Whole genome sequencing reveals the emergence of a *Pseudomonas aeruginosa* shared strain sub-lineage among patients treated within a single cystic fibrosis centre

Bryan A. Wee[1†], Anna S. Tai[2-4†], Laura J. Sherrard[5], Nouri L. Ben Zakour[1], Kirt R. Hanks[1], Timothy J. Kidd[1,6,7], Kay A. Ramsay[2,5], Iain Lamont[8], David M. Whiley[9,10], Scott C. Bell[2,3,5], Scott A. Beatson[1*]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia

[2]Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia

[3]Adult Cystic Fibrosis Centre, Department of Thoracic Medicine, The Prince Charles Hospital, Brisbane, QLD, Australia

[4]Western Australia Adult Cystic Fibrosis Centre, Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Western Australia, Australia

[5]Lung Bacteria Group, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[6]Centre for Experimental Medicine, Queen's University Belfast, Belfast, UK

[7]Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia

[8]Department of Biochemistry, University of Otago, Dunedin, New Zealand

[9]Faculty of Medicine, UQ Centre for Clinical Research, The University of Queensland, Brisbane, QLD, Australia

[10]Microbiology Department, Pathology Queensland Central Laboratory, Brisbane, QLD, Australia

[†]These authors contributed equally to this work

*Correspondence: s.beatson@uq.edu.au

Email addresses

Bryan A. Wee, bryan.wee@uqconnent.edu.au

Anna S. Tai, Sze.Tai@health.wa.gov.au

Laura J. Sherrard, lsherrard03@qub.ac.uk

Nouri L. Ben Zakour, n.benzakour@uq.edu.au

Kirt R. Hanks, hanks.kirt@gmail.com

Timothy J. Kidd, t.m.kidd@uq.edu.au

Kay A. Ramsay, k.ramsay@uq.edu.au

Iain Lamont, iain.lamont@otago.ac.nz

David M. Whiley, d.whiley@uq.edu.au

Scott C. Bell, Scott.Bell@qimrberghofer.edu.au

Scott A. Beatson, s.beatson@uq.edu.au

Abstract

Background: Chronic lung infections by *Pseudomonas aeruginosa* are a significant cause of morbidity and mortality in people with cystic fibrosis (CF). Shared *P. aeruginosa* strains, that can be transmitted between patients, are of concern and in Australia the AUST-02 shared strain is predominant in individuals attending CF centres in Queensland and Western Australia. M3L7 is a multidrug resistant sub-type of AUST-02 that was recently identified in a Queensland CF centre and was shown to be associated with poorer clinical outcomes. The main aim of this study was to resolve the relationship of the emergent M3L7 sub-type within the AUST-02 group of strains using whole genome sequencing.

Results: A whole-genome core phylogeny of 63 isolates indicated that M3L7 is a monophyletic sub-lineage within the context of the broader AUST-02 group. Relatively short branch lengths connected all of the M3L7 isolates. A phylogeny based on nucleotide polymorphisms present across the genome showed that the chronological estimation of the most recent common ancestor was around 2001 (± 3 years). SNP differences between sequential M3L7 isolates collected 3-4 years apart from five patients suggested both continuous infection of the same strain and cross-infection of some M3L7 variants between patients. The majority of polymorphisms that were characteristic of M3L7 (i.e. acquired after divergence from all other AUST-02 isolates sequenced) were found to produce non-synonymous mutations in virulence and antibiotic resistance genes.

Conclusions: M3L7 has recently diverged from a common ancestor indicating descent from a single carrier at a CF treatment centre in Australia. Both adaptation to the lung and transmission of M3L7 between adults attending this centre may have

contributed to its rapid dissemination. The study emphasises the importance of clinical management in controlling the emergence of shared strains in CF.

Keywords: Whole-genome sequencing, cystic fibrosis, *Pseudomonas aeruginosa*, chronic lung infections, evolution, AUST-02

Background

Cystic fibrosis (CF) is the most common recessively lethal inherited disease in people of European ancestry. The majority of mortality and morbidity in people with CF is caused by chronic lung infection with *Pseudomonas aeruginosa* [1]. During chronic infection *P. aeruginosa* adapts to the CF airway microenvironment, which promotes multiple phenotypic and genotypic changes including enhanced resistance to antibiotics, excessive exopolysaccharide production, auxotrophy, auxotrophic metabolism, hypermutability and the loss of motility [2-8]. This evolution strategy has been found to occur in parallel in different *P. aeruginosa* strains, suggesting that these pathoadaptive modifications are important in the transition from an opportunistic pathogen to a specialised pathogen of diseased human lungs [8].

Person-to-person transmission of airway-adapted *P. aeruginosa* strains has also been reported with the acquisition of some “shared strains” correlated with adverse clinical outcomes [9-13]. AUST-02 is a prevalent shared strain in CF centres around Australia and is the dominant strain in Queensland infecting approximately 40% of patients infected with *P. aeruginosa* [10, 14, 15]. A recent surveillance study also described the detection of an AUST-02 strain sub-type in approximately 5% of patients attending a single CF treatment centre in Brisbane (Queensland) [16]. This sub-type could be distinguished from all other AUST-02 strains by a unique *mexZ*

(*mexZ-3*, M3) and *lasR* (*lasR*-7, L7) genotype and therefore, was designated M3L7 [16]. The M3L7 sub-type is of particular importance given that, compared to other AUST-02 and *P. aeruginosa* strains, it is associated with enhanced antibiotic resistance and poorer clinical outcomes (e.g. higher hospitalisation requirements and increased risk of lung transplantation or death) [16].

The aim of this study was to use whole genome sequencing (WGS) to reconstruct the population structure and resolve the relationship of the M3L7 sub-type within the AUST-02 group of strains. The analyses revealed that the M3L7 sub-type has recently diverged from a common ancestor suggesting descent from a single founder within a rapidly growing CF centre population. Genetic mutations exclusive to the M3L7 sub-type were identified and may have aided adaptation to the CF airway microenvironment prior to its dissemination.

Methods

Bacterial isolates and whole genome sequencing

P. aeruginosa isolates (n=25) encompassing the majority of M3L7 isolates (named AUS934 to AUS958) that were recently described, and one isolate considered an outgroup (AUS970) were chosen for sequencing (Additional File 1: Table S1) [16]. These isolates originated from expectorated sputum (collected in 2007, 2008 and 2011) provided by adults with CF (n=20), who attended The Prince Charles Hospital Adult CF centre in Brisbane for their clinical care [16]. Preparation of genomic DNA for WGS was undertaken using the UltraClean® Microbial DNA Isolation Kit as described previously [17]. Library preparation (Truseq), qPCR (TapeStation, Agilent

Genomics) and WGS using the Illumina HiSeq 2500 platform with 100 bp paired-end read chemistry were carried out by the Australian Genome Research Facility, Melbourne, Australia.

In order to reconstruct the M3L7 population structure, a further 37 AUST-02 genomes that were previously sequenced (as part of an ongoing AUST-02 population genetic diversity study) were included in the analyses (Additional File 1: Table S1).

Genome mapping and assembly

Reads were taxonomically assigned with Kraken (v0.10.4) to check for contamination and trimmed using Nesoni clip (v0.128) to filter out adapter sequences and low-quality regions [18, 19]. Reads were mapped to the *P. aeruginosa* PAO1 reference genome (NC_002516) using Shrimp as implemented in Nesoni (v0.128) [18, 20, 21]. The PAO1 genome was chosen as a reference due to its high quality, expert-curated annotation [21]. SNPs and small insertions or deletions (indels) shorter than the read length were called using Nesoni.

Genomes were assembled using Velvet (v1.2.10) and VelvetOptimiser (v2.2.5) [22, 23]. Assembled contigs were reordered against PAO1 using Mauve (v2.4.0) and annotated with Prokka (v1.10) [24, 25]. Gene annotations from PAO1 were used as the primary reference.

Phylogenetic analysis

An alignment of 30,811 core SNPs obtained from mapping against PAO1 was used to reconstruct the phylogeny of the 63 AUST-02 genome sequences. RAxML (v8.1.15) was used to estimate the Maximum Likelihood tree with the rapid bootstrap analysis

option (`-f a`) and GTRGAMMA model of nucleotide substitution with a correction for ascertainment bias (`-m ASC_GTRGAMMA --asc-corr lewis`) [26]. A resolved phylogeny of the M3L7 genome sequences (generated in this study) was constructed from a SNP matrix of 364 SNPs with the same settings as above. Phylogenetic trees were viewed and explored using Dendroscope (v3.4.1) and FigTree (v1.4.2) [27, 28]. Minimum spanning trees were generated using the goeBURST Full MST function in Phyloviz [29].

BEAST analysis

To determine the emergence of the M3L7 sub-type, Bayesian inference of the evolutionary rates was conducted using BEAST 1.8.2 [30]. As input a set of 183 SNPs specific for M3L7 was used, excluding two hypermutator isolates (AUS937 and AUS938) and the AUS970 outgroup. Regions of clustered SNPs, where at least three SNPs were found within 10bp of each other, were also removed. Among the different combinations of the molecular clock model (strict and constant relaxed lognormal), substitution model (HKY, GTR) and population size change (coalescent constant and exponential growth) models, the preferred combination of parameters selected based on stepping stone sampling was strict molecular clock, HKY substitution model with four discrete gamma-distributed rate categories, and exponential population size change. Markov Chain Monte Carlo generations were run in triplicate for 50 million steps, sampling every 5,000 steps, to ensure convergence and an ESS value >200 for all parameters. Replicate runs were combined using LogCombiner with a 10% burn-in and maximum credibility trees reporting mean values were created using TreeAnnotator.

Comparative genomic analyses

Comparative genomic analyses were performed using Parsnp, Gingr, BRIG (BLAST Ring Image Generator), Roary (v3.4.2), ACT (Artemis Comparison Tool) and BLAST [31-34]. Large genomic differences were investigated using PHAST (PHAge Search Tool) and Roary [33, 35]. The effect of amino acid substitutions (functionally important or no change) were predicted *in silico* using PROVEAN (Protein Variation Effect Analyzer) [36].

Results and Discussion

Isolates selected for WGS in this study

One to three *P. aeruginosa* isolates were prospectively collected from single sputum specimens obtained annually for culture as part of the Australian Clonal *P.aeruginosa* in Cystic Fibrosis (ACPinCF) study [10] and a systematic surveillance study was subsequently conducted to assess the prevalence of M3L7 between 2007 and 2011 [16]. M3L7 was identified in 28/509 (5.5%) *P. aeruginosa* isolates from 13/170 (7.6%) patients in 2007-2009 and in 21/519 (4.0%) *P. aeruginosa* isolates from 11/173 (6.4%) patients in 2011 [16]. Twenty-five of these isolates were selected from all patients (patients 24 to 42; Additional File 1: Table S1) infected with M3L7 for WGS in this study.

Of the 13 patients identified with M3L7 in 2007-2009, five underwent lung transplantation and one moved interstate by 2011. Therefore, no follow-up isolates were available. Of the seven patients who had M3L7 infection in 2007 and had samples collected in 2011, five remained infected in 2011, one no longer had the

M3L7 subtype detected, while another tested positive by the M3 allele-specific PCR, but the *lasR* sequence could not be determined because of suboptimal sequence quality (this isolate, AUS947, was subsequently confirmed as M3L7 by WGS). A further five patients acquired M3L7 in 2011 and were identified as incident cases (these patients were infected with other strains in 2007). Finally, one patient infected with M3L7 in 2011 had no previous strain-typing data available.

One M3L7 isolate was randomly selected per patient at each time-point (2007-2009 and 2011) for WGS and comprised: i) 13 M3L7 isolates from 13 patients in 2007-2009; ii) six M3L7 isolates from six patients with persistent M3L7 infection in 2011; iii) five M3L7 isolates from five incident cases in 2011; and iv) one M3L7 isolate from a patient with no previous *P. aeruginosa* strain typing data collected in 2011. One further *P. aeruginosa* isolate from a patient (patient 43; Additional File 1: Table S1) in 2007 which contained the M3 *mexZ* allele and a non-L7 *lasR* genotype was included as an outgroup.

M3L7 is a distinct sub-lineage of the AUST-02 shared strain

On the basis of the whole-genome core phylogeny, the 63 AUST-02 isolates form two major discrete lineages (clades), M2 (n=34) and M3 (n=29), consistent with their possession of *mexZ-2* (codon substitution, A38T) or *mexZ-3* (codon substitution, T12N) alleles, respectively (Figure 1 and Additional File 2: Figure S1). The M3L7 isolates (n=26; including 25 isolates sequenced here and a previously sequenced AUST-02 genome, AUS22) form a monophyletic sub-lineage of AUST-02 within the M3 clade that has diverged from all other AUST-02 isolates sequenced to date (Figure 1). Three isolates (AUS853; AUS854; AUS970) within the M3 clade form deep-

branching relationships at the base of the lineage and do not harbour the *lasR-7* allele (L7: 1 bp deletion, 438delG) that defines the M3L7 sub-type. Of the available AUST-02 sequences, isolate AUS970 (sequenced in this study) represented the AUST-02 genome that was most closely related to the M3L7 sub-lineage. In addition to the *mexZ-3* allele (M3), AUS970 carried a wild-type *lasR-1* (L1) allele (therefore named M3L1).

M3L7 diverged recently from other AUST-02 shared strains

The M3L7 sub-lineage expanded recently following a long period of divergence from the M2 sub-lineage according to the relatively short branch lengths connecting the M3L7 isolates and their relative distance from the root (Figure 1). Using BEAST analysis (Figure 2), we estimated that the most recent common ancestor (MCRA) of M3L7 emerged around 2001 (± 3 years). This is approximately six years prior to the first isolation of an M3L7 sub-type (2007) in people with CF in Brisbane (Additional File 3: Figure S2) [16]. Of note, this time period also corresponds with a relatively high annual increase (approximately 10-15%) in the adult CF population at The Prince Charles Hospital (Additional File 3: Figure S2). This situation, combined with limited capacity to segregate all patients, particularly when admitted to an inpatient ward, may have contributed to shared-strain infections. Taken together, these results support a single founder scenario in which a carrier of the M3L7 sub-type acted as a donor within a CF centre with subsequent rapid dissemination in the resident CF population. Notably, both sub-lineages of AUST-02 are prevalent in Queensland, suggesting that their MCRA originated in this state; however, we cannot rule out

independent introductions of M3L7 and M2 founders into the local CF population from outside the same geographical region (state).

Resolving the within-patient relationships of M3L7 provides evidence for both continuous infection and person-to-person transmission

To investigate genomic evidence for continuous infection within patients and person-to-person transmission, we examined the pattern of SNP differences between isolates in more detail. As described previously, the dataset included six pairs of M3L7 isolates (Additional File 2: Figure S1) collected from people with CF at two time-points in 2007 or 2008 and in 2011 [16].

A pair of M3L7 isolates (AUS937 and AUS938) from patient 38 had accumulated a much higher number of SNPs compared to other M3L7 isolates, as indicated by their relatively longer branches in the phylogeny (Additional File 2: Figure S1). Further analysis revealed that these two isolates had acquired independent non-synonymous mutations within the *mutS* gene (AUS937, L341P; AUS938, 1 bp deletion [1076delC]), encoding a DNA mismatch repair protein [37], and were predicted to be deleterious to protein function based on an *in silico* analysis [36]. Mutations in *mutS* are associated with hypermutation, which frequently occurs during chronic *P. aeruginosa* infection of the CF airway [38, 39].

After removal of the AUS937 and AUS938 isolates from the analysis, the phylogeny of the M3L7 sub-lineage could be resolved further: a maximum of 47 core SNPs separated the most divergent (AUS941 and AUS947) isolates (Additional File 4: Figure S3 and Additional File 5: Figure S4). The closest sequential within-patient isolates (AUS956 and AUS957 from patient 41) differed by only two core SNPs demonstrating a remarkably high degree of genome stability over a 4-year period

(Additional File 4: Figure S3). Within-patient isolates from patients 41 and 28 also grouped together with high bootstrap support on the phylogeny, which is consistent with continuous M3L7 infection across the two sampling time-points. Our analyses also revealed that a very close relationship existed between the early and late isolates from patients 30 and 31, which could also be due to continuous infection of the same strain (Additional File 4: Figure S3).

In contrast, based on the Maximum-Likelihood phylogeny, the second isolate (2011) cultured from patient 39 was grouped together with isolates from other patients (patient 36; patient 37; patient 40) instead of only the earlier 2007 isolate (Additional File 4: Figure S3). This indicates the possibility of multiple M3L7 cross-infection events as has been suggested for other shared strains [40], which may occur via the airborne route or during socialisation between patients [41, 42]. Pairwise SNP comparisons between M3L7 isolates also shows a possible transmission pathway with isolates AUS946 and AUS944 from patient 39 and patient 30, respectively, being the most likely source of cross-infections (Additional File 5: Figure S4).

AUS22 (an M3L7 isolate sequenced as part of a different study) and AUS943 were both isolated from patient 32 just six days apart. However, these two isolates were more closely related to isolates from other patients than to each other, which is also suggestive of direct or indirect cross-infection between those patients (Additional File 2: Figure S1). This finding highlights that short-term within-patient diversity of shared strains during chronic infection (e.g. [17]) needs to be considered in light of the *P. aeruginosa* lung diversity of the local CF population as a whole.

The resolution of WGS data of multiple isolates from individual patients, combined with social interaction data will enable future studies to distinguish between

continuous infection or recent acquisition of M3L7 variants amongst the CF population [14].

M3L7 is characterised by an accumulation of non-synonymous mutations in critical pathways

Clonal lineages are expected to accumulate mutations that enable adaptation of the bacterium to a specific environmental niche of a human host [8]. A total of 44 shared SNPs and nine shared indels were acquired after divergence of the M3L7 sub-lineage from all other AUST-02 isolates of the M2 and M3 clades. Thirty-five SNPs were non-synonymous (80%), resulting in a change of the amino acid sequence including two premature stop codons (Additional File 6. Table S2). Four indels produced in-frame mutations, whilst five indels caused a shift in the reading frame (Additional File 6: Table S2).

Genes containing non-synonymous SNPs and indels exclusive to the M3L7 sub-lineage (n=43) were subsequently categorised according to PseudoCAP (*P. aeruginosa* community annotation project) functions (Figure 3) [43]. Fifteen genes were annotated with at least two functional categories and nine genes were part of regulatory networks (Figure 3), including key global regulators (e.g. *rpoN*, *mexT*), which might impact multiple processes [44]. Approximately 50% of the non-synonymous mutations occurred in genes that encode proteins associated with virulence (e.g. PilR involved in surface attachment and twitching motility [45, 46]; MigA involved in swarming [47]; ZnuA in zinc homeostasis [48]; PchD for iron acquisition [49]) or antibiotic resistance mechanisms (e.g. OprD: carbapenem resistance [50]; GyrB: fluoroquinolone resistance [51]; FtsI, Mpl: β-lactam resistance

[52, 53]; PmrB, ColS: polymyxin resistance [54, 55]; MexA, MexS, MexT: multi-drug efflux pumps [56-59]). The mutations found in genes correlated with antibiotic resistance were also identified in M3L7 isolates collected in 2014 [17] and might help explain the increased resistance described previously in the sub-type compared to other strains [16]. Altogether this analysis suggests that the M3L7 sub-lineage is characterised by mutations in genes that might have aided adaptation to the CF airways prior to dissemination in the CF population [4, 8, 39, 60].

Genomic variation due to large deletions

The draft genomes that comprise the M3L7 sub-lineage ranged in size from 6.15 to 6.25 Mbp, whereas the M3L1 isolate (AUS970) genome was substantially smaller (6.05 Mbp). All genomes within the M3 clade had an average GC content of 66.5%. Genome reduction is a typical evolutionary process that occurs during adaptation within the CF lungs [61]; therefore, large genomic variations (>10 kb) were compared between the AUST-02 isolates.

Major differences in genome size were not due to horizontal gene transfer but attributed to several deletion events, summarised in Table 1. For example, one M3L7 isolate (AUS947) lost a large 93 Kbp region encoding a number of virulence genes including *exoY* (a T3SS secreted toxin), a three-part hydrogen cyanide biosynthesis operon, *hcnABC* and part of the *cupA1-A5* chaperone usher fimbrae operon (Table 1). Deletions of varying sizes encompassing *exoY* and *hcnABC* (of the same region) were also missing in three other isolates within the M3 clade (AUS853; AUS854; AUS970) suggesting that there was a selective pressure to lose the functionality of genes encoded at this position.

A 40 Kbp putative prophage was present in five AUST-02 isolates from Perth (AUS15; AUS874; AUS17; AUS876; AUS877) and one isolate from Brisbane (AUS24) of the M2 clade. This prophage is present in nearly all M3L7 genomes (absent in AUS958) and a previous study also observed the presence of the prophage in 9/11 M3L7 isolates collected in 2014 [17]. Therefore, it is most likely that this prophage was present in the last common ancestor of the M3L7 sub-lineage and was vertically inherited. The prophage is inserted immediately adjacent to the *mutS* gene and a search of the putative prophage sequence using the PhaST webtool predicts a complete prophage that is only 9% identical to the *P. aeruginosa* F10 phage [35]. The prophage is flanked by an 18 bp *att* sequence (TCTCTCAGCACACGCC) that delineates the deletion in AUS958.

Table 1. Comparison of deletion events larger than 10 Kb in the AUST-02 genomes.

Size (Kbp)	PAO1 locus tags	Features		Genome	
		Gene	Function	Absent	Present
50-93	PA2165-PA2217	• *exoY* • *hcnABC* • *cupA1-A5*	T3SS secreted toxin Hydrogen cyanide biosynthesis operon Chaperone usher fimbrae operon	• AUS947 (M3L7) • AUS970 (M3L1) • AUS853 (M2 clade) • AUS854 (M2 clade)	• All other AUST02 isolates
40	PA3619-PA3620	• Adjacent to *mutS*	Putative prophage	• AUS958 (M3L7) • AUS970 (M3L1) • AUS853 (M2 clade) • AUS854 (M2 clade)	• All other M3L7 isolates • Some Perth isolates* • AUS24 (M2 clade)
14	PA1914- PA1923	• PA1914 • *nrdD* • PA1922	Encodes halovibrin Pathogenesis related factor TonB-dependent receptor	• All M3L7 isolates	• All other AUST-02 isolates
12	PA2229- PA2237	• *pslABCDEFG*	Biofilm formation	• All M3L7 isolates • AUS853 (M2 clade)	• All other AUST-02 isolates

*AUS15; AUS874; AUS17; AUS876; AUS877 (M2 clade)

In conclusion, the persistence of the AUST-02 shared strain in the CF population in Australia has led to the emergence of a monophyletic sub-lineage (M3L7) that is distinct from the M2 sub-lineage of AUST-02. Our WGS analysis demonstrated that M3L7 strains are characterised by mutations in genes that are likely to affect antibiotic resistance and virulence phenotypes. The rapid dissemination of this clinically important sub-type is most likely due to a combination of adaptation to the CF airway microenvironment and transmission between people attending the same CF centre. This work highlights the importance of clinical management in the emergence of shared strains amongst people with CF and provides a framework for future efforts in real-time genomic surveillance to monitor the transmission and pathogenicity of AUST-02 amongst the Australian CF population and to detect newly emergent shared strains.

Declarations

Ethics approval and consent to participate

Ethics approval for this project was granted under HREC/07/QRCH/9 and HREC/13/QPCH/127 by The Prince Charles Hospital Human and Research Ethics Committee, Metro North Hospital and Health Service, Brisbane, Queensland, Australia and all participants provided written, informed consent.

Consent for publication

Not applicable (no individual level patient data contained within this manuscript).

Availability of data and material

Genome sequence data generated in this study was deposited in the European Nucleotide Archive under study PRJEB14781 with accession identifiers ERS1249679 to ERS1249704. Other AUST-02 isolates are available as part of a separate study (PRJEB21755).

Competing interests

None to declare.

Funding

This work was supported by project grant funding from the National Health and Medical Research Council (NHMRC: 455919), Australian Infectious Diseases Research Centre (QIMRB-UQ seed grant) and TPCH Foundation grant (MS2013-02). AST is the recipient of NHMRC medical and dental Postgraduate Scholarship, Australian Cystic Fibrosis Research Trust Postgraduate Scholarship and Airways Infections, Inflammation & Cystic Fibrosis Group Scholarship. LJS is the recipient of The Shelley Shephard Memorial Scholarship. KAR is the recipient of an Australian Postgraduate Award, PhD Scholarship. TJK is the recipient of an ERS–EU RESPIRE2 Marie Skłodowska-Curie Postdoctoral Research Fellowship (MC RESPIRE2 first round, grant number 4571-2013) and a NHMRC Early Career Fellowship (GNT1088448). ILL receives grant support from the New Zealand Health Research Council, Curekids, Cystic Fibrosis New Zealand the New Zealand Lotteries Board (Health). SCB is the recipient of a Queensland Health, Health Research Fellowship and receives grant support from the NHMRC, CF Foundation

Therapeutics (USA), TPCH Foundation and Children's Health Foundation, Queensland. SAB is the recipient of a NHMRC Fellowship (APP1090456).

Author's contributions

AST, TJK, DMW, SCB and SAB designed the study. AST, TJK, KAR, DMW and SCB collected, performed genotyping and selected samples for sequencing. BAW and LJS carried out the comparative genome analyses and wrote the original manuscript draft. NLBZ performed BEAST phylodynamic analyses. KRH, NLBZ, IL and SAB contributed to the interpretation of the bioinformatics analyses. KAR and SCB collected information on numbers of CF patients attending TPCH. DMW, SCB and SAB supervised the project. All authors reviewed, edited and approved the final manuscript.

Acknowledgements

We thank Ms Rebecca Stockwell (Lung Bacteria Group, QIMR Berghofer Medical Research Institute, Brisbane, Australia) for extracting the genomic DNA for WGS.

References

1. Cystic Fibrosis Trust. UK Cystic Fibrosis Registry 2014 Annual Report. 2015. Accessed 12th Jan 2016.

2. Cullen L, McClean S. Bacterial adaptation during chronic respiratory infections. Pathogens. 2015;4:66-89.

3. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, et al. Regional isolation drives bacterial diversification within cystic fibrosis lungs. Cell Host Microbe. 2015;18:307-19.

4. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, et al. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A. 2006;103:8487-92.

5. Chung JC, Becq J, Fraser L, Schulz-Trieglaff O, Bond NJ, Foweraker J, et al. Genomic variation among contemporary *Pseudomonas aeruginosa* isolates from chronically infected cystic fibrosis patients. J Bacteriol. 2012;194:4857-66.

6. Thomas SR, Ray A, Hodson ME, Pitt TL. Increased sputum amino acid concentrations and auxotrophy of *Pseudomonas aeruginosa* in severe cystic fibrosis lung disease. Thorax. 2000;55:795-7.

7. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, et al. *Pseudomonas aeruginosa* population diversity and turnover in cystic fibrosis chronic infections. Am J Respir Crit Care Med. 2011;183:1674-9.

8. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. Nat Genet. 2015;47:57-64.

9. Al-Aloul M, Crawley J, Winstanley C, Hart CA, Ledson MJ, Walshaw MJ. Increased morbidity associated with chronic infection by an epidemic *Pseudomonas aeruginosa* strain in CF patients. Thorax. 2004;59:334-6.

10. Kidd TJ, Ramsay KA, Hu H, Marks GB, Wainwright CE, Bye PT, et al. Shared *Pseudomonas aeruginosa* genotypes are common in Australian cystic fibrosis centres. Eur Respir J. 2013;41:1091-1100.

11. Jones AM, Dodd ME, Doherty CJ, Govan JR, Webb AK. Increased treatment requirements of patients with cystic fibrosis who harbour a highly transmissible strain of *Pseudomonas aeruginosa*. Thorax. 2002;57:924-5.

12. Aaron SD, Vandemheen KL, Ramotar K, Giesbrecht-Lewis T, Tullis E, Freitag A, et al. Infection with transmissible strains of *Pseudomonas aeruginosa* and clinical outcomes in adults with cystic fibrosis. JAMA. 2010;304:2145-53.

13. Fothergill JL, Walshaw MJ, Winstanley C. Transmissible strains of *Pseudomonas aeruginosa* in cystic fibrosis lung infections. Eur Respir J. 2012;40:227-38.

14. Kidd TJ, Soares Magalhaes RJ, Paynter S, Bell SC, Group ACI. The social network of cystic fibrosis centre care and shared *Pseudomonas aeruginosa* strain infection: a cross-sectional analysis. Lancet Respir Med. 2015;3:640-50.

15. O'Carroll MR, Syrmis MW, Wainwright CE, Greer RM, Mitchell P, Coulter C, et al. Clonal strains of P*seudomonas aeruginosa* in paediatric and adult cystic fibrosis units. Eur Respir J. 2004;24:101-6.

16. Tai AS, Bell SC, Kidd TJ, Trembizki E, Buckley C, Ramsay KA, et al. Genotypic diversity within a single *Pseudomonas aeruginosa* strain commonly shared by Australian patients with cystic fibrosis. PLoS One. 2015;10:e0144022.

17. Sherrard LJ, Tai AS, Wee BA, Ramsay KA, Kidd TJ, Ben Zakour NL, et al. Within-host whole genome analysis of an antibiotic resistant *Pseudomonas aeruginosa* strain sub-type in cystic fibrosis. PLoS One. 2017;12:e0172179.

18. Harrison P. Nesoni. In., 0.128 edn. 2015. https://github.com/Victorian-Bioinformatics-Consortium/nesoni

19. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46.

20. David M, Dzamba M, Lister D, Ilie L, Brudno M. SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics. 2011;27:1011-2.

21. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature. 2000;406:959-64.

22. Gladman S, Seemann T. VelvetOptimser. In., 2.2.5 edn. 2012. https://github.com/tseemann/VelvetOptimiser

23. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821-9.

24. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5:e11147.

25. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-9.

26. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312-3.

27. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics. 2007;8:460.

28. Rambout A. FigTree. In., 1.4.2 edn. 2014. http://tree.bio.ed.ac.uk/software/figtree/

29. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrico JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics. 2012;13:87.

30. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969-73.

31. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15:524.

32. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402.

33. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691-3.

34. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. Bioinformatics. 2005;21:3422-3.

35. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:W347-52.

36. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012;7:e46688.

37. Oliver A, Mena A. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. Clin Microbiol Infect. 2010;16:798-808.

38. Oliver A, Canton R, Campo P, Baquero F, Blazquez J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science. 2000;288:1251-4.

39. Feliziani S, Marvig RL, Lujan AM, Moyano AJ, Di Rienzo JA, Krogh Johansen H, et al. Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. PLoS Genet. 2014;10:e1004651.

40. Williams D, Evans B, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, et al. Divergent, coexisting *Pseudomonas aeruginosa* lineages in chronic cystic fibrosis lung infections. Am J Respir Crit Care Med. 2015;191:775-85.

41. Knibbs LD, Johnson GR, Kidd TJ, Cheney J, Grimwood K, Kattenbelt JA, et al. Viability of *Pseudomonas aeruginosa* in cough aerosols generated by persons with cystic fibrosis. Thorax. 2014; 69:740-5.

42. Ojeniyi B, Frederiksen B, Hoiby N. *Pseudomonas aeruginosa* cross-infection among patients with cystic fibrosis during a winter camp. Pediatr Pulmonol. 2000;29:177-81.

43. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. Nucleic Acids Res. 2016;44:D646-53.

44. Winstanley C, O'Brien S, Brockhurst MA. *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. Trends Microbiol. 2016;24:327-37.

45. Hobbs M, Collie ES, Free PD, Livingston SP, Mattick JS. PilS and PilR, a two-component transcriptional regulatory system controlling expression of type 4 fimbriae in *Pseudomonas aeruginosa*. Mol Microbiol. 1993;7:669-82.

46. Kohler T, Curty LK, Barja F, van Delden C, Pechere JC. Swarming of *Pseudomonas aeruginosa* is dependent on cell-to-cell signaling and requires flagella and pili. J Bacteriol. 2000;182:5990-6.

47. Lindhout T, Lau PC, Brewer D, Lam JS. Truncation in the core oligosaccharide of lipopolysaccharide affects flagella-mediated motility in *Pseudomonas aeruginosa* PAO1 via modulation of cell surface attachment. Microbiology. 2009;155:3449-60.

48. D'Orazio M, Mastropasqua MC, Cerasi M, Pacello F, Consalvo A, Chirullo B, et al. The capability of *Pseudomonas aeruginosa* to recruit zinc under

conditions of limited metal availability is affected by inactivation of the ZnuABC transporter. Metallomics. 2015;7:1023-35.

49. Serino L, Reimmann C, Visca P, Beyeler M, Chiesa VD, Haas D. Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated *pchDCBA* operon in *Pseudomonas aeruginosa*. J Bacteriol. 1997;179:248-57.

50. Sanbongi Y, Shimizu A, Suzuki T, Nagaso H, Ida T, Maebashi K, et al. Classification of OprD sequence and correlation with antimicrobial activity of carbapenem agents in P*seudomonas aeruginosa* clinical isolates collected in Japan. Microbiol Immunol. 2009;53:361-7.

51. Bruchmann S, Dotsch A, Nouri B, Chaberny IF, Haussler S. Quantitative contributions of target alteration and decreased drug accumulation to *Pseudomonas aeruginosa* fluoroquinolone resistance. Antimicrob Agents Chemother. 2013;57:1361-8.

52. Tsutsumi Y, Tomita H, Tanimoto K. Identification of novel genes responsible for overexpression of *ampC* in *Pseudomonas aeruginosa* PAO1. Antimicrob Agents Chemother. 2013;57:5987-93.

53. Sainsbury S, Bird L, Rao V, Shepherd SM, Stuart DI, Hunter WN, et al. Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. *J Mol Biol* 2011, 405(1):173-184.

54. Gutu AD, Sgambati N, Strasbourger P, Brannon MK, Jacobs MA, Haugen E, et al. Polymyxin resistance of *Pseudomonas aeruginosa phoQ* mutants is

dependent on additional two-component regulatory systems. Antimicrob Agents Chemother. 2013;57:2204-15.

55. Moskowitz SM, Ernst RK, Miller SI. PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to lipid A. J Bacteriol. 2004;186:575-9.

56. Poole K, Tetro K, Zhao Q, Neshat S, Heinrichs DE, Bianco N. Expression of the multidrug resistance operon *mexA-mexB-oprM* in *Pseudomonas aeruginosa*: *mexR* encodes a regulator of operon expression. Antimicrob Agents Chemother. 1996;40:2021-8.

57. Vettoretti L, Plesiat P, Muller C, El Garch F, Phan G, Attree I, et al. Efflux unbalance in *Pseudomonas aeruginosa* isolates from cystic fibrosis patients. Antimicrob Agents Chemother. 2009;53:1987-97.

58. Sobel ML, Neshat S, Poole K. Mutations in PA2491 (*mexS*) promote MexT-dependent *mexEF-oprN* expression and multidrug resistance in a clinical strain of *Pseudomonas aeruginosa*. J Bacteriol. 2005;187:1246-53.

59. Kohler T, Epp SF, Curty LK, Pechere JC. Characterization of MexT, the regulator of the MexE-MexF-OprN multidrug efflux system of *Pseudomonas aeruginosa*. J Bacteriol. 1999;181:6300-5.

60. Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. PLoS Genet. 2013,9:e1003741.

61. Rau MH, Marvig RL, Ehrlich GD, Molin S, Jelsbak L. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. Environ Microbiol. 2012; 14:2200-11.

Figure Legends

Figure 1. Radial phylogeny of the AUST-02 genomes. The relationship of the M3L7 sub-lineage to other sequenced AUST-02 genomes from patients attending CF centres in Brisbane/Queensland (Red), Perth/Western Australia (Green) and Sydney/New South Wales (Blue) is shown. The outgroup (AUS970) is indicated. The major clades (M2 and M3) are represented by pink and yellow shaded boxes and are defined by different *mexZ* alleles (M2, A38T; M3, T12N). The scale bar represents 10 nucleotide substitutions. Phylogeny was reconstructed estimated from an alignment of 30,811 core genome SNPs (relative to PAO1) using RAxML.

Figure 2. Time-calibrated phylogeny of the M3L7 sub-lineage. Ancestral reconstruction was performed using BEAST 1.8.2 based on a 183 bp non-recombinant SNPs alignment for the 23 non-hypermutator M3L7 strains (sequenced in the current study) isolated between 2007 and 2011, with HKY substitution-, strict clock-, and exponential population tree- models preferred. Posterior probability support is indicated for each node. Paired samples from a patient are coloured according to the legend depicted on the bottom left corner. The x-axis represents the years between 2001-2011.

Figure 3. Genes (n=43) with non-synonymous SNPs and indels that characterise the M3L7 sub-lineage. The black squares indicate the PseudoCAP function of the gene. See Additional File 6: Table S2 for detailed information of the specific mutations. *Candidate pathoadaptive genes identified previously [8].

Additional Files

Additional File 1: Table S1. Details of *Pseudomonas aeruginosa* isolates used in this study.

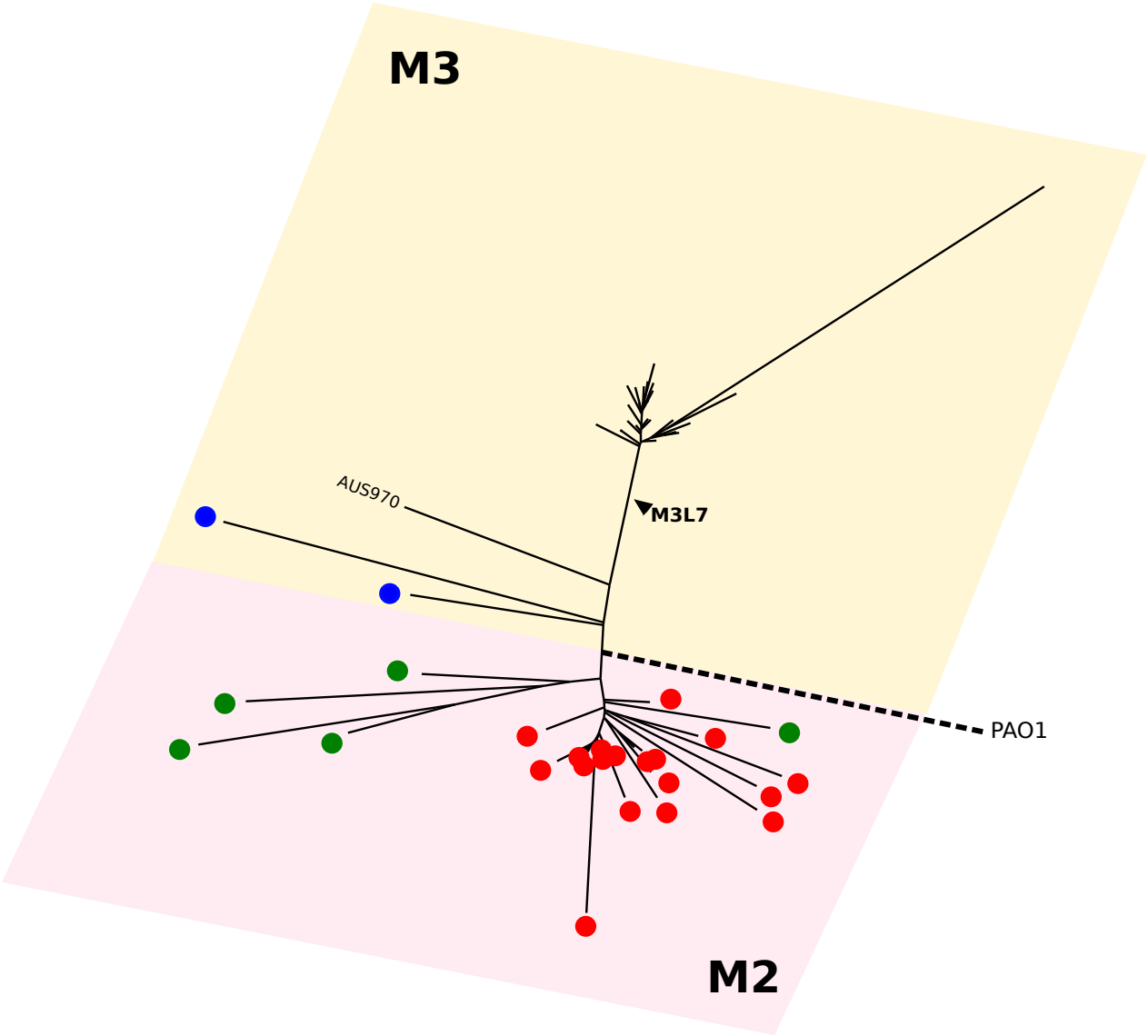
Additional File 2: Figure S1. Phylogeny of sequenced AUST-02 strains showing the M3L7 isolates in relation to other AUST-02 isolates within the major M2 and M3 clades. The Maximum-Likelihood phylogenetic tree was estimated from an alignment of 30,811 core genome SNPs using RAxML. >70% Bootstrap support (*), 100% bootstrap support (**). Scale indicates branch length representing 10 nucleotide substitutions. M2, *mexZ-2* allele (codon substitution, A38T); M3, *mexZ-3* allele (codon substitution, T12N**)**. Pairs of isolates (collected in 2007 or 2008 and 2011) are indicated by colour. Dotted lines indicate branches that have been shortened and are not to scale. Geographic location of CF treatment centre where each isolate was obtained is shown on the right of the tree.

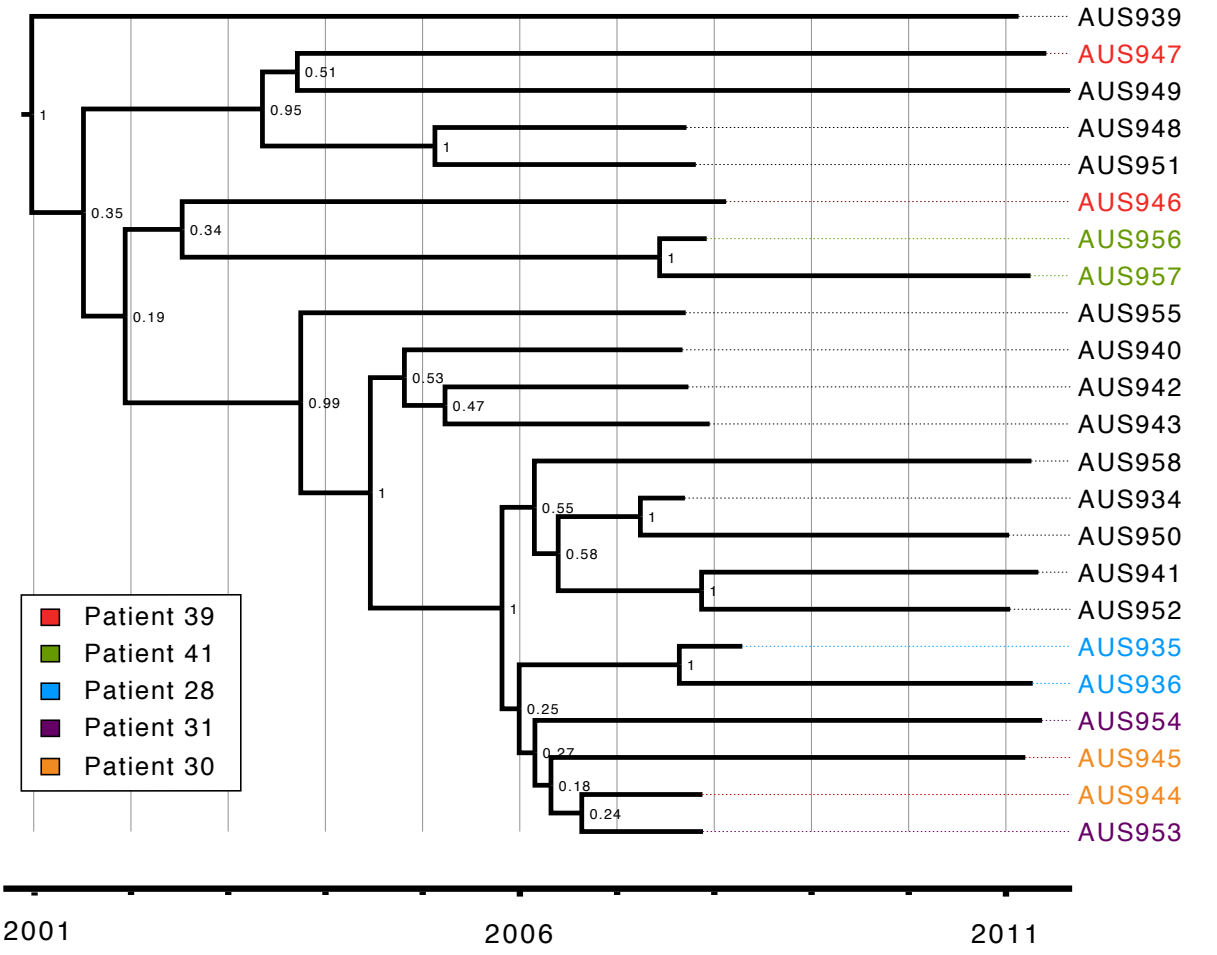
Additional File 3: Figure S2. Growth of the adult CF population at The Prince Charles Hospital between 2001 and 2015.

Additional File 4: Figure S3. A resolved phylogeny inferred from 364 core SNPs. The maximum likelihood tree was generated using RAxML with 1000 bootstrap replicates. Bootstrap values >70% are shown. Scale indicates branch length representing 5 nucleotide substitutions. Non-hypermutator isolates (n=23) sequenced as part of this study were included. Pairs of isolates were collected in 2007 or 2008 and 2011 as indicated by colour. M3, *mexZ-3* allele (codon substitution, T12N); L1, *lasR-1* allele (wild-type); L7, *lasR-7* allele (1 bp deletion, 438delG).

Additional File 5: Figure S4. Core SNP based minimum spanning tree depicting the most likely route of transmission. Isolates sequenced in this study included. Two isolates (AUS946 and AUS944) were predicted to be closest to the source. Non-hypermutator isolates sequenced in this study were included. P, patient. Figure generated using Phyloviz [29].

Additional File 6: Table S2. Position and details of non-synonymous SNPs (n=35) and indels (n=9) that characterise the M3L7 sub-lineage.


M3
AUS970
M3L7
PAO1
M2
10 SNPs


AUS939
AUS947
AUS949
AUS948
AUS951
AUS946
AUS956
AUS957
AUS955
AUS940
AUS942
AUS943
AUS958
AUS934
AUS950
AUS941
AUS952
AUS935
AUS936
AUS954
AUS945
AUS944
AUS953
1
0.51
0.95
1
0.35
0.34
1
0.19
0.99
0.53
0.47
1
0.55
1
0.58
1
1
1
0.25
0.27
0.18
0.24
Patient 39
Patient 41
Patient 28
Patient 31
Patient 30
2001
2006
2011

PseudoCap Function	*gyrB**	*spuG*	*mexA**	*yebB*	*migA*	PA0736	PA0948	*oprD**	*fleR*	PA1186	*hxuC/yebA*	*pmpA/ybiF*	*femA*	*fpvA*	PA2406	*mexS**	*mexT*	*infA*	PA3768	PA3828	*mpl**	PA4027	*bfmS*	*phzS*	*pchD*	*inaA*	*colS*	*ftsI*	*algW*	*rpoN*	PA4530	*pilR*	*phuR*	*folP*	*pmrB*	PA4796	PA4852	*purH*	*mdoH*	PA5174	PA5265	*znuA*	PA5530
Transport of small molecules		X	X					X			X		X	X											X								X									X	X
Membrane proteins		X				X					X	X								X																					X		X
Transcriptional regulators									X				X			X	X													X		X											
Two-component regulatory systems									X														X				X					X			X								
Cell wall / LPS / capsule					X																X		X					X											X				
Putative enzymes					X											X			X					X																			
Adaptation, Protection																							X			X			X														
Motility & Attachment									X																																	X	
Antibiotic resistance and susceptibility			X																		X																						
Translation, post-translational modification, degradation																		X											X														
Secreted Factors (toxins, enzymes, alginate)																									X				X														
Cell division																												X															
Biosynthesis of cofactors, prosthetic groups and carriers																																		X									
Nucleotide biosynthesis and metabolism																																						X					
Fatty acid and phospholipid metabolism																																								X			
DNA replication, recombination, modification and repair	X																																										
Hypothetical, unclassified, unknown				X			X			X					X							X									X					X	X						