**Comparative genomics of clinical isolates of *Pseudomonas fluorescens,* including the discovery of a novel disease-associated subclade.**

by

Brittan Starr Scales

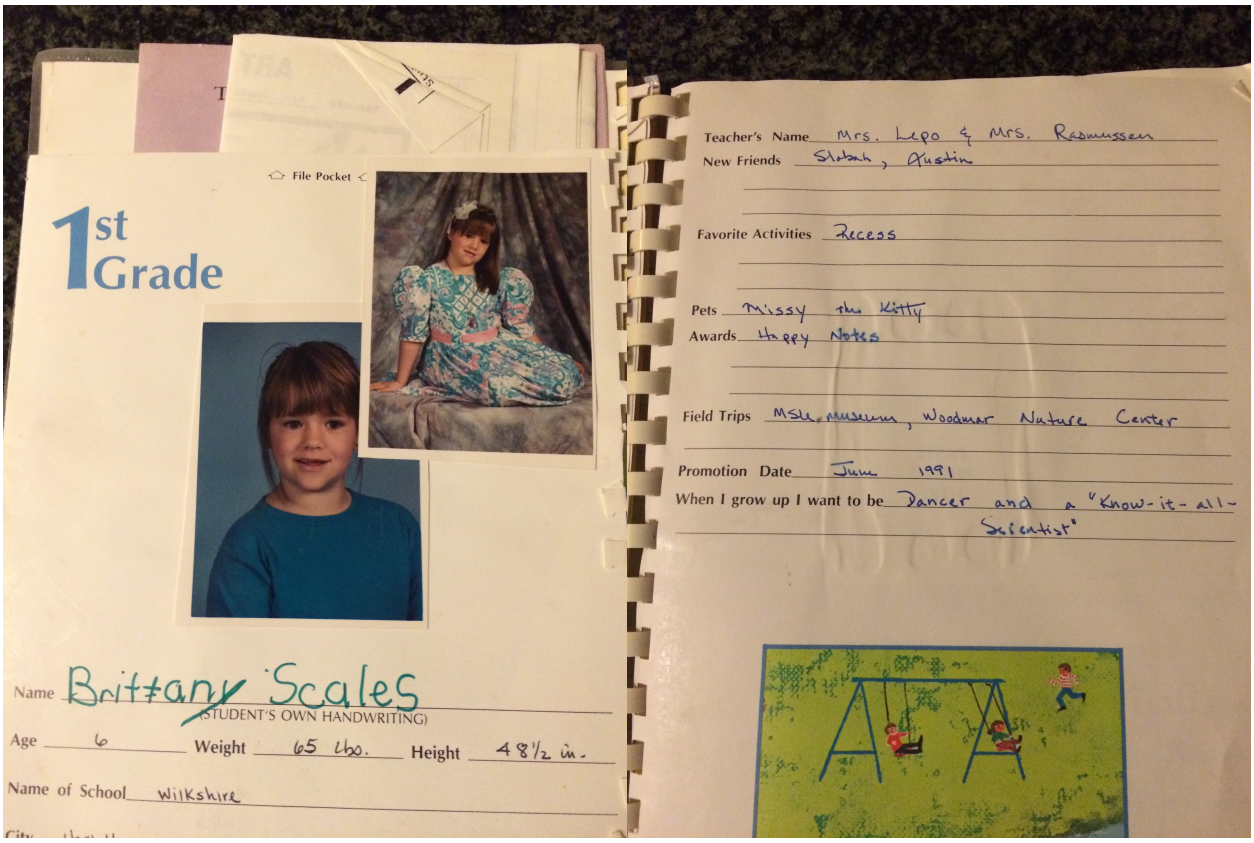
A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in The University of Michigan
2015

Doctoral Committee:

Professor Gary B. Huffnagle, Chair
Professor John Joseph LiPuma
Professor Bethany B. Moore
Professor Michele S. Swanson

## DEDICATION

Twenty- five years ago a little girl told her mother that when she grew up she wanted to be a ‘Know-it-all-Scientist’. This dissertation is dedicated to the single mother who didn’t tell her little girl that such a goal was impossible, but instead, did everything in her power to make it happen.



## TABLE OF CONTENTS


DEDICATION ii

LIST OF FIGURES v

ABSTRACT x

CHAPTER

**I. Microbiology, Genomics and Clinical Significance of the *Pseudomonas fluorescens* Species Complex, an Unappreciated Colonizer of Humans 1**

I. Figures 27

I. References 38

**II. Materials and Methods 62**

II. References 70

**III. Multi-locus Sequence Analysis (MLSA) of newly sequenced clinical *Pseudomonas spp.* isolates 73**

III. Figures 82

III. References 86

III. Appendix 88

**IV. Comparative Genomics of *Pseudomonas fluorescens* Subclade III Strains from Human Lungs 91**

IV. Figures 118

IV. References 129

IV. Appendix 151

**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 164**

V. Figures 184

V. References 203

V. Appendix 211

**VI. Chronic Inflammation in the Lower Airways Favors the Growth of *P. fluorescens* Species Complex Bacteria** **233**

VI. Figures 246

VI. References 255

VI. Appendix 260

**VII. Finale** **269**

VII. Figure 287

VII. References 288

## LIST OF FIGURES

**Chapter I**


| | | |
|---|---|---|
| Figure I.1 | Reported *P. fluorescens* infections | 27 |
| Figure I.2 | *P. fluorescens* isolates cultured over an eleven-year period by the University of Michigan Hospital Microbiology lab. | 28 |
| Figure I.3 | The functional range and environmental niches of the *Pseudomonas* genus, highlighting the broad distribution of the *P. fluorescens* species complex. | 29 |
| Figure I.4 | Species diversity within the *P. fluorescens* species complex. | 30 |
| Figure I.5 | Scanning electron micrograph of *P. fluorescens* (courtesy of Science Source). | 31 |
| Figure I.6 | Characteristics of *P. fluorescens* complex bacteria. | 32 |
| Figure I.7 | Phylogenetic tree of 38 *Pseudomonas* type strain, based on concatenated nine gene MLST analysis. | 33 |
| Figure I.8 | Summary information of fully sequenced bacterial strains from the *P. fluorescens* species complex. | 34 |
| Figure I.9 | Phylogenetic tree of 38 *Pseudomonas* type strains, based on the V3-V5 region sequence of the 16S rRNA gene. | 35 |
| Figure I.10 | Scanning electron micrographs of *P. fluorescens* biofilms. | 36 |
| Figure I.11 | Type-III secretion systems in *P. fluorescens*. | 37 |


**Chapter III**


| | | |
|---|---|---|
| Figure III.1 | Phylogenetic tree of 60 sequenced clinical *Pseudomonas spp*. strains with six-gene MLSA scheme. | 82 |
| Figure III.2 | Phylogenetic tree with eight-gene MLSA scheme to characterize *P. fluorescens* subclade III. | 83 |

Figure III.3 Phylogenetic tree with seven-gene MLSA scheme to characterize *P. fluorescens* subclade IV. 84

Figure III.4 Phylogenetic tree of 56 sequenced clinical *Pseudomonas spp*. strains with seven-gene MLSA scheme. 85

Appendix Figure III.1 Housekeeping genes used for screening and classifying the newly sequenced *Pseudomonas spp*. strains. 88

Appendix Figure III.2 Housekeeping genes in the newly sequenced *Pseudomonas spp.* strains. 89

**Chapter IV**

Figure IV.1 Date, location and source of *P. fluorescens* strains. 118

Figure IV.2 Assembly statistics of newly sequenced *P. fluorescens* strains. 119

Figure IV.3 Genomic features of newly sequenced *P. fluorescens* strains. 120

Figure IV.4 Phenogram based on average nucleotide identity (ANI) between subclade III strains. 121

Figure IV.5 Secondary metabolite genes and gene clusters in subclade III strains. 122

Figure IV.6 Pan, accessory and core analysis of the eleven subclade III strains. 123

Figure IV.7 The czc gene cluster in environmental and clinical subclade III strains. 124

Figure IV.8 G+C content across the genomes of newly sequenced subclade III strains. 125

Figure IV.9 Alignment of the pfiT amino acid sequences from subclade III strains. 126

Figure IV.10 Genes involved in Type-II, III, IV, VI and WCI secretion system in subclade III strains. 127

Figure IV.11 Neighbor-joining tree based on the amino acid sequence of czcA homologues. 128

Appendix Figure IV.1 Coverage map of AU2989 assembled scaffolds and contigs. 151

Appendix Figure IV.2 The nucleotide sequence of the 16S rRNA sequences was 152

used to query the NCBI nucleotide collection (nr/nt).

Appendix Figure IV.3 Average nucleotide identity (ANI). 153

Appendix Figure IV.4 Secondary metabolite genes and reference organisms for blast. 154

Appendix Figure IV.5 Blast results of selected GC Islands. 157

Appendix Figure IV.6 Full annotation of genes involved in secretion systems in subclade III strains. 159

Appendix Figure IV.7 Alignment of FimA homologues. 161

Appendix Figure IV.8 Alignment of VasH homologues. 162

Appendix Figure IV.9 Phylogenetic tree of VasH homologues. 163

**Chapter V**

Figure V.1 Date, location and source *P. fluorescens* strains. 184

Figure V.2 Growth properties of clinical and reference *Pseudomonas* strains. 185

Figure V.3 Assembly statistics of clinical strains of *P. fluorescens* sequenced in this study. 186

Figure V.4 Subclade IV isolation across the United States 188

Figure V.5 Genome features of *P. fluorescens* strains in this study. 189

Figure V.6 Genome size of newly sequenced and reference *Pseudomonas* strains. 190

Figure V.7 % G+C content of newly sequenced and reference *Pseudomonas* strains. 191

Figure V.8a G+C content across the genomes of newly sequenced subclade IV strains. 192

Figure V.8b G+C content across the genomes of subclade I and II strains. 193

Figure V.9 Pan, accessory and core analysis of the ten subclade IV strains. 194

Figure V.10 PCA built on shared gene clusters between Pseudomonas strains. 195

Figure V.11 Alignment of the pfiT amino acid sequences from subclade I and II *P. fluorescens* strains. 196

Figure V.12 Alignment of I2 nucleotide sequence subclade I, II and III *P. fluorescens* strains. 197

Figure V.13 Number of gene homologues annotated as part of the Ysc/Yop- like type-III secretion system. 198

Figure V.14 Amplification of the ExoU protein with *P. fluorescens* subclade IV-specific primers. 199

Figure V.15 ExoU and SpcU protein phylogeny in *Pseudomonas* strains. 200

Figure V.16 Alignment of SpcU homologues. 201

Figure V.17 Model of the type III secretion system in *P. fluorescens* subclade IV strains. 202

Appendix Figure V.1 Average Nucleotide Identity (ANI). 211

Appendix Figure V.2 Genome size of newly sequenced and reference *Pseudomonas* strains. 212

Appendix Figure V.3 % G+C content of newly sequenced and reference *Pseudomonas* strains. 213

Appendix Figure V.4 Nucleotide blast of selected GC Islands. 214

Appendix Figure V.5 Annotated type III secretion genes in *P. fluorescens* subclades IV and *P. aeruginosa* strains. 225

Appendix Figure V.6 Statistics of annotated type III secretion genes. 228

Appendix Figure V.7 Development of P. fluorescens subclade IV-specific ExoU primers. 229

Appendix Figure V.8 Blastp with ExoU Subclade IV protein. 230

Appendix Figure V.9 Blastp with SpcU Subclade IV protein. 232

**Chapter VI**

Figure VI.1 Inflamed mouse lung microbiome from in-house colony. 247

Figure VI.2 Inflamed mouse lung microbiome from Jackson labs. 248

Figure VI.3 Inflamed mouse tongue microbiome from in-house colony. 249

Figure VI.4 Inflamed mouse tongue microbiome from Jackson labs. 250

Figure VI.5 Relative abundance of mouse lung- associated *Pseudomonas* OTUs in five treatment subsets. 251

Figure VI.6 Relative abundance of human lung- associated *Pseudomonas* OTUs in three patient subsets. 252

Figure VI.7 *Pseudomonas* abundance in COPD lung explants. 253

Figure VI.8 V3 - V5 16S rRNA phylogenetic tree. 254

Appendix Figure VI.1 Blast results of *Pseudomonas* OTUs. 260

Appendix Figure VI.2 Blast alignment of *Pseudomonas* OTUs against reference *Pseudomonas* genomes. 264

Appendix Figure VI.3 Blast and alignment of the two mouse lung-associated *Streptococcus* OTUs. 267

**Chapter VII**

Figure VII.1 Core, accessory and pan genomes. 287

## ABSTRACT

Comparative genomics of clinical isolates of *Pseudomonas fluorescens,* including the discovery of a novel disease-associated subclade.

by

Brittan Starr Scales

Chair: Gary Huffnagle

Abstract title: Comparative genomics of clinical isolates of *Pseudomonas fluorescens,* including the discovery of a novel disease-associated subclade.

Taxonomically, there are over 52 *Pseudomonas* species that group within the *Pseudomonas fluorescens* species-complex. They are gram-negative bacteria (gammaproteobacteria) that have the ability to flourish in many different environmental niches due their metabolic adaptability. Prior to beginning the project described in this dissertation, we had identified (using culture-independent analysis) that the lung microbiome of patients with severe chronic obstructive pulmonary disease (COPD) contained abundant levels of *P. fluorescens*. This finding was highly unexpected because *P. fluorescens* colonization had not been previously reported in chronic respiratory disease (human or veterinary). Furthermore, there was no taxonomic or genome information on clinical *P. fluorescens* strains in the published literature or NCBI database. This raised the question of whether there were differences between *P. fluorescens* strains isolated from clinical vs. environmental sources. To begin to address this question, I sequenced a collection of *Pseudomonas spp.* isolates from individuals with chronic lung disease, which included 22 clinical *P. fluorescens* strains. Twelve of the 22 could be grouped within previously defined subclades I, II and III of the *P. fluorescens* species-complex. However, the other 10 strains were distinct and could be clustered as a phylogenetically unique fourth subclade, which lacked any representative genomes (clinical or environmental) in the

NCBI database. I performed additional comparative genomic analyses on all four subclades in order to identify genomic attributes that were associated with growth in humans (vs. the environment). Subclade III clinical isolates were almost indistinguishable from environmental isolates; however, the clinical isolates contained additional homologues of genes involved in metal toxicity resistance, an attribute that has been reported for chronic antibiotic exposure, such as would occur in the cystic fibrosis lung. In sharp contrast, subclade IV strains displayed marked reductions in genome size, gene diversity and GC content, as well as containing all of the elements for a Ysc family type III secretion system that was extremely similar to that found in *P. aeruginosa.* Furthermore, our subclade IV strains, collected from geographic locations across the U.S., had a very high level of shared nucleotide identity and a small accessory genome, suggesting that they may have evolved from living in multi-trophic environments to a life in a much narrower niche, perhaps human airways. We also identified that some colonies of inbred mice contained indigenous *P. fluorescens* in their lung microbiome (subclade III) and chronic inflammation was associated with an outgrowth of *P. fluorescens*. Expanding my analysis to human respiratory disease, I found an increase in *P. fluorescens* (largely subclade III) in lung samples from multiple diseases (COPD, IPF and lung transplant) but not healthy subjects. Altogether, these studies reveal that *P. fluorescens* is an under-appreciated colonizer of humans, particularly in the context of pulmonary inflammation, and lay the groundwork for future studies delineating the contribution and molecular mechanisms of this host-microbe interaction.