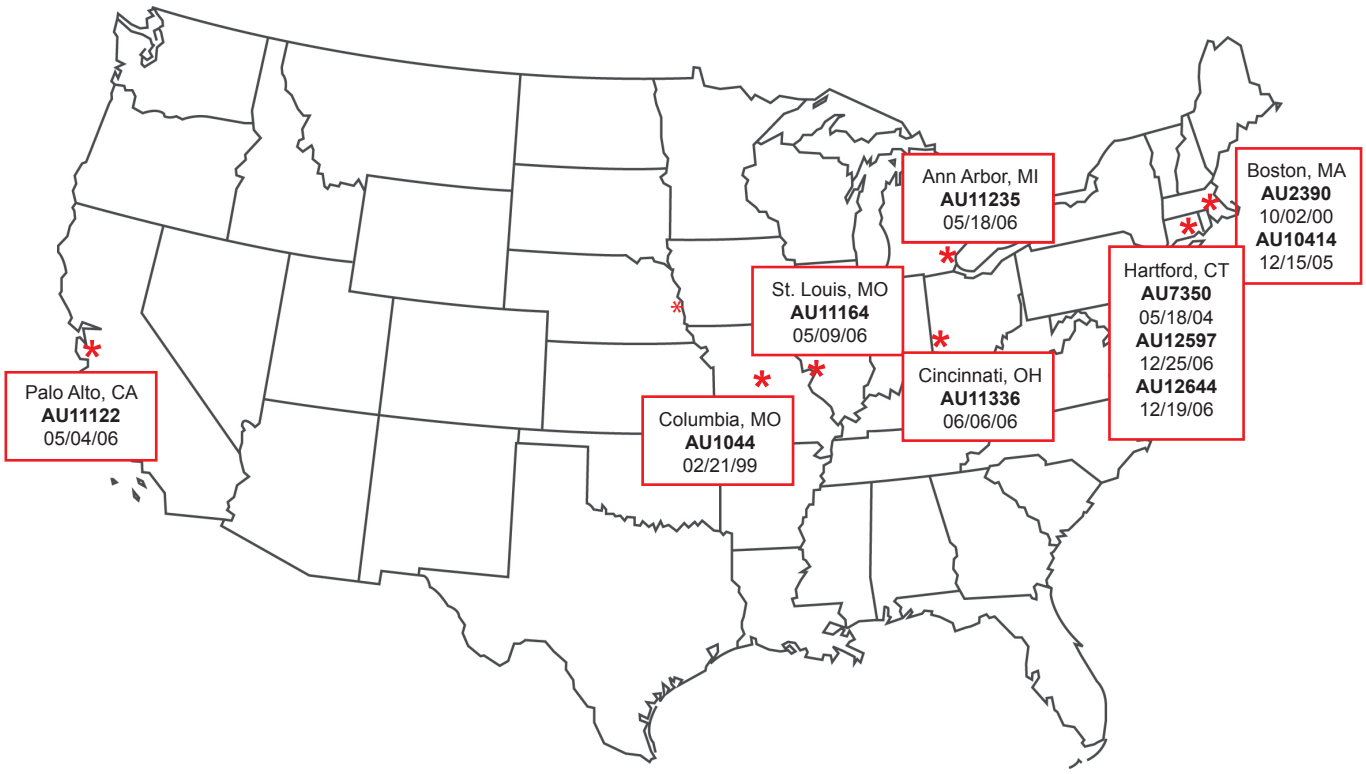


Figure V.4: Subclade IV isolation across the United States.
The geographical location of the treatment center where each subclade IV was isolated. The date of isolation is included under the name of each strain.

Figure V.5: Genome features of *P. fluorescens* strains in this study.

	Strain	Genome Size (Mbp)	G+C Content (%)	# of Coding Sequences	# of 16S rRNAs	Plasmid	Reference
Subclade IV	**AU1044**	4768445	58.7	4309	3	No	this study
	AU2390	5409304	58.3	4925	2	No	this study
	AU7350	4845006	58.7	4394	2	No	this study
	AU10414	4836706	58.6	4386	3	No	this study
	AU11122	4839902	58.6	4384	4	No	this study
	AU11164	4834154	58.7	4380	3	No	this study
	AU11235	4839509	58.7	4392	4	No	this study
	AU11336	4837661	58.6	4387	3	No	this study
	AU12597	4896021	58.8	4441	2	No	this study
	AU12644	4863816	58.8	4417	1	No	this study
Subclade II	**AU5633**	6083404	60.3	5448	7	No	Scales, 2015
	AU11114	6204679	58.8	5586	5	No	Scales, 2015
	***P. fluorescens* Pf0-1**	6438405	60.6	5722	6	No	Loper, 2012
	***P. fluorescens* R124**	6256692	60.3	5500	5	No	Barton, 2013
	***P. fluorescens* Q2-87**	6368168	60.7	5597	6	No	Loper, 2012
	***P. fluorescens* Q8r1-96**	6602611	61	5717	5	No	Loper, 2012
Subclade I	**AU11706**	7256896	62.9	6548	5	No	Scales, 2015
	AU13852	6678920	63.3	6125	3	No	Scales, 2015
	AU20219	7275643	62.8	6574	2	No	Scales, 2015
	***P. protegens* Pf-5**	7074893	63.3	6347	5	No	Paulsen, 2005
	***P. protegens* CHA0**	6867980	63.4	6226	5	No	Tatusova, 2014

Figure V.6: Genome size of newly sequenced and reference *Pseudomonas* strains.

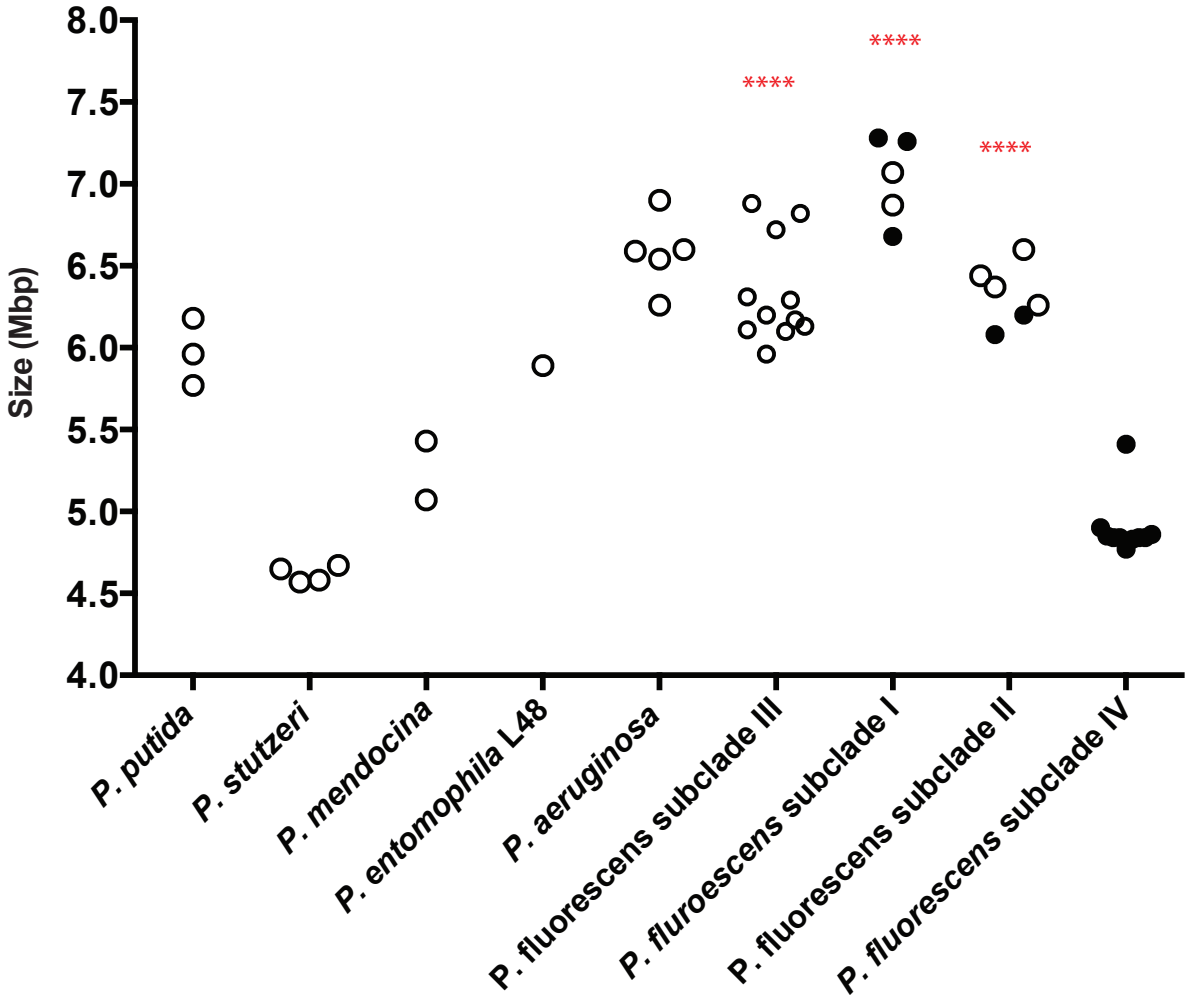


Figure V.6: Genome Size in newly sequenced and reference *Pseudomonas* strains.
The number of nucleotides in the genomes of newly sequenced and reference *Pseudomonas* strains. Closed dots represent strains sequenced in this study; open dots represent previoulsy sequenced *Pseudomonas* strains downloaded from NCBI. One-way ANOVA with Tukey's multiple comparison post test was used to determine statistical relevance. All *P. fluorescens* subclades are statistically different from subclade IV. All statistical analysis is displayed in corresponding.
Table I. ****p <0.00001

Figure V.7: % G+C content of newly sequenced and reference *Pseudomonas* strains**.**

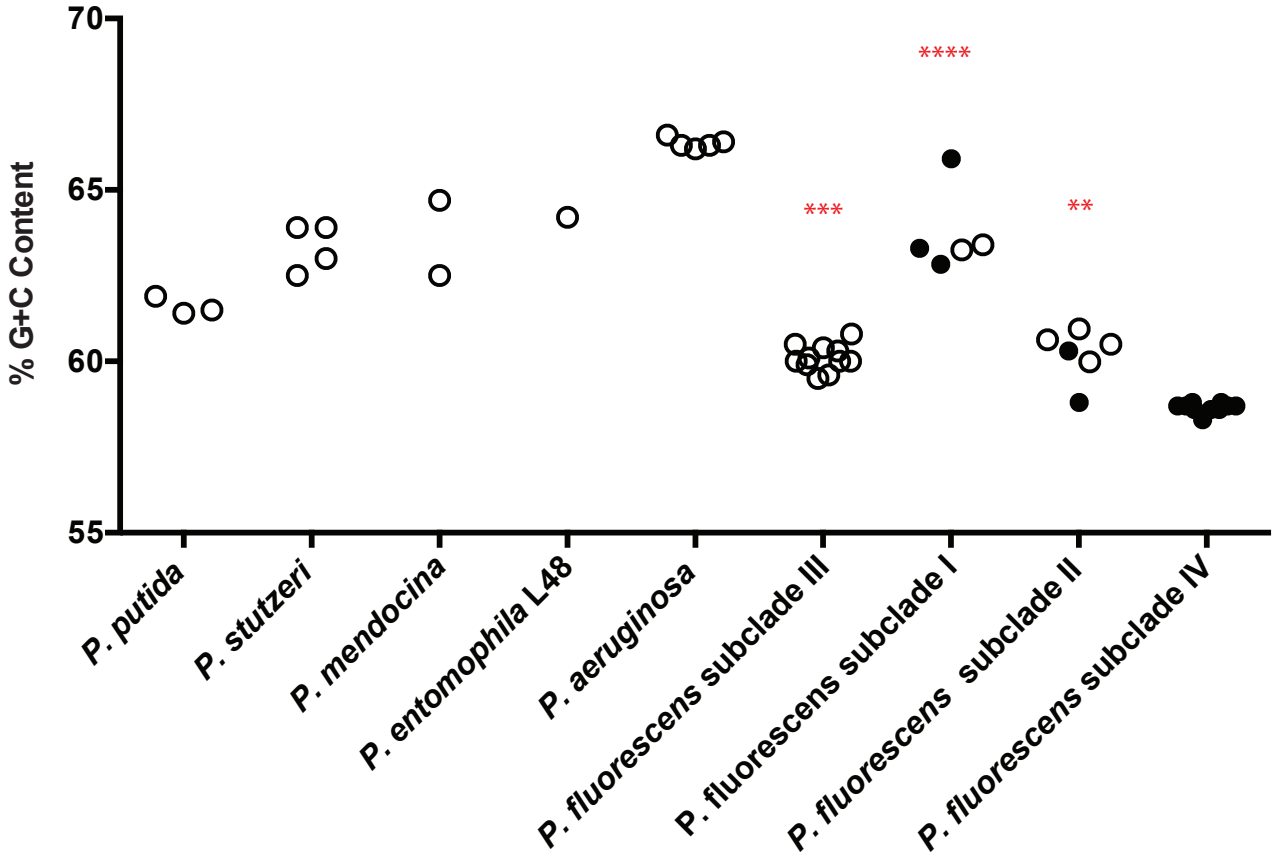


Figure V.7: % G+C content of newly sequenced and reference *Pseudomonas* strains.
The number of G+C occurrences, as percentage of total genome, in newly sequenced and reference *Pseudomonas* strains. Closed dots represent strains sequenced in this study; open dots represent previoulsy sequenced *Pseudomonas* strains downloaded from NCBI. One-way ANOVA with Tukey's multiple comparison post test was used to determine statistical relevance. All *P. fluorescens* subclades are statistically different from subclade IV. All statistical analysis is displayed in corresponding. **Appendix Figure V.3**. **p <0.01; ***p <0.001; ****p <0.00001

Figure.V8a: G+C content across the genomes of newly sequenced subclade IV strains.

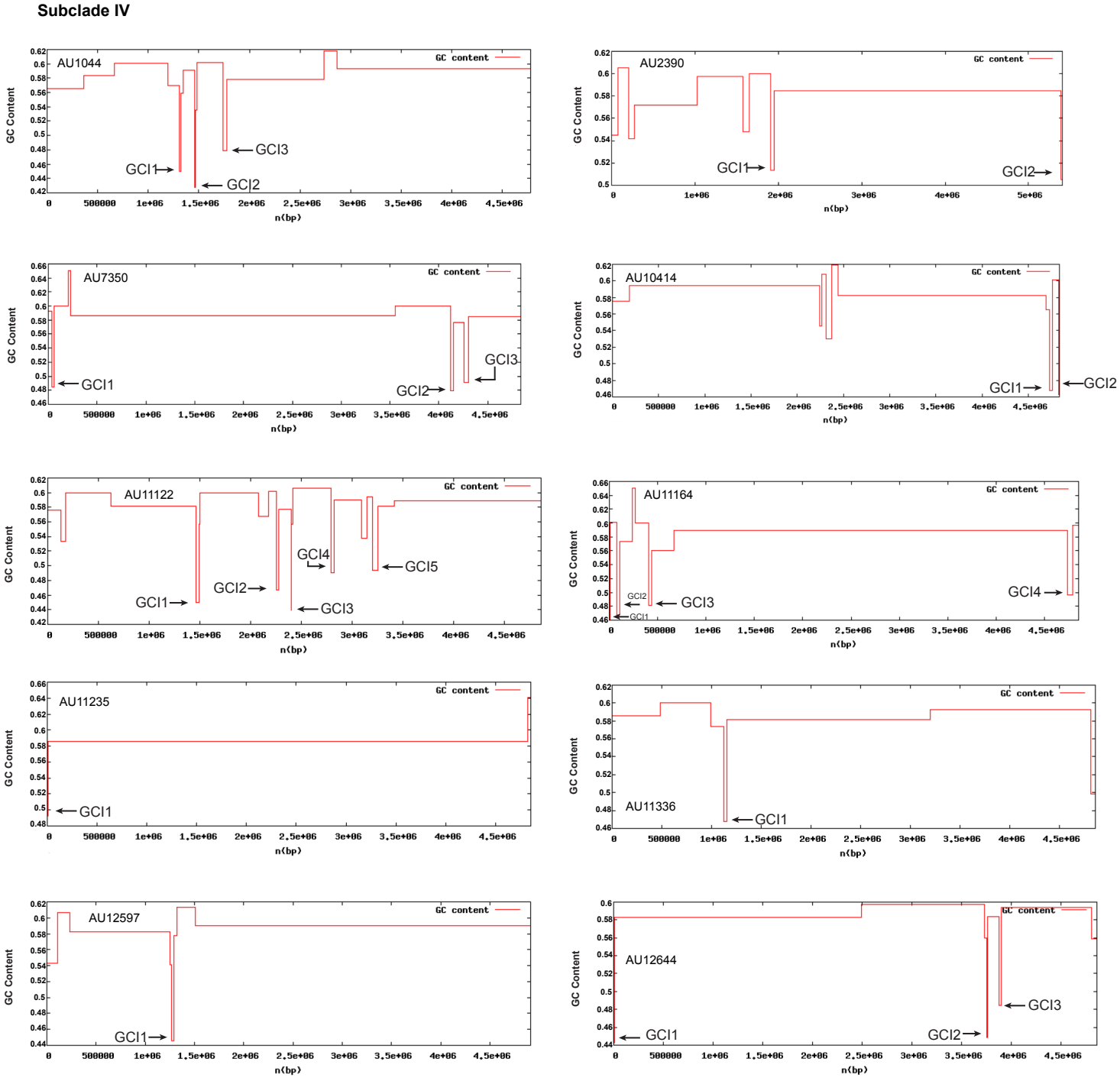


Figure.V8a: G+C content across the genomes of newly sequenced subclade IV strains.
GC content calculated using the online GC profile tool, as described in **Chapter II**. GC content displayed on y-axis and position in draft genome displayed on x-axis. Arrows indicate GC islands that have been further analyzed via NCBI blastn, results in **Appendix Figure V.4.**

Figure V.8b: G+C content across the genomes of subclade I and II strains.

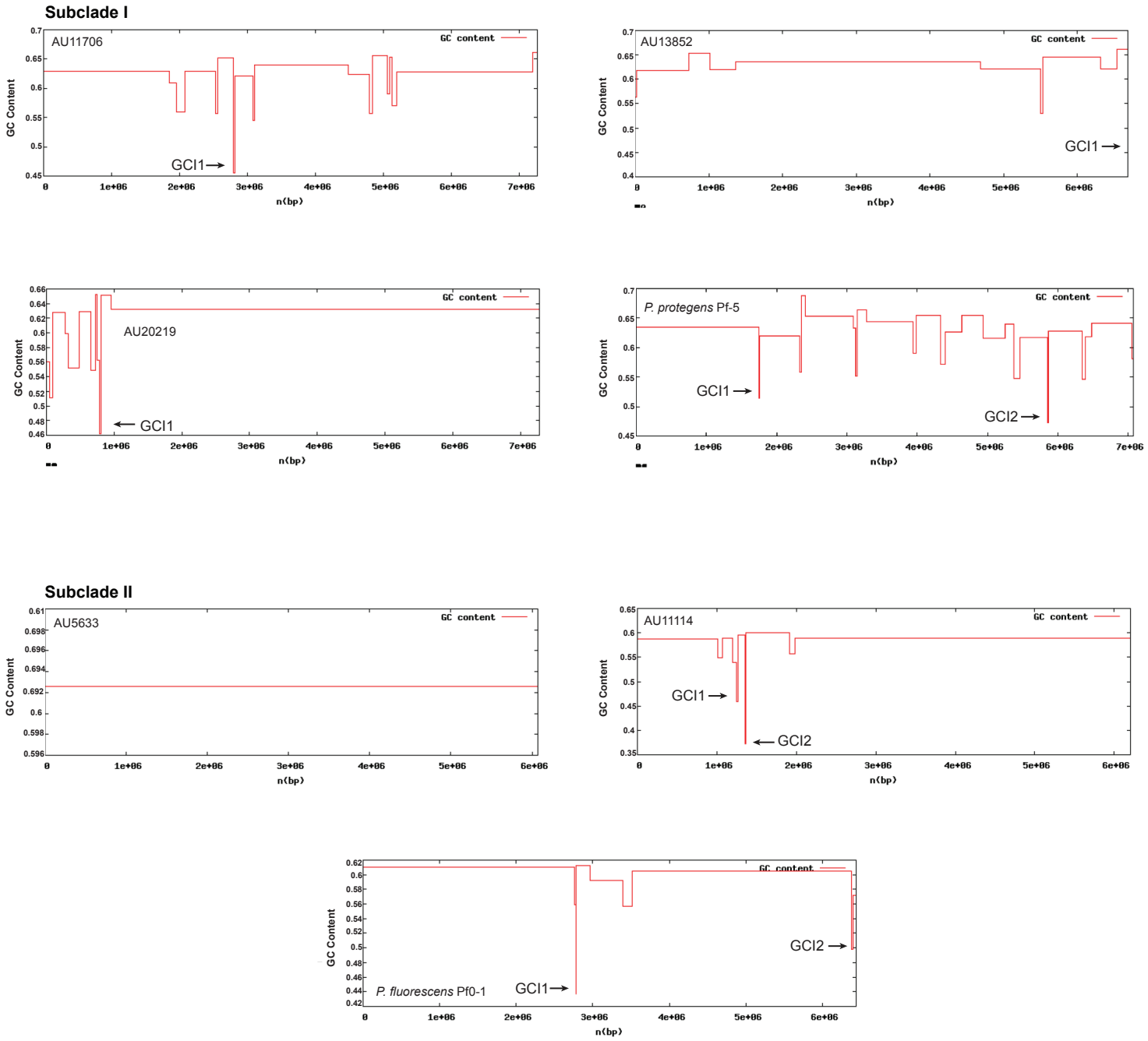


Figure V.8b. **G+C content across the genomes of subclade I and II strains.**
GC content calculated using the online GC profile tool, as described in **Chapter II**. GC content displayed on y-axis and position in draft genome displayed on x-axis. Arrows indicate GC islands that have been further analyzed via NCBI blastn, results in **Appendix Figure V.4**.

Figure V.9: Pan, accessory and core analysis of the ten subclade IV strains.

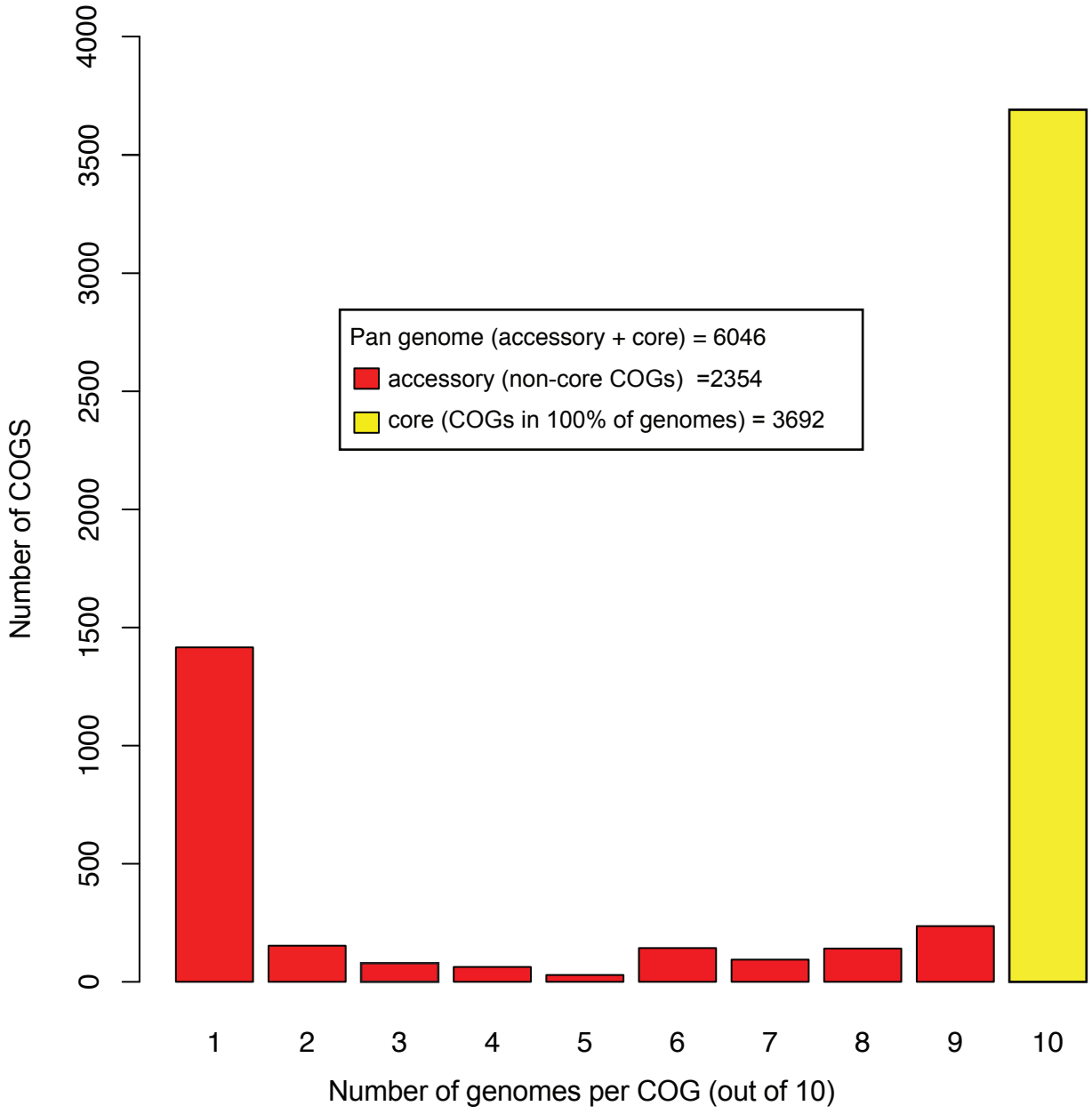


Figure V.9: Pan, accessory and core analysis of the ten subclade IV strains.
Each of the 6046 COGS was analyzed to determine how many genomes encoded that particular COG (1-01, x-axis). Then the number of COGS encoded by only one genome, two genomes, etc was determined (y-axis). The pan genome contains both the core and accessory genomes. The average number of COGs per genome is 4442 (range: 4309-4925). The pan genome and its compartments was calculated using the COGtriangle clustering algorithm in GET_HOMOLOGUES, as described in **Chapter II**.

Figure V.10: PCA built on shared gene clusters between *Pseudomonas* strains.

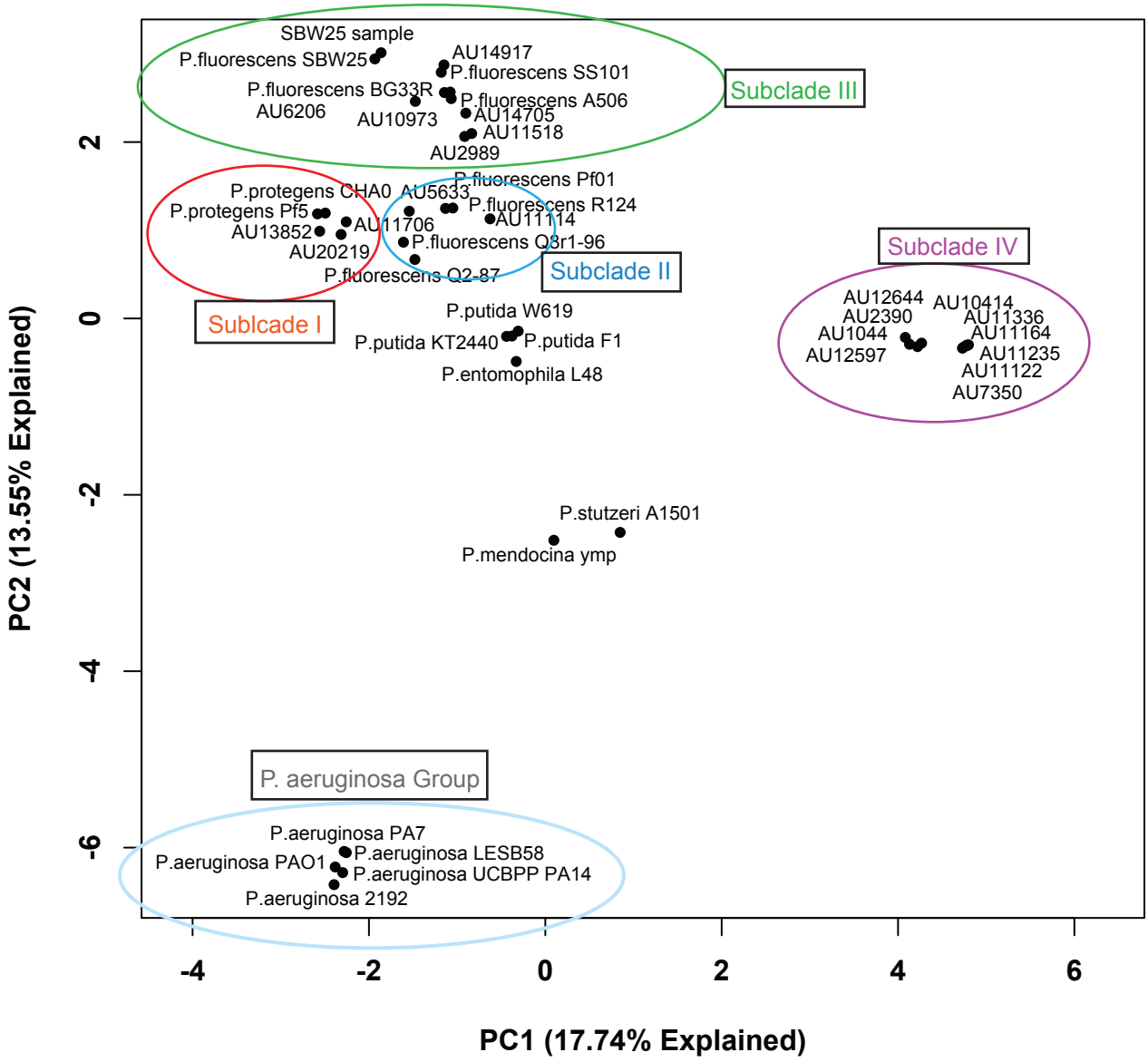


Figure V.10: PCA built on shared gene clusters between *Pseudomonas* strains.
Principal component analysis (PCA) built based on shared gene clusters between *Pseudomonas* strains. Circles are drawn to indicate the natural clustering that occurs between subclades and species groups. The strains beginning with 'AU' in subclades I, II and IV represent clinical strains sequenced in this study. Strains beginning with 'AU' in subclade III represent clinical strains we previously reported.

Figure V.11: Alignment of the *pfi*T amino acid sequences from subclade I and II *P. fluorescens* strains.

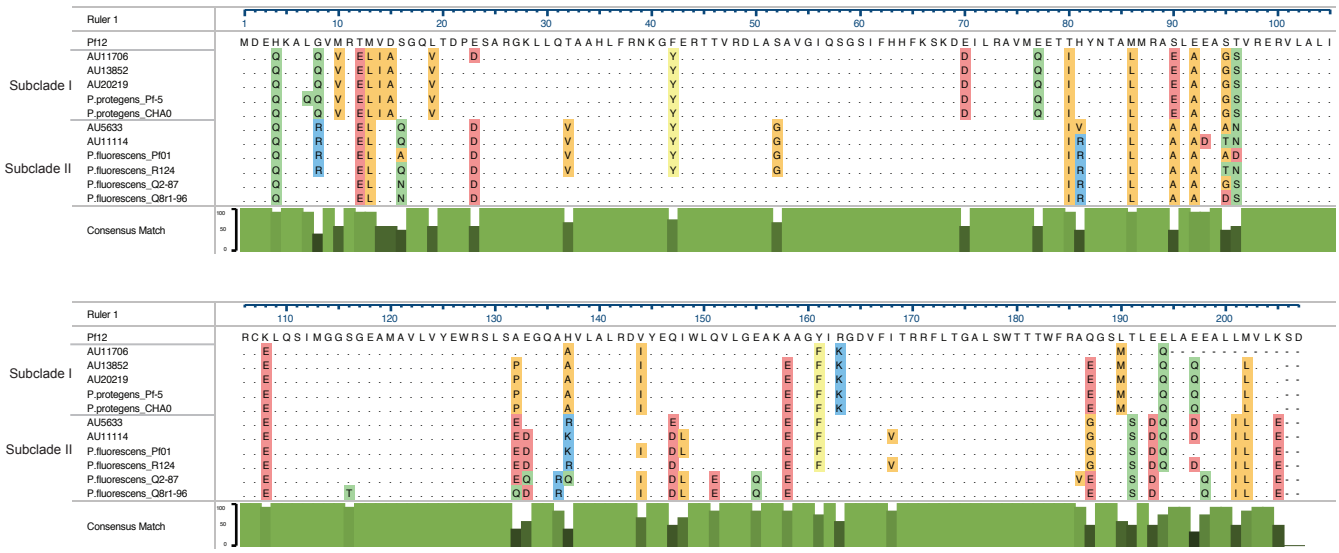


Figure V.11: Alignment of the *pfi*T amino acid sequences from subclade I and II *P. fluorescens* strains.
Amino acid sequence of the *pfi*T protein aligned using MUSCLE, as described in **Chapter II**. Published sequence of *pfi*T used as reference (labeled *pfi*T above). Only amino acids differing from reference are shown and colored. Percent nucleotide consensus match visualized in green underneath alignment. Alignment visualized with DNAstar MegAligPro.

Figure V.12: Alignment of l2 nucleotide sequence subclade I, II and III *P. fluorescens* strains.

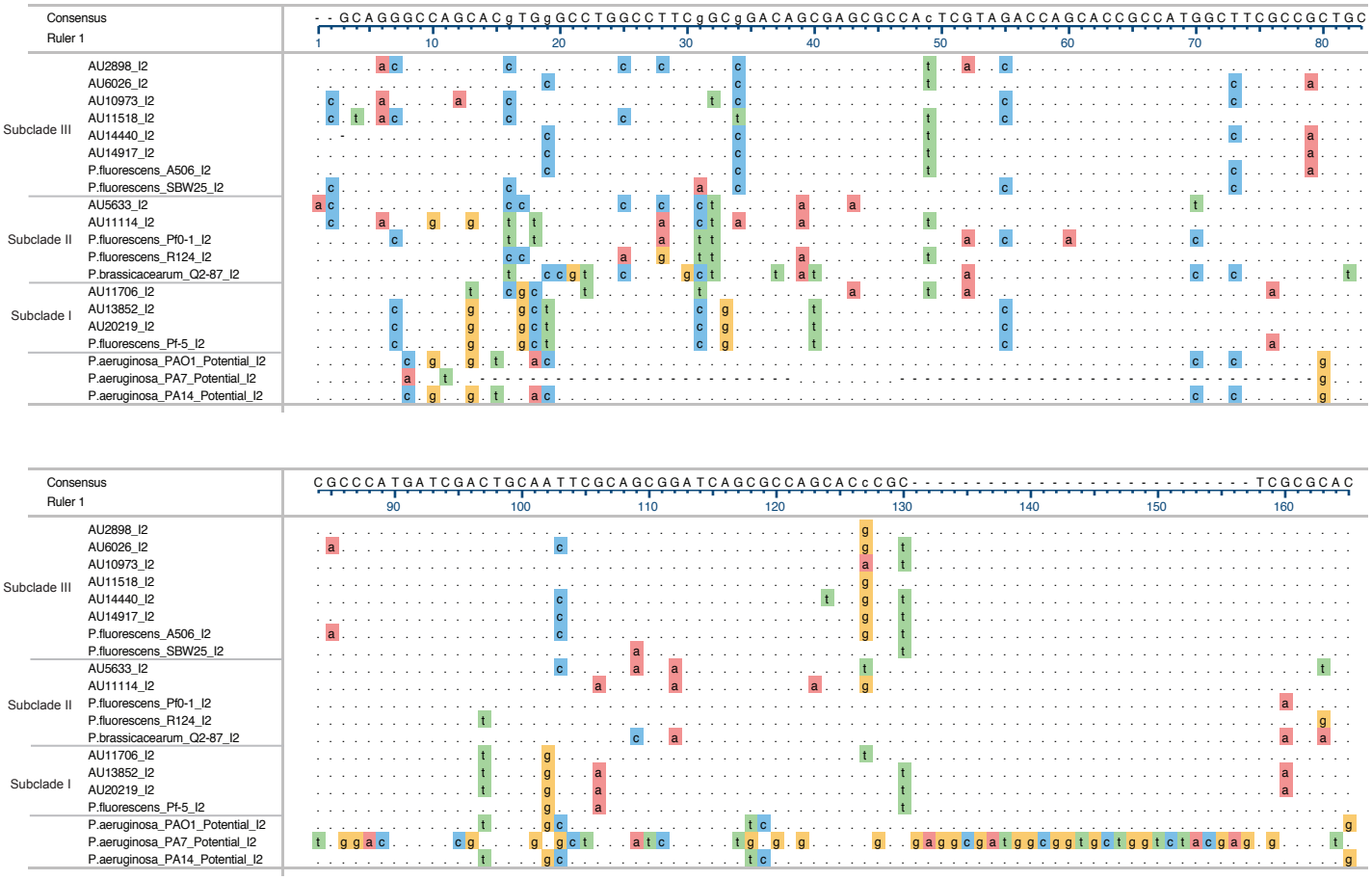


Figure V.12: Alignment of l2 nucleotide sequence subclade I, II, III *P. fluorescens* strains.
Nucleotide sequences of the l2 gene were aligned using the MAFFT algorithm within DNAstar's MegAlign Pro software. Only nucleotides differing from concensus are shown and colored. Alignment visualized with DNAstar MegAligPro.

Figure V.13: Number of gene homologues annotated as part of the Ysc/Yop-like type-III secretion system.

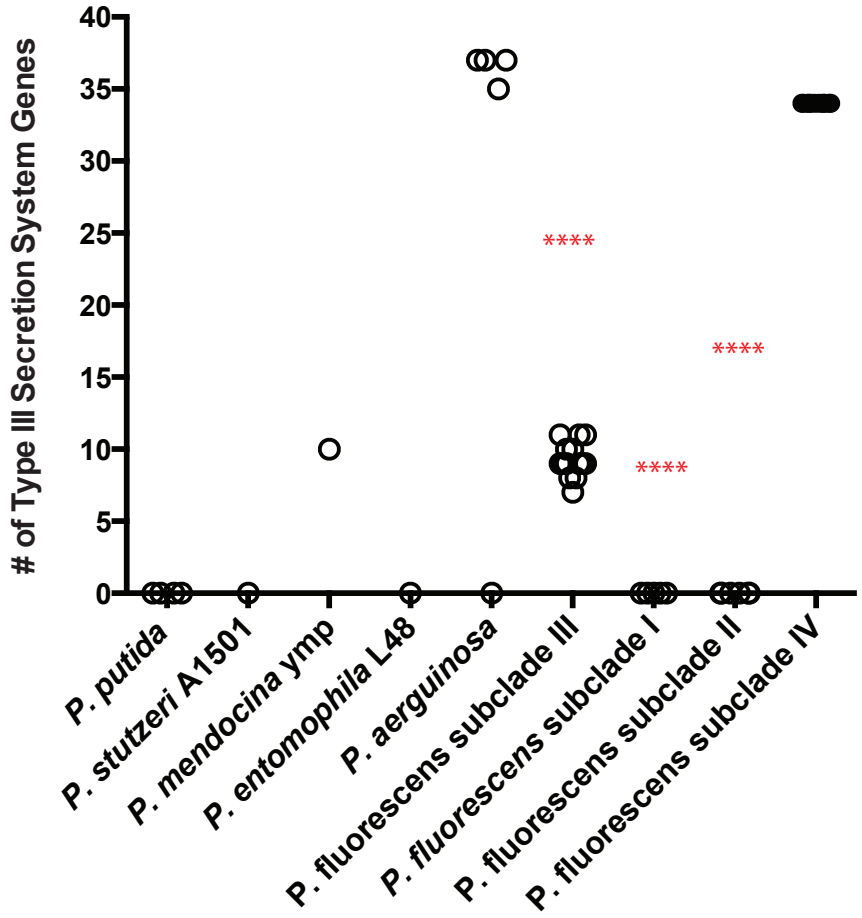


Figure V.13: Number of gene homologues annotated as part of the Ysc/Yop-like type-III secretion system.
Genes were annotated as belonging to the Ysc/Yop-family T3SS using the RAST online annotation software, as described in **Chapter II**. Closed dots represent strains sequenced in this study; open dots represent previoulsy sequenced *Pseudomonas* strains downloaded from NCBI. The *P. putida* strains correspond to F1, KT2440 and W619 The *P. aeruginsoa* strains correspond to PAO1, LESB58 and PA14. The *P. aeruginosa* strain with no T3SS genes is PA7. Two-way ANOVA with Tukey's multiple comparison post test was used to determine statistical relevance. All *P. fluorescens* subclades were significantly different from subclade IV. All statistical analysis is displayed in corresponding **Appendix Figure V.6.** ****p <0.00001

Figure V.14: Amplification of the ExoU protein with *P. fluorescens* subclade IV- specific primers.

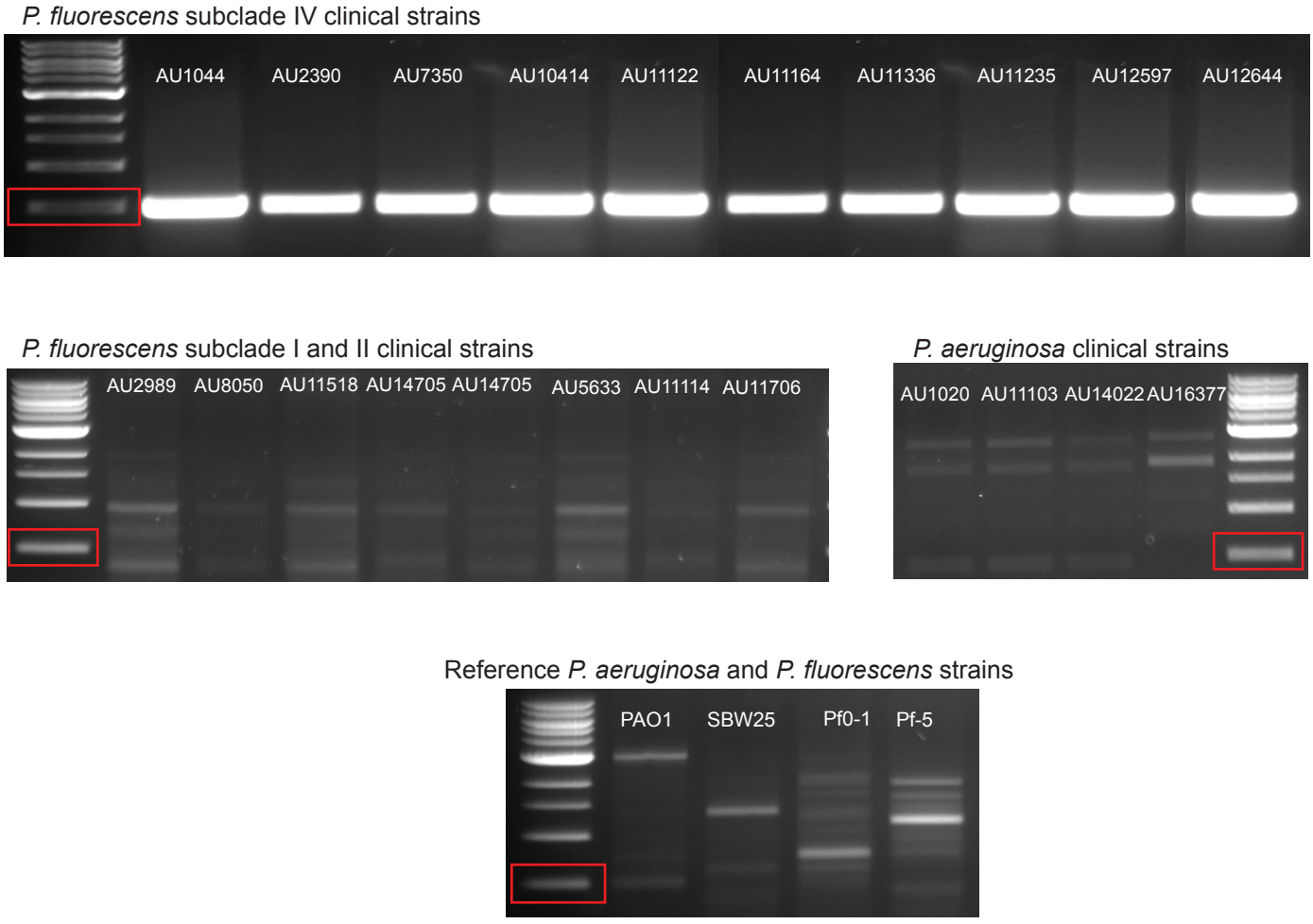


Figure V.14: Amplification of the ExoU protein with *P. fluorescens* subclade IV- specific primers.
Amplified ExoU gene in newly sequenced and previously sequenced *Pseudomonas* strains. Primers designed and PCR performed as detailed in **Appendix File V.8**. Red box indicates predicted size of the ExoU amplified gene.

Figure V.15: ExoU and SpcU protein phylogeny in *Pseudomonas* strains.

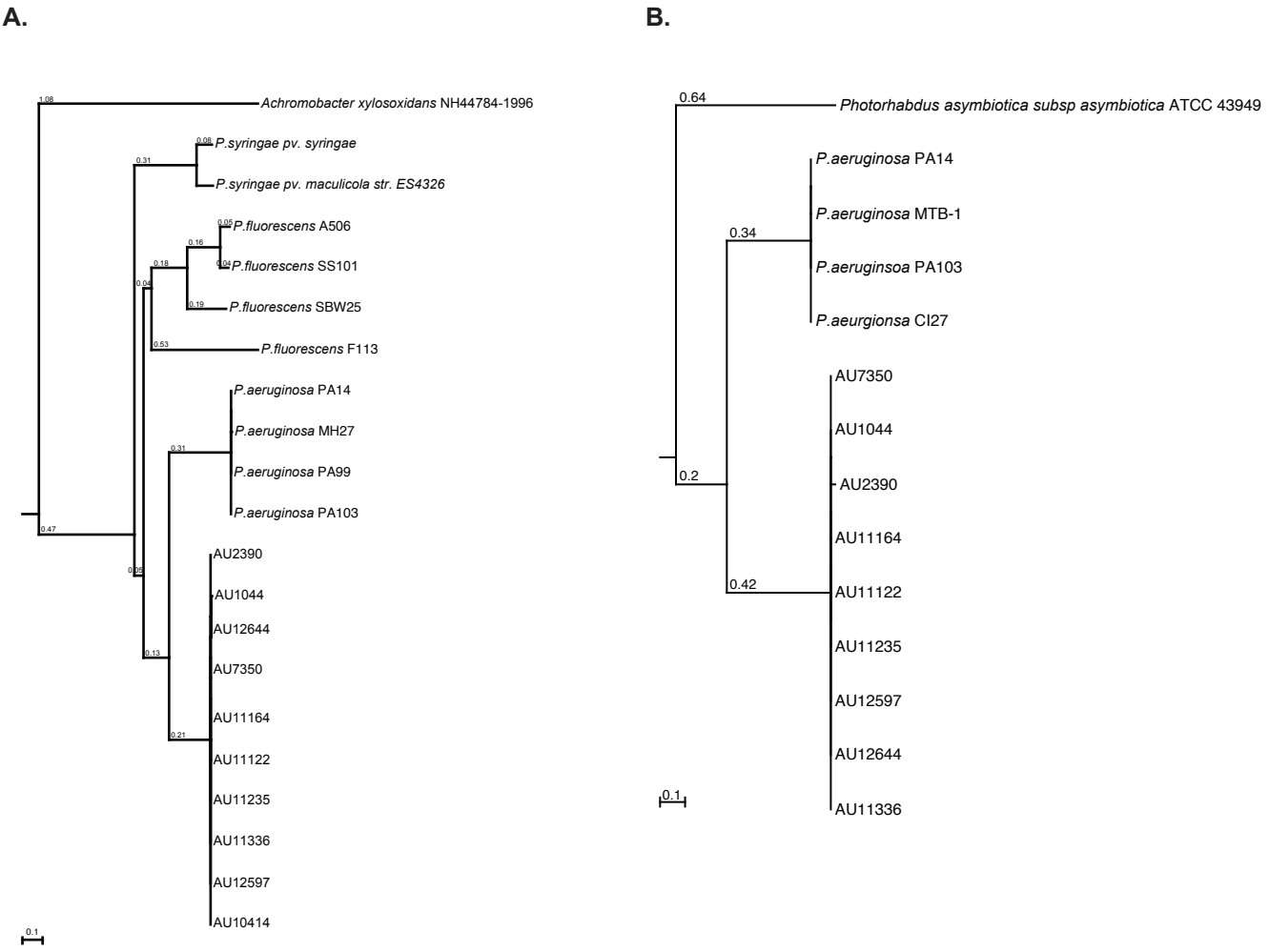


Figure U. ExoU protein phylogeny and amino acid similarity in *Pseudomonas* strains.
A. Phylogenetic tree built on the amino acid sequence of the **A**. ExoU protein and **B.** of newly sequenced *P. fluorescens* subclade IV and reference *Pseudomonas* strains.

Figure V.16: Alignment of SpcU homologues.

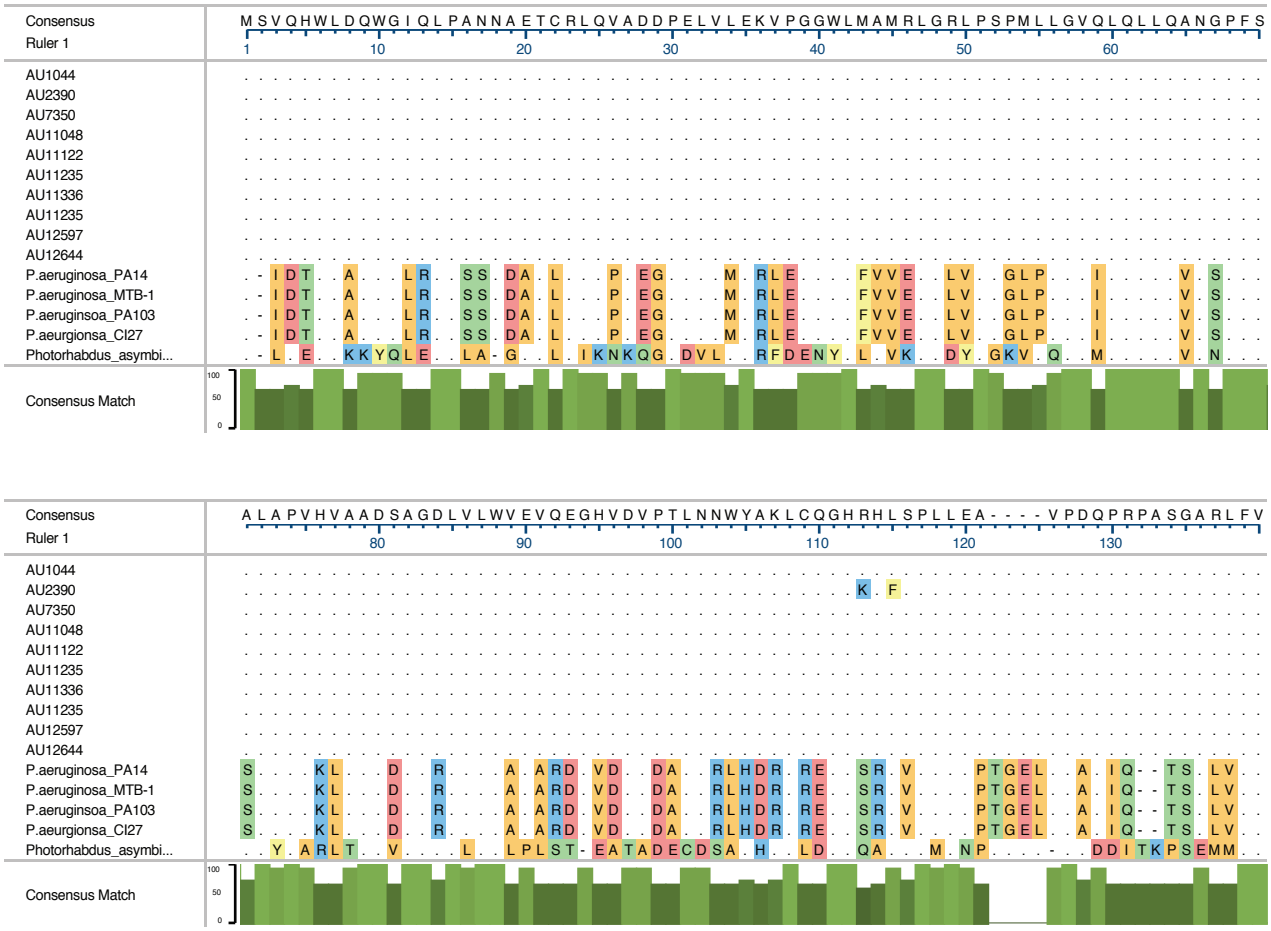


Figure V.16: Alignment of SpcU homologues.
Alignment of SpcU homologues from clinical *P. fluorescens* subclade IV strains and representative *P. aeruginosa* strains performed with the MAFFT algorithm (Katoh, 2002) within DNAstar's MegAlign Pro software. SpcU homologues discovered in RAST, as described in **Chapter II**. Dots represent similarties to concensus; colored amino acids represent individual changes from the concensus.

Figure V.17: Model of the type III secretion system in *P. fluorescens* subclade IV strains.

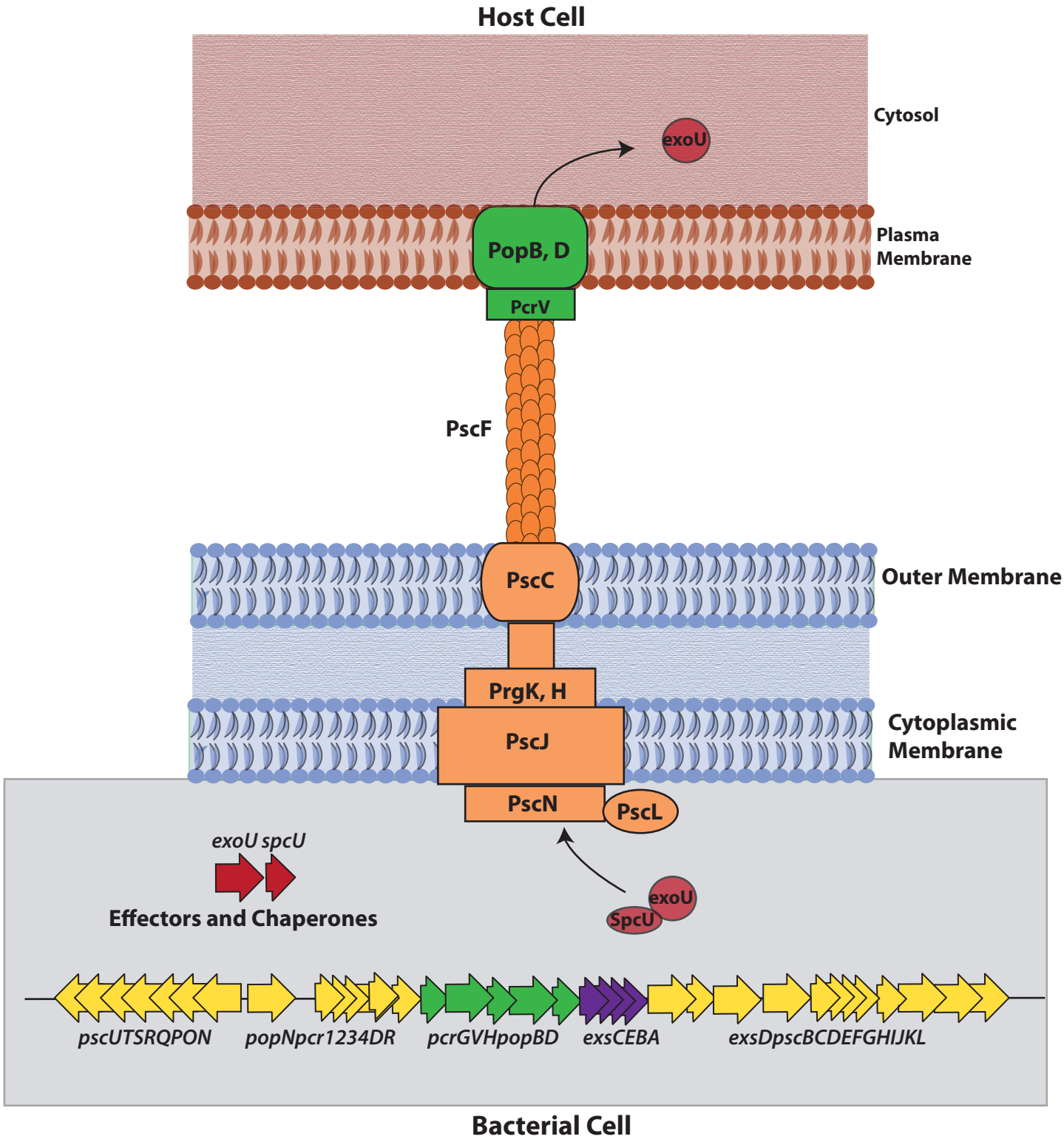


Figure V.17: Model of the type III secretion system in *P. fluorescens* subclade IV strains.
Proposed model of the type III secretion system in the newly sequenced *P. fluorescens* subclade IV strains. The model is built based on the presence of annotated type III secretion system genes in the subclade IV strains and the previously published location of the expressed proteins of those genes in the type III secretion system in *P. aeruginosa* (Hauser, 2009).

Chapter V

References

1. Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW. 2011. Pseudomonas genomes: diverse and adaptable. FEMS Microbiol Rev 35:652-680.
2. Loper JE, Hassan KA, Mavrodi DV, Davis EW, 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT. 2012. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. PLoS Genet 8:e1002784.
3. Scales BS, Dickson RP, LiPuma JJ, Huffnagle GB. 2014. Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. Clin Microbiol Rev 27:927-948.
4. Scales BS, Erb-Downward JR, Huffnagle IM, LiPuma JJ, Huffnagle GB. 2015. Comparative genomics of *Pseudomonas fluorescens* subclade III strains from human lungs. BMC Genomics *In Review*.
5. Dickson RP, Erb-Downward JR, Huffnagle GB. 2014. Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. Lancet Respir Med 2:238-246.
6. Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346-351.

7. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13.
8. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences 101:3160-3165.
9. Aujoulat F, Roger F, Bourdier A, Lotthe A, Lamy B, Marchandin H, Jumas-Bilak E. 2012. From environment to man: genome evolution and adaptation of human opportunistic bacterial pathogens. Genes (Basel) 3:191-232.
10. Georgiades K, Raoult D. 2010. Defining pathogenic bacterial species in the genomic era. Front Microbiol 1:151.
11. Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for alpha-proteobacterial genome evolution. Proceedings of the National Academy of Sciences of the United States of America 101:9722-9727.
12. Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol 15:321-326.
13. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D. 2001. Mechanisms of evolution in Rickettsia conorii and R. prowazekii. Science 293:2093-2098.
14. Pallen MJ, Wren BW. 2007. Bacterial pathogenomics. Nature 449:835-842.
15. Gao F, Zhang CT. 2006. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. Nucleic Acids Res 34:W686-691.
16. Lin G, Bultman J, Johnson EA, Fell JW. 2012. Genetic manipulation of Xanthophyllomyces dendrorhous and Phaffia rhodozyma. Methods Mol Biol 898:235-249.
17. Nagendran R, Lee YH. 2015. Green and Red Light Reduces the Disease Severity by Pseudomonas cichorii JBC1 in Tomato Plants via Upregulation of Defense-Related Gene Expression. Phytopathology 105:412-418.

18. Van Houdt R, Monchy S, Leys N, Mergeay M. 2009. New mobile genetic elements in Cupriavidus metallidurans CH34, their possible roles and occurrence in other bacteria. Antonie Van Leeuwenhoek 96:205-226.

19. Klockgether J, Cramer N, Wiehlmann L, Davenport CF, Tummler B. 2011. Pseudomonas aeruginosa Genomic Structure and Diversity. Front Microbiol 2:150.

20. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.

21. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science 278:631-637.

22. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26:1481-1487.

23. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696-7701.

24. Quinn GP, Keough MJ. 2002. Chapter 17. Principal components and correspondence analysis., p. 537, Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge

25. Liu L, Chen H, Brecher MB, Li Z, Wei B, Nandi B, Zhang J, Ling H, Winslow G, Braun J, Li H. 2013. Pfit is a structurally novel Crohn's disease-associated superantigen. PLoS Pathog 9:e1003837.

26. Sutton CL, Kim J, Yamane A, Dalwadi H, Wei B, Landers C, Targan SR, Braun J. 2000. Identification of a novel bacterial sequence associated with Crohn's disease. Gastroenterology 119:23-31.

27. Wei B, Huang T, Dalwadi H, Sutton CL, Bruckner D, Braun J. 2002. Pseudomonas fluorescens encodes the Crohn's disease-associated I2 sequence and T-cell superantigen. Infect Immun 70:6567-6575.

28. Cianciotto NP. 2005. Type II secretion: a protein secretion system for all seasons. Trends Microbiol 13:581-588.

29. Peabody CR, Chung YJ, Yen MR, Vidal-Ingigliardi D, Pugsley AP, Saier MH, Jr. 2003. Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. Microbiology 149:3051-3072.

30. Sandkvist M. 2001. Type II secretion and pathogenesis. Infect Immun 69:3523-3535.

31. Planet PJ, Kachlany SC, Fine DH, DeSalle R, Figurski DH. 2003. The Widespread Colonization Island of Actinobacillus actinomycetemcomitans. Nat Genet 34:193-198.

32. Craig L, Li J. 2008. Type IV pili: paradoxes in form and function. Curr Opin Struct Biol 18:267-277.

33. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ. 2006. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. Proc Natl Acad Sci U S A 103:1528-1533.

34. Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc Natl Acad Sci U S A 104:15508-15513.

35. Suarez G, Sierra JC, Sha J, Wang S, Erova TE, Fadl AA, Foltz SM, Horneman AJ, Chopra AK. 2008. Molecular characterization of a functional type VI secretion system from a clinical isolate of Aeromonas hydrophila. Microb Pathog 44:344-361.

36. Suarez G, Sierra JC, Erova TE, Sha J, Horneman AJ, Chopra AK. 2010. A type VI secretion system effector protein, VgrG1, from Aeromonas hydrophila that induces host cell toxicity by ADP ribosylation of actin. J Bacteriol 192:155-168.

37. Salmond GP, Reeves PJ. 1993. Membrane traffic wardens and protein secretion in gram-negative bacteria. Trends Biochem Sci 18:7-12.

38. Lindgren PB, Peet RC, Panopoulos NJ. 1986. Gene cluster of Pseudomonas syringae pv. "phaseolicola" controls pathogenicity of bean plants and hypersensitivity of nonhost plants. J Bacteriol 168:512-522.

39. Marchi M, Boutin M, Gazengel K, Rispe C, Gauthier JP, Guillerm-Erckelboudt AY, Lebreton L, Barret M, Daval S, Sarniguet A. 2013. Genomic analysis of the biocontrol strain Pseudomonas fluorescens Pf29Arp with evidence of T3SS and T6SS gene expression on plant roots. Environ Microbiol Rep 5:393-403.

40. Preston GM, Bertrand N, Rainey PB. 2001. Type III secretion in plant growth-promoting Pseudomonas fluorescens SBW25. Mol Microbiol 41:999-1014.

41. Rezzonico F, Binder C, Defago G, Moenne-Loccoz Y. 2005. The type III secretion system of biocontrol Pseudomonas fluorescens KD targets the phytopathogenic Chromista Pythium ultimum and promotes cucumber protection. Mol Plant Microbe Interact 18:991-1001.

42. Anderson LM, Stockwell VO, Loper JE. 2004. An Extracellular Protease of Pseudomonas fluorescens Inactivates Antibiotics of Pantoea agglomerans. Phytopathology 94:1228-1234.

43. Sory MP, Cornelis GR. 1994. Translocation of a hybrid YopE-adenylate cyclase from Yersinia enterocolitica into HeLa cells. Mol Microbiol 14:583-594.

44. He SY, Nomura K, Whittam TS. 2004. Type III protein secretion mechanism in mammalian and plant pathogens. Biochim Biophys Acta 1694:181-206.

45. Hauser AR. 2009. The type III secretion system of Pseudomonas aeruginosa: infection by injection. Nat Rev Microbiol 7:654-665.

46. Lee VT, Smith RS, Tummler B, Lory S. 2005. Activities of Pseudomonas aeruginosa effectors secreted by the Type III secretion system in vitro and during infection. Infect Immun 73:1695-1705.

47. Hauser AR, Kang PJ, Engel JN. 1998. PepA, a secreted protein of Pseudomonas aeruginosa, is necessary for cytotoxicity and virulence. Molecular microbiology 27:807-818.

48. Finck-Barbancon V, Goranson J, Zhu L, Sawa T, Wiener-Kronish JP, Fleiszig SM, Wu C, Mende-Mueller L, Frank DW. 1997. ExoU expression by Pseudomonas aeruginosa correlates with acute cytotoxicity and epithelial injury. Molecular microbiology 25:547-557.

49. Sawa T, Ohara M, Kurahashi K, Twining SS, Frank DW, Doroques DB, Long T, Gropper MA, Wiener-Kronish JP. 1998. In vitro cellular toxicity predicts Pseudomonas aeruginosa virulence in lung infections. Infect Immun 66:3242-3249.

50. Lee EJ, Evans DJ, Fleiszig SM. 2003. Role of Pseudomonas aeruginosa ExsA in penetration through corneal epithelium in a novel in vivo model. Invest Ophthalmol Vis Sci 44:5220-5227.

51. Vance RE, Rietsch A, Mekalanos JJ. 2005. Role of the type III secreted exoenzymes S, T, and Y in systemic spread of Pseudomonas aeruginosa PAO1 in vivo. Infect Immun 73:1706-1713.

52. Holder IA, Neely AN, Frank DW. 2001. Type III secretion/intoxication system important in virulence of Pseudomonas aeruginosa infections in burns. Burns 27:129-130.

53. Koh AY, Priebe GP, Pier GB. 2005. Virulence of Pseudomonas aeruginosa in a murine model of gastrointestinal colonization and dissemination in neutropenia. Infect Immun 73:2262-2272.

54. Schulert GS, Feltman H, Rabin SD, Martin CG, Battle SE, Rello J, Hauser AR. 2003. Secretion of the toxin ExoU is a marker for highly virulent Pseudomonas aeruginosa

isolates obtained from patients with hospital-acquired pneumonia. J Infect Dis 188:1695-1706.

55. Shaver CM, Hauser AR. 2004. Relative contributions of Pseudomonas aeruginosa ExoU, ExoS, and ExoT to virulence in the lung. Infect Immun 72:6969-6977.

56. Ader F, Le Berre R, Faure K, Gosset P, Epaulard O, Toussaint B, Polack B, Nowak E, Viget NB, Kipnis E, Guery BP. 2005. Alveolar response to Pseudomonas aeruginosa: role of the type III secretion system. Infect Immun 73:4263-4271.

57. Roy-Burman A, Savel RH, Racine S, Swanson BL, Revadigar NS, Fujimoto J, Sawa T, Frank DW, Wiener-Kronish JP. 2001. Type III protein secretion is associated with death in lower respiratory and systemic Pseudomonas aeruginosa infections. J Infect Dis 183:1767-1774.

58. Quinaud M, Chabert J, Faudry E, Neumann E, Lemaire D, Pastor A, Elsen S, Dessen A, Attree I. 2005. The PscE-PscF-PscG complex controls type III secretion needle biogenesis in Pseudomonas aeruginosa. J Biol Chem 280:36293-36300.

59. He SY, Nomura K, Whittam TS. 2004. Type III protein secretion mechanism in mammalian and plant pathogens. Biochimica et biophysica acta 1694:181-206.

60. Hueck CJ. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev 62:379-433.

61. Frithz-Lindsten E, Du Y, Rosqvist R, Forsberg A. 1997. Intracellular targeting of exoenzyme S of Pseudomonas aeruginosa via type III-dependent translocation induces phagocytosis resistance, cytotoxicity and disruption of actin microfilaments. Molecular microbiology 25:1125-1139.

62. Garrity-Ryan L, Kazmierczak B, Kowal R, Comolli J, Hauser A, Engel JN. 2000. The arginine finger domain of ExoT contributes to actin cytoskeleton disruption and inhibition of internalization of Pseudomonas aeruginosa by epithelial cells and macrophages. Infect Immun 68:7100-7113.

63. Cowell BA, Chen DY, Frank DW, Vallis AJ, Fleiszig SM. 2000. ExoT of cytotoxic Pseudomonas aeruginosa prevents uptake by corneal epithelial cells. Infect Immun 68:403-406.

64. Yahr TL, Vallis AJ, Hancock MK, Barbieri JT, Frank DW. 1998. ExoY, an adenylate cyclase secreted by the Pseudomonas aeruginosa type III system. Proceedings of the National Academy of Sciences of the United States of America 95:13899-13904.

65. Vallis AJ, Finck-Barbancon V, Yahr TL, Frank DW. 1999. Biological effects of Pseudomonas aeruginosa type III-secreted proteins on CHO cells. Infect Immun 67:2040-2044.

66. Cowell BA, Evans DJ, Fleiszig SM. 2005. Actin cytoskeleton disruption by ExoY and its effects on Pseudomonas aeruginosa invasion. FEMS Microbiol Lett 250:71-76.

67. Sayner SL, Frank DW, King J, Chen H, VandeWaa J, Stevens T. 2004. Paradoxical cAMP-induced lung endothelial hyperpermeability revealed by Pseudomonas aeruginosa ExoY. Circ Res 95:196-203.

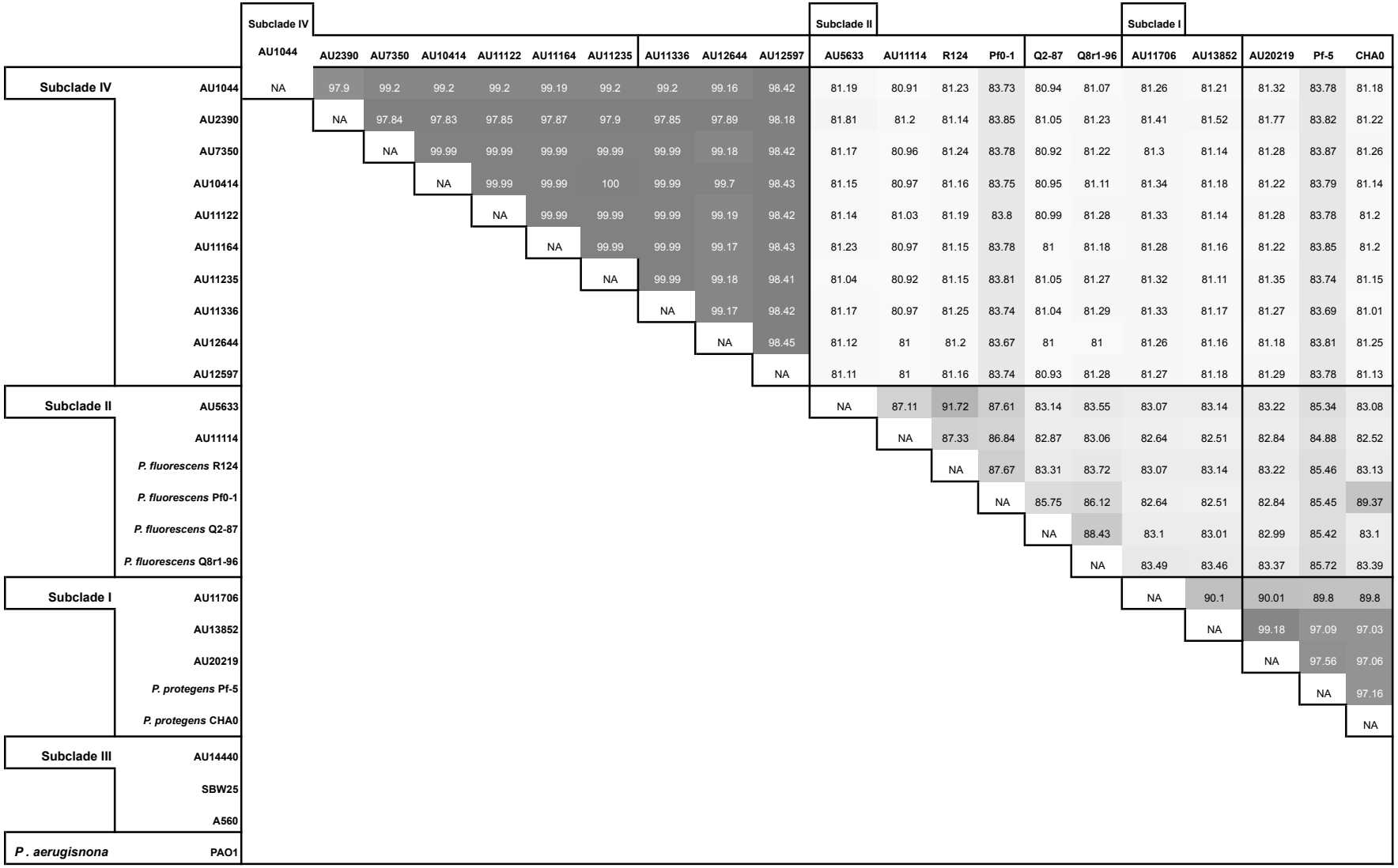
68. Sato H, Frank DW, Hillard CJ, Feix JB, Pankhaniya RR, Moriyama K, Finck-Barbancon V, Buchaklian A, Lei M, Long RM, Wiener-Kronish J, Sawa T. 2003. The mechanism of action of the Pseudomonas aeruginosa-encoded type III cytotoxin, ExoU. Embo J 22:2959-2969.

69. Phillips RM, Six DA, Dennis EA, Ghosh P. 2003. In vivo phospholipase activity of the Pseudomonas aeruginosa cytotoxin ExoU and protection of mammalian cells with phospholipase A2 inhibitors. J Biol Chem 278:41326-41332.

70. Malangoni MA, Crafton R, Mocek FC. 1994. Pneumonia in the surgical intensive care unit: factors determining successful outcome. Am J Surg 167:250-255.

71. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Hoiby N, Molin S. 2012. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Microbiol 10:841-851.

Appendix Figure V.1: Average Nucleoeotide Identity (ANI).

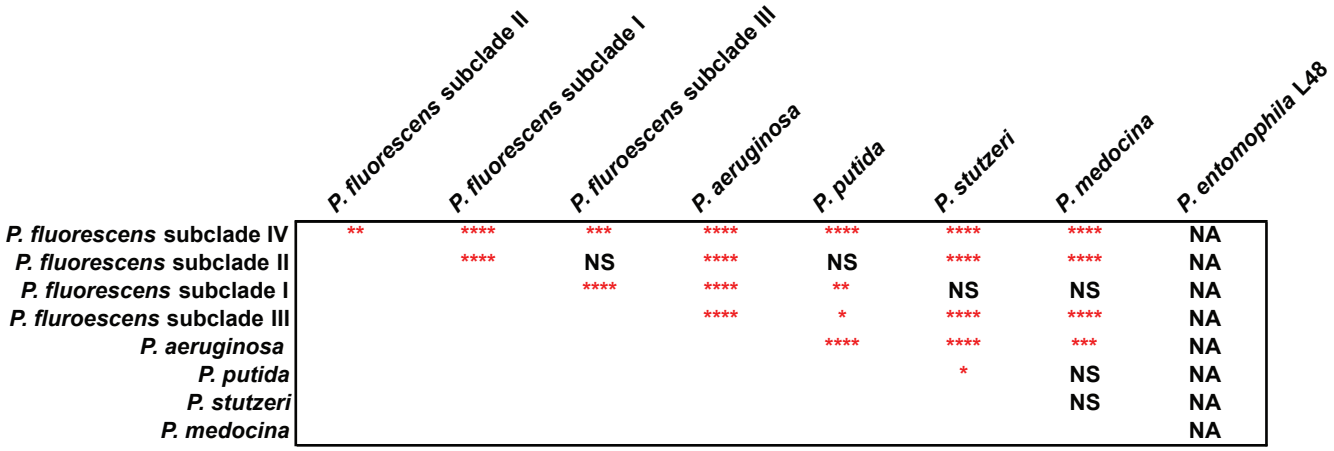


		Subclade IV										Subclade II						Subclade I				
		AU1044	AU2390	AU7350	AU10414	AU11122	AU11164	AU11235	AU11336	AU12644	AU12597	AU5633	AU11114	R124	Pf0-1	Q2-87	Q8r1-96	AU11706	AU13852	AU20219	Pf-5	CHA0
Subclade IV	AU1044	NA	97.9	99.2	99.2	99.2	99.19	99.2	99.2	99.16	98.42	81.19	80.91	81.23	83.73	80.94	81.07	81.26	81.21	81.32	83.78	81.18
	AU2390		NA	97.84	97.83	97.85	97.87	97.9	97.85	97.89	98.18	81.81	81.2	81.14	83.85	81.05	81.23	81.41	81.52	81.77	83.82	81.22
	AU7350			NA	99.99	99.99	99.99	99.99	99.99	99.18	98.42	81.17	80.96	81.24	83.78	80.92	81.22	81.3	81.14	81.28	83.87	81.26
	AU10414				NA	99.99	99.99	100	99.99	99.7	98.43	81.15	80.97	81.16	83.75	80.95	81.11	81.34	81.18	81.22	83.79	81.14
	AU11122					NA	99.99	99.99	99.99	99.19	98.42	81.14	81.03	81.19	83.8	80.99	81.28	81.33	81.14	81.28	83.78	81.2
	AU11164						NA	99.99	99.99	99.17	98.43	81.23	80.97	81.15	83.78	81	81.18	81.28	81.16	81.22	83.85	81.2
	AU11235							NA	99.99	99.18	98.41	81.04	80.92	81.15	83.81	81.05	81.27	81.32	81.11	81.35	83.74	81.15
	AU11336								NA	99.17	98.42	81.17	80.97	81.25	83.74	81.04	81.29	81.33	81.17	81.27	83.69	81.01
	AU12644									NA	98.45	81.12	81	81.2	83.67	81	81	81.26	81.16	81.18	83.81	81.25
	AU12597										NA	81.11	81	81.16	83.74	80.93	81.28	81.27	81.18	81.29	83.78	81.13
Subclade II	AU5633											NA	87.11	91.72	87.61	83.14	83.55	83.07	83.14	83.22	85.34	83.08
	AU11114												NA	87.33	86.84	82.87	83.06	82.64	82.51	82.84	84.88	82.52
	P. fluorescens R124													NA	87.67	83.31	83.72	83.07	83.14	83.22	85.46	83.13
	P. fluorescens Pf0-1														NA	85.75	86.12	82.64	82.51	82.84	85.45	89.37
	P. fluorescens Q2-87															NA	88.43	83.1	83.01	82.99	85.42	83.1
	P. fluorescens Q8r1-96																NA	83.49	83.46	83.37	85.72	83.39
Subclade I	AU11706																	NA	90.1	90.01	89.8	89.8
	AU13852																		NA	99.18	97.09	97.03
	AU20219																			NA	97.56	97.06
	P. protegens Pf-5																				NA	97.16
	P. protegens CHA0																					NA
Subclade III	AU14440																					
	SBW25																					
	A560																					
P. aerugisnona	PAO1																					

Appendix Figure V.1: Average Nuceloeotide Identity (ANI).

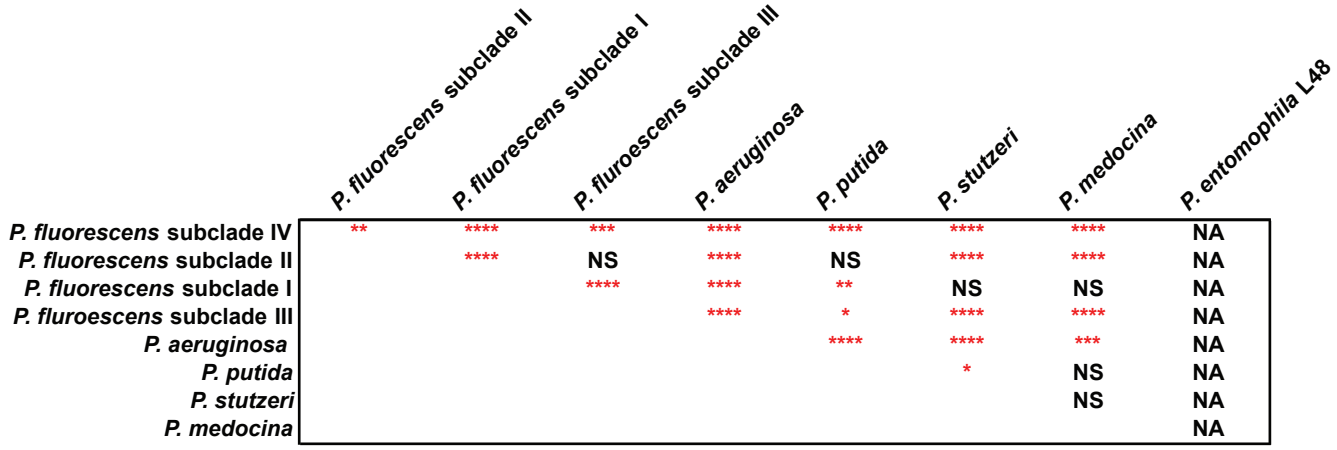
		Subclade III			P. aeruginosa
		AU14440	SBW25	A560	PAO1
Subclade IV	AU1044	83.81	83.74	83.76	77.94
	AU2390	84.11	83.96	83.77	77.88
	AU7350	83.76	83.58	83.81	77.91
	AU10414	83.85	83.7	83.91	77.79
	AU11122	83.83	83.73	83.84	77.85
	AU11164	83.89	83.7	83.93	77.84
	AU11235	83.84	83.71	83.92	77.8
	AU11336	83.95	83.69	83.93	77.84
	AU12644	84.12	83.74	83.81	77.74
	AU12597	84.2	83.72	83.62	77.85
Subclade II	AU5633	84.88	85.01	84.91	78.71
	AU11114	84.84	84.86	84.68	78.28
	P. fluorescens R124	84.93	85.12	85.03	78.94
	P. fluorescens Pf0-1	85.05	85.16	85.02	79
	P. fluorescens Q2-87	85.04	85.04	85.1	79.13
	P. fluorescens Q8r1-96	85.22	85.24	85.35	79.16
Subclade I	AU11706	84.99	85.08	84.96	79.54
	AU13852	85.03	85	84.92	79.48
	AU20219	85.08	84.99	84.93	79.46
	P. protegens Pf-5	85.06	85.04	85.05	79.65
	P. protegens CHA0	85	85.06	85.06	79.66
Subclade III	AU14440	NA	88.84	95.37	78.70
	SBW25		NA	88.8	78.7
	A560			NA	78.81
P . aerugisnona	PAO1				NA

Appendix Figure V.2: Genome size of newly sequenced and reference *Pseudomonas* strains.



Appendix Figure V.2: Genome size of newly sequenced and reference *Pseudomonas* strains.
The number of nucleotides in the genomes of newly sequenced and reference *Pseudomonas* strains. One-way ANOVA with Tukey's multiple comparison post test was used to determine statistical relevance. All groups are statistically different from subclade IV except *P. stutzeri* and *P. mendocina*, as seen in corresponding **Figure V.6.** NS = Not Significant; *p <0.05; **p <0.01; ***p <0.001; ****p <0.00001. One-way ANOVA cannot be applied to a single data point so *P. entomophila* L48 was not included in the statistical analysis (indicated by NA).

Appendix File V.3: % G+C content of newly sequenced and reference *Pseudomonas* strains**.**



Appendix File V.3: % G+C content of newly sequenced and reference *Pseudomonas* strains.
The number of G+C occurrences, as percentage of total genome, in newly sequenced and reference *Pseudomonas* strains. One-way ANOVA with Tukey's multiple comparisonpost test was used to determine statistical relevance. All groups are statistically different from subclade IV, as seen in corresponding **Figure V.5**. NS = Not Significant; *p <0.05; **p <0.01; ***p <0.001; ****p <0.00001. One-way ANOVA cannot be applied to a single data point so *P. entomophila* L48 was not included in the statisticalanalysis (indicated by NA).

Appendix Figure V.4: Nucleotide blast of selected GC Islands.

Subclade IV										
Strain	GC Island	Start	Stop	Length	GC%	Description	Query cover	E value	Identity	Accession
AU1044	GCI1	1310001	1324540	14540	44.57	Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_51	97%	0	100%	LN483141.1
						Amycolatopsis lurida NRRL 2430, complete genome	97%	0	100%	CP007219.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	97%	0	100%	LK996026.1
						Haemonchus placei genome assembly H_placei_MHpl1, scaffold HPLM_contig0002728	47%	0	100%	LM594107.1
						Strongyloides papillosus genome assembly S_papillosus_LIN, scaffold SPAL_contig0001419	74%	0	100%	LM527081.1
						Parastrongyloides trichosuri genome assembly P_trichosuri_KNP, scaffold PTRK_contig0000177	97%	0	100%	LM523540.1
						Enterobius vermicularis genome assembly E_vermicularis_Canary_Islands, scaffold EVEC_contig0000548	68%	0	100%	LM422671.1
						Hymenolepis diminuta genome assembly H_diminuta_Denmark, scaffold HDID_contig0004437	61%	0	100%	LM389718.1
						Taenia asiatica genome assembly T_asiatica_South_Korea, scaffold TASK_contig0001553	89%	0	100%	LM128679.1
Strain	GC Island	Start	Stop	Length	GC%	Echinostoma caproni genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929	97%	0	100%	LL266921.1
AU1044	GCI2	1457070	1466000	8931	43.11	Description	Query cover	E value	Identity	Accession
						Enterobacter cloacae strain GGT036, complete genome	3%	6.00E-39	77%	CP009756.1
						Enterobacter cloacae subsp. dissolvens SDM, complete genome	3%	6.00E-39	77%	CP003678.1
						Enterobacter cloacae subsp. cloacae ATCC 13047, complete genome	3%	6.00E-39	77%	CP001918.1
						Escherichia fergusonii ATCC 35469 chromosome, complete genome	3%	1.00E-30	76%	CU928158.2
						Escherichia coli O55:H7 str. RM12579, complete genome	2%	4.00E-26	77%	CP003109.1
						Escherichia coli O55:H7 str. CB9615, complete genome	2%	4.00E-26	77%	CP001846.1
						Escherichia coli O55:H7 DNA, O55-antigen biosynthesis gene clusters and their flanking regions, strain: TB182A	2%	4.00E-26	77%	AB453015.1
						Escherichia coli O55:H6 DNA for O55-antigen biosynthesis gene clusters and their flanking regions, strain: ICC219	2%	4.00E-26	77%	AB353133.1
						Escherichia coli O55:H7 DNA for O55-antigen biosynthesis gene clusters and their flanking regions, strain: WC416	2%	4.00E-26	77%	AB353132.1
Strain	GC Island	Start	Stop	Length	GC%	Escherichia coli serotype O55:H7 O-antigen gene cluster, complete sequence	2%	4.00E-26	77%	AF461121.1
AU1044	GCI3	1740360	1773414	33055	47.93	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens CHA0, complete genome	9%	0	83%	CP003190.1
						Pseudomonas sp. St29 DNA, complete genome	9%	0	83%	AP014628.1
						Pseudomonas sp. Os17 DNA, complete genome	9%	0	83%	AP014627.1

						Pseudomonas protegens Cab57 DNA, complete genome	9%	0	83%	AP014522.1
						Pseudomonas protegens Pf-5, complete genome	9%	0	83%	CP000076.1
						Pseudomonas sp. UW4, complete genome	9%	0	82%	CP003880.1
						Pseudomonas chlororaphis strain UFB2, complete genome	9%	0	83%	CP011020.1
						Pseudomonas chlororaphis strain PCL1606, complete genome	9%	0	82%	CP011110.1
						Pseudomonas sp. MRSN12121, complete genome	9%	0	82%	CP010892.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas fluorescens Pf29Arp contig191, whole genome shotgun sequence	9%	0	82%	ANOR01000191.1
AU2390	GCI1	1905308	1950300	44993	51.36	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens strain PCL1751, complete genome	12%	0	92%	CP010896.1
						Pseudomonas fluorescens A506, complete genome	12%	0	93%	CP003041.1
						Pseudomonas poae RE*1-1-14, complete genome	12%	0	92%	CP004045.1
						Pseudomonas entomophila str. L48 chromosome,complete sequence	9%	0	83%	CT573326.1
						Pseudomonas chlororaphis strain PA23, complete genome	11%	0	86%	CP008696.1
						Pseudomonas sp. UW4, complete genome	3%	0	86%	CP003880.1
						Pseudomonas alkylphenolia strain KL28, complete genome	2%	0	82%	CP009048.1
						Pseudomonas aeruginosa pathogenicity island PAGI-1 gene, complete sequence	1%	0	86%	AF241171.1
						Pseudomonas aeruginosa DNA, complete genome, strain: NCGM 1984	1%	0	86%	AP014646.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas aeruginosa DNA, complete genome, strain: NCGM 1900	1%	0	86%	AP014622.1
AU2390	GCI2	5386417	5407129	20713	50.42	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens F113, complete genome	46%	0	99%	CP003150.1
						Pseudomonas brassicacearum strain DF41, complete genome	38%	0	99%	CP007410.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	22%	0	87%	CP002585.1
						Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_51	11%	0	100%	LN483141.1
						Enterobacteria phage phiX174, complete genome	11%	0	100%	CP004084.1
						Amycolatopsis lurida NRRL 2430, complete genome	11%	0	100%	CP007219.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	11%	0	100%	LK996026.1
						Strongyloides papillosus genome assembly S_papillosus_LIN, scaffold SPAL_contig0001419	11%	0	100%	LM527081.1
						Parastrongyloides trichosuri genome assembly P_trichosuri_KNP, scaffold PTRK_contig0000177	11%	0	100%	LM523540.1
Strain	GC Island	Start	Stop	Length	GC%	Echinostoma caproni genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929	11%	0	100%	LL266921.1
AU7350	GCI1	35316	61682	26367	48.46	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens F113, complete genome	9%	0	93%	CP003150.1
						Pseudomonas sp. UW4, complete genome	9%	0	93%	CP003880.1
						Pseudomonas brassicacearum strain DF41, complete genome	9%	0	93%	CP007410.1

						Pseudomonas mandelii JR-1, complete genome	17%	0	93%	CP005960.1
						Pseudomonas fluorescens strain UK4, complete genome	16%	0	92%	CP008896.1
						Pseudomonas chlororaphis strain PA23, complete genome	16%	0	92%	CP008696.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	17%	0	92%	CP002585.1
						Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	8%	0	92%	CP009290.1
						Pseudomonas fluorescens Pf29Arp contig107, whole genome shotgun sequence	8%	0	92%	ANOR01000107.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas protegens Cab57 DNA, complete genome	8%	0	92%	AP014522.1
AU7350	GCI2	4116063	4149149	33087	47.92	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens CHA0, complete genome	9%	0	83%	CP003190.1
						Pseudomonas sp. St29 DNA, complete genome	9%	0	82%	AP014628.1
						Pseudomonas sp. Os17 DNA, complete genome	9%	0	82%	AP014627.1
						Pseudomonas protegens Cab57 DNA, complete genome	9%	0	82%	AP014522.1
						Pseudomonas protegens Pf-5, complete genome	9%	0	82%	CP000076.1
						Pseudomonas sp. UW4, complete genome	9%	0	82%	CP003880.1
						Pseudomonas chlororaphis strain UFB2, complete genome	9%	0	83%	CP011020.1
						Pseudomonas fluorescens Pf29Arp contig191, whole genome shotgun sequence	9%	0	82%	ANOR01000191.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	9%	0	82%	CP002585.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas fluorescens F113, complete genome	9%	0	81%	CP003150.1
AU7350	GCI3	4258470	4304008	45539	49.12	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens Pf-5, complete genome	7%	0	84%	CP000076.1
						Pseudomonas protegens Cab57 DNA, complete genome	7%	0	84%	AP014522.1
						Pseudomonas protegens CHA0, complete genome	7%	0	84%	CP003190.1
						Pseudomonas mandelii JR-1, complete genome	7%	0	85%	CP005960.1
						Pseudomonas sp. MRSN12121, complete genome	7%	0	84%	CP010892.1
						Pseudomonas chlororaphis strain PCL1606, complete genome	7%	0	84%	CP011110.1
						Pseudomonas sp. St29 DNA, complete genome	8%	0	83%	AP014628.1
						Pseudomonas alkylphenolia strain KL28, complete genome	7%	0	82%	CP009048.1
						Pseudomonas chlororaphis strain PA23, complete genome	7%	0	87%	CP008696.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	7%	0	87%	CP009290.1
AU10414	GCI1	4731805	4761398	29594	46.72	Description	Query cover	E value	Identity	Accession
						Pseudomonas chlororaphis strain UFB2, complete genome	7%	2.00E-91	69%	CP011020.1
*note: had to use 'more disssimilar sequences (discontiguous megablast)						Pseudomonas cichorii JBC1, complete genome	5%	1.00E-86	66%	CP007039.1
						Pseudomonas syringae pv. phaseolicola 1448A, complete genome	6%	3.00E-81	68%	CP000058.1
						Pseudomonas syringae pv. syringae HS191, complete genome	10%	7.00E-77	68%	CP006256.1
						Pseudomonas syringae pv. syringae B728a, complete	9%	3.00E-75	68%	CP000075.1

						genome				
						Pseudomonas syringae pv. tomato str. DC3000, complete genome	6%	5.00E-73	68%	AE016853.1
						Pseudomonas syringae pv. syringae B301D, complete genome	9%	2.00E-71	67%	CP005969.1
						Pseudomonas putida DOT-T1E, complete genome	6%	4.00E-67	67%	CP003734.1
						Pseudomonas putida KT2440 complete genome	6%	2.00E-65	68%	AE015451.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas syringae CC1557, complete sequence	5%	6.00E-65	65%	CP007014.1
AU10414	GCI2	4833242	487195	3954	46.16	Description	Query cover	E value	Identity	Accession
						Pseudomonas sp. TKP, complete genome	10%	8.00E-169	93%	CP006852.1
						Pseudomonas fluorescens SBW25 complete genome	10%	3.00E-158	92%	AM181176.4
						Pseudomonas fluorescens strain PCL1751, complete genome	10%	2.00E-155	91%	CP010896.1
						Pseudomonas entomophila str. L48 chromosome,complete sequence	7%	3.00E-48	80%	CT573326.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas putida HB3267, complete genome	1%	2.00E-20	95%	CP003738.1
AU11122	GCI1	1461209	1489825	28617	44.54	Description	Query cover	E value	Identity	Accession
						Enterobacteria phage phiX174, complete genome	95%	0	99%	CP004084.1
						Amycolatopsis lurida NRRL 2430, complete genome	95%	0	99%	CP007219.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	95%	0	99%	LK996026.1
						Protopolystoma xenopodis genome assembly P_xenopodis_South_Africa, scaffold PXEA_scaffold0015595	76%	0	99%	LM703067.1
						Onchocerca flexuosa genome assembly O_flexuosa_Cordoba, scaffold OFLC_scaffold0001064	92%	0	99%	LM540398.1
						Escherichia coli genome assembly FHI89, scaffold scaffold-8_contig-14.0_1_3588_[organism:Escherichia	63%	0	99%	LM997026.1
						Escherichia coli genome assembly FHI87, scaffold scaffold-4_contig-19.0_1_5337_[organism:Escherichia	95%	0	99%	LM996980.1
						Escherichia coli genome assembly FHI24, scaffold scaffold-8_contig-33.0_1_3324_[organism:Escherichia	52%	0	99%	LM995715.1
						Acinetobacter baumannii AC12, complete genome	96%	0	99%	CP007549.1
Strain	GC Island	Start	Stop	Length	GC%	Coliphage phi-X174, complete genome	95%	0	99%	J02482.1
AU11122	GCI2	2249703	2279287	29585	46.73	Description	Query cover	E value	Identity	Accession
						Pseudomonas chlororaphis strain UFB2, complete genome	7%	2.00E-91	69%	CP011020.1
						Pseudomonas cichorii JBC1, complete genome	5%	1.00E-86	66%	CP007039.1
						Pseudomonas syringae pv. phaseolicola 1448A, complete genome	6%	3.00E-81	68%	CP000058.1
						Pseudomonas syringae pv. syringae HS191, complete genome	10%	7.00E-77	68%	CP006256.1
						Pseudomonas syringae pv. syringae B728a, complete genome	9%	3.00E-75	68%	CP000075.1

						Pseudomonas syringae pv. tomato str. DC3000, complete genome	6%	5.00E-73	68%	AE016853.1
						Pseudomonas syringae pv. syringae B301D, complete genome	9%	2.00E-71	67%	CP005969.1
						Pseudomonas putida DOT-T1E, complete genome	6%	4.00E-67	67%	CP003734.1
						Pseudomonas putida KT2440 complete genome	6%	2.00E-65	68%	AE015451.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas syringae CC1557, complete sequence	5%	6.00E-65	65%	CP007014.1
AU11122	GCI3	2395906	2402797	6892	43.98	Description	Query cover	E value	Identity	Accession
						Pseudomonas cichorii JBC1, complete genome	2%	2.00E-42	84%	CP007039.1
Strain	GC Island	Start	Stop	Length	GC%	Cupriavidus metallidurans CH34 megaplasmid, complete sequence	2%	5.00E-19	78%	CP000353.2
AU11122	GCI4	2790740	2818869	28130	49.08	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens F113, complete genome	11%	0	93%	CP003150.1
						Pseudomonas sp. UW4, complete genome	14%	0	93%	CP003880.1
						Pseudomonas brassicacearum strain DF41, complete genome	11%	0	93%	CP007410.1
						Pseudomonas mandelii JR-1, complete genome	17%	0	93%	CP005960.1
						Pseudomonas fluorescens strain UK4, complete genome	18%	0	92%	CP008896.1
						Pseudomonas chlororaphis strain PA23, complete genome	18%	0	92%	CP008696.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	17%	0	92%	CP002585.1
						Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	8%	0	92%	CP009290.1
						Pseudomonas fluorescens Pf29Arp contig107, whole genome shotgun sequence	8%	0	92%	ANOR01000107.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas protegens Cab57 DNA, complete genome	8%	0	92%	AP014522.1
AU11122	GCI5	3202879	325479	49601	49.34	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens Pf-5, complete genome	7%	0	84%	CP000076.1
						Pseudomonas protegens Cab57 DNA, complete genome	7%	0	84%	AP014522.1
						Pseudomonas protegens CHA0, complete genome	7%	0	84%	CP003190.1
						Pseudomonas mandelii JR-1, complete genome	7%	0	85%	CP005960.1
						Pseudomonas sp. MRSN12121, complete genome	7%	0	84%	CP010892.1
						Pseudomonas chlororaphis strain PCL1606, complete genome	7%	0	84%	CP011110.1
						Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	11%	0	87%	CP009290.1
						Pseudomonas sp. St29 DNA, complete genome	10%	0	83%	AP014628.1
						Pseudomonas alkylphenolia strain KL28, complete genome	9%	0	82%	CP009048.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas chlororaphis strain PA23, complete genome	10%	0	87%	CP008696.1
AU11164	GCI1	1	5319	5319	45.93	Description	Query cover	E value	Identity	Accession
						Pseudomonas entomophila str. L48 chromosome,complete sequence	10%	5.00E-48	80%	CT573326.1
						Pseudomonas cichorii JBC1, complete genome	3%	5.00E-33	85%	CP007039.1
						Pseudomonas putida HB3267, complete genome	1%	3.00E-20	95%	CP003738.1
						Erwinia sp. Ejp617, complete genome	1%	5.00E-13	91%	CP002124.1

						Erwinia amylovora ATCC 49946 chromosomal sequence	1%	5.00E-13	90%	FN666575.1
						Erwinia amylovora ATCC BAA-2158, whole genome shotgun sequence, contig 13	1%	2.00E-11	89%	FR719193.1
						Erwinia amylovora CFBP1430 complete genome	1%	2.00E-11	90%	FN434113.1
						Erwinia amylovora ATCC BAA-2158, whole genome shotgun sequence, contig 15	1%	1.00E-08	87%	FR719195.1
Strain	GC Island	Start	Stop	Length	GC%	Erwinia tasmaniensis strain ET1/99 complete chromosome	3%	2.00E-06	84%	CU468135.1
AU11164	GCI2	77172	106756	29585	46.73	Description	Query cover	E value	Identity	Accession
						Pseudomonas chlororaphis strain UFB2, complete genome	7%	2.00E-91	69%	CP011020.1
*note: had to use 'more disssimilar sequences (discontiguous megablast)						Pseudomonas cichorii JBC1, complete genome	5%	1.00E-86	66%	CP007039.1
						Pseudomonas syringae pv. phaseolicola 1448A, complete genome	6%	3.00E-81	68%	CP000058.1
						Pseudomonas syringae pv. syringae HS191, complete genome	10%	7.00E-77	68%	CP006256.1
						Pseudomonas syringae pv. syringae B728a, complete genome	9%	3.00E-75	68%	CP000075.1
						Pseudomonas syringae pv. tomato str. DC3000, complete genome	6%	5.00E-73	68%	AE016853.1
						Pseudomonas syringae pv. syringae B301D, complete genome	9%	2.00E-71	67%	CP005969.1
						Pseudomonas putida DOT-T1E, complete genome	6%	4.00E-67	67%	CP003734.1
						Pseudomonas putida KT2440 complete genome	6%	2.00E-65	68%	AE015451.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas syringae CC1557, complete sequence	5%	6.00E-65	65%	CP007014.1
AU11164	GCI3	409676	43450	24675	48.18	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens strain UK4, complete genome	10%	0%	79%	CP008896.1
						Pseudomonas mandelii JR-1, complete genome	11%	0%	78%	CP005960.1
						Pseudomonas chlororaphis strain PA23, complete genome	10%	0%	77%	CP008696.1
						Pseudomonas entomophila str. L48 chromosome,complete sequence	7%	0%	77%	CT573326.1
						Pseudomonas fulva 12-X, complete genome	9%	0%	77%	CP002727.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	11%	0%	77%	CP002585.1
						Pseudomonas stutzeri RCH2, complete genome	15%	0%	76%	CP003071.1
						Pseudomonas putida S13.1.2, complete genome	7%	0%	76%	CP010979.1
						Pseudomonas rhizosphaerae strain DSM 16299, complete genome	10%	0%	76%	CP009533.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas stutzeri DSM 10701, complete genome	9%	0%	75%	CP003725.1
AU11164	GCI4	4739996	4791014	51019	49.66	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens Pf-5, complete genome	7%	0	84%	CP000076.1
						Pseudomonas protegens Cab57 DNA, complete genome	7%	0	84%	AP014522.1
						Pseudomonas protegens CHA0, complete genome	7%	0	84%	CP003190.1
						Pseudomonas mandelii JR-1, complete genome	7%	0	85%	CP005960.1
						Pseudomonas sp. MRSN12121, complete genome	6%	0	84%	CP010892.1
						Pseudomonas chlororaphis strain PCL1606, complete genome	6%	0	84%	CP011110.1

						Pseudomonas sp. St29 DNA, complete genome	7%	0	83%	AP014628.1
						Pseudomonas alkylphenolia strain KL28, complete genome	7%	0	82%	CP009048.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas chlororaphis strain PA23, complete genome	7%	0	87%	CP008696.1
AU11235	GCI1	1	6277	6277	48.72	Description	Query cover	E value	Identity	Accession
						Pseudomonas entomophila str. L48 chromosome,complete sequence	0%	3.00E-11	98%	CT573326.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas sp. VLB120, complete genome	0%	6.00E-08	92%	CP003961.1
AU11336	GCI1	1121581	1151165	29585	46.73	Description	Query cover	E value	Identity	Accession
						Pseudomonas chlororaphis strain UFB2, complete genome	7%	2.00E-91	69%	CP011020.1
						Pseudomonas cichorii JBC1, complete genome	5%	1.00E-86	66%	CP007039.1
						Pseudomonas syringae pv. phaseolicola 1448A, complete genome	6%	3.00E-81	68%	CP000058.1
						Pseudomonas syringae pv. syringae HS191, complete genome	10%	7.00E-77	68%	CP006256.1
						Pseudomonas syringae pv. syringae B728a, complete genome	9%	3.00E-75	68%	CP000075.1
						Pseudomonas syringae pv. tomato str. DC3000, complete genome	6%	5.00E-73	68%	AE016853.1
						Pseudomonas syringae pv. syringae B301D, complete genome	9%	2.00E-71	67%	CP005969.1
						Pseudomonas putida DOT-T1E, complete genome	6%	4.00E-67	67%	CP003734.1
						Pseudomonas putida KT2440 complete genome	6%	2.00E-65	68%	AE015451.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas syringae CC1557, complete sequence	5%	6.00E-65	65%	CP007014.1
AU11336	GCI2	4808967	4858519	49553	49.72	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens strain UK4, complete genome	7%	0	98%	CP008896.1
						Pseudomonas fluorescens strain LBUM223, complete genome	2%	0	83%	CP011117.1
						Pseudomonas mandelii JR-1 plasmid, complete sequence	2%	0	86%	CP005961.1
						Uncultured proteobacterium QS1 genomic sequence	1%	2.00E-164	84%	AY688432.2
						Pseudomonas putida strain DLL-E4, complete genome	2%	5.00E-151	77%	CP007620.1
						Pseudomonas aeruginosa strain 131/2010 insertion sequence ISPa26, complete sequence; and outer membrane porin protein (oprD) gene, partial cds	1%	5.00E-151	77%	KF682462.1
						Pseudomonas monteilii SB3101, complete genome	2%	5.00E-151	77%	CP006979.1
						Pseudomonas monteilii SB3078, complete genome	2%	5.00E-151	77%	CP006978.1
						Pseudomonas putida S16, complete genome	2%	5.00E-151	77%	CP002870.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas aeruginosa outer membrane porin protein (oprD) gene, partial sequence; and insertion sequence ISPa25 TnpA (tnpA) gene, complete cds	1%	5.00E-151	77%	EU000222.1
AU12597	GCI1	1268143	1286773	18631	44.11	Description	Query cover	E value	Identity	Accession
						Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_51	98%	0	99%	LN483141.1
						Enterobacteria phage phiX174, complete genome	98%	0	99%	CP004084.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	98%	0	99%	LK996026.1

						Strongyloides papillosus genome assembly S_papillosus_LIN, scaffold SPAL_contig0001419	91%	0	99%	LM527081.1
						Parastrongyloides trichosuri genome assembly P_trichosuri_KNP, scaffold PTRK_contig0000177	98%	0	99%	LM523540.1
						Echinostoma caproni genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929	98%	0	99%	LL266921.1
						Hymenolepis nana genome assembly H_nana_Japan, scaffold HNAJ_contig0004412	98%	0	99%	LM405563.1
						Toxocara canis genome assembly T_canis_Equador, scaffold TCNE_contig0003098	92%	0	99%	LM047136.1
						Escherichia coli genome assembly FHI92, scaffold scaffold-6_contig-18.0_1_5295_[organism:Escherichia	97%	0	99%	LM997153.1
Strain	GC Island	Start	Stop	Length	GC%	Sphingobacterium sp. PM2-P1-29 genome assembly Sequencing method, scaffold BN1088_Contig_19	98%	0	99%	LK931771.1
AU12644	GCI1	1	8248	8248	44.22	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens CHA0, complete genome	8%	0	93%	CP003190.1
						Pseudomonas protegens Cab57 DNA, complete genome	8%	0	92%	AP014522.1
						Pseudomonas sp. St29 DNA, complete genome	7%	0	90%	AP014628.1
						Pseudomonas sp. Os17 DNA, complete genome	7%	0	90%	AP014627.1
						Pseudomonas poae RE*1-1-14, complete genome	8%	0	89%	CP004045.1
						Pseudomonas sp. StFLB209 DNA, complete genome	2%	5.00E-64	88%	AP014637.1
						Pseudomonas stutzeri A1501, complete genome	1%	2.00E-39	86%	CP000304.1
						Pseudomonas putida transposon Tn4652 transposase (tnpA), putative tpnA repressor protein (tnpC), and putative cointegrate resolution proteins S and T genes	1%	3.00E-37	85%	AF151431.1
						Pseudomonas putida plasmid pWW0 complete genome	1%	9.00E-37	85%	AJ344068.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas putida KT2440 complete genome	1%	9.00E-37	85%	AE015451.1
AU12644	GCI2	3755914	3766388	10475	44.78	Description	Query cover	E value	Identity	Accession
						Pseudomonas syringae pv. phaseolicola 1448A, complete genome	1%	5.00E-45	97%	CP000058.1
						Pseudomonas cichorii JBC1, complete genome	1%	1.00E-41	96%	CP007039.1
						Pseudomonas syringae pv. persicae strain 5846 QueA gene, partial cds; tRNA-Leu gene, complete sequence; HrpK gene, partial cds; and unknown genes	1%	1.00E-41	96%	AY147018.1
						P.syringae insertion element IS53 DNA sequence	1%	1.00E-41	96%	M83932.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	1%	4.00E-36	95%	CP002585.1
						Pseudomonas mandelii JR-1 plasmid, complete sequence	0%	2.00E-35	95%	CP005961.1
						Pseudomonas putida H8234, complete genome	0%	5.00E-30	93%	CP005976.1
						Pseudomonas monteilii SB3101, complete genome	1%	3.00E-23	88%	CP006979.1
						Pseudomonas monteilii SB3078, complete genome	1%	3.00E-22	87%	CP006978.1
Strain	GC Island	Start	Stop	Length	GC%	Pseudomonas putida strain DLL-E4, complete genome	0%	1.00E-21	88%	CP007620.1

AU12644	GCI3	3880788	3907081	26294	48.46	Description	Query cover	E value	Identity	Accession
						Pseudomonas fluorescens F113, complete genome	9%	0	93%	CP003150.1
						Pseudomonas brassicacearum strain DF41, complete genome	9%	0	93%	CP007410.1
						Pseudomonas sp. UW4, complete genome	17%	0	93%	CP003880.1
						Pseudomonas mandelii JR-1, complete genome	16%	0	93%	CP005960.1
						Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	9%	0	93%	CP002585.1
						Pseudomonas fluorescens strain UK4, complete genome	15%	0	93%	CP008896.1
						Pseudomonas chlororaphis strain PA23, complete genome	15%	0	93%	CP008696.1
						Pseudomonas fluorescens Pf29Arp contig107, whole genome shotgun sequence	8%	0	92%	ANOR01000107.1
						Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	8%	0	92%	CP009290.1
						Pseudomonas protegens Pf-5, complete genome	8%	0	92%	CP000076.1
Subclade I										
Strain	GC Island	Start	Stop	Length	GC%					
AU11706	GCI1	2793301	2821744	28444	45.12	Description	Query cover	E value	Identity	Accession
						Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_51	89%	0	100%	LN483141.1
						Amycolatopsis lurida NRRL 2430, complete genome	89%	0	100%	CP007219.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	89%	0	100%	LK996026.1
						Onchocerca flexuosa genome assembly O_flexuosa_Cordoba, scaffold OFLC_scaffold0001064	86%	0	100%	LM540398.1
						Parastrongyloides trichosuri genome assembly P_trichosuri_KNP, scaffold PTRK_contig0000177	89%	0	100%	LM523540.1
						Enterobius vermicularis genome assembly E_vermicularis_Canary_Islands, scaffold EVEC_contig0000548	59%	0	100%	LM422671.1
						Hymenolepis diminuta genome assembly H_diminuta_Denmark, scaffold HDID_contig0004437	52%	0	100%	LM389718.1
						Taenia asiatica genome assembly T_asiatica_South_Korea, scaffold TASK_contig0001553	82%	0	100%	LM128679.1
						Echinostoma caproni genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929	89%	0	100%	LL266921.1
						Hymenolepis nana genome assembly H_nana_Japan, scaffold HNAJ_contig0004412	89%	0	100%	LM405563.1
Strain	GC Island	Start	Stop	Length	GC%	Toxocara canis genome assembly T_canis_Equador, scaffold TCNE_contig0003098	85%	0	100%	LM047136.1
AU13852	GCI2	6680368	6682707	2340	43.68	Description	Query cover	E value	Identity	Accession
						Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_51	94%	0	100%	LN483141.1
						Enterobacteria phage phiX174, complete genome	94%	0	100%	CP004084.1
						Amycolatopsis lurida NRRL 2430, complete genome	94%	0	100%	CP007219.1
						Desulfitobacterium hafniense genome assembly assembly_v1, strain PCS-E, scaffold scaffold9	94%	0	100%	LK996026.1

						Strongyloides papillosus genome assembly S_papillosus_LIN, scaffold SPAL_contig0001419	94%	0	100%	LM527081.1
						Parastrongyloides trichosuri genome assembly P_trichosuri_KNP, scaffold PTRK_contig0000177	94%	0	100%	LM523540.1
						Echinostoma caproni genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929	94%	0	100%	LL266921.1
						Hymenolepis nana genome assembly H_nana_Japan, scaffold HNAJ_contig0004412	94%	0	100%	LM405563.1
						Toxocara canis genome assembly T_canis_Equador, scaffold TCNE_contig0003098	94%	0	100%	LM047136.1
Strain	GC Island	Start	Stop	Length	GC%	Escherichia coli genome assembly FHI97, scaffold scaffold-9_contig-37.0_1_1617_[organism:Escherichia	70%	0	100%	LM997183.1
AU20219	GCI1	782655	898172	25518	45.78	Description	Query cover	E value	Identity	Accession
						Pseudomonas protegens CHA0, complete genome	12%	0	99%	CP003190.1
						Pseudomonas protegens Pf-5, complete genome	12%	0	99%	CP000076.1
						Pseudomonas protegens Cab57 DNA, complete genome	12%	0	99%	AP014522.1
						Pseudomonas sp. Os17 DNA, complete genome	12%	0	99%	AP014627.1
						Pseudomonas sp. St29 DNA, complete genome	12%	0	99%	AP014628.1
						Pseudomonas fluorescens A506, complete genome	12%	0	95%	CP003041.1
						Pseudomonas sp. WCS374 genome	12%	0	95%	CP007638.1
						Pseudomonas chlororaphis strain PA23, complete genome	11%	0	98%	CP008696.1
						Pseudomonas fluorescens strain LBUM223, complete genome	12%	0	95%	CP011117.1
Sublcade II										
Strain	GC Island	Start	Stop	Length	GC%	Description	Query cover	E value	Identity	Accession
AU11114	GCI1	1249884	1265038	15155	45.48	Streptococcus sp. VT 162, complete genome	90%	0	100%	NZ_CP007628.1
						Amycolatopsis lurida NRRL 2430, complete genome	90%	0	100%	CP007219.1
						Acinetobacter baumannii AC12, complete genome	90%	0	100%	CP007549.1
						Acinetobacter baumannii AC30, complete genome	14%	0	100%	CP007577.1
						Pseudomonas fluorescens Pf0-1, complete genome	8%	0	88%	NC_007492.2
						Pseudomonas brassicacearum strain DF41, complete genome	5%	7.00E-148	84%	NZ_CP007410.1
						Pseudomonas chlororaphis strain UFB2, complete genome	2%	1.00E-125	92%	CP011020.1
						Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	2%	4.00E-125	92%	NZ_CP009290.1
						Pseudomonas sp. UW4, complete genome	2%	4.00E-125	92%	NC_019670.1
						Pseudomonas mandelii JR-1, complete genome	2%	2.00E-124	92%	NZ_CP005960.1
Strain	GC Island	Start	Stop	Length	GC%	Description	Query cover	E value	Identity	Accession
AU11114	GCI2	1355409	1362687	7279	37.18	Pseudomonas sp. VLB120, complete genome	5%	1.00E-63	76%	NC_022738.1
						Pseudomonas putida W619, complete genome	2%	1.00E-14	72%	NC_010501.1

Appendix Figure V.4:Nucleotide blast of selected GC Islands.
The nucleotide sequence of selected GC Islands (indicated with arrows in **Figures V.8a** and **V.8b**) were used to query the NCBI nucleotide collection (nr/nt). The top six results are displayed.

Appendix Figure V.5: Annotated type III secretion genes in *P. fluorescens* subclades IV and *P. aeruginosa* strains.

Subclade IV Ysc/Yop-like T3SS

	AU1044	AU2390	AU7350	AU10414	AU11122	AU11235	AU11336	AU12597	AU12644
PcrD	x	x	x	x	x	x	x	x	x
PcrF	x	x	x	x	x	x	x	x	x
PcrG	x	x	x	x	x	x	x	x	x
PcrH	x	x	x	x	x	x	x	x	x
PcrR	x	x	x	x	x	x	x	x	x
PcrV	x	x	x	x	x	x	x	x	x
PopB	x	x	x	x	x	x	x	x	x
PopE									
PopN	x	x							
PopR	x	x	x	x	x	x	x	x	x
PscB	x	x	x	x	x	x	x	x	x
PscC	x	x	x	x	x	x	x	x	x
PscD	x	x	x	x	x	x	x	x	x
PscE									
PscF	x	x	x	x	x	x	x	x	x
PscG	x	x	x	x	x	x	x	x	x
PscH	x	x	x	x	x	x	x	x	x
PscI	x	x	x	x	x	x	x	x	x
PscJ	x	x	x	x	x	x	x	x	x
PscK	x	x	x	x	x	x	x	x	x
PscL	x	x	x	x	x	x	x	x	x
PscN	x	x	x	x	x	x	x	x	x
PscO	x	x	x	x	x	x	x	x	x
PscP	x	x	x	x	x	x	x	x	x
PscQ	x	x	x	x	x	x	x	x	x
PscR	x	x	x	x	x	x	x	x	x
PscS	x	x	x	x	x	x	x	x	x
PscT	x	x	x	x	x	x	x	x	x
PscU	x	x	x	x	x	x	x	x	x
PscW	x	x	x	x	x	x	x	x	x
PscZ	x	x	x	x	x	x	x	x	x
SctX	x	x	x	x	x	x	x	x	x
SycN	x	x	x	x	x	x	x	x	x
TyeA	x	x	x	x	x	x	x	x	x
ExoS									
ExoT									
ExoU	x	x	x	x	x	x	x	x	x

ExoY									
SpcS									
SpcU	x	x	x	x	x	x	x	x	x

Subclade IV Hrp1 T3SS

	AU1044	AU2390	AU7350	AU10414	AU11122	AU11235	AU11336	AU12597	AU12644
HrpA									
HrpB									
HrpD									
HrpF									
HrpG									
HrpJ									
HrpK1									
HrpP									
HrpQ									
HrpR									
HrpS									
HrpT									
HrpW									
HrpZ1									
HopPmaJ	x	x	x	x	x	x	x	x	x

P. aeruginosa Ysc/Yop-like T3SS

	PAO1	PA7	PA14	2192	LESB58
PcrD	x		x	x	x
PcrF	x		x	x	x
PcrG	x		x	x	x
PcrH	x		x	x	x
PcrR	x		x	x	x
PcrV	x		x	x	x
PopB	x		x	x	x
PopE	x		x	x	x
PopN	x		x	x	x
PopR	x		x	x	x
PscB	x		x	x	x
PscC	x		x	x	x
PscD	x		x	x	x
PscE	x		x	x	x
PscF	x		x	x	x
PscG	x		x	x	x
PscH	x		x	x	x
PscI	x		x	x	x
PscJ	x		x	x	x
PscK	x		x	x	x

PscL	x		x	x	x
PscN	x		x	x	x
PscO	x		x	x	x
PscP	x		x	x	x
PscQ	x		x	x	x
PscR	x		x	x	x
PscS	x		x	x	x
PscT	x		x	x	x
PscU	x		x	x	x
PscW	x		x	x	x
PscZ	x		x	x	x
SctX	x		x	x	x
SycN	x		x	x	x
TyeA	x		x	x	x
ExoS	x				x
ExoT	x		x		x
ExoU			x		
ExoY				x	x
SpcS	x				
SpcU			x		

P. aeruginosa Hrp1 T3SS

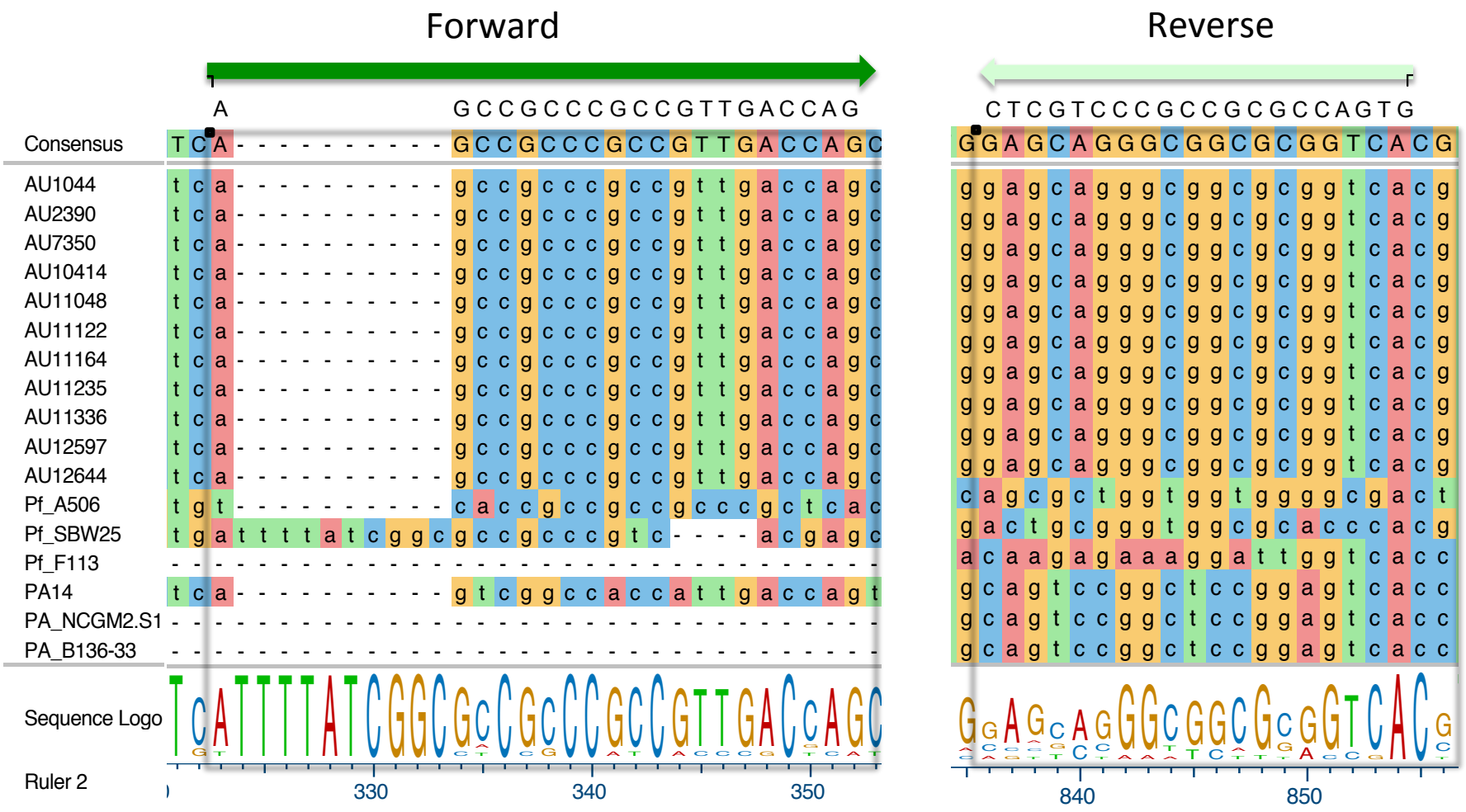
	PAO1	PA7	PA14	2192	LESB58
HrpA					
HrpB					
HrpD					
HrpF					
HrpG					
HrpJ					
HrpK1					
HrpP					
HrpQ					
HrpR					
HrpS					
HrpT					
HrpW					
HrpZ1					
HopPmaJ		x	x	x	x

Appendix Figure V.5: Annotated type III secretion genes in *P. fluorescens* subclades IV, II, I and *P. aeruginosa* strains.
Genes annotated as belonging to either Ysc/Yop- or the Hrp- like type III secretion system using the RAST online annotation software, as detailed in **Chapter II.** An 'x' indicates presence of gene in that genome.

Appendix Figure V.6: Statistics of annotated type III secretion genes.

Tukey's multiple comparisons test	Significant?	Summary
Yop System		
P. fluorescens subclade IV vs. P. fluorescens subclade III	Yes	****
P. fluorescens subclade IV vs. P. fluorescens subclade II	Yes	****
P. fluorescens subclade IV vs. *P. fluorescens* subclade I	Yes	****
P. fluorescens subclade IV vs. *P. aerguinosa*	No	ns
P. fluorescens subclade IV vs. *P. putida*	Yes	****
P. fluorescens subclade IV vs. *P. entomophila* L48	Yes	****
P. fluorescens subclade IV vs. *P. stutzeri* A1501	Yes	****
P. fluorescens subclade IV vs. *P. mendocina* ymp	Yes	***
P. fluorescens subclade III vs. P. fluorescens subclade II	No	ns
P. fluorescens subclade III vs. *P. fluorescens* subclade I	Yes	**
P. fluorescens subclade III vs. *P. aerguinosa*	Yes	****
P. fluorescens subclade III vs. *P. putida*	Yes	**
P. fluorescens subclade III vs. *P. entomophila* L48	No	ns
P. fluorescens subclade III vs. *P. stutzeri* A1501	No	ns
P. fluorescens subclade III vs. *P. mendocina* ymp	No	ns
P. fluorescens subclade II vs. *P. fluorescens* subclade I	No	ns
P. fluorescens subclade II vs. *P. aerguinosa*	Yes	****
P. fluorescens subclade II vs. *P. putida*	No	ns
P. fluorescens subclade II vs. *P. entomophila* L48	No	ns
P. fluorescens subclade II vs. *P. stutzeri* A1501	No	ns
P. fluorescens subclade II vs. *P. mendocina* ymp	No	ns
P. fluorescens subclade I vs. *P. aerguinosa*	Yes	****
P. fluorescens subclade I vs. *P. putida*	No	ns
P. fluorescens subclade I vs. *P. entomophila* L48	No	ns
P. fluorescens subclade I vs. *P. stutzeri* A1501	No	ns
P. fluorescens subclade I vs. *P. mendocina* ymp	No	ns
P. aerguinosa vs. *P. putida*	Yes	****
P. aerguinosa vs. *P. entomophila* L48	Yes	****
P. aerguinosa vs. *P. stutzeri* A1501	Yes	****
P. aerguinosa vs. *P. mendocina* ymp	Yes	*
P. putida vs. *P. entomophila* L48	No	ns
P. putida vs. *P. stutzeri* A1501	No	ns
P. putida vs. *P. mendocina* ymp	No	ns
P. entomophila L48 vs. *P. stutzeri* A1501	No	ns
P. entomophila L48 vs. *P. mendocina* ymp	No	ns
P. stutzeri A1501 vs. *P. mendocina* ymp	No	ns

Appendix Figure V.7: Development of *P. fluorescens* subclade IV-specific ExoU primers.



Appendix Figure V.7: **Development of *P. fluorescens* subclade IV-specific ExoU primers.**
Nucleotide sequences of the ExoU region from newly sequenced subclade IV *P. fluorescens* and reference *Pseudomonas* strains were aligned with the MAFFT algorithmwithin DNAstar's MegAlign Pro software, as described in **Chapter II.**

Appendix Figure V.8: Blastp with ExoU Subclade IV protein

A. Top 10 Results (sorted by Max score)						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
ExoU protein [Pseudomonas aeruginosa]	778	778	97%	0	62%	gi|553765609|WP_023097922.1
MULTISPECIES: exosortase [Pseudomonas]	775	775	97%	0	62%	gi|489225664|WP_003134060.1
exosortase [Pseudomonas aeruginosa]	775	775	97%	0	62%	gi|685977714|WP_031686411.1
exosortase [Pseudomonas aeruginosa]	775	775	97%	0	62%	gi|544849285|WP_021264405.1
ExoU protein [Pseudomonas aeruginosa]	775	775	97%	0	62%	gi|553775226|WP_023107059.1
Chain B, 1.92 Angstrom Resolution Crystal Structure Of The Full-Length Spcu In Complex With Full-Length Exou From The Type Iii Secretion System Of Pseudomonas aeruginosa [Pseudomonas aeruginosa]	775	775	97%	0	62%	gi|388325694|3TU3_B
exosortase [Pseudomonas aeruginosa]	774	774	97%	0	62%	gi|685987152|WP_031693452.1
ExoU protein [Pseudomonas aeruginosa]	771	771	97%	0	62%	gi|553766493|WP_023098783.1
exosortase [Pseudomonas aeruginosa]	770	770	97%	0	62%	gi|640512061|WP_024947856.1
Chain B, Structure Of The Heterodimeric Complex Exou-Spcu From The Type Iii Secretion System (T3ss) Of Pseudomonas aeruginosa [Pseudomonas aeruginosa]	747	747	96%	0	61%	gi|384482468|4AKX_B
type three secretion system effector protein [Photorhabdus asymbiotica]	693	693	97%	0	55%	gi|502387210|WP_012777218.1

B. Top 10 Results within the *P. fluorescens* Species Complex (sorted by Max score)						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
type III secretion system effector protein exou [Pseudomonas fluorescens]	554	554	90%	0	48%	gi|666087612|AIG03532.1
type III secretion system effector protein exou [Pseudomonas fluorescens LMG 5329]	543	543	88%	0	50%	gi|689664204|KGE69246.1
type III secretion system effector protein exou [Pseudomonas fluorescens]	536	536	89%	6.00E-179	48%	gi|696326909|WP_032902294.1
type III secretion system effector protein exou [Pseudomonas fluorescens]	535	535	90%	3.00E-178	48%	gi|665925045|WP_031291096.1
type III secretion system effector protein ExoU [Pseudomonas fluorescens]	534	534	89%	5.00E-178	47%	gi|504533191|WP_014720293.1
hypothetical protein [Pseudomonas fluorescens]	533	533	90%	1.00E-177	47%	gi|506366587|WP_015886306.1
type III secretion system effector protein ExoU [Pseudomonas fluorescens]	523	523	92%	5.00E-174	46%	gi|489287418|WP_003194986.1
type III secretion system effector protein ExoU [Pseudomonas fluorescens]	522	522	91%	2.00E-173	47%	gi|515544709|WP_016977738.1
type III secretion system effector protein ExoU [Pseudomonas synxantha]	519	519	90%	3.00E-172	47%	gi|492258387|WP_005792216.1
type III secretion system effector protein ExoU [Pseudomonas fluorescens EGD-AQ6]	493	493	81%	3.00E-163	48%	gi|542009337|ERH55219.1

C. AU11122 ExoU protein alignment against reference strains						
Reference Strain	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
P.aeruginosa PA14	805	805	97%	0	62%	Query_124677
P.aeruginosa PA99	805	805	97%	0	62%	Query_23921
P.aeruginosa 103	805	805	97%	0	62%	Query_43385
P.aeruginosa MH27	801	801	97%	0	62%	Query_107813
P. fluorescens A506	514	514	89%	6.00E-178	47%	Query_54715
P.fluorescens SBW25	543	577	93%	0	48%	Query_21485
P.fluorescens F113	344	344	74%	2.00E-113	41%	Query_176311
P.fluorescens SS101	516	516	92%	1.00E-178	46%	Query_216103
P.syringae pv. syringae	504	504	89%	2.00E-174	46%	Query_242817
P.syringae pv. maculicola str. ES4326	506	506	90%	6.00E-175	46%	Query_55093
Achromobacter xylosoxidans NH44784-1996	101	140	64%	1.00E-26	30%	Query_8983

Appendix Figure V.8: Blastp with ExoU Subclade IV protein.
The amino acid sequence of the ExoU protein from the *P. fluorescens* subclade IV strain AU11122 was used to screen the Microbial database **A.** The top ten results, as sorted by Max Score. **B.** The top ten results that fall within the *P. fluorescens* species complex. **C.** The amino acid sequence of the ExoU protein from AU11122 was aligned against ExoU proteins from the reference strains in **Figure V.15**.

Appendix Figure V.9: Blastp with SpcU Subclade IV protein.

A. NCBI BLASTp with Subclade IV SpcU protein- Top 10 Results (sorted by Max score)						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
hypothetical protein [Pseudomonas fluorescens]	273	273	100%	1.00E-91	100%	WP_047278298.1
hypothetical protein [Pseudomonas fluorescens]	271	271	100%	5.00E-91	99%	WP_047296065.1
hypothetical protein [Pseudomonas lundensis]	269	269	100%	4.00E-90	97%	WP_048375355.1
MULTISPECIES: hypothetical protein [Pseudomonas]	218	218	100%	5.00E-70	87%	WP_048364079.1
hypothetical protein [Pseudomonas sp. PH1b]	137	137	86%	4.00E-38	62%	WP_025131116.1
SpcU [Pseudomonas aeruginosa]	130	130	88%	2.00E-35	54%	WP_034024593.1
SpcU [Pseudomonas aeruginosa]	129	129	86%	3.00E-35	55%	WP_031755414.1
SpcU [Pseudomonas aeruginosa]	129	129	86%	5.00E-35	55%	WP_031641132.1
MULTISPECIES: SpcU [Pseudomonas]	128	128	86%	8.00E-35	55%	WP_003134054.1
Chain A, 1.92 Angstrom Resolution Crystal Structure Of The Full-Length Spcu In Complex With Full-Length Exou From The Type Iii Secretion System Of Pseudomonas aeruginosa [Pseudomonas aeruginosa]	129	129	86%	9.00E-35	55%	3TU3_A

B. NCBI BLASTp with Subclade IV SpcU protein- Top 10 Results within the *P. fluorescens* Species Complex						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
hypothetical protein [Pseudomonas fluorescens]	273	273	100%	1.00E-91	100%	WP_047278298.1
hypothetical protein [Pseudomonas fluorescens]	271	271	100%	5.00E-91	99%	WP_047296065.1
hypothetical protein [Pseudomonas lundensis]	269	269	1	4E-90	0.97	WP_048375355.1

Appendix Figure V.9: Blastp with ExoU Subclade IV protein.
The amino acid sequence of the SpcU protein from the *P. fluorescens* subclade IV strain AU11122 was used to screen the Microbial database. **A.** The top ten results, as sorted by Max Score. **B.** The top ten results that fall within the *P. fluorescens* species complex.

Chapter VI

Chronic Inflammation in the Lower Airways Favors the Growth of *P. fluorescens* Species Complex Bacteria

BACKGROUND

There is a link between inflammation and changes in the gastrointestinal microbiota. Shifts in microbial communities that have a negative effect on the host, called microbial dysbiosis, have been linked to multiple inflammatory diseases, such as IBD, colitis, and colorectal cancer (1-5). Inflammation creates an environment that favors the outgrowth, or 'blooms', of certain groups of bacteria (6, 7). A common association is found between inflammation and an increase in the relative proportion of Proteobacteria (1, 7-10). The connection between inflammation and blooms of Proteobacteria is seen in mouse models of gastrointestinal inflammation and in human patients within inflammatory gastrointestinal illnesses and colorectal cancer (3-5) (1, 8, 9, 11, 12). What is not known is whether inflammation has a similar affect on the microbial community of the lower respiratory tract.

With advances in sampling and culture-independent identification techniques, investigators can now connect microbial community profiles to diseases of the respiratory

environment (13-16). Gammaproteobacteria, which includes *Pseudomonas spp.* are highly abundant in the microbiome of patients with non-cystic fibrosis bronchiectasis, such as in individuals with Chronic Obstructive Pulmonary Disease (COPD) (17). Periods of increased airway inflammation in patients with COPD are marked with higher levels of gammaproteobacteria when compared to controls (18). Explants from individuals with advanced COPD also contain a high abundance of gammaproteobacteria and, in particular, members of the genus *Pseudomonas* (15). Similarly, the microbiota of some individuals who had received lung transplants are also dominated by the *Pseudomonas* species' *P. aeruginosa* and *P. fluorescens* (14). The lung microbiota in individuals with asthma also contain higher levels of gammaproteobacteria than in healthy controls (18-21). In addition, *Pseudomonas spp.*, particularly *P. aeruginosa,* is the most frequent cause of chronic infection in individuals with Cystic Fibrosis (22).

We hypothesized that pulmonary inflammation would lead to an increase in gammaproteobacteria in the lungs of mice. To test this, we analyzed the lung microbiome in mice via culture-independent methods in an *Aspergilllus fumigatus* model of pulmonary inflammation (23, 24). In this model, repeated weekly exposures to *A. fumigatus* conidia leads to a persistent inflammation without the development of aspergillosis in the mice. By four weeks of *A. fumigatus* exposure there is an immune response marked by the co-development of T(H1), T(H)2 and T(H)17 responses that continues through eight weeks. In analyzing the lung microbiome of mice after four and eight weeks of inflammation we found a significant increase in the relative abundance of *Pseudomonas.* Going back to previous datasets, we found increased levels of *Pseudomonas* in the lung microbiome of patients with idiopathic pulmonary fibrosis, moderate and severe COPD. Taxonomic classification on the mouse and human lung-associated *Pseudomonas* determined that it is *P. fluorescens*-related bacteria that bloomed

during inflammation in the mouse model and was one of the most abundant members in the diseased human datasets.

RESULTS

Figures VI.1 and **VI.2** show the results for the microbiome analysis from both groups of mice (**Figure VI.1** for the in-house sourced mice and **Figure VI.2** for the mice sourced from The Jackson Laboratory). In both groups of mice the untreated lungs had a detectable bacterial signal measured through 16S rRNA qPCR (results not shown) and 16S rRNA pyrosequencing. Both untreated mouse lungs in Figures A and B show a diverse bacterial community, however, the most abundant OTU differs between the two differently sourced mouse groups. The most abundant OTU in the untreated lungs from the in-house colony is a Proteobacteria (*Pseudomonas*, OTU#1, **Figure VI.1**), and the most abundant OTU in the Jackson sourced mice is a Bacteroidetes (*Barnesiella*, OTU#1, **Figure VI.2**). Tongue samples from untreated mice also show a detectable bacterial community via 16S rRNA qPCR (data not shown) and 16S rRNA pyrosequencing (Figures C and D). In contrast to the lung community, untreated mouse tongues were dominated by one OTU, a Streptococcus (Firmicutes, OTU#1 in **Figures VI.3** and **VI.4**). This was the same for both mice sourced from an in-house colony and from those sourced from The Jackson Laboratory.

The microbial communities from the two mouse groups (in-house colony and Jackson) responded differently to inflammation over time. After four weeks of weekly intranasal *A. fumigatus* challenges, the lung community of the in-house sourced mice showed an increase in relative abundance of the top OTU*,* classified as a *Pseudomonas* (labeled as *Pseudomonas* OTU#1 in **Figure VI.1**). This *Pseudomonas* OTU increased from approximately 12% to 40% of the total bacterial community (**Figure VI.1**). A similar increase in a *Pseudomonas* OTU was not

seen in the in-house sourced mice that were treated with four weeks of PBS inoculations. The *Pseudomonas* OTU#1 was nearly undetectable in PBS 4 week control lungs while a *Barnesiella* OTU increased from approximately 12% in the untreated lungs to approximately 35% in the PBS 4 week control lungs (**Figure VI.1**). This suggests that inflammation provided an environment for the increase in *Pseudomonas* in four week inflamed lungs. In the Jackson mice given four weeks of *A. fumigatus* challenges, no single OTU dominated the bacterial community. One *Ruminococcus* OTU (a Firmicute) did increase from nearly undetectable levels in untreated mice to approximately 20% relative abundance by four weeks of inflammation (**Figure VI.2**). The *Pseudomonas* OTU in the top ten relative abundant OTUs in the untreated mouse lungs (labeled in **Figure VI.2** as *Pseudomonas* OTU#2) did not show a significant increase in relative abundance in this group. In the Jackson mice, the microbial community of the PBS control mice lungs is similar to that of the inflamed 4 week lungs; no single OTU increased in relative abundance to dominate the microbial community after four weeks of *A. fumigatus* or PBS (**Figure VI.2**).

The two different mouse groups show an even more pronounced difference in response to inflammation after eight weeks of treatment. In the on-house mouse group, *Pseudomonas* OTU#1 increases in relative abundance from approximately 40% in four weeks inflamed lung to 45% in eight weeks inflamed lungs (**Figure VI.1**). In addition, there is an increase in a *Streptococcus* OTU (*Streptococcus* OTU#1 in **Figure VI.1**) from less than 2% relative abundance in four week inflamed lungs to around 30% relative abundance in the eight week inflamed lungs. These two OTUs, *Pseudomonas* OTU#1 and *Streptococcus* OTU#1, completely dominate the microbiome of the on-house mouse lungs after eight weeks of inflammation but are not dominant in the PBS-control lungs of this mouse group **(Figure VI.1**). In the Jackson mice, eight weeks of inflammation leads to a dramatic shift in the microbial community, with one *Streptococcus* OTU (OTU#2, **Figure VI.2**) increasing from nearly undetectable levels in the

untreated and four week inflamed lungs to over 90% relative abundance in the eight week inflamed lungs (**Figure VI.2**). This OTU does not dominate the PBS control 8 week lungs in the Jackson mice. Interestingly, after eight weeks of inflammation both groups of mice (in-house and Jackson) show a large increase in the relative abundance of a *Streptococcus* OTU (Figures **VI.1** and **VI.2**). However, only *Pseudomonas* OTU#1 from the on-house mice shows an increase in relative abundance after four and eight weeks of inflammation. In contrast to the lungs, the microbial community of the tongue does not change during inflammation. In both mouse groups, the tongue community is dominated by a *Streptococcus* OTU.

The relative abundances of the two *Pseudomonas* OTUs are further displayed in Figure VI.5. Mouse lung-associated *Pseudomonas* OTU#1 corresponds to *Pseudomonas* OTU#1 in the in-house sourced mouse lungs (**Figure VI.1**) and the mouse lung-associated *Pseudomonas* OTU#2 corresponds to the *Pseudomonas* OTU#2 in the Jackson mouse lungs (**Figure VI.2**). The two *Pseudomonas* OTUs show similar abundance in untreated mouse lungs. However, after four weeks of inflammation, *Pseudomonas* OTU#1 increases in relative abundance in four out of eight mice. In three mice, *Pseudomonas* OTU#1 dominates the community at over 60% relative abundance after four weeks of inflammation (**Figure VI.5**). *Pseudomonas* OTU#2 shows no change in relative abundance after four weeks of inflammation. After eight weeks of inflammation, *Pseudomonas* OTU#1 increases in relative abundance such that it was more than 50% of the bacterial community in one mouse, more than 60% in two mice and more than 80% in two additional mice. Interestingly, in four mice from the in-house colony, *Pseudomonas* OTU#1 remains under 10% relative abundance after eight weeks of inflammation (**Figure VI.5**). In the case of *Pseudomonas* OTU#1, inflammation leads to a large increase in relative abundance in some mice that begins at four weeks and continues to increase during eight weeks. In the case of *Pseudomonas* OTU#2, inflammation does not lead to a marked increase in relative abundance (**Figure VI.5**).

Previous studies in our laboratory investigated the lung microbiome of human patient groups diagnosed with different pulmonary diseases or conditions. The groups consisted of patients with idiopathic pulmonary fibrosis (IPF), those who had received a lung transplant (LT), those diagnosed with moderate chronic obstructive pulmonary disease (COPD) and those diagnosed with severe COPD. Also studied is a group of individuals not diagnosed with any lung condition (ie. healthy). The healthy, IPF and LT patients each had their lung community sampled using a bronchoalveolar lavage (BAL) sampling procedure. The moderate COPD patients had their lung communities sampled through a brush procedure. The severe COPD individuals had their lung communities sampled through a lung explant procedure. From all human patient group samples, total DNA was collected for pyrosequencing of the 16S rRNA gene. For the healthy, IPF, LT and moderate COPD brushes samples the V3-V5 variable region of the 16S rRNA gene was amplified; for the severe COPD explant samples the V1-V3 variable region of the 16S rRNA gene was amplified. All amplified 16S rRNA gene sequences were binned into OTUs at 0.03 cutoff (97% sequence similarity) and taxonomically classified for further analysis.

Eight human lung-associated *Pseudomonas* OTUs were identified in the healthy, IPF, LT and moderate COPD patient groups. The relative abundances of each of these *Pseudomonas* OTUs are shown in **Figure VI.6** and are numbered human lung-associated *Pseudomonas* OTUs #1 – #8. Only human lung-associated *Pseudomonas* OTUs #1 and #2 show detectable relative abundance in all four patient groups (healthy, IPF, LT and moderate COPD, Figure VI.6). Human lung-associated *Pseudomonas* OTU#1 is one of the most dominant *Pseudomonas* OTUs in the IPF, LT and moderate COPD patient BALs but is only detectable in one healthy patient BAL. There is a dichotomy with human lung-associated *Pseudomonas* OTU#2 in the LT patient BALs; it is either very dominant (>70% relative abundance) or hardly detectable (<10%). Human lung-associated *Pseudomonas* OTU#2 is also seen at some level in most of the

moderate COPD brush samples and in one healthy patient BAL. Human lung-associated *Pseudomonas* OTUs #3 and #4 are detected in most of the IPF and LT patient BALs and in a few of the healthy patient BALs but always at lower than 30% relative abundance. Human lung-associated *Pseudomonas* OTUs #5, #6 and #7 are barely detectable in any of the human patients. Human lung-associated *Pseudomonas* OTU#8 is only detected in the moderate COPD brushes and is either at >30% relative abundance (2 patient samples) or at <20% relative abundance. Overall, human lung-associated *Pseudomonas* OTU#1 is the human lung *Pseudomonas* OTU with the most consistent presence across the human patient samples.

The relevant abundance of *Pseudomonas* OTUs was also analyzed for the lung explant samples that were taken from the patients with severe COPD. The 16S rRNA gene from these samples was amplified using primers that aligned to the V1-V3 variable regions and, therefore, could not be processed with the samples from the other patient sets. However, the amplified 16S rRNA sequences were also binned into OTUs at a 97% similarity cutoff and three OTUs were identified as *Pseudomonas.* These *Pseudomonas* OTUs were then classified as either *P. fluorescens, P. aeruginosa* or *P. putida* by screening a representative consensus sequence from each *Pseudomonas* OTU against NCBI's microbial nucleotide database. A more detailed description of this taxonomic classification is provided below. For the severe COPD patients, multiple lung explant samples were taken from different portions of the lower airways and the microbial community of each explant samples was processed and analyzed (15). The relative abundance of each *Pseudomonas* OTU was determined for the individual explant samples from each patient. As seen in Figure VI.7, the *P. fluorescens* OTU shows the highest relative abundance of any of the *Pseudomonas* OTUs in each of the lung explant samples. In more than half of the samples, the *P. fluorescens* OTU is more than 70% of the microbial community (**Figure VI.7**). The *P. aeruginosa* and *P. putida* OTUs are less than 10% relative abundance in

each of the lung explant samples. Therefore, the *P. fluorescens* OTU dominated the microbial community in the majority of the lung explant samples from severe COPD patients.

The initial method of taxonomic classification applied to the binned OTU sequences does not delineate past the genus-level of classification. In order to see which *Pseudomonas* species is represented by the mouse and human lung-associated *Pseudomonas* OTUs, we applied several different taxonomic methods. For the taxonomic classification methods, we utilized the census sequence from each *Pseudomonas* OTU that represents 97% of the shared nucleotide identity of the sequence binned within that OTU (referred as the OTU-rep). For the first taxonomic classification method we created a phylogenetic tree based on the nucleotide similarity of the OTU-rep sequences from the *Pseudomonas* OTUs and the nucleotide sequence between the V3 and V5 variable region of the 16S rRNA from select *Pseudomonas* strains (**Figure VI.8**). Even considering the short nucleotide length of the V3-V5 16S rRNA gene sequence (~250 bp), a tree can be inferred that places the reference *Pseudomonas spp.* into previously shown taxonomic clades (25) and indicates likely species-level classification of the mouse and human lung-associated *Pseudomonas* OTUs (**Figure VI.8**).

The two mouse lung-associated *Pseudomonas* OTUs cluster into two different areas of the tree. Mouse lung-associated *Pseudomonas* OTU#2 is grouped in the clade consisting of reference *P. putida, P. entomophila* and *P. mendocina* sequences. In this clade is also human lung-associated *Pseudomonas* OTU#3. This suggests that these two *Pseudomonas* OTUs are most likely related to a *P. putida/entomophila/mendocina* pseudomonad. Mouse lung-associated *Pseudomonas* OTU#1 groups closely together with human lung-associated *Pseudomonas* OTU#4 and these two *Pseudomonas* OTUs branch most closely to a *P. fluorescens* species-complex clade (**Figure VI.8**) (25). The *P. fluorescens*-species complex consists of up to 52 differently named *Pseudomonas spp.,* including, *P. protegens, P.*

chlorarphis and *P. brassicacearum* (25, 26). Within this *P. fluorescens*-species complex clade is human lung-associated *Pseudomonas* OTU#8 that was only present in the moderate COPD patients (**Figures VI.6** and **VI.8**). This suggests that mouse lung-associated *Pseudomonas* OTU#1 and human lung-associated *Pseudomonas* OTUs #4 and #8 are most likely members of the *P. fluorescens* species complex. Another clade consisting of *P. fluorescens* species complex bacteria is located in another position of the tree and contains human lung-associated *Pseudomonas* OTU#1. The *P. fluorescens* species complex is known to consist of multiple subclades (25-27). The relatively short sequence of the 16S rRNA V3-V5 region is not able to divide the *P. fluorescens* species complex into its known three subclades but can indicate which *Pseudomonas* OTUs are or are not part of the *P. fluorescens* species complex (25, 26). The *P. aeruginosa* clade within the tree consists of human lung-associated *Pseudomonas* OTU#2. Finally, three human lung-associated *Pseudomonas* OTUs (#5, 6 and 7) grouped separately on the tree such that no species-level classification could be inferred. The V3-V5 16S rRNA phylogenetic tree was able to identify a primarily species-level classification for both mouse lung-associated *Pseudomonas* OTUs and five of the eight human lung-associated *Pseudomonas* OTUs.

The second method of taxonomic classification used to analyze the *Pseudomonas* OTUs involved screening the OTU-rep sequences against NCBI's online Microbial nucleotide database. The results of the blast screen are shown in **Appendix File VI.1**. The top ten hits from a blastn screen of the OTU-rep sequence from the mouse lung-associated *Pseudomonas* OTU#1 consist of members of the *P. fluorescens*-species complex and two *P. syringae* strains. The top ten hits retrieved from mouse lung-associated *Pseudomonas* OTU#2 consists of *P. putida* and strains that are part of the *P. putida*-species group (27). Both of these results agree with the placement of these two OTUs in the V3-V5 16S rRNA tree (Figure VI.8). Members of the *P. fluorescens* species complex are represented in the results from human lung-associated

Pseudomonas OTU #1, #4, #7 and #8. In all four of these OTUs, strains of *P. syringae* are also in the blast results. *P. syringae* is also grouped in the same clade as members of the *P. fluorescens* species complex in the V3-V5 tree. At the V3-V5 region of the 16S rRNA gene, it is likely not possible using these two methods to distinguish a *Pseudomonas* OTU as being part of the *P. fluorescens* species complex or *P. syringae*. However, it is possible to distinguish a *P. fluorescens/P. syringae* from a *P. aeruginosa* or *P. putida.* The results from the blast table (**Appendix File VI.1**) therefore agree with the tree in Figure VI.8 for human lung-associated *Pseudomonas* OTUs #1, #4 and #8, which place these OTUs closest to members of the *P. fluorescens* species complex/*P. syringae*. However, human lung-associated *Pseudomonas* OTU #7 is placed separately next to *Pseudomonas* OTU#6 and OTU#5 and not within the *P. fluorescens* species complex clade, which disagrees slightly with the blastn results. The blastn search results from human lung- associated *Pseudomonas* OTU#5 show mixed results, with some *P. putida* strains and some additional *Pseudomonas* spp. not included in the tree (ex: *P. monteilli* and *P. plecoglossicida*). The OTU-rep sequence from human lung-associated *Pseudomonas* OTU#6 retrieves mostly *P. stutzeri* results, another *Pseudomonas spp.* also not included on the tree. Therefore, human lung-associated *Pseudomonas* OTUs #5, #6 and #7 likely represent *Pseudomonas* strains that are genetically distinct from those on the tree. Considering that the *Pseudomonas* genus consists of at least 211 recognized species, it was not possible to put every *Pseudomonas* species on the V3-V5 tree (26). However, the combination of the blast table and the V3-V5 16S rRNA tree allows for *Pseudomonas* OTUs to be classified as either *P. aeruginosa, P. putida* or *P. fluorescens*-species complex/ *P. syringae.*

The final method of classification applied to the *Pseudomonas* OTUs was to align each OTU-rep sequence against a panel of representative *Pseudomonas* genomes (**Appendix File VI.2**). The mouse lung- associated *Pseudomonas* OTU#1 showed 100% sequence identity and query coverage only against the two *P. fluorescens* species complex strains SBW25 and Pf-5.

This result agrees with the results found in the V3-V5 phylogenetic tree (**Figure VI.8**) and the blast results in **Appendix File VI.1**. Mouse lung-associated *Pseudomonas* OTU#2 aligned with 100% sequence identity and query coverage to *P. putida* GB-1, which also agrees with the V3-V5 tree and blast table (**Figure VI.8** and **Appendix File VI.1**). Human lung-associated *Pseudomonas* OTU#1 aligned with 100% sequence identity and query coverage to *P. fluorescens* A506 and human lung-associated *Pseudomonas* OTU#2 aligned with 100% sequence identity and query coverage to the *P. aeruginosa* genomes PAO1 and PA14. The remaining human lung-associated *Pseudomonas* OTUs did not align with perfect sequence identity or query coverage to any of the screened *Pseudomonas* genomes. However, the genome with highest match to the human lung-associated *Pseudomonas* OTU rep sequence agreed with the destination provided by the V3-V5 tree and the blast table. Additional genomes would need to be added in order to further delineate a species level classification of these *Pseudomonas* OTUs, however, some basic distinctions between *P. aeruginosa, P. putida* and *P. fluorescens* species complex/*P. syringae* can be made with the combination of these three taxonomic tools.

The final taxonomic analysis we performed was on the *Streptococcus* OTUs from the inflamed mouse lung microbiome. Both the mice from the in-house colony (**Figure VI.1**) and the mice from the Jackson Labs (**Figure VI.2**) showed an increase in relative abundance of a single *Streptococcus* OTU after eight weeks of inflammation (*Streptococcus* OTU #1 and #2, respectively). We used the OTU-rep sequence and screened NCBI's Microbial nucleotide database. The results are displayed in **Appendix File VI.3**. Both *Streptococcus* OTUs had the highest number of hits to *Streptococcus agalactiae*. We then performed a direct alignment between these two *Streptococcus* OTU representative sequences and found that they shared 99% sequence identity over 100% query coverage. At this level of similarity, the two *Streptococcus* OTUs likely represent the same *Streptococcus* species.

DISCUSSION

We set out to investigate how inflammation and inflammatory diseases affect the microbial community of the lower airways. Based on literature describing the gastrointestinal tract, we hypothesized that inflammation would lead to an outgrowth of *Proteobacteria* (1, 6-10). We first tested this hypothesis in a mouse model of respiratory inflammation in two different mouse groups. The microbiota of the mice from the two groups showed marked differences in response to inflammation after four and eight weeks of pulmonary challenge with *A. fumigatus* condia. The microbial community of mice from the in-house colony was dominated by two OTUs at eight weeks of inflammation, a *Pseudomonas* OTU and a *Streptococcus* OTU (**Figures VI.1** and **VI.5**). Further taxonomic classification revealed that the *Pseudomonas* OTU in this mouse group was likely a *P. fluorescens* and the *Streptococcus* OTU was a *S. agalactiae* (**Figure VI.8** and **Appendix Files VI.1** and **VI.2**). After four and eight weeks of inflammation, the lung microbiota of mice from Jackson was dominated by a *Streptococcus* OTU, later identified as *S. agalactiae* (**Figure VI.2** and **Appendix File VI.3**). Interestingly, both mouse groups had a *Pseudomonas* OTU within the top ten most abundant OTUs in the untreated mice lungs, prior to any inflammation (**Figures VI.1, VI.2** and **VI.8**). However, the mouse lung-associated *Pseudomonas* OTU identified as *P. fluorescens* bloomed in the inflamed lung environment while the mouse lung-associated *Pseudomonas* OTU identified as *P. putida* did not bloom in response to inflammation (**Figure VI.8** and **Appendix Files VI.1** and **VI.2**). These findings suggest that inflammation selects for only certain gammaproteobacteria species, such as *P. fluorescens,* that are able flourish in this particular environment.

Both mouse groups (in-house and Jackson) also saw a bloom in a *Streptococcus* OTU that was later identified as *S. agalactiae* (**Figures VI.1** and **VI.2**). The same *S. agalactiae* OTU

that increased in relative abundance after eight weeks of inflammation in both mouse groups also dominated the microbial community of each group's corresponding tongue samples (**Figures VI.3** and **VI.4**). This suggests that the original source of the *Streptococcus* OTU in the lower airways of these two mice groups is the oral cavity. Though often considered 'obligate nasal breathers', air can and does traverse the mouth during respiration in rodents (28). *Streptococcus* non-*pneuomonia spp.* bacteria are also found in the nasal microbiota of mice (29). The anatomical structure of the rodent upper and lower respiratory tract suggests that there is interaction between the microbial communities of the oral cavity, the nares and the lungs, suggesting a model by which *Streptococcus spp.* bacteria could traverse from the upper to lower airways during respiration and get positively selected during inflammation. In contrast, the *P. fluorescens* OTU that dominated the in-house mice lung microbiota after inflammation was not found in the corresponding tongue samples, suggesting instead that it was an indigenous member of the lower airway microbiota that was positively selected during inflammation. While the *S. agalactiae* OTU that bloomed during lung microbiota was also found in the oral cavity of mice, the *P. fluorescens* OTU that bloomed during inflammation was not.

Analysis of the microbiota lower airways from patients with idiopathic pulmonary fibrosis (IPF), moderate and severe chronic obstructive pulmonary disease (COPD), individuals that had received a lung transplant (LT) and healthy controls also revealed associations with the outgrowth of *P. fluorescens*-related bacteria in the context of inflammatory-based respiratory diseases. Eight human lung-associated *Pseudomonas* OTUs were found in the IPF, moderate COPD, LT and healthy datasets and three different human lung-associated *Pseudomonas* OTUs within the severe COPD dataset. Using three different classification methods we were able to apply species-level identification to the *Pseudomonas* OTUs. The human lung-associated *Pseudomonas* OTU (#1) most consistently found across the IPF, moderate COPD, LT and healthy datasets was identified as belonging to the *P. fluorescens*-species

complex/*P.syringae* (**Figures VI.5**, **VI.8** and **Appendix 1**,**2**). Human lung-associated *Pseudomonas* OTU#8, only found in the moderate COPD samples, was also classified as likely belonging to the *P. fluorescens* species complex/*P.syringae* (**Figures VI.5**, **VI.8** and **Appendix 1**,**2**). A *P. fluorescens*-species complex OTU also dominated the lung explant samples from the severe COPD dataset. Microbiota analysis of the diseased, and healthy, human lung datasets found that a predominance of bacteria related to *P. fluorescens*/ *P. syringae* species groups.

Both the human and mouse lung samples found an association with bacteria of the *P. fluorescens/P. syringae* species group and inflammation. The *P. fluorescens* and *P. syringae*-species groups are taxonomically related within the *Pseudomonas* genus, and therefore, it might not be possible to differentiate an OTU as *P. fluorescens* or *P. syringae* due to the short nucleotide sequence that exists between the V3 and V5 variable regions of the 16S rRNA gene (27). However, it is possible to distinguish between *P. fluorescens/P. syringae* and other members of the *Pseudomonas* species. Both *P. fluorescens* and *P. syringae* are typically considered environmental bacteria associated with plant, soil and the rhizosphere, while the *Pseudomonas* species most commonly associated with human disease is *P. aeruginosa* (26, 30-32). However, multiple reports have found associations between *P. fluorescens*-species bacteria and different human disease states (26). Here we present evidence that inflammation promotes the growth of *P. fluorescens/P. syringae*-related bacteria and that these bacteria can also be found in high abundance in the lower airways of individuals with IPF, COPD or who have received lung transplants. Further research is needed to understand by what mechanism inflammation favors the growth of *P. fluorescens/P. syringae* bacteria.

Figure VI.1: Inflamed mouse lung microbiome from in-house colony.

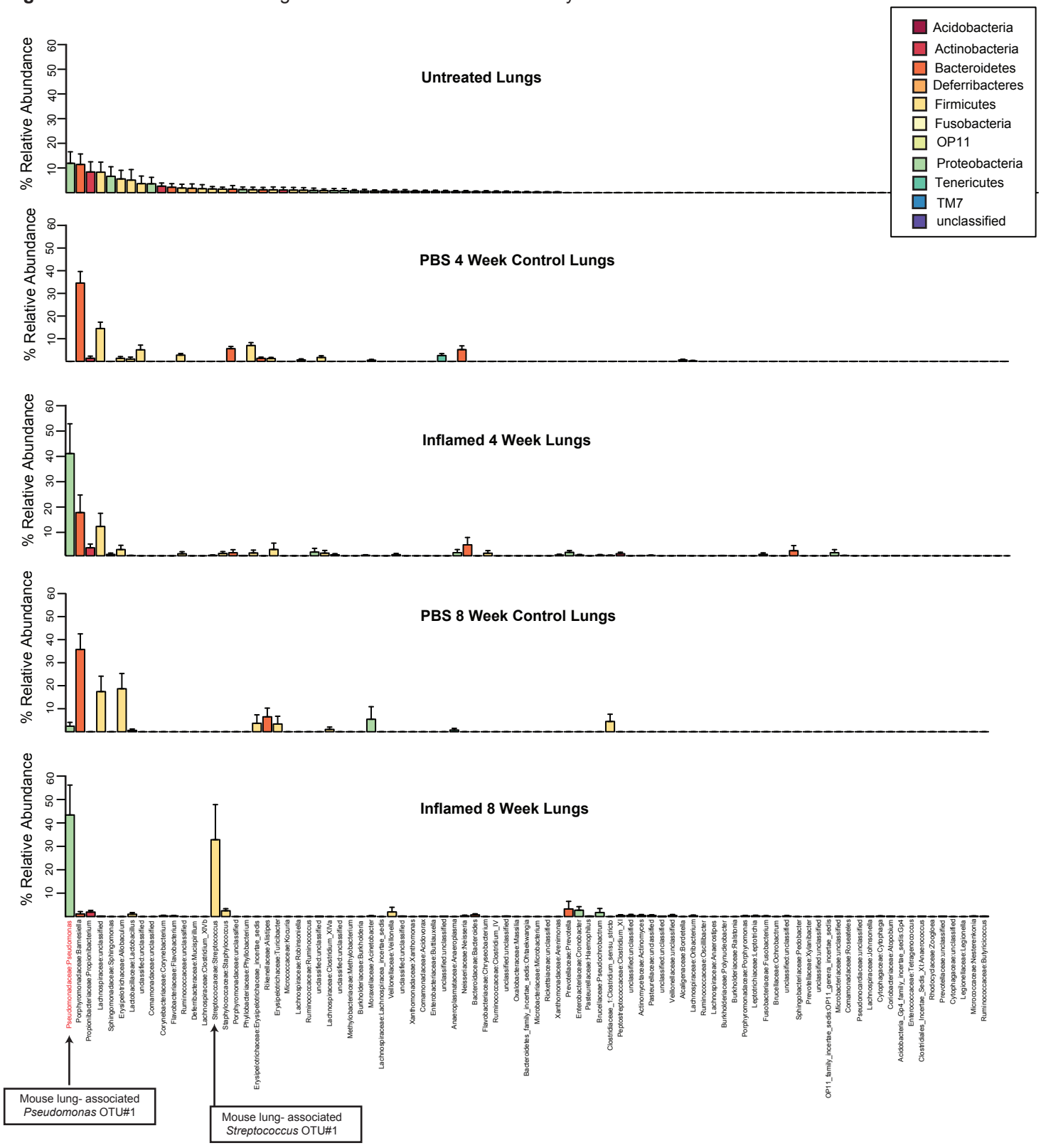


Figure VI.1: Inflamed mouse lung microbiome from in-house colony.
Mice were intranasally inoculated with either 20 ul of 1 x10^8 *A. fumigatus* spores/ml or PBS for four or eight weeks. The following number of mice was in each treatment group: Untreated Control Lungs (n=13); PBS Control 4 Week Lungs (n=8); Inflamed 4 Week Lungs (n=11); PBS Control 8 Week Lungs (n=6); Inflamed 8 Week Lungs (n=9). For each mouse group, lung snips were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. The OTUs classified as *Pseudomonas* and *Streptococcus* are numbered for further analysis.

Figure VI.2: Inflamed mouse lung microbiome from Jackson labs.

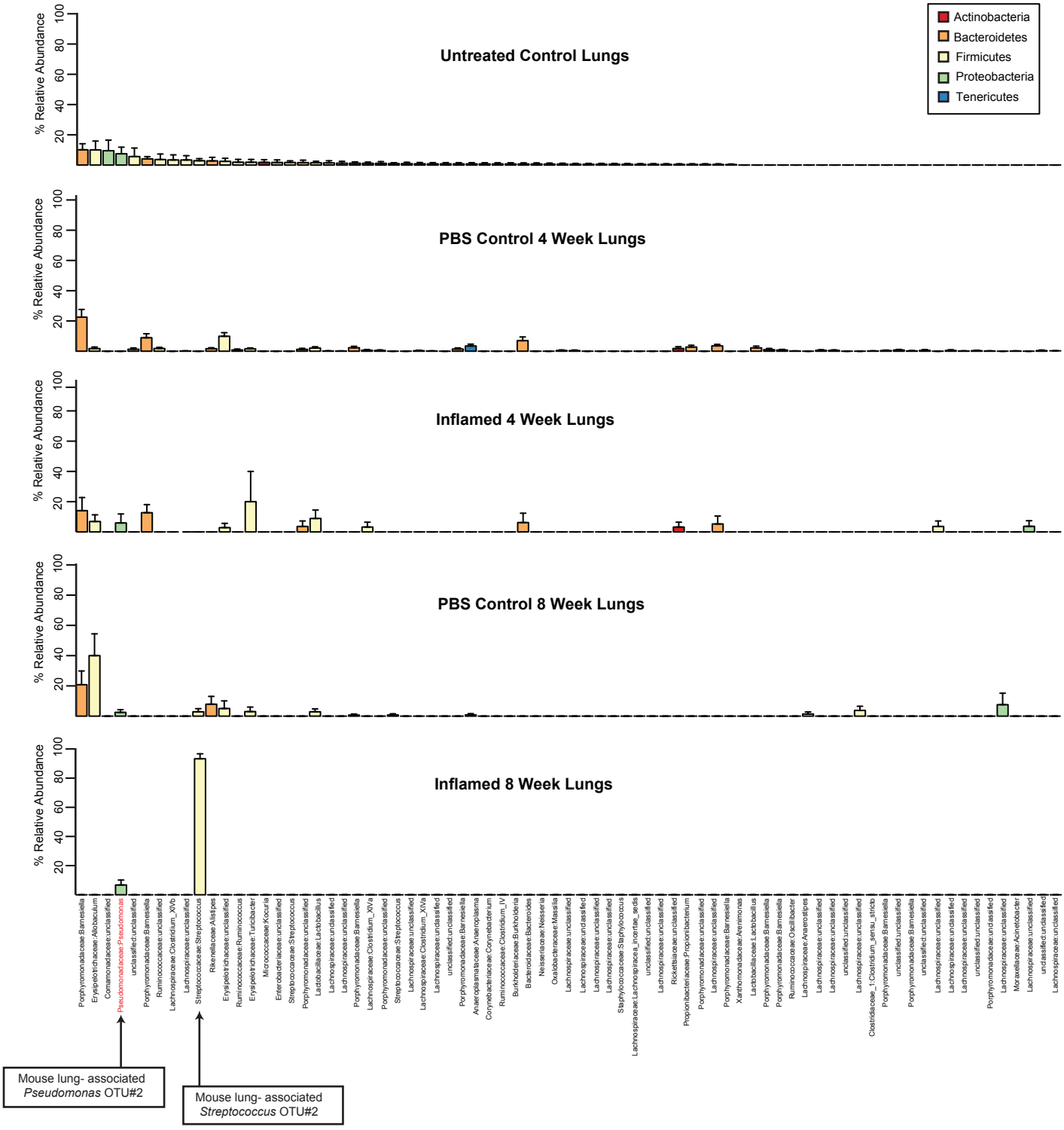


Figure VI.2: Inflamed mouse lung microbiome from Jackson labs.
Mice were intranasally inoculated with either 20 ul of 1 x10^8 *A. fumigatus* spores/ml or PBS for four or eight weeks. The following number of mice was in each treatment group:Untreated Control Lungs (n=7); PBS Control 4 Week Lungs (n=6); Inflamed 4 Week Lungs (n=5); PBS Control 8 Week Lungs (n=6); Inflamed 8 Week Lungs (n=3). For each mouse group, lung snips were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity.The OTUs classified as a *Pseudomonas* and *Streptococcus* are numbered for further analysis.

Figure VI.3: Inflamed mouse tongue microbiome from in-house colony.

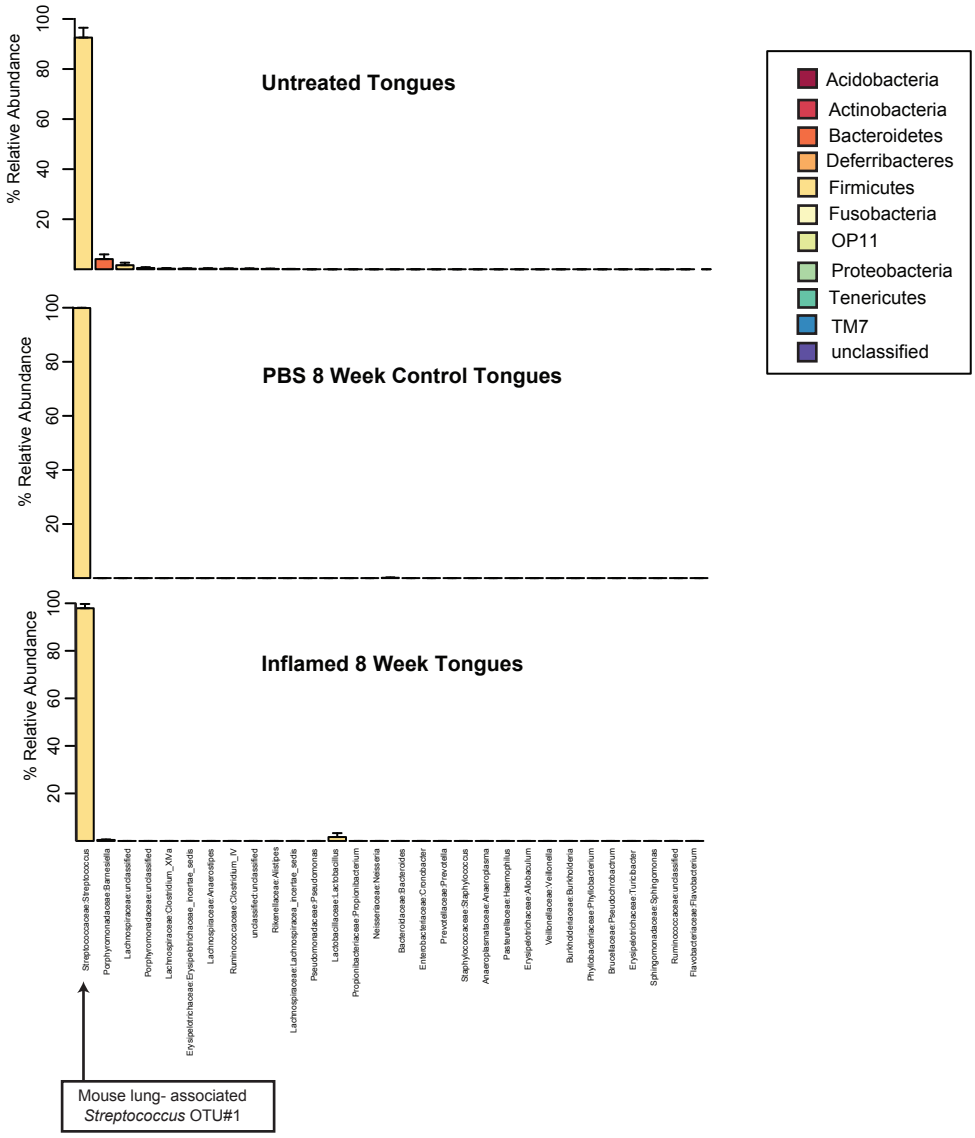


Figure VI.3: Inflamed mouse tongue microbiome from in-house colony.
Mice were intranasally inoculated with either 20 ul of 1 x10^8 *A. fumigatus* spores/ml or PBS for eight weeks. The following number of mice was in each treatment group: Untreated Control Tongues (n=7); PBS Control 8 Week Tongues (n=6); Inflamed 8 Week Tongues (n=3). For each mouse group, tongues were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. The OTUs classified *Streptococcus* are numbered for further analysis.

Figure VI.4: Inflamed mouse tongue microbiome from Jackson labs.

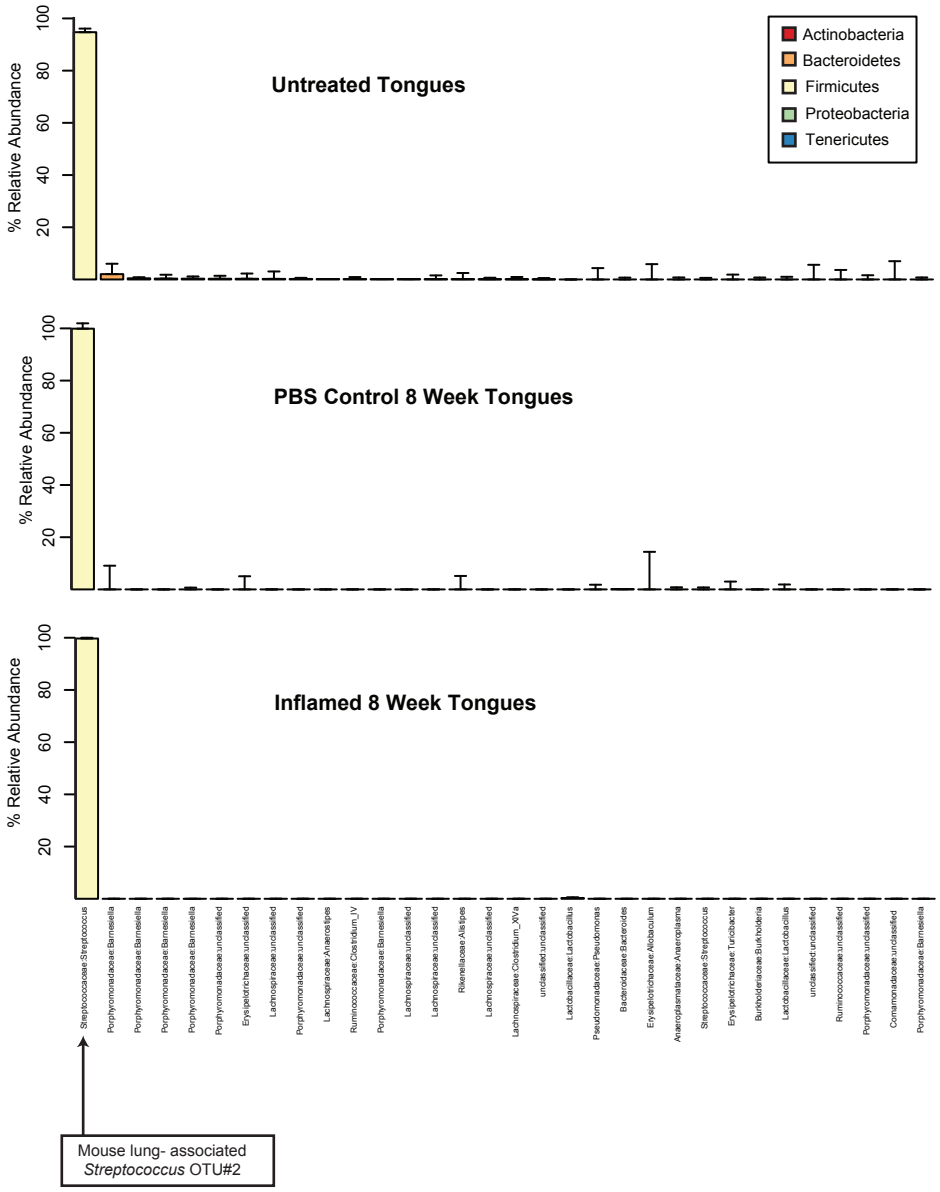


Figure VI.4: Inflamed mouse tongue microbiome from Jackson labs.
Mice were intranasally inoculated with either 20 ul of 1 x10^8 *A. fumigatus* spores/ml or PBS for eight weeks. The following number of mice was in each treatment group: Untreated Control Tongues (n=7); PBS Control 8 Week Tongues (n=6); Inflamed 8 Week Tongues (n=3). For each mouse group, tongues were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. The OTUs classified *Streptococcus* are numbered for further analysis.

Figure VI.5: Relative abundance of mouse lung- associated *Pseudomonas* OTUs in five treatment subsets**.**

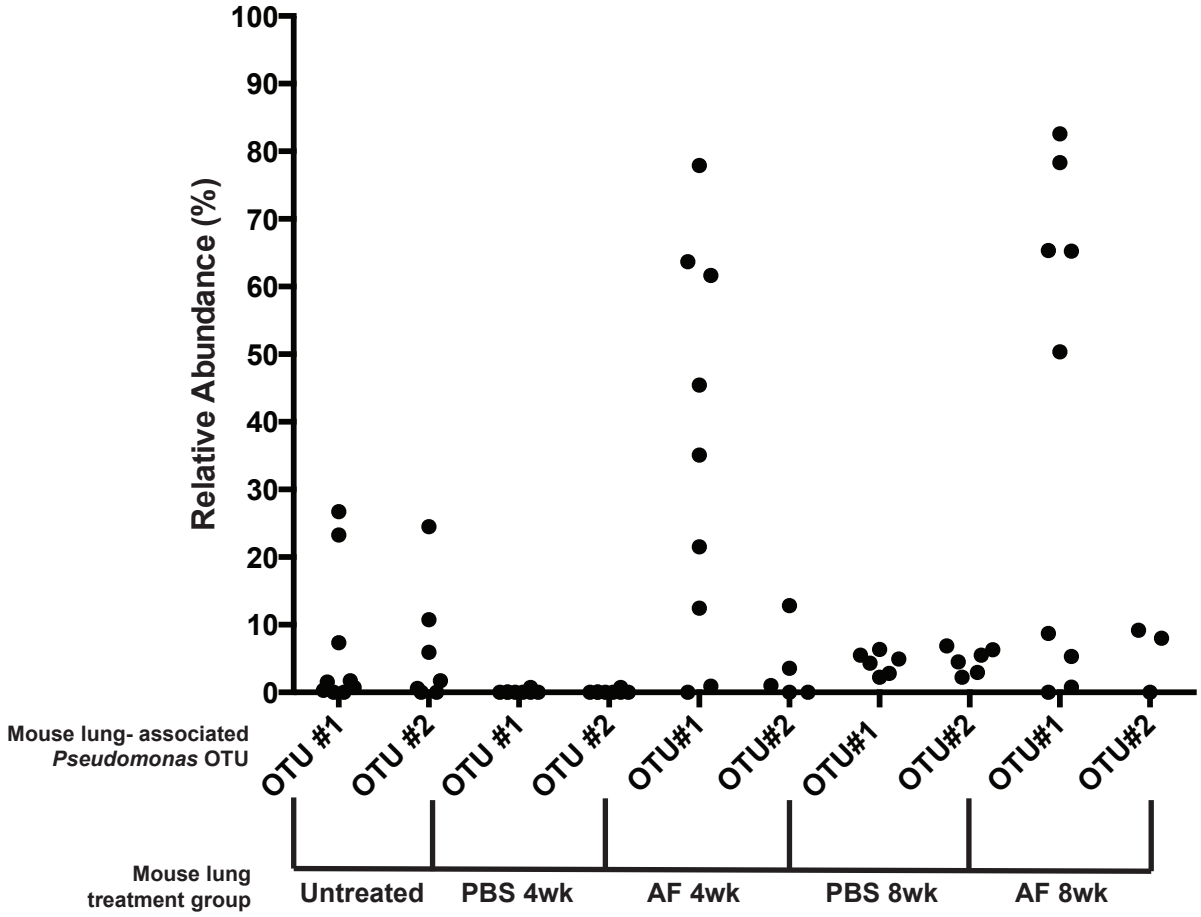


Figure VI.5: Relative abundance of mouse lung- associated *Pseudomonas* OTUs in five treatment subsets.
The relative abundance of *Pseudomonas* OTUs associated with mouse lungs in five difference treatment groups. The following number of mice was in each treatment group: Untreated OTU#1 (n=9); Untreated OTU#2 (n=7); PBS 4wk OTU #1 and #2 (n=8); AF 4 wk OTU#1 (n=8); AF 4wk OTU#2 (n=5); PBS 8wk #1 and #2 (n=6); AF 8wk OTU #1 (n=9); AF 8wk OTU #2 (n=3). The two OTUs were discovered in two separate mice subsets; OTU#1 from mice taken from an in-house colony and OTU #2 from mice purchased from The Jackson Lab. For each mouse group, lung snips were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. The OTUs correspond to the Mouse lung- associated *Pseudomonas* OTUs in **Figure VI.7** and **Appendix Figure VI.1**.

Figure VI.6: Relative abundance of human lung- associated *Pseudomonas* OTUs in three patient subsets**.**

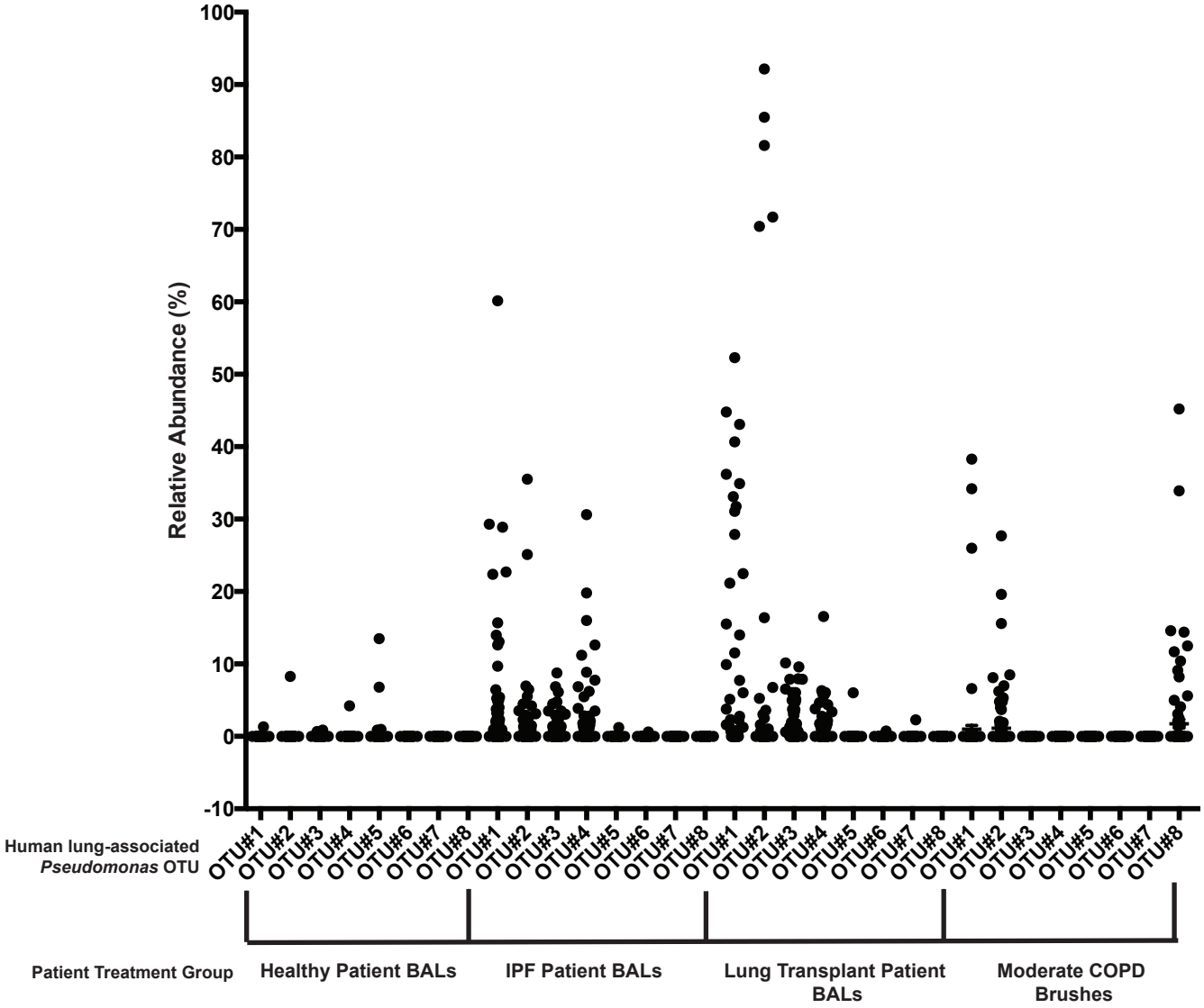


Figure VI.6: Relative abundance of human lung- associated *Pseudomonas* OTUs in three patient subsets.
The relative abundance of *Pseudomonas* OTUs associted with human lungs in three difference patient groups. The following number of patients were in each treatment group: Healthy (n=26); IPF (n=78) and Lung Transplant (n=44). For each treatment group, bronchoalveolar lavages were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. The OTUs correspond to the Human lung- associated *Pseudomonas* OTUs in **Figure VI.8** and **Appendix Figure VI.1**.

Figure VI.7: *Pseudomonas* abundance in COPD lung explants.

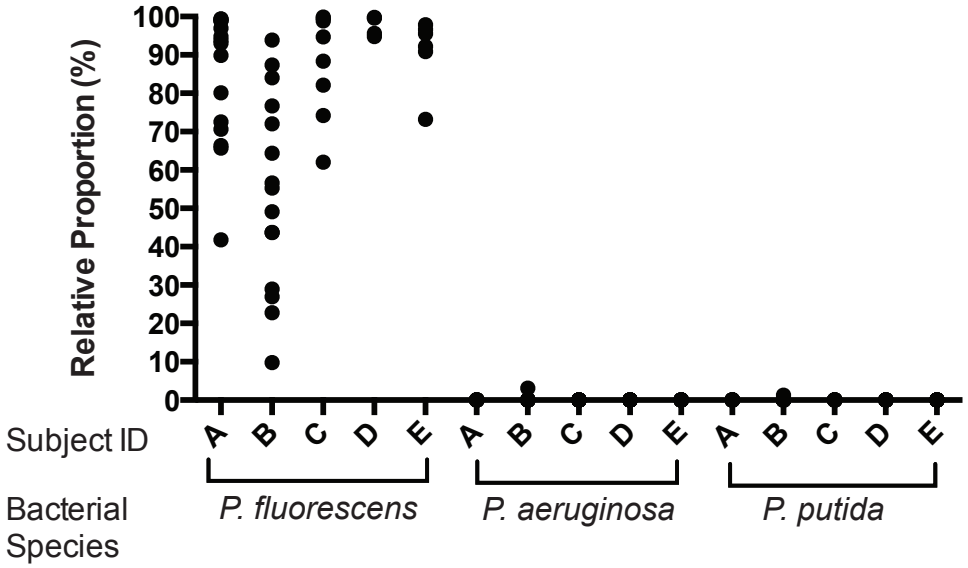


Figure VI.7: *Pseudomonas* abundance in COPD lung explant.
The relative abundance of OTUs classified as *P. fluorescens*, *P. aeruginosa* and *P. putida* within each COPD lung explant sample. Each column corresponds to a single subject and each dot represents a separate sample taken from a different part of their lung. Lung explants were collected for whole DNA isolation and 16S rRNA amplification. OTUs were binned at 97% similarity. OTUs classified as *Pseudomonas* were identified as *P. fluorescens, P. aeruginosa* or *P. putida* through blast search against NCBI's microbial nucleotide database, using the representative 16S rRNA sequence for that OTUs as a query.

Figure VI.8: V3-V5 16S rRNA phylogenetic tree.

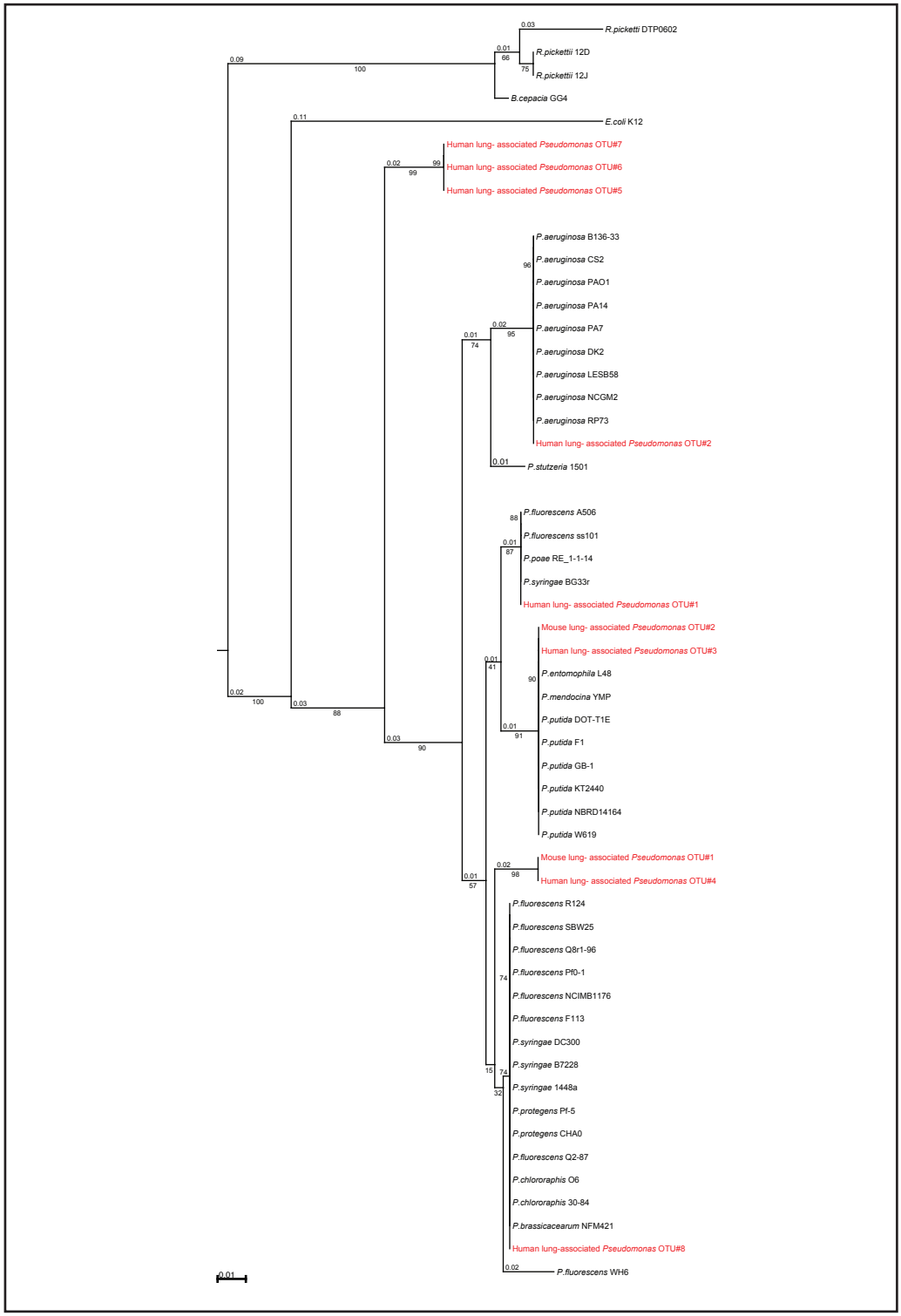


Figure VI.8: V3-V5 16S rRNA phylogenetic tree.
The phylogenetic relationship between newly and previously sequenced *Pseudomonas* isolates, based on the genetic sequence spanning the V3 and V5 variable regions of the 16S rRNA gene. The V3-V5 rRNA sequences were aligned with MAFFT (Katoh, 2013)(Katoh, 2002).

Chapter VI

References

1. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79.
2. Reeves AE, Theriot CM, Bergin IL, Huffnagle GB, Schloss PD, Young VB. 2011. The interplay between microbiome dynamics and pathogen dynamics in a murine model of Clostridium difficile Infection. Gut Microbes 2:145-158.
3. Couturier-Maillard A, Secher T, Rehman A, Normand S, De Arcangelis A, Haesler R, Huot L, Grandjean T, Bressenot A, Delanoye-Crespin A, Gaillot O, Schreiber S, Lemoine Y, Ryffel B, Hot D, Nunez G, Chen G, Rosenstiel P, Chamaillard M. 2013. NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. J Clin Invest 123:700-711.
4. Arthur JC, Jobin C. 2013. The complex interplay between inflammation, the microbiota and colorectal cancer. Gut Microbes 4:253-258.
5. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD. 2013. The gut microbiome modulates colon tumorigenesis. MBio 4:e00692-00613.
6. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338:120-123.
7. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, Sterzenbach T, Tsolis RM, Roth JR, Baumler AJ. 2011. Intestinal inflammation allows Salmonella to

use ethanolamine to compete with the microbiota. Proc Natl Acad Sci U S A 108:17480-17485.

8. Darfeuille-Michaud A, Neut C, Barnich N, Lederman E, Di Martino P, Desreumaux P, Gambiez L, Joly B, Cortot A, Colombel JF. 1998. Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. Gastroenterology 115:1405-1413.

9. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO. 2010. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes 1:138-147.

10. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra AM, Laughlin RC, Gomez G, Wu J, Lawhon SD, Popova IE, Parikh SJ, Adams LG, Tsolis RM, Stewart VJ, Baumler AJ. 2013. Host-derived nitrate boosts growth of E. coli in the inflamed gut. Science 339:708-711.

11. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, Timmis KN, Schreiber S. 2004. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut 53:685-693.

12. Martin HM, Campbell BJ, Hart CA, Mpofu C, Nayar M, Singh R, Englyst H, Williams HF, Rhodes JM. 2004. Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. Gastroenterology 127:80-93.

13. Dickson RP, Erb-Downward JR, Huffnagle GB. 2013. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med 7:245-257.

14. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. 2014. Changes in the lung microbiome following lung transplantation include the emergence of two distinct pseudomonas species with distinct clinical associations. PLoS One 9:e97214.

15. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ, Huffnagle GB. 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One 6:e16384.

16. Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ. 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A 109:5809-5814.

17. Tunney MM, Einarsson GG, Wei L, Drain M, Klem ER, Cardwell C, Ennis M, Boucher RC, Wolfgang MC, Elborn JS. 2013. Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. Am J Respir Crit Care Med 187:1118-1126.

18. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. 2014. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. J Clin Microbiol 52:2813-2823.

19. Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, Good JT, Jr., Gelfand EW, Martin RJ, Leung DY. 2013. The effects of airway microbiome on corticosteroid responsiveness in asthma. Am J Respir Crit Care Med 188:1193-1201.

20. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WO. 2010. Disordered microbial communities in asthmatic airways. PLoS One 5:e8578.

21. Marri PR, Stern DA, Wright AL, Billheimer D, Martinez FD. 2013. Asthma-associated differences in microbial composition of induced sputum. J Allergy Clin Immunol 131:346-352 e341-343.

22. Fitzsimmons S. 1996. The Cystic Fibrosis Foundation Patient Registry Report. Pediatric Pulmonology 1996 Suppl 21:267-275.

23. Murdock BJ, Shreiner AB, McDonald RA, Osterholzer JJ, White ES, Toews GB, Huffnagle GB. 2011. Coevolution of TH1, TH2, and TH17 responses during repeated pulmonary exposure to Aspergillus fumigatus conidia. Infect Immun 79:125-135.

24. Shreiner AB, Murdock BJ, Sadighi Akha AA, Falkowski NR, Christensen PJ, White ES, Hogaboam CM, Huffnagle GB. 2012. Repeated exposure to Aspergillus fumigatus conidia results in CD4+ T cell-dependent and -independent pulmonary arterial remodeling in a mixed Th1/Th2/Th17 microenvironment that requires interleukin-4 (IL-4) and IL-10. Infect Immun 80:388-397.

25. Loper JE, Hassan KA, Mavrodi DV, Davis EW, 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT. 2012. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. PLoS Genet 8:e1002784.

26. Scales BS, Dickson RP, LiPuma JJ, Huffnagle GB. 2014. Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. Clin Microbiol Rev 27:927-948.

27. Mulet M, Lalucat J, Garcia-Valdes E. 2010. DNA sequence-based analysis of the Pseudomonas species. Environ Microbiol 12:1513-1530.

28. Gautam SH, Verhagen JV. 2012. Direct behavioral evidence for retronasal olfaction in rats. PLoS One 7:e44781.

29. Krone CL, Biesbroek G, Trzcinski K, Sanders EA, Bogaert D. 2014. Respiratory microbiota dynamics following Streptococcus pneumoniae acquisition in young and elderly mice. Infect Immun 82:1725-1731.

30. Murray TS, Egan M, Kazmierczak BI. 2007. Pseudomonas aeruginosa chronic colonization in cystic fibrosis patients. Curr Opin Pediatr 19:83-88.

31. Lyczak JB, Cannon CL, Pier GB. 2000. Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist. Microbes Infect 2:1051-1060.

32. Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW. 2011. Pseudomonas genomes: diverse and adaptable. FEMS Microbiol Rev 35:652-680.

Appendix Figure VI.1: Blast results of *Pseudomonas* OTUs.

Mouse lung- associated *Pseudomonas* OTU#1						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas sp. TKP, complete genome	307	1538	98%	5.00E-81	99%	NC_023064.1
Pseudomonas protegens CHA0, complete genome	307	1538	98%	5.00E-81	99%	NC_021237.1
Pseudomonas sp. UW4 chromosome, complete genome	307	2083	98%	5.00E-81	99%	NC_019670.1
Pseudomonas fluorescens F113 chromosome, complete genome	307	1538	98%	5.00E-81	99%	NC_016830.1
Pseudomonas brassicacearum subsp. brassicacearum NFM421 chromosome, complete genome	307	1532	98%	5.00E-81	99%	NC_015379.1
Pseudomonas fluorescens Pf0-1 chromosome, complete genome	307	1845	98%	5.00E-81	99%	NC_007492.2
Pseudomonas fluorescens SBW25 chromosome, complete genome	307	1538	98%	5.00E-81	99%	NC_012660.1
Pseudomonas syringae pv. phaseolicola 1448A chromosome, complete genome	307	1538	98%	5.00E-81	99%	NC_005773.3
Pseudomonas protegens Pf-5 chromosome, complete genome	307	1538	98%	5.00E-81	99%	NC_004129.6
Pseudomonas syringae pv. syringae B728a chromosome, complete genome	307	1538	98%	5.00E-81	99%	NC_007005.1
Mouse lung-associated *Pseudomonas* OTU#2						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas sp. CCOS191 genome assembly Pseudomonas sp. strain CCOS191, chromosome : I	324	2269	100%	5.00E-86	100%	NZ_LN847264.1
Pseudomonas putida S13.1.2, complete genome	324	2269	100%	5.00E-86	100%	NZ_CP010979.1
Pseudomonas plecoglossicida strain NyZ12, complete genome	324	2269	100%	5.00E-86	100%	NZ_CP010359.1
Pseudomonas mosselii SJ10, complete genome	324	2269	100%	5.00E-86	100%	NZ_CP009365.1
Pseudomonas sp. FGI182, complete genome	324	2269	100%	5.00E-86	100%	NZ_CP007012.1
Pseudomonas cremoricolorata strain ND07, complete genome	324	1945	100%	5.00E-86	100%	NZ_CP009455.1
Pseudomonas putida strain DLL-E4, complete genome	324	1606	100%	5.00E-86	100%	NZ_CP007620.1
Pseudomonas parafulva strain CRS01-1, complete genome	324	2269	100%	5.00E-86	100%	NZ_CP009747.1
Pseudomonas putida S12, complete genome	324	1945	100%	5.00E-86	100%	NZ_CP009974.1
Pseudomonas monteilii SB3101, complete genome	324	1945	100%	5.00E-86	100%	NC_023076.1
Human lung- associated *Pseudomonas* OTU#1						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas fluorescens PICF7, complete genome	448	2688	100%	5.00E-123	100%	NZ_CP005975.1
Pseudomonas fluorescens strain PCL1751, complete genome	448	2688	100%	5.00E-123	100%	NZ_CP010896.1
Pseudomonas simiae strain WCS417	448	896	100%	5.00E-123	100%	NZ_CP007637.1
Pseudomonas sp. WCS374	448	2688	100%	5.00E-123	100%	NZ_CP007638.1
Pseudomonas poae RE*1-1-14, complete genome	448	2240	100%	5.00E-123	100%	NC_020209.1
Pseudomonas fluorescens A506, complete genome	448	2688	100%	5.00E-123	100%	NC_017911.1
Janthinobacterium agaricidamnosum NBRC 102515 = DSM 9628, complete genome	442	615	100%	2.00E-121	99%	NZ_HG322949.1
Pseudomonas fluorescens strain LBUM223, complete genome	436	2621	100%	1.00E-119	99%	NZ_CP011117.1
Pseudomonas fluorescens strain UK4, complete genome	431	2588	100%	5.00E-118	99%	NZ_CP008896.1
Pseudomonas sp. TKP, complete genome	431	2156	100%	5.00E-118	99%	NC_023064.1
Pseudomonas fluorescens SBW25 complete genome	431	2156	100%	5.00E-118	99%	NC_012660.1
Human lung- associated *Pseudomonas* OTU#2						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas aeruginosa strain S04 90 genome	448	1792	100%	5.00E-123	100%	NZ_CP011369.1

Pseudomonas aeruginosa strain Carb01 63, complete genome	448	1792	100%	5.00E-123	100%	NZ_CP011317.1
Pseudomonas aeruginosa DNA, complete genome, strain: NCGM 1984	448	1792	100%	5.00E-123	100%	NZ_AP014646.1
Pseudomonas aeruginosa DNA, complete genome, strain: NCGM 1900	448	1792	100%	5.00E-123	100%	NZ_AP014622.1
Pseudomonas aeruginosa YL84, complete genome	448	1792	100%	5.00E-123	100%	NZ_CP007147.1
Pseudomonas aeruginosa strain PSE305	448	448	100%	5.00E-123	100%	NZ_HG974234.1
Pseudomonas aeruginosa strain FRD1, complete genome	448	1792	100%	5.00E-123	100%	NZ_CP010555.1
Pseudomonas aeruginosa LESlike1 sequence	448	1792	100%	5.00E-123	100%	NZ_CP006984.1
Pseudomonas aeruginosa VRFPA04, complete genome	448	3584	100%	5.00E-123	100%	NZ_CP008739.1
Pseudomonas aeruginosa PA96	448	1792	100%	5.00E-123	100%	NZ_CP007224.1
Pseudomonas aeruginosa PAO1H2O	448	1773	100%	5.00E-123	100%	NZ_CP008749.1
Human lung- associated *Pseudomonas* OTU#3						
Description	Max score	Total score	Query cover	E value	Identity	Accession
Pseudomonas plecoglossicida strain NyZ12, complete genome	490	3433	100%	8.00E-136	100%	NZ_CP010359.1
Pseudomonas sp. FGI182, complete genome	490	3433	100%	8.00E-136	100%	NZ_CP007012.1
Pseudomonas putida strain DLL-E4, complete genome	490	2432	100%	8.00E-136	100%	NZ_CP007620.1
Pseudomonas monteilii SB3101, complete genome	490	2942	100%	8.00E-136	100%	NC_023076.1
Pseudomonas monteilii SB3078, complete genome	490	2942	100%	8.00E-136	100%	NC_023075.1
Pseudomonas sp. VLB120, complete genome	490	3433	100%	8.00E-136	100%	NC_022738.1
Pseudomonas putida S16, complete genome	490	2933	100%	8.00E-136	100%	NC_015733.1
Pseudomonas mendocina NK-01, complete genome	490	1961	100%	8.00E-136	100%	NC_015410.1
Pseudomonas mendocina ymp, complete genome	490	1961	100%	8.00E-136	100%	NC_009439.1
Pseudomonas sp. CCOS191 genome assembly Pseudomonas sp. strain CCOS191, chromosome : I	484	3389	100%	4.00E-134	99%	NZ_LN847264.1
Human lung- associated *Pseudomonas* OTU#4						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas syringae pv. syringae B301D, complete genome	468	2341	100%	4.00E-129	98%	NZ_CP005969.1
Pseudomonas syringae pv. syringae HS191, complete genome	468	2341	100%	4.00E-129	98%	NZ_CP006256.1
Pseudomonas protegens Cab57 DNA, complete genome	468	2341	100%	4.00E-129	98%	NZ_AP014522.1
Pseudomonas syringae CC1557, complete sequence	468	2341	100%	4.00E-129	98%	NZ_CP007014.1
Pseudomonas protegens CHA0, complete genome	468	2341	100%	4.00E-129	98%	NC_021237.1
Pseudomonas syringae pv. phaseolicola 1448A, complete genome	468	2341	100%	4.00E-129	98%	NC_005773.3
Pseudomonas protegens Pf-5, complete genome	468	2341	100%	4.00E-129	98%	NC_004129.6
Pseudomonas syringae pv. syringae B728a chromosome, complete genome	468	2341	100%	4.00E-129	98%	NC_007005.1
Pseudomonas syringae pv. tomato str. DC3000 chromosome, complete genome	468	2341	100%	4.00E-129	98%	NC_004578.1
Pseudomonas sp. St29 DNA, complete genome	468	2809	100%	4.00E-129	98%	AP014628.1
Human lung- associated *Pseudomonas* OTU#5						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas putida S13.1.2, complete genome	435	3045	97%	4.00E-119	97%	NZ_CP010979.1
Pseudomonas pseudoalcaligenes CECT 5344 complete genome	435	1740	97%	4.00E-119	97%	NZ_HG916826.1
Pseudomonas pseudoalcaligenes genome assembly Ppseudo_Pac, chromosome : I	435	1740	97%	4.00E-119	97%	NZ_LK391695.1
Pseudomonas plecoglossicida strain NyZ12, complete genome	435	3045	97%	4.00E-119	97%	NZ_CP010359.1
Pseudomonas sp. FGI182, complete genome	435	3045	97%	4.00E-119	97%	NZ_CP007012.1

Pseudomonas putida strain DLL-E4, complete genome	435	2155	97%	4.00E-119	97%	NZ_CP007620.1
Pseudomonas parafulva strain CRS01-1, complete genome	435	3045	97%	4.00E-119	97%	NZ_CP009747.1
Pseudomonas putida S12, complete genome	435	2610	97%	4.00E-119	97%	NZ_CP009974.1
Pseudomonas monteilii SB3101, complete genome	435	2610	97%	4.00E-119	97%	NC_023076.1
Pseudomonas monteilii SB3078, complete genome	435	2610	97%	4.00E-119	97%	NC_023075.1
Human lung- associated *Pseudomonas* OTU#6						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas stutzeri strain 19SMN4, complete genome	490	1961	100%	8.00E-136	100%	NZ_CP007509.1
Pseudomonas stutzeri RCH2, complete genome	490	1961	100%	8.00E-136	100%	NC_019936.1
Pseudomonas stutzeri DSM 4166, complete genome	490	1961	100%	8.00E-136	100%	NC_017532.1
Pseudomonas stutzeri ATCC 17588 = LMG 11199, complete genome	490	1945	100%	8.00E-136	100%	NC_015740.1
Pseudomonas stutzeri A1501, complete genome	490	1961	100%	8.00E-136	100%	NC_009434.1
Pseudomonas sp. MT-1	484	1590	100%	4.00E-134	99%	NZ_AP014655.1
Pseudomonas stutzeri strain 28a24, complete genome	484	1939	100%	4.00E-134	99%	NZ_CP007441.1
Pseudomonas stutzeri CCUG 29243, complete genome	484	1939	100%	4.00E-134	99%	NC_018028.1
Pseudomonas balearica DSM 6083, complete genome	479	1910	97%	2.00E-132	100%	NZ_CP007511.1
Pseudomonas stutzeri DSM 10701, complete genome	473	1895	100%	8.00E-131	99%	NC_018177.1
Human-associated Pseudomonas OTU#7						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas pseudoalcaligenes genome assembly Ppseudo_Pac, chromosome : I	446	1773	100%	2.00E-122	97%	NZ_LK391695.1
Pseudomonas pseudoalcaligenes CECT 5344 complete genome	440	1762	100%	8.00E-121	97%	NZ_HG916826.1
Pseudomonas chlororaphis strain PCL1606, complete genome	435	2175	100%	4.00E-119	96%	NZ_CP011110.1
Pseudomonas sp. MRSN12121, complete genome	435	2175	100%	4.00E-119	96%	NZ_CP010892.1
Pseudomonas chlororaphis strain PA23, complete genome	435	2169	100%	4.00E-119	96%	NZ_CP008696.1
Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	435	2164	100%	4.00E-119	96%	NZ_CP009290.1
Pseudomonas fluorescens Pf0-1, complete genome	435	2610	100%	4.00E-119	96%	NC_007492.2
Pseudomonas sp. CCOS191 genome assembly Pseudomonas sp. strain CCOS191, chromosome : I	429	3001	100%	2.00E-117	96%	NZ_LN847264.1
Pseudomonas syringae pv. syringae B301D, complete genome	429	2147	100%	2.00E-117	96%	NZ_CP005969.1
Pseudomonas syringae pv. syringae HS191, complete genome	429	2147	100%	2.00E-117	96%	NZ_CP006256.1
Human-associated Pseudomonas OTU#8						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Accession**
Pseudomonas chlororaphis strain PCL1606, complete genome	486	2433	100%	1.00E-134	100%	NZ_CP011110.1
Pseudomonas sp. MRSN12121, complete genome	486	2433	100%	1.00E-134	100%	NZ_CP010892.1
Pseudomonas chlororaphis strain PA23, complete genome	486	2428	100%	1.00E-134	100%	NZ_CP008696.1
Pseudomonas chlororaphis subsp. aurantiaca strain JD37, complete genome	486	2422	100%	1.00E-134	100%	NZ_CP009290.1
Pseudomonas fluorescens Pf0-1, complete genome	486	2920	100%	1.00E-134	100%	NC_007492.2
Pseudomonas brassicacearum subsp. brassicacearum NFM421, complete genome	484	2419	99%	4.00E-134	100%	NC_015379.1
Pseudomonas syringae pv. syringae B301D, complete genome	481	2406	98%	5.00E-133	100%	NZ_CP005969.1
Pseudomonas syringae pv. syringae HS191, complete genome	481	2406	98%	5.00E-133	100%	NZ_CP006256.1
Pseudomonas mandelii JR-1, complete genome	481	2887	98%	5.00E-133	100%	NZ_CP005960.1
Pseudomonas protegens Cab57 DNA, complete genome	481	2406	98%	5.00E-133	100%	NZ_AP014522.1

Appendix Figure VI.1: Blast results of *Pseudomonas* OTUs.
The nucleotide sequence of the V3-V5 16S rRNA region of each *Pseudomonas* was used to screen against NCIB's online Microbial nucleotide database.

Appendix Figure VI.2: Blast alignment of *Pseudomonas* OTUs against reference *Pseudomonas* genomes.

Mouse lung-associated *Pseudomonas* OTU#1						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	427	1699	100%	1.00E-119	98%	0/249
Pseudomonas aeruginosa UCBPP-PA14	427	1710	100%	3.00E-120	98%	0/249
Pseudomonas fluorescens SBW25	460	2304	100%	3.00E-130	100%	0/249
Pseudomonas fluorescens A506	444	2665	100%	2.00E-125	99%	0/249
Pseudomonas protegens Pf-5	460	2304	100%	3.00E-130	100%	0/249
Pseudomonas putida GB-1	438	3071	100%	1.00E-123	98%	0/249
Mouse lung-associated *Pseudomonas* OTU#2						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	427	1699	100%	3.00E-120	98%	0/249
Pseudomonas aeruginosa UCBPP-PA14	427	1710	100%	3.00E-120	98%	0/249
Pseudomonas fluorescens SBW25	438	2193	100%	1.00E-123	98%	0/249
Pseudomonas fluorescens A506	444	2665	100%	2.00E-125	99%	0/249
Pseudomonas protegens Pf-5	438	2193	100%	1.00E-123	98%	0/249
Pseudomonas putida GB-1	460	3226	100%	2.00E-130	100%	0/249
Human lung- associated *Pseudomonas* OTU#1						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	383	1522	99%	5.00E-107	95%	1/241
Pseudomonas aeruginosa UCBPP-PA14	383	1533	99%	6.00E-107	95%	1/241
Pseudomonas fluorescens SBW25	431	2156	100%	2.00E-121	99%	0/242
Pseudomonas fluorescens A506	448	2688	100%	2.00E-126	100%	0/242
Pseudomonas protegens Pf-5	418	2092	97%	2.00E-117	99%	0/235
Pseudomonas putida GB-1	412	2890	97%	7.00E-116	98%	0/235
Human lung- associated *Pseudomonas* OTU#2						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	448	1780	100%	2.00E-126	100%	0/242
Pseudomonas aeruginosa UCBPP-PA14	448	1792	100%	2.00E-126	100%	0/242
Pseudomonas fluorescens SBW25	399	1999	99%	6.00E-112	97%	1/241
Pseudomonas fluorescens A506	383	2300	99%	5.00E-107	95%	1/241
Pseudomonas protegens Pf-5	399	1999	97%	6.00E-112	97%	0/237
Pseudomonas putida GB-1	405	2838	97%	1.00E-113	97%	0/237
Human lung- associated *Pseudomonas* OTU#3						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	449	1788	98%	6.00E-127	98%	0/261
Pseudomonas aeruginosa UCBPP-PA14	449	1799	98%	6.00E-127	98%	0/261
Pseudomonas fluorescens SBW25	449	2249	97%	6.00E-127	98%	0/258
Pseudomonas fluorescens A506	455	2737	97%	1.00E-128	98%	0/258
Pseudomonas protegens Pf-5	457	2286	100%	4.00E-129	98%	0/265

Pseudomonas putida GB-1	484	3394	100%	2.00E-137	99%	0/265
Human lung- associated *Pseudomonas* OTU#4						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	431	1714	98%	2.00E-121	97%	0/260
Pseudomonas aeruginosa UCBPP-PA14	431	1725	98%	2.00E-121	97%	0/260
Pseudomonas fluorescens SBW25	455	2276	97%	1.00E-128	98%	0/258
Pseudomonas fluorescens A506	438	2632	97%	1.00E-123	97%	0/258
Pseudomonas protegens Pf-5	468	2341	100%	2.00E-132	98%	0/265
Pseudomonas putida GB-1	451	3161	100%	2.00E-127	97%	0/265
Human lung- associated *Pseudomonas* OTU#5						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	424	1684	98%	4.00E-119	96%	1/263
Pseudomonas aeruginosa UCBPP-PA14	424	1696	98%	4.00E-119	96%	1/263
Pseudomonas fluorescens SBW25	429	2147	98%	8.00E-121	96%	0/262
Pseudomonas fluorescens A506	412	2477	98%	7.00E-116	95%	0/262
Pseudomonas protegens Pf-5	429	2147	97%	9.00E-121	97%	0/259
Pseudomonas putida GB-1	435	3045	97%	2.00E-122	97%	0/259
Human lung- associated *Pseudomonas* OTU#6						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	449	1788	98%	6.00E-127	98%	0/261
Pseudomonas aeruginosa UCBPP-PA14	449	1799	98%	6.00E-127	98%	0/261
Pseudomonas fluorescens SBW25	438	2193	97%	1.00E-123	97%	0/258
Pseudomonas fluorescens A506	433	2599	97%	6.00E-122	97%	0/258
Pseudomonas protegens Pf-5	446	2230	100%	8.00E-126	97%	0/265
Pseudomonas putida GB-1	440	3084	100%	3.00E-124	97%	0/265
Human lung- associated *Pseudomonas* OTU#7						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	405	1611	98%	1.00E-113	95%	0/261
Pseudomonas aeruginosa UCBPP-PA14	405	1622	985	1.00E-113	95%	0/261
Pseudomonas fluorescens SBW25	422	2110	97%	1.00E-118	96%	0/258
Pseudomonas fluorescens A506	416	2499	97%	6.00E-117	96%	0/258
Pseudomonas protegens Pf-5	429	2147	100%	9.00E-121	96%	0/265
Pseudomonas putida GB-1	418	2929	100%	2.00E-117	95%	0/265
Human lung- associated *Pseudomonas* OTU#8						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Identity**	**Gaps**
Pseudomonas aeruginosa PAO1	444	1766	99%	3.00E-125	97%	0/261
Pseudomonas aeruginosa UCBPP-PA14	444	1777	99%	3.00E-125	97%	0/261
Pseudomonas fluorescens SBW25	477	2387	98%	3.00E-135	100%	0/258
Pseudomonas fluorescens A506	460	2765	98%	3.00E-130	99%	0/258
Pseudomonas protegens Pf-5	481	2406	98%	2.00E-136	100%	0/260
Pseudomonas putida GB-1	453	3174	98%	4.00E-128	98%	0/260

Appendix Figure VI.2: Blast alignment of *Pseudomonas* OTUs against reference *Pseudomonas* genomes.
The nucleotide sequence of the V3-V5 rRNA region of each *Pseudomonas* OTU was used to screen against reference *Pseudomonas* genomes using the blastn tool on NCBI's online database.

Appendix Figure VI.3: Blast and alignment of the two mouse lung- associated *Streptococcus* OTUs.

Mouse lung-associated *Streptococcus* OTU#1						
Description	Max score	Total score	Query cover	E value	Identity	Accession
Streptococcus agalactiae strain 138spar, complete genome	435	2610	100%	4.00E-119	98%	NZ_CP007565.1
Streptococcus agalactiae 138P, complete genome	435	2610	100%	4.00E-119	98%	NZ_CP007482.1
Streptococcus iniae strain YSFST01-82, complete genome	435	2169	100%	4.00E-119	98%	NZ_CP010783.1
Streptococcus agalactiae strain GBS2-NM, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP007571.1
Streptococcus agalactiae strain GBS1-NY, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP007570.1
Streptococcus agalactiae strain GBS6, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP007572.1
Streptococcus agalactiae CNCTC 10/84, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP006910.1
Streptococcus agalactiae strain NGBS572, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP007632.1
Streptococcus agalactiae, strain COH1, complete genome	435	3045	100%	4.00E-119	98%	NZ_HG939456.1
Streptococcus agalactiae strain NGBS061, complete genome	435	3045	100%	4.00E-119	98%	NZ_CP007631.1
Streptococcus lutetiensis 033, complete genome	435	2610	100%	4.00E-119	98%	NC_021900.1
Streptococcus agalactiae ILRI005 complete genome	435	3045	100%	4.00E-119	98%	NC_021486.1
Streptococcus agalactiae 09mas018883 complete genome	435	3045	100%	4.00E-119	98%	NC_021485.1
Streptococcus agalactiae 2-22, complete genome	435	5220	100%	4.00E-119	98%	NC_021195.1
Streptococcus agalactiae SA20-06, complete genome	435	3045	100%	4.00E-119	98%	NC_019048.1
Streptococcus agalactiae GD201008-001, complete genome	435	3014	100%	4.00E-119	98%	NC_018646.1
Streptococcus gallolyticus subsp. gallolyticus ATCC 43143 DNA, complete genome	435	2175	100%	4.00E-119	98%	NC_017576.1
Mouse lung-associated *Streptococcus* OTU#2						
Description	Max score	Total score	Query cover	E value	Identity	Accession
Streptococcus gordonii str. Challis substr. CH1, complete genome	446	1784	100%	2.00E-122	99%	NC_009785.1
Streptococcus agalactiae strain 138spar, complete genome	440	2643	100%	8.00E-121	98%	NZ_CP007565.1
Streptococcus agalactiae 138P, complete genome	440	2643	100%	8.00E-121	98%	NZ_CP007482.1
Streptococcus iniae strain YSFST01-82, complete genome	440	2197	100%	8.00E-121	98%	NZ_CP010783.1
Streptococcus agalactiae strain GBS2-NM, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP007571.1
Streptococcus agalactiae strain GBS1-NY, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP007570.1
Streptococcus agalactiae strain GBS6, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP007572.1
Streptococcus agalactiae CNCTC 10/84, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP006910.1
Streptococcus agalactiae strain NGBS572, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP007632.1
Streptococcus agalactiae, strain COH1, complete genome	440	3084	100%	8.00E-121	98%	NZ_HG939456.1
Streptococcus agalactiae strain NGBS061, complete genome	440	3084	100%	8.00E-121	98%	NZ_CP007631.1
Streptococcus lutetiensis 033, complete genome	440	2643	100%	8.00E-121	98%	NC_021900.1
Streptococcus agalactiae ILRI005 complete genome	440	3084	100%	8.00E-121	98%	NC_021486.1
Streptococcus agalactiae 09mas018883 complete genome	440	3084	100%	8.00E-121	98%	NC_021485.1
Streptococcus agalactiae 2-22, complete genome	440	2643	100%	8.00E-121	98%	NC_021195.1
Streptococcus agalactiae SA20-06, complete genome	440	3084	100%	8.00E-121	98%	NC_019048.1

Streptococcus agalactiae GD201008-001, complete genome	440	3052	100%	8.00E-121	98%	NC_018646.1
Alignment between the two *Streptococcus* OTUs	Max score	Total score	Query cover	E value	Identity	Gaps
	457	457	100%	1.00E-133	99%	0/250

Appendix Figure VI.3: Blast and alignment of the two mouse lung- associated *Streptococcus* OTUs.
The nucleotide sequence of the V3-V5 16S rRNA region of the two mouse lung-associated *Streptococcus* OTUs were used to screen against NCIB's online Microbial nucleotide database. The two *Streptococcus* OTUs were also aligned together using NCBI's online blastn tool.

Chapter VII

Finale

THESIS SUMMARY

The aim of this dissertation is to utilize a comparative genomic approach to gain insight into potential mechanisms involved in *P. fluorescens*-human interactions in health and disease. The first step in achieving this aim was to broaden the limited taxonomy of the *P. fluorescens* species complex (Specific Aim #1, **Chapter I**). Prior to the work presented here, all sequenced strains of *P. fluorescens* were isolated from non-human host samples and, therefore, all knowledge on the phylogeny of the *P. fluorescens* species complex was based on the genetics of environmentally-isolated strains. As presented in **Chapter III**, we sequenced 60 clinical *Pseudomonas spp.* isolates from individuals with chronic lung disease. Using a MLSA approach, we inferred phylogenetic trees based on the concatenated sequences of housekeeping genes. This analysis placed 22 of the 60 strains into subclades of the *P. fluorescens* species complex. Three strains were placed in subclade I, two strains in subclade II, seven strains in subclade III and ten strains were placed in a new fourth subclade. While the other three subclades contain previously sequenced environmental strains, subclade IV contained only newly sequenced clinical strains. The branching pattern of the phylogenetic trees presented in **Figures III.1, III.3**

III.4 suggest that subclade IV strains are highly dissimilar genetically from other members of the *P. fluorescens* species complex but highly similar to each other. The data in **Chapter III** successfully achieves Specific Aim #1 in widening the taxonomy of the *P. fluorescens* species complex to include strains isolated from human clinical samples.

The next goal was to predict functional attributes related to survival in the human host through comparative genomics of the newly sequenced clinical *P. fluorescens* species strains (Specific Aim #2, **Chapter I**). **Chapter IV** presents detailed genomic analysis of the clinical subclade III strains and compares them to the genetic features of subclade III environmental strains. One of the key findings from this analysis is that clinical subclade III strains contain additional homologues to genes involved in metal resistance. Metal resistance in bacteria is linked to resistance to antibiotics (1-4). Therefore, the increased number of metal resistance genes in the clinical subclade III strains could provide a potential fitness advantage in the antibiotic-laden environment of the CF lung.

Chapter V investigated the genomics of the ten clinical strains that formed the novel subclade IV in **Chapter III** and discovered multiple features that distinguish the subclade IV strains from the rest of the *P. fluorescens* species complex. Subclade IV strains have significantly smaller genomes, lower GC content, and decreased diversity in unique predicted coding regions. All three aspects are commonly seen as bacteria that have adapted from thriving in a multitrophic environment to a single, more narrow environmental niche. In addition, subclade IV strains contained the complete consortium of genes involved in expressing the same type-III secretion system found in *P. aeruginosa,* a *Pseudomonas spp.* commonly associated with human disease. Together, **Chapters IV** and **V** provide the first detailed genetic analysis of *P. fluorescens* strains from clinical samples and outline genetic attributes related to survival in the human host.

The final goal was to investigate how inflammation associates with *P. fluorescens* growth in mammalian hosts (Specific Aim #3, **Chapter I**). We first tested how chronic inflammation affects the lung microbiota in genetically inbreed mice sourced either from a colony that contained indigenous *P. fluorescens* or a colony that does not (**Chapter VII**). Mice that contain indigenous *P. fluorescens* saw a *P. fluorescens* bloom after both four and eight weeks of inflammation. Mice that do not contain *P. fluorescens* do not see such a bloom in *Pseudomonas spp.* bacteria. We next taxonomically classified *Pseudomonas* OTUs from previous human lung microbiome datasets. The human lung datasets represented individuals with idiopathic pulmonary fibrosis (IPF), those who had received a lung transplant (LT), those diagnosed with moderate chronic obstructive pulmonary disease (COPD), those diagnosed with severe COPD and those without disease (healthy). We identified multiple *P. fluorescens* OTUs in the human lung microbiome datasets and found that lung disease was associated with an increase in relative abundance of *P. fluorescens* OTU reads. The data presented in **Chapter VII** addresses Specific Aim #3 and finds that inflammation in the lungs favors the growth of *P. fluorescens* species bacteria.

What makes the conclusions presented in this thesis particularly interesting is that, prior to this undertaking, strains within the *P. fluorescens* species complex were almost completely identified as environmental bacteria and only accidental colonizers of humans. We hypothesized that strains of *P. fluorescens* contain genetic adaptations that favor their selection and growth in a diseased lung but do not readily cause acute disease, and therefore they are under-appreciated colonizers of humans. The idea that changes in the environment of the human host (i.e. inflammation) could lead to the selection of pre-adapted bacterial strains from environmental samples is not new but is often ignored in medical microbiology, where microbes are typically given finite labels of ‘pathogenic’ or ‘non-pathogenic’. The conclusions from the

comparative genomics of *P. fluorescens* clinical strains suggests that, rather than trying to reduce bacteria to simplistic titles, a more complex conceptual framework is needed that takes into account the constant adaptation of bacteria to constantly changing environment that may or may not pre-select them for survival or infection in a human host.

FUTURE DIRECTIONS

Verifying comparative genomics findings

Comparative genomics revealed multiple genomic attributes in clinical *P. fluorescens* strains potentially related to colonization and survival in a human host. The power of *in silco* genomic analysis is that it can reveal specific features of importance for future investigation. However, comparative genomics often can only point to potential biological mechanisms and to prove functionality additional *in vitro* work is required.

We would test the relevance of the key findings from comparative genomics of subclade III and IV strains by answering the following questions:

1. Does the increased number of metal resistance genes lead to increase antibiotic resistance in the clinical subclade III strains?

Prediction: Clinical strains would have increased resistance to antibiotics due to their increased number of metal resistance genes.

Experiments:

a. Compare the ability of clinical and environmental subclade III strains to resist antibiotics in an *in vitro* stetting. Plate the strains in differing concentrations of antibiotics, focusing on those in the carbapenem class (i.e. imipenem).

b. Knockout the additional copies of metal resistance genes in the clinical strains and test whether this modulates their ability to resist antibiotics *in vitro.*

2. Is the Yop-like T3SS found in subclade IV clinical strains functional (i.e., does it secrete the effector ExoU)? If so, is the effector ExoU cytotoxoic against human cells in culture?

Prediction: Due to completeness of the T3SS gene cluster in subclade IV strains, the secretion apparatus is functional and induction leads to ExoU secretion and cytotoxity against human cells.

Experiments:

a. To measure the secretion of ExoU, we will grow the subclade IV strains in low calcium media (which induces ExoU secretion) or standard minimal media, collect the supernatants and separate the concentrated supernatants on a SDS-PAGE gel and identify a protein that separates at the known molecular weight for ExoU (5).

b. To measure the phospholipase activity of ExoU, we will use a fluorescence-based assay where cleavage of a dye-labeled molecule leads to a measurable fluorescent signal (5).

c. To test the cytotoxicity of the secreted T3SS effector, we will culture subclade IV strains in low calcium media, apply the subclade IV strains to human epithelial cells grown *in vitro* and measure cell rounding (6).

If our prediction is correct, these future experiments will represent the first instance of a functional Yop-like T3SS in *P. fluorescens* strains.

3. Are *P. fluorescens* subclade IV strains found in habitats other than the diseased human lung?

Prediction: Subclade IV strains are found in both environmental (ie. soil, water, plants) and non-environmental (ie. showerheads, hospital equipment) niches outside of the human host.

Experiments:

a. To test this prediction, samples are collected from environment and non-environmental sources for DNA isolation. The isolated DNA is screened with a series of PCR primer sets (*Pseudomonas*-genus specific, *P. aeruginosa*-species specific and ExoU-subclade IV specific) to look for subclade IV specific genetic signatures. A positive result for the *Pseudomonas*-genus and ExoU-specific primer sets, combined with a negative result for the *P. aeruginosa*-specific primer set, suggests that the collected sample contains a subclade IV strain.

b. A portion of each sample is also saved for *in vitro* cultivation. When a sample is shown to contain a potential subclade IV strain, that sample is plated on standard media (LB agar). Colonies are selected for colony PCR with a partial 16S rRNA primer set. The amplified genetic region is isolated with gel electrophoresis, excised and sequenced. The sequenced partial 16S rRNA gene is then used to screen NCBI's nonredundant nucleotide database to verify it is likely a *P. fluorescens*-like strain. That same partial 16S rRNA gene sequence is also compared to the same region in the draft genomes of the previously sequenced subclade IV strains.

Verifying findings from lung inflammation mouse model

Additional work is also required to understand how inflammation promotes the growth of *P. fluorescens* species bacteria. Inflammation has been shown to increase the growth of Proteobacteria in mouse models of gastrointestinal inflammation (7-10), in human patients (11-14) and in the mouse model of pulmonary inflammation in **Chapter VII.** Catecholamines provide one possible connection between inflammation and growth promotion in bacteria. Catecholamines are organic compounds that contain a benzene with two hydroxyl side groups (called a catechol) and a side chain amine (15). Two well known catecholamines are epinephrine (adrenaline) and norepinephrine (noradrenaline) (15). These two molecules are important components of the 'fight and flight' response and interact directly with the immune system. Lymphocytes and phagocytes synthesize, release and respond to epinephrine and norepinephrine, leading to lymphocyte proliferation, differentiation and even cytokine production (15). These two molecules also interact with autoinducer-3 (AI-3) quorum sensing (QS) signaling through direct binding to the QscC sensor kinase that also bind AI-3 (16). *P. aeruginosa* is an example of a bacteria species that utilizes a AI-3 QS signaling pathway and has been shown to respond to the release of norepinephrine in gut-derived sepsis by increasing the expression of virulence traits (17). Both epinephrine and norepinephrine increase growth of *P. aeruginosa, K. pneumonaie* and *E. coli* in *in vitro* culture (18). Future experiments in our lab will apply these findings to clinical strains of *P. fluorescens.*

Prediction: The presence of epinephrine and norepinephrine will promote the growth of clinical *P. fluorescens* strains in media.

Experiments:

a. Inoculate clinical strains of *P. fluorescens* from each subclade in the presence of a range of concentrations of epinephrine and norepinephrine in culture and measure growth after 1, 2, 6, 12, 24 and 48 hours.

b. Repeat the above experiment, including a group that also receives LED209, a compound that blocks catecholamine binding to the QscC sensor kinase.

c. Pre-treat clinical *P. fluorescens* strains with epinephrine, norepinephrine or LED209 and inoculate in mice whose lungs have been chronically inflamed for four weeks. After 6, 12 and 24 hours post-inoculate, collect mouse lungs and lung supernatants. Look for growth after 24 and 48 hours and use colonyPCR to verify the identity of any growth.

These experiments will further out understanding on how inflammation in the host affects the growth of bacteria and help explain the association between *P. fluorescens* and chronically inflamed lungs.

CONCEPTUAL FRAMEWORK

The rhizosphere vs. the human host

While it is common in medical microbiology to reduce the non-human 'environment' to simple niches of water, air and soil, environmental habitats are continuous gradients that contain diverse communities of bacteria, fungi, invertebrates, plants, protozoa and other eukaryotic species. In these environmental habitats, the non-microbial organisms that bacteria live on, in and alongside are often referred to as 'shelter' organisms. Close associations between bacteria and their shelter organisms create hotspots for the exchange of genetic material and can lead to the acquisition of new genetic material into the bacterial world. (19-22).

Genetic exchange between environmental bacteria and their shelter organisms is an important contributing factor towards the emergence of many human pathogens and has generally shaped the overall genetic evolution of bacteria (19-23). Environmental habitats thus act as 'genetic melting pots' and are a ready source of new bacterial strains that can potentially colonize and infect human hosts (19, 21, 24-26).

One of the most important environmental hotspots for the exchange of genetic material of new bacterial traits is the rhizosphere. The term 'rhizosphere' was coined by Hiltner in 1904 and is defined as 'the layer of soil influenced by root metabolism. (26). Plant roots release a long list of substances, such as organic acids, amino acids, vitamins, sugars and carbohydrates, that lead to a nutrient-rich niche for microorganisms to thrive (27). Compared to non-root soil, the bulk mass and diversity of microorganisms is significantly higher in the rhizosphere and this provides many opportunities for genetic transfer between different bacteria strains and species (26). Evidence exits that conjugative transfer of chromosomal genes can occur between *P. aeruginosa* PAO1, known for causing facultative disease in compromised human hosts, and *P. aeruginosa* CHA0, known for its biocontrol abilities (28, 29). Strains of gram-negative *Stentrophomonas* have acquired antibiotic and heavy metal resistance genes that are isoforms of genes found on gram-positive *Staphylococcus aureus* plasmids (29). Genes that are related to the pathogenicity or fitness of bacteria in human hosts often occur in genetic blocks, called islands, which carry signatures of recent horizontal gene transfer (30). In our analysis of newly sequenced *P. fluorescens* strains in subclades I, II, III and IV, we found many genomic islands with significantly lower G+C content than the rest of the genome (referenced to as G+C Islands or GCIs), suggesting these genomic regions were acquired recently through horizontal gene transfer (**Figure 5, Subclade III chapter**). When the genetic sequences of these GCIs were screened against NCBI's Microbial nucleotide database it was revealed that many of these GCI were acquired from other rhizosphere bacteria species (**Additional File 5, Subclade III**

chapter). Even though all of the *P. fluorescens* strains sequenced in this study were isolated from human patients, their genomes still carry evidence of life in the rhizosphere, including the genomic islands recently acquired from other rhizosphere bacteria.

The process of adapting to certain environmental habitats, such as the rhizosphere, can indirectly acclimate bacteria to a life within a human host. Many of the same cell-surface proteins that required for the initial attachment and colonization of plant roots, such as fimbriae, flagella and O-antigen lipopolysaccharides (LPS), are also utilized in the attachment and colonization of human cells (31-33). Bacterial compounds that trigger that plant innate immune response, e.g. flagella and LPS, also triggers the innate immune response in humans (34). In order to thrive in association with plants and the rhizosphere, bacteria adapt to form biofilms, produce bio-surfactants and degrade pathogenicity factors produced by phytopathogenic bacteria. All of these attributes also benefit a bacterium when surviving in a human host. The genera *Burkholderia, Enterobacter, Ralstonia, Staphylococcus, Stenotrophomonas* and *Pseudomonas* all contain plant and root- associated bacteria that also cause infections in human hosts (26). Many of the attributes that rhizobacteria, such as fluorescent pseudomonads, acquire in order to thrive in the rhizosphere can also enhance their ability to thrive in a human host.

One common genetic adaptation that benefits survival in both the rhizosphere and human host is the acquisition of gene clusters involved in the production of antibiotics and resistance to antibiotics. Rhizobacteria, including members of the *P. fluorescens* species complex, are well known for their production of antibiotics (35). Phenazine (36-38); hydrogen cyanide (HCN) (39); 2,4-diacetylphloroglucinol (DAPG) (40); rhizoxin (41-43); and pyolutorin (44, 45) are all examples of anti-microbial secondary metabolites produced by strains within the *P. fluorescens* species complex. As many other bacteria in the rhizosphere produce antibiotics,

pseudomonads also adapt the counter-measure tool of antibiotic resistance (46). There is also a strong connection between the development of antibiotic resistance and heavy metal resistance (1, 4). In our analysis of subclade III *P. fluorescens* species strains, we found that strains isolated from CF lungs contained more genes involved in heavy metal resistance than strains isolated from the rhizosphere. In the treatment of CF, patients are often given a long list of different antibiotics, making the CF lung an environment where antibiotic resistance is an important strategy for survival. Bacteria that had already acquired tools for antibiotic resistance would be 'pre-adapted' and have a competition advantage in the antibiotic-laden CF lung. We hypothesize that the unique environment of the cystic fibrosis lung selects for strains of *P. fluorescens* species complex bacteria from environmental sources, such as the rhizosphere, that have acquired genomic traits that assist in their survival within this particular niche of the human host.

The concept that a diseased lung could positively select for specific strains of bacteria was manifested in our animal models of lung inflammation and analysis of human lung datasets. Mice with indigenous *P. fluorescens* bacteria in their respiratory microflora saw a bloom of *P. fluorescens* after four and eight weeks of inflammation (**Figure A, Inflamed mouse lung chapter**) while mice whose microflora lacked indigenous *P. fluorescens* saw no bloom during inflammation (**Figure B, Inflamed mouse lung chapter**). The bloom of *Pseudomonas* bacteria during inflammation was species-specific, as a *Pseudomonas* OTU identified as *P. putida/P. mendocina* did not bloom during inflammation (**Figure B**, **Inflamed mouse lung chapter**). Analysis of human patient datasets that included patients with idiopathic pulmonary fibrosis (IPF), those who had received a lung transplant (LT), those diagnosed with moderate chronic obstructive pulmonary disease (COPD) and those diagnosed with severe COPD also found increased signal of *P. fluorescens* bacteria when compared to healthy human samples (**Figures G and I, Inflamed mouse lung chapter)**. *P. fluorescens*-species bacteria, whose typical

environment includes the rhizosphere, bloomed in relative abundance during lung inflammation, suggesting that they contain specific traits that have pre-adapted them to the particular environment of the diseased mammalian lung.

P. fluorescens subclade IV strains; a snapshot in evolution to a human host

Comparative genomics of bacterial isolates from individuals with cystic fibrosis and other respiratory diseases lead to the discovery of a group of ten strains that formed a distinct, novel subclade within the *P. fluorescens* species complex. The ten strains within the newly formed fourth subclade showed multiple genomic aspects that are significantly different from any other strains within the *P. fluorescens* species complex and suggest that they are undergoing genetic adaptation to the human host. One of the most pronounced differences between the subclade IV strains and other members of the *P. fluorescens* species complex was in genomic size. The subclade IV strains sequenced and analyzed in this study have genomes that, on average, are 1.5 million Mbps smaller than other members of the *P. fluorescens* species complex (**Subclade IV chapter, Figure I**). Reduction in genome size and gene diversity both occur as bacteria evolve from a multi-trophic, environmental lifestyle to a purely human host based lifestyle. Larger bacterial genomes are associated with survival among varied biotic populations in diverse habitats (47, 48). Human pathogens, in contrast, are often associated with smaller genomes, and the reduction of genome content is related to the close association these bacteria have with host cells (49-52). Subclade IV strains also had significantly lower overall G+C content then other members of the *P. fluorescens* species complex. Lower genomic G+C content is another characteristic of an adaptation to life within a human host (**Subclade IV chapter, Figure J**) (53). The reduced size and G+C content of genomes of the subclade IV strains are similar to trends seen in bacteria that evolve from an environmental a human host niche.

The main driving force for the reduced genome size of bacterial strains that have specialized for growth within a human host is that of gene loss (54). Members of the *Rickettsia* genus lost more than 2,000 genes during the evolution to an intracellular lifestyle (55). Which genes were lost during this evolution is reflected in the adaptation of *Rickettsia spp.* to specific host cells and the development of new *Rickettsia* species (55). The subclade IV strains not only have reduced genome size, but a reduction in the diversity of predicted protein coding genes. **Figure A of this chapter** displays the pan, accessory and core of each of the *P. fluorescens* subclades, all the clinical *P. fluorescens* strains, all sequenced (both clinical and environmental) *P. fluorescens* species complex strains; and all fully sequenced *Pseudomonas* genus genomes publically available on NCBI (as of March, 2014). The groups are sorted in **Figure A** from lowest to highest in the number of strains included in each group. The core genome compartment contains the predicted protein coding regions (in this analysis, referred to as COGs) shared between every strain analyzed. The accessory genome compartment is all the gene clusters not the core and contains COGs found in at least one of the strains but not all of the strains. The pan genome contains both the core and accessory genome compartments. As the number of genomes/strains analyzed increases, the core genome typically decreases while the accessory and pan genomes increase, a pattern seen in every group analyzed in **Figure A of this chapter.** However, this is not true for subclade IV (56).

As the number of strains included in the analysis increases from five strains in *P. fluorescens* subclade I to 34 strains in all *P. fluorescens* species complex, the core genome decreases from 4802 to 2488 COGs and the accessory genome increases from 3864 to 17114 COGs. When all 90 sequenced *Pseudomonas* genomes from NCBI are analyzed, the accessory genome jumps to 32737 unique COGs and the core genome shrinks to only 967 shared COGs. Based only on the number of strains per group, subclade IV strains should share an accessory genome between that of subclade II (n=6, accessory genome = 6590 COGs) and subclade III

(n=14, accessory genome = 8183 COGs). Instead, the ten subclade IV strains share an accessory genome of only 2354 COGs, the smallest accessory genome of any of the groups compared. The significantly reduced size of the accessory genome in subclade IV strains reveals that there are fewer unique COGs across the ten strains. The fewer number of unique COGs reveals a narrowing of gene diversity in the subclade IV strains, which alludes to evolution from a varied, environmental habitat (such as the rhizosphere) to a more narrowed habitat (such as the a mammalian host).

The relationship between the newly sequenced clinical subclade IV strains and other members of the *P. fluorescens* species complex is similar to the relationship between human-adapted *Photohabdus asymbioticia* strains and environmental-adapted *Photohabdus luminescens* strains (48). *P. luminescens* is a bioluminescent bacterial species that lives in symbiosis with soil nematodes (57). Together, this bacteria-nematode combination is used as successful biological control agent (58, 59). *P. asymbioticia* is a related species within the same genus that is known to cause soft tissue infections and bacterium in humans (60, 61). On average, the genome size of *P. asymbioticia* strains is ~600 kb smaller than that of *P. luminescens* strains (62). Comparative genomics between a representative strains from each species also found a reduction in the number of protein-coding sequences in the human-associated *P. asymbiotica* than the environmental-associated *P. luminescens*. Though the core and accessory genomes were not directly compared between the two, there was a significant reduction in diversity in genes associated with insecticidal toxin production in the human-associated *P. asymbiotica*. Insecticidal toxin production is an important adaptation in *P. luminescens* for surviving in the non-human environment (62).

Similar to subclade III vs IV strains, *P. luminescens* and *P. asymbiotica* differ in the gene clusters that encode their T3SSs. The T3SS in *P. luminescens* inhibits phagocytosis following

translocation into insect hemocytes (63). *P. asymbiotica* has a similar T3SS but has acquired a T3SS-related effector not found in *P. luminescens* that is a homolog to ExoU (62). Of the four possible effector proteins associated with the Yop-like T3SS in *P. aeruginosa* (ExoS, T, U and Y), ExoU is considered toxic to host tissue (64-66). The T3SS found in *P. asymbiotica*, and its corresponding effector ExoU, fall within the Yop-like family of T3SS, so called because this secretion was first discovered in *Yersinia spp.* bacteria (62, 64). And while ExoU and its corresponding secretion apparatus serve as a human virulence trait during *P. asymbiotica* and *P. aeruginosa*- related infections, it is an example of an pre-adaptation likely acquired in a non-human based habitat. *P. aeruginosa* strains isolated from the environment often carry a Yop-like T3SS that has been implicated as important for defense against grazing amoebas in *P. aeruginosa* biofilms (67). The same T3SS that serves as a weapon against amoeba in the environment becomes an important pathogenicity factor during infection in a human host (67). The newly sequenced *P. fluorescens* subclade IV strains represent the first report of any *P. fluorescens* species bacteria to contain the full gene assortment necessary to express a Yop-like T3SS and contain the corresponding effector/chaperone combination ExoU/SpcU. (**Figure T, Figure U**).

Much like *P. asymbiotica* and *P. aeruginosa,* we propose that the subclade IV strains originated in the rhizosphere and genomic traits, such as a Yop-like T3SS, lead to positive selection in the unique environment of a diseased human lung. Genetic signatures, such as a reduction in genome size, genomic G+C content and overall gene diversity suggest that the subclade IV strains have undergone specific adaptation to the narrowed niche of the human host. Additional *in vitro* and *in vivo* studies are required to assess this theory, such as competition assays in varied environments (rhizosphere, mammalian host) and T3SS/ExoU expression assays. However, through comparative genomics alone we can conclude that the subclade IV strains represent a novel group of *P. fluorescens* bacteria that can colonize a

human host, contain human-associated ‘virulence’ factors and have genomic-wide signs of human-host adaptation.

Damage-response Framework

How does one then classify these novel, human-adapted subclade IV *P. fluorescens* species bacteria? Are they opportunistic pathogens? They carry the same primary T3SS that is considered an important virulence factor that is found in *P. aeruginosa.* Are they human commensals? Or are they adapted to another host and the CF lung provides a similar environment? The reduced size of the subclade IV genomes and loss of genetic diversity suggest that they have adapted, or are adapting, to a host restricted lifestyle. In fact, these two states- opportunistic and commensal- are not individual characteristics of a microorganism, but operate on a continuum that is based on the state of the host. The damage-response framework was put forth to explain the required contribution of host and microbe to defining virulence, pathogenicity and disease.

Virulence, pathogenicity and disease are not singular traits of a microbe, but instead require interaction with a host. Both host and microbe contribute to the final outcome of their interaction, and when the outcome is damage to the host, then disease is said to have occurred (68). The concept that virulence and pathogenicity are not independent entities of a microorganism but can only occur, and be measured, when host and microbe interact, is the central tenant of the damage-response framework (68)**.** The damage-response framework was first introduced to explain the phenomenon that the same bacteria strains can be pathogenic or nonpathogenic depending on the context of the host (69). For example, *Candida albicans* and *Staphylococcus epidermidis*, are leading causes of bloodstream infections in patients whose immune systems are compromised through HIV-infection but are also normal members of the microbial flora that seldom cause disease in non-immune compromised individuals (6969).

Outside of the damage-response framework, a microbe such as *Candida albicans* is given varied labels depending on the situation. In most hosts it is a commensal; in an immune compromised hosts it is an opportunistic pathogen; and in healthy women that suffer from recurrent vaginitis, it is a primary pathogen. The damage-response framework questions the need for this multiple terminology; every microorganism has the potential of acting in an opportunistic, commensal or primary pathogen manner but none of those states are intrinsic aspects of that microbe.

Using the damage-response framework as a guide, we suggest that the subclade IV *P. fluorescens* bacteria represent microorganisms with higher potential to cause damage in a human host than other members of the *P. fluorescens* species complex, but more often, they are living in a state of commensalism with a human host. Due to the presence of a Yop-like T3SS, subclade IV strains have a higher probability of causing host damage than other members of the *P. fluorescens* species complex that do not carry this T3SS. Our mouse model of pulmonary inflammation revealed that *P. fluorescens* bacteria found at low abundance in the lung microbiota of healthy mice will bloom and overwhelm the rest of the microflora during chronic inflammation. The reduced size and gene diversity of the subclade IV strain genomes suggest that they have at least partially adapted to a lifestyle within a human host and this increased contact with a human host also increases the likelihood that a change in the human host immune system could lead the subclade IV strains to transition from a commensal state to a state where host damage occurs (ie. pathogenic state). Together, all these observations come together to portray a model where commensal, host-adapted strains of *P. fluorescens* subclade IV bacteria thrive quietly in the background of a healthy human adult but, during inflammation, bloom to overwhelm the microbiota and may or may not cause host damage with their Yop-like T3SS. Whether or not this model is accurate will require future research. Even in absence of validating this model, the comparative genomics presented here provides the first report of

sequenced *P. fluorescens*-species complex strains from human patients, revealing the existence of a novel group of clinical *P. fluorescens*-species strains that contain genome-wide signatures of human-host adaptation and contain a virulence factor associated with human-host pathogenesis.

Figure VII.1: Core, accessory and pan genomes.

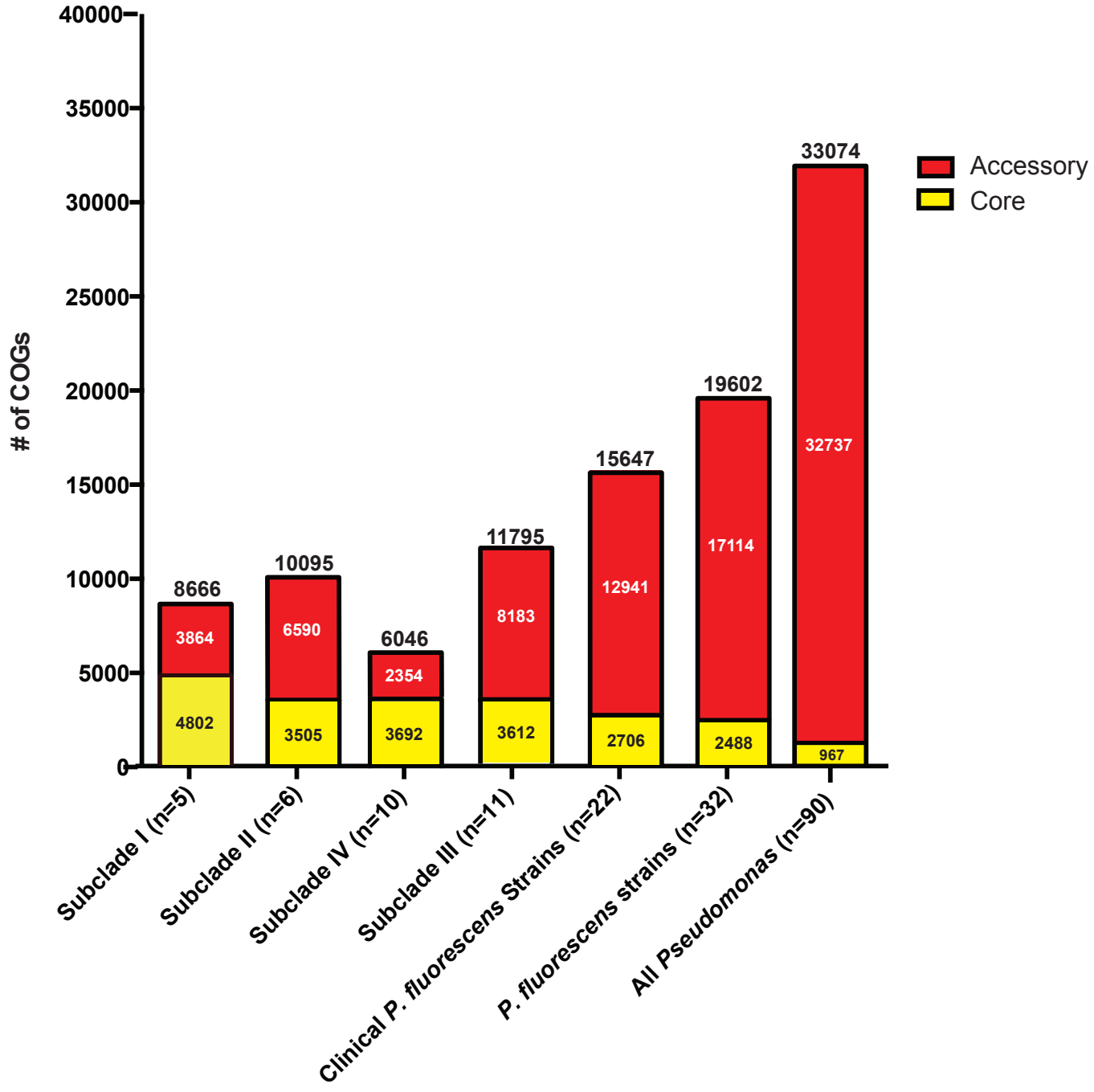


Figure VII.1: **Core, accessory and pan genomes.**
The number of COGS in each genome compartment. The softcore corresponds to 95% of considered genomes and contains the core genome. The pan genome (number above the bars) contains the accessory and core compartments. The pan genome and its compartments was calculated using the COGtriangle clustering algorithm (Kristense, 2010) in GET_HOMOLOGUES (Contreras-Moreira, 2013).

Chapter VII

References

1. Alonso A, Rojo F, Martinez JL. 1999. Environmental and clinical isolates of Pseudomonas aeruginosa show pathogenic and biodegradative properties irrespective of their origin. Environmental microbiology 1:421-430.
2. McArthur JV, Tuckfield RC. 2000. Spatial patterns in antibiotic resistance among stream bacteria: effects of industrial pollution. Appl Environ Microbiol 66:3722-3726.
3. Rasmussen LD, Sorensen SJ. 1998. The effect of longterm exposure to mercury on the bacterial community in marine sediment. Curr Microbiol 36:291-297.
4. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q, Miyada CG, Lory S. 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences of the United States of America 100:8484-8489.
5. Rolsma SL, Frank DW. 2014. In vitro assays to monitor the activity of Pseudomonas aeruginosa Type III secreted proteins. Methods Mol Biol 1149:171-184.
6. Sato H, Frank DW. 2014. Intoxication of host cells by the T3SS phospholipase ExoU: PI(4,5)P2-associated, cytoskeletal collapse and late phase membrane blebbing. PLoS One 9:e103127.
7. Arthur JC, Jobin C. 2013. The complex interplay between inflammation, the microbiota and colorectal cancer. Gut Microbes 4:253-258.
8. Arthur JC, Gharaibeh RZ, Muhlbauer M, Perez-Chanona E, Uronis JM, McCafferty J, Fodor AA, Jobin C. 2014. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun 5:4724.
9. Couturier-Maillard A, Secher T, Rehman A, Normand S, De Arcangelis A, Haesler R, Huot L, Grandjean T, Bressenot A, Delanoye-Crespin A, Gaillot O, Schreiber S, Lemoine

Y, Ryffel B, Hot D, Nunez G, Chen G, Rosenstiel P, Chamaillard M. 2013. NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. J Clin Invest 123:700-711.

10. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD. 2013. The gut microbiome modulates colon tumorigenesis. MBio 4:e00692-00613.
11. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, Timmis KN, Schreiber S. 2004. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut 53:685-693.
12. Darfeuille-Michaud A, Neut C, Barnich N, Lederman E, Di Martino P, Desreumaux P, Gambiez L, Joly B, Cortot A, Colombel JF. 1998. Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. Gastroenterology 115:1405-1413.
13. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79.
14. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO. 2010. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes 1:138-147.
15. Flierl MA, Rittirsch D, Huber-Lang M, Sarma JV, Ward PA. 2008. Catecholamines-crafty weapons in the inflammatory arsenal of immune/inflammatory cells or opening pandora's box? Mol Med 14:195-204.
16. Sperandio V, Torres AG, Jarvis B, Nataro JP, Kaper JB. 2003. Bacteria-host communication: the language of hormones. Proceedings of the National Academy of Sciences of the United States of America 100:8951-8956.

17. Alverdy J, Holbrook C, Rocha F, Seiden L, Wu RL, Musch M, Chang E, Ohman D, Suh S. 2000. Gut-derived sepsis occurs when the right pathogen with the right virulence genes meets the right host: evidence for in vivo virulence expression in Pseudomonas aeruginosa. Ann Surg 232:480-489.

18. Belay T, Sonnenfeld G. 2002. Differential effects of catecholamines on in vitro growth of pathogenic bacteria. Life Sci 71:447-456.

19. Jackson RW, Johnson LJ, Clarke SR, Arnold DL. 2011. Bacterial pathogen evolution: breaking news. Trends Genet 27:32-40.

20. Saile E, Koehler TM. 2006. Bacillus anthracis multiplication, persistence, and genetic exchange in the rhizosphere of grass plants. Appl Environ Microbiol 72:3168-3174.

21. Simoes RR, Poirel L, Da Costa PM, Nordmann P. 2010. Seagulls and beaches as reservoirs for multidrug-resistant Escherichia coli. Emerg Infect Dis 16:110-112.

22. Moliner C, Raoult D, Fournier PE. 2009. Evidence of horizontal gene transfer between amoeba and bacteria. Clin Microbiol Infect 15 Suppl 2:178-180.

23. Waterfield NR, Wren BW, ffrench-Constant RH. 2004. Invertebrates as a source of emerging human pathogens. Nature Reviews Microbiology 2:833-841.

24. Greub G, Raoult D. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413-433.

25. Scully LR, Bidochka MJ. 2006. Developing insect models for the study of current and emerging human pathogens. FEMS Microbiol Lett 263:1-9.

26. Berg G, Eberl L, Hartmann A. 2005. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. Environmental microbiology 7:1673-1685.

27. Neumann G, Römheld V. 2007. The Release of Root Exudates as Affected by the Plant Physiological Status. 20072634:23-72.

28. Troxler J, Azelvandre P, Zala M, Defago G, Haas D. 1997. Conjugative Transfer of Chromosomal Genes between Fluorescent Pseudomonads in the Rhizosphere of Wheat. Appl Environ Microbiol 63:213-219.

29. Alonso A, Sanchez P, Martinez JL. 2000. Stenotrophomonas maltophilia D457R contains a cluster of genes from gram-positive bacteria involved in antibiotic and heavy metal resistance. Antimicrob Agents Chemother 44:1778-1782.

30. Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep 2:376-381.

31. Lugtenberg BJJ, Dekkers LC. 1999. What makes Pseudomonas bacteria rhizosphere competent? Environmental microbiology 1:9-13.

32. Rahme LG, Stevens EJ, Wolfort SF, Shao J, Tompkins RG, Ausubel FM. 1995. Common virulence factors for bacterial pathogenicity in plants and animals. Science 268:1899-1902.

33. Cao H, Baldini RL, Rahme LG. 2001. Common mechanisms for pathogens of plants and animals. Annual review of phytopathology 39:259-284.

34. van Loon LC, Bakker PA, Pieterse CM. 1998. Systemic resistance induced by rhizosphere bacteria. Annual review of phytopathology 36:453-483.

35. Scales BS, Dickson RP, LiPuma JJ, Huffnagle GB. 2014. Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. Clin Microbiol Rev 27:927-948.

36. Laursen JB, Nielsen J. 2004. Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. Chem Rev 104:1663-1686.

37. Weller DM, Landa BB, Mavrodi OV, Schroeder KL, De La Fuente L, Blouin Bankhead S, Allende Molar R, Bonsall RF, Mavrodi DV, Thomashow LS. 2007. Role of 2,4-diacetylphloroglucinol-producing fluorescent Pseudomonas spp. in the defense of plant roots. Plant Biol (Stuttg) 9:4-20.

38. Mavrodi DV, Blankenfeldt W, Thomashow LS. 2006. Phenazine compounds in fluorescent Pseudomonas spp. biosynthesis and regulation. Annu Rev Phytopathol 44:417-445.

39. Ramette A, Moenne-Loccoz Y, Defago G. 2003. Prevalence of fluorescent pseudomonads producing antifungal phloroglucinols and/or hydrogen cyanide in soils naturally suppressive or conducive to tobacco black root rot. FEMS Microbiol Ecol 44:35-43.

40. Keel C, Weller DM, Natsch A, Defago G, Cook RJ, Thomashow LS. 1996. Conservation of the 2,4-diacetylphloroglucinol biosynthesis locus among fluorescent Pseudomonas strains from diverse geographic locations. Appl Environ Microbiol 62:552-563.

41. Takahashi M, Matsumoto S, Iwasaki S, Yahara I. 1990. Molecular basis for determining the sensitivity of eucaryotes to the antimitotic drug rhizoxin. Mol Gen Genet 222:169-175.

42. Gross H, Loper JE. 2009. Genomics of secondary metabolite production by Pseudomonas spp. Nat Prod Rep 26:1408-1446.

43. Tsuruo T, Oh-hara T, Iida H, Tsukagoshi S, Sato Z, Matsuda I, Iwasaki S, Okuda S, Shimizu F, Sasagawa K, et al. 1986. Rhizoxin, a macrocyclic lactone antibiotic, as a new antitumor agent against human and murine tumor cells and their vincristine-resistant sublines. Cancer Res 46:381-385.

44. Schnider-Keel U, Seematter A, Maurhofer M, Blumer C, Duffy B, Gigot-Bonnefoy C, Reimmann C, Notz R, Defago G, Haas D, Keel C. 2000. Autoinduction of 2,4-diacetylphloroglucinol biosynthesis in the biocontrol agent Pseudomonas fluorescens CHA0 and repression by the bacterial metabolites salicylate and pyoluteorin. J Bacteriol 182:1215-1225.

45. Sarniguet A, Kraus J, Henkels MD, Muehlchen AM, Loper JE. 1995. The sigma factor sigma s affects antibiotic production and biological control activity of Pseudomonas fluorescens Pf-5. Proc Natl Acad Sci U S A 92:12255-12259.

46. Mousa WK, Raizada MN. 2015. Biodiversity of genes encoding anti-microbial traits within plant associated microbes. Front Plant Sci 6:231.

47. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences 101:3160-3165.

48. Aujoulat F, Roger F, Bourdier A, Lotthe A, Lamy B, Marchandin H, Jumas-Bilak E. 2012. From environment to man: genome evolution and adaptation of human opportunistic bacterial pathogens. Genes (Basel) 3:191-232.

49. Georgiades K, Raoult D. 2010. Defining pathogenic bacterial species in the genomic era. Front Microbiol 1:151.

50. Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for alpha-proteobacterial genome evolution. Proceedings of the National Academy of Sciences of the United States of America 101:9722-9727.

51. Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol 15:321-326.

52. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D. 2001. Mechanisms of evolution in Rickettsia conorii and R. prowazekii. Science 293:2093-2098.

53. Pallen MJ, Wren BW. 2007. Bacterial pathogenomics. Nature 449:835-842.

54. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13.

55. Georgiades K, Merhej V, El Karkouri K, Raoult D, Pontarotti P. 2011. Gene gain and loss events in Rickettsia and Orientia species. Biol Direct 6:6.

56. Ozer EA, Allen JP, Hauser AR. 2014. Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. BMC Genomics 15:737.

57. Clarke DJ. 2014. The genetic basis of the symbiosis between Photorhabdus and its invertebrate hosts. Adv Appl Microbiol 88:1-29.

58. Ehlers R-U. 1996. Current and Future Use of Nematodes in Biocontrol: Practice and Commercial Aspects with Regard to Regulatory Policy Issues. Biocontrol Science and Technology 6:303-316.

59. Kaya HK, Gaugler R. 1993. Entomopathogenic Nematodes. Annual Review of Entomology 38:181-206.

60. Costa SC, Girard PA, Brehelin M, Zumbihl R. 2009. The emerging human pathogen Photorhabdus asymbiotica is a facultative intracellular bacterium and induces apoptosis of macrophage-like cells. Infect Immun 77:1022-1030.

61. Gerrard J, Waterfield N, Vohra R, ffrench-Constant R. 2004. Human infection with Photorhabdus asymbiotica: an emerging bacterial pathogen. Microbes Infect 6:229-237.

62. Wilkinson P, Waterfield NR, Crossman L, Corton C, Sanchez-Contreras M, Vlisidou I, Barron A, Bignell A, Clark L, Ormond D, Mayho M, Bason N, Smith F, Simmonds M, Churcher C, Harris D, Thompson NR, Quail M, Parkhill J, Ffrench-Constant RH. 2009. Comparative genomics of the emerging human pathogen Photorhabdus asymbiotica with the insect pathogen Photorhabdus luminescens. BMC Genomics 10:302.

63. Brugirard-Ricaud K, Duchaud E, Givaudan A, Girard PA, Kunst F, Boemare N, Brehelin M, Zumbihl R. 2005. Site-specific antiphagocytic function of the Photorhabdus luminescens type III secretion system during insect colonization. Cell Microbiol 7:363-371.

64. Hauser AR. 2009. The type III secretion system of Pseudomonas aeruginosa: infection by injection. Nat Rev Microbiol 7:654-665.

65. Malangoni MA, Crafton R, Mocek FC. 1994. Pneumonia in the surgical intensive care unit: factors determining successful outcome. Am J Surg 167:250-255.

66. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Hoiby N, Molin S. 2012. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Microbiol 10:841-851.

67. Matz C, Moreno AM, Alhede M, Manefield M, Hauser AR, Givskov M, Kjelleberg S. 2008. Pseudomonas aeruginosa uses type III secretion system to kill biofilm-associated amoebae. Isme J 2:843-852.

68. Casadevall A, Pirofski LA. 2003. The damage-response framework of microbial pathogenesis. Nat Rev Microbiol 1:17-24.

69. Pirofski L-a, Casadevall A. 2008. The Damage-Response Framework of Microbial Pathogenesis and Infectious Diseases. 635:135-146.