Comparative Genomics of *Pseudomonas aeruginosa* Populations

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Matthew Phillip Moore

September 2017

Abstract

Pseudomonas aeruginosa causes a wide range of infections, is often associated with antimicrobial resistance and is the primary cause of chronic lung infection in cystic fibrosis (CF) and the overall morbidity and mortality associated with the disease. As *P. aeruginosa* is an environmental bacterium that opportunistically infects CF patients most infecting lineages are distinct. The determination of common adaptive routes during infection is further complicated by within-lineage heterogeneity, multi-lineage infections and the emergence of transmissible strains. In order to better understand the genetic basis of varying pathogenicity, a diverse dataset of 1,407 *P. aeruginosa* genomes from the environment, non-CF bronchiectasis and from CF samples was analysed. Detailed analysis of mutations potentially associated with the CF lung environment was conducted by comparison of the genomes from CF and environmental isolates and by including a geographically and historically diverse panel of a transmissible lineage, the Liverpool Epidemic Strain (LES). Sequencing of isolates from non-CF bronchiectasis showed for the first time the commonality in adaptive routes in non-CF chronic lung diseases and the utility of genome sequencing to infer within and between host diversity and population structure. Many antibiotic resistance genes (ARGs) in all samples were observed to be adaptive and a sample of genomes from hospital isolates from Thailand was used to assess international multi-drug resistance lineages with mobile genetic element associated ARGs. This study has shown, by analysis of diverse datasets, how genome sequencing can reveal the genetic basis of phenotypic heterogeneity and subsequent varying patient outcomes with this diverse infection.

Acknowledgements

I would like to thank the Cystic Fibrosis Trust for funding this project and all of those who contributed, helped and advised who are too numerous to list in full so particularly:

My supervisors Craig and Jo for their support throughout this project, your experience and patience make it impossible to imagine having had anyone else guiding this effort. I would also like to thank everyone in Winstanley group over the years particularly David for all his help, Emily and Christina. I'd especially like to thank all of you that helped me through my brief stint in the laboratory, all with good humour, sometimes amazement and plenty of celebratory lunches and coffees. I would also like to thank everyone at IGH that made it a great place to be. Roger Levesque and his team, Julie, Luca, Jean-Guilliame and Irena for hosting me in their lab and making me extremely welcome in Quebec and the bonus of Rogers soundtracks. Thank you to the Microbiology society for funding the visit.

Finally I would like to thank family and friends who have supported me throughout and even before I began with the PhD. I would particularly like to thank my partner Jenny without whom this thesis would be poorly formatted and probably not exist in the first place. Your constant reassurance and thoughtfulness got me through and your enthusiasm inspired and reminded me why I wanted to do this in the first place.

Declaration

This thesis is the result of my own work. The material presented here has not been presented and is not being present, either wholly or in part for any other degree or qualification.

This project was supervised by Professor Craig Winstanley and Dr Jo Fothergill at the University of Liverpool.

Matthew Moore

Publications

Hilliam Y*, **Moore MP***, Lamont IL, Bilton D, Haworth CS, Foweraker J, et al. Pseudomonas aeruginosa adaptation and diversification in the non-cystic fibrosis bronchiectasis lung. Eur Respir J. 2017;49(4). Available from: http://www.ncbi.nlm.nih.gov/pubmed/28446558

*authors contributed equally

Abbreviations

ANI	Average nucleotide identity
ARG	Antibiotic resistance gene
BHL	N-butyrylhomoserine lactone
CGR	Centre for Genomic Research (Liverpool)
Cup	Chaperone/usher/pathway
GATK	The Genome Analysis Toolkit
GWAS	Genome wide association studies
HHQ	2-heptyl-4-quinolone
IBIS	Sequencing facility (Quebec, Canada)
IQS	Integrated quorum sensing
LES	Liverpool Epidemic Strain
Mbp	Million base pairs
MES	Manchester Epidemic Strain
Mid	Midlands Epidemic Strain
MLST	Multi locus sequence typing
MRSA	Methicillin resistance Staphylococcus aureus
MST	Multi-spacer typing
MVLA	Multiple locus variable-number tandem repeat analysis
NRPS	Non-ribosomal peptide synthases
OdDHL	N-(3-oxododecanoyl)-homoserine lactone
OHHL	N-(3-oxohexanoyl)-L-homoserine lactone
ORF	Open reading frame
PAI	P. aeruginosa autoinducer

PCR-RFLP PCR-restriction fragment length polymorphism

PFGE Pulse field gel electrophoresis

QS Quorum Sensing

RNA-seq RNA sequencing

RND Resistance nodule division

sRNA Small RNA

SSH Suppression subtractive hybridization

TSS Type n secretion system

WGS Whole genome Sequencing

SNP Single nucleotide Polymorphism

Indel Insertion or deletion

Table of Contents

Chapter One, Introduction

Chapter Two, General Methods

Chapter Three, Adaptation to the cystic fibrosis lung by loss of function

Chapter Four, Whole Genome Sequencing of Historically and Geographically Diverse *Pseudomonas aeruginosa* Liverpool Epidemic Strain Isolates

Chapter Seven, General Discussion

Chapter One

1. Introduction

1.1. Microbial Genomics

Chain termination 'Sanger' sequencing was conceived by Fred Sanger in 1977. His method became the first commercially available DNA sequencing technology and was later replaced by high-throughput 'second generation' sequencers, facilitating a methodological revolution in molecular biology. The first free living organism to have its entire genome sequenced was that of bacterium *Haemophilus influenzae* in 1995[1]. The whole genome took 13 months to complete and represented a landmark technical achievement. Microbiological practices such as genotyping, molecular epidemiological surveillance, identifying antibiotic resistance, taxonomy have been revolutionised by available genome sequences[2]. For these breakthroughs to be possible sequence data analysis has developed and shifted with the change in platforms, type and amount of output. Algorithms have been developed for *de novo* genome assembly and scaffolding[3,4], read mapping approaches[5], variant calling[6] and gene annotation[7] from which most microbial bioinformatics workflows are built. Importantly, the majority of informatics software, code libraries and databases, from which the majority of advances have been possible are open source[8]. New insights have also been obtained through the application of modeling, statistical approaches and machine learning to genomics data[9–12].

1.1.1. Taxonomy and Genotyping

Fundamental microbiological practices are being updated by the availability of sequencing. The taxonomic placement of microbes before whole genome sequencing (WGS) primarily relied upon the 'gold standard' DNA-DNA hybridization and more recently 16S rRNA similarity. Now phylogenetic approximation using some or all of the genome enable reconstruction of detailed taxonomies[13]. Additionally average nucleotide identity (ANI) has replaced the former 'gold standard'[14]. As a result of these advances numerous microbial species have been reclassified[15]; ANI was, for example, successfully used to identify new species of *Vibrio* and *Burkholderia*[16]. Microbiological typing methods, such as serotyping have been largely replaced by molecular typing methods, which can be sequenced based or non-sequence based. Non-sequence based molecular genotyping methods include PCR-restriction fragment length polymorphism (PCR-RFLP), pulse field gel electrophoresis (PFGE), Microarrays and multiple locus variable-number tandem repeat analysis (MVLA). Multiple genotyping methods have since been developed based on the DNA sequence data. These include multi-spacer typing (MST)[17] and multi locus sequence typing (MLST)[18]. MLST was developed for the identification of pathogenic clones[19]. MLST is based on variation in seven housekeeping genes thought to be non-recombining, not under selection and present (due to their essential functions) in all members of the species. MLST allows for surveillance of pathogenic strains even when more discriminatory WGS data is available as the sequence type can quickly be determined from the whole genome and provide useful information from samples. As such it has been used, for example, in the

typing and surveillance of *Staphylococcus areus*[20], extended spectrum β-lactamase producing *Escherichia coli*[21], and *Streptococcus pneumoniae*[22].

1.1.2. Bacterial Core, Accessory and Pan-genomes

The pangenome is the entire set of genes that comprise a species; wherever there is variation in the gene families present between members of a species their sum is the pangenome. The pangenome can be context defined such as strain or genus' pangenome, but unless otherwise stated it will be the species pangenome that is being referred to. Gene families that are core are present in all members of the species and may be summarised as the core genome. Gene families that are accessory, range from two isolates to n-1 isolates, where n is the total number of isolate genomes. Finally, unique gene families are found only in one. The amount of intraspecies, pangenome, diversity varies (table 1.1). It can be clinically useful to determine the pangenome and the field of vaccinology now includes 'reverse vaccinology'; the process of identifying from whole genome sequencing gene products as vaccine candidates. The term was first coined when the method was developed to identify new surface exposed antigen vaccine candidates from the genome of *Neisseria meningitidis* serogroup B[23]. Reverse vaccinology studies have also been conducted for *Mycobacterium tuberculosis, Ricksettsia prowazekii, E. coli, Bacillus anthracis, Chlamydia pneumonia, Leptospira interrogans, Porphyromonas gingivalis, Streptococcus pneumoniae* and *Streptococcus pyogenes*[24–26].

Table 1.1. A small selection of pangenome statistics for five organisms

Species	Number of genomes	Number and percent of genes that are core	Number and percent of genes that are accessory	Description/comment
Salmonella enterica	206	1,717 (16%)	8,550 (83%)	Pathogen associated with soil, water, rumen
Streptococcus pyogenes	49	986 (37%)	1,644 (62%)	Obligate human pathogen
Chlamydia trachomatis	67	153 (15%)	821 (84%)	Obligate intracellular pathogen
Acinetobacter baumannii	40	1,997 (28%)	5,042 (71%)	Soil associated opportunistic pathogen
Bacillus cereus	30	3,245 (27%)	8,666 (72%)	Associated with rumen, mammalian and insect intestinal tracts

1.1.3. Determining the genetic basis of bacterial phenotypes

Identification of the virulence factors and antibiotic resistance gene repertoire of the organism is crucial to understanding microbial pathogens. Genomics is enabling researchers to make inroads into the prediction of complex microbial phenotypes relevant to the major features of infectious disease causation: infectivity, attachment, invasiveness, persistence, resistance and virulence.

Genomic identification of virulence factors and antibiotic resistance has primarily involved the curation of databases of genes and alleles with known phenotypic outcomes and associations with or putative virulence or resistance phenotype. Antibiotic resistance genes (ARGs) for example were identified in *Acinetobacter baumannii* genomes by alignment to the Resfinder[27] or ARDB[28] databases. Similarly it was possible to identify 829 virulence genes in *E. coli*

strain PA45 by alignment to the virulence factor database, vfDB[29]. The method Mykrobe predictor was developed for prediction of antibiotic resistance in *S. aureus* and *Mycobacterium tuberculosis*[11]. Diagnostic tests to determine antibiotic resistance and susceptibility in *M. tuberculosis* can take weeks due its slow growth rate. In *S. aureus,* Mykrobe predictor achieves >99% sensitivity and specificity and in *M. tuberculosis* 98.5% specificity and 82.6% sensitivity[11]. The adoption of rapid sequencing for identification of virulence and resistance genes from genomic and metagenomic sequences from clinical samples will transform how clinical microbiology is conducted[13] and provide more information than ever before for medical microbiological research. However these methods ultimately depend on established an understanding of genes and their association with phenotype. In many cases complex gene regulation and pleiotropy result in gene function being unclear.

Methods are emerging that have the potential to predict bacteria phenotypes with a complex genetic basis. Methods such as genome wide association studies (GWAS) developed for human genome sequencing projects[30] take a top down approach towards eliminating bias and potentially revealing associations between genes or alleles and phenotypes. GWAS in bacteria is typically based on sequence presence or absence to identify whether there is an association with the observed effect (phenotype). Phenotypic classifications such as resistance can be associated with responsible genes or islands[31–34]. Alternatively SNP data may associate alleles and particular mutations with phenotypic outcomes[10,35–37]. A major development has been the method proposed by Sheppard *et al.*,[31] of a kmer based approach from which the association of both SNPs and sequence

presence/absence can be assessed allowing for core and accessory genome diversity to be considered. The future of GWAS is promising and the issues of recombination, linkage disequilibrium, strength of selection and population stratification that may confound it are being addressed[9]. Further, there is the potential for the development of machine learning algorithms capable of inferring non-linear associations, such as the random forest approach. The random forest algorithm is an ensemble machine learning approach capable of determining the predictive capacity of sequences and variants for subsequent causality testing[38].

1.1.4. Genomic epidemiology

Molecular epidemiology can be counted among the disciples that have most benefitted from the availability of WGS data. Comparison of whole genomes allows discrimination based on SNPs, providing greater discriminatory power than molecular epidemiological approaches and certainly more than molecular and microbiological typing approaches formerly used to inform epidemiology[39,40].

The data provided by whole genome sequencing, so far primarily from 'second generation' sequencers, has allowed outbreaks to be mapped in greater detail than ever before[41]. Further the portable sequencing platform, Oxford Nanopores MinIon expands the potential of genomic epidemiology and broader investigation of pathogens in the healthcare setting worldwide. In 2012 genomic epidemiology successfully identified 6 cases of methicillin resistance *Staphylococcus aureus* (MRSA) to a common source, where corresponding PFGE

profiles were unable to distinguish[42]. Genomic epidemiology has revealed the chain of transmission in *Acinetobacter baumanni*[43,44], *Klebsiella pneumoniae*[45] and MRSA[46,47]. In non-communicable infections such as burns infections with *Pseudomonas aeruginosa,* WGS was used to determine the source, the showers in a burns unit[45]. Infection dynamics may also be resolved from sequencing, such as in *Clostridium difficile*[48]. It was determined that infection could not be explained as transmission between symptomatic carriers but that either asymptomatic carriers have introduced *C. difficille* to the hospital environment or that there's a broader environmental contamination with *C. dificile*[48]. In 2016 the average nucleotide identity (ANI) method was used to resolve *Elizabethkingia* strains during an outbreak. *El. anophelis*, *El. meningoseptica* and *El. miricola* were identified as distinct species and a core genome MLST scheme developed[49]. As such *El. meningoseptica* were reassigned as *El. anophelis*[49]. A community associated outbreak from 2015 to 2016 occurred in Wisconsin, Illinois and Michigan, US, of the recently defined *El. anophelis*[50]. Genomic epidemiology determined that the outbreak was a single strain that had been growing and adapting in an environmental reservoir for ~1 year before the first infection occurred[50]. Observations of temporal accumulations of mutations allowed the authors to further hypothesise that patients were either exposed to a low diversity population or that transmission bottlenecks were occurring, further demonstrating the potential epidemiological insights possible with WGS data[50]. These retrospective epidemiological studies have clearly demonstrated the power of WGS to track outbreaks, determine the chain of transmission or common source and further yield clinically useful information about the bacteria involved. The potential for real-time genomic epidemiology has also

been demonstrated[12,51]. In 2015 real time nanopore sequencing in a UK hospital demonstrated there was no common source of *Salmonella* enteriditis. Outbreaks on different, unlinked, wards harboured genetically distinct isolates[52].

1.1.5. Bacterial genome evolution

Bacteria are asexually reproducing, unicellular organisms. Though they clonally reproduce there is vast observable genetic diversity across the bacterial kingdom and discernible species, strains and clonal groups (though the boundaries are sometimes harder to define than is the case with animals for instance). There are a number of mechanisms contributing to extant diversity of bacteria such as mutation. Mutations occur as a result of replicative errors or physical damage to bacterial DNA. Once present the relative fitness conferred by a mutation will determine whether it will be selected against (purifying selection) selected for (positive selection) or be neutral to selection. Mutations in coding sequences may also be classified independent of their impact on fitness as synonymous, where the base or codon change does not result in encoding of a different amino acid or non-synonymous where the result of a mutation is a change in the encoded amino acid. Further, non-synonymous mutations can be categorised as nonsense mutations if they're disruptive to the encoded protein by introducing a frame-shift, premature stop codon or loss of a start or stop codon. A further mechanism for genetic change in bacteria is recombination by horizontal gene transfer (HGT). Though prokaryotes lack sexual reproduction, a uniquely eukaryotic trait, they do engage in homologous and non- homologous recombination, where the resulting exchange of genetic material that may be adaptive, neutral or disadvantageous as with mutational

genetic change. Genetic diversity may also result from gene duplications. If a copy of a gene loses function or gains a function as a result then this is categorised as a paralog, whereas orthologs are genes that have diverged along with organisms (have common ancestry) and finally, xenologs are those acquired through HGT.

The relative fitness of genetic variation in a population (fitness coefficient) is determined by numerous molecular evolutionary processes. Fitness is best conceptualised as a landscape wherein there are fitness peaks and valleys; there are numerous ways to be on a peak, including suboptimal ones[53]. Genetic drift is any random change of allele frequency in a population, resulting from chance rather than selective pressures. The effective population size (*Ne*) best predicts what impact genetic drift may have on the selection coefficient. In a population with small *Ne* alleles may drift to fixation even when there's strong selective pressure acting, for example to maintain heterogeneity, because chance changes in allele frequency become less unlikely. However, the majority of mutations accumulated are deleterious. A large accumulation of mutations is the leading contributor of a high genetic load, suboptimal fitness as a result of genotypes present. Another phenomena determining potential moves on a fitness landscape is a form of drift: 'genetic hitchhiking'. A neutral change in allele frequency can be led by the physical proximity (linkage disequilibrium) of an allele to another that's under selection[54]. Mutation rates can also be adaptive, especially when populations are placed under stress[55]. Though the genetic load is increased, a heightened mutation rate can be crucial to adapting rapidly to a high stress environment such as within a host. Transient hypermutable states

can result from DNA mismatch repair deficiencies, recombination or polymerase fidelity[56]. The process of clonal interference occurs when multiple advantageous mutations accumulate in a population and when recombination is absent result in competition and potential loss of beneficial mutations[57]. Occurring on their own these mutations would not be lost. Homologous recombination can mitigate phenomena such as clonal interference and combine advantageous mutations in a lineage. How much homologous recombination drives adaptation in bacterial lineages varies[58], for example, from low in *Bacillus cereus* populations[59,60] to high in *Neisseria gonorrhoaeae* populations[61].

1.2. Pseudomonas aeruginosa

Discovered in 1882 *Pseudomonas aeruginosa* is a Gram negative, motile, rod-shaped bacterium belonging to the genus *Pseudomonas.* The pseudomonads can be found in soil or marine environments or associated with plants and animals. *P. aeruginosa* is of particular interest due to its status as a major opportunistic pathogen of humans and other animals. In humans *P. aeruginosa* is involved in a variety of infections including burn and wound infections[62], infection of cancer patients or immunosuppressed patients[63], eye infections[64] and chronic lung infections of CF[65–68] and non-CF bronchiectasis patients[69]. Worldwide *P. aeruginosa* is a major nosocomial pathogen and in 2017 was placed in the highest priority category (critical) by the World Health Organisation (WHO) of bacteria for which new antibiotics are urgently needed. The versatility of the species has been attributed to its ability to thrive such a variety of environmental and host niches[70].

1.2.1. The *Pseudomonas aeruginosa* genome

In 2000 *P. aeruginosa* PAO1, an isolate that was originally sampled from a wound and that became a well-studied reference laboratory strain, was whole genome sequenced[71]. At the time it also represented the largest bacterial genome to be sequenced at 6.3 million base pairs (Mbp). The genome sequence revealed a GC content of 66.6% and 5,570 open reading frames (ORFs), of which approximately 8% were predicted to encode transcriptional regulators and two-component systems[71]. Subsequently, PA14, a similarly well-studied *P. aeruginosa* strain was sequenced revealing a larger genome of 6.5Mbp and a larger proportion of genes (~12%) involved in regulation. *P. aeruginosa* has a

larger genome and number of genes than other common nosocomial pathogens such as *E. coli* and *S. areus* that carry an average of 4,721 and 3,118 genes respectively[72].

In order to assess the *P. aeruginosa* population structure using whole genome sequencing, Stewart *et al*,[73] approximated the phylogeny of 55 *P. aeruginosa* genomes and determined that there were three distinct groups containing PAO1, PA14 and taxonomic outlier PA7 respectively[74]. Group 1 was by far the largest and contains strains PAO1, LESB58, DK2 and PAK, while group 2 contains PA14. Group 3 was the smallest containing only PA7 and one other genome at the time[73]. With a greater number of genomes sequenced a minority do not fit clearly into these three phylogenetic groupings, though not all are PA7-like phylogenetic outliers[75]. It remains that the majority place within the three major phylogenetic groups and that they are distributed with the largest number in group 1, 2 and then a minority in 3[75,76].

A study of 200 *P. aeruginosa* genomes revealed 16,000 non redundant genes in the pangenome, 15% of which comprised the core genome[77]. On average 40% of a *P. aeruginosa* genome is constant, with the remaining 60% of genes per genome being variable to different extents[77]. Genes associated with metabolism and respiration are core in the *P. aeruginosa* pangenome, but also frequently present are antibiotic resistance-associated genes including multidrug efflux pumps, secretion systems, ABC transporters and two-component systems[77]. This confirmed previous conclusions from smaller pangenome studies of *P. aeruginosa*[78,79]. In the accessory genome type I and IV secretion systems

associated with the export of toxins, proteases, lipases and other effectors are present, in addition to a wide range of antibiotic resistance genes, including genes that encode products that confer resistance to vancomycin[77]. The complete gene repertoire confirms the view that *P. aeruginosa* carries a high level of versatility in a range of niches, intrinsic resistance to common anti-pseudomonad therapies, complex gene regulation further enhancing its versatility, a variety of secretion systems and an ability to acquire genes, including antibiotic resistance genes.

In 2017 the Pseudomonas Database listed 1,636 *P. aeruginosa* and 2,560 pseudomonad genomes sequenced overall[80].

1.2.2. Complex gene regulation of *Pseudomonas aeruginosa*

In addition to the diverse pangenome of *P. aeruginosa,* gene regulation expands the potential adaptability on a transcriptional, translational and protein level. *P. aeruginosa* has a large number of genes overall for a bacterium and among the highest coding density of regulatory genes[71,81]; comprised mainly of sigma factors, two component systems and small RNA (sRNA). Individual genes may be regulated by multiple regulators, further complicating interconnected networks and the potential for interpreting the outcome of mutations accumulated in these genes and inference of their adaptive potential[82].

Sigma factors are involved in a range of core processes in *P. aeruginosa*, from RpoD involved in house-keeping and RpoS involved in the stationary phase and stress[83], to AlgU involved in mucoidy. The *mucA* gene encodes an anti-sigma

factor that regulates alginate biosynthesis and may drastically alter the lifestyle of a *P. aeruginosa* and its capabilities as a pathogen. Some major *P. aeruginosa* sigma factors are summarised in table 1.2.

Table 1.2. Major *P. aeruginosa* sigma factors and anti-sigma factors

Sigma factor	Gene name	Description	Anti sigma factor
RpoH	*rpoH*	Heat shock response	Na
RpoS	*rpoS*	Biofilm, quorum sensing and	Na
		antibiotic resistance	Na
RpoD	*rpoD*	Housekeeping	Na
RNA pol	*rpoN*	Quorum sensing	Na
AlgU	*algU*	Alingate biosynthesis, mucoidy	MucA
FliA	*fliA*	Flagella biosynthesis	FlgM
PvdS	*pvdS*	Pyoverdine	FpvR
FiuL	*fiuL*	Ferrichrome production	FiuR

Two component response systems are the primary means by which bacteria sense their environment, allowing bacteria when stimulated to transcriptionally regulate gene expression in response to the environment. In *P. aeruginosa* PAO1 there are 64 response regulators and 63 histidine kinases[84] with which to sense its environment and respond with transcriptional activation. Histidine kinases have variable detection and input domains and as such can respond to a range of environmental stimuli such as nutrient, phosphate and oxygen availability, antimicrobial peptides and osmolarity[85,86].

Small RNAs regulate gene expression at the level of translation, binding mRNA and resulting in repression or activation[87]. Transcriptional regulation by sRNAs may affect genes associated with iron and cell envelope homeostasis, virulence

and metabolism[88]. In *P. aeruginosa* 44 sRNAs have been discovered[89] and in PAO1 RNA sequencing (RNA-seq) determined a further 500 intergenic sRNAs[90].

GacAS serves as an example of the further complexity and versatility made possible by the regulatory networks of *P. aeruginosa*. GacAS is a two component system associated with secondary metabolites, biofilm formation and quorum sensing[91,92], which might themselves re-model gene expression. GacS contains a histidine kinase domain, response regulator and phosphotransfer domain[93,94] and is controlled by phosphorylation by GacA or LadS[94] and prevention of phosphorylation by RetS[95,96]. In turn phosphorylated GacA activates transcription of small RNAs RsmZ and RsmY. Activated RsmZ and RsmY block the negative regulator of RsmA that can then positively regulate a variety of features including antibiotic resistance, quorum sensing, type 3 secretion, other transcription factors and iron homeostasis[97–100].

1.2.3. Cell to cell Signaling

Cell to cell signaling, or Quorum Sensing (QS) was discovered in *Vibrio fischeri* which utilises the *luxI-luxR* system. In the *luxI-luxR* system *luxI* encodes the acylhomoserine lactone (AHL) signal N-(3-oxohexanoyl)-L-homoserine lactone (OHHL) and *luxR* encodes an AHL dependent transcription factor[101–103]. In *P. aeruginosa* LuxR homolog, LasR was discovered and subsequently shown to be a key regulator of *lasB*, *lasA, toxA* and *aprA* in addition to being a potential global regulator of virulence factors[104]. The LuxI homolog, LasI[105] (the *P. aeruginosa* autoinducer (PAI)) was determined as N-(3-oxododecanoyl)-homoserine lactone (OdDHL)[105].

A second *P. aeruginosa* autoinducer, N-butyrylhomoserine lactone (BHL) was discovered[106] and its cognate receptor RhlR subsequently discovered[107]. RhlR is encoded by the *rhlABR* gene cluster with BHL encoding *rhlI* downstream[107]. Autoinducers OdHL and BHL activate and form complexes with LasR and RhlR and these complexes bind with *las-rhl* boxes in the promoters of target genes to start their expression. The response to the respective complexes varies across genes affected that overall comprise ~10% of the *P. aeruginosa* genome.

The third QS system of *P. aeruginosa* was discovered to be structurally distinct from the AHL systems of *las* and *rhl*[108]. The 2-heptyl-3-hydroxy-4-quinolone molecule represents the first 4-quinolone bacterial signaling molecule determined, encoded by *pqsABCD, phnAB* and *pqsH*[109]. The genes *pqsABCD* encode PQS biosynthesis, producing a precursor 2-heptyl-4-quinolone (HHQ)[109]. Finally, PQS biosynthesis HHQ is converted to PQS by PqsH, which is controlled by LasR[110]. The cognate receptor of PQS signaling is PqsR, a LysR-type transcriptional regulator that binds to the promoter region of the *pqsABCDE* operon[109,111].

A fourth system, the integrated quorum sensing (IQS) system is comprised of non-ribosomal peptide synthases (NRPS's). The *iqs* system is encoded by the gene cluster *ambBCDE* and has structurally been established as 2-(2-hydroxyphenyl)-thiazole-4-carbaledhyde[112]. The *iqs* system is capable of compensating for a non-functional *las* system and when disrupted a decrease in

PQS and BHL systems is observed as well as a decrease in virulence factors expressed.

The role of the respective signaling systems is complex, involving transcriptional regulation of many genes in *P. aeruginosa*. The manner in which the four systems are also interconnected further complicates comprehension of these circuits[112]. The *P. aeruginosa* QS system however is known to be involved in regulation of virulence[113], phosphate deprivation stress[114,115], iron acquisition[116], oxygen deprivation[117], nutrient deprivation[117] and response to host factors[118]. Quorum sensing is an integral component of *P. aeruginosa*'s ability to infect host plants, protists and animals[119].

1.2.4. Metabolic capabilities

P. aeruginosa is metabolically versatile owing much of its ability to thrive in diverse niches and its subsequent success as an opportunistic pathogen in the ability to make use of varied carbon and nitrogen sources[78]. *P. aeruginosa* is able to catabolise amino acids, polyamines, carboxylic acids, short-chain fatty acids and complex xenobiotics such as halogenated n-alkanes and aromatic compounds even as sole carbon sources[120,121]. It is also possible for *P. aeruginosa* to grow aerobically or anaerobically making use of nitrate as an electron acceptor through dentrification or by use of arginine and pyruvate through fermentation[78]. Whole genome metabolic reconstruction, modeling and transposon mutagenesis of PA14 and PAO1 further shows the interconnectivity between metabolic pathways and virulence factors[122]. The metabolic model for

PAO1 (iPae1146) reveals 1,146 genes, 1,493 reactions and 1,284 metabolites while that of PA14 (iPae1129) reveals 1,129 genes, 1,495 reactions and 1,286 metabolites[122]. In these models 112 and 108 metabolic genes were linked to virulence functions in PAO1 and PA14 respectively[122].

Pseudomonads produce a number of secondary metabolites[123] involved in a variety of functions. *P. aeruginosa* produces secondary metabolites with antibiotic functions[124], hydrogen cyanide[117], those involved with iron acquisition[125] and phenazines[126]. The *P. aeruginosa* phenazine pyocyanin has been implicated in virulence in a variety of models and during CF lung infection and ear infection, otitis externa[127,128]. Pyocyanin deficiency has been associated with suppressed virulence in mouse models of acute pneumonia and burns[111,129], *Drosophila melanogaster*[130], plant infection of *Arabidopsis thaliana*[91,131] and nematode *Caenorhabditis elegans*[132].

1.2.5. Uptake of Iron

The abundance of the element iron and its ability to exist in two states Fe^{2+} and Fe^{3+} make it useful in numerous oxidation and reduction reactions in bacteria[133]. Within a host setting, such as during lung infections the availability of iron is reduced compared with the environmental niches from which *P. aeruginosa* is thought to have been acquired typically. As such the ability of a bacterial population ability to scavenge iron is crucial to its survival and pathogenicity. Broadly *P. aeruginosa* may import iron by production of chelating molecules (siderophores) that scavenge Fe^{3+}, the import of xenosiderophores

not produced by *P. aeruginosa*, import of host heme molecules and extracellular reduction of Fe^{3+} and subsequent import of Fe^{2+}.

P. aeruginosa produces two main chelating siderophores low affinity pyochelin[82,134] and high affinity pyoverdine, that *P. aeruginosa* is observed to switch to in low-iron environments[135]. Pyoverdine production is characteristic of the fluorescent *Pseudomonas* species, *P. aeruginosa, P. flourescens, P. putida,* and *P. syringae.* Types of pyoverdine in *P. aeruginosa* (I, II and III) and corresponding receptors have been identified, from which typing schemes have been developed[136,137]. Pyoverdine is comprised of a chromophore and a peptide chain that may be variable generating the different pyoverdine types[125,136]. *P. aeruginosa* possesses TonB-dependent receptors and transducers such as FpvA which allow the import of ferrisiderophores[138]. In contrast, the Has and Phu systems in *P. aeruginosa* enable heme import from hemoproteins[139].

1.2.6. Secretion systems

P. aeruginosa secretion systems are essential to exploiting a variety of environments. There are numerous bacterial secretion systems (T1SS-T3SS and T5SS-T6SS) that conduct extracellular release of proteins and direct injection into host eukaryotic cells. *P. aeruginosa* lacks T4SS's but is otherwise considered a model organism for their study[140].

T1SS are simple trimeric structures[141] that secrete molecules such as AprA and HasAp, involved in collagen degradation and iron acquisition respectively and AprX which is of unknown function[141,142]. T2SS, unlike T1SS, are also able to

allow folded proteins to pass[142]. T2SS secrete multiple proteins such as lipases LipA, LipC, PlcH and PhlN[143], exotoxin A (ToxA) and elastase (LasB)[144]. While T1SS and T2SS function by extracellular secretion of proteins T3SS inject proteins directly into cells. The 'needle' structure that constitutes the T3SS is facilitates effector secretion of exotoxins ExoS, ExoT, ExoU and ExoY[145]. *P. aeruginosa* isolates usually contain either ExoS or ExoU paired with ExoT[146].

1.2.7. Biofilm Lifestyle

Biofilms are extracellular polymeric substances within which bacterial communities can live[147,148], largely composed of exopolysaccarides and extracellular DNA[149]. Biofilms develop through a process of adherence, microcolony formation, maturation and then dispersal, dependent on quorum sensing[150,151]. *P. aeruginosa* biofilms utilise the Pel system, encoded by six genes in the *pelABCDEF* operon or Psl systems, encoded by 15 genes in the *pslABCDEFGHIJKL*[152] system. Psl biofilms are rich in galactose and mannose[153]. A major benefit of the biofilm rather than planktonic lifestyle is protection from immune defences and antibiotics[154–157]. This is not only due to slowed progression through the biofilm's extracellular matrix but also due to the varying metabolic activities within the biofilm, such as dormancy[158,159]. Alginate overproduction in *P. aeruginosa* can be observed in comparison with smooth colonies[160]; mucoid colonies may overexpress alginate as much as 6x that of smooth colonies[161]. Alginate can provide protection from phagocytosis and antibiotics and is a free radical scavenger[162,163]. Though not crucial for biofilm formation[164,165] it represents a major component of it[164]. Disruptive mutations

in the gene *mucA* that encodes anti-sigma factor gene MucA results in alginate overexpression due to lack of negative regulation of alginate biosynthesis[166].

1.2.8. Antibiotic Resistance

Core to the *P. aeruginosa* species are intrinsic antibiotic resistance mechanisms such as efflux pumps and less directly, the biofilm lifestyle[167]. MexAB-OprM, MexCD-OprJ, MexEF-OprN and MexXY resistance nodule division (RND) efflux pumps present in *P. aerguinosa* can confer resistance to aminoglycosides, novobicin, β-lactams, tetracycline, chloramphenicol, cephalosporins, fluoroquinolones, macrolides, tigecycline and trimethoprim[168–171]. These major RND efflux pumps in *P. aeruginosa* are encoded by operons in the core genome[78]. In addition to intrinsic resistance its possible for *P. aeruginosa* to acquire resistance to almost all antibiotics available[172,173].

1.3. Cystic Fibrosis

Cystic Fibrosis (CF) results from disruption in the cystic fibrosis transmembrane conductance regulator (CFTR). There are multiple different autosomal recessive mutations that result in the disruption of CFTR, which normally functions to regulate ion transport across epithelial surfaces[174,175]. The most common causative mutation of CF is deltaF508 CFTR[176,177], a mutation that results in an absence of CFTR in epithelial cells due to post-translation degradation[176,177]. CFTR disruption results in a dehydration and thickening of mucosa in the lungs and digestive tract and susceptibility to lung infections[178] resulting in pulmonary exacerbations[176], the need for aggressive antibiotic treatment regimes[177], progressive lung disease[176,179] and premature death[179].

Increasingly the picture of CF lung infection is one of complex, polymicrobial infections[180], comprised of a mixture of aerobic and anaerobic infections[181]. Adult sputum samples, assessed by 16S rRNA analysis, reveal core genera present in the CF lung, *Streptoccocus, Pseudomonas, Prevotella Veillonella, Neisseria, Porphorymonas* and *Canontella* in abundance[182]. Organisms such as the *Streptoccous* milleri group spp, *Trichosporon* spp, nontuberculosis Mycobacterium and rhinovirus are emerging as pathogens, while *Stenotrophomonous maltophilia, Achromobacter* spp, *Ralstonia* spp, *Burkholderia gladioli* and *Streptococcus pneumoniae,* widely believed to be pathogenic, are having that status questioned[182]. It remains however that *Haemophilus influenzae, Staphylococcus aureus* and above all, *P. aeruginosa* predominate in adult CF lung infections[68,183,184].

1.3.1. *Pseudomonas aeruginosa* Lung Infections in Cystic Fibrosis

P. aeruginosa is the leading cause of overall morbidity and mortality in adult CF patients as a result of progressive lung damage and chronic infection[185]. Patients as young as three years old can show signs of previous lung infection[186], typically followed by intermittent infections during adolescence and finally a chronic *P. aeruginosa* infection[187]. The risk of morbidity is increased once the lungs are colonised with *P. aeruginosa*[188] and chronic infection will result in progressive lung disease and ultimately premature death[189]. By age 20 approximately 80% of adult CF patients are chronically infected with *P. aeruginosa*[190,191] and the 8 year risk of death is 2.6x higher than uninfected CF patients as a result[68].

Though in 1996 it was discovered that select lineages of *P. aeruginosa* were capable of cross infecting patients[192], it remains that the majority of patients harbour unique lineages. In Canada *P. aeruginosa* infected CF patients sampled presented a total of 157 genetic types (RAPD and PFGE typing) of *P. aeruginosa*, 123 of which were unique to individual patients[193]. This is the typical picture of *P. aeruginosa* acquisition though in some clinics transmissible strains can predominate. Where there is no cross infection or common source of infection a genotype can be observed at a higher frequency where it represents a common environmental lineage such as clone C[194]. There is no known bias towards a particular environmental 'type', or phylogenetic group and the unique lineages infecting largely reflect the general diversity of environmentally sampled *P. aeruginosa*[195].

Antibiotic resistance in *P. aeruginosa* CF lung infections is a major problem and as such samples are routinely tested for resistance and susceptibility. CF patients are treated with broad spectrum antibiotics, Ceftazidime[172], Meropenem[196], Piperacillin/Tazobactam and Ciprofloxacin[197]. CF patients are also treated with Tobramycin[198] that acts on Gram negatives and Colistin, a polymyxin[199].

1.3.2. Transmissible Strains

Most often CF patients acquire an infecting *P. aeruginosa* from the environment, patients harbour unique strains of which their source is unknown. Multiple patients have been identified to be infected by the clone C strain, but is also highly abundant in a variety of environments sampled, for example[200]. It was first suggested that cross infection may be occurring in the 1980's in Danish patients[201] and had previously been indicated that siblings share strains[202–204]. Additionally, there was a precedent for cross infection in CF clinics with the *Burkholderia cenocepacia* ET12 clone outbreaks[205]. Figure 1.1 shows a timeline of milestones into transmissible CF pathogen research.

In 1996 PFGE and flagellin typing of *P. aeruginosa* samples suspected due to antibiogram profiles between pediatric patients attending Alder Hey Children's Hospital in Liverpool provided the first molecular evidence for a transmissible strain of *P. aeruginosa* in CF, The Liverpool Epidemic Strain (LES)[192]. Further, the LES was discovered to be the single most common strain of *P. aeruginosa* infecting CF patients across 31 centres in England and Wales[206] and was later discovered in samples from multiple centres in Canada[207].

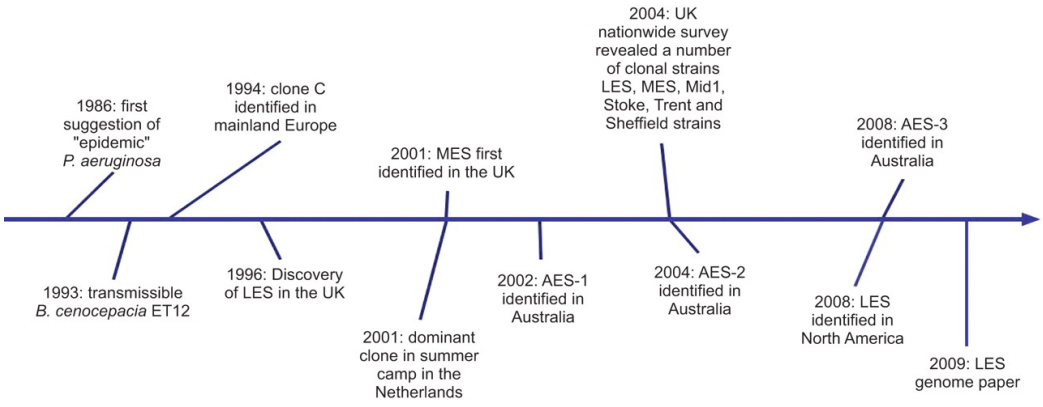


Figure 1.1. A timeline of events in research into transmissible strains of *P. aeruginosa* and the emergence of transmissible *B. cenocepacia* ET12. Adapted from Fothergill *et al.*[208]

Following the discovery of the transmissible *P. aeruginosa*, the LES and its association with poorer prognosis than other chronic *P. aeruginosa* infections[209] and antibiotic resistance[210], other independently evolved transmissible strains were discovered in various countries' CF clinics. In the UK the Manchester Epidemic Strain (MES) and Midlands (Mid1) stains were discovered [211], in Denmark the transmissible DK2, in Australia AES-1, AES-2 and AES-3[191,212,213] and in Canada, Canadian Epidemic Strain B[214] and the Prairie Epidemic Strain[215] and in the Netherlands, the Dutch Epidemic Strain[216]. In 2016 Mansfeld *et al.,* genomically and phenotypically assessed the Dutch Epidemic Strain (ST-406) within patient diversity, a strain that chronically infects 15% of CF patients in the Netherlands[216]. SNPs were identified in regulator genes *mexT, exsA, mexE, mexF* and the T3SS and T6SS, similar to adaptative genes in chronic CF strains[216]. Importantly, apart from the Danish and Manchester epidemic strains which share a common ancestry[216], genotyping and later core gene and genome phylogenetics confirm that the emergence of transmissibility has occurred independently in CF clinics internationally[73,75,76,212,214,217]. Though epidemic

strains carry genomic and phenotypic markers of chronic adaptation to the CF lung there is as yet no shared feature indicative of acquired transmissibility.

In response to the discovery of strains capable of cross infection diagnostics have been developed to identify known transmissible strains in the clinic. For the LES suppression subtractive hybridization (SSH) was used initially[206,218–220] followed by similar diagnostics for the MES, Mid1 and AES-1 strains[221–223]. In the Liverpool CF clinics the UK isolates are routinely typed by multiplex PCR tests designed to identify the MES, Mid1 and LES[224]. In response to knowledge of transmissible strains centres have instituted segregation policies to prevent cross infection between CF patients. In Liverpool segregation policies have resulted in a sharp decrease of LES positive samples in the adult and children's CF unit[65]. In Australia segregation of AES-1 positive patients yielded a decrease of patients observed to harbour the strain of one third[225,226].

1.3.3. Virulence during *Pseudomonas aeruginosa* infections

Many of the features that are characteristic of *P. aeruginosa* are potentially implicated in virulence during CF lung infections. These include the T3SS, flagella, LPS and secreted factors that result in inflammation and lung damage. These features and its regulatory plasticity determine its ability to survive within-host stresses such as iron, oxygen and nitrogen availability, heat and osmolarity. In particular, the QS systems of *P. aeruginosa* regulate up to 10% of the genome[94], including many virulence determinants with potential roles in the pathogenicity associated with CF[227].

Features of acute infections in *P. aeruginosa*, fimbriae[85,228], type IV pili[229] and LPS[230] are involved in establishment of CF lung infections. Fimbriae are a pili type that utilise the chaperone/usher/pathway (Cup)[228] of which three types have been identified in *P. aeruginosa* (Cup A, B, C)[85,228]. Cup attachment factors are essential for attachment to surfaces for subsequent biofilm formation[228]. Type IV pili are the primary adhesin responsible for epithelial cell adhesion[229] and there's further interaction observed *in vitro* with the glycosylated cell surface regions GM1 and GM2 of lung epithelial cells[228]. The two-component system *pilS/pilR* controls transcription of the type IV pili[231]. The LPS is comprised of three parts: Lipid A responsible for immune response[232], an oligosaccharide core and a variable O antigen[230]. *P. aeruginosa* produces two LPS types, A and B[233]. Similarly to type IV pili LPS attaches to GM1[234] but also CD14 and MD2 TLR4's[235].

Following establishment of infection in the CF lungs, *P. aeruginosa* has a range of virulence factors at its disposal, including the soluble lectins LecA and LecB[236,237], the T3SS and its exotoxins. The T3SS of *P. aeruginosa* enables direct injection of exotoxins ExoS, ExoU, ExoT and ExoY into host cells and further, T3SS exhibits cytotoxicity in absence of these exotoxins[145]. ExoS has been demonstrated to cause pulmonary damage in animal models[238]. In the presence of a host cell cofactor[239], ExoS causes cytoskeleton disruption in the target cell[240]. ExoS also interacts with immune cells directly by binding TLR4 and TLR2 domains[241] and indirectly resulting in T-cell activation[242] and apoptosis[243]. ExoS likely contributes to the damaging inflammation observed in the CF lung as a result of *P. aeruginosa* infection by activation of a pro-inflammatory cytokine

response[241]. Cytotoxin ExoS differs from ExoT, though similar in structure, in the pathways it targets and overall reduced potency[244]. ExoU is the most potent cyotoxin produced by *P. aeruginosa*[245] through the ability to disrupt host cell membranes. Soluble lectins LecA (PA-IL) and LecB (PA-IIL) are virulence factors secreted by *P. aeruginosa*[246]. Both have been observed to be causative of pulmonary lesions and an increase in in alveolar permeability[237].

Many features of *P. aeruginosa* during infection are dependent upon the QS system. *P. aeruginosa* QS-regulated virulence factors include secondary metabolites pyocyanin[247] and pyoverdine[125], elastases[215], proteases, exotoxins and lectins. Exotoxin A, the most toxic *P. aeruginosa* exotoxin to mice[248] and is capable of targeted host cell necrosis similar to diphtheria toxin[249–252]. The secreted phenazine pyocyanin is a blue pigment with a number of pathogenic properties[126,129]. Pyocyanin oxidises glutathione increasing oxidative and necrotic stress in epithelial cells[253]. Pyocyanin has also been observed to inactivate ATPase resulting in further chloride channel dysfunction in CF[126,129]. The immune response is also suppressed by pyocyanin by the induction of neutrophil apoptosis[254], resulting in increase IL-8 production[240]. Many isolates of the transmissible *P. aeruginosa* strain, the LES, are pyocyanin overproducers[255]. Elastase is a zinc metalloprotease with proteolytic activity[256], also known as LasB. LasB virulence occurs due to its pro-inflammatory effect by the increase of IL-8[144,257]. Additionally, elastase can cleave proteinase-activated receptor 2, reducing the host response[258] and targets cytokines such as interferon gamma and tumor necrosis factor[259]. Pyoverdine is a major iron chelating siderophore in *P. aeruginosa* that is also a virulence factor during CF

lung infections[240]. Pyoverdine has a major role in regulating its own secretion and also the secretion of exotoxin A and proteases[240,260–262]. *P. aeruginosa* proteases have a well understood virulence role during eye infections[263] including keratitis[264] and are increasingly understood to have a role in respiratory pathogenicity. The T1SS secretes protease virulence factor, alkaline protease[263] . Alkaline protease degrades fibrin and has been implicate in pneuomonia[265]. Protease IV virulence is due to its degradation of surfactant proteins A, B and D[266]. Rhamnolipids produced by *P. aeruginosa* have a role in motility similar to that of type IV and flagella[267] and are involed in biofilm formation[268]. Rhamnolipids are glycolipids secreted by *P. aeruginosa* and they impact virulence by their detergent qualities on lung surfactant facilitating phospholipases[269], *P. aeruginosa* invasion[270], inhibition of phagocytosis[271] and reduction of mucociliary clearance[272].

1.3.4. Adaptation to the cystic fibrosis lung environment

The study of evolving *P. aeruginosa* populations has been advanced by genome sequencing, confirming prior knowledge of genetic changes associated with chronicity and generating genetic datasets of genes and alleles inferred to be adaptive with known and unknown functional roles[35,77,273,274]. Diverse environmental lineages of *P. aeruginosa* infect CF patients resulting in considerable phenotypic background diversity confounding attempts to identify features causative of varying clinical outcomes. Further, background genetic variation confounds the determination of features associated with clinical outcomes such as persistence, antibiotic resistance, pathogenicity and capacity

immune evasion. Understanding how *P. aeruginosa* adapts to the lung has been derived from sequential sampling of patients, inference of convergence and experimental evolution[66]. Phenotypic observations over the course of infections have identified trends in *P. aeruginosa* CF lung populations. During an infection many non-mucoid lineages will adopt the alginate over producing mucoid phenotype and is associated particularly with chronic infections[166]. Further, the adoption of a biofilm lifestyle using the alginate, pel or psl exopolysaccharides[153,161,165], auxotrophy[275], loss of motility[276], hypermutability[277], antibiotic resistance[167,187] and a general reduction in the expression of *P. aeruginosa* virulence factors[66]. QS systems and key virulence feature, the T3SS have also been observed to become defective during infections[278].

Within the CF lung different environmental lineages may coexist; lineages that are distinct by genotyping, phylogenetic placement or genetic distance, representing lineages inferred to be distinct at the point of infection[279]. Further, genomic analysis demonstrates that lineages may diversify into sub-lineages with distinct evolutionary trajectories[280]. It has been shown for example that hypermutable sub-lineages that emerge can coexist with normal mutators[279,281]. Whether there are multiple environmental lineages present or not, the study of a lineage or sub-lineage is itself obscured by the emergence significant heterogeneity within populations[279,282]. Understanding this complexity also has practical implications. Studies of distinct colonies sampled from a single patient such as within those identified as differentially mucoid or non-mucoid there can be a range of susceptible and resistant isolates to antibiotics tested[283,284].

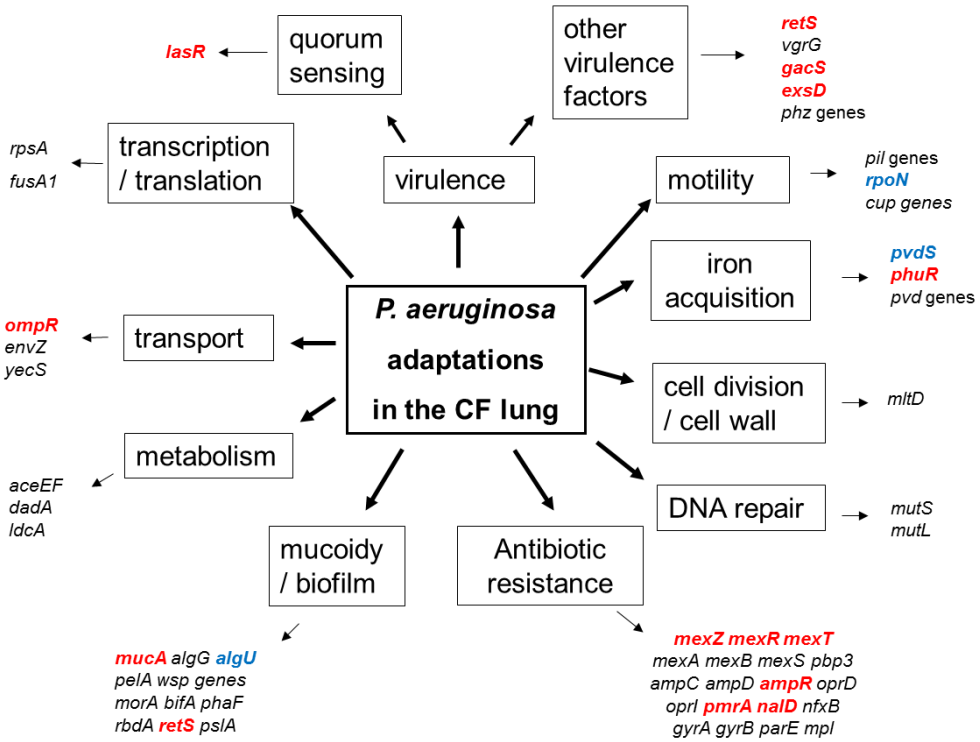


Figure 1.2. Genes that are commonly inferred to be adapted during CF lung infection. Genes encoding regulators are in red and those that encode sigma factors in blue. Adapted from Winstanley *et al.* [66]

Genome sequencing is increasingly revealing the genetic basis of evolutionary change during CF infections[273,274,281,285–287], figure 1.2 from Winstanley *et al.,* summarises genes across multiple studies inferred to be adaptive to the CF lung. Some genes are directly linked to phenotypic changes such as loss of function mutations in anti-sigma factor encoding *mucA* that normally negatively regulates alingate production[166]. The loss of MucA function results in alginate overproduction and has been observed to occur independently in different infections[166]. Additionally compensatory mutations have been observed to revert alginate overproduction[161,288]. Other genetic changes commonly inferred to be adaptive have known functions. Hypermutability can be achieved by

numerous potential nonsense mutations; disruptive mutations in either DNA mismatch repair genes *mutS, mutL, muM* and *uvrD* may result in a hypermutator lineage[289]. Many genes that can be inferred to be adaptive have regulatory roles including *lasR* that encodes the QS regulator LasR and genes *rpoN, mexT, retS, exsD* and *ampR*[66].

1.4. Aims of this study

This study uses >1,000 *P. aeruginosa* genomes to better understand genetic variation that predisposes lineages to establishing CF lung infections and the genetic basis for adaptation in the lung. CF and non-CF bronchiectasis associated isolate genomes, the transmissible LES and multidrug resistant isolates from Thailand were sequenced in addition to environmental genomes. The aims were to use comparative genomics in order to:

- Identify genes that may bias acquisition in CF
- Identify genes that are adaptive during CF lung infections
- Map genetic diversity accounting for phenotypic heterogeneity observed during *P. aeruginosa* infections
- Reveal the genetic bases of antibiotic resistance in *P. aeruginosa*

Chapter Two

2. General Methods

2.1. Whole genome sequencing

Sequencing was conducted at the Centre for Genomic Research (CGR) on the Illumina HiSeq platform and University of Liverpool and Laval University, Quebec (IBIS) on the Illumina MiSeq platform. Additional *P. aeruginosa* genomes were downloaded via the Pseudomonas.com database[290]. The 'wider *P. aeruginosa* population' for phylogenetic analyses refers to 423 genomes comprising the Kos *et al.* dataset[76] and reference genomes PAO1[71], PA14, PA7[74] and LESB58[291].

2.2. Adapter Trimming and Quality Filtering

Genomes sequenced at the CGR on the Illumina HiSeq platform where Illumina sequence adapters were also removed (adapter trimming) and sequencing reads were quality filtered according to Phred scores. Genomes sequenced at IBIS were adapter trimmed and quality filtered after download using Trim Galore! (-q 20 –o trimmed --paired)[292].

2.3. Genome Assembly and Draft Genome Quality Assessment

All genomes that were *de novo* assembled and scaffolded from sequencing reads were using the A5-miseq assembly pipeline[293]. When working with the LES genomes (Chapter 5), A5 assembled contigs were instead scaffolded to reference genome LESB58 with abacas[294]. Genome assembly quality was assessed using quast-3.1[295] and interpreted as a combination of contig length based metrics: number of contigs, N50 and longest contig length.

2.4. Read Mapping, Variant Calling and Annotation

Sequencing reads were aligned to a reference *P. aeruginosa* genomes PAO1 or LESB58 (read mapping) using bwa0.7.5a mem[296] with standard parameters. The reference genome (*.fasta*) was first indexed with bwa-0.7.5a index[296] and samtools-0.1.18-r580[297] faidx. The sequence dictionary was created with picard-tools-1.135[298] CreateSequenceDictionary. The resulting sequence alignment map (*.sam*) file from read mapping with bwa mem was converted to a binary alignment map (*.bam*) file using picard-tools SortSam. Duplicates were marked using picard-tools MarkDuplicates and a bam file index created with picard-tools BuildBamIndex. The Genome Analysis Toolkit-3.4 (GATK)[299] Realignor Target Creator was used to designate targets for indel realignment and indels are realigned with GATK IndelRealigner. Variants were called using GATK HaplotypeCaller (-ploidy 1 –emitRefConfidence GVCF) to produce a variant call (*.vcf*) file prepared for genotyping. The vcf file was genotyped using GATK GenotypeGVCFs and filtered with vcffilter[300], basic filtering (DP >9 and QUAL >10).

Variant annotation was performed by snpEff-4.1l[301] with the default parameters for gatk output (eff -gatk) to the reference genome databases PAO1 (uid57945) or LESB58 (uid59275). Additionally it was evaluated whether a gene has a larger deletion not detectable due to lack of sequencing reads and absence of genomic context in vcf files. First bam files were indexed with samtools index and the reads aligned to open reading frames as determined by snpEff database regions were assessed for coverage using samtools depth. The samtools depth

results were processed to get an approximate 'alignment' length from which larger deletions can be determined by subtracting from the total gene length (*awk '{if($3>0) total+=1}END{print total}*).

2.5. Multi Locus Sequence Typing *in silico*

In all chapters except Chapter 3, multi locus sequence type (MLST) was performed using a combination of mlst[302] and the CGE webserver[303]. If profiles were incomplete they were checked for the closest ST to their partial profile using the pubMLST database[304]. In Chapter 3 SRST2[305] was instead used for extracting MLST profiles from short reads.

2.6. Whole genome pairwise alignments

Mummer was used for whole genome pairwise alignments (nucmer –showsnps)[306]. Ambiguous bases were removed before counting variant sites and insertions and deletions collapsed.

2.7. Core Genome Extraction

For all *P. aeruginosa* core SNP trees, core genome polymorphic sites were extracted by Panseq[307]. Fragments (500bp) that are present in all genomes to at least 90% similarity were determined as core. For assessment of recombination core locus alleles were extracted from Panseq and aligned using clustalW[308], implemented in the biopython AlignIO module[309]. Core genome multiple alignments were concatenated into a core genome supermatrix. Maximum likelihood (ML) phylogenetic trees were approximated using IQtree with automatic model selection and 1000 ultrafast bootstraps or MEGA-6[310], using

the HKY85 substitution model and 1000 bootstraps. Gubbins was used for prediction of recombinant SNPs in the full core genome supermatrix, using a starting tree before final (clonal frame) tree approximation using RaxML[311,312], using the CAT+GTR substitution model. Gubbins was set to a maximum of five iterations is RAxML trees didn't converge after two iterations. Trees were visualised with the interactive tree of life software[313].

Chapter Three

Adaptation to the Cystic Fibrosis Lung by Loss of Function

3. Introduction

3.1. Adaptive loss of function

Bacterial populations can be expected to adapt to a changing environment or when transplanted into a new environment. The genetic basis generally of this adaptation comprises mutations that fine tune existing features[314–316], gain of genes by horizontal gene transfer[317] and by loss or disruption of genes. A mutation resulting in loss of function of that gene may be adaptive if the feature is detrimental or simply without benefit in a new or changing environment at some cost of production. Indirectly, loss of regulatory function can increase fitness due to subsequent remodeling of genes expressed[318]. Another phenomenon in which loss of functions may be advantageous is in the emergence of sub-populations utilising 'public goods' without the cost of contributing themselves. These 'cheats' thus gain a competitive advantage but are mechanistically self-limiting (as in quorum sensing dependent public goods due to the loss of sufficient signal density) or as a replacing 'cooperators' resulting in loss of the public good upon which the adaptation depends[319]. An example of adaptation as a result of loss of function mutations is the resulting loss of features that are detrimental upon the introduction of antibiotics. Loss of features that allow antibiotic entry, such as loss of function in the outer

membrane protein OprF in *E. coli*[320]. The adaptive landscape by loss of function is expansive may be rapidly utilised in response to change[321].

The large number of genes in a typical *P. aeruginosa*, genome coding density dedicated to regulatory genes[71,200] and observed metabolic versatility[322] suggest that *P. aeruginosa* is well placed to rapidly adapt to a new environment by loss of function. In addition to features developed in the environment that are beneficial during infection, the standing genetic variation of *P. aeruginosa* and its plasticity make it a formidable opportunistic pathogen[322].

3.1.1. Developing a representative panel of *P. aeruginosa*

In 2013 De Soyza *et al*. reported the development of a panel of 43 *P. aeruginosa* for collaborative study of diverse lineages[323]. The panel was assembled as part of a European Cooperation in Science and technology (COST) project and will be referred to as the COST panel. Isolates selected comprised commonly studied clones, isolates of transmissible strains, and sequential isolates from CF patients. Isolates that were diverse in virulence, serotype, genotype and geography were also included. Isolates IST27 and IST27N a mucoid isolate and spontaneously non-mucoid variant of IST27 respectively were included to assess the phenotypic switch in this important characteristic. Additional mucoid phenotype isolates NH57388A and 968333S were also included. Another important phenotype to pathogenesis is lipopolysaccharide (LPS) composition and as such isolates with defined LPS structures were included. Phenotyping of all of the 43 panel strains was carried out to also assess growth density at 15 hours, motility by swimming, swarming and twitching, pyocyanin production,

antibiotic resistance, auxotrophy, LPS composition, quorum sensing, biofilm formation and mucoidy. Phage typing was also performed. While there was no genotype or set of genotypes specifically associated to CF/non-CF reflecting the fact that they're primarily acquired from the, principle components analysis (PCA) showed a positive correlation between CF/non-CF status and *in vivo* virulence (*G. mellonella* model), biofilm formation and alginate production. PCA also confirmed the positive correlation of pyocyanin production and *in vivo* virulence[324].

3.1.2. The International *Pseudomonas* Genomics Consortium Dataset

Expanding on the development of a representative panel of *P. aeruginosa* and more recent studies showing the diversity even within hosts[282,325,326] the International *Pseudomonas aeruginosa* Genomics Consortium (led by Roger Levesque, Universite Laval, Quebec, Canada) developed the International *Pseudomonas* Genomics Consortium Database (IPCD)[75]. This database containing >1,000 newly sequenced *P. aeruginosa* genomes aims to integrate genomic data with phenotypic, virulence, clinical and environmental metadata. In addition to the representative panel this complex database attempts to unravel the genetic bases of diversity in *P. aeruginosa*. Provisionally these genomes were published and their phylogeny reconstructed[75] analysis of the core and pan genome is currently being conducted (unpublished).

3.1.3. Aims

This study made use of the COST representative panel of *P. aeruginosa* and IPCD datasets to assess adaptation by loss of function in CF infections. The aims of this chapter were to:

- Assess the presence of loss of function mutations in the COST International Panel of phenotypically and genotypically diverse *P. aeruginosa*
- Investigate environmental *P. aeruginosa* genomes for loss of function mutations characteristic of adaptation to the CF lung environment
- Identify new genes that appear to be adaptive by loss of function
- Identify genes that are maintained during CF infections

3.2. Methods

Genome sequencing, raw sequence read filtering and adapter trimming, read mapping and variant calling were performed as described in Chapter Two: General Methods. All variant results in this chapter were obtained by read mapping and variant calling to PAO1. Additionally a select few genomes were *de novo* assembled and scaffolded as described in General Methods.

3.2.1. Loss of function mutations

For both the COST International Panel genomes and the IPCD genomes genomes were mapped to reference genome PAO1 as described in General Methods, variants called and the effect of variants predicted. Nonsense mutations were then combined with evidence for deletions 30bp+ in size, up to complete absences as evidence of loss of function. Statistical significance of difference between environmental and CF genome nonsense mutations, larger deletions and absences were performed with the chi squared goodness of fit test with Yates continuity correction in R. The Benjamini-Hochberg (BH) multiple testing correction was applied to the significant p values with a false detection rate (FDR) of 5%.

3.3. Results

3.3.1. COST International Panel of *Pseudomonas aeruginosa*

Isolate genomes LESB58 and PA14 were not included in read mapping and variant calling analysis as they're complete genomes in addition to PAO1 as it's the reference genome used for read mapping, variant calling and effect prediction for these datasets. Because of problems with contamination, strain NN2 has been omitted from further inclusion in the COST panel and 2192 is not analysed here, because the data are a composite of different isolates (table 3.1).

Table 3.1. International COST panel isolate genomes mapped to PAO1 for variant calling

Isolate	Analysed by variant calling
RP1	**Yes**
AUS23	**Yes**
DK2	**Yes**
AMT 0060-1	**Yes**
AMT 0060-2	**Yes**
AMT 0060-3	**Yes**
PAK	**Yes**
IST 27 mucoid	**Yes**
IST 27N	**Yes**
968333S	**Yes**
679	**Yes**
NH57388A	**Yes**
1709-12	**Yes**
Jpn 1563	**Yes**
LMG 14084	**Yes**
Pr335	**Yes**
U018a	**Yes**
CPHL 9433	**Yes**
LES400	**Yes**
LES431	**Yes**
39016	**Yes**
Mi162	**Yes**
C3719	**Yes**
AES-1R	**Yes**
AUS52	**Yes**
AA2	**Yes**
AMT0023-30	**Yes**
AMT0023-34	**Yes**
CHA	**Yes**
15108/-1	**Yes**
13121/-1	**Yes**
39177	**Yes**
KK1	**Yes**
A5803	**Yes**
TBCF10839	**Yes**
57P31PA	**Yes**
AA44	**Yes**
AA43	**Yes**
LESB58	No
PA01	No (reference)
PA14	No
NN2	No
2192	No

3.3.2. Loss of Function in the COST Panel Genomes

All genomes analysed were investigated for loss of function events (nonsense mutations and larger deletions) that are thought to be associated with CF. Genes involved in mucoidy, DNA mismatch repair, *gacA/S* switch to chronicity, biofilm dispersal gene *rbdA* and motility gene *fleR* were among those investigated for loss of function. The approach was also taken to identify genes that appear to be independently acquiring nonsense mutations to discover new genes associated with adaptation to the CF lungs by loss of function.

Loss of function in anti-sigma factor gene, *mucA* is associated with the mucoidy phenotype. Nine isolate genomes contained evidence for loss of function in MucA (table 3.2). All isolates with loss of function in MucA, except 968333S from a non-cystic fibrosis bronchiectasis patient sample, were sampled from CF infections. Compensatory mutations are possible that revert alginate overproduction once negative regulation by *mucA* is lost. Genes involved in alginate biosynthesis a*lgD* (PA4530), *alg8* (PA3541), *alg44* (PA3542), *algK* (PA3542), *algE* (PA3544), *algG* (PA3545) , *algX* (PA3546), *algL* (PA3547), *algI* (PA3548), *algJ* (PA3549), *algF* (PA3550), *algA* (PA3551) and sigma factor gene *algU* that could be compensatory by loss of function to the mucoid phenotype, were investigated for nonsense mutations and larger deletions/absences. The genome of strain AUS23 contains a 2bp insertion in sigma factor *algU*. The genome of strain Mi162 contains a 2bp insertion at reference position 3975435 and a 2bp deletion at reference position 3975437, so it is unlikely these mutations are entirely disruptive. LES400 and LES431 contain a nonsense mutation in *algF*.

Table 3.2. Evidence for loss of function in *mucA* in the COST panel genomes

Isolate	Mutation and impact	Reference position	Source of isolation
AMT 0060-2	1bp frameshift deletion	832286	CF (early)
IST 27 mucoid	1bp frameshift deletion	832338	CF
IST 27N	1bp frameshift deletion	832338	CF
968333S	7bp frameshift deletion	832140	Non-CF Bronchiectasis
NH57388A	89bp deletion		CF
AUS23	5bp frameshift insertion	832197	CF
DK2	1bp frameshift deletion	832338	CF (transmissible)
AUS52	C-A premature stop codon*	832194	CF
	1bp frameshift deletion	832358	
AMT0060-1	1bp frameshift insertion	832274	CF
	1bp frameshift deletion	832286	

Nonsense mutation in DNA mismatch repair genes, *mutS, mutL, mutM* and *uvrD* can be evidence for hypermutable lineages of *P. aeruginosa*. Isolate 968333S presents an 11bp frame-shifting deletion in *mutS*.

The quorum sensing transcriptional regulator gene *lasR* has acquired a nonsense mutation in seven of the panel genomes and was the gene was completely absent in a further two (table 3.3). CF isolate genome AMT0023-34 carries a 1bp frame-shifting insertion in *lasR* whereas an earlier isolate from the same patient does not carry a loss of function in *lasR* (AMT0023-30). Isolates carrying a nonsense mutation in *lasR* were primarily sourced from CF patients, but also included isolates sourced from a community acquired pneumonia (CAP) patient (A5803) and a non-CF burn patient isolate (Mi162). Additionally

an environmental isolate from a tobacco plant (CPHL9433) carried a nonsense mutation in *lasR* (table 3.3).

Table 3.3. Evidence for loss of function in *lasR* in the COST panel genomes

Isolate	**Mutation and impact**	**Reference position**	**Source of Isolation**
CPHL 9433	2bp frameshift insertion	1558835	Tobacco plant
LES400	7bp frameshift insertion	1558527	CF
Mi162	168bp frameshift insertion	1558840	Non-CF burn
AMT0023-34	1bp frameshift deletion	1558316	CF (late)
AUS23	G-T, premature stop codon	1558480	CF (transmissible)
AUS52	G-T, premature stop codon	1558603	CF (transmissible)
A5803	G-A, premature stop codon	1558349	CAP*
KK1	Complete absence of *lasR*	1558171-1558890	CF
DK2	Complete absence of *lasR*	1558171-1558890	CF (transmissible)

Four isolate genomes contained a nonsense mutation in the biofilm dispersal gene *rbdA*. They were isolated from CF (C3719, TBCF10839), the hospital environment (Pr335) and a (keratitis) eye infection (39177). Transmissible isolates LES400 and LES431 were observed to have acquired a premature stop codon in *fleR*. Transmissible strain isolates C3719 and AUS52 were phenotypically non-motile but the potential genetic basis could not be determined.

3.3.2.1. Independent acquisition of nonsense mutations

The independent acquisition of nonsense mutations was determined to identify genes that may be adaptive by loss of function. Different loss of function alleles were identified by nonsense mutations presenting in the same genes but in a different position or where they are represented by a different mutation at the same position. As such unique single nonsense mutations carried in different genomes must represent multiple acquisitions and if associated with a particular phenotype or isolate source, may be inferred to be adaptive.

In total there were 799 different genes with nonsense mutations, 410 of which were present in only one genome. For those in two or more genomes an assessment of whether they have been independently acquired or are shared by closely related isolates was carried out. Those that have acquired loss of function mutations independently were inferred to be adaptive, table 3.4 displays those in which there were three or more independent acquisitions of loss of function in the COST panel genomes. Genes, *mucA*, *lasR* that have already been evaluated for loss of function due to their known roles in CF lung infections were also identified by this method as the two genes most often acquiring nonsense mutations independently.

Table 3.4. Genes in which nonsense mutations have been acquired independently 3 or more times across the COST panel genomes

Gene	Independent occurrences	Genomes	Gene description
lasR	7	7	Transcriptional regulator of quorum sensing
mucA	6	9	Anti-sigma factor for alginate biosynthesis
mexA	4	6	*mexA*B-oprM efflux pump
rbdA	4	4	Biofilm dispersal
wbpM	3	17	O-antigen lipopolysaccharide
PA3124	3	10	Probable transcriptional regulator
exoY	3	7	Adenylate cyclase
PA4798	3	5	Hypothetical protein
PA2037	3	3	Hypothetical protein
PA0313	3	3	L-cysteine transporter of ABC system YecS
mexB	3	3	*mexA*B-oprM efflux pump
lpxO2	3	3	lpxO2 lipopolysaccharide biosynthetic protein
PA4469	3	3	Hypothetical protein
PA1221	3	3	Hypothetical protein
ladS	3	3	Histidine kinase
PA4063	3	3	Hypothetical protein
PA0987	3	3	Hypothetical protein

Both *mexB* and *mexA* of the *mexAB-oprM* multidrug efflux pump were observed to have independently acquired nonsense mutations 3 and 4 times respectively (table 3.5). When the entire operon was investigated for loss of function, eleven isolate genomes contained evidence for loss of function in the MexAB-OprM multidrug efflux pump (table 3.5). Sequential isolates AA2 (early) AA43 and AA44 (late) contained the same frameshift mutation in *mexA*. Isolates IST27 (mucoid) and IST27N (spontaneous non-mucoid) contained the same frameshift mutation in the *oprM* gene of the operon. The 'CF late' isolate AMT0023-34 contained a premature stop codon in *mexB* whereas early isolate AMT0023-30

did not contain any loss of function mutations in the operon. The transmissible strain isolate DK2 had a 78bp deletion in *mexB* and also a frameshift in *mexA*.

Table 3.5. Evidence for loss of function in the *mexAB-oprM* operon across the COST panel genomes

Gene	Isolate	Mutation and impact	Reference position	Source of isolation
mexB	AUS23	C-T premature stop codon	475507	CF (transmissible)
	LES431	1bp frameshift insertion	473231	CF (transmissible)
	AMT0023-34	C-T premature stop codon	474952	CF (late)
	DK2	78bp deletion		
mexA	AA2	1bp frameshift deletion	472324	CF (early)
	AA43	1bp frameshift deletion	472324	CF (late)
	AA44	1bp frameshift deletion	472324	CF (late)
	DK2	2bp frameshift insertion	472744	CF (transmissible)
	NH57388A	1bp frameshift deletion	472570	CF
	AUS52	1bp frameshift deletion	472223	CF
oprM	IST27	2bp frameshift deletion	477519	CF
	IST27N	2bp frameshift deletion	477519	CF

3.3.3. International *Pseudomonas* Genomics Consortium Dataset

To assess whether there was an association with loss of function of individual genes and CF infections, gene absences including genes carrying inactivating nonsense mutations were compared between an environmental population of *P. aeruginosa* and those isolated from CF lung infections. Reads were mapped to *P. aeruginosa* reference genome PAO1 and assessed for absence, large deletions and small frameshifting indels and premature stop codons as described.

Isolate genomes from CF infections and those sampled in the environment were obtained from the IPCD dataset. Some CF patients had contributed multiple isolates so to avoid bias (enrichment of results by multiple isolates of the same within patient population) one genome was randomly selected from each. Where patients had provided isolates representing multiple ST's it was determined whether the ST's were sufficiently different to most likely represent independent environmental lineages and a representative of each randomly selected. The CF genome set was corrected as such resulting in 388 genomes from a total of 458. The environmental isolates were obtained from multiple countries and diverse ecological niches including river samples from Australia, West Africa and Belgium, a poultry water line sample from Canada, lake samples, Japanese coastal and open ocean sea water samples, soil and plants samples from the U.S.A., sea water samples from Tunisia and shallow well water sample from Puerto Rico. Environmental genomes isolated from the hospital environment were excluded due to potential overlap with infecting populations resulting in a final environmental genome dataset of 137 from a total of 178.

3.3.4. Genes overrepresented for loss of function in CF compared with the environment

There were 90 different genes in the CF genomes that were significantly (chi squared $p < 0.05$) overrepresented for loss of function mutations amongst the genomes of CF isolates (table 3.6) compared with those from the environment. Of these, 34 were significant to $p < 0.005$. The most significantly overrepresented gene for loss of function in CF compared with the environment was type IV fimbrial biogenesis encoding gene, *pilC* ($p = 3.74 \times 10^{-9}$). There were 36 genes of unknown function (hypothetical or conserved hypothetical), including genes among the most significantly overrepresented for loss of function such as PA0561 ($p=2.4 \times 10^{-7}$) and PA1153 ($p= 2.9 \times 10^{-6}$). The genes that were overrepresented for loss of function in CF compared with the environment were associated with a variety of functions including motility (*flhB*), adhesion (*cup* genes), pyoverdine biosynthesis (*pvdO, pvdE, pvdF*), pyoverdine receptor (*fpvA*), antiobiotic resistance (*mpl*) and alginate biosynthesis (*mucA*). Due to multiple genes being tested for a significant difference in absence and inactivating mutations (n = 5606) the Benjamimi-Hochberg (BH) multiple testing criteria was applied with a false detection rate of 5%. The top 19 chi squared p-values were lower than the BH critical value and so remained significant ($p = 3.74 \times 10^{-9}$ – 1.14×10^{-4}); these rows are in bold in table 3.6.

Table 3.6. All genes significantly (chi squared p <.05) overrepresented for loss of function in CF genomes compared with the environment. Significant P values smaller than the Benjamimi-Hochberg (BH) critical value (FDR 5%) are highlighted in bold

Locus tag	Gene name	Description of function	Chisq P Value	BH critical value
PA4527	*pilC*	**still frameshift type 4 fimbrial biogenesis protein PilC**	**3.74E-09**	**9.09E-06**
PA4526	*pilB*	**type 4 fimbrial biogenesis protein PilB**	**8.81E-08**	**1.82E-05**
PA4312		**conserved hypothetical protein**	**2.36E-07**	**2.73E-05**
PA0561		**hypothetical protein**	**2.66E-06**	**3.64E-05**
PA1153		**hypothetical protein**	**2.92E-06**	**4.55E-05**
PA0619		**probable bacteriophage protein**	**3.44E-06**	**5.45E-05**
PA2566		**conserved hypothetical protein**	**4.19E-06**	**6.36E-05**
PA2403		**hypothetical protein**	**6.81E-06**	**7.27E-05**
PA2398	*fpvA*	**ferripyoverdine receptor**	**1.68E-05**	**8.18E-05**
PA1887		**hypothetical protein**	**1.69E-05**	**9.09E-05**
PA2397	*pvdE*	**pyoverdine biosynthesis protein PvdE**	**2.98E-05**	**1.00E-04**
PA2564		**hypothetical protein**	**3.23E-05**	**1.09E-04**
PA5516	*pdxY*	**pyridoxamine kinase**	**3.32E-05**	**1.18E-04**
PA3486	*vgrb4 G*	**Type VI secretion system**	**3.43E-05**	**1.27E-04**
PA2395	*pvdO*	**PvdO**	**5.55E-05**	**1.36E-04**
PA3487	*tle5*	**Lipid metabolism**	**7.35E-05**	**1.45E-04**
PA3488	*tli5*		**7.35E-05**	**1.55E-04**
PA1889		**hypothetical protein**	**8.14E-05**	**1.64E-04**
PA2396	*pvdF*	**pyoverdine synthetase F**	**1.14E-04**	**1.73E-04**
PA4075		hypothetical protein	1.95E-04	1.82E-04
PA4020	*mpl*	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	2.54E-04	1.91E-04
PA0763	*mucA*	anti-sigma factor MucA	3.22E-04	2.00E-04
PA4514	*piuA*	probable outer membrane receptor for iron transport	5.66E-04	2.09E-04
PA2256	*pvcC*	paerucumarin biosynthesis protein PvcC	6.09E-04	2.18E-04
PA2457		Hypothetical protein	1.29E-03	2.27E-04
PA1150	*pys2*	Pyocin S2	1.31E-03	2.36E-04
PA4999	*waaL*	O-antigen ligase	1.34E-03	2.45E-04
PA1628		probable 3-hydroxyacyl-CoA dehydrogenase	1.51E-03	2.55E-04
PA2319		Probable transposase	2.11E-03	2.64E-04

PA3434		Probable transposase	2.11E-03	2.73E-04
PA4310	*pctB*	chemotactic transducer PctB	2.22E-03	2.82E-04
PA3993		Probable transposase	2.96E-03	2.91E-04
PA4797		Probable transposase	2.96E-03	3.00E-04
PA2388	*fpvR*	FpvR	3.59E-03	3.09E-04
PA4133	*cytN*	cytochrome c oxidase subunit (cbb3-type)	3.81E-03	3.18E-04
PA2300	*chiC*	chitinase	5.12E-03	3.27E-04
PA3346		two-component response regulator	5.12E-03	3.36E-04
PA5226		hypothetical protein	5.65E-03	3.45E-04
PA2386	*pvdA*	L-ornithine N5-oxygenase	5.68E-03	3.55E-04
PA0558		Hypothetical protein	5.70E-03	3.64E-04
PA0445		Probable transposase	5.74E-03	3.73E-04
PA2690		Probable transposase	5.74E-03	3.82E-04
PA2387	*fpvI*	FpvI	5.84E-03	3.91E-04
PA2426	*pvdS*	sigma factor PvdS	6.65E-03	4.00E-04
PA2427		hypothetical protein	6.65E-03	4.09E-04
PA0626		hypothetical protein	6.66E-03	4.18E-04
PA2456		Hypothetical protein	7.75E-03	4.27E-04
PA0623		probable bacteriophage protein	7.76E-03	4.36E-04
PA3867		Probable DNA invertase	9.40E-03	4.45E-04
PA3866		Pyocin S4	9.56E-03	4.55E-04
PA4092	*hpaC*	4-hydroxyphenylacetate 3-monooxygenase small chain	1.01E-02	4.64E-04
PA5255	*algQ*	Alginate regulatory protein AlgQ	1.01E-02	4.73E-04
PA1888		hypothetical protein	1.05E-02	4.82E-04
PA0614		hypothetical protein	1.20E-02	4.91E-04
PA0618		probable bacteriophage protein	1.23E-02	5.00E-04
PA0624		hypothetical protein	1.23E-02	5.09E-04
PA0628		conserved hypothetical protein	1.23E-02	5.18E-04
PA0981		Hypothetical protein	1.53E-02	5.27E-04
PA4071		hypothetical protein	1.56E-02	5.36E-04
PA0627		conserved hypothetical protein	1.63E-02	5.45E-04
PA0648		Hypothetical protein	1.85E-02	5.55E-04
PA4113	*ydeA*	probable major facilitator superfamily (MFS) transporter	1.87E-02	5.64E-04
PA2565		hypothetical protein	1.91E-02	5.73E-04
PA0647		Hypothetical protein	2.03E-02	5.82E-04
PA1154		conserved hypothetical protein	2.19E-02	5.91E-04
PA0728		probable bacteriophage integrase	2.48E-02	6.00E-04
PA0622		probable bacteriophage protein	2.50E-02	6.09E-04
PA0724	*coaA*	Probable coat protein A of bacteriophage Pf1	2.62E-02	6.18E-04
PA2393		Putative dipeptidase, involved	2.64E-02	6.27E-04

		in pyoverdine biosynthesi		
PA1339	*aatP*	AatP	2.75E-02	6.36E-04
PA1937		conserved hypothetical protein	2.80E-02	6.45E-04
PA0615		hypothetical protein	2.85E-02	6.55E-04
PA0616		hypothetical protein	2.85E-02	6.64E-04
PA0617		probable bacteriophage protein	2.85E-02	6.73E-04
PA0625		hypothetical protein	2.85E-02	6.82E-04
PA1151	*imm2*	Pyocin S2 immunity protein	2.95E-02	6.91E-04
PA3869		Hypothetical protein	3.03E-02	7.00E-04
PA2203		probable amino acid permease	3.19E-02	7.09E-04
PA0977		hypothetical protein	3.23E-02	7.18E-04
PA0983		Conserved hypothetical protein	3.89E-02	7.27E-04
PA0081	*fha1*	Fha1	3.97E-02	7.36E-04
PA4082	*cupB5*	adhesive protein CupB5	4.14E-02	7.45E-04
PA3825		Cyclic-guanylate-specific phosphodiesterase	4.15E-02	7.55E-04
PA0729		Hypothetical protein	4.18E-02	7.64E-04
PA1449	*flhB*	flagellar biosynthetic protein FlhB	4.48E-02	7.73E-04
PA1626		probable major facilitator superfamily (MFS) transporter	4.53E-02	7.82E-04
PA5546		conserved hypothetical protein	4.53E-02	7.91E-04
PA0980		Hypothetical protein	4.79E-02	8.00E-04
PA2392	*pvdP*	Pyoverdine biosynthesis	4.80E-02	8.09E-04
PA3865.1		pyocin S4 immunity protein	4.91E-02	8.18E-04

The majority of associated functions to genes significantly ($p < 0.5$) overrepresented for loss of function in CF compared with the environment are hypothetical genes of unknown function (24.32%), followed by those related to phage, transposon and plasmid (19.81%), adaptation and protection (9.09%) and transport of small molecules (8.10%) (figure 3.1).

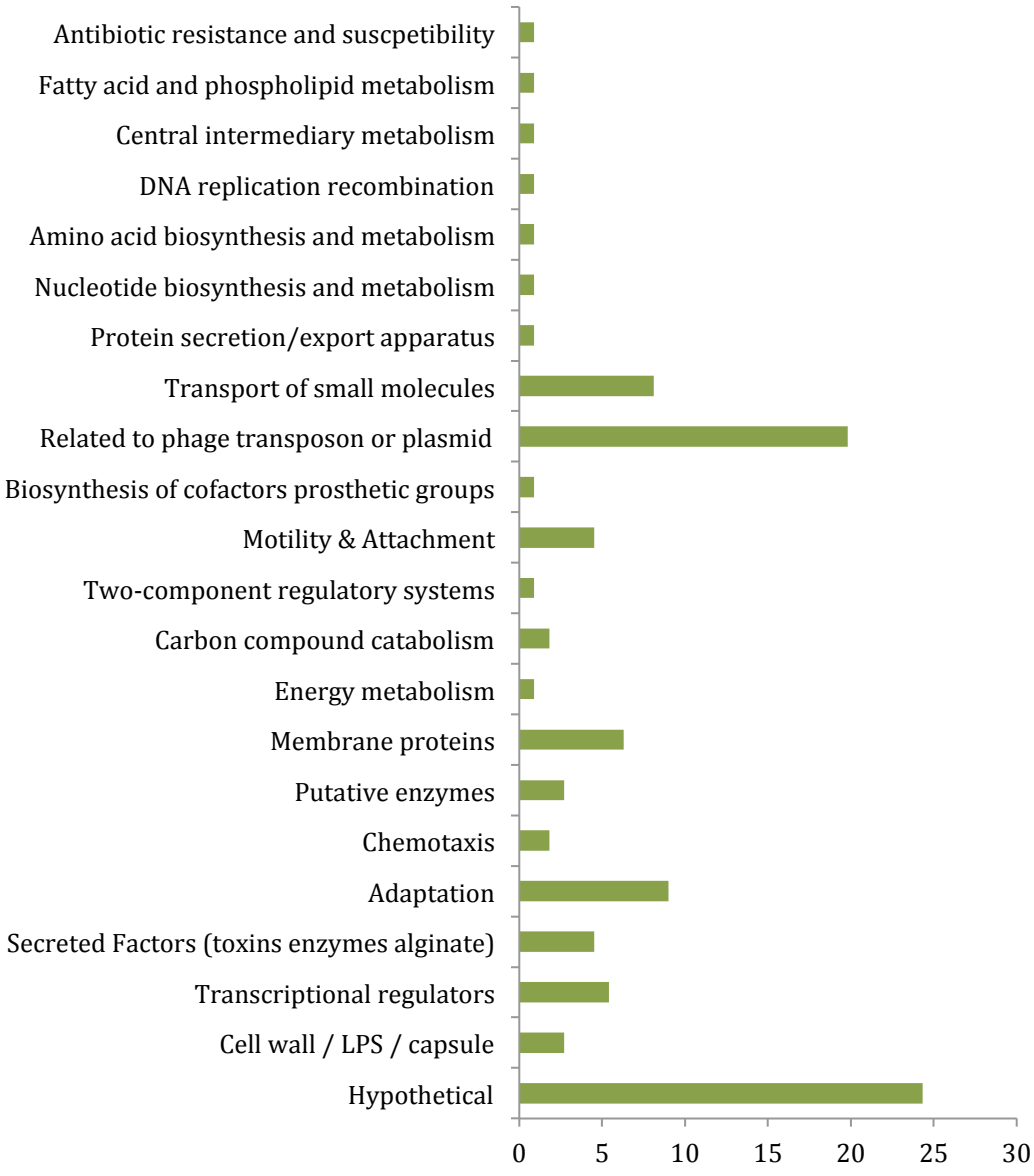


Figure 3.1. PseudoCAP functional categories associated with the genes significantly overrepresented for loss of function in the CF isolate genomes (<0.05 chi squared significance)

For the 35 genes that were overrepresented for loss of function in CF compared with the environment and are hypothetical genes, those of unknown function, the functions of full gene homologs in the Pseudomonas Database orthologous gene set were sought. Table 3.7 displays the details of orthologous group membership and function within other members of the species or genus. For eleven of these genes of unknown function in PAO1, between 24 and 332 members of various orthologous groups in various *Pseudomonas* were found but were also of unknown function. The majority of associated funtions (n = 11) were related to phage and transposable elements (PA0628, PA0627, PA1154, PA1937, PA1153, PA0626, PA0614, PA0624, PA0615, PA0616, PA0625 and PA0983).

Table 3.7. Orthologs and their functions of the genes significantly overrepresented for loss of function (chi squared $p < 0.05$) that are of unknown function in PAO1

PAO1 locus ID	**Orthologous group ID**	**Functional description**
PA4312	POG003910	Putative deacetylase
PA2566	POG003005	Pyridine nucleotide-disulfide oxidoreductase
PA0628	POG000611	phage late control gene D protein
PA0627	POG000610	putative phage protein X
PA1154	POG001113	Transposase
PA1937	POG000943	Transposase
PA5546	POG005077	Putative cyclopropan-fatty-acyl-phospholipid synthase
PA0561	POG000545	Hypothetical
PA1153	POG001112	Prophage antirepressor
PA1887	POG003523	Hypothetical
PA2564	POG001050	Methyltransferase type 12
PA1889	POG003521	Hypothetical
PA4075	POG001611	Methyltransferase
PA5226	POG004807	putative chromosome segregation ATPase
PA2427	POG003109	putative pvdY

PA0626	POG000609	Phage tail protein
PA1888	POG003522	Hypothetical
PA0614	POG000597	Putative holin
PA0624	POG000607	mu-like prophage FluMu gp41 family protein
PA4071	POG001615	Hypothetical
PA2565	POG003006	Hypothetical
PA0615	POG000598	phage protein
PA0616	POG000599	putative baseplate assembly protein V
PA0625	POG000608	tail length determinator protein
PA0977	POG000942	Hypothetical
PA0647	POG000627	Hypothetical
PA0648	POG000628	Hypothetical
PA0729	POG000705	RelE/ParE family plasmid stabilisation system protein
PA0980	POG000944	Hypothetical
PA0981	POG000945	Hypothetical
PA0983	POG000943	Transposase
PA2457	POG002259	yd repeat containing protein
PA2564	POG001050	Trans-aconitate 2-methyltransferase
PA3869	POG001802	Type II secretory pathway, component ExeA
PA4312	POG003910	Putative deacetylase

3.3.5. Genes underrepresented for loss of function in CF compared with the environment

There were 1409 different genes significantly (chi squared, $p < .05$) underrepresented for loss of function in CF. A little over half of these genes (n = 749) were significant in the range of p = 0.01-0.046. There were 375 genes that were significant to $p < 0.005$. Table 3.10 displays the 50 most significantly underrepresented for loss of function. BH multiple testing correction was performed also for genes underrepresented for loss of function in CF compared with the environment. The top 300 chi squared p-values remained significant were smaller than the BH critical value. The full list of genes underrepresented in CF for loss of function are provided in supplementary table 1.

Table 3.8. The top 50 most significantly (<0.05 chi squared P value) underrepresented genes for loss of function in CF genomes compared with the environment. Significant P values smaller than the Benjamimi-Hochberg (BH) critical value (FDR 5%) are in highlighted in bold

Locus tag	Gene name	Description of function	Chisq p-value	BH Critical value
PA4586		hypothetical protein	1.49E-19	8.92E-06
PA4162		probable short-chain dehydrogenase	2.75E-18	1.78E-05
PA3105	*xcpQ*	general secretion pathway protein D	9.22E-17	2.68E-05
PA4192		probable ATP-binding component of ABC transporter	1.09E-16	3.57E-05
PA4193		probable permease of ABC transporter	1.09E-16	4.46E-05
PA4194		probable permease of ABC transporter	1.09E-16	5.35E-05
PA4195		probable binding protein component of ABC transporter	1.09E-16	6.24E-05
PA4191		probable iron/ascorbate oxidoreductase	3.04E-16	7.14E-05
PA1858	*str*	streptomycin 3''-phosphotransferase	8.22E-16	8.03E-05
PA3469		conserved hypothetical protein	2.16E-15	8.92E-05
PA5318		hypothetical protein	4.56E-15	9.81E-05
PA3841	*exoS*	exoenzyme S	1.44E-14	1.07E-04
PA4879		conserved hypothetical protein	3.28E-14	1.16E-04
PA3362		hypothetical protein	4.91E-14	1.25E-04
PA2130	*cupA3*	usher CupA3	5.15E-14	1.34E-04
PA2771		diguanylate cyclase with a self-blocked I-site	5.79E-13	1.43E-04
PA2772		hypothetical protein	8.97E-13	1.52E-04
PA1036		hypothetical protein	2.49E-12	1.61E-04
PA5321	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase	6.52E-12	1.69E-04
PA4985		Uncharacterized protein	8.84E-12	1.78E-04
PA0244		hypothetical protein	2.69E-11	1.87E-04
PA2182		hypothetical protein	3.13E-11	1.96E-04
PA0347	*glpQ*	glycerophosphoryl diester phosphodiesterase	4.21E-11	2.05E-04
PA0191		probable transcriptional regulator	4.31E-11	2.14E-04
PA2583		probable sensor/response regulator hybrid	6.50E-11	2.23E-04
PA2190		conserved hypothetical protein	7.91E-11	2.32E-04
PA1868	*xqhA*	secretion protein XqhA	8.24E-11	2.41E-04
PA3387	*rhlG*	beta-ketoacyl reductase	2.48E-10	2.50E-04
PA4222		probable ATP-binding component of ABC transporter	2.48E-10	2.59E-04
PA2188		probable alcohol dehydrogenase (Zn-	2.49E-10	2.68E-04

		dependent)		
PA2183		hypothetical protein	2.64E-10	2.77E-04
PA2127	*cgrA*	cupA gene regulator A	3.11E-10	2.85E-04
PA1576		probable 3-hydroxyisobutyrate dehydrogenase	3.35E-10	2.94E-04
PA2184		conserved hypothetical protein	3.80E-10	3.03E-04
PA2200		Cyclic-guanylate-specific phosphodiesterase	3.90E-10	3.12E-04
PA2185	*katN*	non-heme catalase KatN	4.42E-10	3.21E-04
PA0261		hypothetical protein	5.74E-10	3.30E-04
PA5404		hypothetical protein	6.17E-10	3.39E-04
PA2135		probable transporter	6.55E-10	3.48E-04
PA2112		conserved hypothetical protein	7.62E-10	3.57E-04
PA4063		hypothetical protein	7.64E-10	3.66E-04
PA2187		hypothetical protein	8.11E-10	3.75E-04
PA2296		hypothetical protein	1.15E-09	3.84E-04
PA2186		hypothetical protein	1.35E-09	3.93E-04
PA2440		hypothetical protein	1.41E-09	4.01E-04
PA0499		probable pili assembly chaperone	3.30E-09	4.10E-04
PA2295		probable permease of ABC transporter	3.39E-09	4.19E-04
PA3063	*pelB*	PelB	3.49E-09	4.28E-04
PA2167		hypothetical protein	3.59E-09	4.37E-04
PA2116		conserved hypothetical protein	3.64E-09	4.46E-04

The majority of associated functions to genes underrepresented for loss of function in CF compared with the environment were hypothetical genes of unknown function (32.75%). Associated functions included membrane proteins (9.37%), putative enzymes (7.91%), transport of small molecules (7.32%) and transcriptional regulators (7.67%); these represent approximately a third of the associated functions (figure 3.2).

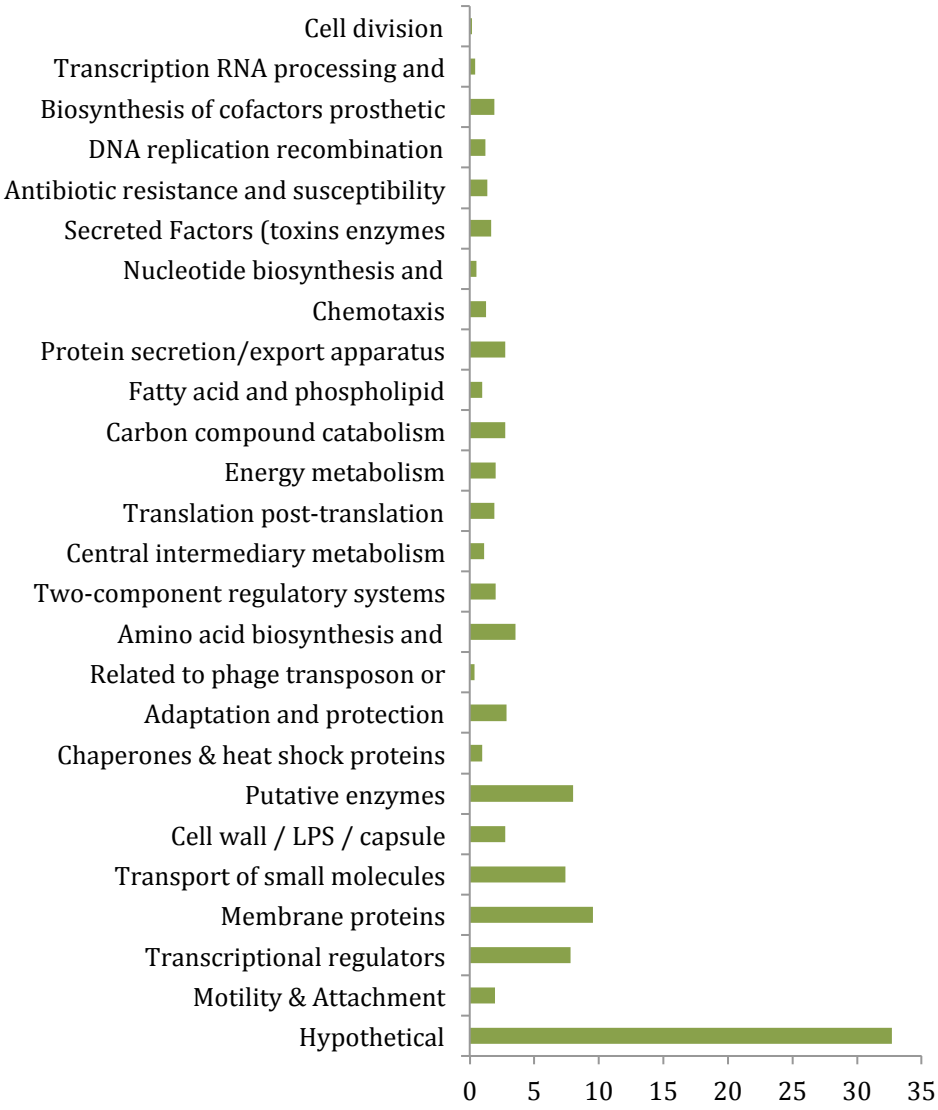


Figure 3.2. Functional categories represented by the genes underrepresented for loss of function in the CF isolate genomes (<0.05 chi squared significance)

3.3.5.1. Comparison of associated functional categories in over and underrepresented genes in CF isolates compared with environmental isolates

There were differences in the functional categories associated with over or underrepresented genes for loss of function (supplementary figure 1). Genes associated with cell division, chaperones and heat shock proteins, transcription RNA processing and degradation and also translation post-transcriptional modification and degradation were never overrepresented for loss of function in the CF isolate genomes. The largest associated function is unknown in both genes overrepresented and underrepresented for loss of function. Hypothetical genes represented 32.67% of all functional associations of underrepresented genes and 24.32% of overrepresented. Genes associated with functions transport of small molecules, related to phage transposon or plasmid, motility and attachment adaptation & protection and cell wall/LPS/capsule represent a greater proportion of functional associations for genes that are overrepresented for loss of function rather than underrepresented.

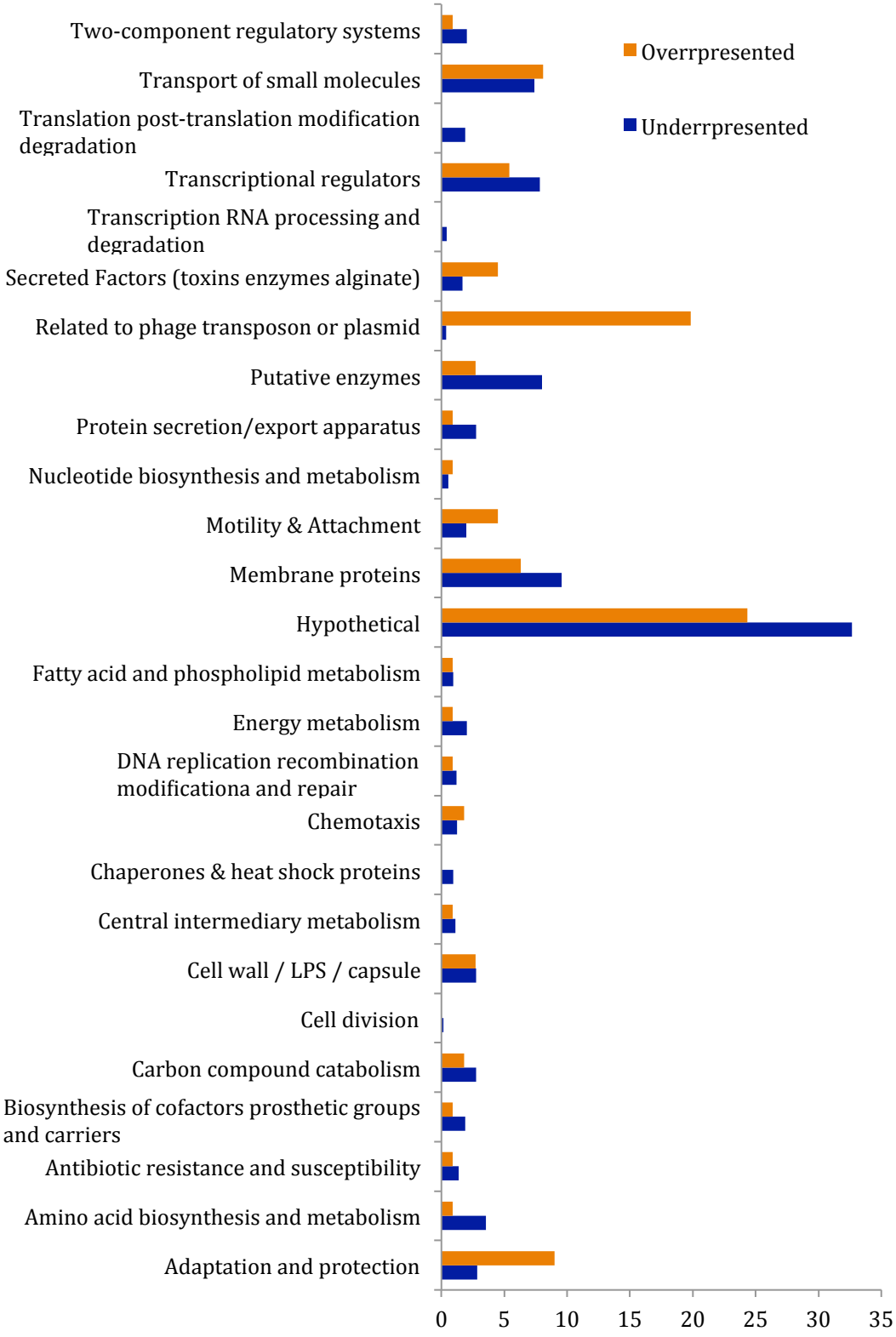


Figure 3.3. Comparison of associated pseudoCAP functional categories from overrepresented and underrepresented genes for loss of function in CF (<0.05 chi squared significance)

3.4. Discussion

This chapter focused on two approaches, inference of adaptive loss of function by the number of independent acquisitions in the CF population and comparison of a panel of CF isolate genomes and environmental isolate genomes. The first depends on the expectation that for a particular gene it should be unlikely that a variety of environmental niches from which *P. aeruginosa* infections are acquired would have the same selective pressures compared with the comparably discrete environment of the CF lung. Comparison of the CF and environmental isolates has been made possible by sequencing a large number of genomes. This has the benefit of potentially capturing bias in acquisition and also the ability to assess larger deletions and gene absences.

3.4.1. Genes underrepresented for Loss of Function in CF Compared with the Environment

There were 1409 different genes significantly ($p=<0.05$) underrepresented in the CF genomes compared with the environment. Compared with those overrepresented for loss of function they were more involved in amino acid biosynthesis and metabolism, antibiotic resistance and susceptibility, biosynthesis of cofactors and prosthetic groups and carries, cell division, central intermediary metabolism, chaperones and heat shock proteins, DNA replication recombination modification and repair, energy metabolism, fatty acid and phospholipid metabolism, membrane proteins, protein secretion/export apparatus, putative enzymes, transcriptional regulators and two component

systems. These results provide a wealth of information of genes that might be important for *P. aeruginosa* to maintain in CF.

The top 50 most significantly underrepresented genes in CF for loss of function were associated with diverse functions. For example the gene *xcpQ* associated with protein secretion[140] is present and without nonsense mutations in 53.28% of environmental genomes compared with 87.63% of those from CF. Specific antibiotic resistance genes also appear to be more conserved in CF than in the environment such as the streptomycin resistance gene *str*. The *str* gene (PA1858) is present and without loss of function mutations in 93.29% of CF genomes compared with 64.96%. Sensor/response regulator hybrid gene PA2583 has been observed to be upregulated in colistin resistant strains and knocking it out has a detrimental effect on expression of *pmrA* and *arnB* involved in polymyxin resistance[327]. PA2583 is significantly underrepresented for loss of function in CF ($p = 6.5 \times 10^{-11}$); it was present and without nonsense mutations in 85.56% of CF genomes compared with 58.39% of environmental. LysR-type regulator gene PA0191 has been observed to be induced in the presence of monosulfactam[328] and is underrepresented for loss of function in CF (chi squared, $p = 4.31 \times 10^{-11}$).

3.4.2. Genes Overrepresented for Loss of Function in CF Compared with the Environment

Genes previously identified to be adaptive by loss of function in CF were observed to be overrepresented for loss of function in this dataset[66]. Loss of function in the *mucA* gene was significantly overrepresented in the CF genomes compared with the environment ($p = 3.2 \times 10^{-4}$). Of 388 *P. aeruginosa* genomes from CF, 58 carried loss of function mutations in *mucA* compared with 4 of 137 from the environment. Seven CF isolate genomes with loss of function in *mucA* however were observed to carry potential compensatory mutations by loss of function in at least one of these genes involved in alginate biosynthesis. There wasn't however evidence for an overrepresentation of loss of function in DNA mismatch repair genes *mutS, mutL, mutM* and *uvrD* ($p = 0.32$). Loss of function was observed in at least one of these genes in 21 of 388 CF genomes and 4 of 137 environmental. It has previously been observed that iron acquisition by the heme, rather than the more common pyoverdine mechanism is preferred during CF[326] infections. Consistent with this 13 different pyoverdine biosynthesis and receptor genes were observed to be overrepresented for loss of function in CF (table 3.6). Loss of quorum sensing regulator LasR has been associated with CF lung infections[278] and though inferred to be adaptive by loss of function in the representative panel it was not observed to be significantly overrepresented for loss of function in the IPCD CF genomes.

Genes associated with antibiotic resistance and susceptibility were also observed to be overrepresented for loss of function in CF, both where loss of function is associated with increased and decreased antibiotic resistance. UDP-

N-acetylmuramate-alanyl-gamma-D-glutamyl-meso-diaminopimelate-ligase gene (*mpl*) was observed to be significantly overrepresented for loss of function in CF isolate genomes compared with the environment ($p = 2.53 \times 10^{-4}$). Loss of function in *mpl* has previously been identified in a transmissible CF strain, the LES[280]. Disruption of *mpl* has been associated with piperacillin and ceftazidime resistance due to an accumulation of immature cell wall components and up-regulation of AmpC expression[329]. The TonB dependent siderophore receptor PiuA is observed to be significantly overrepresented for loss of function. Previously disruption of the *piuA* gene has been observed to increase MICs 8 fold to the BAL30072 monosulfactam siderophore[328], though another study demonstrated the potential for compensatory iron import mechanisms *in vivo*[330]. Though the MexAB-oprM efflux pump is a major multidrug resistance mechanism the operon *mexAB-oprM* is observed in 65 CF genomes carry nonsense mutations, larger deletions and absences in the operon compared with 6 environmental genomes.

3.4.3. Limitations and future work

Different members of the species may carry orthologous genes of different types. These can be associated with the phylogenetic grouping, but so far the evidence suggests that the population involved with opportunistic CF infections (not including transmissible lineages) mirrors that of the environment[195]. Therefore bias in the species towards types shouldn't explain a significant overrepresentation of absence of that gene in CF isolate genomes. It does however reduce the sample size as the comparison should be considered as between sub-populations of the potential proportion carrying that homolog in

the environment. For example, separate types of *P. aeruginosa* siderophore receptor gene *fpvA* have been identified (corresponding to type I, II, or III pyoverdine)[125]. The reference genome PAO1 that all genomes were mapped to carries the *fpvA* gene encoding type I. Type I FpvA was carried in 25% of the CF genomes compared with 55.26% of those from the environment. Additionally a further two genomes from CF carry the type I *fpvA* gene but have each independently acquired nonsense mutations. Overall lack of type I FpvA is significant in CF compared with the environment (chi squared, $p = 1.68 \times 10^{-5}$). Mapping against a reference genome carrying one type of FpvA however reduces the sample size rendering the result more likely to be as a result of chance alone. The evolutionary considerations are also changed by this knowledge, an alternative hypothesis being that CF isolates carrying other types of FpvA are favoured rather than the absence of function. Greater environmental sampling to increase knowledge of such cases is desirable, or identification of orthologous types and subsequent mapping and variant calling against a representative panel of reference genomes.

Within host evolutionary dynamics may also explain some observations. *P. aeruginosa* secretes many costly metabolites during CF lung infection. Cheats, members of the population that enjoy the benefit of these products with none of the cost[331], can have a competitive advantage over producers though this is clearly self-limiting as the cheats to producers (upon which they depend) ratios may rapidly change[319,332]. In the CF lung, non-producers of costly siderophore pyoverdine, occur due to loss of function[333]. In some cases the retention of pyoverdine receptor genes suggest that these are cheats rather than simple

non-producers of a costly and unnecessary or actively disadvantageous feature[333].

3.4.4. Chapter summary

The COST representative panel was whole genome sequenced to investigate CF associated mutations. It was demonstrated that *lasR* and *mucA* commonly associated with CF infections were adaptive by loss of function and new genes potentially associated such as *mexB* proposed.

Subsequent analysis of loss of function in the IPCD CF compared with environmental isolate genomes showed:

- 92 genes overrepresented for loss of function in CF
- MucA loss of function significantly overrepresented in CF
- Genes where loss of function is commonly associated with CF such as *mutS, mutL, mutM, uvrD* and *lasR* not significantly overrepresented compared with the environment
- >1,000 genes underrepresented for loss of function to be further investigated

Chapter Four

Whole Genome Sequencing of Historically and Geographically Diverse *Pseudomonas aeruginosa* Liverpool Epidemic Strain Isolates

4. Introduction

4.1. Background of LES strains

The discovery of the Liverpool Epidemic Strain (LES) provided the first molecular evidence to confirm suspicions of *P. aeruginosa* cross infection between CF patients[192]. Molecular surveys of CF clinics following its discovery revealed that the LES was the most common single clone in CF clinics nationwide[211]. Further, intercontinental spread of the strain was observed as the LES and the 'LES-like' were isolated and typed from CF patients in Canada[334]. Distinct transmissible strains have since been observed in the Midlands and Manchester U.K., the 'Midlands' and 'Manchester Epidemic Strains'[335], in Australia strains 'AES1'[288,336], 'AESII' and 'AESIII'[337], Denmark the 'DK2'[279] strain, Belgium[338] and in Canada the 'Prairie Epidemic Strain'[339]. Prior to the discovery of cross infecting lineages of *P. aeruginosa* in CF the *Bukrholderia cenocepacia* ET12 lineage had been discovered to transmit between patients prompting segregation of patients in affected clinics[205,340]. Following the discovery of the LES segregation of patients was implemented for LES-positive clinics, successfully diminishing new instances in some U.K. clinics[225]. The LES also appears to super-infect patients, replacing existing chronic *P. aeruginosa* lung infections[341]. Patients chronically infected with the

LES also carry a poorer prognosis and isolates present elevated levels of antibiotic resistance[210].

4.1.1. Epidemiology

The LES was discovered in 1996 when, prompted by patterns of resistance across *P. aeruginosa* positive CF patients, a molecular investigation was conducted at Alder Hey Children's Hospital, Liverpool, U.K. Molecular typing revealed that 55/65 patients were harbouring the same strain making cross infection highly likely; it was suggested that ceftazidime monotherapy may have resulted in the emergence of the strain[192]. Following the discovery of the LES, a molecular epidemiological survey was undertaken in England that was published in 2004[211]. In total, 1225 *P. aeruginosa* isolates from CF infections, from 31 clinical centres were genotyped. It was demonstrated that while the majority (78%) of patient's harboured individual genotypes, transmissible strains were the most common single genotypes[211]. Fifteen centres presented 93 isolates of the LES making it the most common single clone. These initial findings were corroborated by later studies, finding that in Scotland the LES was also present in 10/43 patients sampled[342] and that in the adult CF centre in Liverpool 63/80 patients were LES positive[206]. Later, in 2008 the LES was sampled from clinics in Canada[343].

4.1.2. Microbiology and clinical features

It has been shown that some isolates of the LES over-produce quorum sensing regulated secreted virulence factors such as pyocyanin LasA and elastase,

raising their virulence in infection models[344,255]. However, the quorum sensing regulator LasR is also frequently observed to be independently mutated resulting in loss of function[104,154,278]. The LES is additionally an unusual *P. aeruginosa* for having lost its motility and always lacking a flagella[345], presumably due to a fixed nonsense mutation in the gene *fleR.*

The mechanism of transmission by the LES is not yet understood but studies of the hospital environment have found no lasting environmental reservoirs[346]. Repeated sampling of nearby dry-surfaces demonstrates that the LES is present only initially[346]. Similarly the LES does present in air samples for a short time[346]. In addition to the ability to cross infect CF patients, the LES has demonstrated a number of interesting clinical features. The LES can super-infect existing chronic *P. aeruginosa* lung infections in CF[341], infection of both non-CF parents of a CF patient has been reported[347], the LES has also been implicated in non-CF bronchiectasis lung infections and transmission to a pet cat, Osiris, has occurred[348].

4.1.3. The LES genome

Isolates pre-dating the discovery of the LES were identified from frozen samples by typing. LES reference isolate LESB58 is the oldest dated isolate as it was archived in 1988. LESB58 was the first LES isolate genome to be sequenced and this was published in 2009 by Winstanley *et al.*[291]. LESB58 was observed to have a 6.7Mbp genome carrying 6,027 genes, 5,931 of which are protein coding. The LES genome was aligned to *P. aeruginosa* reference genomes already

sequenced, PA01, PA14 and PA7 and determined to contain 574, 528 and 825 orthologs respectively[291]. It was also possible to determine that the LESB58 genome carries 5 genomic islands and 6 prophages (table 4.1). A 10,344bp region of prophage 2 is homologous with a region prophage 3. Prophage 3 and 5 have two homologous regions[349]. Additionally prophage 4 shows homology with D3112 phage and prophage 6 belongs to the filamentous phage family *Inoviridae* (Pf1)[349]. A rat infection model determined that LES prophages 2, 3 and 5 demonstrated that they were associated with *in vivo* competitiveness while prophage1 is defective[291].

Table 4.1. Genomic island and prophage regions identified in LESB58[291]

Region name	Start	End	No. of genes
LESGI-1	2504700	2551100	31
LESGI-2	2751800	2783500	18
LESGI-3	2796836	2907406	107
LESGI-4	3392800	3432228	32
LESGI-5	4931528	4960941	26
Prophage 1	665561	680385	19
Prophage 2	863875	906018	44
Prophage 3	1433756	1476547	53
Prophage 4	1684045	1720850	48
Prophage 5	2690450	2740350	65
Prophage 6	4545190	4960941	12

Additional LES genomes were sequenced for comparative genomic studies assessing within and between strain diversity[214,334]. Dettman *et al*., conducted a comparative genomics study of epidemic (LES) and non-epidemic strain genomes. Reconstructing a phylogenetic tree of polymorphic sites in the core genome of LES and non-LES isolate genomes combined they showed that the British and Canadian LES formed a monophyletic group[214]. Further, it was

determined which genes contributed the most diversity on the branch leading to the LES clade to infer, by their functional categories, mutational changes that may be characteristic of the strain. Functional categories, kinase activity, arginine metabolism, redox, protein secretion and cation transport were most overrepresented for mutations by function in the core genome[214].

A detailed comparison was then conducted by Jeukens *et al.,* of British and Canadian LES isolate genomes. Isolates LES400, LES431 and LESB65 from the U.K., combined with reference LESB58 and LESlike1, LESlike4, LESlike5 and LESlike7 isolates from Canada (Ontario, Ottawa, Toronto and Hamilton respectively) were whole genome sequenced[334]. In the accessory genome of these strains it was shown that LESB58 genomic islands and prophages varied. LES prophage 5 was absent in 5/7 genomes while genomic island 4 was present in all. Analysis of core genome polymorphisms determined that LES400 and LES431 the oldest isolates after LESB58 were genetically most similar to LESB58[334]. The genes *lasR, vfr, exsA, ropN, cyaB* and *ampD* were identified to be the most diverse in the core genome. Further 1 in 6 mutations in genes across the genomes were carried in regulatory genes[334].

4.1.4. Aims of this study

The aim of this chapter was to use whole genome sequencing of a historically and geographically diverse panel of Liverpool Epidemic Strain genomes to assess:

- If the Canadian and UK LES are genetically distinct
- The genomic epidemiology of the isolates
- Adaptive strategies
- Pangenome structure of the most diverse set of the LES yet

4.2. Methods

Genome sequencing, raw sequence read filtering and adapter trimming, read mapping and variant calling and genome assembly were performed as described in Chapter Two: General Methods. All sequencing reads mapped for variant calling in this chapter were mapped to LESB58 as described in General Methods. Variant calling from mapped reads was used for all analysis of mutations whereas assembled genomes were used for all other analyses.

4.2.1. Scaffolding to reference LESB58

The syntenies of all *de novo* assembled contigs were determined by scaffolding against the genome of LESB58[291] using ABACAS[294]. Contigs containing sequence accessory to the LESB58 genome were appended to the end of the draft genome.

4.2.2. Percent identity of LESB58 prophages and genomic islands

A database of LES genomes was generated using blast+ makeblastdb against which each LESB58 prophage and genomic island was aligned using blast+, blastn (window size 28). The contribution to percent identity by duplications was corrected for by reciprocally aligning those regions with known duplications with blastn. By aligning the duplications from respective prophages it could be determined how much of the overall alignment for that prophage was a result of the duplicated sequence. The results were visualised as a heatmap according to final percent identity with the LESB58 prophages and

genomic islands and generated using the interactive tree of life software (iTol)[313].

4.2.3. Pan genome analysis

The pan-genome was determined using large scale blast score ratio pipeline (LS-BSR)[350], with standard parameters and alignment by blastn. The pan-genome matrix was randomly sampled at different depths for 100 iterations to determine the core and accumulation curves using LS-BSR post processing script *BSR_to_gene_accumulation_scatter.py*[350] and visualised respectively using the boxplot package in R. Pan-genome statistics were generated with LS-BSR post-processing tool *pan_genome_stats.py*[350]. Core and accessory gene curves were visualised using boxplots in R.

4.2.4. Inference of positive selection

Pseudoalignments were generated of LESB58 genes to which reads were aligned to the full length of the gene and incorporating variants from GATK. Multiple alignments were produced by clustalW. Maximum likelihood gene trees were approximated using RaxML with the GTR substitution model[311]. Inference of codon selection was peformed using hyphy[351], dNdS ratios averaged across all codons (omega) of <1 and purifying selection is inferred whereas >1, positive selection is inferred.

4.3. Results

A geographically diverse and historic sample of the LES was selected for this study and a total of 93 such isolates were sequenced. In order to analyse the most diverse sample possible, these were combined with 50 LES genomes already published from comparative genomics studies and representative sub-samples from within-patient studies. The genomes that have been sequenced are from a diverse range of hosts too, such as bronchiectasis patients, primary cilliary dykinesis (PCD) patients, a pet cat, the non-CF parents of a CF patient and CF patients. The geographic source of these 143 LES, primarily from various centres in the UK and also Canada are listed in table 4.2.

Table 4.2. LES genomes analysed in this study, the source of isolation according to PHE area designations, pathology and sequencing centre or publication from which they were obtained (Darch *et al.*[352], Williams *et al.*[280], Jeukens *et al.*[334])

Isolate	Year if known	Country	City or region	Pathology	Sequencing centre or data source
49070	2004	UK	Liverpool	CF	Liverpool CGR
CF8S1_27	2009	UK	Liverpool	CF	Liverpool CGR
72G	2003	UK	Liverpool	CF	Liverpool CGR
o59100	2005	UK	Liverpool	CF	Liverpool CGR
49400	2004	UK	Liverpool	CF	Liverpool CGR
49492	2004	UK	Liverpool	CF	Liverpool CGR
49348	2003	UK	Liverpool	CF	Liverpool CGR
30567S	2003	UK	Liverpool	CF	Liverpool CGR
89577	2008	UK	Liverpool	CF	Liverpool CGR
89570	2008	UK	Liverpool	CF	Liverpool CGR
CF3S1_08	2009	UK	Liverpool	CF	Liverpool CGR
89244	2008	UK	Liverpool	CF	Liverpool CGR
89347	2008	UK	Liverpool	CF	Liverpool CGR
89093	2008	UK	Liverpool	CF	Liverpool CGR
89568	2008	UK	Liverpool	CF	Liverpool CGR
89224	2008	UK	Liverpool	CF	Liverpool CGR
89221	2008	UK	Liverpool	CF	Liverpool CGR
89150	2008	UK	Liverpool	CF	Liverpool CGR
89212	2008	UK	Liverpool	CF	Liverpool CGR
89204	2008	UK	Liverpool	CF	Liverpool CGR

89153	2008	UK	Liverpool	CF	Liverpool CGR
CF3S1_21	2009	UK	Liverpool	CF	Liverpool CGR
10333	2004	UK	Nottingham	CF	Liverpool CGR
10262	2004	UK	Nottingham	CF	Liverpool CGR
11112	2004	UK	Birmingham	CF	Liverpool CGR
8835	2004	UK	Wrexham	CF	Liverpool CGR
9857	2004	UK	London	CF	Liverpool CGR
8995	2004	UK	Sheffield	CF	Liverpool CGR
9920	2004	UK	Clatterbridge	CF	Liverpool CGR
417GLY	2000	UK	Liverpool	Non-CF	Liverpool CGR
B9	1995	UK	Liverpool	CF	Liverpool CGR
CF4S1_08	2009	UK	Liverpool	CF	Williams *et al.*
OSIRIS	2007	UK	Liverpool	Pet cat	Liverpool CGR
119579	2013	UK	Liverpool	PCD	Liverpool CGR
139124	2013	UK	Liverpool	PCD	Liverpool CGR
139125	2013	UK	Liverpool	PCD	Liverpool CGR
79426	2007	UK	Liverpool	Empyema	Liverpool CGR
79427	2007	UK	Liverpool	Empyema	Liverpool CGR
PHE_LES1	2010	UK	Scotland	CF	Liverpool CGR
PHE_LES2	2013	UK	London	CF	Liverpool CGR
B35	1998	UK	Liverpool	CF	Liverpool CGR
CF4S1_32	2009	UK	Liverpool	CF	Liverpool CGR
PHE_LES3	2012	UK	East Midlands	CF	Liverpool CGR
PHE_LES4	2013	UK	London	CF	Liverpool CGR
PHE_LES5	2014	UK	West Midlands	CF	Liverpool CGR
PHE_LES6	2011	UK	East Midlands	CF	Liverpool CGR
PHE_LES7	2014	UK	West Midlands	CF	Liverpool CGR
PHE_LES8	2011	UK	North East	CF	Liverpool CGR
PHE_LES9	2011	UK	South East	CF	Liverpool CGR
PHE_LES10	2014	UK	East	CF	Liverpool CGR
PHE_LES11	2013	UK	Yorkshire	CF	Liverpool CGR
PHE_LES12	2013	UK	London	CF	Liverpool CGR
49388	2004	UK	Liverpool	CF	Liverpool CGR
CF5S1_04	2009	UK	Liverpool	CF	Williams *et al.*
PHE_LES13	2013	UK	Wales	CF	Liverpool CGR
PHE_LES14	2012	UK	North West	CF	Liverpool CGR
PHE_LES15	2012	UK	South West	CF	Liverpool CGR
PHE_LES16	2014	UK	East	CF	Liverpool CGR
PHE_LES17	2012	UK	North West	CF	Liverpool CGR
PHE_LES18	2012	UK	London	CF	Liverpool CGR
PHE_LES19	2012	UK	London	non-CF	Liverpool CGR
PHE_LES20	2012	UK	North East	CF	Liverpool CGR
PHE_LES21	2013	UK	East	CF	Liverpool CGR
PHE_LES22	2013	UK	East Midlands	Bronchiectasis	Liverpool CGR
30642g	2003	UK	Liverpool	CF	Liverpool CGR

CF5S1_20	2009	UK	Liverpool	CF	Williams *et al.*
PHE_LES23	2013	UK	London	CF	Liverpool CGR
PHE_LES24	2013	UK	London	CF	Liverpool CGR
PHE_LES25	2013	UK	Wales	CF	Liverpool CGR
PHE_LES26	2013	UK	South East	CF	Liverpool CGR
PHE_LES27	2013	UK	West Midlands		Liverpool CGR
PHE_LES28	2013	UK	East	CF	Liverpool CGR
PHE_LES29	2013	UK	West Midlands		Liverpool CGR
PHE_LES30		UK	Liverpool		Liverpool CGR
89192	2008	UK	Liverpool	CF	Liverpool CGR
CF7S1_10	2009	UK	Liverpool	CF	Williams *et al.*
CF9S1_24	2009	UK	Liverpool	CF	Williams *et al.*
120MO	2003	UK	Liverpool	CF	Liverpool CGR
CF7S1_15	2009	UK	Liverpool	CF	Williams *et al.*
49513	2004	UK	Liverpool	CF	Liverpool CGR
CF8S1_02	2009	UK	Liverpool	CF	Williams *et al.*
CF177_P138		UK	Birmingham	CF	Birmingham Nick Loman
CF54_NCTC13415		UK	Birmingham		Birmingham Nick Loman
CF60_P44		UK	Birmingham	CF	Birmingham Nick Loman
Md1_LES		UK	Birmingham	CF	Birmingham Nick Loman
CF7S1_39	2009	UK	Liverpool	CF	Williams *et al.*
CF8S1_33	2009	UK	Liverpool	CF	Williams *et al.*
CF8S1_36	2009	UK	Liverpool	CF	Williams *et al.*
CF9S1_11	2009	UK	Liverpool	CF	Williams *et al.*
CF10S1_11	2009	UK	Liverpool	CF	Williams *et al.*
CF10S1_21	2009	UK	Liverpool	CF	Williams *et al.*
LESB58_2	1988	UK	Liverpool	CF	Liverpool CGR
CF1S1_20	2009	UK	Liverpool	CF	Dettman *et al.*
JD304		Canada	Ottawa	CF	Dettman *et al.*
JD310		Canada	Ottawa	CF	Dettman *et al.*
JD315		Canada	Hamilton	CF	Dettman *et al.*
JD322		Canada	Toronto	CF	Dettman *et al.*
JD324		Canada	Toronto	CF	Dettman *et al.*
JD326		Canada	Hamilton	CF	Dettman *et al.*
JD329		Canada	London	CF	Dettman *et al.*
JD335		Canada	Toronto	CF	Dettman *et al.*
LES400	1998	UK	Liverpool	CF	Jeukens et al
LES431	2000	UK	Liverpool	Non-CF	Jeukens et al
LESB58	1988	UK	Liverpool	CF	Jeukens et al
LESB65	2003	UK	Liverpool	CF	Jeukens et al
LESlike_1	2005	Canada	Ottawa	CF	Williams *et al.*
LESlike_4	2005	Canada	Toronto	CF	Williams *et al.*
LESlike_5	2007	Canada	Toronto	CF	Williams *et al.*

LESlike_7	2006	Canada	Hamilton	CF	Williams *et al.*
CF1S1_33	2009	UK	Liverpool	CF	Williams *et al.*
CF3S1_12	2009	UK	Liverpool	CF	Williams *et al.*
CF3S1_24	2009	UK	Liverpool	CF	Williams *et al.*
CF4S1_2	2009	UK	Liverpool	CF	Williams *et al.*
CF4S1_26	2009	UK	Liverpool	CF	Williams *et al.*
CF5S1_34	2009	UK	Liverpool	CF	Williams *et al.*
CF5S1_38	2009	UK	Liverpool	CF	Williams *et al.*
CF6S1_1	2009	UK	Liverpool	CF	Williams *et al.*
CF6S1_18	2009	UK	Liverpool	CF	Williams *et al.*
CF7S1_36	2009	UK	Liverpool	CF	Williams *et al.*
CF3S2_19		UK	Liverpool	CF	Williams *et al.*
CF3S2_25		UK	Liverpool	CF	Williams *et al.*
CF3S2_35		UK	Liverpool	CF	Williams *et al.*
CF3S2_24		UK	Liverpool	CF	Williams *et al.*
CF3S2_31		UK	Liverpool	CF	Williams *et al.*
CF3S2_5		UK	Liverpool	CF	Williams *et al.*
LiP9		UK	East Midlands	CF	Quebec
LiP13		UK	East Midlands	CF	Quebec
LiP10a		UK	East Midlands	CF	Quebec
LiP10b		UK	East Midlands	CF	Quebec
LiP8		UK	East Midlands	CF	Quebec
LiP1a		UK	East Midlands	CF	Quebec
LiP1c		UK	East Midlands	CF	Quebec
LiP2c		UK	East Midlands	CF	Quebec
LiP2a		UK	East Midlands	CF	Quebec
LiP2b		UK	East Midlands	CF	Quebec
LiP11		UK	East Midlands	CF	Quebec
LiP14		UK	East Midlands	CF	Quebec
LiP12		UK	East Midlands	CF	Quebec
LiP5		UK	East Midlands	CF	Quebec
LiP7		UK	East Midlands	CF	Quebec
SED11		UK	East	CF	Darch *et al.*
SED1		UK	East	CF	Darch *et al.*
SED20		UK	East	CF	Darch *et al.*

4.3.1. Genome Assembly quality

Genome sizes ranged from 6,256,335bp (PHELES12) to 6,764,430bp (PHELES27), a range of 508,095bp and mean overall of 6,473,573.9bp. The GC content of the genomes ranged from 65.22% (PHELES27) to 66.54% (JD315), a range of 1.32% and a mean of 65.92%.

To assess genome assembly quality, contig length based metrics, N50, number of contigs and largest contig were used. LESB58 and LES431 are complete genomes so were excluded from this comparison. N50 ranged from 5342 (JD324) to 4,752,040 (LES400) with a mean of 428722.17. Number of contigs (above 1000bp) ranged from 1488 (JD324) to 3 (LESB65) (Table 4.3).

Table. 4.3. Ranges for genome contig length based assembly quality metrics, genome size and GC% content

Genome size	6256335-6764430
GC content%	65.22-66.54
Number of contigs*	3-1488
N50	5342-4,752,040

*contigs >1000bp counted

4.3.2. Core genome SNP phylogeny

The core genome was 3,761,492bp and contained 972 phylogenetically informative sites. An unrooted maximum likelihood phylogeny was approximated from the 143 taxa LES core genome resulting in 653 splits using the TVM+R3 substitution model (IQ-TREE-1.5.6) selected by ModelFinder with 1000 ultrafast bootstraps (supplementary figure 4). Two major clades containing genomes of isolates primarily from Canada and the second with genomes exclusively isolated from the UK with LESB58 and LESB58_2 as an outgroup to both. Gubbins was used, using the LES core genome alignment and ML phylogenetic tree as a starting tree, to approximate a core genome phylogeny with branch lengths and topology corrected for the presence of recombinant SNPs (the clonal frame) (figure 4.1). Recombinant SNPs were predicted to account for 2.87% of all core genome SNPs. Supplementary figure 5 displays the core genome tree as a cladogram with branch lengths displayed. Supplementary figure 6 displays recombinant SNPs based on their position in the core genome alignment alongside the tree. The core genome was also extracted for the 143 LES taxa and non-LES outgroup PAO1 (supplementary figure 5), confirming LESB58 as an outgroup to the UK LES for further analysis.

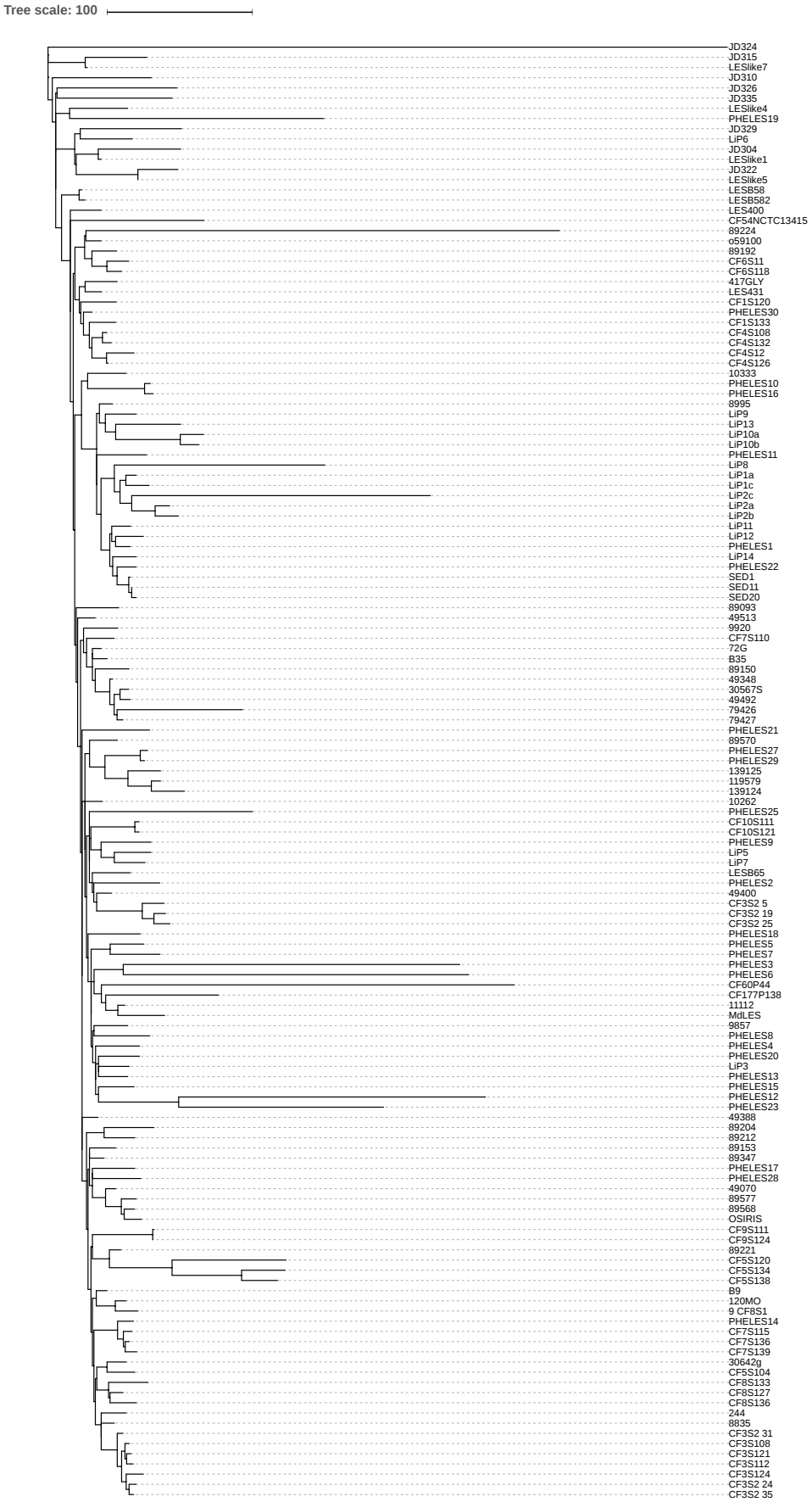


Figure 4.1. LES core genome phylogeny (n = 143). Maximum likelihood approximation with IQ-TREE-1.5.6 using the TVM+R3 substitution model and 1000 ultrafast bootstraps

In addition to phylogenetic clustering of isolate genomes from Canada and the UK respectively, there is clustering of genomes within the UK clade (figure 4.2). However, placement in the core SNP tree is not consistent with source of isolation in every case. For all inconsistencies where further patient data was available, the patients had previously lived or at least presented a sample at the location with which their isolate genome clusters on the tree. All UK locations referred to are Public Health England (PHE) designated regions.

PHELES10 and PHELES16 are isolates provided from patients in the East of England in 2014 that cluster instead with isolates from the East Midlands. Both patients presented samples prior to 2014 in the East Midlands CF Unit. The patient that provided PHELES11 in Scotland, that also clusters with genomes from the East Midlands previously provided a *P. aeruginosa* sample in the East Midlands with the LES associated serotype (O6). The greatest discrepancies between locations of isolation and genomic epidemiology were for isolates LiP6 and PHELES19 clustering with the isolates in Canada. Further investigation by colleagues at PHE revealed that LiP6 was isolated from a patient in the East Midlands that had previously lived in Canada.

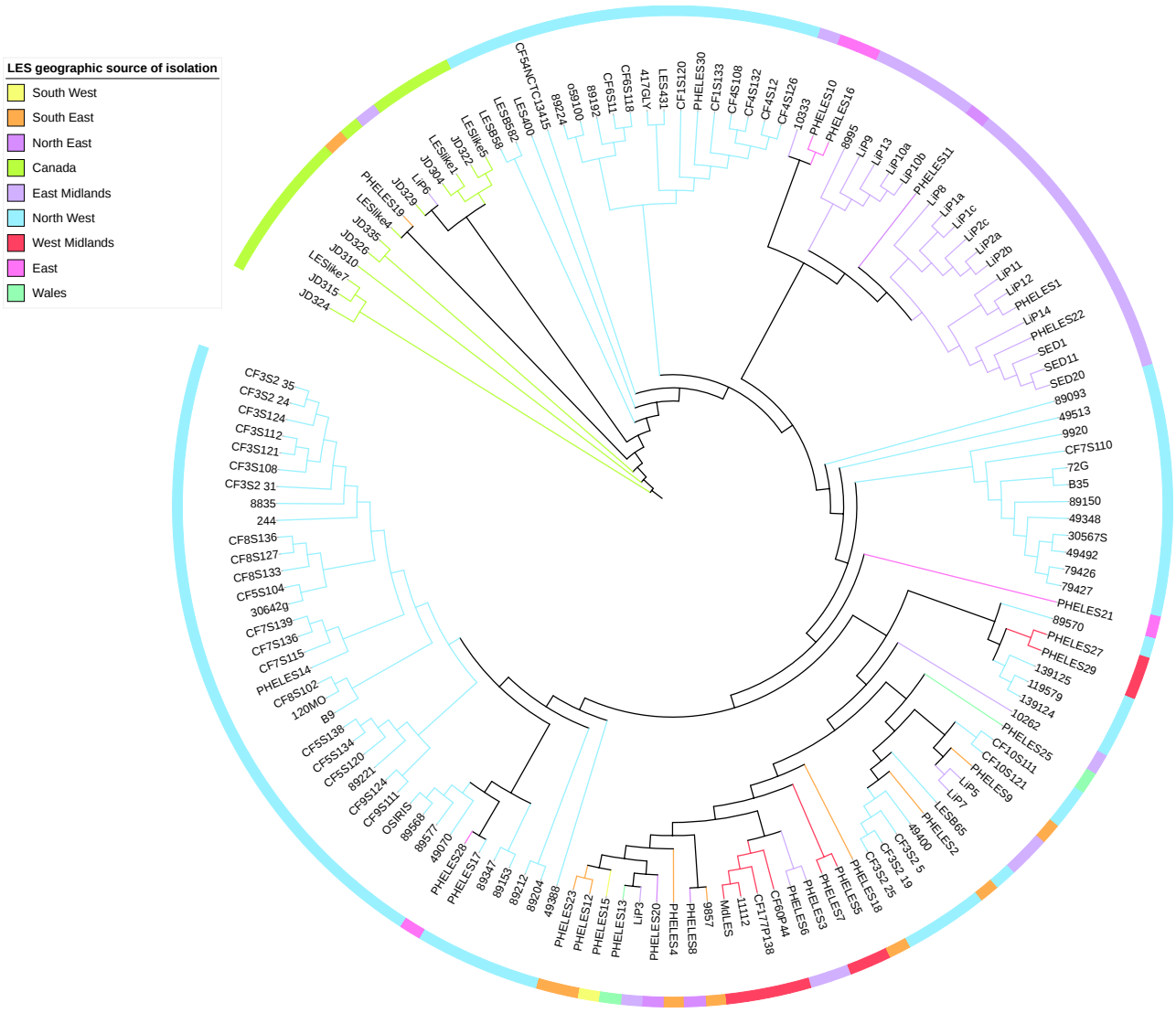
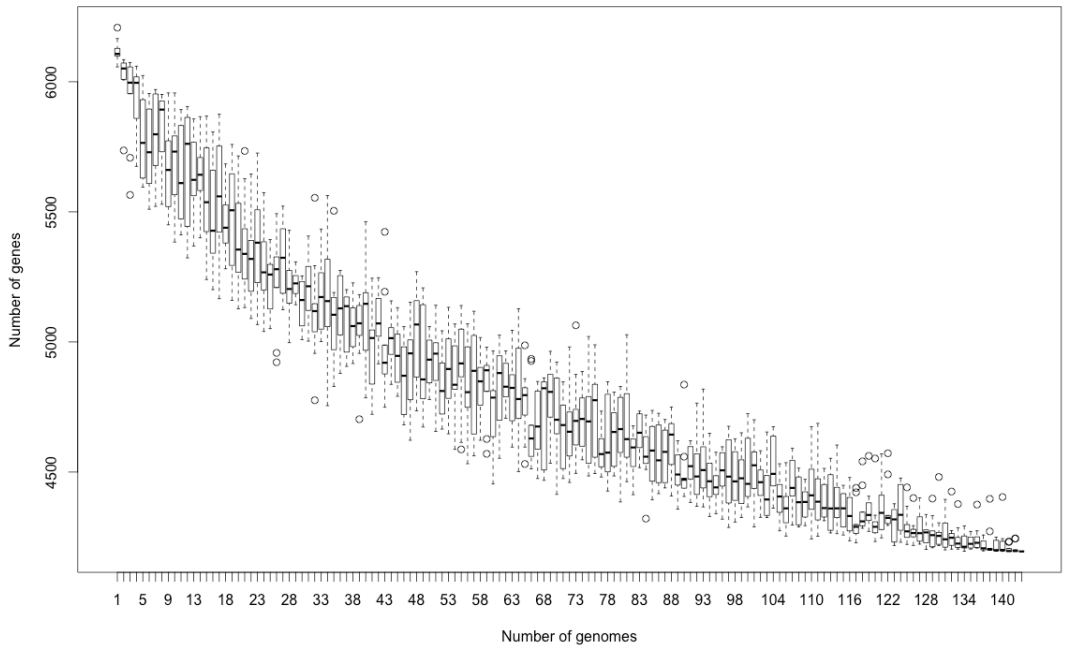


Figure 4.2. LES core genome (n = 143 genomes) RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins. Colour strip and branches coloured by PHE regions according to geographic source of isolation

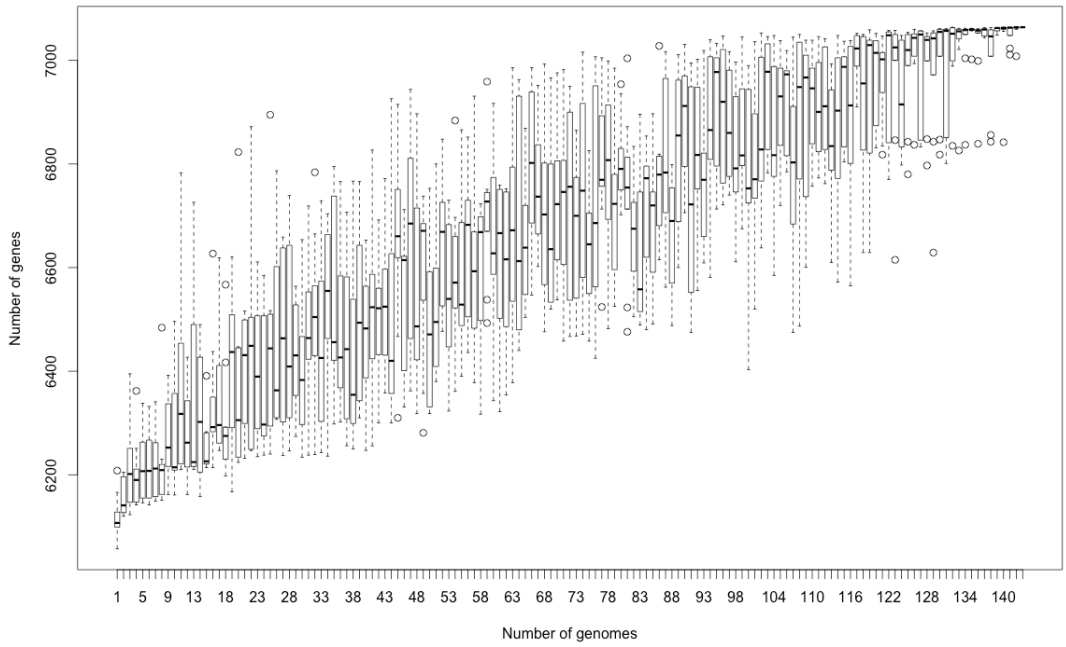
4.3.3. LES Pangenome

A matrix of gene content in the LES pangenome was determined. The pangenome matrix was sampled at various depths per genome (n=143) for 100 iterations and plotted for core and accumulated genes (figure 4.4). Figure 4.4a shows the conservation of genes within the LES. The trend of accumulated genes (figure 4.4b) indicates that the LES has an open pan-genome. The number of genes in the core genome was 4195 while 466 were unique. The LES core genome therefore accounts for 90% of the pangenome and an average of 3.26 new genes are added with each new genome sequenced.

Pan genome matrix values were designated as group1 if they belonged to LES phylogenetic group 1 (figure 4.2a) or group 2 if they belonged to LES phylogenetic group 2 (figure 4.2b). There were no set differences observed between the groups in presence or absence of genes and similarly no set difference following exclusion of LESB58 and re-sequenced LESB58 (LESB58_2) are an outgroup to both groups.



a.



b.

Figure 4.3. Pangenome matrix randomly sampled at various depths for 100 iterations. (a) shows the core genome and (b) the accumulation of genes overall. On the x axis number of genes added and on the y axis number of genomes added

4.3.4. Distribution of LESB58 Prophage and Genomic Islands amongst the wider LES collection

The reference genome LESB58 carries 5 genomic islands and 6 prophages; these were each aligned to each of the 143 LES genomes to assess percent identity. Figure 4.4 presents a heatmap of the percent identity for each genomic island or prophage. In the majority of genomes LESB58 genomic islands and prophages are mostly present but also vary from complete or near complete absence to 100% identity. Prophage 5 is absent in the majority of genomes, percent identity 0-14% in 124/143 genomes. Prophage 5 is present at >99% in LES431 and LES400.

Table 4.4. Mean and range of percent identity of blast alignment with LESB58 genomic islands and prophages across the LES genomes (n=143)

LESB58 region	Mean percent identity	Range of percent identity
LESGI1	91.9	36.6-100
LESGI2	99.89	89.14-100
LESGI3	98.51	0-100
LESGI4	85.72	6.98-100
LESGI5	65.03	0-100
prophage2	70.78	0-100
prophage3	71.23	0-100
prophage4	98.94	0-100
prophage5	12.35	0-100
prophage6	94.91	0-100

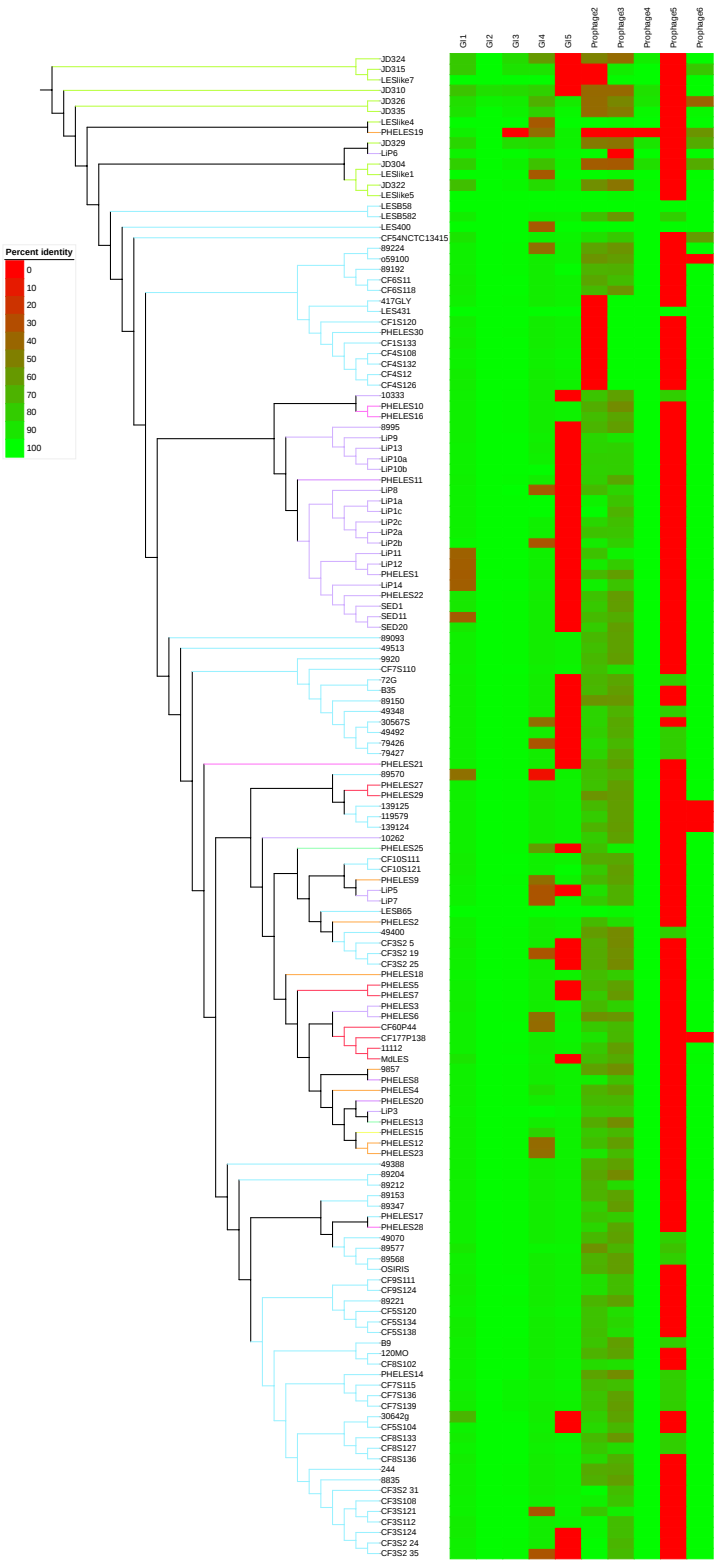


Figure 4.4. LES core genome (n = 143 genomes) RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins. Colour strip and branches coloured by PHE regions according to geographic source of isolation. Alongside the tree is a heatmap for each corresponding genome of LESB58 genomic island and prophage percent identity

4.3.5. Mutational Diversity within the UK LES Genomes

To analyse the mutational diversity across the UK LES genomes, those with short reads available (n=122) (table 4.2) were mapped to reference genome LESB58, variants called and annotated. Mutations were classified as synonymous or non-syonymous. Non-synonymous mutations were further classified according to whether they were predicted to be nonsense mutations, missense mutations (causing a codon change and subsequent amino acid substitution), in-frame insertions and deletions and upstream or downstream modifiers. The genome with the largest number of mutations compared with LESB58 was that of isolate CF9S1_24 with 1202, whilst the genome with the least was B9 with 100. In total, 8 isolate genomes contained >500 mutations, 43 contained >200 but <500 and the remaining 70 genomes excluding B9 contained between 100 and 200 mutations. The mean number of mutations per genome relative to LESB58 was 223.

4.3.6. Nonsense Mutations

There were a total of 1812 nonsense mutations in the 122 genomes analysed, 1347 frame-shifting indels and 465 premature stop codons. There are 24 genes with nonsense mutations carried in 10 or more genomes in the population and as many as carrying a nonsense mutation in 120 genomes (table 4.4.). Functionally, these genes are involved in small molecule transport (PLES_56861, *mexB*, *glpT*, PALES_34011, *oprD*), the cell wall/LPS/capsule (*mucA*, *mpl, mraW,* PALES_46251, *mltD*) in addition to redox (PLES_20241), motility and attachment (*fleR*), carbon compound metabolism and hypothetical

genes of unknown function. The LES is characteristically non-motile as a result of loss of flagellin encoding gene *fleR*. In this study 120 of the 122 genomes analysed carried the same nonsense mutation in *fleR*. The genes PLES_46401 and PLES_47271 encoding hypothetical proteins also present with nonsense mutations in >50% of the genomes. Additionally anti sigma factor *mucA* has a nonsense mutation in 23 isolate genomes.

In addition to nonsense mutations, mapped reads were assessed for larger gene deletions and absences. There were 892 different genes that appear to have lost function in at least 1 genome. Of the 892 genes with any loss of function 293 appear to have lost function in only 1 genome. There are 69 different genes with loss of function in >50% of the genomes. Motility gene *fleR* has a premature stop codon in 120/122 genomes analysed and a 28bp deletion, spanning the same region in a further genome, that of isolate 8835. There is no evidence for loss of function in *fleR* in a single isolate genome, that of LiP1a.

Table 4.5. Genes across the UK LES genomes carrying a nonsense mutation in ten or more alleles

Gene name	Number of genomes	Gene description
fleR	120	Flagellar motility gene
PLES_46401	96	Hypothetical protein
PLES_47271	84	Hypothetical protein
oprD	49	Imipenem outer membrane porin
mltD	44	Murein transglycosylase D precursor
PLES_56861	27	Putative choline transporter
mexB	25	Multidrug efflux pump
lasR	24	Transcriptional regulator of quorum sensing
mucA	23	Anti-sigma factor for alginate biosynthesis
PLES_59241	20	Hypothetical protein
PLES_26281	20	Hydrophobe/amiphile efflux
PLES_07981	19	Hypothetical protein
PLES_50281	15	Hypothetical protein
PLES_48481	13	Hypothetical protein
glpT	13	Sn-glycerol-3-phosphate transporter
ambB	13	AMB* biosynthesis
pqqF	12	pyrroloquinoline quinone biosynthesis protein F
exsA	12	Transcriptional regulator of type3 secretion
mraW	10	S-adenosyl-methyltransferase
lpxO2	10	Lipopolysaccaride biosynthesis
PLES_46251	10	Glycosyl transferase
mpl	10	Inactivation leads to overproduction of beta-lactamases
PLES_20241	10	Putative oxidase
PLES_34011	10	TonB dependent receptor

*Antimetabolite L-2-Amino-4-methoxy-trans-3-butenoic Acid

To determine which functions the genes most often carrying loss of function genes were associated with their pseudoCAP categories, summarised in figure 4.5. The categories are displayed as associations to genes that have evidence for loss of function as a function of the number of genomes harbouring disrupted alleles. The majority of associated functions are unknown, hypothetical genes (71.3%). The proportions of associated categories without considering hypothetical genes are summarised in figure 4.8. The category associated with the greatest proportion is 'non-coding RNA genes' (16.25%) followed by 'transport of small molecules' (14.07%) and 'transcriptional regulators' (13.54%). Of the proportion of associated functional categories that are transcriptional regulators, two-component systems represent 37.7%.

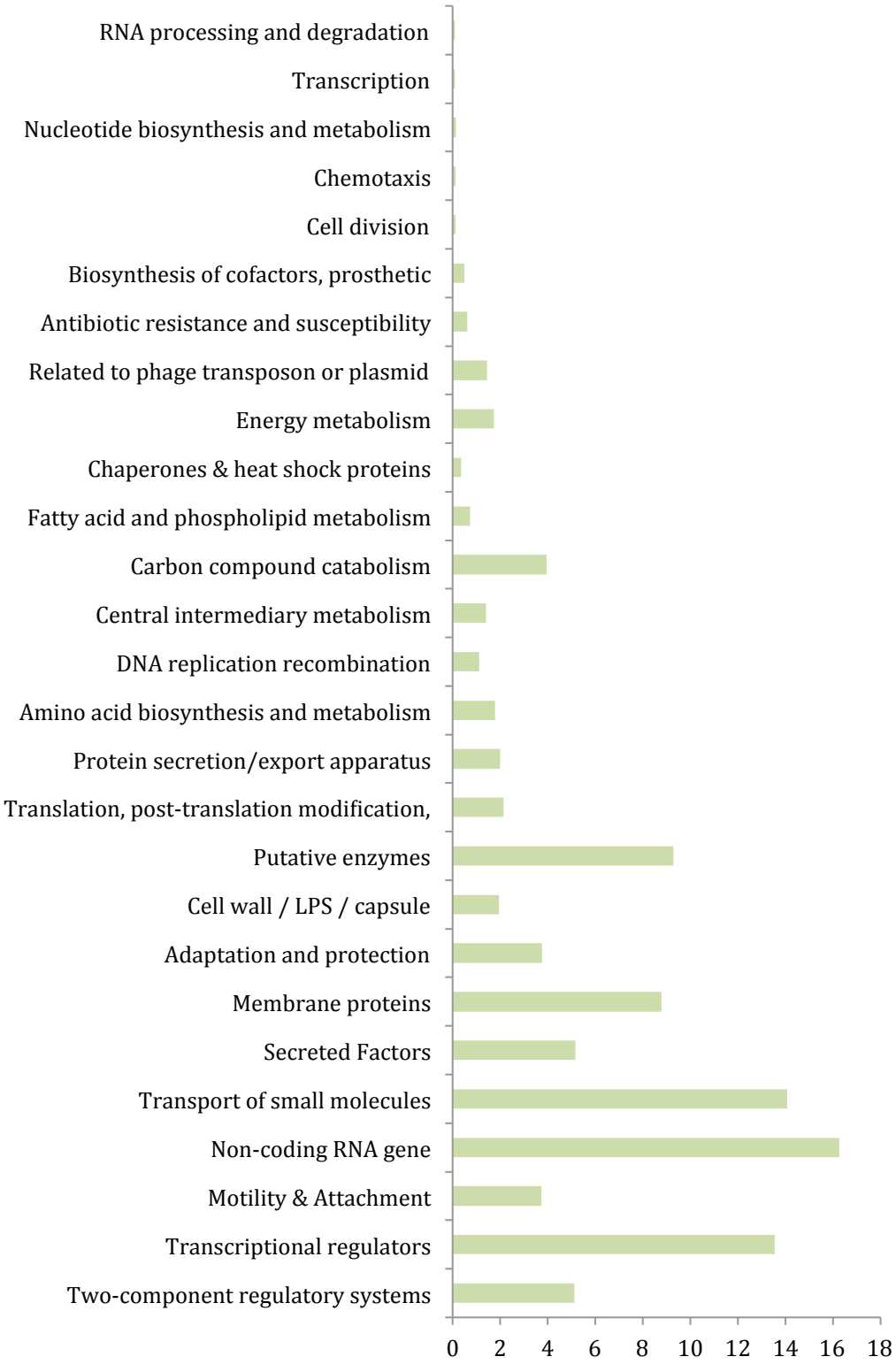


Figure 4.5. PseudoCAP functional categories associated with genes carrying nonsense mutations in ten or more of the UK LES genomes

4.3.6.1. Independent acquisition of loss of function

To identify genes that may be adapting by loss of function in the UK population, the independent acquisition of nonsense mutations was inferred by number of unique nonsense mutations per gene in the population. Where a mutation was the same by position they were inferred to have been inherited. In genes in the population where per genome there is a combination of shared and unique nonsense mutations it was parsimoniously inferred that the shared positions precede the uniques. 600 different genes in the population were inferred to have lost function by acquisition of a nonsense mutation. The majority of genes affected (347/600) present a nonsense mutation in only one isolate genome. As such these represent 347/500 genes for which there is only one independent acquisition. The remaining 153 genes with a mutation in two genomes or more range from being maintained in 2-120 of 122 UK LES genomes. Table 4.6 displays this distribution.

Table 4.6. The number of different genes that have independently acquired loss of function the same amount of times in the population

Number of independent acquisitions	Number of genes
30	1
18	1
15	1
12	1
10	1
8	2
7	1
6	1
5	4
4	5
3	11
2	71
1	500

The number of genomes in which these genes carried a nonsense mutation also vary and was not associated with the number of acquisitions. Table 4.6 displays the genes that have most often independently acquired nonsense mutations (4 or more acquisitions). The number of acquisitions of nonsense mutations per gene ranges from 4 to 30 across 16 different genes that have acquired nonsense mutations 4 or more times. The number of genomes in which nonsense mutations are present in these genes range from PLES_16021 with 5 independent acquisitions in 5 genomes, to PLES_46401 (with 4 acquisitions maintained in 96/122 genomes). The gene most often independently mutated *oprD,* was observed to carry a nonsense mutation in 49/122 genomes. The genes are associated with functions such as antibiotic resistance, (*mpl, mexB, oprD,* PALES_16021, *glpT*), quorum sensing (*lasR*), mucoidy (*mucA*), iron acquisition (PLES_34011*, pchE*). There were also hypothetical genes with unknown functions, PLES_59241, PLES_48481 and PLES_46401.

Table 4.7. Genes that have independently acquired a loss of function mutation =>4 times in the UK LES population

Independent acquisitions	Gene name or LESB58 locus tag	Genomes in which these mutations are observed	Gene description
30	*oprD*	49	Imipenem outer membrane porin
18	*lasR*	24	Transcriptional regulator of quorum sensing
12	PLES_59241	20	Hypothetical membrane protein
15	*mexB*	25	Multidrug efflux pump transporter, part of the *mexAB-oprM* operon
10	*mucA*	23	Anti-sigma factor for alginate biosynthesis
8	*mltD*	44	Murein transglycosylase D precursor
8	*glpT*	13	sn-glycerol-3-phosphate transporter
7	*pncA*	9	nicotinamidase
6	PLES_34011	10	tonB dependent receptor
5	PLES_48481	13	Hypothetical protein
5	PLES_46251	10	Putative glycosyl transferase
5	*opmQ*	7	Putative outer membrane protein precursor
5	PLES_16021	5	Transcriptional regulator (*marR* ortholog, involved in repression of MDR mar operon)
4	PLES_46401	96	Hypothetical protein with filamentous haemagglutinin outer membrane domain)
4	*mpl*	10	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase
4	PLES_03151	9	Transcriptional regulator
4	PLES_47771	8	Two-component sensor
4	*pchE*	7	dihydroaeruginoic acid synthase

The porin gene *oprD* had independently lost function more than any other gene in the population. 49 genomes contain nonsense mutations in *oprD* due to at least 30 independent events. In total there is evidence for inactivation of *oprD* in 55 genomes; an additional 6 genomes contain larger deletions ranging from 141bp to 647bp of the 1,326bp gene with an absence of mapped reads. Small insertions ranging 1-30bp comprised the majority of instances of loss of function in the gene. Loss of function was achieved by frame-shifting insertions in 12 of 30 independent events and maintained in 24 of 55 genomes. Five independent acquisitions maintained in 16 genomes were by substitution resulting in a premature stop codon. Finally, there were 7 occurrences of single base pair frame shifting deletions, resulting in loss of function in 9/55 genomes.

In 26 of 122 genomes there was a loss of function mutation by nonsense mutations in the gene encoding an anti-sigma factor for alginate production, *mucA*. Loss of function in MucA is known to be associated with the mucoidy phenotype by subsequent overproduction of alginate. No genetic evidence for compensatory loss of function mutations in alginate biosynthesis genes was observed.

Loss of function mutations were observed in *mexB* in 26 genomes, among the highest number of genomes for any of the genes carrying loss of function mutations. Additionally Loss of function was observed to have been acquired 15 times independently (table 4.6). Loss of function in the efflux pump can also occur via mutations in *mexA* or *oprM*. Combined evidence of nonsense mutations demonstrates evidence for loss of function in the operon in 32 isolate

genomes. For 31 of these genomes loss of function is evidenced by nonsense mutations (6 in *mexB* and 5 in *mexA*), rather than larger deletions or absences and as such is possible to infer how often the operon has independently lost function. Overall the operon has independently acquired loss of function mutations at least 21 times.

Quorum sensing regulator *lasR* was found to harbor nonsense mutations in 24 genomes, acquired 18 times in the population. Further there is complete absence of *lasR* in isolates 10262, CF5S1_04 and Md1-264 in addition to a 31bp deletion in isolate 244. In total there were 28 of 122 with loss of function in *lasR*.

Of the 122 UK LES genomes analysed, 8 genomes were identified as having evidence of being hypermutable by the presence of nonsense mutations in DNA mismatch repair genes: *mutS, mutM, mutL* or *uvrD* (table 4.8). Some of these mutations appear to be shared in closely related genomes. LiP10a and LiP10b (from the same patient) and LiP13 (from a different patient) contain the same 4bp frame-shifting deletion in *mutM*. PHELES3 shares with PHELES6 57bp of the 58bp deletion in *mutS*, but PHELES6 that has an additional 1bp frame-shifting indel in *mutL*.

Table 4.8. Nonsense mutations in DNA mismatch repair genes, *mutS, mutL, mutM, uvrD*

Isolate	Genetic evidence for hypermutability
LiP2c	*mutS* 1bp insertion frameshift
79426	*mutS* premature stop codon gained
LiP10a	*mutM* 4bp deletion frameshift
LiP10b	*mutM* 4bp deletion frameshift
LiP13	*mutM* 4bp deletion frameshift
89224	*mutL* 5bp insertion frameshift
PHELES6	*mutS* 57bp deletion & *mutL* 1bp deletion frameshift
PHELES3	*mutS* 58bp deletion

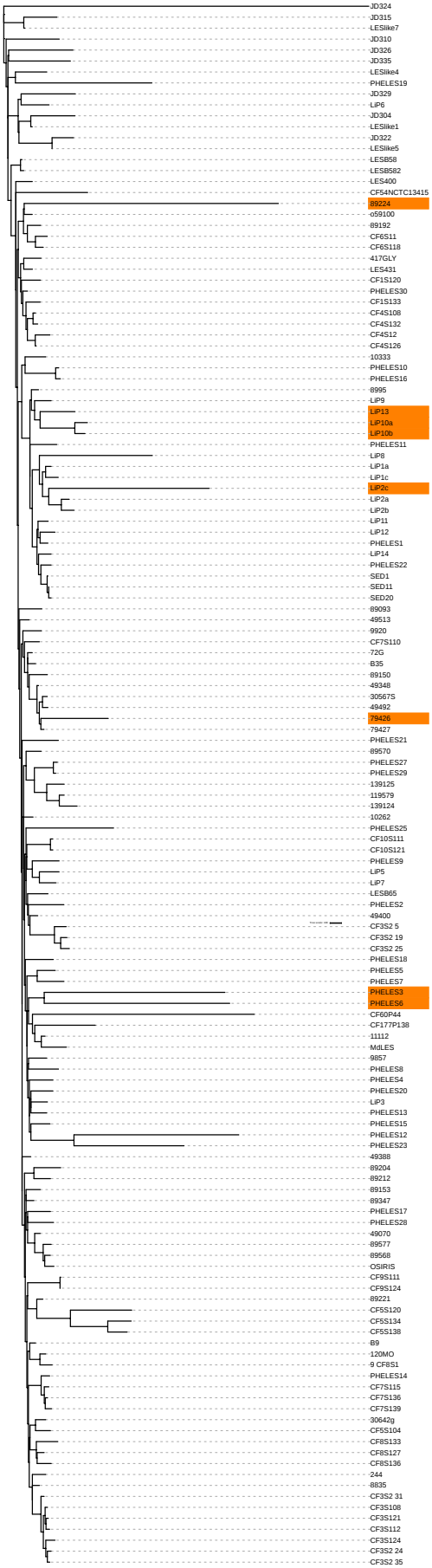


Figure 4.6. LES core genome phylogeny (figure 4.1). Genomes with genetic evidence for DNA mismatch repair deficiencies labeled in orange

In total the 18 genes most often independently acquiring nonsense mutations are associated with 12 different pseudoCAP categories. For each category the number of associations is a function of the number of genomes in which that gene has a nonsense mutation. Each gene may also be associated with multiple categories. In total there are 593 associations to functional categories (figure 4.7). The function most often associated with the genes most often independently acquiring nonsense mutations, as a percentage of all functional associations, is that of unknown or hypothetical (21.75%), followed by transport of small molecules (18.71%) and cell wall/LPS/capsule (12.98%).

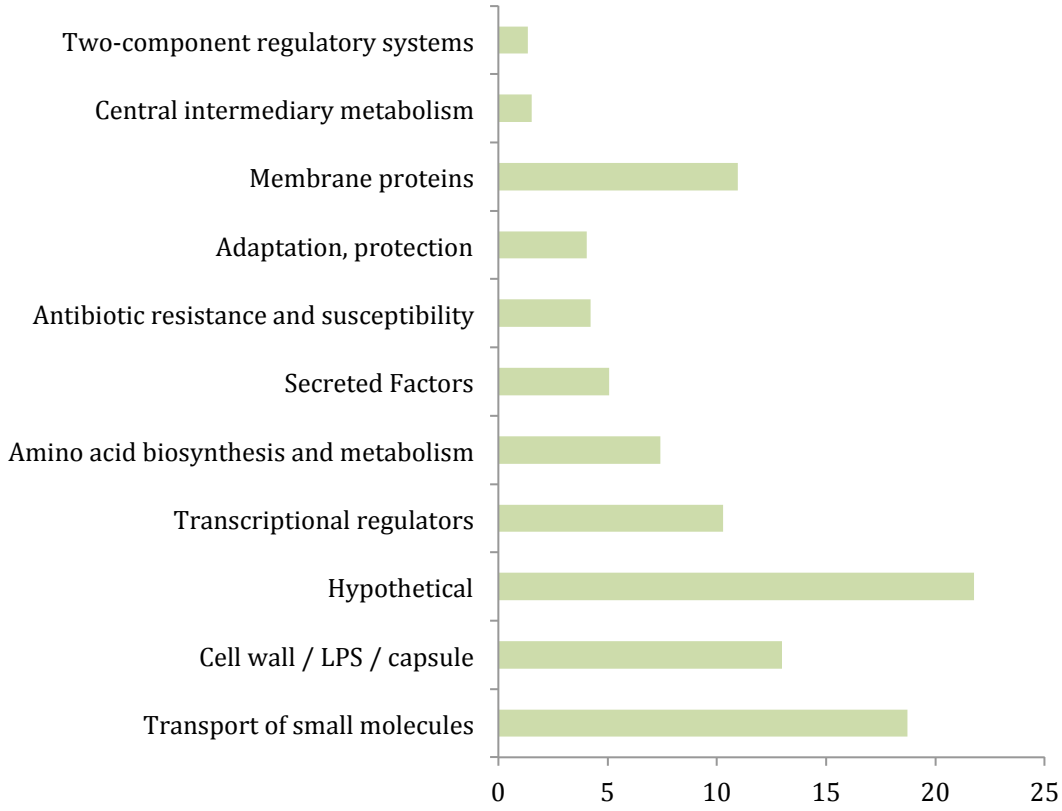


Figure 4.7. PseudoCAP functional categories associated with the genes most often independently acquiring nonsense mutations as a function of the number of genomes in which a loss of function allele is carried

4.3.7. Missense Mutations

To assess missense mutations that might be adaptive in the UK LES population, genes with loss of function were not considered. Across the remaining genes there were 13,691 synonymous and non-syonymous mutations. Table 4.9 shows how many of these mutations are in-frame insertions or deletions, synonymous mutations, upstream and downstream modifiers of coding regions and those that are missense mutations. In genes without evidence for loss of function, non-synonymous changes were greater in number than those that are synonymous.

Table 4.9. Count of mutations by functional impact in the UK LES population

Predicted functional impact of mutation	Number of SNPs compared with LESB58
Synonymous	3212
Upstream modifier	1251
Downstream modifier	24
In frame insertion	167
In frame deletion	141
Missense	8895
Total	13690

There are 2427 different genes with at least one missense mutation compared with LESB58. Of these 60% (1478 genes) have a single missense variation, which may be present in a varying number of genomes (table 4.9).

Table 4.10. The number of unique missense mutations per gene in the population and the number of different genes with that many unique missense mutations

Number of unique missense mutations	Number of genes
20	1
19	1
17	1
15	1
14	2
13	1
12	3
10	5
9	1
8	7
7	12
6	12
5	31
4	76
3	229
2	566
1	1478

For all genes with at least one missense mutation the associations to pseudoCAP functional categories were determined as a percent of the overall categories associated (figure 4.9). The highest proportion of associations was those of unknown function due to hypothetical genes (20.28%), followed by transport of small molecules (10.96%), membrane proteins (9.37%) and putative enzymes (9.31%).

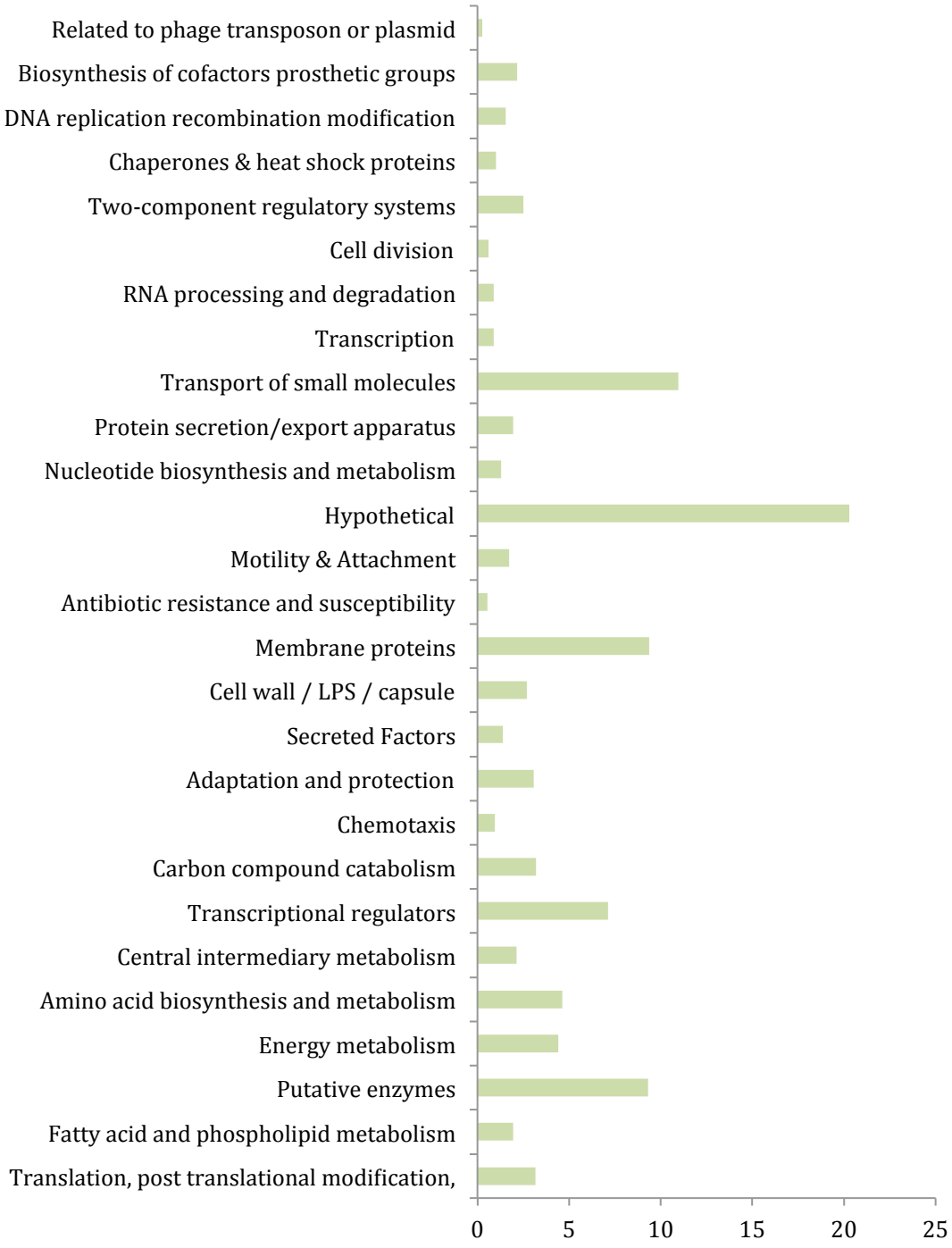


Figure 4.8. PseudoCAP functional categories associated with genes with one or more missense mutations across the UK LES genomes

There are 2,428 genes with at least one missense mutation in the UK LES genomes. Missense mutations range from being present in only one genome to fixation in the UK population compared with LESB58. Genes with a missense mutation in >100 genomes are presented in table 4.11. There were 11 different genes with a missense mutation fixed in the population.

Table 4.11. The genes in which a non-syonymous (missense) mutation is shared in >100 isolates across the UK LES genomes. Fixed mutations (in all 122 genomes) in bold

Gene name or locus tag	Number of genomes	Gene function
gyrB	**122**	**DNA gyrase subunit B**
PLES_28161	**122**	**putative two-component sensor**
PLES_05421	**122**	**hypothetical protein**
PLES_45841	**122**	**hypothetical protein**
PLES_54891	**122**	**putative transporter**
PLES_31551	**122**	**hypothetical protein**
fis	**122**	**DNA-binding protein Fis**
PLES_15441	**122**	**putative mechanosensitive ion channel**
PLES_44361	**122**	**putative ring-cleaving dioxygenase**
PLES_48951	**122**	**putative hydroxylase**
PLES_13941	**122**	**putative amino acid permease**
PLES_52231	121	hypothetical protein
PLES_11151	121	carboxylesterase
nosR	120	regulatory protein NosR
PALES_44581	120	putative ATP-binding/permease fusion ABC transporter
metX	120	homoserine O-acetyltransferase
PLES_19231	120	putative secretion protein
PLES_48781	119	putative metallopeptidase
PLES_58511	119	hypothetical protein
PLES_30291	116	putative alcohol dehydrogenase (Zn-dependent)
PLES_01041	116	putative sulfate transporter
PLES_48791	115	putative transcriptional regulator
PLES_29121	102	putative ATP-binding/permease fusion ABC transporter

4.3.7.1. Genes with a diversity of missense mutations in the population

Genes that carry 10 or more different missense mutations in the population are displayed in table 4.12. They include genes that are associated with with quorum sensing (*lasR*), alginate biosynthesis (*algG),* antibiotic resistance (*pmrB, pmrA exsD mpl, ftsl, mexB*), a hypothetical protein (PLES_19111). Of the 14 genes, 5 are predicted to encode regulators.

Table 4.12. Genes with ten or more unique missense mutations in the population

Gene name or locus tag	Unique mutations	Number of genomes	Gene function
lasR	20	35	transcriptional regulator LasR
algG	19	31	alginate-c5-mannuronan-epimerase AlgG
pmrB	17	16	PmrB: two-component regulator system signal sensor kinase PmrB
spoT	15	20	guanosine-3'
PLES_19111	14	18	hypothetical protein
rpoD	14	17	RNA polymerase sigma factor RpoD
ftsI	13	85	penicillin-binding protein 3
gyrA	12	56	DNA gyrase subunit A
mpl	12	24	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl- meso-diaminopimelate ligase
mexB	12	16	Resistance-Nodulation-Cell Division (RND) multidrug efflux transporter MexB
exsA	10	33	transcriptional regulator ExsA
pmrA	10	17	PmrA: two-component regulator system response regulator PmrA
exsD	10	12	ExsD
rpoB	10	10	DNA-directed RNA polymerase subunit beta

Diversification of a gene in a population can be due to positive selection or drift, as such it the dN/dS ratios were calculated for the genes with the highest number of unique missense mutations. Alleles from genomes without missense or without mutations altogether were included in the multiple alignments. A dNdS ratio >1 can indicate positive selection, while a ratio of <1 can indicate purifying selection. Table 4.12 displays the genes with 10 or more unique missense mutations. The dN/dS ratio for the majority (10/14) revealed a ratio of >1 that indicates positive selection (table 4.13).

Table 4.13. Genes ten or more unqiue missense mutations in the population. Average dNdS ratios across codons calculated

Gene	Average dN/dS over all codons	Gene function
spoT	**1.442445966**	guanosine-3'
rpoD	**1.0721435**	RNA polymerase sigma factor RpoD
rpoB	0.957809229	DNA-directed RNA polymerase subunit beta
pmrB	**1.127049428**	PmrB: two-component regulator system signal sensor kinase PmrB
pmrA	**1.424881821**	PmrA: two-component regulator system response regulator PmrA
PLES_19111	0.87522896	hypothetical protein
mpl	0.982156966	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl- meso-diaminopimelate ligase
mexB	**1.428139835**	Resistance-Nodulation-Cell Division (RND) multidrug efflux transporter MexB
lasR	**1.393989122**	transcriptional regulator LasR
gyrA	0.156205563	DNA gyrase subunit A
ftsL	**1.49594221**	penicillin-binding protein 3
exsD	**1.468284208**	ExsD
exsA	**1.304668628**	transcriptional regulator ExsA
algG	**1.430029314**	alginate-c5-mannuronan-epimerase AlgG

4.4. Discussion

Previously a limited sample of LES genomes was available for comparative genomics[214,334] and studies focused on the within-host diversity and adaptation of LES lineages during chronic lung infections[352,280]. The panel of LES isolates for this study have greater geographic diversity, a larger sample size, are historically diverse and represent a diversity of associated pathologies than previous studies. Using this larger dataset it has been possible to assess the genomic epidemiology, pangenome structure and potential genetic basis of ongoing adaptation in greater detail.

The U.K. and Canadian LES form distinct clades with the exception of LiP6 and PHELES19, sourced in the U.K. but clustering phylogenetically with the majority Canadian clade. For LiP6 further patient information was available that they had previously lived in Canada, so is credible that this isolate was instead acquired there. Generally within the UK too genomes cluster according to their source. Upon rooting the tree with a non-LES outgroup, PAO1, the Canadian genomes cluster with the root of the LES tree and LESB58, the oldest known isolate, roots the U.K. genomes. Previously it wasn't possible to determine this ordering of Canadian and U.K. LES genomes as only a small set were available for comparison[334]. The Canadian clade further encompasses ST-146's (the majority of LES isolates) and ST-683's. The inner node confidence separating Canadian ST-683's and ST-146's is 94% (supplementary figure 5).

Deficiency in DNA mismatch repair genes *mutS, mutL, mutM* and *uvrD* have been observed in *P. aeruginosa* to result in the hypermutable phenotype[353]. The

benefits of rapid evolution in hypermutable lineages comes at the cost of an accumulation of disadvantageous mutations also[353]. Though in one study of the Danish transmissible lineage of *P. aeruginosa* as many as 48% of lineages harboured DNA mismatch repair deficiency mutations[282]. In this dataset 8/122 genomes are identified as potentially hypermutable due to DNA mismatch repair deficiency (table 4.8). There also evidence for cross infection of potentially hypermutable strains between the patient infected with LiP10a/LiP10b and the patient infected with LiP13 due to their adjacency in the core SNP tree (figure 4.6) and of *mutM* sharing the same 4bp deletion. There is also potential that PHELES3 and PHELES6 from different patients but adjacent in the core SNP tree share the same 58bp deletion where PHELES3 harbours a 57bp deletion otherwise that is otherwise the same. Hypermutator lineages of *P. aeruginosa* have been implicated in convergent accumulation of mutations associated with chronic CF lung infections[325], have previously been observed in the LES[280] and other transmissible strains[282].

The LES tonB-dependent receptor gene PLES_34011 was observed to have independently acquired nonsense mutations 6 times in the population, carried in 10 genomes. TonB-dependent receptor gene homolog (PA1922) disruptive mutations have previously been associated with chronic *P. aeruginosa* CF lung infections[354,355].

Loss of motility has been associated with chronic *P. aeruginosa* isolates[181] and a defining characteristic of the Liverpool Epidemic Strain is its lack of motility due to a fixed nonsense mutation flagellar motility gene *fleR*[334]. Of 122 U.K. LES

isolates analysed 120 have the fixed C-A mutation at LESB58 reference position 4656647 that introduces a premature stop codon in *fleR,* truncating the gene. An additional isolate (8835) has a 28bp deletion in *fleR* spanning the same locus of the nonsense mutation. LiP1a is the only isolate genome analysed without a nonsense mutation in *fleR*. However closely related LiP1c from the same patient as LiP1a does harbour the nonsense mutation in *fleR* (figure 4.2).

Changes to *oprD* likely resulting in loss of function have occurred independently by nonsense mutations more than any other gene analysed in the U.K. LES genomes. There is genetic evidence for loss of function in 55/122 genomes. Inactivation of *oprD* in *P. aeruginosa* has been observed to result in carbapenemase resistance[356]. Complete disruption in *oprD* has not previously been reported in the LES. However, decreased expression of *oprD* in LES400 and LES431 compared with PAO1 has previously been linked with increased MICs for meropenem[357]. Loss of function in membrane-bound lytic murein transporter gene *mltD* is observed (table 4.7). Disruptive mutations in *mltD* have not previously been observed in the LES. However disruptive mutations in PAO1 *mltD* homolog (PA1812) have been associated with increased beta-lactam MIC's[358]. Fosfomycin transporter gene *glpT* has independently lost function 8 times and loss of function mutations are carried in 13 isolate genomes. Loss of function has previously been observed in the LES[280]. In *P. aeruginosa* PA14 disruptive mutations in *glpT* gene result in fosfomycin resistance but limits the carbon source to glycerol-3-P[359]. Loss of function is observed in UDP-N-acetylmuramate-alanyl-gamma-D-glutamyl-meso-diaminopimelate-ligase gene *mpl* (how many times) and have previously been identified in the LES[280].

Disruption of homologous *mpl* in *P. aeruginosa* has been associated with piperacillin and ceftazidime resistance due to an accumulation of immature cell wall components and up-regulation of AmpC expression[329]. LES gene PLES_59241 carries a loss of function mutation in 20 genomes that have been acquired independently at least 12 times. Homolog of PLES_59241 in PAO1, PA5528 encodes a protein of unknown function but that is associated with the *amgRS* two component system[360]. The amgRS two component system is involved in aminoglycoside resistance. Loss of function mutations have not previously been identified in the LES in PLES_59241, or associated with other chronic *P. aeruginosa* infections. Instead, mutations in PLES_59241 have been inferred to cause gain of function in the LES previously[352]. In this study there is genetic evidence for loss of function in *mexB* in 26/122 genomes, or 31/122 genomes for the entire *mexAB-oprM* operon due to additional genomes with disruptions in *mexA*.

Ten different genes among the genes with the most unique missense mutations (n = 13) were inferred to be under positive selection in the alleles in the population without nonsense mutations (table 4.12). Many of these genes were adaptive by loss of function in other members of the population suggesting that there are different, within-patient, selection pressures on lineages of the LES or that constraint is relaxed or deleterious missense mutations are adaptive consistent with the nonsense results. Others such as genes *spoT, rpoD, pmrA, pmrB, algG exsD* and *exsA* did not appear to be adaptive by loss of function in the population.

4.4.1. Chapter Summary

Whole genome sequencing of the most geographically and historically diverse set of LES isolates yet was sequenced and it was determined that:

- Canadian and UK LES are genetically distinct
- Genomic Epidemiology correctly associates genomes to source of isolation
- Diversity within the LES in the LESB58 prophages and genomic islands
- Convergent evolution within the LES by loss of function

Chapter 5

Pseudomonas aeruginosa Adaptation to the non-Cystic Fibrosis Lung: Parallels with Cystic Fibrosis Chronic Lung Infections

5. Introduction

5.1. Background

Bronchiectasis is a chronic, progressive lung disease associated with a widening of the bronchi [69] that affects 26.9% of males and 35.2% of females in the UK, per 100,000 in the UK[361]. CF-associated bronchiectasis may by CF associated is detected early in life and results in life-long clinical interventions, including exposure to antibiotics from an early age[187]. Infection with *P. aeruginosa* occurs in ~30% of cases of Bronchiectasis overall and has been associated with increased morbidity and mortality[362,363].

Bacterial infections are a persistent problem for CF patients and in many cases worsen overall prognosis[362]. *P. aeruginosa* is the leading cause of overall morbidity and mortality in CF, presenting early in life[190] and often establishing a chronic infection later in life. CF patients may also harbour heterogeneous populations complicating treatment[66,285] or communities of distinct, potentially heterogeneous *P. aeruginosa* populations acquired from different sources, coexisting in the lungs. Additionally transmissible lineages of *P. aeruginosa* exist so that cross infection between patients whom otherwise attend the same clinics has resulted in segregation procedures[192,342]. In the CF lung chronic *P. aeruginosa* adapts in characteristic ways, including the adoption of a biofilm

lifestyle, switch to mucoidy, hypermutability, quorum sensing deficiency and lack of motility[66]. These adaptations are associated with worsening prognoses and chronicity[68,362]. For bronchiectasis patients without CF the disease presents later in life, typically between 60 and 70 years old[363]. It is estimated that in the UK of ~150,000 people living with bronchiectasis, 90% is non-CF bronchiectasis[361].

As with non-CF bronchiectasis, primary ciliary dyskinesia (PCD) patients are treated with intensive antibiotic chemotherapy, the lungs demonstrate poor mucociliary clearance, clinical risk factors of persistent infection[364]. A study of *P. aeruginosa* infection in PCD identified genetic similarities in adaptive routes to chronic infection as in CF. By whole genome sequencing of isolates from 12 PCD patients infected with *P. aeruginosa*, 10 were determined to be a chronic lineage and convergent evolution was inferred in 8 genes, *mexZ*, *algU*, *mucA, mexA, pscP*, *pilG*, *lasR* and *mexS*[364].

5.1.1. Aims

The aims of this chapter were to use whole genome sequencing of bronchiectasis associated *P. aeruginosa* to assess:

- The representation of different lineages of *P. aeruginosa* in a bronchiectasis patient populations
- Whether there is genetic evidence for adaptation within the lung comparable to that of cystic fibrosis isolates
- Whether there is evidence for transmissible lineages or lineages acquired by different patients from a common nosocomial source
- Genetic diversity within *P. aeruginosa* populations in the lungs of bronchiectasis patients

5.2. Methods

Genomes were *de novo* assembled and scaffolded, reads were mapped and variants called and annotated to the PAO1 reference genome, core genomes extracted, phylogenies approximated and annotated, MLST derived and genomes compared by pairwise alignment as described in Chapter Two: General Methods.

5.2.1. Samples and isolates used in this study

189 *P. aeruginosa* isolates were isolated from sputum samples of 93 bronchiectasis patients across 16 UK centres (supplementary table 2) determined to have a chronic infection (two or more positive respiratory tract cultures in the previous 12 months). For 24 patients two or more isolates were obtained and for three of these patients (patients 147, 148, 149) 14 or 15 isolates were obtained for sequencing. A subset of isolates representative isolates per patient, or multiple if they harboured multiple lineages (determined by carrying separate MLST profiles) of 99 isolates from 91 patients.

5.2.2. Inference of genes adaptive by loss of function

To infer that genes are adaptive by loss of function in the representative panel of genomes variants were annotated as being nonsense mutations (frame-shifting insertions and deletions and premature stop codons). How many times a gene was determined to have lost function independently across the population was taken as evidence for inference of that change being adaptive in non-CF bronchiectasis infections. It is shown in supplementary figure 1 how the assumptions that large numbers of independent acquisitions are more likely

due to adaptation in the lung than genetic background. Parsimoniously if an allele includes a nonsense mutation or mutations that are both unique and shared by position and base change, conservatively that which is shared is taken as having occurred first. As such the model conservatively estimates the minimum amount of independence thereby ignoring that the same change at the same position is not only possible but in, some genes, likely[274].

5.3. Results

5.3.1. Genome Assembly Quality

In order to characterise *P. aeruginosa* during chronic non-CF bronchiectasis lung infections 189 isolates were obtained from 91 patients attending 16 adult bronchiectasis clinics in the UK. Following adapter trimming and quality filtering, 189 genomes were *de novo* assembled and scaffolded. The quality of these draft genomes was assessed by contig length metrics the N50 and largest contig size. Further the GC% content and draft genome sizes were determined.

The number of contigs (longer than 1000bp in length) ranged from that of isolate C88 with 32 to that of A163 with 589. The median number of contigs per genome was 66. In total 74.07% of genomes were assembled into <100 contigs. The largest single contigs assembled in the genomes ranged from that of isolate A163 with 124,349bp to that of isolate C118 with 1,231,324bp. The median size of the largest contig assembled was 662,406bp. The N50 ranged from that of A163 with an N50 of 18,934bp to that of C73 with an N50 548,081bp. The median N50 was 242,365bp.

Genome sizes ranged from that of isolate C4 with 6,108,491bp to that of isolate C124 with 7,917,225bp. The median genome size across the range is 6,685,627bp. The GC content as a percentage of the genomes ranged from 65.27% to 66.21%.

5.3.2. Across Patient Population Diversity

MLST profiles were determined for all 189 genomes. The most common ST was that of 'clone C' (ST-17), representing 22 of 189 genomes. A further 19 sequence types represented just one genome respectively (Supplementary Table 2).

Two isolates, A36 and A163, presented the LES associated MLST profile, ST-146. Both genomes short reads were mapped to reference LES genome LESB58 and variants called. A36 and A163 were determined to be 113 SNPs & 25 indels and 109 SNPs and 22 indels different from LESB58 respectively. They were also adjacent to the LES on the core SNP tree of the wider *P. aeruginosa* population.

Some sequence types were partially determined and these are provisionally characterised as sequence type not found (NF). Of 189 genomes, 29 were ST-NF and as such were further assessed to deduce the potential ST (or ST's) from MLST loci profiles available. Alternatively, an ST may be genuinely novel where all loci have profiles assigned but in combination do not pertain to an ST. It was determined that 12 genomes were ST-novel (table 5.1), while the remainder carried 6/7 which combined suggested potential profiles to assign (table 5.1).

Table 5.1. Incomplete MLST profiles to be determined legitimately novel or 'nearest to' an establish profile by the loci that can be resolved

Isolate	ST	Closest ST	MLST loci						
			acs	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*
A19	NF	92 or 261	105	5	30	-	1	4	14
B37	Novel		107	4	3	27	12	7	128
B62	NF	1404	16	-	6	3	4	7	1
C101	NF	303 (4 loci)	16	-	12	18	3	4	9
C102	NF	304 (4 loci)	16	-	12	18	3	4	9
C103	NF	305 (4 loci)	16	-	12	18	3	4	9
C106	NF	156,179,353,1494 (6 loci)	-	27	28	3	4	13	7
C119	Novel		5	1	109	3	1	1	47
C120	Novel		17	5	11	5	4	29	2
C164	NF	1240,1985 (4 loci)	28	5	46	5	1	-	61
C16	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C17	NF	27,120,2314 (5 loci)	6	-	6	113	4	6	7
C18	NF	158,179,180,1496,2063,2109 (6 loci)	36	27	28	-	4	13	7
C21	Novel		22	6	1	3	1	76	1
C22	Novel		22	6	1	3	1	76	1
C23	Novel		22	6	1	3	1	76	1
C31	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C32	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C33	NF	155,677,1276 (5 loci)	28	5	36	-	3	13	7
C4	NF	260 (6 loci)	14	5	-	7	4	13	7
C55	Novel		28	5	11	11	15	75	1
C56	Novel		28	5	11	11	15	75	1
C57	Novel		28	5	11	11	15	75	1
C58	Novel		28	5	11	11	15	75	1
C59	Novel		28	5	11	11	15	75	1
C60	Novel		28	5	11	11	15	75	1
C69	NF	1707, 2055 (6 loci)	16	24	1	149	4	-	19
C83	NF	620 (6 loci)	9	7	63	13	8	-	8
C85	NF	156, 179, 353, 1494 (6 loci)	-	27	28	3	4	13	7

All genomes sequenced were included with the wider *P. aeruginosa* population and the core genome extracted. The core genome was 861,928bp. The phylogeny was approximated (supplementary figure 6). Gubbins was used to predict and filter recombinant SNPs from the alignment. There was 507,635 SNPs in the core genome, 6.32% of which were recombinant. After filtering the core genome contained 7,194 polymorphic sites for approximating the clonal frame. The resulting clonal frame from gubbins is displayed in figure 5.1. The placement of the genomes with diverse associated MLST was disparate, with isolates spread across the tree (figure 5.1).

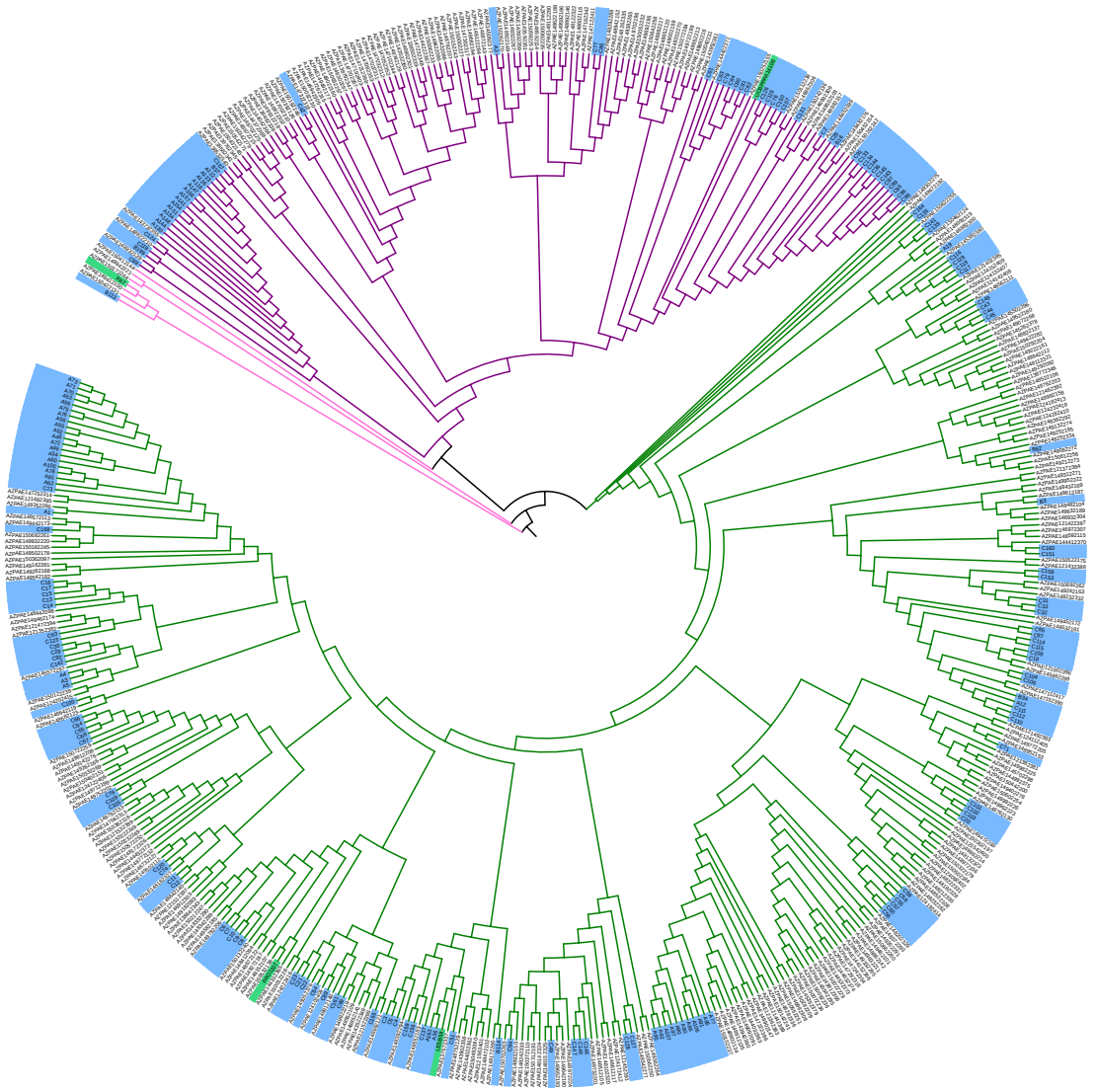


Figure 5.1. Core genome of the bronchiectasis genomes and the wider *P. aeruginosa* population with recombinant SNPs filtered out. Green represents *P. aeruginosa* phylogenetic group1, blue group 2 and pink group 3. Bronchiectasis genomes are labeled in blue

5.3.3. Multi-lineage infections

A subset of 24 patients provided more than one sample to this study and as such it was possible to determine that, as in CF, whether patients were harbouring multiple distinct, coexisting, infecting lineages. This was determined by identifying those patients with more than one MLST determined from sequenced isolates (supplementary table 2).

Patients 42, 72, 73 84, 85 and 148 presented with two different sequence types and Patient 92 carried three different sequence types ST-235, ST164 and ST-871. Isolates representative of each sequence type are highlighted per patient in supplementary table 2. Polymorphism in no more than MLST loci in the lung could result in distinct sequence types and false determination of separate populations. Figure 5.2 highlights by linking isolates provided by the same patient that when placed in the wider *P. aeruginosa* population (figure 5.1) the isolates with different sequence types are phylogenetically distinct, further demonstrating that these patients are harbouring multi-lineage infections.

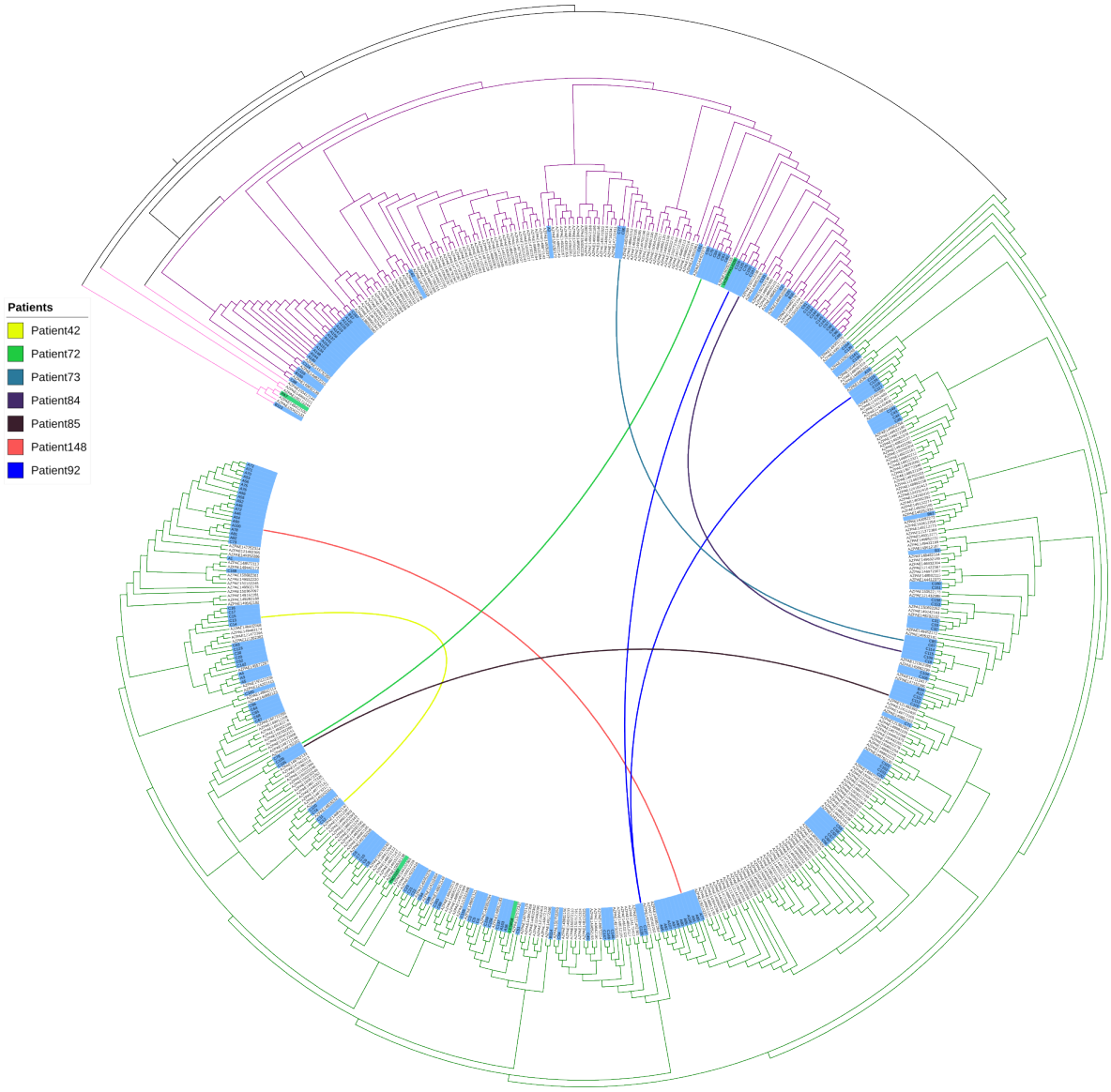


Figure 5.2. Cladogram representing a maximum likelihood approximation of the *P. aeruginosa* core genome. Green represents *P. aeruginosa* phylogenetic group1, blue group 2 and pink group 3. Representative genomes of each ST where patients harbor >1 are connected per patient to demonstrate relatedness

5.3.4. Lineage diversity

In order to assess the diversity of *P. aeruginosa* populations in the non-CF bronchiectasis lung two or more isolates were sequenced from 24 patients. Further 15 isolates were sequenced from each of patients 147, 148 and 149. Patient 148 however was determined to be carrying two distinct lineages of ST-17 (n = 4) and ST-175 (n = 11) (figure 5.2) and as such were treated as separate populations in the analysis. Table 5.2 shows the descriptive statistics of SNP and Indel differences of all-against-all comparisons of patients 147, 148 (ST-17/ST-175) and 149's isolates. The greatest median SNP and indel difference was that of patient 148 (ST-17) with 482.5 and 25.5 respectively. The least within median population diversity of SNPs and indels was patient 148 (ST-175) with 179 and 5 respectively.

Table 5.2. The number of single nucleotide polymorphism and small insertion and deletion differences between genomes isolated from the same patients respectively

Patients	Median SNPs	SNP range	Median indels	Indel range
Patient 147	261.00	88-640	14.00	0-35
Patient 148 (ST17)	482.50	159-654	25.50	6-34
Patient 148 (ST175)	179.00	79-403	5.00	0-36
Patient 149	206.00	68-327	10.00	3-28

Overall three or more isolates were sequenced from nine patients (figure 5.3). The within population diversity of isolates sequenced from patient 42 displayed the largest range of 206-949 SNPs and indels and the highest median with 875.5. The population sampled from patient 65 demonstrated the smallest range with 100-162 SNPs and indels and a median of 130.

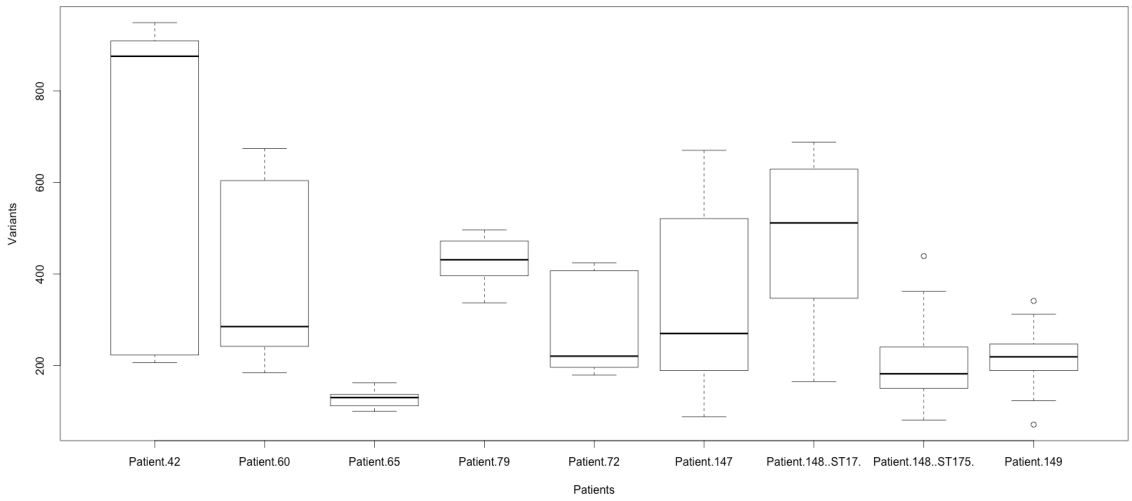


Figure 5.3. Boxplots of all against all whole genome pairwise variant (SNPs and indels) differences within patient lineages

5.3.5. Representative panel of bronchiectasis isolate genomes

A representative panel was created containing one isolate genome sequence from each patient or one of each sequence type if there were multiple infecting lineages. The resulting panel consisted of 99 isolate genomes from 91 patients. Supplementary table 2 highlights these 99 isolates, their MLST profiles and their associated patient identifiers. The core genome of the representative panel was 3,847,369bp containing 7,698 polymorphic sites, of which 7,534 were complete (without gaps or ambiguous nucleotides), of which 4030 were phylogenetically informative. An ML phylogenetic tree was approximated using the HKY85 substitution model (supplementary figure 7) and used as the starting tree to predict recombination. Gubbins predicted that 6.98% of SNPs in the core genome were as a result of recombination. The clonalframe is displayed in figure 5.4. Of the 99 panel genomes, 27 placed phylogenetically in *P. aearuginosa* group 2 of the tree, 71 in *P. aeruginosa* phylogenetic group 1 and a single isolate (B113) is a phylogenetic outlier but is not PA7-like that bifurcates from the branch between the two major clades. Bifurcation of *P. aeruginosa* phylogenetic groups 1 and 2 showed 100% bootstrap support.

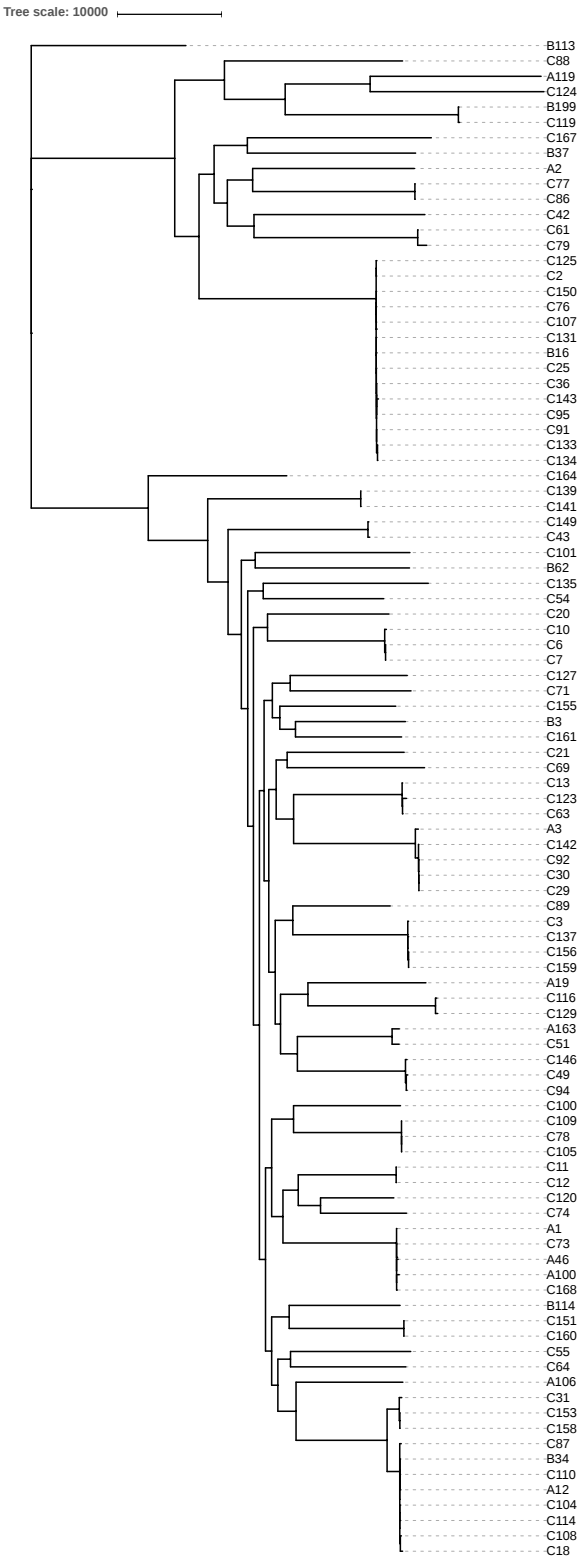


Figure 5.4. Core genome SNP phylogeny of representative panel of non-CF bronchiectasis genomes (n = 99). RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins. Tree scale is in number of SNPs, the branch to B113 is truncated for clarity

5.3.6. Evidence for potential common source of isolates or cross infection

Where MLST profiles were the same between isolates from different patients in the same centre, these were further investigated for genetic similarity by whole genome pairwise alignment, identifying SNP and indel differences. In five cases pairs of genomes differed by <200 SNPs and indels, (C6/C7, C29/C30, C105/C109, C139/C141, and C156/C159). The representative bronchiectasis panel genomes core SNP tree is labeled per centre in figure 5.5.

The between patient diversity of isolates sampled was lower in the five cases (C6/C7, C29/C30, C105/C109, C139/C141, and C156/C159) identified than between some isolate genomes from within the same patient. Isolates from within individual patients determined to be of the same lineage by MLST and placement on the core SNP tree (figure 5.1) varied by as much as >750 SNPs and indels (C125/C126) (table 5.3).

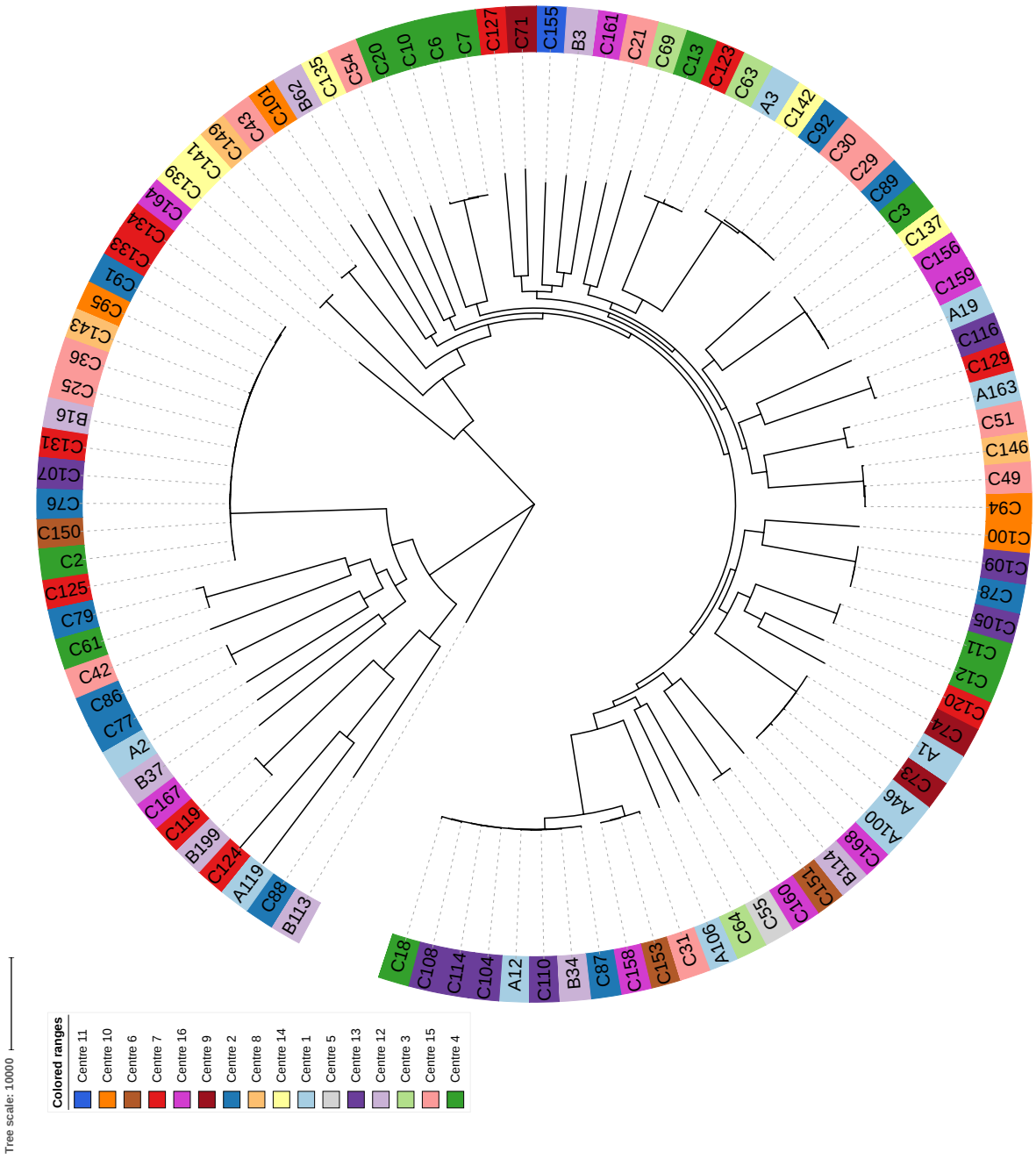


Figure 5.5. Core genome phylogeny of representative panel of non-CF bronchiectasis genomes (n = 99). Final ML tree approximated by RaxML after recombinant sites in the core genome were filtered out by Gubbins (figure 5.4). Leaf labels are coloured according to centre where isolates were sampled

Table 5.3. Whole genome all against all pairwise comparisons of isolate genomes with common MLSTs within the same centre. SNP and indel counts displayed per pairwise comparison

Isolate	Centre	Isolation date	Patient	MLST	Comparison	SNPs	INDELs
C6	4	03/03/2010	38	244	C6-C7	179	8
					C6-C8	3790	79
C7	4	23/03/2010	39	244	C6-C9	3733	74
					C6-C10	3833	62
C8	4	16/04/2010	40	244	C7-C8	3736	82
					C7-C9	3714	75
C9	4	16/04/2010	40	244	C8-C9	281	5
					C7-C10	3929	65
C10	4	16/04/2010	40	244	C8-C10	603	14
					C9-C10	515	11
C11	4	19/08/2010	41	282	C11-C12	340	8
C12	4	20/08/2010	42	282			
C29	15	03/06/2009	48	252	C29-C30	168	3
C30	15	04/06/2009	49	252			
C25	15	20/05/2009	46	253	C25-C36	3428	43
C36	15	21/05/2010	52	253			
C91	2	13/01/2011	76	253	C76-C91	846	5
C76	2	12/05/2009	70	253			
C77	2	03/07/2009	71	308	C77-C86	277	1
C86	2	06/08/2009	73	308			
C105	13	25/07/2009	83	840	C105-C109	131	8
C109	13	11/08/2009	85	840			
C104	13	16/05/2009	82	179	C104-C108	10,551	114
					C104-C110	2863	39
C108	13	11/08/2009	84	179	C104-C111	2765	52
					C104-C112	1913	25
C110	13	11/08/2009	85	179	C104-C114	6795	96
					C104-C115	6972	105
C111	13	11/08/2009	85	179	C108-C110	4780	78
					C108-C111	4852	66
					C108-C112	3780	62
					C108-C114	10,963	136
					C108-C115	10,833	133
					C110-C111	176	8

C112	13	11/08/2009	85	179	C110-C112 C110-C114 C110-C115 C111-C112 C111-C114 C111-C115 C112-C114 C112-C115 C114-C115	198 3115 3107 148 2838 2789 1941 1895 281	7 40 45 4 43 57 29 31 4
C114	13	05/12/2009	86	179			
C115	13	05/12/2009	86	179			
C125	7	29/04/2010	92	253	C125-C126 C125-C131 C125-C133 C125-C134 C126-C131 C126-C133 C126-C134 C131-C133 C131-C134 C133-C134	736 1159 872 817 9636 5847 5883 3486 3400 330	27 22 22 21 99 71 70 40 38 4
C126	7	29/04/2010	92	253			
C131	7	08/05/2010	93	253			
C133	7	28/05/2010	94	253			
C134	7	19/12/2010	95	253			
C156	16	16/11/2010	108	260	C156-C159	160	3
C159	16	09/12/2010	110	260			
C139	14	08/09/2010	98	ST2102	C139-C141	177	3
C141	14	01/10/2010	99	ST2102			

5.4. Adaptation by Loss of Function

To identify genes that may be adapting by loss of function in the representative panel of non-CF bronchiectasis isolate genomes the independent acquisition of nonsense mutations was inferred. Of the non-CF bronchiectasis representative panel of genomes there were 1,480 different genes with a nonsense mutation in at least one genome. There were 308 genes with nonsense mutations in two or more genomes (table 5.4). One gene, *mexB* has acquired a nonsense mutation 16 times in the population, also frequently independently acquiring loss of function were *mucA* with 13 acquisitions and *betT2* with 9 acquisitions. All genes with five or more acquisitions are displayed in table 5.5.

Table 5.4. The number of different genes with 2 or more independently acquired nonsense mutations in the representative panel population

Number of genes	Number of independent acquisitions of a nonsense mutation in each gene in the panel
205	2
47	3
37	4
8	5
3	6
5	7
1	9
1	13
1	16

Table 5.5. Genes in which nonsense mutations have occurred independently >=5 times in the representative panel population

Gene	Independent Acquisitions	Maintained in	Function / Comment
mexB	16	18	Transporter from MexAB-OprM efflux pump, antibiotic resistance, virulence
mucA	13	31	Anti-sigma factor, mutations can lead to mucoidy.
betT2	9	11	Transporter, uptake of small molecules such as choline and glycine betaine, contributing to growth via phosphatidyl choline metabolism and osmoprotection
bifA	7	9	Cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation
mexA	7	8	Membrane fusion protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
pcoA	7	15	Copper resistance
PA4469	7	19	Hypothetical protein in same operon as and upstream of sodM (superoxide dismutase; response to oxidative stress)
rbdA	7	24	Cyclic-di-GMP phosphodiesterase, modulation of biofilm dispersal, negative regulation of Pel production
pilJ	6	6	Methyl accepting chemotaxis receptor-like protein involved in twitching motility and biofilm formation
oprM	6	8	Outer membrane protein from MexAB-OprM efflux pump, antibiotic resistance, virulence
oprF	6	10	Major porin, biofilm formation
chpA	5	6	Chemotaxis-like chemosensory protein involved in twitching motility
fimV	5	5	Peptidoglycan-binding protein, promotes type IV pilin assembly, twitching motility
ladS	5	5	Sensor kinase, implicated in switch between acute and chronic infection
mutL	5	6	Mismatch repair system, DNA repair, mutation can lead to mutator phenotype
gmd	5	6	GDP-mannose 4,6-dehydratase,
mexS	5	6	Mutations promote MexT-dependent *mexEF-oprN* expression and multidrug resistance
pchE	5	7	Pyochelin synthesis
PA0054	5	6	Hypothetical protein

Genes inferred to be adaptive by loss of function (table 5.5) are involved with diverse functions such as antibiotic resistance (*oprM, mexB, mexA, mexS, oprF*), hypermutability (*mutL*), regulation of alingate biosynthesis and the mucoidy phenotype (*mucA*) and motility (*fimV*). Nonsense mutations in DNA mismatch repair genes are observed in 11 of the 99 panel isolates, including phylogenetic outlier B113 (table 5.4, supplementary figure 9). In 7 of the 11 isolates there is a frameshifting indel in *mutL*, in the remaining frame-shifting indels in *mutS,* including a 10bp deletion in C123.

Table 5.6. Isolate genomes with nonsense mutations in DNA mismatch repair genes *mutS, mutL, uvrD* or *mutM*

Isolate	Genetic basis
C123	10bp deletion in *mutS*
C31	Frameshifting indel in *mutL*
C78	Frameshifting indel in *mutL*
B113	Frameshifting indel in *mutL*
C129	Frameshifting indel in *mutL*
C133	Frameshifting indel in *mutS*
C134	Frameshifting indel in *mutS*
C127	Frameshifting indel in *mutS*
C79	Stop codon gained in *mutL*
C18	Stop codon gained in *mutL*
65-C74	Stop codon gained *mutL*

In total 31 genomes presented with a nonsense mutation in anti-sigma factor for alginate biosynthesis gene *mucA*. It was determined that nonsense mutations have been acquired at least 13 times in the gene in the bronchiectasis panel. Further, an additional three genomes have larger deletions in *mucA* so that there was a total of 34 of 99 genomes with evidence for loss of function in MucA.

The genes *mexA, mexB* and *oprM* that comprise the *mexAB-oprM* efflux pump were individually inferred to be adaptive by loss of function. Gene *mexA* carries a nonsense mutation in 8 genomes and they've been acquired a minimum of 7 times. Eighteen genomes carried a nonsense mutation in *mexB* from which it could be determined that there has been at least 16 independent acquisitions. A further four genomes carry a larger deletion in *mexB*, ranging from an estimated 43 to 280bp. Finally gene *oprM* carries a nonsense mutation in 8 genomes, acquired at least 6 times.

Figure 5.6 plots the associated pseudoCAP functional categories of the genes inferred to be adaptive by loss of function (table 5.4). There may be multiple functional associations per gene and each are counted according to the number of genomes in which loss of function is maintained in the respective genes. The function most often associated was 'Transport of Small Molecules' representing 15.95% of all associations followed by 'Membrane Proteins' that represent 13.36% of associations. Regulatory genes are associated with inferred adaptation by loss of function with 'Transcriptional Regulators' representing 8.4% and 'Two Component Systems' representing 3.23% of functional associations.

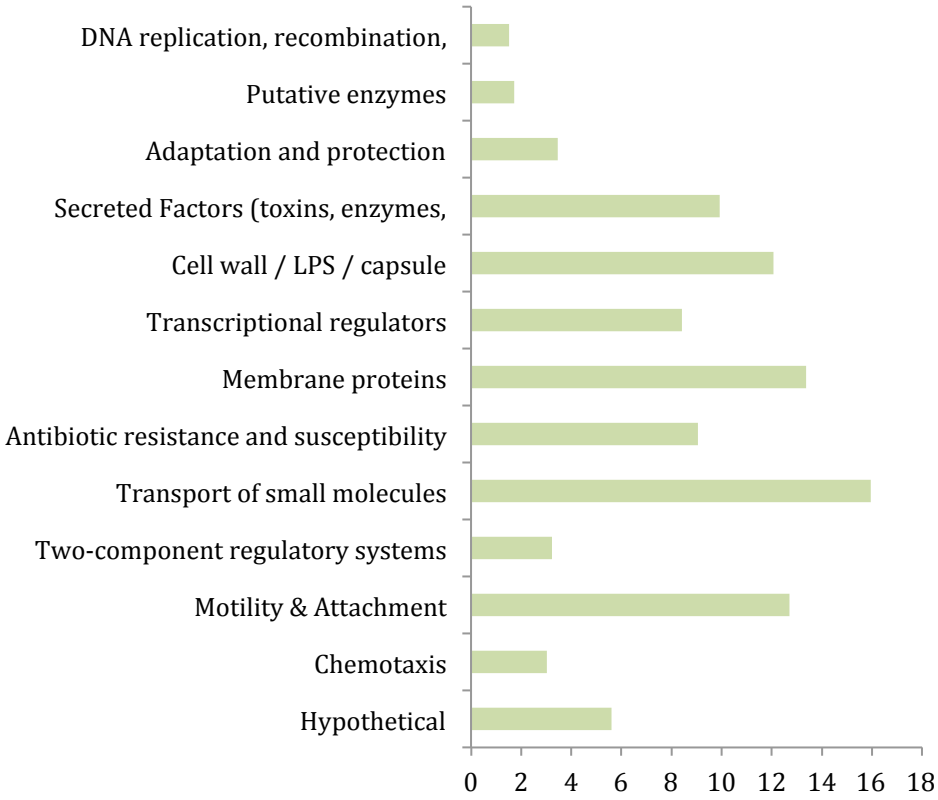


Figure 5.6. PseudoCAP functional categories associated with genes most often independently acquiring nonsense mutations (table 5.4) as a function of the number of genomes in which the loss of function allele is carried

5.5. Discussion

This study represents the first application of genomics to analyse chronically infecting *P. aeruginosa* from non-CF Bronchiectasis patients. It was possible to determine that multi-lineage infections occur, that within centre and within patient diversity can be similar suggesting potential common source or cross infection of isolates not know to be transmissible.

Compared with CF there is no evidence of widespread cross-infection of known transmissible isolates in the non-CF bronchiectasis set[208]. There is however evidence for the presence of the LES in one patient. The LES has previously been observed to have spread to non-CF patients[208,348]. However, the pairwise distance between isolates from different patients within the same centre was at times less than the mutational diversity of isolates within-patients. Lineages or 'strains' may be more abundant in the environment such as PA14 and clone C[289] were among the most abundant in the non-CF bronchiectasis set too. As such, their abundance isn't indicative of cross infection or common source of infection. However, by pairwise comparisons of isolate genomes sampled from different patients attending the same centre it was revealed that many differed by <200 variants. SNP cutoffs have been proposed in bacterial genomics for the inference of cross infection but lack a conceivable method to falsify their conclusions. In *P. aeruginosa* reservoirs around a hospital have been reliably determined from WGS[365] however; demonstrating the potential for the same approach to non-CF Bronchiectasis *P. aeruginosa* infections. Where the SNP diversity between patients is small relative to typical environmental sampling and especially when relative to within patient diversity, sampling of the hospital

environment could reliably rule-out or confirm a common source of infection within the clinic. Further segregation policies have been successful in CF clinics to stem cross infection of putative transmissible strains[366]; segregation of patients across which there is the lowest SNP diversity and further sampling of other patients may confirm the patient to patient infectivity of these lineages by observing a reduction in new cases.

Previously paired isolates from non-CF Bronchiectasis, 9/10 patients presented a common genotype while one patient harboured two distinct genotypes[367]. With a subsample of 24 patients multiple isolates were obtained and sequenced demonstrating the presence of multi-lineage infections in non-CF Bronchiectasis with as many as 3 infecting lineages. Figure 5.1 demonstrates how diverse the isolates carrying different MLST profiles are, rendering significantly less probable simple mutation within-lineage of an informative MLST loci resulting in different genotypes.

The first *P. aeruginosa* genomic study of non-CF *P. aeruginosa* opportunistically infecting diseased lungs, in PCD patients, confirmed pathoadaptive mutations typically observed in CF[364]. Similarly loss of function in *mucA* resulting in mucoidy and *lasR* are observed in non-CF bronchiectasis *P. aeruginosa* infections as in CF[66,214,325]. Other loss of function mutations common in genes *pliJ, chpA* and *fimV* are involved in twitching motility and have been observed in CF[66] and artificial sputum[349] meaning that this adaptation may pertain to survival in the physical conditions (comparable mucus viscosity) in non-CF Bronchiectasis as in CF.

Adaptation by loss of function in the non-CF Bronchiectasis *P. aeruginosa* population was inferred from independence as described in general methods. It was assumed that the genes with the greatest number of independent acquisitions of loss of function were more likely to have occurred in their common lung environment than in varied source niches. However, as each genome was mapped to reference genome PAO1, genetic distance could be a factor. Genomes with the greatest number of mutations compared to PAO1 would also have the greatest number of genes with nonsense mutations. Supplementary figure 9 demonstrates that this is the case. Genomes belonging phylogenetically to *P. aeruginosa* phylogenetic group 3 (+ other outliers to the two major groups) that typically have the greatest distance to PAO1 also have the greatest number of genes with nonsense mutations. The trend follows that phylogenetic group 2, that of PA14, has the next largest number of genes with nonsense mutations and the fewest are derived from genomes belonging to *P. aeruginosa* phylogenetic group 1 within which PAO1 is nested. By comparison inclusion of genes inferred to be adaptive by loss of function (table 5.5) only, does not display this effect (supplementary figure 1).

Genes that commonly lose function in chronic CF infections, *mucA* and *lasR*[66] were observed to have lost function. Regulatory genes typically adaptive in CF such as *mexT, retS, exsD,* and *ampR* were not so in non-Bronchiectasis. In CF there are multiple adaptive routes to chronic infection[66,279] and this is most likely the same with non-CF Bronchiectasis. Other genes commonly carrying loss of function mutations (*pilJ*, *chpA*, *fimV*) are implicated in disruption of twitching motility, observed in CF[66] and artificial sputum media[349].

P. aeruginosa will produce exopolysaccharides alginate, Psl and Pel that contribute to biofilms to protect itself from hostile environments[152]. Psl is integral to surface attachment and in CF it's been suggested that free-floating biofilms are more typical during infection[158]. In non-CF bronchiectasis this preference towards Pel is also observed with 36 genomes with one or more loss of function mutations in Psl genes with additional loss of function in genes *bifA*[368], *rbdA*[369], *oprF*[370] and *ladS*[371] that support this view.

5.5.1. Chapter Summary

By sequencing *P. aeruginosa* isolates sampled from non-CF bronchiectasis lung infections it could be determined:

- Multi-lineage infections are present in multiple patients sampled, as in CF
- Adaptive routes such as mucoidy are present in non-CF Bronchiectasis
- Evidence of a common source of isolates within centres
- Significant within-patient diversity

Chapter Six

Integron Mediated Antibiotic Resistance in *Pseudomonas aeruginosa* from a Hospital in Thailand

6. Introduction

6.1. Background

Pseudomonas aeruginosa poses a public health problem as a major nosocomial pathogen with a standing diversity of intrinsic antibiotic resistance mechanisms[167]. Increasingly, horizontally acquired antibiotic resistance, in particular to carbapenems, is of serious concern. As a result the World Health Organisation (WHO) has designated *P. aeruginosa* a critical threat to public health; its highest category concerned with antibiotic resistant bacteria. Antibiotic resistance genes (ARGs) acquired horizontally are present in hospital and environmental *P. aeruginosa*[79,138,200].

6.1.1. Antibiotic Resistance

Awareness of antibiotics increased through the 1930's to 1950's and in the 1950's the potential for horizontally acquired resistance became apparent[372]. Bacteria may acquire genes horizontally by conjugation, transformation, transduction and via transposable elements such as integrons. Conjugation and transformation have been observed to transfer ARGs in a variety of environments including soil, marine sediment, seawater and sewage wastewater[373]. Bacteriophage viral alterations of bacterial genomes occur[374] across broad host ranges and the role of transduction in environmental transfer of ARGs has also been observed in marine environments[375]. The gene encoding

methicillin resistance *mecA*, for instance, famously acquired by methicillin resistant *S. aureus* (MRSA), has been observed within bacteriophage DNA in a wastewater systems[376].

While antibiotic resistance genes and antibiotics have naturally been present in the environment the scale and impact of human pollution of the environment with antibiotics[377] is well evidenced. During 2013 for example, 53,800 tons of antibiotics were released into rivers and waterways in China[378]. Environmental pollution with antibiotics can be demonstrated to be associated with in an increased abundance of ARGs[379–381]. Since the widespread manufacture of antibiotics ARGs in a variety of environments has increased[173,382,383]. A study by Ali Khan *et al.*,[384] in Pakistan provides a clear demonstration of the anthropogenic impact on environmental antibiotic resistance. Antibiotics measured from five rivers matched unpolluted levels from other studies, whereas downstream of Lahore oxytetracycline, trimethoprim and sulfamethoxazole were measured at 1100, 1700 and 2700ng L^{-1}[384]. The range near drug formulation facilities was even higher with 1,100-49,000ng L^{-1} of antibiotics erythromycin, lincomycin, ciprofloxacin, ofloxacin, levofloxacin, oxytetracycline, trimethoprim and sulfamethoxazole[384]. The levels of *sul1* and *drfA1* genes were associated with this antibiotic pollution[384]. A study by Yong-Guan-Zhu *et al.*,[385] recently further demonstrated that in a sample of 18 estuaries in China 248 antibiotic resistance genes were present representing all major mechanisms of resistance with antibiotic deactivation, efflux pumps and cellular protection[385]. Network analysis showed an association between the

presence of ARGs and an integron integrase gene (*int1*) demonstrating the probable mobility of these genes[385].

6.1.2. Mobile genetic element associated antibiotic resistance in *P. aeruginosa*

P. aeruginosa, largely as a result of, intrinsic and acquired resistances is a complicated nosocomial pathogen[167]. The high level of resistance in various *P. aeruginosa* worldwide[386–388] leaves only colistin as a 'last resort' antibiotic. However, in 2016 a study by Liu *et al*., reported the discovery of the MCR-1 colistin resistance conferring plasmid in *E. coli*[389]. MCR-1 was subsequently demonstrated to confer resistance to *P. aeruginosa in vitro* and *in vivo*[390]. Rather than acquisition *P. aeruginosa* has also be observed to inherit resistance to polymyxins including colistin vertically; mutations accumulated in *pmrAB* have resulted in resistance in isolates from colistin treated CF patients[391]. Lineages of MDR *P. aeruginosa* have been observed to be overrepresented in hospitals worldwide; ST-235, ST-132 and ST-111 are enriched for integron antibiotic resistance gene casettes[392–394]. While ST-235's are typically still susceptible to colistin, ST-235s with colistin resistance have also have been observed in South Korea[395].

In a study across tertiary centres in Thailand, 261 isolates were assessed for antibiotic resistance and it was demonstrated that 71.65% were multidrug resistant with carbapenem resistance and with >50% of isolates in each hospital carrying resistance to meropenem[396]. Molecular investigation of multidrug and

carbapenem resistant *P. aeruginosa* in Thailand revealed that the genetic basis was not always clear and was multi-factorial[397]. However, decreased expression of OprD was observed in 93.56% of isolates and increased expression of MexXY and MexAB-oprM[397]. Integron associated genes IMP-1, IMP-14 and VIM-2 were additionally identified [397]. These findings are consistently observed in hospitals in other countries[76,398,399].

6.1.3. Aims

The aims of this chapter are to assess the antibiotic resistance of clinical *Pseudomonas aeruginosa* isolates from a hospital in Bangkok, Thailand. A selection of isolates are to be sequenced to assess:

- The genetic basis of different resistance profiles, particularly
- The presence of integrons with antibiotic resistance genes
- Potential presence of international MDR strains
- Potential for transmission of lineages or integrons within the hospital

6.2. Methods

A sample of *P. aeruginosa* isolates was provided by Dr Pitak Santanirand from Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, to investigate antibiotic resistance profiles and the potential for integron mediated resistance. Genomes were assembled, reads mapped and variants called, MLST determined and core SNP phylogenies approximated as described in the general methods.

6.2.1. Antibiotic susceptibility and resistance

Antimicrobial susceptibility testing was carried out according to the British Society of Antimicrobial Chemotherapy (BSAC) guidelines. Briefly, isolates to be tested were cultured onto Columbia plates (overnight 37 °C). From these, single colonies were mixed with sterile distilled water to attain a standard optical density (10 MacFarland units), and 10 µl spread onto iso-sense plates and incubated overnight at 37 °C with Meropenem (10 µg), Ceftazidime (30 µg), Piperacillin/Tazobactam (75/10 µg)/, ciprofloxacin (5 µg), tobramycin (10 µg) and Colistin (25 µg) antibiotic discs. Antibiotic sensitivity was determined by measuring the zone of inhibition and the results were interpreted as sensitive, intermediate resistant, or resistant consistent standardised BSAC protocols (table 6.1).

Table 6.1. Zone diameter breakpoints for *Pseudomonas spp.*

	Interpretation of zone diameters (mm)			
Antibiotic	**R>**	**I**	**S≤**	**Disc content (µg)**
Meropenem	15	16-19	20	10
Ceftazidime	23	-	24	30
Piperacillin/Tazobactam	24	-	25	75/10
Ciprofloxacin	19	20-29	30	5
Tobramycin	19	-	20	10
Colistin	4	-	4	25*

*Disc diffusion is no longer recommended for colistin susceptibility testing

6.2.2. Antibiotic Resistance Gene Predictions

Integrons, whether complete In0 elements (presence of a integrase gene and no *attC* sites), CALIN elements (clusters of *attC* sites lacking integron integrases) and complete integrons (an integrase and at least one *attC* site) were discovered from assembled genomes using integron finder[400]. The --local_max option was used for optimum sensitivity and --func_annot that matches predicted functional sequences with integron or integron like elements to the Resfams database[401].

6.3. Results

P. aeruginosa isolates from Ramathibodi Hospital, Mahidol University, Bangkok, Thailand were acquired from a range of patients and types of infection (Table 6.2) to assess the presence of antibiotic resistance and the potential for determining its genetic basis. In total 49 isolates were received for analysis, all of which were tested for resistance against commonly used anti-pseudomonal antibiotics.

Table 6.2. All Isolates sampled for this study with patient age, gender and relevant pathology

Isolate	Age	Sex	Specimen
A1(U2448)	67	M	Urine
A10(P2744)	62	M	Stoma
A2(H882)	66	M	Blood
A3(H638)	90	M	Blood
A4(P2337)	55	F	Chest
A5(P2338)	55	F	Chin
A6(P2087)	56	M	Wound swab
A7(P2010)	61	M	Wound swab
A8(P1885)	74	M	Foot
A9(P2733)	53	F	Chest
B1(P2356)	1	F	PCD
B10(T2867)	72	F	Sputum
B2(E2758)	48	F	Eye (Plate)
B3(T2101)	81	M	Sputum
B4(U2816)	63	M	Urine
B6(P2404)	72	F	Foot
B7(T2709)	67	F	Sputum
B8(T2584)	80	F	Sputum
B9(T2436)	90	M	Sputum
C1(T2548)	98	F	Sputum
C10(T3359)	74	M	Nasopharynx
C2(T2284)	66	F	Sputum
C3(T2924)	90	M	Sputum
C4(3139)	52	M	Urine (catheter)
C9(P1814)	1	F	Ascitic
C6(T3044)	81	M	Nasal
C7(T3582)	72	F	Sputum
C8(T3532)	83	M	Sputum
C9(T3498)	82	M	Sputum
D1(T3354)	55	F	Sputum
D10(U4045)	82	F	Urine
D2(T3266)	72	F	Sputum
D3(T3141)	52	M	Sputum
D4(T3649)	87	F	Sputum
D5(T3677)	82	M	Maxillary sinus

D6(T3664)	67	F	Sputum
D7(U3171)	78	M	Urine
D8(T3382)	79	F	Sputum
D9(T3362)	66	M	Sputum
E1(T4189)	55	F	Sputum
E10(P4321)	61	M	Wound swab
E2(E4656)	59	F	Eye (Plate)
E3(T3979)	92	M	Sputum
E4(T4247)	67	M	Sputum
E5(U4030)	8	M	Urine
E6(U4117)	77	F	Urine
E7(E3885)	39	F	Eye (Plate)
E8(U4082)	62	M	Urine
E9(U4068)	81	M	Urine (catheter)

6.3.1. Antibiotic Resistance Profiles

Isolates were tested for resistance to meropenem, ceftazidine, piperacillin/tazobactam, ciprofloxacin, tobramycin and colistin. Figure 6.1 shows the respective resistance and sensitivity profiles to the six antibiotics. Of 49 isolates, 13 were susceptible to all antibiotics tested except colistin. A further two isolates were susceptible to all except colistin and intermediary resistance to meropenem. The majority (n = 34) were resistant to at least one antibiotic tested and of those 23 are resistant to two or more. The most common resistance observed was to piperacillin/tazobactam present in 29/49 isolates, followed by ciprofloxacin observed in 19/49 isolates and thirdly ceftazidine observed in 18/49 isolates. Intermediary resistance is present for meropenem and ciprofloxacin in 7/49 and 4/49 isolates respectively (table 6.3).

Table 6.3. Count of isolates resistance, susceptible or displaying intermediary results to each antbiotic tested

Number of isolates	MEM10	CAZ30	PTZ85	CIP	TN	CO25
Resistant	9	18	29	19	17	0
Susceptible	33	31	20	26	32	49
Intermediary	7	0	0	4	0	0

Isolates tested for resistance and susceptibility to the 6 antibiotics also exhibited multiple resistances at once (figure 6.1). Five isolates, A1U2448, B3T2101, B9T2436, E1T4189 and E9U4068 were resistance to all antibiotics tested except colistin. A further three isolates were resistant to four antibiotics and intermediary for meropenem.

Figure 6.1. Antibiotic resistance and susceptibility patterns to meropenem, ceftazidine, piperacillin/tazobactam, ciprofloxacin, tobramycin and colistin

Strains	MEM10	CAZ30	PTZ85	CIP	TN	CO25
A10(P2744)	S	R	S	S	S	S
A1(U2448)	R	R	R	R	R	S
A2(H882)	S	S	S	S	S	S
A3(H638)	I	R	R	R	R	S
A4(P2337)	S	S	R	S	R	S
A5(P2338)	R	S	R	R	R	S
A6(P2087)	S	R	R	S	S	S
A7(P2010)	S	S	S	R	S	S
A8(P1885)	S	S	S	S	S	S
A9(P2733)	S	S	S	S	S	S
B10(T2867)	S	S	R	S	S	S
B1(P2356)	S	R	R	S	S	S
B2(E2758)	S	S	S	S	S	S
B3(T2101)	R	R	R	R	R	S
B4(U2816)	I	S	S	S	S	S
B6(P2404)	S	S	R	I	R	S
B7(T2709)	S	S	S	S	S	S
B8(T2584)	R	S	R	R	R	S
B9(T2436)	R	R	R	R	R	S
C10(T3359)	S	S	S	S	S	S
C1(T2548)	S	R	R	R	R	S
C2(T2284)	S	S	R	I	S	S
C3(T2924)	S	R	R	S	S	S
C4(3139)	S	S	R	R	S	S
C6(T3044)	S	S	S	S	S	S
C7(T3582)	I	R	R	R	R	S
C8(T3532)	R	R	R	R	S	S
C9(P1814)	I	R	R	R	R	S
C9(T3498)	S	S	S	S	S	S
D10(U4045)	S	S	S	S	S	S
D1(T3354)	S	S	R	I	S	S
D2(T3266)	I	R	R	R	R	S
D3(T3141)	S	S	R	R	S	S
D4(T3649)	S	S	R	S	S	S
D5(T3677)	S	S	S	S	S	S
D6(T3664)	S	S	R	S	S	S
D7(U3171)	S	S	R	I	S	S
D8(T3382)	I	S	R	R	S	S
D9(T3362)	S	S	S	S	S	S
E10(P4321)	S	S	S	S	R	S
E1(T4189)	R	R	R	R	R	S
E2(E4656)	I	S	S	S	S	S
E3(T3979)	S	S	S	S	S	S
E4(T4247)	S	S	S	S	S	S
E5(U4030)	S	R	R	R	R	S
E6(U4117)	S	S	R	S	S	S
E7(E3885)	R	R	S	R	R	S
E8(U4082)	S	R	S	S	S	S
E9(U4068)	R	R	R	R	R	S

6.3.2. Genome assemblies

A sample of isolates with a variety of resistance profiles (n=24) was selected for sequencing (table 6.4). The genomes ranged in size from 6,250,924bp to 7,916,909bp. Genome quality measured by the N50 metric ranged from 84,855 to 765,594 and number of contigs from 20 to 617.

Table 6.4. Genome assembly size and length based quality metrics, longest scaffold length, number of contigs and N50

OriginalID	Contigs	Scaffolds	Genome Size	Longest Scaffold	N50
D9(3362)	94	93	7184787	555697	250354
C7(T3583)	60	58	6908303	679729	270739
A1(A2448)	94	89	7092398	423130	210888
B3(T2101)	62	58	6919171	504101	347699
A3(H638)	78	73	7150152	1003143	328126
B9(T2436)	84	80	7150150	604449	262087
E1(T4189)	37	32	6334840	951957	507745
C8(T3532)	67	63	6912106	784647	269661
C9(P1814)	89	82	7142873	931611	255757
E9(4068)	74	71	7068852	1411812	554462
D3(3141)	56	52	6880848	1121322	480584
E3(T3979)	51	47	6272937	692536	385450
A9(P2733)	20	20	6250924	1331236	765594
B6(P2404)	40	36	6675891	848636	608561
D5(T3677)	43	40	6474838	972162	354473
C6(T3044)	47	45	6733566	923035	477284
A8(P1885)	44	38	6273981	1230438	564768
B4(U2816)	49	46	6369227	615979	287370
E6(U4117)	617	609	7916909	282325	84855
B2(E2758)	41	35	6398247	772742	426159
B8(T2584)	46	39	6993634	870481	355083
D8(T3382)	32	27	6582611	775206	478039
E4(T4242)	43	36	6408319	779058	472459
D1(T3354)	41	34	6937757	868611	583246

6.3.3. Antibiotic resistance/susceptibility of those sequenced

A sample of isolates tested for antibiotic resistance and susceptibility were genome sequenced; table 6.5 lists the isolates, patient information and pathologies.

Table 6.5. Patient information of those that provided isolates that were sequenced

Isolates	ST	Patient Age	Patient Gender	Pathology
D1(T3354)	233	55	F	Sputum
C8(T3532)	235	83	M	Sputum
C9(P1814)	235	82	M	Sputum
D3(3141)	244	52	M	Sputum
D9 3362	253	66	M	Sputum
B8(T2584)	260	80	F	Sputum
B4(U2816)	270	63	M	Urine
A1(A2448)	273	67	M	Urine
C6(T3044)	277	81	M	Nasal
B6(P2404)	309	72	F	Foot
D5(T3677)	357	82	M	Maxillary sinus
A9(P2733)	379	53	F	Chest
E3(T3979)	381	92	M	Sputum
D8(T3382)	491	79	F	Sputum
B3(T2101)	708	81	M	Sputum
C7(T3583)	708	72	F	Sputum
A3(H638)	1121	90	M	Blood
B9(T2436)	1121	90	M	Sputum
B2(E2758)	1330	48	F	Eye (Plate)
E1(T4189)	1459	55	F	Sputum
E9(4068)	2191	81	M	Urine (catheter)
A8(P1885)	NF	74	M	Foot
E6(U4117)	NF	77	F	Urine
E4(T4242)	Novel	67	M	Sputum

Multidrug resistant isolates A1(U2448), B3(T2101), B9(T2436), E1(T4189) and E9(U4068) were among those sequenced. Eight isolates were also sequenced that are entirely susceptible to the antibiotics tested. An additional isolate, B4(U2816) was completely susceptible except for intermediary resistance to meropenem reported.

Isolates	MEM10	CAZ30	PTZ85	CIP	TN	CO25
D9T3362	S	S	S	S	S	S
C7T3582	I	R	R	R	R	S
A1U2448	R	R	R	R	R	S
B3T2101	R	R	R	R	R	S
A3H638	I	R	R	R	R	S
B9T2436	R	R	R	R	R	S
E1T4189	R	R	R	R	R	S
C8T3532	R	R	R	R	S	S
C9P1814	I	R	R	R	R	S
E9U4068	R	R	R	R	R	S
D3T3141	S	S	R	R	S	S
E3T3979	S	S	S	S	S	S
A9P2733	S	S	S	S	S	S
B6P2404	S	S	R	I	R	S
D5T3677	S	S	S	S	S	S
C6T3044	S	S	S	S	S	S
A8P1885	S	S	S	S	S	S
B4U2816	I	S	S	S	S	S
E6U4117	S	S	R	S	S	S
B2E2758	S	S	S	S	S	S
B8T2584	R	S	R	R	R	S
D8T3382	I	S	R	R	S	S
E4T4247	S	S	S	S	S	S
D1T3354	S	S	R	I	S	S

Figure 6.2. Antibiotic resistance and susceptibility patterns to meropenem, ceftazidine, piperacillin/tazobactam, ciprofloxacin, tobramycin and colistin of the isolates sequenced

Of the isolates sequenced five were only susceptible to colistin, presenting resistance to 5/6 antibiotics tested: A1(U2488), B3(T2101), B9(T2436), E1(T4189) and E9(U4068) (table 6.6). A further two isolates sequenced, C7(T3582) and A3(H638) were resistant to 5/6 antibiotics and were similarly only susceptible to colistin but also were of only intermediary resistance to meropenem (Table 6.6).

Table 6.6. Count of sequenced isolates resistance, susceptible or displaying intermediary results to each antibiotic tested

Number of isolates	MEM10	CAZ30	PTZ85	CIP	TN	CO25
Resistant	7	9	15	12	10	0
Susceptible	12	15	9	10	14	24
Intermediary	5	0	0	2	0	0

6.3.4. Genome Diversity

The MLST profiles for all 24 genomes were determined and it was observed that they represent 21 distinct sequence types (table 6.7). For isolates A8(P1885) and E6(U4117) the ST could not be determined from 6/7 and 5/7 loci identified respectively and isolate E4(T4242) carried a novel combination of profiles. Sequence types 235, 1121 and 708 were each represented by two isolates (table 6.7).

Table 6.7. *P. aeruginosa* MLST profiles for all genomes sequenced

Isolate	ST	*acs*	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*
D9 3362	253	4	4	16	12	1	6	3
D1(T3354)	233	16	5	30	11	4	31	41
D8(T3382)	491	16	55	7	61	1	4	45
C7(T3583)	708*	11	3	-	3	1	4	60
A1(A2448)	273	104	4	36	71	4	-	53
B3(T2101)	708	11	3	11	3	1	4	60
A3(H638)	1121*	67	-	65	31	1	6	26
B9(T2436)	1121*	67	-	65	31	1	6	26
E1(T4189)	1459	11	2	6	3	4	38	3
C8(T3532)	235	38	11	3	13	1	2	4
C9(P1814)	235	38	11	3	13	1	2	4
E9(4068)	2191*	111	30	64	26	-	59	7
D3(3141)	244	17	5	12	3	14	4	7
E3(T3979)	381	11	20	1	65	4	4	10
A9(P2733)	379	39	5	11	28	4	4	63
B6(P2404)	309	13	8	9	3	1	17	15
D5(T3677)	357	2	4	5	3	1	6	11
C6(T3044)	277	39	5	9	11	27	5	2
A8(P1885)	NF*	111	30	64	58	48	24	-
B4(U2816)	270	22	3	17	5	2	10	7
E6(U4117)	NF*	-	5	-	7	3	12	19
B2(E2758)	1330	76	5	30	3	2	12	18
B8(T2584)	260	14	5	10	7	4	13	7
E4(T4242)	Novel*	6	76	1	3	3	28	3

*MLST's may be unresolved (not found, NF), novel if they match 7 loci without a defined ST or they may be approximated based on the loci present using the pubMLST database. Where ST's are approximate the only possible ST for the combination available is designated though it's possible they too are novel.

To further analyse the diversity of genomes sequenced they were placed in the context of the wider population of *P. aeruginosa* genomes, core genome extracted and a maximum likelihood tree approximated (Figure 6.3). The core genome was comprised of 2109 complete polymorphic sites (gaps ignored), 1566 of which were phylogenetically informative.

Isolates identified with the same sequence type, ST-708 isolates C7(T3583)-B3(T2101) and ST-1121 isolates A3(H638)-B9(T2436) respectively were adjacent on the core SNP tree and ST-235 isolates C8(T3532)-C9(P1814) were additionally adjacent to AZPAE14726_2315.

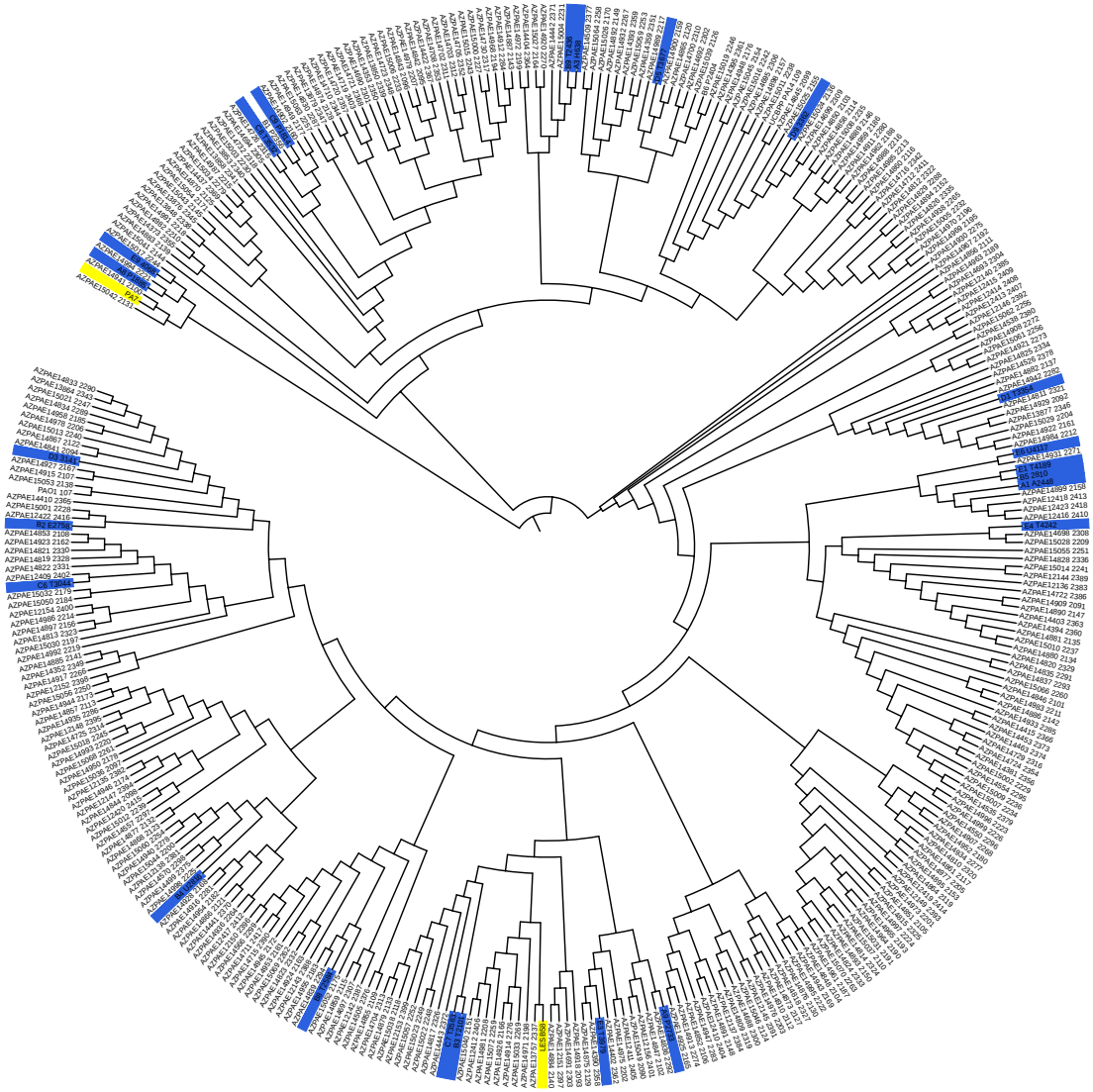


Figure 6.3. Genomes sequenced and the wider *P. aeruginosa* genome population core SNP phylogeny. RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins. Thai genomes leaf labels are in blue and reference *P. aeruginosa* genomes in yellow

6.3.4.1. Whole genome pairwise differences within sequence types

Three pairs of isolates carried the same sequence type C8(T3532)/(C9(P1814) (ST-235), C7(T3583)/B3(T2101) (ST-708) and A3(H638)/B9(T2436) (ST-1121). These were further assessed for genetic similarity by whole genome pairwise alignment and resulting SNP and indel counts. ST-1121 isolates A3(H638) and B9(T2436) were 23 SNPs and 5 indels different, ST-708 isolates C7(T3583) and B3(T2101) are 12 SNPs and 7 indels different whereas ST235 isolates C8(T3532) and C9(P1814) were 6,249 SNPs and 79 indels different.

The pairs of isolates most genetically similar also presented similar resistance/susceptibility profiles. A3(H638) and B9(T2436) sampled from the same patient, were 23SNPs and 5 indels different and displayed resistance to ceftazidime, piperacillin/tazobactam, ciprofloxacin and tobramycin but B9(T2436) is resistant to meropenem while A3(H638) displays intermediary resistance. C7(T3583) and B3(T2101) from different patients, were 12SNPs and 7 indels different show the same antibiogram except where C7(T3583) is intermediary for resistance to meropenem (figure 6.2).

6.3.5. CARD database Antibiotic Resistance Gene Prediction

All genomes were aligned to the CARD database and it was determined whether there was genetic evidence for antibiotic resistance (figure 6.4). Genomes ranged from A8(P1885) that perfectly matched four antibiotic resistance genes or gene variants to A1(A2448) that matched 21. The mean number of elements perfectly matched was 14.39. Three of the most susceptible isolates, A1(A2448), B3(T2101) and C7(T3583) matched the most elements perfectly with 21, 20 and 20 respectively. E1(T4189), B9(T2436) and E9(4068) also among the most resistance isolates varied in their predictions however from 8 to 16 elements.

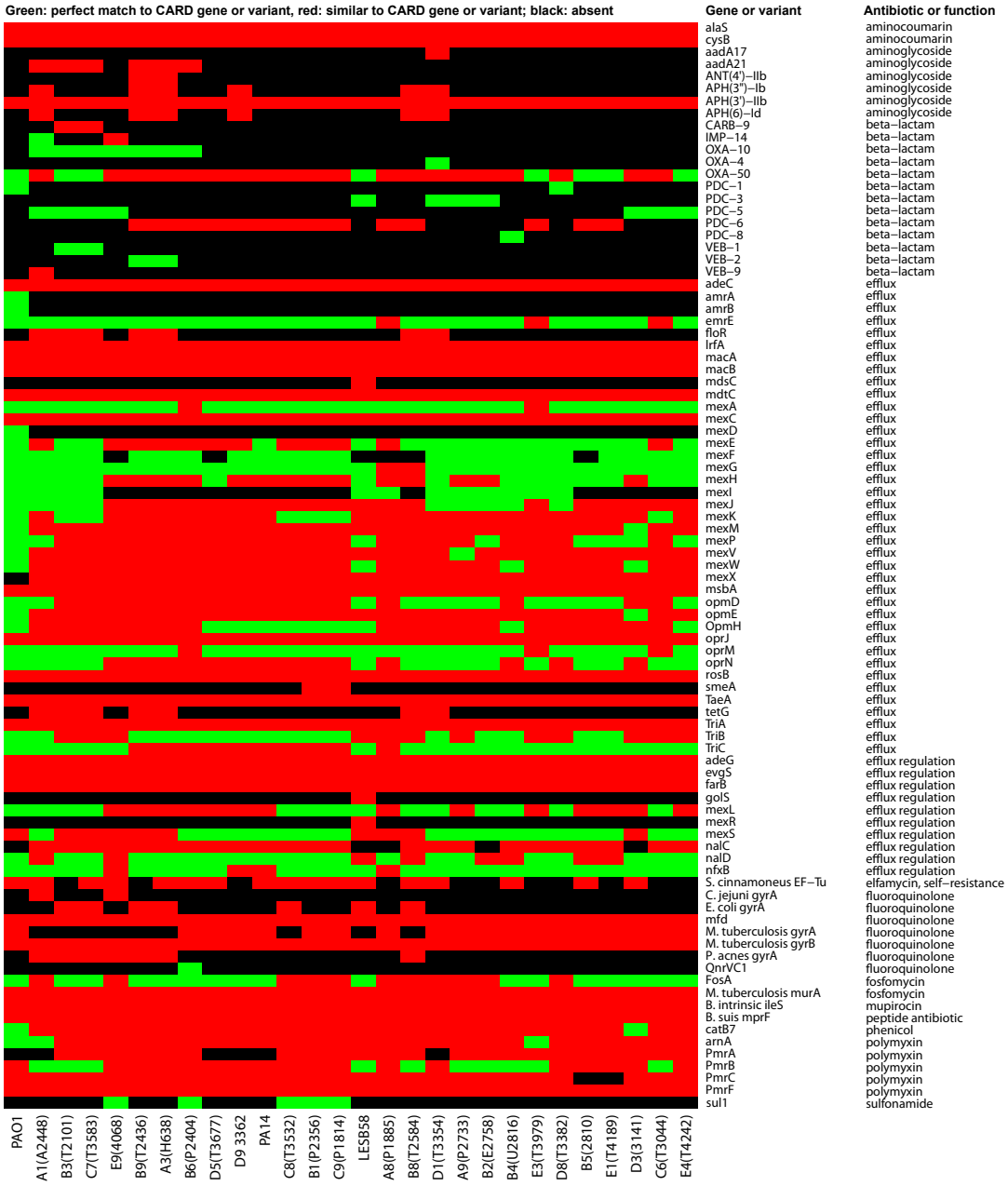


Figure 6.4. Results of search all genomes against the CARD database for the presence of resistance genes and variants

6.3.6. Integrons Carrying Antibiotic Resistance Genes

The presence of integrons was determined with integron finder and functional annotation using the resfams database to determine antibiotic resistance genes. It was determined that 12 genomes carried a complete integron and two of these carried an one additional CALIN element and A1(A2448) a complete integron and two CALIN elements (table 6.8).

Table 6.8. Number of integron, CALIN elements and In0 elements predicted per genome by Integron Finder

Isolate	Complete Integrons	CALIN	In0
D1(T3354)	1	0	0
B8(T2584)	1	0	0
E6(U4117)	1	0	0
B6(P2404)	1	0	0
E9(4068)	1	0	0
C9(P1814)	1	0	0
C8(T3532)	1	1	0
B9(T2436)	1	0	0
A3(H638)	1	0	0
B3(T2101)	1	1	0
A1(A2448)	1	2	0
C7(T3583)	1	1	0
D9(3362)	0	0	0
E4(T4242)	0	0	0
D8(T3382)	0	0	0
B2(E2758)	0	0	0
B4(U2816)	0	0	0
A8(P1885)	0	0	0
C6(T3044)	0	0	0
D5(T3677)	0	0	0
A9(P2733)	0	0	0
E3(T3979)	0	0	0
D3(3141)	0	0	0
E1(T4189)	0	0	0

Functional annotation of elements within integrons and CALIN elements revealed that in addition to *attC* sites, integrases, *attI* sites and promoters there are gene cassettes of antibiotic resistance associated genes, (Table 6.9). All integrons and CALIN elements present contain an antibiotic resistance gene. The antibiotic resistance genes annotated include extended spectrum beta-lactamases (VEB-PER), quinolone resistance gene (*qnr*), metallo beta-lactamase (IMP), class D beta-lactamases, beta lactamases (PSE), carbapenemases (CARB), chlor efflux pumps and *emrE* a multidrug resistance associated transporter.

Table 6.9. Integron elements present as predicted from sequenced genomes by Integron Finder

	Element size (bp)	**Strand**	**Element type**	**Annotation**
C7(T3583) Integron01 CALIN	671	1	protein	protein
	59	-1	attC	attC
	899	-1	protein	VEB-PER
	59	-1	attC	attC
	815	-1	protein	ANT3
	800	-1	protein	ClassD
	69	-1	attC	attC
C7(T3583) Integron_02 complete	110	-1	attC	attC
	914	-1	protein	CARB-PSE
	58	-1	attI	attI_1
	34	1	Promoter	Pint_1
	26	-1	Promoter	Pc_1
	1013	1	protein	intI
A1(A2448) integron_01 CALIN	59	1	attC	attC
	113	1	attC	attC
	1259	1	protein	Chlor_Efflux_Pump
	69	1	attC	attC
	800	1	protein	ClassD
	815	1	protein	ANT3
	59	1	attC	attC
A1(A2448) integron_02 complete	57	-1	attC	attC
	131	1	protein	protein
	132	-1	attC	attC
	740	-1	protein	IMP

	58	-1	attI	attI_1
	34	1	Promoter	Pint_1
	26	-1	Promoter	Pc_1
	1070	1	protein	intI
A1(A2448) integron03 CALIN	59	1	attC	attC
	1238	1	protein	emrE
B3(T2101) integron_01 complete	110	-1	attC	attC
	914	-1	protein	CARB-PSE
	58	-1	attI	attI_1
	34	1	Promoter	Pint_1
	26	-1	Promoter	Pc_1
	1013	1	protein	intI
B3(T2101) integron_02 CALIN	671	1	protein	protein
	59	-1	attC	attC
	899	-1	protein	VEB-PER
	59	-1	attC	attC
	815	-1	protein	ANT3
	800	-1	protein	ClassD
	69	-1	attC	attC
	476	-1	protein	protein
A3(H638) integron_01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	1208	-1	protein	protein
	899	1	protein	VEB-PER
	59	1	attC	attC
	113	1	attC	attC
	1259	1	protein	Chlor_Efflux_Pump
	69	1	attC	attC
	800	1	protein	ClassD
	815	1	protein	ANT3
	59	1	attC	attC
B9(T2436) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	1208	-1	protein	protein
	899	1	protein	VEB-PER
	59	1	attC	attC
	113	1	attC	attC
	1259	1	protein	Chlor_Efflux_Pump
	69	1	attC	attC
	800	1	protein	ClassD
	815	1	protein	ANT3
	59	1	attC	attC
C8(T3532) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1

	58	1	attI	attI_1
	203	-1	protein	protein
	60	1	attC	attC
	59	1	attC	attC
C8(T3532) integron_02 CALIN	59	-1	attC	attC
C9(P1814) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	203	-1	protein	protein
	60	1	attC	attC
	59	1	attC	attC
E9(4068) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	740	1	protein	IMP
	132	1	attC	attC
	59	1	attC	attC
	800	1	protein	ClassD
	71	1	attC	attC
B6(P2404) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	656	1	protein	Qnr
	71	1	attC	attC
	800	1	protein	ClassD
	815	1	protein	ANT3
	59	1	attC	attC
E6(U4117) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	656	1	protein	Qnr
	71	1	attC	attC
	800	1	protein	ClassD
B8(T2584) integron01 complete	1013	-1	protein	intI
	26	1	Promoter	Pc_1
	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	59	1	attC	attC
	1259	1	protein	Chlor_Efflux_Pump
	69	1	attC	attC
D1(T3354) integron01	1013	-1	protein	intI
	26	1	Promoter	Pc_1

complete	34	-1	Promoter	Pint_1
	58	1	attI	attI_1
	830	1	protein	ClassD
	89	1	attC	attC
	779	1	protein	ANT3
	59	1	attC	attC
	1259	1	protein	Chlor_Efflux_Pump
	69	1	attC	attC

Lineages suspected of being common source/infection carried the same integrons. Integron amino acid sequences for ST-235 isolates C8(T3532) and C9(P1814) intgegron01 respectively were extracted and determined to have no variable sites (figure 6.5). ST-708 isolates C7(T3583) and B3(T2101) integron02 and integron01 respectively amino acid sequences were extracted and determined to have no variable positions (figure 6.6).

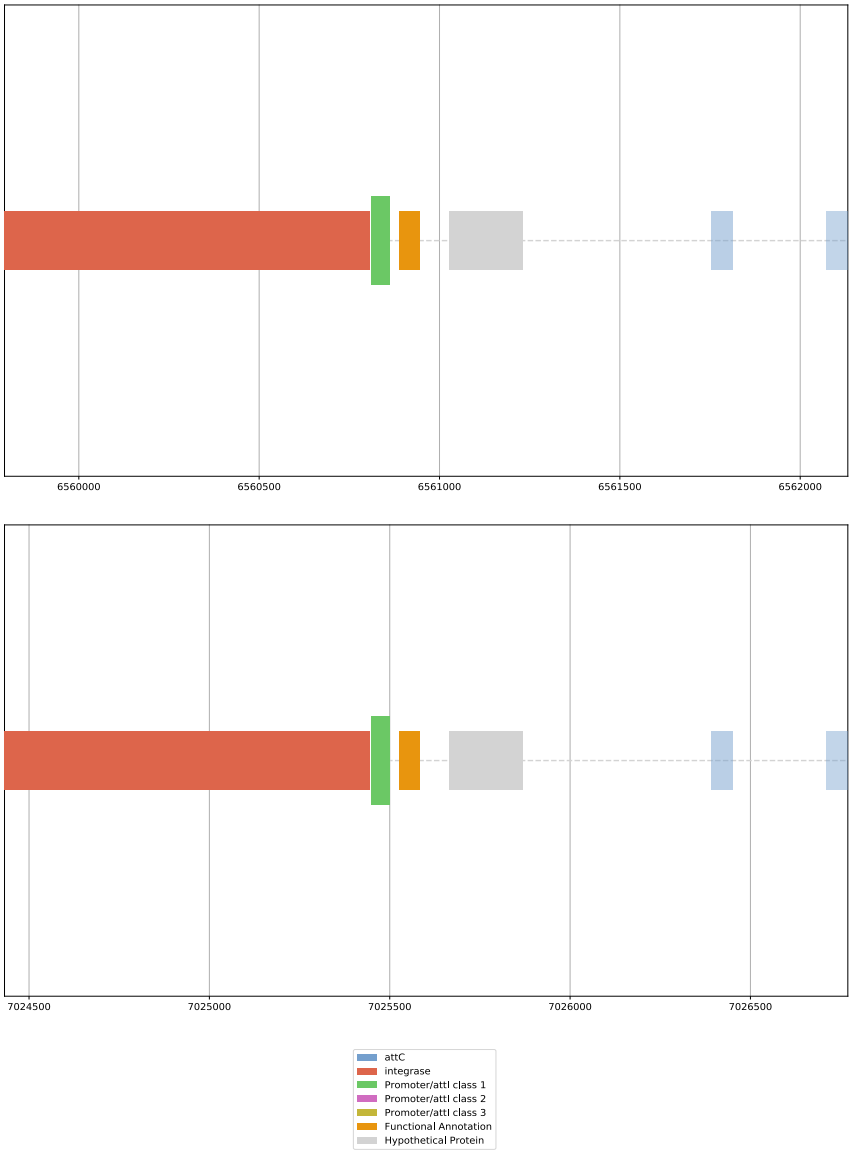


Figure 6.5. C8(T3532) integron01 (top) and C9(P1814) integron01 (bottom). Graphical representation of integron elements as determined by integron finder

Fgure 6.6. C7(T3583) integron02 (top) and B3(T2101) integron01 (bottom). Graphical representation of integron elements as determined by integron finder

6.4. Discussion

Using whole genome sequencing of a small sample of patient isolates whole genome sequenced clinically useful information could be determined and findings added to the knowledge of the genetic basis of antibiotic resistance in *P. aeruginosa* worldwide. While the genetic landscape of *P. aeruginosa* remains complex, it is increasingly being mapped[76,356,390,402,403]. In this chapter, genomic investigation has contributed to the knowledge of resistance genes in *P. aeruginosa* in a clinic in Thailand. Analysis of antibiotic resistance genes has shown whether they're integron associated; MLST and core genome phylogeny has demonstrated the strain diversity and the potential for a common source of infection, revealing the presence of at least one international MDR lineage.

Colistin was the only antibiotic to which none of the isolates were resistant. Numerous isolates were also multidrug resistant to the five antibiotics for which there was any resistance (A1U2448, B3T2101, B9T2436, E1T4189 and E9U4068). These were selected from a sample of diverse patients in terms of pathology, age and gender. The sensitivity in all isolates tested to colistin is reassuring but colistin is not only increasingly a 'last resort' antibiotic[67,404] but reports of *P. aeruginosa* and other bacteria acquiring resistance to polymyxins are emerging[391,395].

Determining the MLST's of the isolates sequenced (table 6.4), alongside phylogenetic placement (figure 6.3) demonstrated the diversity of the isolates but also that some were potentially closely related. Further, some ST's have been identified as overrepresented in the hospital setting and being associated

with multidrug resistance. The genetic background of MDR isolates is diverse[405] but lineages are emerging such as ST-244[406–408] and ST-235, identified internationally is an MDR lineage[392,394,395,404,409–411], though antibiograms may vary. The genome sequence data indicates that there have been at least two introductions of ST235's into the clinic and there is evidence for within-hospital transmission of one ST235 lineage carrying an antibiotic resistance integron. The contributions of various factors to growing antibiotic resistance are not assessed within this study but evidence is mounting elsewhere describing the extent of the problem in *P. aeruginosa* alone, that it's increasing and the potential for anthropogenic causes of increasing resistance[199,391,406,410,411].

In the CARD results (figure 6.5) *pmrB* is in many isolates flagged as a resistance gene, or for the mutations it carries, yet no isolates carry colistin resistance (figure 6.4). Entirely susceptible isolate E3(T3979) is reported to carry fifteen resistance conferring. Many of these genes such as *mexS*, *pmrA* and *pmrB* are present in most *P. aeruginosa* but their association with resistance is not always clear[402]. Aminoglycoside nucleotidyltransferase (ANT3) were contained in integrons carried by seven isolates, C7(T3583), A1(A2448), B3(T2101), A3(H6838), B9(T2436), B6(P2404) and D1(T3354) (Table 6.9). Aminoglycoside nucleotidyltransferase have been identified transferring between *Acinetobacter* spp. and confers resistance to streptomycin and spectinomycin[412,413]. Two isolate genomes C7(T3583) and B3(T2101) carry the same two integrons containing between them, an ambler class A beta-lactamase (ANT3), VEB-PER beta-lactamase genes and a gene encoding a chloramphenicol efflux pump. B3(T2101) displayed resistance to all antibiotics tested except colistin while

C7(T3583) displayed intermediary resistance to meropenem as well. Isolates A3(H6838) and B9(T2436) from the same patient also carried a VEB-PER and phenotypically A3(H638) was resistant to all antibiotics tested except colistin and B9(T2436) all except colistin and intermediary resistance to meropenem. Isolate genomes B6(P2404) and E6(U4117) both contain integrons that carry quinolone resistance gene *qnrVC1*. The *qnrVC1* gene has been observed in various bacteria such as *Vibrio cholerae* in Brazil[414], *E. coli* in Europe[415] and in *P. aeruginosa* previously[416–418]. Though B6(P2404) and E6(U4117) carry the *qnr* gene they display intermediary resistance to ciprofloxacin and susceptibility respectively (Figure 6.2). One isolate genome, that of A1(A2448), carried and integron with the *emrE* multidrug efflux encoding gene[419]. CARD analysis reports *emrE* presence in the majority genomes due to the commonly present *P. aeruginosa emrE* homolog *Pae-Emre*[420]. Acquired metallo-beta-lactamases (MBLs) in *P. aeruginosa* are associated with a heightened mortality rate[421], isolates A1(A2448) and E9(4068) were observed to have integrons containing IMP14 and IMP54. Both isolates were only susceptible to colistin including presenting resistance to meropenem.

6.4.1. Chapter Summary

Whole genome sequencing of isolates from Ramathibodi Hospital, Thailand successfully demonstrated:

- Antibiotic resistance, including multi-drug resistance is highly prevalent in the patient isolate population in this clinic
- International, high risk MDR lineages were present
- Potential for a common source of infection within the hospital or transmission between patients

Chapter Seven

7. General discussion

7.1. Comparative genomics to determine the genetic basis of phenotypic heterogeneity in *P. aeruginosa* CF lung infections

In CF the leading cause of overall morbidity and mortality is bacterial infections and in adult CF patients *P. aeruginosa* predominates[65]. As *P. aeruginosa* is an environmental bacterium that opportunistically infects CF patients the diversity of lineages infecting is expansive[75,195]. There are features of adaptation to various environmental niches and through due to interactions with other organisms that have selected for virulence factors that by chance are problematic in CF lung infections. Some *P. aeruginosa* lung infections adapt to the lung environment and establish chronic infections[161,285]. As an opportunistic pathogen the typical *P. aeruginosa* carries acute virulence features and the versatility due to complex gene regulation and a large genome and number of genes to establish a long-term infection[285,422]. The aim is to better understand what predisposes *P. aeruginosa* to successfully establishing in the lungs, the transition to chronicity and the changing state of pathogenicity along this evolutionary trajectory. Lineages have also independently evolved worldwide capable of cross infection between patients[208], patients harbouring these epidemic strains are not infected by distinct lineages but a single population that has emerged, already exposed to the CF lung environment and treatment regimes. Identifying commonality across *P. aeruginosa* CF lung infections is further confounded by multi-lineage infections, emerging sub-lineages, different treatment regimes across patients, host differences and genetic confounders as

neutral changes could be linked to adaptive changes[66]. This is all in the context of a changing lung environment and microbial community, much of which is not understood.

This study combined comparative genomics approaches applied to various CF associated *P. aeruginosa* genomes (Chapter Three) and the transmissible lineage, the LES (Chapter Four). Analysis was also expanded to non-CF bronchiectasis, a lung condition similar to CF that *P. aeruginosa* opportunistically infects (Chapter Five) and isolate genomes from a hospital in Thailand to assess antibiotic resistance, particularly MBE associated (Chapter Six). Genetic and phenotypic studies have observed that during CF infections *P. aeruginosa* can rapidly adapt by loss of function. Disruptive mutations and gene losses can generally be directly adaptive or result in remodeling of gene regulation[70], though no dedicated study so far has evaluated loss of function specifically across a diverse panel of CF isolates.

In Chapter Three the approach was to determine disruptive mutations and gene absence compared with PAO1 in CF and environmental genomes. It was determined which genes were overrepresented for loss of function, with the aim to identify those that might be biased for acquisition or have convergently adapted. Overall 90 genes were identified to be overrepresented for loss of function in CF encoding functionally diverse features. Taking the opposite approach it was then determine which genes are underrepresented for loss of function in CF, where loss of the gene product may more often be advantageous in various environments or at least constraints be relaxed more often than in

CF. Many genes identified to be overrepresented for loss of function were consistent with previous observations in CF. Loss of MucA results in the chronicity-associated change to the mucoidy phenotype and CF genomes were overrepresented for nonsense mutations and absences of its gene *mucA*. It has previously been observed *P. aeruginosa* may adapt to preference iron acquisition from heme rather than pyoverdine[326]. Consistent with this it was observed that pyovderine biosynthesis and receptor genes *fpvG*, *fpvA, pvdE, pvcC, pvdF, pvdA, fpvL, fpvR, pvdO, tonB, piuA, pvdS* and *pvdP* were each overrepresented for loss of function in CF (table 3.6). In the COST representative panel it was inferred that genes *mexB* and *mexA* of the *mexAB-oprM* efflux pump encoding operon were adaptive by loss of function (table 3.4). As such the entire operon was considered for overrepresentation for loss of function (table 3.6) and determined to be significantly overrepresented for loss of function. Because the encoded system is involved in multidrug resistance it has been suggested that it could be targeted as a treatment for *P. aeruginosa* infections[423].

Chronic lung diseases, CF, CF associated bronchiectasis, non-CF bronchiectasis, COPD and PCD include overlapping pathologies and the risk of *P. aeruginosa* lung infections. In PCD a study was conducted identifying by sequencing of infecting *P. aeruginosa* that there was shared mutational signatures between PCD and CF[364]. The aim in Chapter Five was to sequence infecting *P. aeruginosa* from multiple non-CF bronchiectasis patients to determine the between and within host genetic diversity, the potential for any overrepresentation of environmental lineages, the presence of transmissible lineages and to compare

adaptive mutations to CF. It was determined that there was probably common source or cross infection of some lineages (figure 5.5), multi-lineage infections (figure 5.2) and adaptive similarities with CF (table 5.5). The model for inference of adaptive loss of function developed (table 3.6), was expanded for the non-CF bronchiectasis dataset. The larger dataset afforded greater power to identify genes adaptive by loss of function and assess the potential impact of population structure (supplementary figure 9). Many genes were determined to be adaptive by loss of function in common with CF such as quorum sensing regulator *lasR* and *mucA*. Others that have been linked with adaptation in CF were observed infrequently or not at all such as *mexT, retS, exsD* and *ampR*[66,424]. Consistent with previous findings in CF[326] and the genes overrepresented for loss of function in Chapter Three in CF associated isolate genomes, convergence of loss of function in pyoverdine genes was inferred. In this representative panel of isolate genomes across multiple non-CF bronchiectasis patients *mexB* was identified to carry more independent nonsense mutations than any other gene, *mexA* and *oprM* were also among the genes that have most often acquired nonsense mutations (table 5.5).

In CF lung infections with *P. aeruginosa* transmissible lineages have independently evolved globally[208]. In Chapter Four geographically and historically diverse isolates of the LES were sequenced (table 4.2). The number and diversity of these genomes enabled detailed analysis of the lineages epidemiology, adaptation and potential origins. Previous studies have identified the LES as a monophyletic group and for those sampled in Canada and the U.K. to cluster separately[214,334]. For the first time the LES core genome phylogeny

was confidently rooted with a non-LES ougroup, the PAO1 genome. Further phylogeographic analysis could be used to assess whether the LES has its origins in Canada[425]. A number of discrepancies were identified within the genomic epidemiology but resolved with follow-up epidemiology of the patients backgrounds, demonstrating the power of this approach.

As the LES is a single population, multiple detailed approaches could be taken to identify potentially adaptive mutations: convergent evolution of non-synonymous mutations and maintenance and fixation of mutations since the common ancestor with LESB58. Genes *mucA* and *lasR,* previously associated with adaptation by loss of function in CF, were observed to be independently acquiring nonsense mutations and potentially being maintained following further transmission (table 4.7). The LES has been observed to be increasingly resistant to antibiotics[208,210]. Antibiotic resistance can be increased by disruption of porin encoding gene *oprD* . In the UK LES population *oprD* independently acquired nonsense mutations more than any other gene. Similarly disruption of genes *mpl* and *piuA* have been linked to increased resistance and were inferred to be adaptive by loss of function in the LES (table 4.7). Alternatively antibiotic efflux pump operon *mexAB-oprM* is disrupted in 32/122 UK LES genomes and only *oprD* and *lasR* independently acquired nonsense mutations more than *mexB*. Disruption of *mexAB-oprM*, particularly *mexB* is consistent with the findings in the IPCD CF genomes and the bronchiectasis genomes datasets in this study.

Genes *lasR, mexB* and *mpl* were identified to be adaptive by loss of function but also inferred to be under positive selection in the population (discounting the alleles with nonsense mutations) (table 4.13). Either the gene is under different selective pressures in different isolates or the inference of positive selection actually results from relaxed constraint or selection for disruptive missense mutations. Detailed study of potentially adaptive genes in transmissible lineages can also be informative to wider *P. aeruginosa* infections analogous to the study of within host, sequential isolate genomes[273,280,282].

Across chapters 3-5 genomes were analysed for evidence of loss of function mutations that were adaptive. The methods varied from identifying independent acquisitions of nonsense mutations across lineages in various chronic lung disease infections (Chapter 3, COST panel genomes), to identifying an association of nonsense mutations and larger loss of function events with acute and chronic CF infection isolates when compared with environmentally sampled isolate genomes (Chapter 3, IPCD genomes), independent acquisition of nonsense mutations in a single transmissible lineage infecting primarily CF patients and a few non-CF patients (Chapter 4, the LES genomes) and independent acquisitions across lineages of chronic lung disease non-CF bronchiectasis infection isolates (Chapter 5, non-CF bronchiectasis panel genomes). Table 7.1 displays the loss of function mutations inferred to be adaptive with any commonality across this study.

Table 7.1. Loss of function mutations inferred to be adaptive across chapters 3-5 present in at least two of the COST panel, IPCD overrepresented in CF, LES genomes or non-CF bronchiectasis datasets. PAO1 locus tags are used where there is not an accepted gene name across lineages on pseudomonas.com[290]. Loss of function of genes associated with CF compared with the environment included are all significant after BH multiple testing correction

	COST Panel (Chapter 3)	IPCD overrepresented in CF (Chapter 3)	LES (Chapter 4)	Bronchiectasis (Chapter 5)
lasR	+		+	
mucA	+		+	+
mexA	+			+
rbdA	+			+
mexB	+		+	+
PA4469	+		+	+
ladS	+			+
pchE			+	+

The genetic basis of antibiotic resistance in *P. aeruginosa* is complex and in CF increasing antibiotic resistance during chronic infections is typical[76,167,426,427]. In this study analysis of mutations have identified adaptations associated with antibiotic resistance such as positive selection in *pmr* genes, resistance associated with loss of function in *oprD*, *piuA* and *mpl*[328,329,356,391,399]. These results were mirrored in the analysis non-CF bronchiectasis associated isolate genomes. In Chapter Six a sample of clinical isolates from a range of patients and pathologies were obtained from a hospital in Bangkok, Thailand (Table 6.1). Except for complete colistin susceptibility in the isolates tested, antibiotic resistance was frequently observed in antibiotics commonly used in CF treatments, meropenem, ceftazidime, piperacillin/tazobactam, ciprofloxacin and tobramycin. By aligning the genomes to the CARD database resistance was

predicted but many genetic elements intrinsic to *P. aeruginosa* that don't always convey resistance were also identified, as demonstrated by the inclusion of reference strains PAO1, PA14 and LESB58 (figure 6.4). Genotyping of the strains and phylogenetic placement was informative, demonstrating the presence of international MDR lineage ST-235[76,409]. Integron Finder analysis revealed the ARGs associated with MBEs demonstrating the risk not only of vertical adaptation towards antibiotic resistance but the expansive potential of horizontally acquired resistance in *P. aeruginosa*.

P. aeruginosa infections are complex and predicting clinical outcomes remains difficult. Genome sequencing is making inroads into this complexity though, predicting antibiotic resistance[428], the potential for rapid typing of strains and detailed genomic epidemiology[194]. A wealth of information has been generated in this study and as Chapter Three demonstrates, more data can yield new insights into the potential genetic basis of pathogenicity. It also demonstrates that further sampling of environmental genomes would be beneficial for comparison. Many genes analysed were also of unknown function and often limited to comparison to the gene content of high quality reference genomes. GWAS has demonstrated the potential for associating genes and alleles with outcomes and in the future the predictive capacity of genes, alleles and their combinations could be assessed with machine learning. Higher order classifications too could be assessed such as transmissible/non-transmissible, impact on patient FEV1 or whether a lineage is predisposed to establishing a chronic infection.

8. Bibliography

1. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269,** 496–512 (1995).
2. Loman, N. J. & Pallen, M. J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **13,** 787–794 (2015).
3. Salzberg, S. L. *et al.* GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res.* **22,** 557–67 (2012).
4. Lin, Y. *et al.* Comparative studies of de novo assembly tools for next-generation sequencing technologies. *Bioinformatics* **27,** 2031–7 (2011).
5. Ruffalo, M., Laframboise, T. & Koyutürk, M. Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* **27,** 2790–2796 (2011).
6. Olson, N. D. *et al.* Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front. Genet.* **6,** 235 (2015).
7. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 1–2 (2014). doi:10.1093/bioinformatics/btu153
8. Stajich, J. E. & Lapp, H. Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief. Bioinform.* **7,** 287–296 (2006).
9. Chen, P. E. & Shapiro, B. J. The advent of genome-wide association studies for bacteria. *Curr. Opin. Microbiol.* **25,** 17–24 (2015).
10. Laabei, M. *et al.* Predicting the virulence of MRSA from its genome sequence. *Genome Res.* **24,** 839–849 (2014).
11. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat. Commun.* **6,** 10063 (2015).
12. Tang, P. & Gardy, J. L. Stopping outbreaks with real-time genomic epidemiology. *Genome Med.* **6,** 104 (2014).
13. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **13,** 601–612 (2012).
14. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2567–2572 (2005).
15. Varghese, N. J. *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* **43,** 6761–6771 (2015).
16. Vanlaere, E. *et al.* Taxon K, a complex within the Burkholderia cepacia complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. *Int. J. Syst. Evol. Microbiol.* **59,** 102–111 (2009).
17. Fournier, P.-E. & Raoult, D. Identification of rickettsial isolates at the species level using multi-spacer typing. *BMC Microbiol.* **7,** 72 (2007).
18. Maiden, M. & Bygraves, J. Multilocus sequence typing : A portable approach to the identification of clones within populations of pathogenic organisms. *Proc. …* **95,** 3140–3145 (1998).
19. Maiden, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 3140–3145 (1998).
20. Rincon, S. *et al.* Cefazolin high-inoculum effect in methicillin-susceptible Staphylococcus aureus from South American hospitals. *J. Antimicrob.*

Chemother. **68,** 2773–2778 (2013).
21. Kang, C.-I. *et al.* Clinical and Molecular Epidemiology of Community-Onset Bacteremia Caused by Extended-Spectrum β-Lactamase-Producing Escherichia coli over a 6-Year Period. *J. Korean Med. Sci.* **28,** 998–1004 (2013).
22. Zhang, Y.-J. *et al.* Serological and molecular capsular typing, antibiotic susceptibility and multilocus sequence typing of Streptococcus pneumoniae isolates from invasive and non-invasive infections. *Chin. Med. J. (Engl).* **126,** 2296–2303 (2013).
23. Mora, M., Veggi, D., Santini, L., Pizza, M. & Rappuoli, R. Reverse vaccinology. *Drug Discov. Today* **8,** 459–464 (2003).
24. Fournier, P.-E., Drancourt, M. & Raoult, D. Bacterial genome sequencing and its use in infectious diseases. *Lancet. Infect. Dis.* **7,** 711–723 (2007).
25. Maione, D. *et al.* Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* **309,** 148–150 (2005).
26. Moriel, D. G. *et al.* Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9072–9077 (2010).
27. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67,** 2640–2644 (2012).
28. Liu, B. & Pop, M. ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **37,** D443-7 (2009).
29. Chen, L. *et al.* VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **33,** (2005).
30. Topol, E. J., Murray, S. S. & Frazer, K. a. The genomics gold rush. *JAMA* **298,** 218–221 (2007).
31. Sheppard, S. K. *et al.* Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11923–7 (2013).
32. van Hemert, S. *et al.* Identification of Lactobacillus plantarum genes modulating the cytokine response of human peripheral blood mononuclear cells. *BMC Microbiol.* **10,** 293 (2010).
33. Salipante, S. J. *et al.* Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. *Genome Res.* **25,** 119–128 (2015).
34. Chaston, J. M., Newell, P. D. & Douglas, A. E. Metagenome-wide association of microbial determinants of host phenotype in drosophila melanogaster. *MBio* **5,** (2014).
35. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nat. Genet.* **45,** 1183–9 (2013).
36. Alam, M. T. *et al.* Dissecting vancomycin-intermediate resistance in staphylococcus aureus using genome-wide association. *Genome Biol. Evol.* **6,** 1174–1185 (2014).
37. Chewapreecha, C. *et al.* Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam Resistance within Pneumococcal Mosaic Genes. *PLoS Genet.* **10,** (2014).
38. Bayjanov, J. R., Starrenburg, M. J. C., van der Sijde, M. R., Siezen, R. J. & van Hijum, S. A. F. T. Genotype-phenotype matching analysis of 38

Lactococcus lactisstrains using random forest methods. *BMC Microbiol.* **13,** 68 (2013).
39. Versalovic, J. & Lupski, J. R. Molecular detection and genotyping : more accurate and of pathogens : rapid answers. *Trends Microbiol.* **10,** 15–21 (2002).
40. Sabat, A. J. *et al.* Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eur. Commun. Dis. Bull.* **18,** 20380 (2013).
41. Croucher, N. J. & Didelot, X. The application of genomics to tracing bacterial pathogen transmission. *Current Opinion in Microbiology* **23,** 62–67 (2015).
42. Eyre, D. W. *et al.* A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. *BMJ Open* **2,** e001124–e001124 (2012).
43. Halachev, M. R. *et al.* Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. *Genome Med.* **6,** 70 (2014).
44. Lewis, T. *et al.* High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. *J. Hosp. Infect.* **75,** 37–41 (2010).
45. Snitkin, E. S. *et al.* Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumoniae with Whole-Genome Sequencing. *Sci. Transl. Med.* **4,** 148ra116-148ra116 (2012).
46. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: A descriptive study. *Lancet Infect. Dis.* **13,** 130–136 (2013).
47. Köser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N.Engl.J.Med.* **366,** 2267–2275 (2012).
48. Eyre, D. W. *et al.* Diverse sources of C. difficile infection identified on whole-genome sequencing. *N. Engl. J. Med.* **369,** 1195–205 (2013).
49. Breurec, S. *et al.* Genomic epidemiology and global diversity of the emerging bacterial pathogen Elizabethkingia anophelis. *Sci. Rep.* **6,** 30379 (2016).
50. Perrin, A. *et al.* Evolutionary dynamics and genomic features of the Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. *Nat. Commun.* **8,** 15483 (2017).
51. Eyre, D. W. *et al.* Detection of Mixed Infection from Bacterial Whole Genome Sequence Data Allows Assessment of Its Role in Clostridium difficile Transmission. *PLoS Comput. Biol.* **9,** (2013).
52. Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol.* **16,** 114 (2015).
53. Cravens, H. Sewall Wright and Evolutionary Biology. By William B. Provine. (Chicago: University of Chicago Press, 1986. xvi + 545 pp. $30.00.). *J. Am. Hist.* **74,** 544–545 (1987).
54. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23,** 23–35 (1974).
55. Metzgar, D. & Wills, C. Evidence for the Adaptive Evolution of Mutation Rates. *Cell* **101,** 581–584 (2018).
56. Weigand, M. R. & Sundin, G. W. General and inducible hypermutation

facilitate parallel adaptation in Pseudomonas aeruginosa despite divergent mutation spectra. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13680–13685 (2012).
57. Imhof, M. & Schlotterer, C. Fitness effects of advantageous mutations in evolving Escherichia coli populations. *Proc. Natl. Acad. Sci.* **98,** 1113–1117 (2001).
58. Achtman, M. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annu. Rev. Microbiol.* **62,** 53–70 (2008).
59. Hanage, W. P., Fraser, C. & Spratt, B. G. The impact of homologous recombination on the generation of diversity in bacteria. *J. Theor. Biol.* **239,** 210–219 (2006).
60. Fraser, C., Hanage, W. P. & Spratt, B. G. Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl. Acad. Sci.* **102,** 1968–1973 (2005).
61. Pérez-Losada, M. *et al.* Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect. Genet. Evol.* **6,** 97–112 (2006).
62. Church, D., Elsayed, S., Reid, O., Winston, B. & Lindsay, R. Burn wound infections. *Clinical Microbiology Reviews* **19,** 403–434 (2006).
63. Hauser, A. R. & Ozer, E. A. Pseudomonas aeruginosa. *Nat. Rev. Microbiol.* **9,** 2011 (2011).
64. Fleiszig, S. M. J. & Evans, D. J. The pathogenesis of bacterial keratitis: Studies with Pseudomonas aeruginosa. *Clinical and Experimental Optometry* **85,** 271–278 (2002).
65. Understanding CF lung disease and its consequences. *Thorax* **64,** A109 LP-A114 (2009).
66. Winstanley, C., O'Brien, S. & Brockhurst, M. A. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends in Microbiology* **24,** 327–337 (2016).
67. Taccetti, G., Campana, S., Festini, F., Mascherini, M. & Döring, G. Early eradication therapy against Pseudomonas aeruginosa in cystic fibrosis patients. *Eur. Respir. J.* **26,** 458–61 (2005).
68. Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B. & Gibson, R. L. Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr. Pulmonol.* **34,** 91–100 (2002).
69. De Soyza, A., Brown, J. S. & Loebinger, M. R. Research priorities in bronchiectasis. *Thorax* **68,** 695–696 (2013).
70. Girard, G. & Bloemberg, G. V. Central role of quorum sensing in regulating the production of pathogenicity factors in Pseudomonas aeruginosa. *Future Microbiol.* **3,** 97–106 (2008).
71. Stover, C. K. *et al.* Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* **406,** 959–64 (2000).
72. He, J. *et al.* The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2530–5 (2004).
73. Stewart, L. *et al.* Draft genomes of 12 host-adapted and environmental isolates of Pseudomonas aeruginosa and their positions in the core genome phylogeny. *Pathog. Dis.* **71,** 20–25 (2014).

74. Roy, P. H. *et al.* Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. *PLoS One* **5,** e8842 (2010).
75. Freschi, L. *et al.* Clinical utilization of genomics data produced by the international Pseudomonas aeruginosa consortium. *Front. Microbiol.* **6,** 1–8 (2015).
76. Kos, V. N. *et al.* The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrob. Agents Chemother.* **59,** 427–36 (2015).
77. Mosquera-Rendón, J. *et al.* Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species. *BMC Genomics* **17,** 45 (2016).
78. Valot, B. *et al.* What It Takes to Be a Pseudomonas aeruginosa? The Core Genome of the Opportunistic Pathogen Updated. *PLoS One* **10,** e0126468 (2015).
79. Pohl, S. *et al.* The extensive set of accessory Pseudomonas aeruginosa genomic components. *FEMS Microbiol. Lett.* 1–7 (2014). doi:10.1111/1574-6968.12445
80. Winsor, G. L. *et al.* Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Res.* **39,** D596-600 (2011).
81. Galán-Vásquez, E., Luna, B. & Martínez-Antonio, A. The Regulatory Network of Pseudomonas aeruginosa. *Microb. Inform. Exp.* **1,** 3 (2011).
82. Balasubramanian, D., Schneper, L., Kumari, H. & Mathee, K. A dynamic and intricate regulatory network determines Pseudomonas aeruginosa virulence. *Nucleic Acids Res.* **41,** 1–20 (2013).
83. Schulz, S. *et al.* Elucidation of sigma factor-associated networks in Pseudomonas aeruginosa reveals a modular architecture with limited and function-specific crosstalk. *PLoS pathogens* **11,** e1004744 (2015).
84. Rodrigue, A., Quentin, Y., Lazdunski, A., Méjean, V. & Foglino, M. Two-component systems in Pseudomonas aeruginosa: why so many? *Trends Microbiol.* **8,** 498–504 (2000).
85. Sivaneson, M., Mikkelsen, H., Ventre, I., Bordi, C. & Filloux, A. Two-component regulatory systems in Pseudomonas aeruginosa: An intricate network mediating fimbrial and efflux pump gene expression. *Mol. Microbiol.* **79,** 1353–1366 (2011).
86. Otto, M. Bacterial sensing of antimicrobial peptides. *Contrib. Microbiol.* **16,** 136–149 (2009).
87. Wurtzel, O. *et al.* The Single-Nucleotide Resolution Transcriptome of Pseudomonas aeruginosa Grown in Body Temperature. *PLoS Pathog.* **8,** (2012).
88. Repoila, F. & Darfeuille, F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol. Cell* **101,** 117–131 (2009).
89. Sonnleitner, E. & Haas, D. Small RNAs as regulators of primary and secondary metabolism in Pseudomonas species. *Applied Microbiology and Biotechnology* **91,** 63–79 (2011).
90. Gómez-Lozano, M., Marvig, R. L., Molin, S. & Long, K. S. Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa. *Environ. Microbiol.* **14,** 2006–2016 (2012).

91. Rahme, L. G. *et al.* Plants and animals share functionally common bacterial virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8815–8821 (2000).
92. Coleman, F. T. *et al.* Hypersusceptibility of cystic fibrosis mice to chronic Pseudomonas aeruginosa oropharyngeal colonization and lung infection. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1949–54 (2003).
93. Hrabak, E. M. & Willis, D. K. The lemA gene required for pathogenicity of Pseudomonas syringae pv. syringae on bean is a member of a family of two-component regulators. *J. Bacteriol.* **174,** 3011–3020 (1992).
94. Goodman, a L. *et al.* A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in Pseudomonas aeruginosa. *Dev Cell* **7,** 745–754 (2004).
95. Goodman, A. L. *et al.* Direct interaction between sensor kinase proteins mediates acute and chronic disease phenotypes in a bacterial pathogen. *Genes Dev.* **23,** 249–259 (2009).
96. Ventre, I. *et al.* Multiple sensors control reciprocal expression of Pseudomonas aeruginosa regulatory RNA and virulence genes. *Proc. Natl. Acad. Sci.* **103,** 171–176 (2006).
97. Heeb, S. & Haas, D. Regulatory Roles of the GacS/GacA Two-Component System in Plant-Associated and Other Gram-Negative Bacteria. *Mol. Plant-Microbe Interact.* **14,** 1351–1363 (2001).
98. Wolfgang, M. C., Lee, V. T., Gilmore, M. E. & Lory, S. Coordinate regulation of bacterial virulence genes by a novel adenylate cyclase-dependent signaling pathway. *Developmental Cell* **4,** 253–263 (2003).
99. Brencic, A. & Lory, S. Determination of the regulon and identification of novel mRNA targets of Pseudomonas aeruginosa RsmA. *Mol. Microbiol.* **72,** 612–632 (2009).
100. Linares, J. F., Gustafsson, I., Baquero, F. & Martinez, J. L. Antibiotics as intermicrobial signaling agents instead of weapons. *Proc. Natl. Acad. Sci.* **103,** 19484–19489 (2006).
101. Eberhard, A. Inhibition and activation of bacterial luciferase synthesis. *J. Bacteriol.* **109,** 1101–1105 (1972).
102. Nealson, K. H., Platt, T. & Hastings, J. W. Cellular control of the synthesis and activity of the bacterial luminescent system. *J. Bacteriol.* **104,** 313–22 (1970).
103. Meighen, E. A. Bacterial bioluminescence: organization, regulation, and application of the lux genes. *FASEB J.* **7,** 1016–22 (1993).
104. Kiratisin, P., Tucker, K. D. & Passador, L. LasR, a transcriptional activator of Pseudomonas aeruginosa virulence genes, functions as a multimer. *J. Bacteriol.* **184,** 4912–4919 (2002).
105. Jones, S. *et al.* The lux autoinducer regulates the production of exoenzyme virulence determinants in Erwinia carotovora and Pseudomonas aeruginosa. *EMBO J.* **12,** 2477–82 (1993).
106. Pearson, J. P., Passador, L., Iglewski, B. H. & Greenberg, E. P. A second N-acylhomoserine lactone signal produced by Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci.* **92,** 1490–1494 (1995).
107. Ochsner, U. A. & Reiser, J. Autoinducer-mediated regulation of rhamnolipid biosurfactant synthesis in Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci.* **92,** 6424–6428 (1995).
108. Pesci, E. C. *et al.* Quinolone signaling in the cell-to-cell communication

system of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci.* **96,** 11229–11234 (1999).
109. Gallagher, L. A., McKnight, S. L., Kuznetsova, M. S., Pesci, E. C. & Manoil, C. Functions required for extracellular quinolone signaling by Pseudomonas aeruginosa. *J. Bacteriol.* **184,** 6472–6480 (2002).
110. Schertzer, J. W., Boulette, M. L. & Whiteley, M. More than a signal: non-signaling properties of quorum sensing molecules. *Trends Microbiol.* **17,** 189–195 (2009).
111. Cao, H. *et al.* A quorum sensing-associated virulence gene of Pseudomonas aeruginosa encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. *Proc. Natl. Acad. Sci.* **98,** 14613–14618 (2001).
112. Lee, J. & Zhang, L. The hierarchy quorum sensing network in Pseudomonas aeruginosa. *Protein Cell* **6,** 26–41 (2014).
113. Calfee, M. W., Coleman, J. P. & Pesci, E. C. Interference with Pseudomonas quinolone signal synthesis inhibits virulence factor expression by Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 11633–7 (2001).
114. Lee, J. *et al.* A cell-cell communication signal integrates quorum sensing and stress response. *Nat. Chem. Biol.* **9,** 339–343 (2013).
115. Chugani, S. & Greenberg, E. P. The influence of human respiratory epithelia on Pseudomonas aeruginosa gene expression. *Microb. Pathog.* **42,** 29–35 (2007).
116. Ochsner, U. A., Wilderman, P. J., Vasil, A. I. & Vasil, M. L. GeneChip® expression analysis of the iron starvation response in Pseudomonas aeruginosa: Identification of novel pyoverdine biosynthesis genes. *Mol. Microbiol.* **45,** 1277–1287 (2002).
117. Castric, P. a. Hydrogen cyanide production by Pseudomonas aeruginosa at reduced oxygen levels. *Can. J. Microbiol.* **29,** 1344–1349 (1983).
118. Strempel, N. *et al.* Human host defense peptide LL-37 stimulates virulence factor production and adaptive resistance in Pseudomonas aeruginosa. *PLoS One* **8,** e82240 (2013).
119. Smith, R. S. & Iglewski, B. H. Pseudomonas aeruginosa quorum sensing as a potential antimicrobial target. *Journal of Clinical Investigation* **112,** 1460–1465 (2003).
120. Hickey, W. J. & Focht, D. D. Degradation of mono-, di-, and trihalogenated benzoid acids by Pseudomonas aeruginosa JB2. *Appl. Environ. Microbiol.* **56,** 3842–3850 (1990).
121. Marín, M. M., Yuste, L. & Rojo, F. Differential expression of the components of the two alkane hydroxylases from Pseudomonas aeruginosa. *J. Bacteriol.* **185,** 3232–3237 (2003).
122. Bartell, J. A. *et al.* Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate virulence factor synthesis. *Nat. Commun.* **8,** 14631 (2017).
123. Budzikiewicz, H. Secondary metabolites from fluorescent pseudomonads. *FEMS Microbiol. Lett.* **104,** 209–228 (1993).
124. WELLS, I. C. Antibiotic substances produced by Pseudomonas aeruginosa; syntheses of Pyo Ib, Pyo Ic, and Pyo III. *J. Biol. Chem.* **196,** 331–340 (1952).

125. Meyer, J. M. *et al.* Use of siderophores to type pseudomonads: The three Pseudomonas aeruginosa pyoverdine systems. *Microbiology* **143,** 35–43 (1997).
126. Lau, G. W., Hassett, D. J., Ran, H. & Kong, F. The role of pyocyanin in Pseudomonas aeruginosa infection. *Trends in Molecular Medicine* **10,** 599–606 (2004).
127. Wilson, R. *et al.* Measurement of Pseudomonas aeruginosa phenazine pigments in sputum and assessment of their contribution to sputum sol toxicity for respiratory epithelium. *Infect. Immun.* **56,** 2515–2517 (1988).
128. Reimer, a, Edvaller, B. & Johansson, B. Concentrations of the Pseudomonas aeruginosa toxin pyocyanin in human ear secretions. *Acta Otolaryngol. Suppl.* **543,** 86–88 (2000).
129. Lau, G. W., Ran, H., Kong, F., Hassett, D. J. & Mavrodi, D. Pseudomonas aeruginosa pyocyanin is critical for lung infection in mice. *Infect. Immun.* **72,** 4275–4278 (2004).
130. Lau, G. W. *et al.* The Drosophila melanogaster toll pathway participates in resistance to infection by the gram-negative human pathogen Pseudomonas aeruginosa. *Infect. Immun.* **71,** 4059–4066 (2003).
131. Rahme, L. G. *et al.* Use of model plant hosts to identify Pseudomonas aeruginosa virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13245–50 (1997).
132. Mahajan-Miklos, S., Tan, M. W., Rahme, L. G. & Ausubel, F. M. Molecular mechanisms of bacterial virulence elucidated using a Pseudomonas aeruginosa-Caenorhabditis elegans pathogenesis model. *Cell* **96,** 47–56 (1999).
133. Andrews, S. C., Robinson, A. K. & Rodriguez-Quinones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **27,** 215–237 (2003).
134. Brandel, J. *et al.* Pyochelin, a siderophore of Pseudomonas aeruginosa: physicochemical characterization of the iron(III), copper(II) and zinc(II) complexes. *Dalton Trans.* **41,** 2820–2834 (2012).
135. Dumas, Z., Ross-Gillespie, A. & Kummerli, R. Switching between apparently redundant iron-uptake mechanisms benefits bacteria in changeable environments. *Proc. R. Soc. B Biol. Sci.* **280,** 20131055–20131055 (2013).
136. Cornelis, P., Hohnadel, D. & Meyer, J. M. Evidence for different pyoverdine-mediated iron uptake systems among Pseudomonas aeruginosa strains. *Infect. Immun.* **57,** 3491–3497 (1989).
137. de Chial, M. *et al.* Identification of type II and type III pyoverdine receptors from Pseudomonas aeruginosa. *Microbiology* **149,** 821–831 (2003).
138. Hartney, S. L. *et al.* TonB-dependent outer-membrane proteins and siderophore utilization in Pseudomonas fluorescens Pf-5. *BioMetals* **24,** 193–213 (2011).
139. Ochsner, U. A., Johnson, Z. & Vasil, M. L. Genetics and regulation of two distinct haem-uptake systems, phu and has, in Pseudomonas aeruginosa. *Microbiology* **146,** 185–198 (2000).
140. Filloux, A. Protein secretion systems in Pseudomonas aeruginosa: An essay on diversity, evolution, and function. *Frontiers in Microbiology* **2,** (2011).

141. Holland, I. B., Schmitt, L. & Young, J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol. Membr. Biol.* **22,** 29–39 (2005).
142. Bleves, S. *et al.* Protein secretion systems in Pseudomonas aeruginosa: A wealth of pathogenic weapons. *International Journal of Medical Microbiology* **300,** 534–543 (2010).
143. König, B., Jaeger, K. E., Sage, A. E., Vasil, M. L. & König, W. Role of Pseudomonas aeruginosa lipase in inflammatory mediator release from human inflammatory effector cells (platelets, granulocytes, and monocytes). *Infect. Immun.* **64,** 3252–3258 (1996).
144. Azghani, A. O. Pseudomonas aeruginosa and Epithelial Permeability: Role of Virulence Factors Elastase and Exotoxin A. *Am. J. Respir. Cell Mol. Biol.* **15,** 132–140 (1996).
145. Frank, D. W. The exoenzyme S regulon of Pseudomonas aeruginosa. *Mol. Microbiol.* **26,** 621–629 (1997).
146. Duong, F. *et al.* The AprX protein of Pseudomonas aeruginosa: A new substrate for the Apr type I secretion system. *Gene* **262,** 147–153 (2001).
147. Yang, L. *et al.* Distinct roles of extracellular polymeric substances in Pseudomonas aeruginosa biofilm development. *Environ. Microbiol.* **13,** 1705–1717 (2011).
148. Costerton, J. W. Bacterial Biofilms: A Common Cause of Persistent Infections. *Science (80-.).* **284,** 1318–1322 (1999).
149. Wei, Q. & Ma, L. Z. Biofilm matrix and its regulation in Pseudomonas aeruginosa. *International Journal of Molecular Sciences* **14,** 20983–21005 (2013).
150. Parsek, M. R. & Peter Greenberg, E. [3] Quorum sensing signals in development of Pseudomonas aeruginosa biofilms. *Methods Enzymol.* **310,** 43–55 (1999).
151. Davies, D. G. *et al.* The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm. *Science (80-.).* **280,** 295 LP-298 (1998).
152. Franklin, M. J., Nivens, D. E., Weadge, J. T. & Lynne Howell, P. Biosynthesis of the pseudomonas aeruginosa extracellular polysaccharides, alginate, Pel, and Psl. *Frontiers in Microbiology* **2,** (2011).
153. Ma, L., Lu, H., Sprinkle, A., Parsek, M. R. & Wozniak, D. J. Pseudomonas aeruginosa Psl is a galactose- and mannose-rich exopolysaccharide. in *Journal of Bacteriology* **189,** 8353–8356 (2007).
154. Heurlier, K. *et al.* Quorum-sensing-negative (lasR) mutants of Pseudomonas aeruginosa avoid cell lysis and death. *J. Bacteriol.* **187,** 4875–4883 (2005).
155. Moskowitz, S. M., Foster, J. M., Emerson, J. C., Gibson, R. L. & Burns, J. L. Use of Pseudomonas biofilm susceptibilities to assign simulated antibiotic regimens for cystic fibrosis airway infection. *J. Antimicrob. Chemother.* **56,** 879–886 (2005).
156. Yau, Y. C. W. *et al.* Randomized controlled trial of biofilm antimicrobial susceptibility testing in cystic fibrosis patients. *J. Cyst. Fibros.* **14,** 262–6 (2015).
157. Høiby, N. *et al.* The clinical impact of bacterial biofilms. *Int J Oral Sci Int. J. Oral Sci.* **3,** 55–65 (2011).

158. Bjarnsholt, T. *et al.* Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr. Pulmonol.* **44,** 547–558 (2009).
159. Borriello, G. *et al.* Oxygen limitation contributes to antibiotic tolerance of Pseudomonas aeruginosa in biofilms. *Antimicrob. Agents Chemother.* **48,** 2659–2664 (2004).
160. DOGGETT, R. G., HARRISON, G. M. & WALLIS, E. S. COMPARISON OF SOME PROPERTIES OF PSEUDOMONAS AERUGINOSA ISOLATED FROM INFECTIONS IN PERSONS WITH AND WITHOUT CYSTIC FIBROSIS. *J. Bacteriol.* **87,** 427–431 (1964).
161. Mathee, K. *et al.* Mucoid conversion of Pseudomonas aeruginosa by hydrogen peroxide: A mechanism for virulence activation in the cystic fibrosis lung. *Microbiology* **145,** 1349–1357 (1999).
162. Cobb, L. M., Mychaleckyj, J. C., Wozniak, D. J. & Lopez-Boado, Y. S. Pseudomonas aeruginosa Flagellin and Alginate Elicit Very Distinct Gene Expression Patterns in Airway Epithelial Cells: Implications for Cystic Fibrosis Disease. *J. Immunol.* **173,** 5659–5670 (2004).
163. Simpson, J. A., Smith, S. E. & Dean, R. T. Scavenging by alginate of free radicals released by macrophages. *Free Radic. Biol. Med.* **6,** 347–353 (1989).
164. Boyd, A. & Chakrabarty, A. M. Pseudomonas aeruginosa biofilms: role of the alginate exopolysaccharide. *J. Ind. Microbiol.* **15,** 162–168 (1995).
165. Wozniak, D. J. *et al.* Alginate is not a significant component of the extracellular polysaccharide matrix of PA14 and PAO1 Pseudomonas aeruginosa biofilms. *Proc. Natl. Acad. Sci.* **100,** 7907–7912 (2003).
166. Govan, J. & Deretic, V. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiol. Rev.* **60,** (1996).
167. Breidenstein, E. B. M., de la Fuente-Núñez, C. & Hancock, R. E. W. Pseudomonas aeruginosa: All roads lead to resistance. *Trends in Microbiology* **19,** 419–426 (2011).
168. Cao, L., Srikumar, R. & Poole, K. MexAB-OprM hyperexpression in NalC-type multidrug-resistant Pseudomonas aeruginosa: Identification and characterization of the nalC gene encoding a repressor of PA3720-PA3719. *Mol. Microbiol.* **53,** 1423–1436 (2004).
169. Li, X. Z., Nikaido, H. & Poole, K. Role of MexA-MexB-OprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **39,** 1948–1953 (1995).
170. Köhler, T., Epp, S. F., Curty, L. K. & Pechère, J. C. Characterization of MexT, the regulator of the MexE-MexF-OprN multidrug efflux system of Pseudomonas aeruginosa. *J. Bacteriol.* **181,** 6300–6305 (1999).
171. Morita, Y., Tomida, J. & Kawamura, Y. Mexxy multidrug efflux system of Pseudomonas aeruginosa. *Frontiers in Microbiology* **3,** (2012).
172. Lambert, P. Mechanisms of antibiotic resistance in Pseudomonas aeruginosa. *J. R. Soc. Med.* (2002).
173. Drudge, C. N. *et al.* Diversity of integron- and culture-associated antibiotic resistance genes in freshwater floc. *Appl. Environ. Microbiol.* **78,** 4367–4372 (2012).
174. Perez, A. *et al.* CFTR inhibition mimics the cystic fibrosis inflammatory profile. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **292,** L383–L395 (2007).

175. Smith, J. J., Travis, S. M., Greenberg, E. P. & Welsh, M. J. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* **85,** 229–236 (1996).
176. Goss, C. H. & Burns, J. L. Exacerbations in cystic fibrosis · 1: Epidemiology and pathogenesis. *Thorax* **62,** 360 (2007).
177. Regelmann, W. E., Elliott, G. R., Warwick, W. J. & Clawson, C. C. Reduction of sputum Pseudomonas aeruginosa density by antibiotics improves lung function in cystic fibrosis more than do bronchodilators and chest physiotherapy alone. *Am. Rev. Respir. Dis.* **141,** 914–21 (1990).
178. Knowles, M. *et al.* Abnormal ion permeation through cystic fibrosis respiratory epithelium. *Science (80-.).* **221,** 1067–1070 (1983).
179. Elborn, J. S. Cystic fibrosis. *The Lancet* **388,** 2519–2531 (2016).
180. Rogers, G. B. *et al.* Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. *J. Clin. Microbiol.* **41,** 3548–3558 (2003).
181. Worlitzsch, D. *et al.* Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *J. Clin. Invest.* **109,** 317–325 (2002).
182. Filkins, L. M. & O'Toole, G. A. Cystic Fibrosis Lung Infections: Polymicrobial, Complex, and Hard to Treat. *PLoS Pathogens* **11,** (2015).
183. FitzSimmons, S. The changing epidemiology of cystic fibrosis. *J. Pediatr.* **122,** (1993).
184. Razvi, S. *et al.* Respiratory microbiology of patients with cystic fibrosis in the United States, 1995 to 2005. *Chest* **136,** 1554–60 (2009).
185. Ruzal-Shapiro, C. Cystic fibrosis: An overview. *Radiologic Clinics of North America* **36,** 143–161 (1998).
186. Burns, J. L. *et al.* Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. *J. Infect. …* **183,** 444–52 (2001).
187. Doring, G. *et al.* Antibiotic therapy against Pseudomonas aeruginosa in cystic fibrosis: A European consensus. *Eur. Respir. J.* **16,** 749–767 (2000).
188. Pier, G. B. CFTR mutations and host susceptibility to Pseudomonas aeruginosa lung infection. *Current Opinion in Microbiology* **5,** 81–86 (2002).
189. Brennan, A. L. & Geddes, D. M. Cystic fibrosis. *Curr. Opin. Infect. Dis.* **15,** 175–82 (2002).
190. Koch, C. Early infection and progression of cystic fibrosis lung disease. *Pediatr. Pulmonol.* **34,** 232–236 (2002).
191. Tingpej, P. *et al.* Phenotypic characterization of clonal and nonclonal Pseudomonas aeruginosa strains isolated from lungs of adults with cystic fibrosis. *J. Clin. Microbiol.* **45,** 1697–1704 (2007).
192. Cheng, K. *et al.* Spread of ?-lactam-resistant Pseudomonas aeruginosa in a cystic fibrosis clinic. *The Lancet* **348,** 639–642 (1996).
193. Speert, D. P. *et al.* Epidemiology of Pseudomonas aeruginosa in cystic fibrosis in British Columbia, Canada. *Am J Respir Crit Care Med* **166,** 988–93. (2002).
194. Dinesh, S. D., Grundmann, H., Pitt, T. L. & Römling, U. European-wide distribution of Pseudomonas aeruginosa clone C. *Clin. Microbiol. Infect.* **9,** 1228–1233 (2003).

195. Pirnay, J. P. *et al.* Pseudomonas aeruginosa population structure revisited. *PLoS One* **4,** (2009).
196. Blumer, J. L., Saiman, L., Konstan, M. W. & Melnick, D. The efficacy and safety of meropenem and tobramycin vs ceftazidime and tobramycin in the treatment of acute pulmonary exacerbations in patients with cystic fibrosis. *Chest* **128,** 2336–2346 (2005).
197. Hancock, R. E. & Speert, D. P. Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and impact on treatment. *Drug Resist. Updat.* **3,** 247–255 (2000).
198. Li, Z. *et al.* Longitudinal development of mucoid Pseudomonas aeruginosa infection and lung disease progression in children with cystic fibrosis. *Jama* **293,** 581–8 (2005).
199. Li, J. *et al.* Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infectious Diseases* **6,** 589–601 (2006).
200. Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B. & Jackson, R. W. Pseudomonas genomes: Diverse and adaptable. *FEMS Microbiol. Rev.* **35,** 652–680 (2011).
201. Jones, A. M. *et al.* Spread of a multiresistant strain of Pseudomonas aeruginosa in an adult cystic fibrosis clinic. *Lancet* **358,** 557–558 (2001).
202. Thomassen, M. J., Demko, C. A., Doershuk, C. F. & Root, J. M. Pseudomonas aeruginosa isolates: Comparisons of isolates from campers and from sibling pairs with cystic fibrosis. *Pediatr. Pulmonol.* **1,** 40–45 (1985).
203. Grothues, D., Koopmann, U., Von der Hardt, H. & Tummler, B. Genome fingerprinting of Pseudomonas aeruginosa indicates colonization of cystic fibrosis siblings with closely related strains. *J. Clin. Microbiol.* **26,** 1973–1977 (1988).
204. Wolz, C. *et al.* Pseudomonas aeruginosa cross-colonization and persistence in patients with cystic fibrosis. Use of a DNA probe. *Epidemiol Infect* **102,** 205–214 (1989).
205. Livesley, M. A. *et al.* Subspecific differentiation of Burkholderia cepacia isolates in cystic fibrosis. *J. Med. Microbiol.* **47,** 999–1006 (1998).
206. Panagea, S., Winstanley, C. & Parsons, Y. PCR-based detection of a cystic fibrosis epidemic strain of Pseudomonas aeruginosa. *Mol. …* **7,** 195–200 (2003).
207. Kakinuma, Y. *et al.* Molecular conservation within LES9F and PS21 liverpool epidemic strain (LES) markers in wild-type clinical pseudomonas aeruginosa isolated from the sputum of adult patients with cystic fibrosis. *Br. J. Biomed. Sci.* **67,** 87–88 (2010).
208. Fothergill, J. L., Walshaw, M. J. & Winstanley, C. Transmissible strains of Pseudomonas aeruginosa in cystic fibrosis lung infections. *Eur. Respir. J.* **40,** 227–38 (2012).
209. Al-Aloul, M. Increased morbidity associated with chronic infection by an epidemic Pseudomonas aeruginosa strain in CF patients. *Thorax* **59,** 334–336 (2004).
210. Ashish, A., Shaw, M., Winstanley, C., Ledson, M. J. & Walshaw, M. J. Increasing resistance of the Liverpool Epidemic Strain (LES) of Pseudomonas aeruginosa (Psa) to antibiotics in cystic fibrosis (CF)-A cause for concern? *J. Cyst. Fibros.* **11,** 173–179 (2012).

211. Scott, F. W. Identification and characterization of transmissible Pseudomonas aeruginosa strains in cystic fibrosis patients in England and Wales. *J. Med. Microbiol.* **53,** 609–615 (2004).
212. Naughton, S. *et al.* Pseudomonas aeruginosa AES-1 exhibits increased virulence gene expression during chronic infection of cystic fibrosis lung. *PLoS One* **6,** e24526 (2011).
213. Tingpej, P. *et al.* Clinical profile of adult cystic fibrosis patients with frequent epidemic clones of Pseudomonas aeruginosa. *Respirology* **15,** 923–9 (2010).
214. Dettman, J. R., Rodrigue, N., Aaron, S. D. & Kassen, R. Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 21065–70 (2013).
215. Duong, J. *et al.* Phenotypic and genotypic comparison of epidemic and non-epidemic strains of pseudomonas aeruginosa from individuals with cystic fibrosis. *PLoS One* **10,** (2015).
216. Van Mansfeld, R. *et al.* Within-Host evolution of the Dutch high-prevalent pseudomonas aeruginosa clone ST406 during Chronic colonization of a patient with cystic fibrosis. *PLoS One* **11,** (2016).
217. Kaca, W., Drulis-kawa, Z., Stoitsova, S. R. & Toth, V. Developing an international Pseudomonas aeruginosa reference panel . 1–36
218. Smart, C. H. M., Walshaw, M. J., Hart, C. A. & Winstanley, C. Use of suppression subtractive hybridization to examine the accessory genome of the Liverpool cystic fibrosis epidemic strain of Pseudomonas aeruginosa. *J. Med. Microbiol.* **55,** 677–88 (2006).
219. Parsons, Y. N. *et al.* Use of subtractive hybridization to identify a diagnostic probe for a cystic fibrosis epidemic strain of Pseudomonas aeruginosa. *J. Clin. Microbiol.* **40,** 4607–4611 (2002).
220. Fothergill, J. L., Upton, A. L., Pitt, T. L., Hart, C. A. & Winstanley, C. Diagnostic multiplex PCR assay for the identification of the Liverpool, Midlands 1 and Manchester CF epidemic strains of Pseudomonas aeruginosa. *J. Cyst. Fibros.* **7,** 258–261 (2008).
221. Lewis, D. a *et al.* Identification of DNA markers for a transmissible Pseudomonas aeruginosa cystic fibrosis strain. *Am. J. Respir. Cell Mol. Biol.* **33,** 56–64 (2005).
222. Smart, C. H. M. *et al.* Development of a diagnostic test for the Midlands 1 cystic fibrosis epidemic strain of Pseudomonas aeruginosa. *J. Med. Microbiol.* **55,** 1085–91 (2006).
223. Williams, H. L. *et al.* A diagnostic PCR assay for the detection of an Australian epidemic strain of Pseudomonas aeruginosa. *Ann. Clin. Microbiol. Antimicrob.* **9,** 18 (2010).
224. Fothergill, J. L. *et al.* Impact of Pseudomonas aeruginosa genomic instability on the application of typing methods for chronic cystic fibrosis infections. *J. Clin. Microbiol.* **48,** 2053–2059 (2010).
225. Griffiths, A. L. *et al.* Effects of segregation on an epidemic Pseudomonas aeruginosa strain in a cystic fibrosis clinic. *Am. J. Respir. Crit. Care Med.* **171,** 1020–1025 (2005).
226. Griffiths, A. L., Wurzel, D. F., Robinson, P. J., Carzino, R. & Massie, J. Australian epidemic strain pseudomonas (AES-1) declines further in a cohort segregated cystic fibrosis clinic. *J. Cyst. Fibros.* **11,** 49–52 (2012).

227. Winstanley, C. & Fothergill, J. L. The role of quorum sensing in chronic cystic fibrosis Pseudomonas aeruginosa infections. *FEMS Microbiology Letters* **290,** 1–9 (2009).
228. Vallet, I., Olson, J. W., Lory, S., Lazdunski, A. & Filloux, A. The chaperone/usher pathways of Pseudomonas aeruginosa: identification of fimbrial gene clusters (cup) and their involvement in biofilm formation. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6911–6 (2001).
229. Hahn, H. P. The type-4 pilus is the major virulence-associated adhesin of Pseudomonas aeruginosa - A review. in *Gene* **192,** 99–108 (1997).
230. Rocchetta, H. L., Burrows, L. L. & Lam, J. S. Genetics of O-antigen biosynthesis in Pseudomonas aeruginosa. *Microbiol. Mol. Biol. Rev.* **63,** 523–553 (1999).
231. Hobbs, M., Collie, E. S. R., Free, P. D., Livingston, S. P. & Mattick, J. S. PilS and PilR, a two-component transcriptional regulatory system controlling expression of type 4 fimbriae in Pseudomonas aeruginosa. *Mol. Microbiol.* **7,** 669–682 (1993).
232. Lynn, W. A. & Golenbock, D. T. Lipopolysaccharide antagonists. *Immunology Today* **13,** 271–276 (1992).
233. Lam, M. Y. C. *et al.* Occurrence of a common lipopolysaccharide antigen in standard and clinical strains of Pseudomonas aeruginosa. *J. Clin. Microbiol.* **27,** 962–967 (1989).
234. Gupta, S. K., Berk, R. S., Masinick, S. & Hazlett, L. D. Pili and lipopolysaccharide of Pseudomonas aeruginosa bind to the glycolipid asialo GM1. *Infect. Immun.* **62,** 4572–4579 (1994).
235. Hajjar, A. M., Ernst, R. K., Tsai, J. H., Wilson, C. B. & Miller, S. I. Human Toll-like receptor 4 recognizes host-specific LPS modifications. *Nat. Immunol.* **3,** 354–359 (2002).
236. Adam, E. C., Mitchell, B. S., Schumacher, D. U., Grant, G. & Schumacher, U. Pseudomonas aeruginosa II lectin stops human ciliary beating: Therapeutic implications of fucose. *Am. J. Respir. Crit. Care Med.* **155,** 2102–2104 (1997).
237. Chemani, C. *et al.* Role of LecA and LecB lectins in Pseudomonas aeruginosa-induced lung injury and effect of carbohydrate ligands. *Infect. Immun.* **77,** 2065–2075 (2009).
238. Goehring, U. M., Schmidt, G., Pederson, K. J., Aktories, K. & Barbieri, J. T. The N-terminal domain of Pseudomonas aeruginosa exoenzyme S is a GTPase-activating protein for Rho GTPases. *J. Biol. Chem.* **274,** 36369–36372 (1999).
239. Fu, H., Coburn, J. & Collier, R. J. The eukaryotic host factor that activates exoenzyme S of Pseudomonas aeruginosa is a member of the 14-3-3 protein family. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2320–4 (1993).
240. Kipnis, E., Sawa, T. & Wiener-Kronish, J. Targeting mechanisms of Pseudomonas aeruginosa pathogenesis. *Medecine et Maladies Infectieuses* **36,** 78–91 (2006).
241. Epelman, S., Bruno, T. F., Neely, G. G., Woods, D. E. & Mody, C. H. Pseudomonas aeruginosa exoenzyme S induces transcriptional expression of proinflammatory cytokines and chemokines. *Infect. Immun.* **68,** 4811–4814 (2000).
242. Bruno, T. F., Woods, D. E., Storey, D. G. & Mody, C. H. Recombinant

Pseudomonas exoenzyme S and exoenzyme S from Pseudomonas aeruginosa DG1 share the ability to stimulate T lymphocyte proliferation. *Can. J. Microbiol.* **45,** 607–611 (1999).
243. Bruno, T. F., Woods, D. E. & Mody, C. H. Exoenzyme S from Pseudomonas aeruginosa induces apoptosis in T lymphocytes. *J. Leukoc. Biol.* **67,** 808–16 (2000).
244. Shaver, C. M. & Hauser, A. R. Relative contributions of Pseudomonas aeruginosa ExoU, ExoS, and ExoT to virulence in the lung. *Infect. Immun.* **72,** 6969–6977 (2004).
245. Lee, V. T., Smith, R. S., Tümmler, B. & Lory, S. Activities of Pseudomonas aeruginosa effectors secreted by the type III secretion system in vitro and during infection. *Infect. Immun.* **73,** 1695–1705 (2005).
246. Tielker, D. *et al.* Pseudomonas aeruginosa lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* **151,** 1313–1323 (2005).
247. Fothergill, J. L., Mowat, E., Ledson, M. J., Walshaw, M. J. & Winstanley, C. Fluctuations in phenotypes and genotypes within populations of Pseudomonas aeruginosa in the cystic fibrosis lung during pulmonary exacerbations. *J. Med. Microbiol.* **59,** 472–481 (2010).
248. Iglewski, B. H. & Sadoff, J. C. Toxin inhibitors of protein synthesis: production, purification, and assay of Pseudomonas aeruginosa toxin A. *Methods Enzymol.* **60,** 780–793 (1979).
249. Apidianakis, Y. & Rahme, L. G. Drosophila melanogaster as a model host for studying Pseudomonas aeruginosa infection. *Nat. Protoc.* **4,** 1285–94 (2009).
250. Gallant, C. V., Raivio, T. L., Olson, J. C., Woods, D. E. & Storey, D. G. Pseudomonas aeruginosa cystic fibrosis clinical isolates produce exotoxin A with altered ADP-ribosyltransferase activity and cytotoxicity. *Microbiology* **146,** 1891–1899 (2000).
251. Bell, C. E. & Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35,** 1137–49 (1996).
252. Wick, M. J., Frank, D. W., Storey, D. G. & Iglewski, B. H. Structure, function, and regulation of Pseudomonas aeruginosa exotoxin A. *Annu. Rev. Microbiol.* **44,** 335–363 (1990).
253. Nikaido, H. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science (80-.).* **264,** 382–388 (1994).
254. Lee, W. *et al.* Pyocyanin production by Pseudomonas aeruginosa induces neutrophil apoptosis and impairs neutrophil-mediated host defenses in vivo. *J. Immunol.* **174,** 3643–3649 (2015).
255. Fothergill, J. L. *et al.* Widespread pyocyanin over-production among isolates of a cystic fibrosis epidemic strain. *BMC Microbiol.* **7,** 45 (2007).
256. Galloway, D. R. Pseudomonas aeruginosa elastase and elastolysis revisited: recent developments. *Mol. Microbiol.* **5,** 2315–2321 (1991).
257. Azghani, A. O., Miller, E. J. & Peterson, B. T. Virulence factors from Pseudomonas aeruginosa increase lung epithelial permeability. *Lung* **178,** 261–269 (2000).
258. Dulon, S. *et al.* Pseudomonas aeruginosa elastase disables proteinase-activated receptor 2 in respiratory epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **32,** 411–419 (2005).

259. Park, P. W., Biedermann, K., Mecham, L., Bissett, D. L. & Mecham, R. P. Lysozyme binds to elastin and protects elastin from elastase-mediated degradation. *J. Invest. Dermatol.* **106,** 1075–80 (1996).
260. Lamont, I. L., Beare, P. A., Ochsner, U., Vasil, A. I. & Vasil, M. L. Siderophore-mediated signaling regulates virulence factor production in Pseudomonasaeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7072–7 (2002).
261. Meyer, J. M., Neely, A., Stintzi, A., Georges, C. & Holder, I. A. Pyoverdin is essential for virulence of Pseudomonas aeruginosa. *Infect. Immun.* **64,** 518–523 (1996).
262. Takase, H., Nitanai, H., Hoshino, K. & Otani, T. Impact of Siderophore Production on Pseudomonas aeruginosa Infections in Immunosuppressed Mice. *Infect. Immun.* **68,** 1834–1839 (2000).
263. Guzzo, J., Pages, J. M., Duong, F., Lazdunski, A. & Murgier, M. Pseudomonas aeruginosa alkaline protease: Evidence for secretion genes and study of secretion mechanism. *J. Bacteriol.* **173,** 5290–5297 (1991).
264. Matsumoto, K. Role of bacterial proteases in pseudomonal and serratial keratitis. *Biological Chemistry* **385,** 1007–1016 (2004).
265. Kipnis, E. *et al.* Massive alveolar thrombin activation in Pseudomonas aeruginosa-induced acute lung injury. *Shock* **21,** 444–451 (2004).
266. Malloy, J. L., Veldhuizen, R. a W., Thibodeaux, B. a, O'Callaghan, R. J. & Wright, J. R. Pseudomonas aeruginosa protease IV degrades surfactant proteins and inhibits surfactant host defense and biophysical functions. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **288,** 409–418 (2005).
267. Kohler, T., Curty, L. K., Barja, F., Van Delden, C. & Pechere, J. C. Swarming of Pseudomonas aeruginosa is dependent on cell-to-cell signaling and requires flagella and pili. *J. Bacteriol.* **182,** 5990–5996 (2000).
268. Davey, M. E., Caiazza, N. C. & O'Toole, G. A. Rhamnolipid surfactant production affects biofilm architecture in Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **185,** 1027–1036 (2003).
269. Maier, R. M. & Soberón-Chávez, G. Pseudomonas aeruginosa rhamnolipids: biosynthesis and potential applications. *Appl. Microbiol. Biotechnol.* **54,** 625–633 (2000).
270. Kownatzki, R., Tummler, B. & Doring, G. Rhamnolipid of Pseudomonas aeruginosa in sputum of cystic fibrosis patients. *Lancet (London, England)* **1,** 1026–1027 (1987).
271. McClure, C. D. & Schiller, N. L. Inhibition of macrophage phagocytosis by Pseudomonas aeruginosa rhamnolipids in vitro and in vivo. *Curr. Microbiol.* **33,** 109–117 (1996).
272. Read, R. C. *et al.* Effect of Pseudomonas aeruginosa rhamnolipids on mucociliary transport and ciliary beating. *J. Appl. Physiol.* **72,** 2271–2277 (1992).
273. Marvig, R. L. *et al.* Within-host microevolution of Pseudomonas aeruginosa in Italian cystic fibrosis patients. *BMC Microbiol.* **15,** 218 (2015).
274. Dettman, J. R. *et al.* Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular Ecology* **21,** 2058–2077 (2012).
275. Barth, A. L. & Pitt, T. L. The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas

aeruginosa. *J. Med. Microbiol.* **45,** 110–119 (1996).
276. Mahenthiralingam, E., Campbell, M. E. & Speert, D. P. Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infect. Immun.* **62,** 596–605 (1994).
277. Oliver, a. High Frequency of Hypermutable Pseudomonas aeruginosa in Cystic Fibrosis Lung Infection. *Science (80-.).* **288,** 1251–1253 (2000).
278. Hoffman, L. R. *et al.* Pseudomonas aeruginosa lasR mutants are associated with cystic fibrosis lung disease progression. *J. Cyst. Fibros.* **8,** 66–70 (2009).
279. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7481–6 (2011).
280. Williams, D. *et al.* Divergent, Coexisting Pseudomonas aeruginosa Lineages in Chronic Cystic Fibrosis Lung Infections. *Am. J. Respir. Crit. Care Med.* **191,** 775–785 (2015).
281. Bragonzi, A. *et al.* Pseudomonas aeruginosa microevolution during cystic fibrosis lung infection establishes clones with adapted virulence. *Am. J. Respir. Crit. Care Med.* **180,** 138–145 (2009).
282. Marvig, R. L., Johansen, H. K., Molin, S. & Jelsbak, L. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genet.* **9,** e1003741 (2013).
283. Mowat, E. *et al.* Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *Am. J. Respir. Crit. Care Med.* **183,** 1674–1679 (2011).
284. Workentine, M. L. *et al.* Phenotypic Heterogeneity of Pseudomonas aeruginosa Populations in a Cystic Fibrosis Patient. *PLoS One* **8,** (2013).
285. Smith, E. E. *et al.* Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8487–92 (2006).
286. Cramer, N. *et al.* Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ. Microbiol.* **13,** 1690–1704 (2011).
287. Hoboth, C. *et al.* Dynamics of adaptive microevolution of hypermutable Pseudomonas aeruginosa during chronic pulmonary infection in patients with cystic fibrosis. *J. Infect. Dis.* **200,** 118–30 (2009).
288. Anthony, M. *et al.* Genetic analysis of Pseudomonas aeruginosa isolates from the sputa of Australian adult cystic fibrosis patients. *J Clin Microbiol* **40,** 2772–2778 (2002).
289. Bezuidt, O. K. *et al.* Intraclonal genome diversity of Pseudomonas aeruginosa clones CHA and TB. *BMC Genomics* **14,** 416 (2013).
290. Winsor, G. L. *et al.* Pseudomonas Genome Database: Improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Res.* **39,** (2011).
291. Craig Winstanley, Morgan G.I. Langille, Joanne L. Fothergill, Irena Kukavica-Ibrulj, Catherine Paradis-Bleau, François Sanschagrin, Nicholas R. Thomson, Geoff L. Winsor, Michael A. Quail, 4 Nicola Lennard, Alexandra Bignell, Louise Clarke, Kathy Seeger, D, and R. C. L. Newly introduced genomic prophage islands are critical determinants of in vivo

competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome …* **1,** 12–23 (2009).
292. Krueger, F. Trim Galore!. *[http://www.bioinformatics.babraham.ac.uk/projects/trim galore/]* (2012). doi:http://www.bioinformatics.babraham.ac.uk/projects/trim galore/
293. Coil, D., Jospin, G. & Darling, A. E. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31,** 587–9 (2015).
294. Assefa, S., Keane, T. M., Otto, T. D., Newbold, C. & Berriman, M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* **25,** 1968—1969 (2009).
295. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* (2013).
296. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).
297. Li, H., Handsaker, B., Wysoker, A. & Fennell, T. The sequence alignment/map format and SAMtools. … (2009).
298. Broad Institute. Picard tools. *https://broadinstitute.github.io/picard/* (2016). Available at: https://broadinstitute.github.io/picard/%5Cnhttp://broadinstitute.github.io/picard/.
299. McKenna, A., Hanna, M. & Banks, E. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome …* 1297–1303 (2010). doi:10.1101/gr.107524.110.20
300. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).
301. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. 80–92 (2012).
302. Seemann, T. mlst.
303. Larsen, M. V. *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **50,** 1355–1361 (2012).
304. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11,** 595 (2010).
305. Inouye, M., Dashnow, H., Raven, L., Schultz, M. B. & Bernard, J. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs SRST2 : Rapid genomic surveillance for public health and hospital microbiology labs. 0–16 (2014).
306. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. MUMmer: comparative applications\nFast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30,** 2478–83 (2002).
307. Laing, C. *et al.* Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* **11,** 461 (2010).
308. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.

Nucleic Acids Res. **22,** 4673–80 (1994).
309. Chapman, B. & Chang, J. Biopython. *ACM SIGBIO Newsl.* **20,** 15–19 (2000).
310. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–9 (2013).
311. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
312. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43,** e15 (2015).
313. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23,** 127–8 (2007).
314. Hoekstra, H. E. & Coyne, J. A. THE LOCUS OF EVOLUTION: EVO DEVO AND THE GENETICS OF ADAPTATION. *Evolution (N. Y).* **61,** 995–1016 (2007).
315. Jacob, F. Evolution and tinkering. *Science* **196,** 1161–1166 (1977).
316. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **175,** 315–332 (2007).
317. Wiedenbeck, J. *et al.* Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35,** 957–76 (2011).
318. Sahl, J. W. *et al.* Evolution of a Pathogen: A Comparative Genomics Analysis Identifies a Genetic Pathway to Pathogenesis in Acinetobacter. *PLoS One* **8,** e54287 (2013).
319. Dumas, Z. & Kümmerli, R. Cost of cooperation rules selection for cheats in bacterial metapopulations. *J. Evol. Biol.* **25,** 473–484 (2012).
320. Thanassi, D. G., Suh, G. S. B. & Nikaido, H. Role of outer membrane barrier in efflux-mediated tetracycline resistance of Escherichia coli. *J. Bacteriol.* **177,** 998–1007 (1995).
321. Hottes, A. K. *et al.* Bacterial Adaptation through Loss of Function. *PLoS Genet.* **9,** (2013).
322. Lyczak, J. B., Cannon, C. L. & Pier, G. B. Establishment of Pseudomonas aeruginosa infection: Lessons from a versatile opportunist. *Microbes and Infection* **2,** 1051–1060 (2000).
323. De Soyza, A. *et al.* Developing an international Pseudomonas aeruginosa reference panel. *Microbiologyopen* **2,** 1010–23 (2013).
324. Cullen, L. *et al.* Phenotypic characterization of an international pseudomonas aeruginosa reference panel: Strains of cystic fibrosis (cf) origin show less in vivo virulence than non-cf strains. *Microbiol. (United Kingdom)* **161,** 1961–1977 (2015).
325. Feliziani, S. *et al.* Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. *PLOS Genet.* **10,** e1004651 (2014).
326. Marvig, R. L. *et al.* Within-host evolution of pseudomonas aeruginosa reveals adaptation toward iron acquisition from hemoglobin. *MBio* **5,** (2014).
327. Lee, J. Y., Na, I. Y., Park, Y. K. & Ko, K. S. Genomic variations between colistin-susceptible and-resistant pseudomonas aeruginosa clinical isolates and their effects on colistin resistance. *J. Antimicrob. Chemother.*

69, 1248–1256 (2014).
328. van Delden, C., Page, M. G. P. & Köhler, T. Involvement of Fe Uptake Systems and AmpC β-Lactamase in Susceptibility to the Siderophore Monosulfactam BAL30072 in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **57,** 2095–2102 (2013).
329. Tsutsumi, Y., Tomita, H. & Tanimoto, K. Identification of Novel Genes Responsible for Overexpression of ampC in Pseudomonas aeruginosa PAO1. *Antimicrob. Agents Chemother.* **57,** 5987–5993 (2013).
330. McPherson, C. J. *et al.* Clinically Relevant Gram-Negative Resistance Mechanisms Have No Effect on the Efficacy of MC-1, a Novel Siderophore-Conjugated Monocarbam. *Antimicrob. Agents Chemother.* **56,** 6334–6342 (2012).
331. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4,** 597–606 (2006).
332. Griffin, A. S., West, S. a & Buckling, A. Cooperation and competition in pathogenic bacteria. *Earth* **430,** 1024–1027 (2004).
333. Andersen, S. B., Marvig, R. L., Molin, S., Krogh Johansen, H. & Griffin, A. S. Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc. Natl. Acad. Sci.* **112,** 201508324 (2015).
334. Jeukens, J. *et al.* Comparative genomics of isolates of a Pseudomonas aeruginosa epidemic strain associated with chronic lung infections of cystic fibrosis patients. *PLoS One* **9,** e87611 (2014).
335. Jones, A., Govan, J. & Doherty, C. Spread of a multiresistant strain of Pseudomonas aeruginosa in an adult cystic fibrosis clinic. *Lancet* **358,** 557–558 (2001).
336. Nixon, G. M. *et al.* Clinical outcome after early Pseudomonas aeruginosa infection in cystic fibrosis. *J. Pediatr.* **138,** 699–704 (2001).
337. Syrmis, M. W. Rapid genotyping of Pseudomonas aeruginosa isolates harboured by adult and paediatric patients with cystic fibrosis using repetitive-element-based PCR assays. *J. Med. Microbiol.* **53,** 1089–1096 (2004).
338. Dingemans, J. *et al.* Genome sequence analysis of a Belgian epidemic cystic fibrosis Pseudomonas aeruginosa isolate. *Int. J. Infect. Dis.* **21,** 351 (2017).
339. Parkins, M. D. *et al.* Twenty-five-year outbreak of Pseudomonas aeruginosa infecting individuals with cystic fibrosis: Identification of the prairie epidemic strain. *J. Clin. Microbiol.* **52,** 1127–1135 (2014).
340. LiPuma, J. J. *et al.* Disproportionate distribution of Burkholderia cepacia complex species and transmissibility markers in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **164,** 92–96 (2001).
341. McCallum, S., Corkill, J. & Gallagher, M. Superinfection with a transmissible strain of Pseudomonas aeruginosa in adults with cystic fibrosis chronically colonised by P aeruginosa. *Lancet* **358,** 558–60 (2001).
342. Edenborough, F. P. *et al.* Genotyping of Pseudomonas aeruginosa in cystic fibrosis suggests need for segregation. *J. Cyst. Fibros.* **3,** 37–44 (2004).
343. Aaron, S. D. *et al.* Infection with transmissible strains of Pseudomonas aeruginosa and clinical outcomes in adults with cystic fibrosis. *JAMA J. Am. Med. Assoc.* **304,** 2145–53 (2010).

344. Gilbert, K. B., Kim, T. H., Gupta, R., Greenberg, E. P. & Schuster, M. Global position analysis of the Pseudomonas aeruginosa quorum-sensing transcription factor LasR. *Mol. Microbiol.* **73,** 1072–1085 (2009).
345. Blohmke, C. J. *et al.* Innate immunity mediated by TLR5 as a novel antiinflammatory target for cystic fibrosis lung disease. *J. Immunol.* **180,** 7764–7773 (2008).
346. Panagea, S., Winstanley, C., Walshaw, M. J., Ledson, M. J. & Hart, C. a. Environmental contamination with an epidemic strain of Pseudomonas aeruginosa in a Liverpool cystic fibrosis centre, and study of its survival on dry surfaces. *J. Hosp. Infect.* **59,** 102–7 (2005).
347. McCallum, S., Gallagher, M. & Corkill, J. Spread of an epidemic Pseudomonas aeruginosa strain from a patient with cystic fibrosis (CF) to non-CF relatives. *Thorax* 559–561 (2002).
348. Mohan, K. *et al.* Transmission of Pseudomonas aeruginosa epidemic strain from a patient with cystic fibrosis to a pet cat. *Thorax* **63,** 839–40 (2008).
349. Davies, E. V. *et al.* Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proc. Natl. Acad. Sci.* **113,** 8266–8271 (2016).
350. Sahl, J. W., Caporaso, J. G., Rasko, D. a & Keim, P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* **2,** e332 (2014).
351. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
352. Darch, S. E. *et al.* Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Sci. Rep.* **5,** 7649 (2015).
353. Montanari, S. *et al.* Biological cost of hypermutation in Pseudomonas aeruginosa strains from patients with cystic fibrosis. *Microbiology* **153,** 1445–54 (2007).
354. Dingemans, J. *et al.* Effect of Shear Stress on Pseudomonas aeruginosa Isolated from the Cystic Fibrosis Lung. *MBio* **7,** e00813-16 (2016).
355. Dingemans, J. *et al.* The deletion of TonB-dependent receptor genes is part of the genome reduction process that occurs during adaptation of Pseudomonas aeruginosa to the cystic fibrosis lung. *Pathog. Dis.* **71,** 26–38 (2014).
356. Li, H., Luo, Y.-F., Williams, B. J., Blackwell, T. S. & Xie, C.-M. Structure and function of OprD protein in Pseudomonas aeruginosa: From antibiotic resistance to novel therapies. *Int. J. Med. Microbiol.* **302,** 10.1016/j.ijmm.2011.10.001 (2012).
357. Salunkhe, P. *et al.* A Cystic Fibrosis Epidemic Strain of Pseudomonas aeruginosa Displays Enhanced Virulence and Antimicrobial Resistance. *J. Bacteriol.* **187,** 4908–4920 (2005).
358. Lamers, R. P., Nguyen, U. T., Nguyen, Y., Buensuceso, R. N. C. & Burrows, L. L. Loss of membrane-bound lytic transglycosylases increases outer membrane permeability and β-lactam sensitivity in Pseudomonas aeruginosa . *Microbiologyopen* **4,** 879–895 (2015).
359. Castañeda-García, A., Rodríguez-Rojas, A., Guelfo, J. R. & Blázquez, J. The Glycerol-3-Phosphate Permease GlpT Is the Only Fosfomycin Transporter in Pseudomonas aeruginosa . *J. Bacteriol.* **191,** 6968–6974 (2009).

360. Poole, K. *et al.* Potentiation of Aminoglycoside Activity in Pseudomonas aeruginosa by Targeting the AmgRS Envelope Stress-Responsive Two-Component System. *Antimicrob. Agents Chemother.* **60,** 3509–3518 (2016).
361. Quint, J. K. *et al.* Changes in the incidence, prevalence and mortality of bronchiectasis in the UK from 2004 to 2013: a population-based cohort study. *Eur. Respir. J.* **47,** 186–193 (2016).
362. Finch, S., McDonnell, M. J., Abo-Leyah, H., Aliberti, S. & Chalmers, J. D. A comprehensive analysis of the impact of pseudomonas aeruginosa colonization on prognosis in adult bronchiectasis. *Ann. Am. Thorac. Soc.* **12,** 1602–1611 (2015).
363. Chalmers, J. D., Aliberti, S. & Blasi, F. Management of bronchiectasis in adults. *Eur. Respir. J.* **45,** 1446–1462 (2015).
364. Sommer, L. M. *et al.* Bacterial evolution in PCD and CF patients follows the same mutational steps. *Sci. Rep.* **6,** 28732 (2016).
365. Quick, J. *et al.* Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing. *BMJ Open* **4,** (2014).
366. Jones, A. M. *et al.* Prospective surveillance for Pseudomonas aeruginosa cross-infection at a cystic fibrosis center. *Am. J. Respir. Crit. Care Med.* **171,** 257–60 (2005).
367. De Soyza, A. *et al.* S126 Molecular epidemiological analysis suggests cross infection with pseudomonas aeruginosa is rare in non-cystic fibrosis bronchiectasis. *Thorax* **68,** A65–A65 (2013).
368. Kuchma, S. L. *et al.* BifA, a cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14. in *Journal of Bacteriology* **189,** 8165–8178 (2007).
369. An, S., Wu, J. & Zhang, L. H. Modulation of pseudomonas aeruginosa biofilm dispersal by a cyclic-di-gmp phosphodiesterase with a putative hypoxia-sensing domain. *Appl. Environ. Microbiol.* **76,** 8160–8173 (2010).
370. Bouffartigues, E. *et al.* The absence of the Pseudomonas aeruginosa OprF protein leads to increased biofilm formation through variation in c-di-GMP level. *Front. Microbiol.* **6,** (2015).
371. Records, A. R. & Gross, D. C. Sensor kinases RetS and LadS regulate Pseudomonas syringae Type VI secretion and virulence factors. *J. Bacteriol.* **192,** 3584–3596 (2010).
372. Davies, J. Vicious circles: looking back on resistance plasmids. *Genetics* **139,** 1465–1468 (1995).
373. Davison, J. Genetic exchange between bacteria in the environment. *Plasmid* **42,** 73–91 (1999).
374. Brüssow, H. & Hendrix, R. W. Phage Genomics: Small is beautiful. *Cell* **108,** 13–16 (2002).
375. Jiang, S. C. & Paul, J. H. Gene Transfer by Transduction in the Marine Environment Gene Transfer by Transduction in the Marine Environment. **64,** 2780–2787 (1998).
376. Colomer-Lluch, M., Jofre, J. & Muniesa, M. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One* **6,** (2011).
377. Khan, G. A., Berglund, B., Khan, K. M., Lindgren, P. E. & Fick, J. Occurrence

and Abundance of Antibiotics and Resistance Genes in Rivers, Canal and near Drug Formulation Facilities - A Study in Pakistan. *PLoS One* **8,** (2013).
378. Zhang, Q. Q., Ying, G. G., Pan, C. G., Liu, Y. S. & Zhao, J. L. Comprehensive evaluation of antibiotics emission and fate in the river basins of China: Source analysis, multimedia modeling, and linkage to bacterial resistance. *Environ. Sci. Technol.* **49,** 6772–6782 (2015).
379. Knapp, C. W., Dolfing, J., Ehlert, P. A. I. & Graham, D. W. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environ. Sci. Technol.* **44,** 580–587 (2010).
380. Chen, B., Liang, X., Huang, X., Zhang, T. & Li, X. Differentiating anthropogenic impacts on ARGs in the Pearl River Estuary by using suitable gene indicators. *Water Res.* **47,** 2811–2820 (2013).
381. Cytryn, E. The soil resistome: The anthropogenic, the native, and the unknown. *Soil Biology and Biochemistry* **63,** 18–23 (2013).
382. Segawa, T. *et al.* Distribution of antibiotic resistance genes in glacier environments. *Environ. Microbiol. Rep.* **5,** 127–134 (2013).
383. McKinney, C. W. & Pruden, A. Ultraviolet disinfection of antibiotic resistant bacteria and their antibiotic resistance genes in water and wastewater. *Environ. Sci. Technol.* **46,** 13393–400 (2012).
384. Xu, Y. *et al.* Occurrence and distribution of antibiotics, antibiotic resistance genes in the urban rivers in Beijing, China. *Environ. Pollut.* **213,** 833–840 (2016).
385. Zhu, Y.-G. *et al.* Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat. Microbiol.* **2,** 16270 (2017).
386. Master, R. N., Deane, J., Opiela, C. & Sahm, D. F. Recent trends in resistance to cell envelope-active antibacterial agents among key bacterial pathogens. *Ann. N. Y. Acad. Sci.* **1277,** 1–7 (2013).
387. Souli, M., Galani, I. & Giamarellou, H. Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Euro surveillance : bulletin europ??en sur les maladies transmissibles = European communicable disease bulletin* **13,** (2008).
388. Cortés, J. A., Garzón, D. C., Navarrete, J. A. & Contreras, K. M. Impact of inappropriate antimicrobial therapy on patients with bacteremia in intensive care units and resistance patterns in Latin America. *Revista Argentina de Microbiologia* **42,** 230–234 (2010).
389. Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **16,** 161–168 (2016).
390. Lee, J.-Y., Park, Y. K., Chung, E. S., Na, I. Y. & Ko, K. S. Evolved resistance to colistin and its loss due to genetic reversion in Pseudomonas aeruginosa. *Sci. Rep.* **6,** 25543 (2016).
391. Moskowitz, S. M. *et al.* PmrB mutations promote polymyxin resistance of Pseudomonas aeruginosa isolated from colistin-treated cystic fibrosis patients. *Antimicrob. Agents Chemother.* **56,** 1019–1030 (2012).
392. Guzvinec, M. *et al.* Sequence types 235, 111, and 132 predominate among multidrug-resistant pseudomonas aeruginosa clinical isolates in Croatia. *Antimicrob. Agents Chemother.* **58,** 6277–6283 (2014).

393. Livermore, D. M. Multiple mechanisms of antimicrobial resistance in Pseudomonas aeruginosa: our worst nightmare? *Clin. Infect. Dis.* **34,** 634–640 (2002).
394. Cholley, P. *et al.* The majority of multi-drug resistant Pseudomonas aeruginosa isolates from hospitals in eastern France belongs to a few clonal types. *J Clin Microbiol* (2011). doi:10.1128/jcm.00102-11
395. Wi, Y. M. *et al.* Emergence of colistin resistance in Pseudomonas aeruginosa ST235 clone in South Korea. *Int. J. Antimicrob. Agents* **49,** 767–769 (2017).
396. Khuntayaporn, P., Montakantikul, P., Mootsikapun, P., Thamlikitkul, V. & Chomnawang, M. Prevalence and genotypic relatedness of carbapenem resistance among multidrug-resistant P. aeruginosa in tertiary hospitals across Thailand. *Ann. Clin. Microbiol. Antimicrob.* **11,** 25 (2012).
397. Khuntayaporn, P., Montakantikul, P., Santanirand, P., Kiratisin, P. & Chomnawang, M. T. Molecular investigation of carbapenem resistance among multidrug-resistant Pseudomonas aeruginosa isolated clinically in Thailand. *Microbiol. Immunol.* **57,** 170–178 (2013).
398. Gutiérrez, O. *et al.* Molecular epidemiology and mechanisms of carbapenem resistance in Pseudomonas aeruginosa isolates from Spanish hospitals. *Antimicrob. Agents Chemother.* **51,** 4329–4335 (2007).
399. Wolter, D. J., Hanson, N. D. & Lister, P. D. Insertional inactivation of oprD in clinical isolates of Pseudomonas aeruginosa leading to carbapenem resistance. *FEMS Microbiol. Lett.* **236,** 137–143 (2004).
400. Cury, J., Jové, T., Touchon, M., Néron, B. & Rocha, E. P. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res.* **44,** 4539–4550 (2016).
401. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9,** 1–10 (2014).
402. López-Causapé, C. *et al.* Evolution of the Pseudomonas aeruginosa mutational resistome in an international Cystic Fibrosis clone. *Sci. Rep.* **7,** 5555 (2017).
403. Strateva, T. & Yordanov, D. Pseudomonas aeruginosa - a phenomenon of bacterial resistance. *J. Med. Microbiol.* **58,** 1133–48 (2009).
404. Viedma, E. *et al.* Nosocomial spread of colistin-only-sensitive sequence type 235 Pseudomonas aeruginosa isolates producing the extended-spectrum β-lactamases GES-1 and GES-5 in Spain. *Antimicrob. Agents Chemother.* **53,** 4930–4933 (2009).
405. Fan, X. *et al.* Diverse Genetic Background of Multidrug-Resistant Pseudomonas aeruginosa from Mainland China, and Emergence of an Extensively Drug-Resistant ST292 Clone in Kunming. *Sci. Rep.* **6,** 26522 (2016).
406. Chen, Y., Sun, M., Wang, M., Lu, Y. & Yan, Z. Dissemination of IMP-6-producing Pseudomonas aeruginosa ST244 in multiple cities in China. *Eur. J. Clin. Microbiol. Infect. Dis.* **33,** 1181–1187 (2014).
407. Empel, J. *et al.* Outbreak of Pseudomonas aeruginosa infections with PER-1 extended-spectrum β-lactamase in Warsaw, Poland: Further evidence for an international clonal complex. *J. Clin. Microbiol.* **45,** 2829–2834 (2007).

408. Moyo, S. *et al.* Identification of VIM-2-producing Pseudomonas aeruginosa from Tanzania is associated with sequence types 244 and 640 and the location of blaVIM-2 in a TniC integron. *Antimicrob. Agents Chemother.* **59,** 682–685 (2015).
409. Cho, H. H., Kwon, K. C., Sung, J. Y. & Koo, S. H. Prevalence and genetic analysis of multidrug-resistant Pseudomonas aeruginosa ST235 isolated from a hospital in Korea, 2008-2012. *Ann. Clin. Lab. Sci.* **43,** 414–419 (2013).
410. Yoo, J. S. *et al.* Dissemination of genetically related IMP-6-producing multidrug-resistant Pseudomonas aeruginosa ST235 in South Korea. *Int. J. Antimicrob. Agents* **39,** 300–304 (2012).
411. Nemec, A., Krizova, L., Maixnerova, M. & Musilek, M. Multidrug-resistant epidemic clones among bloodstream isolates of Pseudomonas aeruginosa in the Czech Republic. *Res. Microbiol.* **161,** 234–42 (2010).
412. Zhang, G. *et al.* A new subclass of intrinsic aminoglycoside nucleotidyltransferases, ANT(3")-II, is horizontally transferred among Acinetobacter spp. by homologous recombination. *PLoS Genet.* **13,** (2017).
413. Shaw, K. J., Rather, P. N., Hare, R. S. & Miller, G. H. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol. Rev.* **57,** 138–163 (1993).
414. Fonseca, É. L., Freitas, F. D. S., Vieira, V. V. & Vicente, A. C. P. New qnr gene cassettes associated with superintegron repeats in Vibrio cholerae O1. *Emerg. Infect. Dis.* **14,** 1129–1131 (2008).
415. Mammeri, H., Van De Loo, M., Poirel, L., Martinez-Martinez, L. & Nordmann, P. Emergence of plasmid-mediated quinolone resistance in Escherichia coli in Europe. *Antimicrob. Agents Chemother.* **49,** 71–76 (2005).
416. Michalska, A. D., Sacha, P. T., Ojdana, D., Wieczorek, A. & Tryniszewska, E. Prevalence of resistance to aminoglycosides and fluoroquinolones among Pseudomonas aeruginosa strains in a University Hospital in Northeastern Poland. *Brazilian J. Microbiol.* **45,** 1455–1458 (2014).
417. Belotti, P. T. *et al.* Description of an original integron encompassing bla<inf>VIM-2</inf>, qnrVC1 and genes encoding bacterial group II intron proteins in Pseudomonas aeruginosa. *J. Antimicrob. Chemother.* **70,** 2237–2240 (2015).
418. Fluit, A. C. & Schmitz, F. J. Resistance integrons and super-integrons. *Clinical Microbiology and Infection* **10,** 272–288 (2004).
419. Yelin, R., Rotem, D. & Schuldiner, S. EmrE, a small Escherichia coli multidrug transporter, protects Saccharomyces cerevisiae from toxins by sequestration in the vacuole. *J. Bacteriol.* **181,** 949–956 (1999).
420. Li, X. Z., Poole, K. & Nikaido, H. Contributions of MexAB-OprM and an EmrE homolog to intrinsic resistance of Pseudomonas aeruginosa to aminoglycosides and dyes. *Antimicrob. Agents Chemother.* **47,** 27–33 (2003).
421. Fallah, F., Borhan, R. S. & Hashemi, A. Detection of bla(IMP) and bla(VIM) metallo-β-lactamases genes among Pseudomonas aeruginosa strains. *Int. J. Burns Trauma* **3,** 122–124 (2013).
422. Winstanley, C. & Kaye, S. Genotypic and phenotypic characteristics of Pseudomonas aeruginosa isolates associated with ulcerative keratitis. *J.*

Med. … **54,** 519–26 (2005).
423. Gong, F. Y. *et al.* SiRNA-mediated gene silencing of MexB from the MexA-MexB-OprM efflux pump in Pseudomonas aeruginosa. *BMB Rep.* **47,** 203–208 (2014).
424. Marvig, R. L., Sommer, L. M., Jelsbak, L., Molin, S. & Johansen, H. K. Evolutionary insight from whole-genome sequencing of Pseudomonas aeruginosa from cystic fibrosis patients. *Future Microbiol.* **10,** 599–611 (2015).
425. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** (2007).
426. Khaledi, A. *et al.* Transcriptome profiling of antimicrobial resistance in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **60,** 4722–4733 (2016).
427. Selezska, K. *et al.* Pseudomonas aeruginosa population structure revisited under environmental focus: Impact of water quality and phage pressure. *Environ. Microbiol.* **14,** 1952–1967 (2012).
428. Jeukens, J., Kukavica-Ibrulj, I., Emond-Rheault, J. G., Freschi, L. & Levesque, R. C. Comparative genomics of a drug-resistant Pseudomonas aeruginosa panel and the challenges of antimicrobial resistance prediction from genomes. *FEMS Microbiol. Lett.* **364,** fnx161-fnx161 (2017).

9. Appendix

Supplementary table 1. Genes significantly (chi squared $p < 0.05$) underrepresented for loss of function (gene absence, large deletions and nonsense mutations) compared with the environmental genomes. Naïve p values and BH critical values shown, p values < BH critical values in bold

PAO1 locus tag	naïve p value	BH critical value	PAO1 locus tag	naïve p value	BH critical value
PA4586	**1.49E-19**	8.92E-06	PA2719	5.63E-03	4.20E-03
PA4162	**2.75E-18**	1.78E-05	PA2750	5.63E-03	4.21E-03
PA3105	**9.22E-17**	2.68E-05	PA2767	5.63E-03	4.22E-03
PA4192	**1.09E-16**	3.57E-05	PA2778	5.63E-03	4.23E-03
PA4193	**1.09E-16**	4.46E-05	PA2780	5.63E-03	4.24E-03
PA4194	**1.09E-16**	5.35E-05	PA2789	5.63E-03	4.25E-03
PA4195	**1.09E-16**	6.24E-05	PA2803	5.63E-03	4.26E-03
PA4191	**3.04E-16**	7.14E-05	PA2806	5.63E-03	4.26E-03
PA1858	**8.22E-16**	8.03E-05	PA2839	5.63E-03	4.27E-03
PA3469	**2.16E-15**	8.92E-05	PA2852	5.63E-03	4.28E-03
PA5318	**4.56E-15**	9.81E-05	PA2855	5.63E-03	4.29E-03
PA3841	**1.44E-14**	1.07E-04	PA2892	5.63E-03	4.30E-03
PA4879	**3.28E-14**	1.16E-04	PA2958	5.63E-03	4.31E-03
PA3362	**4.91E-14**	1.25E-04	PA3025	5.63E-03	4.32E-03
PA2130	**5.15E-14**	1.34E-04	PA3113	5.63E-03	4.33E-03
PA3368_1	**1.99E-13**	1.43E-04	PA3197	5.63E-03	4.34E-03
PA2771	**5.79E-13**	1.52E-04	PA3198	5.63E-03	4.34E-03
PA2772	**8.97E-13**	1.61E-04	PA3203	5.63E-03	4.35E-03
PA1036	**2.49E-12**	1.69E-04	PA3209	5.63E-03	4.36E-03
PA5321	**6.52E-12**	1.78E-04	PA3238	5.63E-03	4.37E-03
PA4985	**8.84E-12**	1.87E-04	PA3262	5.63E-03	4.38E-03
PA0244	**2.69E-11**	1.96E-04	PA3281	5.63E-03	4.39E-03
PA2182	**3.13E-11**	2.05E-04	PA3289	5.63E-03	4.40E-03
PA0347	**4.21E-11**	2.14E-04	PA3295	5.63E-03	4.41E-03
PA0191	**4.31E-11**	2.23E-04	PA3332	5.63E-03	4.42E-03
PA2583	**6.50E-11**	2.32E-04	PA3334	5.63E-03	4.42E-03
PA2190	**7.91E-11**	2.41E-04	PA3336	5.63E-03	4.43E-03
PA1868	**8.24E-11**	2.50E-04	PA3352	5.63E-03	4.44E-03
PA3387	**2.48E-10**	2.59E-04	PA3355	5.63E-03	4.45E-03
PA4222	**2.48E-10**	2.68E-04	PA3378	5.63E-03	4.46E-03
PA2188	**2.49E-10**	2.77E-04	PA3379	5.63E-03	4.47E-03
PA2183	**2.64E-10**	2.85E-04	PA3401	5.63E-03	4.48E-03
PA2127	**3.11E-10**	2.94E-04	PA3424	5.63E-03	4.49E-03

PA1576	**3.35E-10**	3.03E-04	PA3558	5.63E-03	4.50E-03
PA2184	**3.80E-10**	3.12E-04	PA3559	5.63E-03	4.50E-03
PA2200	**3.90E-10**	3.21E-04	PA3578	5.63E-03	4.51E-03
PA2185	**4.42E-10**	3.30E-04	PA3586	5.63E-03	4.52E-03
PA0261	**5.74E-10**	3.39E-04	PA3662	5.63E-03	4.53E-03
PA5404	**6.17E-10**	3.48E-04	PA3667	5.63E-03	4.54E-03
PA2135	**6.55E-10**	3.57E-04	PA3668	5.63E-03	4.55E-03
PA2112	**7.62E-10**	3.66E-04	PA3706	5.63E-03	4.56E-03
PA4063	**7.64E-10**	3.75E-04	PA3733	5.63E-03	4.57E-03
PA2187	**8.11E-10**	3.84E-04	PA3748	5.63E-03	4.58E-03
PA2296	**1.15E-09**	3.93E-04	PA3793	5.63E-03	4.59E-03
PA2186	**1.35E-09**	4.01E-04	PA3809	5.63E-03	4.59E-03
PA2440	**1.41E-09**	4.10E-04	PA3810	5.63E-03	4.60E-03
PA0499	**3.30E-09**	4.19E-04	PA3830	5.63E-03	4.61E-03
PA2295	**3.39E-09**	4.28E-04	PA3859	5.63E-03	4.62E-03
PA3063	**3.49E-09**	4.37E-04	PA3863	5.63E-03	4.63E-03
PA2167	**3.59E-09**	4.46E-04	PA3870	5.63E-03	4.64E-03
PA2116	**3.64E-09**	4.55E-04	PA3882	5.63E-03	4.65E-03
PA1223	**3.99E-09**	4.64E-04	PA3898	5.63E-03	4.66E-03
PA1865	**4.72E-09**	4.73E-04	PA3903	5.63E-03	4.67E-03
PA4914	**4.72E-09**	4.82E-04	PA3906	5.63E-03	4.67E-03
PA2109	**4.85E-09**	4.91E-04	PA3907	5.63E-03	4.68E-03
PA2164	**5.30E-09**	5.00E-04	PA3926	5.63E-03	4.69E-03
PA5218	**5.47E-09**	5.08E-04	PA3951	5.63E-03	4.70E-03
PA4225	**5.81E-09**	5.17E-04	PA3956	5.63E-03	4.71E-03
PA4344	**7.53E-09**	5.26E-04	PA3964	5.63E-03	4.72E-03
PA1930	**7.86E-09**	5.35E-04	PA3981	5.63E-03	4.73E-03
PA4282	**8.21E-09**	5.44E-04	PA4003	5.63E-03	4.74E-03
PA1225	**9.89E-09**	5.53E-04	PA4033	5.63E-03	4.75E-03
PA2154	**9.89E-09**	5.62E-04	PA4037	5.63E-03	4.75E-03
PA1224	**1.02E-08**	5.71E-04	PA4279	5.63E-03	4.76E-03
PA0497	**1.23E-08**	5.80E-04	PA4284	5.63E-03	4.77E-03
PA0013	**1.60E-08**	5.89E-04	PA4285	5.63E-03	4.78E-03
PA4796	**1.60E-08**	5.98E-04	PA4287	5.63E-03	4.79E-03
PA0498	**1.67E-08**	6.07E-04	PA4302	5.63E-03	4.80E-03
PA3368	**2.22E-08**	6.16E-04	PA4305	5.63E-03	4.81E-03
PA2055	**2.29E-08**	6.24E-04	PA4341	5.63E-03	4.82E-03
PA2834	**2.36E-08**	6.33E-04	PA4342	5.63E-03	4.83E-03
PA5219	**2.36E-08**	6.42E-04	PA4347	5.63E-03	4.83E-03
PA3222	**2.36E-08**	6.51E-04	PA4357	5.63E-03	4.84E-03
PA4849	**3.07E-08**	6.60E-04	PA4365	5.63E-03	4.85E-03
PA0251	**3.58E-08**	6.69E-04	PA4369	5.63E-03	4.86E-03

PA2117	**4.53E-08**	6.78E-04	PA4384	5.63E-03	4.87E-03
PA0258	**5.14E-08**	6.87E-04	PA4414	5.63E-03	4.88E-03
PA1309	**5.45E-08**	6.96E-04	PA4416	5.63E-03	4.89E-03
PA3132	**9.43E-08**	7.05E-04	PA4439	5.63E-03	4.90E-03
PA0557	**9.49E-08**	7.14E-04	PA4459	5.63E-03	4.91E-03
PA0457_1	**1.27E-07**	7.23E-04	PA4472	5.63E-03	4.92E-03
PA0319	**1.31E-07**	7.31E-04	PA4486	5.63E-03	4.92E-03
PA0689	**1.87E-07**	7.40E-04	PA4510	5.63E-03	4.93E-03
PA2005	**2.15E-07**	7.49E-04	PA4542	5.63E-03	4.94E-03
PA2567	**3.74E-07**	7.58E-04	PA4575	5.63E-03	4.95E-03
PA2790	**3.74E-07**	7.67E-04	PA4591	5.63E-03	4.96E-03
PA5548	**3.83E-07**	7.76E-04	PA4596	5.63E-03	4.97E-03
PA5342	**4.34E-07**	7.85E-04	PA4612	5.63E-03	4.98E-03
PA3032	**5.82E-07**	7.94E-04	PA4627	5.63E-03	4.99E-03
PA3065	**7.33E-07**	8.03E-04	PA4669	5.63E-03	5.00E-03
PA3067	**7.33E-07**	8.12E-04	PA4721	5.63E-03	5.00E-03
PA2119	**8.50E-07**	8.21E-04	PA4725	5.63E-03	5.01E-03
PA2073	**8.91E-07**	8.30E-04	PA4727	5.63E-03	5.02E-03
PA2597	**9.19E-07**	8.39E-04	PA4728	5.63E-03	5.03E-03
PA3317	**9.19E-07**	8.47E-04	PA4730	5.63E-03	5.04E-03
PA3312	**1.11E-06**	8.56E-04	PA4750	5.63E-03	5.05E-03
PA2074	**1.14E-06**	8.65E-04	PA4767	5.63E-03	5.06E-03
PA1118	**1.30E-06**	8.74E-04	PA4778	5.63E-03	5.07E-03
PA2253	**1.34E-06**	8.83E-04	PA4792	5.63E-03	5.08E-03
PA0361	**1.38E-06**	8.92E-04	PA4813	5.63E-03	5.08E-03
PA0257	**1.89E-06**	9.01E-04	PA4820	5.63E-03	5.09E-03
PA4265	**2.01E-06**	9.10E-04	PA4824	5.63E-03	5.10E-03
PA4930	**2.66E-06**	9.19E-04	PA4834	5.63E-03	5.11E-03
PA0721	**3.63E-06**	9.28E-04	PA4871	5.63E-03	5.12E-03
PA2594	**4.77E-06**	9.37E-04	PA4877	5.63E-03	5.13E-03
PA2835	**4.77E-06**	9.46E-04	PA4979	5.63E-03	5.14E-03
PA3066	**5.03E-06**	9.55E-04	PA4980	5.63E-03	5.15E-03
PA1351	**5.48E-06**	9.63E-04	PA4983	5.63E-03	5.16E-03
PA0718	**6.67E-06**	9.72E-04	PA4984	5.63E-03	5.17E-03
PA4177	**9.57E-06**	9.81E-04	PA4992	5.63E-03	5.17E-03
PA1933	**1.08E-05**	9.90E-04	PA5002	5.63E-03	5.18E-03
PA1314	**1.11E-05**	9.99E-04	PA5023	5.63E-03	5.19E-03
PA5266	**1.15E-05**	1.01E-03	PA5026	5.63E-03	5.20E-03
PA2598	**1.40E-05**	1.02E-03	PA5038	5.63E-03	5.21E-03
PA2599	**1.40E-05**	1.03E-03	PA5073	5.63E-03	5.22E-03
PA2601	**1.40E-05**	1.03E-03	PA5079	5.63E-03	5.23E-03
PA1316	**1.55E-05**	1.04E-03	PA5082	5.63E-03	5.24E-03

PA2336	**1.56E-05**	1.05E-03	PA5144	5.63E-03	5.25E-03
PA2335	**1.60E-05**	1.06E-03	PA5157	5.63E-03	5.25E-03
PA1771	**1.71E-05**	1.07E-03	PA5185	5.63E-03	5.26E-03
PA4127	**1.77E-05**	1.08E-03	PA5196	5.63E-03	5.27E-03
PA1348	**2.03E-05**	1.09E-03	PA5220	5.63E-03	5.28E-03
PA2503	**2.03E-05**	1.10E-03	PA5230	5.63E-03	5.29E-03
PA2774	**2.03E-05**	1.11E-03	PA5249	5.63E-03	5.30E-03
PA5184	**2.29E-05**	1.12E-03	PA5277	5.63E-03	5.31E-03
PA3298	**2.40E-05**	1.12E-03	PA5294	5.63E-03	5.32E-03
PA1169	**2.56E-05**	1.13E-03	PA5310	5.63E-03	5.33E-03
PA0014	**2.80E-05**	1.14E-03	PA5330	5.63E-03	5.33E-03
PA2220	**2.93E-05**	1.15E-03	PA5349	5.63E-03	5.34E-03
PA1488	**3.26E-05**	1.16E-03	PA5352	5.63E-03	5.35E-03
PA3340	**3.32E-05**	1.17E-03	PA5357	5.63E-03	5.36E-03
PA1312	**3.46E-05**	1.18E-03	PA5358	5.63E-03	5.37E-03
PA2192	**3.56E-05**	1.19E-03	PA5379	5.63E-03	5.38E-03
PA1984	**3.62E-05**	1.20E-03	PA5383	5.63E-03	5.39E-03
PA1932	**5.01E-05**	1.20E-03	PA5476	5.63E-03	5.40E-03
PA2602	**6.55E-05**	1.21E-03	PA5495	5.63E-03	5.41E-03
PA5542	**7.31E-05**	1.22E-03	PA5502	5.63E-03	5.41E-03
PA0206	**7.43E-05**	1.23E-03	PA5517	5.63E-03	5.42E-03
PA2294	**7.58E-05**	1.24E-03	PA5532	5.63E-03	5.43E-03
PA2334	**7.74E-05**	1.25E-03	PA5543	5.63E-03	5.44E-03
PA2669	**9.39E-05**	1.26E-03	PA0095	5.74E-03	5.45E-03
PA1922	**1.02E-04**	1.27E-03	PA4119	5.74E-03	5.46E-03
PA2096	**1.02E-04**	1.28E-03	PA2175	5.74E-03	5.47E-03
PA1397	**1.04E-04**	1.28E-03	PA4101	5.92E-03	5.48E-03
PA2180	**1.07E-04**	1.29E-03	PA0631	6.08E-03	5.49E-03
PA2333	**1.12E-04**	1.30E-03	PA1297	6.11E-03	5.50E-03
PA0573	**1.13E-04**	1.31E-03	PA2145	6.11E-03	5.50E-03
PA1315	**1.19E-04**	1.32E-03	PA2682	6.11E-03	5.51E-03
PA0205	**1.39E-04**	1.33E-03	PA4805	6.11E-03	5.52E-03
PA0207	**1.39E-04**	1.34E-03	PA0246	6.23E-03	5.53E-03
PA2155	**1.42E-04**	1.35E-03	PA1247	6.23E-03	5.54E-03
PA2596	**1.42E-04**	1.36E-03	PA1599	6.23E-03	5.55E-03
PA2158	**1.44E-04**	1.36E-03	PA1970	6.23E-03	5.56E-03
PA4802	**1.64E-04**	1.37E-03	PA2922	6.23E-03	5.57E-03
PA0202	**1.70E-04**	1.38E-03	PA4960	6.23E-03	5.58E-03
PA2603	**1.97E-04**	1.39E-03	PA5099	6.23E-03	5.58E-03
PA0844	**2.03E-04**	1.40E-03	PA5442	6.23E-03	5.59E-03
PA3884	**2.03E-04**	1.41E-03	PA0092	6.59E-03	5.60E-03
PA1471	**2.15E-04**	1.42E-03	PA2315	6.87E-03	5.61E-03

PA1472	**2.15E-04**	1.43E-03	PA0093	7.46E-03	5.62E-03
PA4554	**2.22E-04**	1.44E-03	PA0101	7.58E-03	5.63E-03
PA2170	**2.35E-04**	1.45E-03	PA1195	7.59E-03	5.64E-03
PA3053	**2.53E-04**	1.45E-03	PA0187	8.38E-03	5.65E-03
PA0203	**2.54E-04**	1.46E-03	PA4798	8.41E-03	5.66E-03
PA1349	**2.55E-04**	1.47E-03	PA0183	8.62E-03	5.66E-03
PA1569	**2.64E-04**	1.48E-03	PA0199	8.62E-03	5.67E-03
PA0688	**2.67E-04**	1.49E-03	PA0997	8.62E-03	5.68E-03
PA1196	**2.72E-04**	1.50E-03	PA1145	8.62E-03	5.69E-03
PA3354	**2.72E-04**	1.51E-03	PA1286	8.62E-03	5.70E-03
PA5427	**2.72E-04**	1.52E-03	PA1480	8.62E-03	5.71E-03
PA1313	**2.90E-04**	1.53E-03	PA1508	8.62E-03	5.72E-03
PA2159	**2.90E-04**	1.53E-03	PA1531	8.62E-03	5.73E-03
PA1079	**3.10E-04**	1.54E-03	PA2060	8.62E-03	5.74E-03
PA1841	**3.10E-04**	1.55E-03	PA2431	8.62E-03	5.74E-03
PA2677	**3.10E-04**	1.56E-03	PA2470	8.62E-03	5.75E-03
PA3715	**3.10E-04**	1.57E-03	PA3170	8.62E-03	5.76E-03
PA4806	**3.10E-04**	1.58E-03	PA3321	8.62E-03	5.77E-03
PA2177	**3.27E-04**	1.59E-03	PA3796	8.62E-03	5.78E-03
PA0096	**3.42E-04**	1.60E-03	PA4573	8.62E-03	5.79E-03
PA0295	**3.57E-04**	1.61E-03	PA5101	8.62E-03	5.80E-03
PA0737	**3.57E-04**	1.61E-03	PA5211	8.62E-03	5.81E-03
PA4603	**3.57E-04**	1.62E-03	PA1246	8.71E-03	5.82E-03
PA0719	**3.63E-04**	1.63E-03	PA2838	8.71E-03	5.83E-03
PA0734	**3.71E-04**	1.64E-03	PA3714	8.71E-03	5.83E-03
PA1366	**3.73E-04**	1.65E-03	PA4083	8.71E-03	5.84E-03
PA0791	**3.77E-04**	1.66E-03	PA4126	8.71E-03	5.85E-03
PA1132	**3.77E-04**	1.67E-03	PA4900	8.71E-03	5.86E-03
PA1740	**3.77E-04**	1.68E-03	PA1797	8.85E-03	5.87E-03
PA2636	**3.77E-04**	1.69E-03	PA2218	9.47E-03	5.88E-03
PA2786	**3.77E-04**	1.69E-03	PA1914	9.94E-03	5.89E-03
PA3957	**3.77E-04**	1.70E-03	PA0099	1.02E-02	5.90E-03
PA1827	**3.91E-04**	1.71E-03	PA2807	1.09E-02	5.91E-03
PA1931	**3.96E-04**	1.72E-03	PA4147	1.09E-02	5.91E-03
PA4482	**3.99E-04**	1.73E-03	PA4151	1.09E-02	5.92E-03
PA2195	**4.11E-04**	1.74E-03	PA3961	1.18E-02	5.93E-03
PA4577	**4.31E-04**	1.75E-03	PA0121	1.18E-02	5.94E-03
PA0204	**4.52E-04**	1.76E-03	PA0323	1.18E-02	5.95E-03
PA1194	**4.99E-04**	1.77E-03	PA0348	1.18E-02	5.96E-03
PA1346	**4.99E-04**	1.78E-03	PA0415	1.18E-02	5.97E-03
PA1086	**5.43E-04**	1.78E-03	PA0803	1.18E-02	5.98E-03
PA1088	**5.43E-04**	1.79E-03	PA1020	1.18E-02	5.99E-03

PA1089	**5.43E-04**	1.80E-03	PA1056	1.18E-02	5.99E-03
PA1090	**5.43E-04**	1.81E-03	PA1168	1.18E-02	6.00E-03
PA1095	**5.73E-04**	1.82E-03	PA1211	1.18E-02	6.01E-03
PA0996	**5.74E-04**	1.83E-03	PA1212	1.18E-02	6.02E-03
PA2760	**5.74E-04**	1.84E-03	PA1217	1.18E-02	6.03E-03
PA0574	**6.00E-04**	1.85E-03	PA1218	1.18E-02	6.04E-03
PA0051	**6.63E-04**	1.86E-03	PA1242	1.18E-02	6.05E-03
PA1096	**6.89E-04**	1.86E-03	PA1260	1.18E-02	6.06E-03
PA1021	**7.11E-04**	1.87E-03	PA1266	1.18E-02	6.07E-03
PA5284	**7.71E-04**	1.88E-03	PA1280	1.18E-02	6.07E-03
PA4149	**8.01E-04**	1.89E-03	PA1515	1.18E-02	6.08E-03
PA2673	**8.62E-04**	1.90E-03	PA1693	1.18E-02	6.09E-03
PA4649	**8.62E-04**	1.91E-03	PA1729	1.18E-02	6.10E-03
PA1091	**8.91E-04**	1.92E-03	PA1810	1.18E-02	6.11E-03
PA1271	**9.22E-04**	1.93E-03	PA1898	1.18E-02	6.12E-03
PA4100	**9.38E-04**	1.94E-03	PA2430	1.18E-02	6.13E-03
PA0819	**9.83E-04**	1.94E-03	PA2524	1.18E-02	6.14E-03
PA1093	**9.91E-04**	1.95E-03	PA2920	1.18E-02	6.15E-03
PA1094	**9.91E-04**	1.96E-03	PA2927	1.18E-02	6.16E-03
PA2131	**1.02E-03**	1.97E-03	PA2931	1.18E-02	6.16E-03
PA1001	**1.09E-03**	1.98E-03	PA2934	1.18E-02	6.17E-03
PA1015	**1.09E-03**	1.99E-03	PA2940	1.18E-02	6.18E-03
PA2665	**1.09E-03**	2.00E-03	PA3175	1.18E-02	6.19E-03
PA2678	**1.09E-03**	2.01E-03	PA3207	1.18E-02	6.20E-03
PA2924	**1.09E-03**	2.02E-03	PA3329	1.18E-02	6.21E-03
PA2926	**1.09E-03**	2.02E-03	PA3669	1.18E-02	6.22E-03
PA3871	**1.09E-03**	2.03E-03	PA3910	1.18E-02	6.23E-03
PA4539	**1.09E-03**	2.04E-03	PA4239	1.18E-02	6.24E-03
PA5399	**1.09E-03**	2.05E-03	PA4278	1.18E-02	6.24E-03
PA4896	**1.10E-03**	2.06E-03	PA4374	1.18E-02	6.25E-03
PA0995	**1.11E-03**	2.07E-03	PA4448	1.18E-02	6.26E-03
PA1350	**1.11E-03**	2.08E-03	PA4650	1.18E-02	6.27E-03
PA2764	**1.13E-03**	2.09E-03	PA5102	1.18E-02	6.28E-03
PA1149	**1.18E-03**	2.10E-03	PA5181	1.18E-02	6.29E-03
PA1092	**1.19E-03**	2.11E-03	PA5260	1.18E-02	6.30E-03
PA2172	**1.24E-03**	2.11E-03	PA5466	1.18E-02	6.31E-03
PA2291	**1.26E-03**	2.12E-03	PA0100	1.18E-02	6.32E-03
PA4102	**1.31E-03**	2.13E-03	PA1147	1.24E-02	6.32E-03
PA4103	**1.31E-03**	2.14E-03	PA1251	1.24E-02	6.33E-03
PA0400	**1.32E-03**	2.15E-03	PA2676	1.24E-02	6.34E-03
PA1680	**1.32E-03**	2.16E-03	PA2921	1.24E-02	6.35E-03
PA2675	**1.32E-03**	2.17E-03	PA3842	1.24E-02	6.36E-03

PA2722	**1.32E-03**	2.18E-03	PA4084	1.27E-02	6.37E-03
PA2723	**1.32E-03**	2.19E-03	PA0690	1.31E-02	6.38E-03
PA2775	**1.32E-03**	2.19E-03	PA2360	1.31E-02	6.39E-03
PA2908	**1.32E-03**	2.20E-03	PA2191	1.37E-02	6.40E-03
PA4826	**1.32E-03**	2.21E-03	PA1511	1.45E-02	6.40E-03
PA5254	**1.32E-03**	2.22E-03	PA1606	1.54E-02	6.41E-03
PA1087	**1.41E-03**	2.23E-03	PA1715	1.57E-02	6.42E-03
PA1681	**1.46E-03**	2.24E-03	PA2125	1.57E-02	6.43E-03
PA5034	**1.46E-03**	2.25E-03	PA2150	1.57E-02	6.44E-03
PA5314	**1.46E-03**	2.26E-03	PA4080	1.57E-02	6.45E-03
PA0508	**1.48E-03**	2.27E-03	PA5252	1.57E-02	6.46E-03
PA0774	**1.48E-03**	2.27E-03	PA0078	1.58E-02	6.47E-03
PA0828	**1.48E-03**	2.28E-03	PA0112	1.58E-02	6.48E-03
PA0939	**1.48E-03**	2.29E-03	PA0133	1.58E-02	6.49E-03
PA1133	**1.48E-03**	2.30E-03	PA0135	1.58E-02	6.49E-03
PA1235	**1.48E-03**	2.31E-03	PA0168	1.58E-02	6.50E-03
PA1293	**1.48E-03**	2.32E-03	PA0170	1.58E-02	6.51E-03
PA1818	**1.48E-03**	2.33E-03	PA0175	1.58E-02	6.52E-03
PA2441	**1.48E-03**	2.34E-03	PA0184	1.58E-02	6.53E-03
PA2556	**1.48E-03**	2.35E-03	PA0186	1.58E-02	6.54E-03
PA2681	**1.48E-03**	2.36E-03	PA0270	1.58E-02	6.55E-03
PA2765	**1.48E-03**	2.36E-03	PA0271	1.58E-02	6.56E-03
PA3135	**1.48E-03**	2.37E-03	PA0297	1.58E-02	6.57E-03
PA3331	**1.48E-03**	2.38E-03	PA0325	1.58E-02	6.57E-03
PA3339	**1.48E-03**	2.39E-03	PA0467	1.58E-02	6.58E-03
PA3827	**1.48E-03**	2.40E-03	PA0482	1.58E-02	6.59E-03
PA3902	**1.48E-03**	2.41E-03	PA0486	1.58E-02	6.60E-03
PA3908	**1.48E-03**	2.42E-03	PA0535	1.58E-02	6.61E-03
PA3920	**1.48E-03**	2.43E-03	PA0590	1.58E-02	6.62E-03
PA3969	**1.48E-03**	2.44E-03	PA0610	1.58E-02	6.63E-03
PA4220	**1.48E-03**	2.44E-03	PA0666	1.58E-02	6.64E-03
PA4405	**1.48E-03**	2.45E-03	PA0672	1.58E-02	6.65E-03
PA4580	**1.48E-03**	2.46E-03	PA0732	1.58E-02	6.65E-03
PA4708	**1.48E-03**	2.47E-03	PA0733	1.58E-02	6.66E-03
PA4832	**1.48E-03**	2.48E-03	PA0736	1.58E-02	6.67E-03
PA5231	**1.48E-03**	2.49E-03	PA0843	1.58E-02	6.68E-03
PA5329	**1.48E-03**	2.50E-03	PA0862	1.58E-02	6.69E-03
PA5359	**1.48E-03**	2.51E-03	PA0864	1.58E-02	6.70E-03
PA5361	**1.48E-03**	2.52E-03	PA0879	1.58E-02	6.71E-03
PA5511	**1.48E-03**	2.52E-03	PA0893	1.58E-02	6.72E-03
PA5519	**1.48E-03**	2.53E-03	PA0894	1.58E-02	6.73E-03
PA5520	**1.48E-03**	2.54E-03	PA0925	1.58E-02	6.74E-03

PA2514	**1.53E-03**	2.55E-03	PA0927	1.58E-02	6.74E-03
PA4585	**1.53E-03**	2.56E-03	PA0951	1.58E-02	6.75E-03
PA2126	**1.60E-03**	2.57E-03	PA0952	1.58E-02	6.76E-03
PA5267	**1.60E-03**	2.58E-03	PA0955	1.58E-02	6.77E-03
PA1826	**1.78E-03**	2.59E-03	PA0998	1.58E-02	6.78E-03
PA0089	**2.08E-03**	2.60E-03	PA1028	1.58E-02	6.79E-03
PA0842	**2.08E-03**	2.60E-03	PA1038	1.58E-02	6.80E-03
PA1646	**2.08E-03**	2.61E-03	PA1043	1.58E-02	6.81E-03
PA1917	**2.08E-03**	2.62E-03	PA1054	1.58E-02	6.82E-03
PA2018	**2.08E-03**	2.63E-03	PA1066	1.58E-02	6.82E-03
PA3843	**2.08E-03**	2.64E-03	PA1069	1.58E-02	6.83E-03
PA4118	**2.08E-03**	2.65E-03	PA1109	1.58E-02	6.84E-03
PA2219	**2.21E-03**	2.66E-03	PA1110	1.58E-02	6.85E-03
PA4148	**2.27E-03**	2.67E-03	PA1136	1.58E-02	6.86E-03
PA2600	**2.56E-03**	2.68E-03	PA1137	1.58E-02	6.87E-03
PA3294	2.71E-03	2.69E-03	PA1163	1.58E-02	6.88E-03
PA0480	2.79E-03	2.69E-03	PA1205	1.58E-02	6.89E-03
PA0496	2.79E-03	2.70E-03	PA1231	1.58E-02	6.90E-03
PA2002	2.79E-03	2.71E-03	PA1237	1.58E-02	6.90E-03
PA2546	2.79E-03	2.72E-03	PA1244	1.58E-02	6.91E-03
PA2563	2.79E-03	2.73E-03	PA1250	1.58E-02	6.92E-03
PA2672	2.79E-03	2.74E-03	PA1285	1.58E-02	6.93E-03
PA2925	2.79E-03	2.75E-03	PA1289	1.58E-02	6.94E-03
PA3176	2.79E-03	2.76E-03	PA1290	1.58E-02	6.95E-03
PA5097	2.79E-03	2.77E-03	PA1330	1.58E-02	6.96E-03
PA4690_5	2.92E-03	2.77E-03	PA1446	1.58E-02	6.97E-03
PA0759	2.98E-03	2.78E-03	PA1467	1.58E-02	6.98E-03
PA4104	3.00E-03	2.79E-03	PA1474	1.58E-02	6.98E-03
PA2417	3.00E-03	2.80E-03	PA1491	1.58E-02	6.99E-03
PA4150	3.00E-03	2.81E-03	PA1537	1.58E-02	7.00E-03
PA0262	3.05E-03	2.82E-03	PA1542	1.58E-02	7.01E-03
PA0696	3.13E-03	2.83E-03	PA1563	1.58E-02	7.02E-03
PA0910	3.19E-03	2.84E-03	PA1592	1.58E-02	7.03E-03
PA2128	3.19E-03	2.85E-03	PA1600	1.58E-02	7.04E-03
PA5032	3.19E-03	2.85E-03	PA1604	1.58E-02	7.05E-03
PA5113	3.19E-03	2.86E-03	PA1652	1.58E-02	7.06E-03
PA4105	3.25E-03	2.87E-03	PA1653	1.58E-02	7.07E-03
PA4107	3.65E-03	2.88E-03	PA1672	1.58E-02	7.07E-03
PA0050	3.67E-03	2.89E-03	PA1689	1.58E-02	7.08E-03
PA0273	3.67E-03	2.90E-03	PA1739	1.58E-02	7.09E-03
PA1014	3.67E-03	2.91E-03	PA1792	1.58E-02	7.10E-03
PA1131	3.67E-03	2.92E-03	PA1817	1.58E-02	7.11E-03

PA2437	3.67E-03	2.93E-03	PA1850	1.58E-02	7.12E-03
PA2923	3.67E-03	2.93E-03	PA1861	1.58E-02	7.13E-03
PA3125	3.67E-03	2.94E-03	PA1876	1.58E-02	7.14E-03
PA3140	3.67E-03	2.95E-03	PA1885	1.58E-02	7.15E-03
PA3589	3.67E-03	2.96E-03	PA1886	1.58E-02	7.15E-03
PA3829	3.67E-03	2.97E-03	PA1891	1.58E-02	7.16E-03
PA3840	3.67E-03	2.98E-03	PA1895	1.58E-02	7.17E-03
PA4498	3.67E-03	2.99E-03	PA2015	1.58E-02	7.18E-03
PA5477	3.67E-03	3.00E-03	PA2499	1.58E-02	7.19E-03
PA1292	3.68E-03	3.01E-03	PA2506	1.58E-02	7.20E-03
PA3288	3.86E-03	3.02E-03	PA2516	1.58E-02	7.21E-03
PA2152	4.28E-03	3.02E-03	PA2547	1.58E-02	7.22E-03
PA2165	4.28E-03	3.03E-03	PA2569	1.58E-02	7.23E-03
PA0292	4.47E-03	3.04E-03	PA2570	1.58E-02	7.23E-03
PA0479	4.47E-03	3.05E-03	PA2579	1.58E-02	7.24E-03
PA1987	4.47E-03	3.06E-03	PA2582	1.58E-02	7.25E-03
PA2421	4.47E-03	3.07E-03	PA2588	1.58E-02	7.26E-03
PA2930	4.47E-03	3.08E-03	PA2591	1.58E-02	7.27E-03
PA4088	4.47E-03	3.09E-03	PA2664	1.58E-02	7.28E-03
PA4540	4.47E-03	3.10E-03	PA2679	1.58E-02	7.29E-03
PA4584	4.47E-03	3.10E-03	PA2721	1.58E-02	7.30E-03
PA0630	4.68E-03	3.11E-03	PA2724	1.58E-02	7.31E-03
PA0674	4.68E-03	3.12E-03	PA2761	1.58E-02	7.31E-03
PA0753	4.68E-03	3.13E-03	PA2766	1.58E-02	7.32E-03
PA1120	4.68E-03	3.14E-03	PA2810	1.58E-02	7.33E-03
PA1875	4.68E-03	3.15E-03	PA2881	1.58E-02	7.34E-03
PA1884	4.68E-03	3.16E-03	PA2887	1.58E-02	7.35E-03
PA2522	4.68E-03	3.17E-03	PA2888	1.58E-02	7.36E-03
PA2716	4.68E-03	3.18E-03	PA2900	1.58E-02	7.37E-03
PA2720	4.68E-03	3.18E-03	PA2959	1.58E-02	7.38E-03
PA2785	4.68E-03	3.19E-03	PA2969	1.58E-02	7.39E-03
PA2837	4.68E-03	3.20E-03	PA3016	1.58E-02	7.40E-03
PA3035	4.68E-03	3.21E-03	PA3034	1.58E-02	7.40E-03
PA3078	4.68E-03	3.22E-03	PA3056	1.58E-02	7.41E-03
PA3738	4.68E-03	3.23E-03	PA3080	1.58E-02	7.42E-03
PA3771	4.68E-03	3.24E-03	PA3196	1.58E-02	7.43E-03
PA3806	4.68E-03	3.25E-03	PA3219	1.58E-02	7.44E-03
PA4478	4.68E-03	3.26E-03	PA3243	1.58E-02	7.45E-03
PA4613	4.68E-03	3.26E-03	PA3267	1.58E-02	7.46E-03
PA4651	4.68E-03	3.27E-03	PA3310	1.58E-02	7.47E-03
PA4716	4.68E-03	3.28E-03	PA3325	1.58E-02	7.48E-03
PA4785	4.68E-03	3.29E-03	PA3326	1.58E-02	7.48E-03

PA5017	4.68E-03	3.30E-03	PA3330	1.58E-02	7.49E-03
PA5393	4.68E-03	3.31E-03	PA3335	1.58E-02	7.50E-03
PA1283	4.79E-03	3.32E-03	PA3341	1.58E-02	7.51E-03
PA4653	4.79E-03	3.33E-03	PA3358	1.58E-02	7.52E-03
PA4106	4.90E-03	3.34E-03	PA3365	1.58E-02	7.53E-03
PA0188	5.00E-03	3.35E-03	PA3386	1.58E-02	7.54E-03
PA2129	5.29E-03	3.35E-03	PA3474	1.58E-02	7.55E-03
PA0005	5.63E-03	3.36E-03	PA3515	1.58E-02	7.56E-03
PA0007	5.63E-03	3.37E-03	PA3518	1.58E-02	7.56E-03
PA0012	5.63E-03	3.38E-03	PA3539	1.58E-02	7.57E-03
PA0020	5.63E-03	3.39E-03	PA3549	1.58E-02	7.58E-03
PA0028	5.63E-03	3.40E-03	PA3552	1.58E-02	7.59E-03
PA0048	5.63E-03	3.41E-03	PA3555	1.58E-02	7.60E-03
PA0070	5.63E-03	3.42E-03	PA3557	1.58E-02	7.61E-03
PA0072	5.63E-03	3.43E-03	PA3584	1.58E-02	7.62E-03
PA0075	5.63E-03	3.43E-03	PA3588	1.58E-02	7.63E-03
PA0122	5.63E-03	3.44E-03	PA3591	1.58E-02	7.64E-03
PA0276	5.63E-03	3.45E-03	PA3596	1.58E-02	7.64E-03
PA0293	5.63E-03	3.46E-03	PA3613	1.58E-02	7.65E-03
PA0316	5.63E-03	3.47E-03	PA3643	1.58E-02	7.66E-03
PA0324	5.63E-03	3.48E-03	PA3671	1.58E-02	7.67E-03
PA0346	5.63E-03	3.49E-03	PA3721	1.58E-02	7.68E-03
PA0357	5.63E-03	3.50E-03	PA3795	1.58E-02	7.69E-03
PA0368	5.63E-03	3.51E-03	PA3826	1.58E-02	7.70E-03
PA0397	5.63E-03	3.51E-03	PA3905	1.58E-02	7.71E-03
PA0405	5.63E-03	3.52E-03	PA3915	1.58E-02	7.72E-03
PA0412	5.63E-03	3.53E-03	PA3944	1.58E-02	7.73E-03
PA0421	5.63E-03	3.54E-03	PA3954	1.58E-02	7.73E-03
PA0422	5.63E-03	3.55E-03	PA3984	1.58E-02	7.74E-03
PA0431	5.63E-03	3.56E-03	PA3985	1.58E-02	7.75E-03
PA0463	5.63E-03	3.57E-03	PA3988	1.58E-02	7.76E-03
PA0464	5.63E-03	3.58E-03	PA4028	1.58E-02	7.77E-03
PA0492	5.63E-03	3.59E-03	PA4111	1.58E-02	7.78E-03
PA0493	5.63E-03	3.60E-03	PA4229	1.58E-02	7.79E-03
PA0529	5.63E-03	3.60E-03	PA4281	1.58E-02	7.80E-03
PA0530	5.63E-03	3.61E-03	PA4340	1.58E-02	7.81E-03
PA0531	5.63E-03	3.62E-03	PA4392	1.58E-02	7.81E-03
PA0539	5.63E-03	3.63E-03	PA4423	1.58E-02	7.82E-03
PA0546	5.63E-03	3.64E-03	PA4424	1.58E-02	7.83E-03
PA0556	5.63E-03	3.65E-03	PA4428	1.58E-02	7.84E-03
PA0584	5.63E-03	3.66E-03	PA4543	1.58E-02	7.85E-03
PA0649	5.63E-03	3.67E-03	PA4565	1.58E-02	7.86E-03

PA0663	5.63E-03	3.68E-03	PA4601	1.58E-02	7.87E-03
PA0667	5.63E-03	3.68E-03	PA4602	1.58E-02	7.88E-03
PA0772	5.63E-03	3.69E-03	PA4640	1.58E-02	7.89E-03
PA0784	5.63E-03	3.70E-03	PA4663	1.58E-02	7.89E-03
PA0806	5.63E-03	3.71E-03	PA4682	1.58E-02	7.90E-03
PA0827	5.63E-03	3.72E-03	PA4700	1.58E-02	7.91E-03
PA0829	5.63E-03	3.73E-03	PA4705	1.58E-02	7.92E-03
PA0831	5.63E-03	3.74E-03	PA4706	1.58E-02	7.93E-03
PA0841	5.63E-03	3.75E-03	PA4772	1.58E-02	7.94E-03
PA0846	5.63E-03	3.76E-03	PA4777	1.58E-02	7.95E-03
PA0852	5.63E-03	3.76E-03	PA4800	1.58E-02	7.96E-03
PA0882	5.63E-03	3.77E-03	PA4814	1.58E-02	7.97E-03
PA0921	5.63E-03	3.78E-03	PA4838	1.58E-02	7.98E-03
PA0932	5.63E-03	3.79E-03	PA4878	1.58E-02	7.98E-03
PA0942	5.63E-03	3.80E-03	PA4894	1.58E-02	7.99E-03
PA0946	5.63E-03	3.81E-03	PA4905	1.58E-02	8.00E-03
PA0953	5.63E-03	3.82E-03	PA4949	1.58E-02	8.01E-03
PA0975	5.63E-03	3.83E-03	PA4950	1.58E-02	8.02E-03
PA1031	5.63E-03	3.84E-03	PA4975	1.58E-02	8.03E-03
PA1049	5.63E-03	3.84E-03	PA4976	1.58E-02	8.04E-03
PA1051	5.63E-03	3.85E-03	PA4978	1.58E-02	8.05E-03
PA1053	5.63E-03	3.86E-03	PA4982	1.58E-02	8.06E-03
PA1065	5.63E-03	3.87E-03	PA4994	1.58E-02	8.06E-03
PA1085	5.63E-03	3.88E-03	PA5001	1.58E-02	8.07E-03
PA1138	5.63E-03	3.89E-03	PA5027	1.58E-02	8.08E-03
PA1162	5.63E-03	3.90E-03	PA5047	1.58E-02	8.09E-03
PA1185	5.63E-03	3.91E-03	PA5092	1.58E-02	8.10E-03
PA1226	5.63E-03	3.92E-03	PA5109	1.58E-02	8.11E-03
PA1227	5.63E-03	3.93E-03	PA5131	1.58E-02	8.12E-03
PA1228	5.63E-03	3.93E-03	PA5132	1.58E-02	8.13E-03
PA1233	5.63E-03	3.94E-03	PA5163	1.58E-02	8.14E-03
PA1234	5.63E-03	3.95E-03	PA5177	1.58E-02	8.14E-03
PA1322	5.63E-03	3.96E-03	PA5183	1.58E-02	8.15E-03
PA1363	5.63E-03	3.97E-03	PA5234	1.58E-02	8.16E-03
PA1473	5.63E-03	3.98E-03	PA5248	1.58E-02	8.17E-03
PA1528	5.63E-03	3.99E-03	PA5313	1.58E-02	8.18E-03
PA1529	5.63E-03	4.00E-03	PA5328	1.58E-02	8.19E-03
PA1530	5.63E-03	4.01E-03	PA5457	1.58E-02	8.20E-03
PA1545	5.63E-03	4.01E-03	PA5471	1.58E-02	8.21E-03
PA1561	5.63E-03	4.02E-03	PA5484	1.58E-02	8.22E-03
PA1565	5.63E-03	4.03E-03	PA5487	1.58E-02	8.22E-03
PA1586	5.63E-03	4.04E-03	PA5541	1.58E-02	8.23E-03

PA1635	5.63E-03	4.05E-03	PA0912	1.74E-02	8.24E-03
PA1639	5.63E-03	4.06E-03	PA0118	1.77E-02	8.25E-03
PA1733	5.63E-03	4.07E-03	PA0911	1.77E-02	8.26E-03
PA1753	5.63E-03	4.08E-03	PA1435	1.77E-02	8.27E-03
PA1760	5.63E-03	4.09E-03	PA1710	1.77E-02	8.28E-03
PA1834	5.63E-03	4.09E-03	PA2126_1	1.77E-02	8.29E-03
PA1867	5.63E-03	4.10E-03	PA2568	1.77E-02	8.30E-03
PA1870	5.63E-03	4.11E-03	PA2617	1.77E-02	8.31E-03
PA1877	5.63E-03	4.12E-03	PA2674	1.77E-02	8.31E-03
PA2432	5.63E-03	4.13E-03	PA2932	1.77E-02	8.32E-03
PA2448	5.63E-03	4.14E-03	PA2976	1.77E-02	8.33E-03
PA2560	5.63E-03	4.15E-03	PA3520	1.77E-02	8.34E-03
PA2566_1	5.63E-03	4.16E-03	PA4008	1.77E-02	8.35E-03
PA2580	5.63E-03	4.17E-03	PA4087	1.77E-02	8.36E-03
PA2581	5.63E-03	4.17E-03	PA4089	1.77E-02	8.37E-03
PA2713	5.63E-03	4.18E-03	PA4623	1.77E-02	8.38E-03
PA2714	5.63E-03	4.19E-03	PA4822	1.77E-02	8.39E-03
PA5264	1.77E-02	8.39E-03	PA1705	2.51E-02	1.05E-02
PA0252	1.83E-02	8.40E-03	PA1722	2.51E-02	1.05E-02
PA1359	1.83E-02	8.41E-03	PA1725	2.51E-02	1.05E-02
PA4152	1.83E-02	8.42E-03	PA1738	2.51E-02	1.05E-02
PA4153	1.83E-02	8.43E-03	PA1928	2.51E-02	1.05E-02
PA2137	2.02E-02	8.44E-03	PA2056	2.51E-02	1.05E-02
PA1221	2.04E-02	8.45E-03	PA2062	2.51E-02	1.05E-02
PA4109	2.04E-02	8.46E-03	PA2496	2.51E-02	1.05E-02
PA5191	2.05E-02	8.47E-03	PA2829	2.51E-02	1.05E-02
PA0025	2.05E-02	8.47E-03	PA2909	2.51E-02	1.06E-02
PA0033	2.05E-02	8.48E-03	PA3446	2.51E-02	1.06E-02
PA0037	2.05E-02	8.49E-03	PA3479	2.51E-02	1.06E-02
PA0086	2.05E-02	8.50E-03	PA3734	2.51E-02	1.06E-02
PA0116	2.05E-02	8.51E-03	PA3750	2.51E-02	1.06E-02
PA0328	2.05E-02	8.52E-03	PA4040	2.51E-02	1.06E-02
PA0340	2.05E-02	8.53E-03	PA4296	2.51E-02	1.06E-02
PA0393	2.05E-02	8.54E-03	PA4351	2.51E-02	1.06E-02
PA0460	2.05E-02	8.55E-03	PA4447	2.51E-02	1.06E-02
PA0477	2.05E-02	8.55E-03	PA4648	2.51E-02	1.06E-02
PA0542	2.05E-02	8.56E-03	PA4899	2.51E-02	1.06E-02
PA0562	2.05E-02	8.57E-03	PA4955	2.51E-02	1.07E-02
PA0577	2.05E-02	8.58E-03	PA5006	2.51E-02	1.07E-02
PA0609	2.05E-02	8.59E-03	PA5014	2.51E-02	1.07E-02
PA0658	2.05E-02	8.60E-03	PA5033	2.51E-02	1.07E-02
PA0798	2.05E-02	8.61E-03	PA5118	2.51E-02	1.07E-02

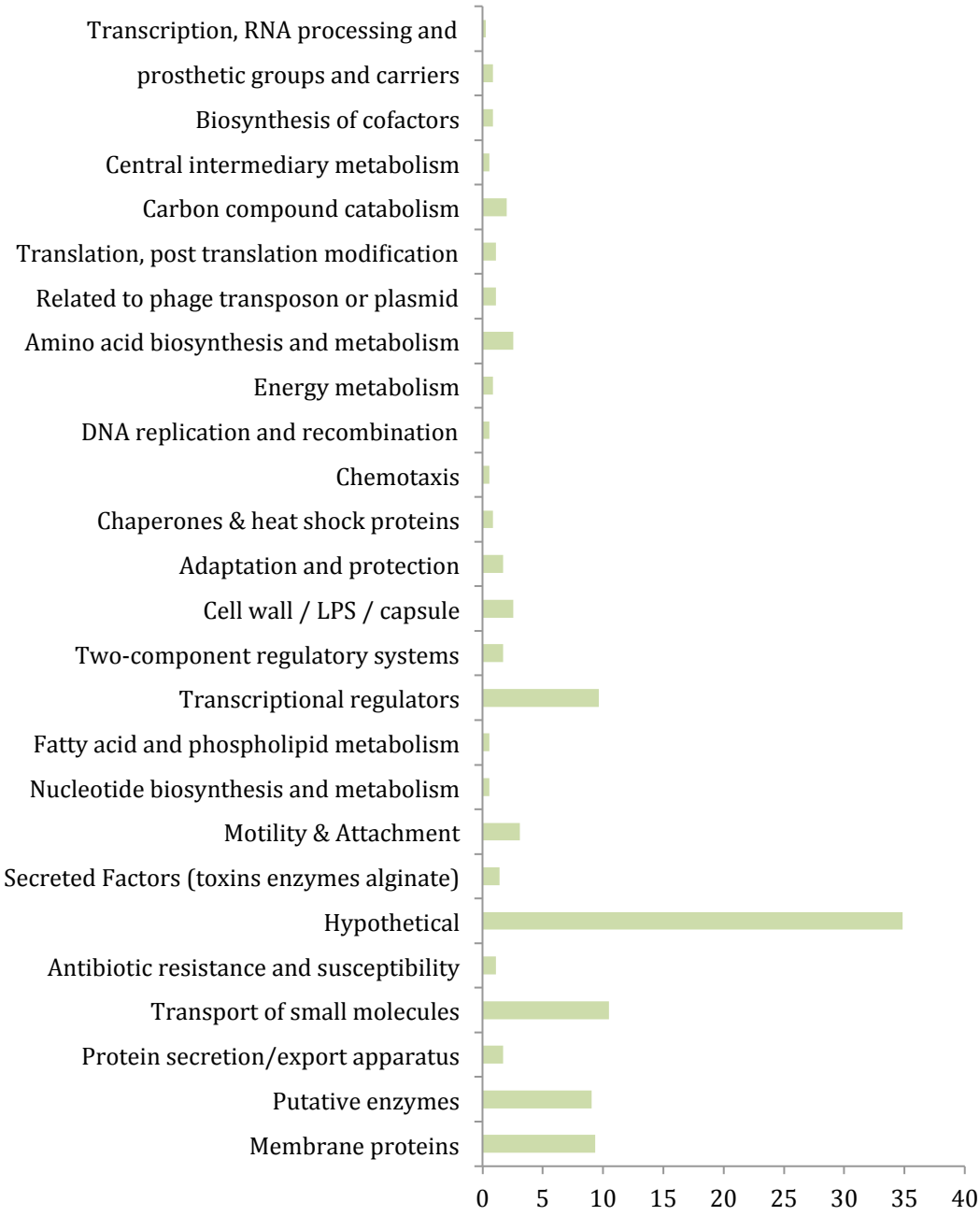
PA0807	2.05E-02	8.62E-03	PA5209	2.51E-02	1.07E-02
PA0808	2.05E-02	8.63E-03	PA0343	2.62E-02	1.07E-02
PA0836	2.05E-02	8.64E-03	PA1352	2.62E-02	1.07E-02
PA0851	2.05E-02	8.64E-03	PA1618	2.62E-02	1.07E-02
PA0886	2.05E-02	8.65E-03	PA2141	2.62E-02	1.07E-02
PA0901	2.05E-02	8.66E-03	PA2144	2.62E-02	1.07E-02
PA0906	2.05E-02	8.67E-03	PA2147	2.62E-02	1.07E-02
PA0930	2.05E-02	8.68E-03	PA2461	2.88E-02	1.08E-02
PA0957	2.05E-02	8.69E-03	PA1148	2.90E-02	1.08E-02
PA1002	2.05E-02	8.70E-03	PA4915	2.90E-02	1.08E-02
PA1004	2.05E-02	8.71E-03	PA3292	2.90E-02	1.08E-02
PA1030	2.05E-02	8.72E-03	PA2138	3.08E-02	1.08E-02
PA1042	2.05E-02	8.72E-03	PA2595	3.17E-02	1.08E-02
PA1050	2.05E-02	8.73E-03	PA2098	3.24E-02	1.08E-02
PA1075	2.05E-02	8.74E-03	PA0804	3.24E-02	1.08E-02
PA1117	2.05E-02	8.75E-03	PA1016	3.24E-02	1.08E-02
PA1329	2.05E-02	8.76E-03	PA1130	3.24E-02	1.08E-02
PA1483	2.05E-02	8.77E-03	PA1146	3.24E-02	1.08E-02
PA1492	2.05E-02	8.78E-03	PA1213	3.24E-02	1.09E-02
PA1496	2.05E-02	8.79E-03	PA2377	3.24E-02	1.09E-02
PA1572	2.05E-02	8.80E-03	PA2439	3.24E-02	1.09E-02
PA1607	2.05E-02	8.80E-03	PA3833	3.24E-02	1.09E-02
PA1615	2.05E-02	8.81E-03	PA4637	3.24E-02	1.09E-02
PA1732	2.05E-02	8.82E-03	PA3923	3.25E-02	1.09E-02
PA1805	2.05E-02	8.83E-03	PA2794	3.26E-02	1.09E-02
PA1844	2.05E-02	8.84E-03	PA5384	3.50E-02	1.09E-02
PA2449	2.05E-02	8.85E-03	PA0030	3.56E-02	1.09E-02
PA2532	2.05E-02	8.86E-03	PA0126	3.56E-02	1.09E-02
PA2535	2.05E-02	8.87E-03	PA0156	3.56E-02	1.09E-02
PA2544	2.05E-02	8.88E-03	PA0173	3.56E-02	1.10E-02
PA2552	2.05E-02	8.88E-03	PA0176	3.56E-02	1.10E-02
PA2558	2.05E-02	8.89E-03	PA0236	3.56E-02	1.10E-02
PA2574	2.05E-02	8.90E-03	PA0238	3.56E-02	1.10E-02
PA2629	2.05E-02	8.91E-03	PA0253	3.56E-02	1.10E-02
PA2649	2.05E-02	8.92E-03	PA0267	3.56E-02	1.10E-02
PA2705	2.05E-02	8.93E-03	PA0451	3.56E-02	1.10E-02
PA2711	2.05E-02	8.94E-03	PA0669	3.56E-02	1.10E-02
PA2718	2.05E-02	8.95E-03	PA0752	3.56E-02	1.10E-02
PA2739	2.05E-02	8.96E-03	PA0758	3.56E-02	1.10E-02
PA2831	2.05E-02	8.97E-03	PA0883	3.56E-02	1.10E-02
PA2850	2.05E-02	8.97E-03	PA0934	3.56E-02	1.11E-02
PA2866	2.05E-02	8.98E-03	PA0935	3.56E-02	1.11E-02

PA2894	2.05E-02	8.99E-03	PA0954	3.56E-02	1.11E-02
PA2898	2.05E-02	9.00E-03	PA0959	3.56E-02	1.11E-02
PA2899	2.05E-02	9.01E-03	PA1000	3.56E-02	1.11E-02
PA2912	2.05E-02	9.02E-03	PA1052	3.56E-02	1.11E-02
PA2937	2.05E-02	9.03E-03	PA1135	3.56E-02	1.11E-02
PA2941	2.05E-02	9.04E-03	PA1166	3.56E-02	1.11E-02
PA2964	2.05E-02	9.05E-03	PA1188	3.56E-02	1.11E-02
PA2972	2.05E-02	9.05E-03	PA1210	3.56E-02	1.11E-02
PA2977	2.05E-02	9.06E-03	PA1364	3.56E-02	1.11E-02
PA2979	2.05E-02	9.07E-03	PA1374	3.56E-02	1.12E-02
PA3008	2.05E-02	9.08E-03	PA1395	3.56E-02	1.12E-02
PA3029	2.05E-02	9.09E-03	PA1469	3.56E-02	1.12E-02
PA3042	2.05E-02	9.10E-03	PA1489	3.56E-02	1.12E-02
PA3045	2.05E-02	9.11E-03	PA1536	3.56E-02	1.12E-02
PA3084	2.05E-02	9.12E-03	PA1578	3.56E-02	1.12E-02
PA3111	2.05E-02	9.13E-03	PA1608	3.56E-02	1.12E-02
PA3127	2.05E-02	9.13E-03	PA1634	3.56E-02	1.12E-02
PA3177	2.05E-02	9.14E-03	PA1656	3.56E-02	1.12E-02
PA3182	2.05E-02	9.15E-03	PA1660	3.56E-02	1.12E-02
PA3233	2.05E-02	9.16E-03	PA1678	3.56E-02	1.12E-02
PA3260	2.05E-02	9.17E-03	PA1679	3.56E-02	1.12E-02
PA3264	2.05E-02	9.18E-03	PA1737	3.56E-02	1.13E-02
PA3273	2.05E-02	9.19E-03	PA1799	3.56E-02	1.13E-02
PA3301	2.05E-02	9.20E-03	PA1853	3.56E-02	1.13E-02
PA3314	2.05E-02	9.21E-03	PA1913	3.56E-02	1.13E-02
PA3428	2.05E-02	9.21E-03	PA1927	3.56E-02	1.13E-02
PA3493	2.05E-02	9.22E-03	PA1929	3.56E-02	1.13E-02
PA3576	2.05E-02	9.23E-03	PA1976	3.56E-02	1.13E-02
PA3597	2.05E-02	9.24E-03	PA1979	3.56E-02	1.13E-02
PA3615	2.05E-02	9.25E-03	PA2028	3.56E-02	1.13E-02
PA3619	2.05E-02	9.26E-03	PA2083	3.56E-02	1.13E-02
PA3678	2.05E-02	9.27E-03	PA2425	3.56E-02	1.13E-02
PA3724	2.05E-02	9.28E-03	PA2466	3.56E-02	1.14E-02
PA3736	2.05E-02	9.29E-03	PA2469	3.56E-02	1.14E-02
PA3820	2.05E-02	9.30E-03	PA2518	3.56E-02	1.14E-02
PA3847	2.05E-02	9.30E-03	PA2557	3.56E-02	1.14E-02
PA3881	2.05E-02	9.31E-03	PA2587	3.56E-02	1.14E-02
PA3935	2.05E-02	9.32E-03	PA2625	3.56E-02	1.14E-02
PA4017	2.05E-02	9.33E-03	PA2699	3.56E-02	1.14E-02
PA4032	2.05E-02	9.34E-03	PA2804	3.56E-02	1.14E-02
PA4038	2.05E-02	9.35E-03	PA2816	3.56E-02	1.14E-02
PA4068	2.05E-02	9.36E-03	PA2867	3.56E-02	1.14E-02

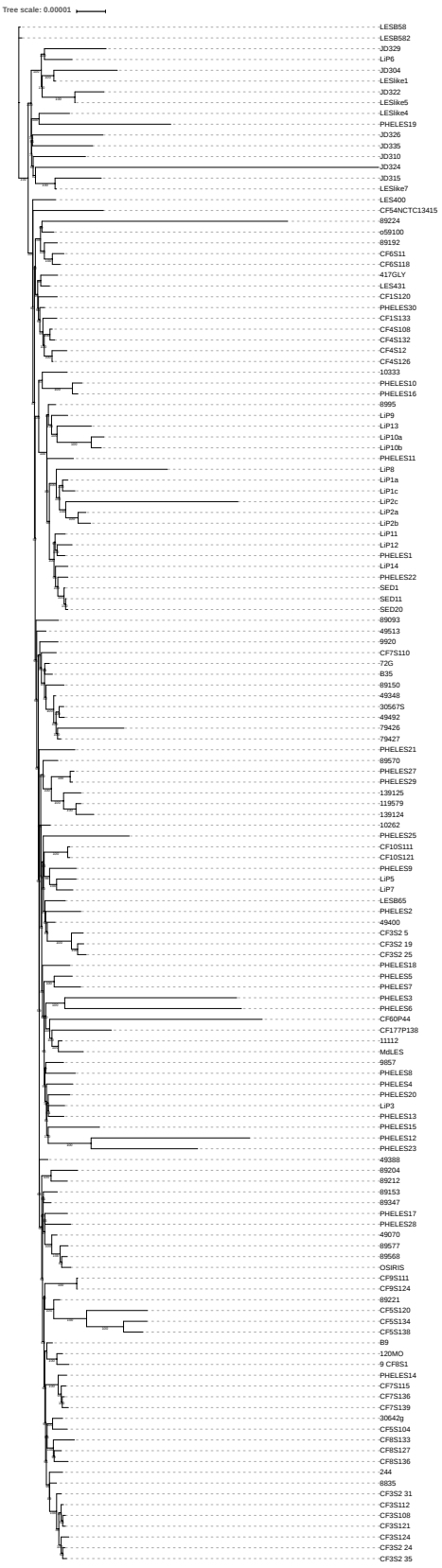
PA4280	2.05E-02	9.37E-03	PA2872	3.56E-02	1.14E-02
PA4289	2.05E-02	9.38E-03	PA2895	3.56E-02	1.15E-02
PA4298	2.05E-02	9.38E-03	PA2910	3.56E-02	1.15E-02
PA4318	2.05E-02	9.39E-03	PA3023	3.56E-02	1.15E-02
PA4343	2.05E-02	9.40E-03	PA3165	3.56E-02	1.15E-02
PA4346	2.05E-02	9.41E-03	PA3236	3.56E-02	1.15E-02
PA4400	2.05E-02	9.42E-03	PA3328	3.56E-02	1.15E-02
PA4401	2.05E-02	9.43E-03	PA3333	3.56E-02	1.15E-02
PA4507	2.05E-02	9.44E-03	PA3342	3.56E-02	1.15E-02
PA4521	2.05E-02	9.45E-03	PA3372	3.56E-02	1.15E-02
PA4531	2.05E-02	9.46E-03	PA3392	3.56E-02	1.15E-02
PA4533	2.05E-02	9.46E-03	PA3494	3.56E-02	1.15E-02
PA4620	2.05E-02	9.47E-03	PA3548	3.56E-02	1.16E-02
PA4628	2.05E-02	9.48E-03	PA3553	3.56E-02	1.16E-02
PA4633	2.05E-02	9.49E-03	PA3556	3.56E-02	1.16E-02
PA4662	2.05E-02	9.50E-03	PA3593	3.56E-02	1.16E-02
PA4717	2.05E-02	9.51E-03	PA3629	3.56E-02	1.16E-02
PA4732	2.05E-02	9.52E-03	PA3679	3.56E-02	1.16E-02
PA4734	2.05E-02	9.53E-03	PA3718	3.56E-02	1.16E-02
PA4738	2.05E-02	9.54E-03	PA3725	3.56E-02	1.16E-02
PA4744	2.05E-02	9.55E-03	PA3780	3.56E-02	1.16E-02
PA4749	2.05E-02	9.55E-03	PA3804	3.56E-02	1.16E-02
PA4766	2.05E-02	9.56E-03	PA3828	3.56E-02	1.16E-02
PA4819	2.05E-02	9.57E-03	PA3851	3.56E-02	1.17E-02
PA4827	2.05E-02	9.58E-03	PA3893	3.56E-02	1.17E-02
PA4831	2.05E-02	9.59E-03	PA3904	3.56E-02	1.17E-02
PA4860	2.05E-02	9.60E-03	PA3932	3.56E-02	1.17E-02
PA4921	2.05E-02	9.61E-03	PA3937	3.56E-02	1.17E-02
PA4961	2.05E-02	9.62E-03	PA3992	3.56E-02	1.17E-02
PA4993	2.05E-02	9.63E-03	PA4293	3.56E-02	1.17E-02
PA5065	2.05E-02	9.63E-03	PA4294	3.56E-02	1.17E-02
PA5096	2.05E-02	9.64E-03	PA4321	3.56E-02	1.17E-02
PA5104	2.05E-02	9.65E-03	PA4339	3.56E-02	1.17E-02
PA5117	2.05E-02	9.66E-03	PA4375	3.56E-02	1.17E-02
PA5122	2.05E-02	9.67E-03	PA4467	3.56E-02	1.17E-02
PA5123	2.05E-02	9.68E-03	PA4499	3.56E-02	1.18E-02
PA5195	2.05E-02	9.69E-03	PA4835	3.56E-02	1.18E-02
PA5208	2.05E-02	9.70E-03	PA4903	3.56E-02	1.18E-02
PA5221	2.05E-02	9.71E-03	PA4917	3.56E-02	1.18E-02
PA5245	2.05E-02	9.71E-03	PA4926	3.56E-02	1.18E-02
PA5281	2.05E-02	9.72E-03	PA4939	3.56E-02	1.18E-02
PA5293	2.05E-02	9.73E-03	PA4977	3.56E-02	1.18E-02

PA5317	2.05E-02	9.74E-03	PA5010	3.56E-02	1.18E-02
PA5368	2.05E-02	9.75E-03	PA5072	3.56E-02	1.18E-02
PA5382	2.05E-02	9.76E-03	PA5112	3.56E-02	1.18E-02
PA5405	2.05E-02	9.77E-03	PA5166	3.56E-02	1.18E-02
PA5432	2.05E-02	9.78E-03	PA5186	3.56E-02	1.19E-02
PA5433	2.05E-02	9.79E-03	PA5206	3.56E-02	1.19E-02
PA5443	2.05E-02	9.79E-03	PA5278	3.56E-02	1.19E-02
PA5460	2.05E-02	9.80E-03	PA5292	3.56E-02	1.19E-02
PA5465	2.05E-02	9.81E-03	PA5295	3.56E-02	1.19E-02
PA5480	2.05E-02	9.82E-03	PA5297	3.56E-02	1.19E-02
PA5489	2.05E-02	9.83E-03	PA5302	3.56E-02	1.19E-02
PA5566	2.05E-02	9.84E-03	PA5400	3.56E-02	1.19E-02
PA5567	2.05E-02	9.85E-03	PA5447	3.56E-02	1.19E-02
PA1046	2.25E-02	9.86E-03	PA5513	3.56E-02	1.19E-02
PA1243	2.25E-02	9.87E-03	PA5539	3.56E-02	1.19E-02
PA1394	2.25E-02	9.88E-03	PA5265	3.67E-02	1.20E-02
PA1965	2.25E-02	9.88E-03	PA0274	3.78E-02	1.20E-02
PA2148	2.25E-02	9.89E-03	PA2136	3.78E-02	1.20E-02
PA2153	2.25E-02	9.90E-03	PA2140	3.78E-02	1.20E-02
PA2161	2.25E-02	9.91E-03	PA3168	3.78E-02	1.20E-02
PA2438	2.25E-02	9.92E-03	PA3268	3.78E-02	1.20E-02
PA3911	2.25E-02	9.93E-03	PA4189	3.78E-02	1.20E-02
PA0009	2.47E-02	9.94E-03	PA5037	3.78E-02	1.20E-02
PA0266	2.47E-02	9.95E-03	PA1355	4.15E-02	1.20E-02
PA0330	2.47E-02	9.96E-03	PA2132	4.15E-02	1.20E-02
PA0409	2.47E-02	9.96E-03	PA2160	4.15E-02	1.20E-02
PA0555	2.47E-02	9.97E-03	PA5263	4.17E-02	1.21E-02
PA0826_1	2.47E-02	9.98E-03	PA3939	4.38E-02	1.21E-02
PA0962	2.47E-02	9.99E-03	PA4172	4.38E-02	1.21E-02
PA0973	2.47E-02	1.00E-02	PA4897	4.38E-02	1.21E-02
PA1102	2.47E-02	1.00E-02	PA3039	4.53E-02	1.21E-02
PA1533	2.47E-02	1.00E-02	PA0017	4.69E-02	1.21E-02
PA1564	2.47E-02	1.00E-02	PA0741	4.69E-02	1.21E-02
PA1581	2.47E-02	1.00E-02	PA1018	4.69E-02	1.21E-02
PA2023	2.47E-02	1.00E-02	PA1024	4.69E-02	1.21E-02
PA2603_1	2.47E-02	1.01E-02	PA1025	4.69E-02	1.21E-02
PA2640	2.47E-02	1.01E-02	PA1121	4.69E-02	1.21E-02
PA2743	2.47E-02	1.01E-02	PA1126	4.69E-02	1.21E-02
PA2883	2.47E-02	1.01E-02	PA1214	4.69E-02	1.22E-02
PA2905	2.47E-02	1.01E-02	PA1219	4.69E-02	1.22E-02
PA2953	2.47E-02	1.01E-02	PA1248	4.69E-02	1.22E-02
PA2960	2.47E-02	1.01E-02	PA1284	4.69E-02	1.22E-02

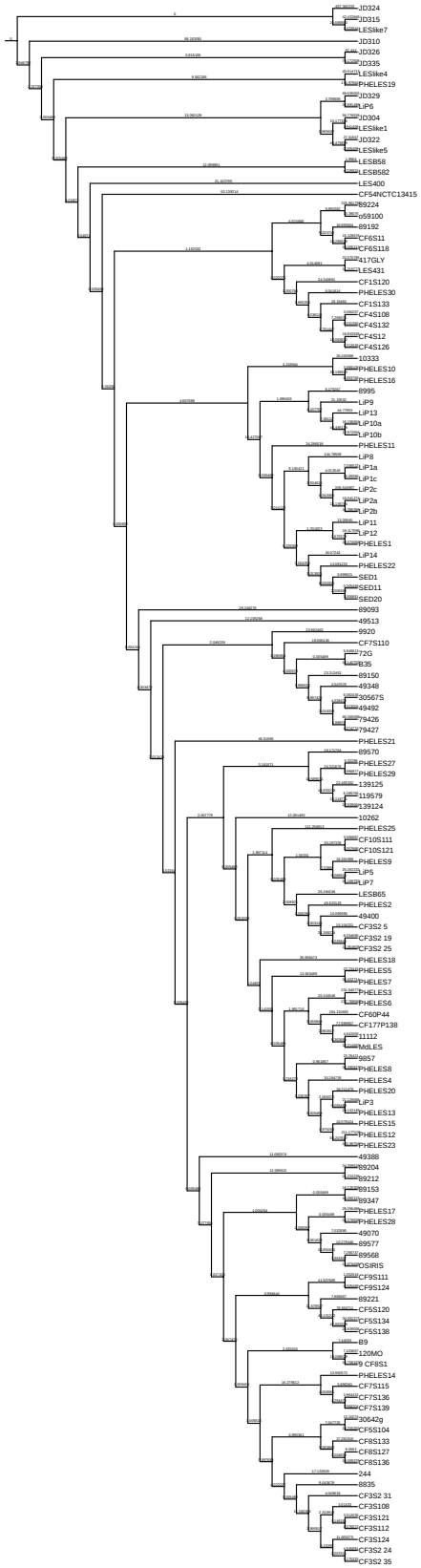
PA3001	2.47E-02	1.01E-02	PA1301	4.69E-02	1.22E-02
PA3088	2.47E-02	1.01E-02	PA1362	4.69E-02	1.22E-02
PA3572	2.47E-02	1.01E-02	PA1605	4.69E-02	1.22E-02
PA3636	2.47E-02	1.01E-02	PA1636	4.69E-02	1.22E-02
PA3745	2.47E-02	1.02E-02	PA1657	4.69E-02	1.22E-02
PA3815	2.47E-02	1.02E-02	PA1690	4.69E-02	1.22E-02
PA3823	2.47E-02	1.02E-02	PA1698	4.69E-02	1.22E-02
PA3838	2.47E-02	1.02E-02	PA1699	4.69E-02	1.23E-02
PA3979	2.47E-02	1.02E-02	PA1702	4.69E-02	1.23E-02
PA4246	2.47E-02	1.02E-02	PA1707	4.69E-02	1.23E-02
PA4266	2.47E-02	1.02E-02	PA1711	4.69E-02	1.23E-02
PA4449	2.47E-02	1.02E-02	PA1718	4.69E-02	1.23E-02
PA4524_1	2.47E-02	1.02E-02	PA1721	4.69E-02	1.23E-02
PA4672	2.47E-02	1.02E-02	PA1723	4.69E-02	1.23E-02
PA4694	2.47E-02	1.02E-02	PA1842	4.69E-02	1.23E-02
PA4704_1	2.47E-02	1.02E-02	PA1871	4.69E-02	1.23E-02
PA4752	2.47E-02	1.03E-02	PA2001	4.69E-02	1.23E-02
PA4764	2.47E-02	1.03E-02	PA2418	4.69E-02	1.23E-02
PA4868	2.47E-02	1.03E-02	PA2419	4.69E-02	1.24E-02
PA4937_1	2.47E-02	1.03E-02	PA2495	4.69E-02	1.24E-02
PA4938	2.47E-02	1.03E-02	PA2529	4.69E-02	1.24E-02
PA5054	2.47E-02	1.03E-02	PA2573	4.69E-02	1.24E-02
PA5119	2.47E-02	1.03E-02	PA2692	4.69E-02	1.24E-02
PA5315	2.47E-02	1.03E-02	PA2704	4.69E-02	1.24E-02
PA5331	2.47E-02	1.03E-02	PA2708	4.69E-02	1.24E-02
PA5360	2.47E-02	1.03E-02	PA2776	4.69E-02	1.24E-02
PA5396	2.47E-02	1.03E-02	PA2933	4.69E-02	1.24E-02
PA5553	2.47E-02	1.04E-02	PA3012	4.69E-02	1.24E-02
PA0027	2.51E-02	1.04E-02	PA3059	4.69E-02	1.24E-02
PA0495	2.51E-02	1.04E-02	PA3091	4.69E-02	1.25E-02
PA0533	2.51E-02	1.04E-02	PA3311	4.69E-02	1.25E-02
PA1019	2.51E-02	1.04E-02	PA3404	4.69E-02	1.25E-02
PA1032	2.51E-02	1.04E-02	PA3799	4.69E-02	1.25E-02
PA1216	2.51E-02	1.04E-02	PA4110	4.69E-02	1.25E-02
PA1458	2.51E-02	1.04E-02	PA4142	4.69E-02	1.25E-02
PA1459	2.51E-02	1.04E-02	PA4659	4.69E-02	1.25E-02
PA1692	2.51E-02	1.04E-02	PA4807	4.69E-02	1.25E-02
PA1696	2.51E-02	1.04E-02	PA4864	4.69E-02	1.25E-02
PA1700	2.51E-02	1.05E-02	PA4869	4.69E-02	1.25E-02
PA1704	2.51E-02	1.05E-02	PA4901	4.69E-02	1.25E-02
PA5346	4.69E-02	1.26E-02	PA4927	4.69E-02	1.26E-02
PA5431	4.69E-02	1.26E-02			



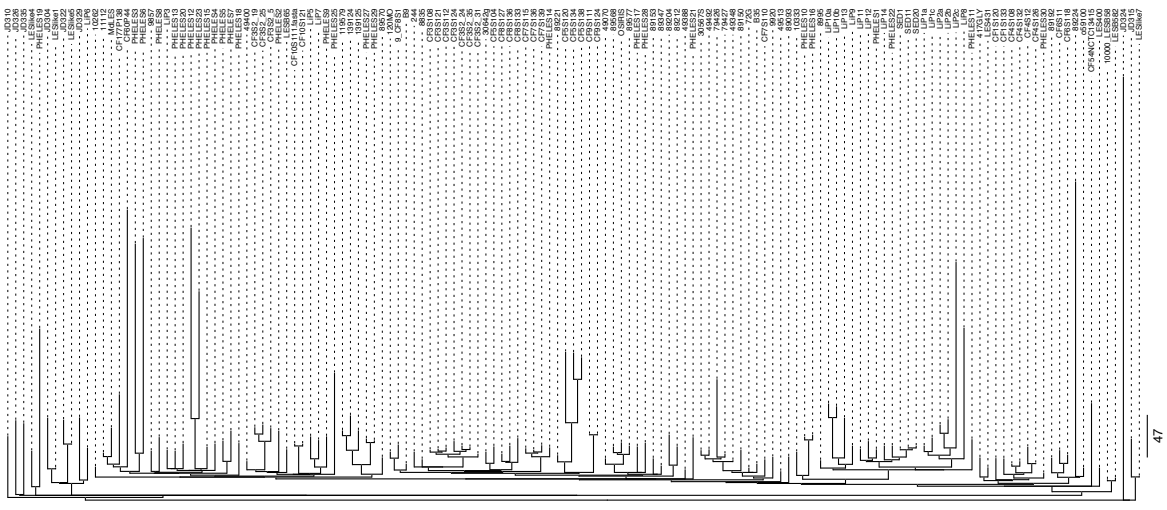
Supplementary figure 1. PseudoCAP functional categories associated with genes significantly underrepresented for loss of function in CF compared with the environment with p values smaller than the BH critical value



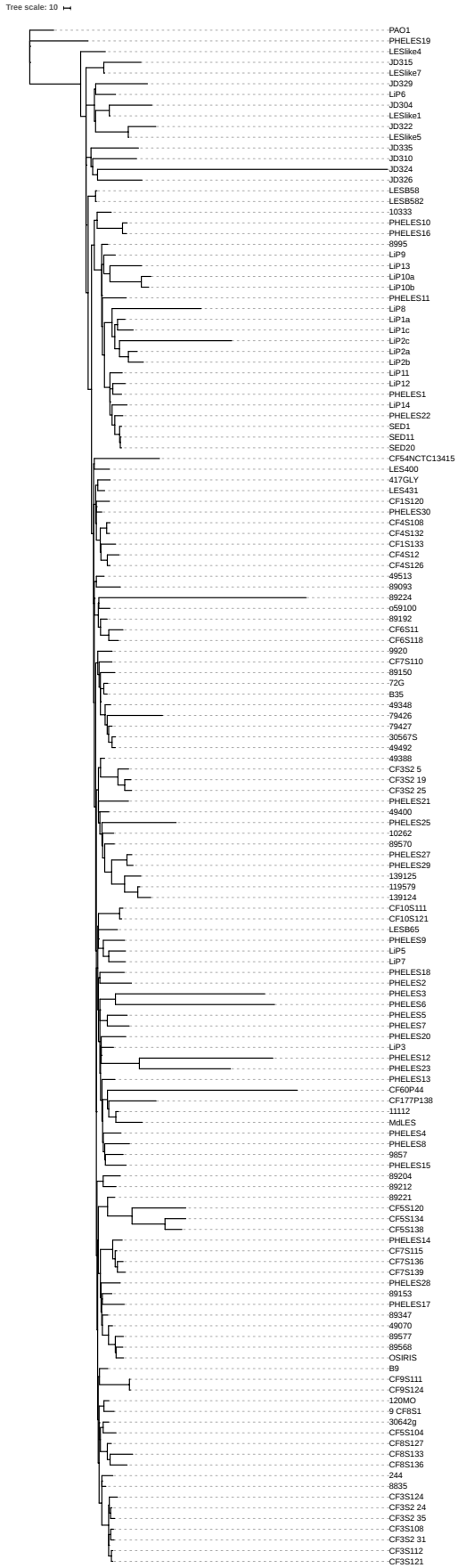
Supplementary figure 2. ML approximation of LES core genome alignment (n = 143) with IQtree, TVM+R3 substituion model and 1000 ultrafast bootstraps. Tree scale substitutions per base



Supplementary figure 3. LES core genome (n = 143 genomes) RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins



Supplementary figure 4. LES core genome recombination events alongside RaxML tree based on polymorphic sites corrected for recombination. Blocks represent recombination events based on position in the core genome. Red blocks are present in multiple genomes while blue are unique



Supplementary figure 5. LES (n = 143 genomes) and PAO1 core genome RAxML tree based on SNP sites filtered to remove recombinant SNPs by gubbins

Supplementary table 2. Non-CF bronchiectasis isolate genomes. Representative bronchiectasis panel in green. MLSTs provided, patient and centre information provided and whether DNA mismatch repair genes *mutS, mutL, mutM* or *uvrD* contain a likely loss of function mutation

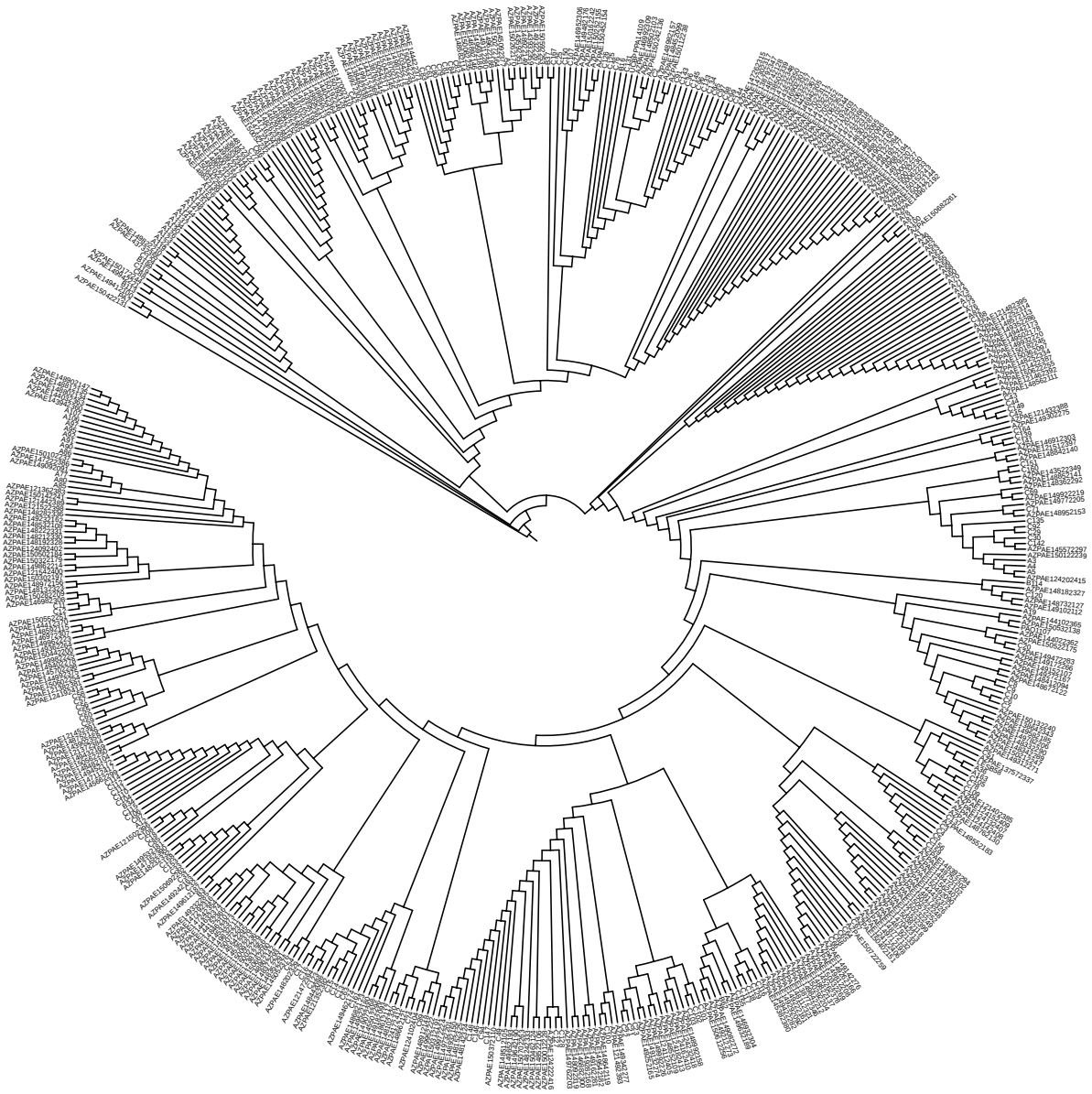
Isolate ID	Center	Date	Patient	ST	Mutator
A1	1	09/10/2014	1	17	
A2	1	16/10/2014	2	207	
A3	1	10/10/2014	3	252	
A4	1	10/10/2014	3	252	
A5	1	10/10/2014	3	252	
B3	12	01/11/2008	8	281	
B16	12	25/11/2013	9	253	
B34	12	03/09/2014	11	179	
B37	12	18/01/2012	12	-	
B62	12	03/09/2014	15	-	
B113	12	23/10/2014	18	1328	*mutL* (INDEL)
B114	12	11/04/2012	19	198	
B199	12	18/10/2011	32	1182	
A12	1	14/11/2014	35	179	
C2	4	14/10/2009	36	253	
C3	4	25/02/2010	37	260	
C4	4	25/02/2010	37	-[a]	
C5	4	25/02/2010	37	260 [a]	
C6	4	03/03/2010	38	244	
C7	4	23/03/2010	39	244	
C10	4	16/04/2010	40	244	
C8	4	16/04/2010	40	244	
C9	4	16/04/2010	40	244	
C11	4	19/08/2010	41	282	
C12	4	20/08/2010	42	282	
C13	4	20/08/2010	42	27 [b]	
C14	4	20/08/2010	42	27 [b]	
C15	4	20/08/2010	42	27 [b]	
C16	4	20/08/2010	42	-[b]	
C17	4	20/08/2010	42	- [b]	
C18	4	15/04/2011	43	-	*mutL* (STOP)
C20	4	01/07/2011	44	878	
C21	15	14/04/2009	45	-[c]	
C22	15	14/04/2009	45	- [c]	
C23	15	14/04/2009	45	- [c]	

C25	15	20/05/2009	46	253	
C29	15	03/06/2009	48	252	
C30	15	04/06/2009	49	252	
C31	15	25/08/2009	50	-[d]	*mutL* (INDEL)
C32	15	25/08/2009	50	-[d]	
C33	15	25/08/2009	50	-[d]	
C36	15	21/05/2010	52	253	
C42	15	12/07/2010	54	309	
C43	15	28/07/2010	55	108	
C44	15	28/07/2010	55	108	
C45	15	28/07/2010	55	108	
C49	15	22/02/2011	58	395	
C51	15	23/02/2011	59	683	
C54	15	15/06/2011	61	1342	
C55	5	26/06/2009	62	-[e]	
C56	5	26/06/2009	62	-[e]	
C57	5	26/06/2009	62	-[e]	
C58	5	26/06/2009	62	-[e]	
C59	5	26/06/2009	62	-[e]	
C60	5	26/06/2009	62	-[e]	
C61	5	17/11/2009	63	620	
C63	3	02/09/2009	64	27	
C64	3	25/11/2009	65	274	
C65	3	25/11/2009	65	274	
C66	3	25/11/2009	65	274	
C67	3	25/11/2009	65	274	
C68	3	25/11/2009	65	274	
C69	3	12/05/2010	66	-	
C71	9	04/09/2009	67	968	
C73	9	05/11/2010	68	17	
C74	9	03/12/2010	69	1202	*mutL* (STOP)
C76	2	12/05/2009	70	253	
C77	2	03/07/2009	71	308	
C78	2	14/07/2009	72	840	*mutL* (INDEL)
C79	2	14/07/2009	72	620[f]	*mutL* (STOP)
C80	2	14/07/2009	72	620[f]	
C81	2	14/07/2009	72	620[f]	
C82	2	14/07/2009	72	620[f]	
C83	2	14/07/2009	72	-[f]	
C84	2	14/07/2009	72	620[f]	

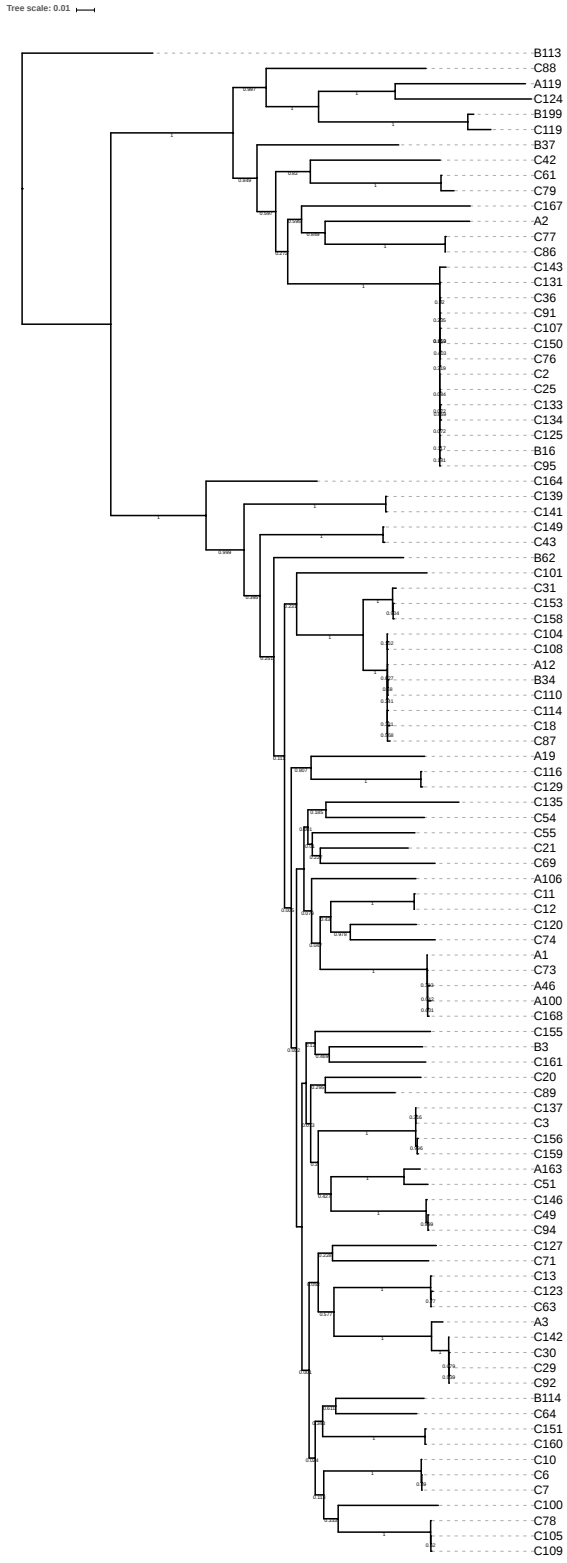
C85	2	06/08/2009	73	-[g]	
C86	2	06/08/2009	73	308	
C87	2	06/08/2009	73	179[g]	
C88	2	15/12/2009	74	1251	
C89	2	25/03/2010	75	1239	
C91	2	13/01/2011	76	253	
C92	2	01/02/2011	77	252	
C94	10	03/07/2009	78	395	
C95	10	29/07/2009	79	253	
C96	10	29/07/2009	79	253	
C97	10	29/07/2009	79	253	
C98	10	29/07/2009	79	253	
C99	10	29/07/2009	79	253	
C100	10	13/10/2009	80	612	
C101	10	21/10/2009	81	-[h]	
C102	10	21/10/2009	81	-[h]	
C103	10	21/10/2009	81	-[h]	
C104	13	16/05/2009	82	179	
C105	13	25/07/2009	83	840	
C106	13	11/08/2009	84	-[i]	
C107	13	11/08/2009	84	253	
C108	13	11/08/2009	84	179[i]	
C109	13	11/08/2009	85	840	
C110	13	11/08/2009	85	179	
C111	13	11/08/2009	85	179	
C112	13	11/08/2009	85	179	
C114	13	05/12/2009	86	179	
C115	13	05/12/2009	86	179	
C116	13	04/06/2010	87	871	
C117	13	04/06/2010	87	871	
C118	13	04/06/2010	87	871	
C119	7	23/01/2010	88	-	
C120	7	29/01/2010	89	-	
C123	7	02/04/2010	90	27	*mutS* (INDEL)
C124	7	08/04/2010	91	1753	
C125	7	29/04/2010	92	253	
C126	7	29/04/2010	92	253	
C127	7	29/04/2010	92	164	*mutS* (INDEL)
C128	7	29/04/2010	92	164	
C129	7	29/04/2010	92	871	*mutL*

					(INDEL)
C131	7	08/05/2010	93	253	
C133	7	28/05/2010	94	253	*mutS* (INDEL)
C134	7	19/12/2010	95	253	*mutS* (INDEL)
C135	14	02/11/2009	96	160	
C137	14	16/04/2010	97	260	
C139	14	08/09/2010	98	2102	
C141	14	01/10/2010	99	2102	
C142	14	11/02/2011	100	252	
C143	8	08/07/2010	101	253	
C144	8	08/07/2010	101	253	
C145	8	08/07/2010	101	253	
C146	8	25/06/2010	102	395	
C147	8	25/06/2010	102	395	
C148	8	25/06/2010	102	395	
C149	8	22/03/2011	103	108	
C150	6	27/08/2010	104	253	
C151	6	03/03/2011	105	1244	
C153	6	07/04/2011	106	155	
C155	11	04/12/2010	107	1211	
C156	16	16/11/2010	108	260	
C158	16	03/12/2010	109	155	
C159	16	09/12/2010	110	260	
C160	16	09/12/2010	111	1244	
C161	16	03/03/2011	112	110	
C164	16	08/12/2010	113	-	
C167	16	21/04/2011	114	296	
C168	16	21/12/2010	115	17	
A19	1	16/12/2014	120	-	
A163	1	19/05/2015	137	146	
A36	1	17/02/2015	137	146	
A46	1	07/04/2015	147	17	
A48	1	07/04/2015	147	17	
A52	1	07/04/2015	147	17	
A53	1	07/04/2015	147	17	
A54	1	07/04/2015	147	17	
A55	1	07/04/2015	147	17	
A56	1	07/04/2015	147	17	
A58	1	07/04/2015	147	17	
A60	1	07/04/2015	147	17	

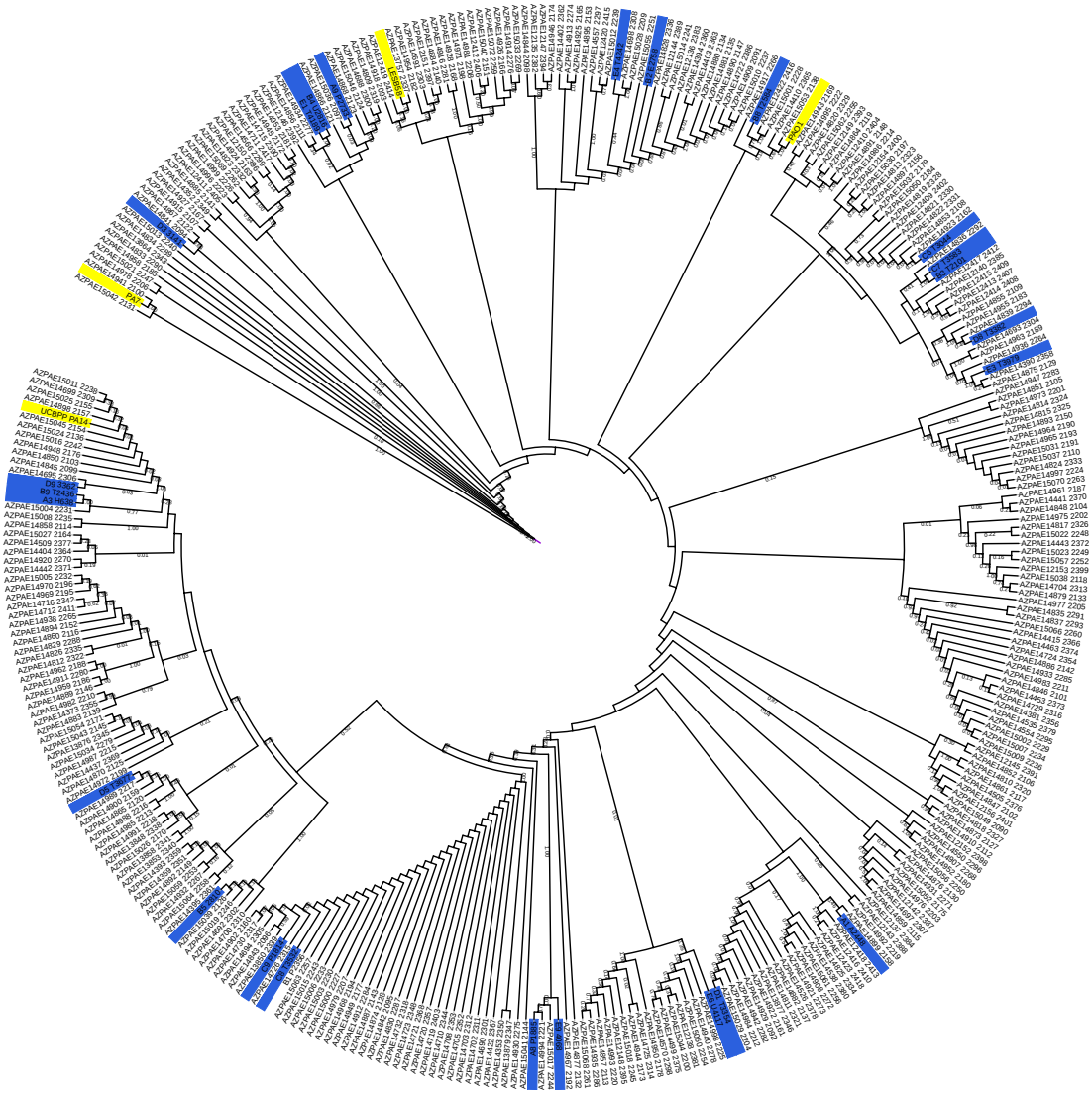
A70	1	07/04/2015	147	17	
A71	1	07/04/2015	147	17	
A72	1	07/04/2015	147	17	
A73	1	07/04/2015	147	17	
A75	1	07/04/2015	147	17	
A76	1	07/04/2015	147	17	
A100	1	07/04/2015	148	17	
A106	1	07/04/2015	148	175	
A107	1	07/04/2015	148	175	
A77	1	07/04/2015	148	175	
A78	1	07/04/2015	148	17	
A80	1	07/04/2015	148	175	
A81	1	07/04/2015	148	17	
A82	1	07/04/2015	148	17	
A85	1	07/04/2015	148	175	
A86	1	07/04/2015	148	175	
A90	1	07/04/2015	148	175	
A91	1	07/04/2015	148	175	
A92	1	07/04/2015	148	175	
A95	1	07/04/2015	148	175	
A97	1	07/04/2015	148	175	
A119	1	15/05/2015	149	667	
A122	1	15/05/2015	149	667	
A123	1	15/05/2015	149	667	
A126	1	15/05/2015	149	667	
A130	1	15/05/2015	149	667	
A134	1	15/05/2015	149	667	
A137	1	15/05/2015	149	667	
A141	1	15/05/2015	149	667	
A144	1	15/05/2015	149	667	
A147	1	15/05/2015	149	667	
A148	1	15/05/2015	149	667	
A151	1	15/05/2015	149	667	
A154	1	15/05/2015	149	667	
A156	1	15/05/2015	149	667	



Supplementary figure 6. Non-CF bronchiectasis and the wider *P. aeruginosa* population core genome SNP phylogeny (n = 522). HKY85 substitution model and ML approximation (MEGA) based on complete core genome polymorphic sites



Supplementary figure 7. Non-CF bronchiectasis representative panel genomes core SNP ML tree with HKY85 substitution model (MEGA) with 1000 bootstrap replicates (shown). Branch to B113 truncated for clarity



Supplementary figure 8. Genomes sequenced and the wider *P. aeruginosa* genome population core SNP phylogeny. HKY85 substitution model and ML approximation (MEGA) based on complete core genome polymorphic sites and 1000 bootstrap replicates. Thai genomes sequenced are marked in blue and reference *P. aeruginosa* genomes in yellow

Number of genes
0
50
100
150
200
Isolates
Number of genes
0
1
2
3
4
5
6
Isolates

Supplementary figure 9. Bar charts of number of genes with nonsense mutations per genome in the representative panel. Yellow bars are genomes belonging to phylogenetic group 3, blue bars phylogenetic group 2 and green phylogenetic group 1. (A) Shows a median line of count of genes (x axis) with nonsense mutations, above which all group 2 and 3 genomes reside. (B) Shows how for the count of genes inferred to be adaptive by loss of function (table 5.5) there is no detectable trend associated with phylogenetic placement and associated genetic distance to the reference