



OPEN

SUBJECT AREAS:

MICROBIOLOGY
MICROBIAL GENETICS

Received

2 October 2014

Accepted

27 November 2014

Published

12 January 2015

Correspondence and
requests for materials
should be addressed to
S.P.D. (steve.diggle@
nottingham.ac.uk)

Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection

Sophie E. Darch¹, Alan McNally², Freya Harrison¹, Jukka Corander³, Helen L. Barr⁴, Konrad Paszkiewicz⁵, Stephen Holden⁶, Andrew Fogarty⁷, Shanika A. Cruz^{1,6} & Stephen P. Diggle¹

¹School of Life Sciences, University of Nottingham, Nottingham, NG7 2RD, U.K., ²Pathogen Research Group, Nottingham Trent University, Nottingham, U.K., ³Department of Mathematics and Statistics, University of Helsinki, Finland, ⁴Division of Respiratory Medicine, Nottingham City Hospital, Nottingham, NG5 1PB, U.K., ⁵College of Life and Environmental Sciences, University of Exeter, Exeter, ⁶Department of Clinical Microbiology, Nottingham University NHS Trust, U.K., ⁷Division of Epidemiology & Public Health, Nottingham City Hospital, Nottingham, NG5 1PB, U.K.

The Cystic Fibrosis (CF) lung harbors a complex, polymicrobial ecosystem, in which *Pseudomonas aeruginosa* is capable of sustaining chronic infections, which are highly resistant to multiple antibiotics. Here, we investigate the phenotypic and genotypic diversity of 44 morphologically identical *P. aeruginosa* isolates taken from a single CF patient sputum sample. Comprehensive phenotypic analysis of isolates revealed large variances and trade-offs in growth, virulence factors and quorum sensing (QS) signals. Whole genome analysis of 22 isolates revealed high levels of intra-isolate diversity ranging from 5 to 64 SNPs and that recombination and not spontaneous mutation was the dominant driver of diversity in this population. Furthermore, phenotypic differences between isolates were not linked to mutations in known genes but were statistically associated with distinct recombination events. We also assessed antibiotic susceptibility of all isolates. Resistance to antibiotics significantly increased when multiple isolates were mixed together. Our results highlight the significant role of recombination in generating phenotypic and genetic diversification during *in vivo* chronic CF infection. We also discuss (i) how these findings could influence how patient-to-patient transmission studies are performed using whole genome sequencing, and (ii) the need to refine antibiotic susceptibility testing in sputum samples taken from patients with CF.

The Cystic Fibrosis (CF) lung presents a complex polymicrobial ecology, which in turn complicates the treatment of chronic bacterial respiratory infection. The pathogen most commonly associated with CF is *Pseudomonas aeruginosa*. Ubiquitous in the environment, *P. aeruginosa* is one of the leading causes of opportunistic infections in humans¹, and is a major cause of increased morbidity and mortality in patients with CF². Its ability to colonise pulmonary epithelial cells, communicate using quorum sensing (QS) signals and form multicellular biofilms, plus its innate resistance to many antimicrobials, results in chronic infections that are almost impossible to eradicate, leading to a clinical pattern of intermittent exacerbations and an eventual decline in lung function^{2–4}. Furthermore, initial colonising strains of *P. aeruginosa*, thought to be acquired from the environment, have the potential to be displaced by more virulent and transmissible epidemic strains^{5–7}.

It is now well established that founder *P. aeruginosa* populations evolve over many years of chronic CF infection, leading to high levels of temporal phenotypic and genetic diversity within a single patient^{5,8–15}. Longitudinal genomic studies of single colonies isolated from individual CF patients suggest such variation is derived from a mixture of single base pair mutations, insertion and deletion events and recombination events^{10,16}. Additionally, recent studies have examined the diversity of *P. aeruginosa* populations isolated from individual patients at a single time point¹⁴. These suggest that considerable phenotypic variation exists at any given time and Pulsed-Field Gel Electrophoresis (PFGE) is suggestive of some level of genetic variation¹⁴. Despite these studies, it is still not clear how such diversity arises and how this impacts on clinically important factors such as identifying transmission events between patients and performing reliable antibiotic susceptibility testing.



Recent studies examining diversity within *P. aeruginosa* populations isolated from the CF lung have described significant variation in antibiotic susceptibility profiles in isolates which vary in morphological appearance. The most commonly described difference is that observed between mucoid and non-mucoid colonies^{2,9,17}. However, no study has conducted a detailed examination of a single, morphologically homogeneous population of *P. aeruginosa*, nor has any study provided a comprehensive map at the genome level of phenotypic variation within an extant CF lung population.

Here we address this gap in our understanding of how diversity evolves in the CF lung and how this may impact on decisions made in a clinical setting, where microbiological analyses are often based on a single colony. We have comprehensively analyzed the levels of phenotypic and genotypic diversity within a single CF lung ecosystem at a single snapshot in time. We reveal previously unobserved correlations and trade-offs between the production of different virulence-related exoproducts, quorum sensing (QS) signals and resistances to a range of clinically-relevant antibiotics. Furthermore, we show significant increases in antibiotic resistance when multiple isolates are cultured together. Finally, we use a range of genomic analyses to quantify genotypic diversity within this single patient. These reveal (i) the dominant role of recombination over spontaneous mutation in the generation of *in vivo* diversity; (ii) unexpected and complex relationships between genotype and phenotype; (iii) the difficulty of classifying patient-to-patient transmission events based on the analysis of single colonies; and (iv) the likely underestimation of antibiotic resistance based on testing a single or few colonies per patient.

Results

***P. aeruginosa* isolates from a single sputum sample display phenotypic tradeoffs.** We randomly selected 44 morphologically identical, non-mucoid *P. aeruginosa* colonies (hereafter referred to as isolates) from a single spontaneously expectorated sputum sample from a clinically stable CF patient with a chronic *P. aeruginosa* infection. We assayed each isolate for overnight growth in standard laboratory medium and for a selection of phenotypes that have previously been associated with virulence in CF infection. The isolates displayed considerable variation in growth and also in the production of tissue-degrading proteases (LasA protease and LasB elastase), the redox-active toxin pyocyanin and the QS signal molecules *N*-butyryl-L-homoserine lactone (C4-HSL), *N*-(3-oxododecanoyl)-L-homoserine lactone (3O-C12-HSL), 2-heptyl-3-hydroxy-4(1*H*)-quinolone (PQS), 2-heptyl-4-hydroxyquinoline *N*-oxide (HQNO) and 2-heptyl-4-quinolone (HHQ). These data are summarised in Fig. 1a and b.

We then wished to determine the extent to which the expression of different virulence factors covaried. On the one hand, the variation shown in Fig. 1a and b could reflect a set of isolates with phenotypes that range from generally low virulence (poor growth and low levels of virulence factor expression) to generally high virulence (extensive growth and high levels of virulence factor expression). On the other hand, the isolates could have qualitatively different phenotypes – e.g. some may produce lots of protease but very little pyocyanin, while others may show the opposite pattern, suggesting that expression of one phenotype could be traded off against another. To address this question, we conducted principal component analyses (PCA).

The results of a PCA on growth and three virulence-associated exoproducts (LasA, LasB and pyocyanin) are shown in Fig. 1c. This revealed negative correlations between the traits, such that no individual isolate demonstrated high values for all four variables. In particular, higher growth was associated with lower per-cell production of LasA protease and LasB elastase, whilst higher pyocyanin production entailed lower LasA protease production: these relationships are illustrated by the vectors for the original variables pointing away from each other on the PCA plot. The first two principal com-

ponents explained approximately 70% of the total variation in these phenotypes. Pairwise Spearman's rank correlations produced results consistent with the PCA and the results for phenotype-related traits were unchanged by the exclusion of one outlier (SED23) (Supplementary Fig. S1).

We conducted a separate PCA on the production of five key QS signals, which between them regulate the expression of a range of virulence factors. QS signals have previously been extracted from the sputum of CF patients^{18,19}. Production levels of the five molecules were generally positively correlated, reflected by the close alignment of the vectors representing the original variables on the PCA plot (Fig. 1d). The first two principal components account for almost 90% of the total variation in the data set and the majority of the data points varied mainly in their level of 3O-C12-HSL production. Consistent with the PCA results, pairwise Spearman's rank correlation coefficients (Supplementary Fig. S1) revealed an almost perfect correlation between HQNO and PQS production ($r_s = 0.96$, $p < 0.001$), both of which were also strongly correlated with the production of C4-HSL ($r_s = 0.89$ and 0.85 respectively, $p < 0.001$) (Supplementary Fig. S1).

Whole genome SNP typing shows high levels of diversity and signatures of adaptive events in the extant *P. aeruginosa* population.

To contextualize the high levels of phenotypic diversity observed within this single sputum sample, we performed whole genome sequencing on 22 of the 44 isolates (SED 1–22), which were selected to represent the full spectrum of phenotypic diversity. Both methods confirmed that all isolates were close relatives of the Liverpool Epidemic Strain (LESB58), a particularly aggressive epidemic clone of *P. aeruginosa*²⁰. As such, we performed mapping of raw sequence data for all 22 variants against the LESB58 reference genome using SNP calling parameters equivalent to those used in recent genomic epidemiological studies of microbial infection^{21,22}. The resulting phylogeny (Fig. 2) and SNP distance matrix (Supplementary Fig. S2) confirmed that the patient was initially infected with a clone very similar to LESB58, which then most probably diversified within the lung, resulting in considerable levels of intra-population diversity. There were a total of 121 high-resolution SNPs called across the population, with 24 of those present across all isolates compared to the reference and so considered ancestral variants. None of the 22 isolates had a SNP difference range within 5 SNPs of the LESB58 reference strain and the most extreme differed by 64 SNPs (Supplementary Fig. S2 and Supplementary Table S1). Of the 77 SNPs which varied across the isolates, 45 were present in multiple isolates randomly distributed across the phylogeny (Table 1).

Despite all of these variants originating from the same single founder clone, such dissemination of SNPs is reminiscent of homoplasic mutation, where mutations in the same nucleotide occur independently in phylogenetically distinct individuals. A number of the homoplasic-like sites also appeared to be under the force of positive selection as determined by dN/dS ratios. This indicated that non-synonymous mutations in *mexB*, PLES_11151, PLES_28341, *glyA2*, and PLES_59241 all displayed signatures of adaptive evolution in this extant population with dN/dS ratios far in excess of 1. Attempts to superimpose the observed phenotypic diversity of the 22 isolates onto the phylogeny failed to show any pattern of acquisition or accrual of phenotypes along the evolutionary trajectory of the phylogeny, with phenotypes appearing randomly across the tree (Fig. 2). This is consistent with our observations of homoplasic-like accrual of mutations which can occur as a result either of fixation of selectively advantageous mutations, or by intra-population recombination. None of the high stringency SNPs identified in the mapping process showed a meaningful biological correlation with the phenotypic diversity. The vast majority of isolate-unique SNPs occurred in SED8, however this isolate showed no mutations associated with hypermutation (nor did any other isolate) and all isolates failed to

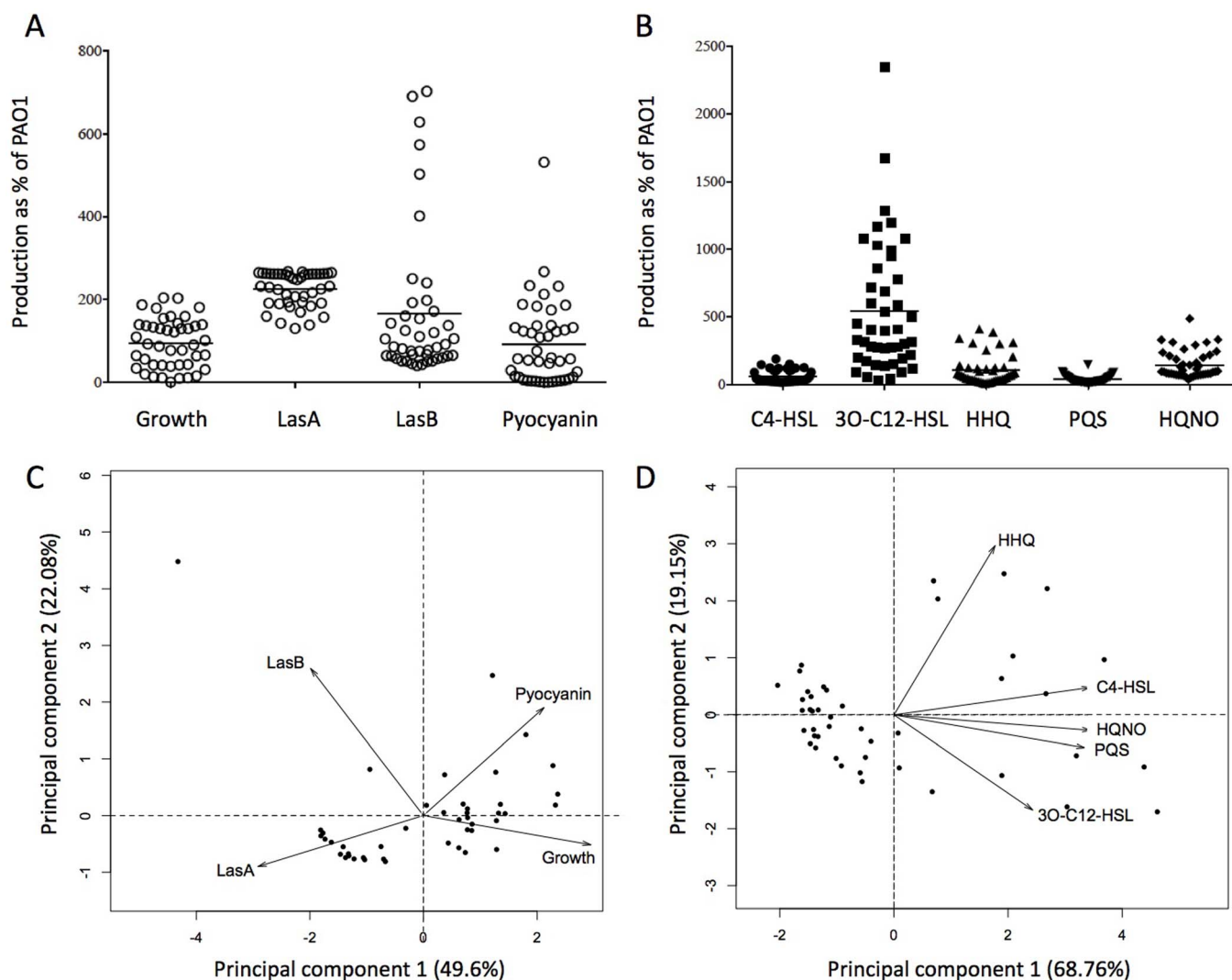


Figure 1 | Phenotypic diversity of *P. aeruginosa* populations. We measured growth, exoproducts and QS signal molecules of individual isolates and expressed these as a percentage of the values obtained for a PAO1 wildtype. Each data point represents the mean of triplicate assays for an individual isolate. (A) Phenotypic assays for growth and exoproduct production show that isolates display large variation between individuals when compared with the PAO1 reference strain. (B) QS Signal molecule measurements show that variation occurs between isolates when compared to the PAO1 reference strain. Panels (C) and (D) show the results of PCA on phenotypic and QS signal data respectively. PCA reduces multiple variables (four phenotypic traits or five QS molecules) to two dimensions, allowing us to plot multivariate data on simple x,y coordinates. The arrows are vectors that show how the original variables relate to the new x and y axes. The PCA plots show that QS signals are linked and that tradeoffs exist between other phenotypes measured.

display classical hypermutation phenotypes when analysed using standard rifampicin spontaneous resistance testing (data not shown).

Analysis of *de novo* assembled genomes indicates a dominant role for recombination in generating diversity in the extant population. Whilst high resolution SNP typing gives the most accurate determination of variation within the population, it does not take into account large-scale insertion, deletion, and chromosomal rearrangement events which may generate diversity. Corrected *de novo* genome assemblies were constructed for each isolate using Velvet and PAGIT, and from these, whole genome alignments were performed using progressiveMauve and pairwise Blast comparisons using Blast Ring Image Generator (BRIG). Both comparative methods showed a lack of large-scale insertion or deletion events responsible for the phenotypic diversity. The loss of a 26.4 Kbp cryptic phage in isolates SED2, SED11 and SED19 compared to LESB58 is the most obvious region of difference (Supplemental Fig. S3). However this event shows no correlation

to any phenotypic pattern seen in these isolates. Perhaps more significantly, there are no acquisition events in any of the 22 variants sequenced as determined by a Progressive Mauve alignment of the *de novo* assembled sequences (Fig. 3). Given that large-scale acquisitions and deletions are well documented in longitudinal CF *P. aeruginosa* populations, we further confirmed our finding by constructing a pan-genome of the *de novo* assembled genomes, our 22 variants and LESB58, and plotted the presence and absence of every locus as a heat map (Supplemental Fig. S4). Our data clearly shows there are no large genomic islands that have been acquired by any of the variants. There are a small number of genetic loci which appear to be strain specific, however BLAST analysis identified these as genes with highly divergent sequences compared to the ortholog in LESB58.

We also checked for major chromosomal inversions and rearrangements which may confer intra-population diversity by utilising our paired-end sequencing reads and performing break-point analysis on each *de novo* assembly compared to the reference LESB58 genome (Supplemental Fig. S5). Our analysis suggests the presence of

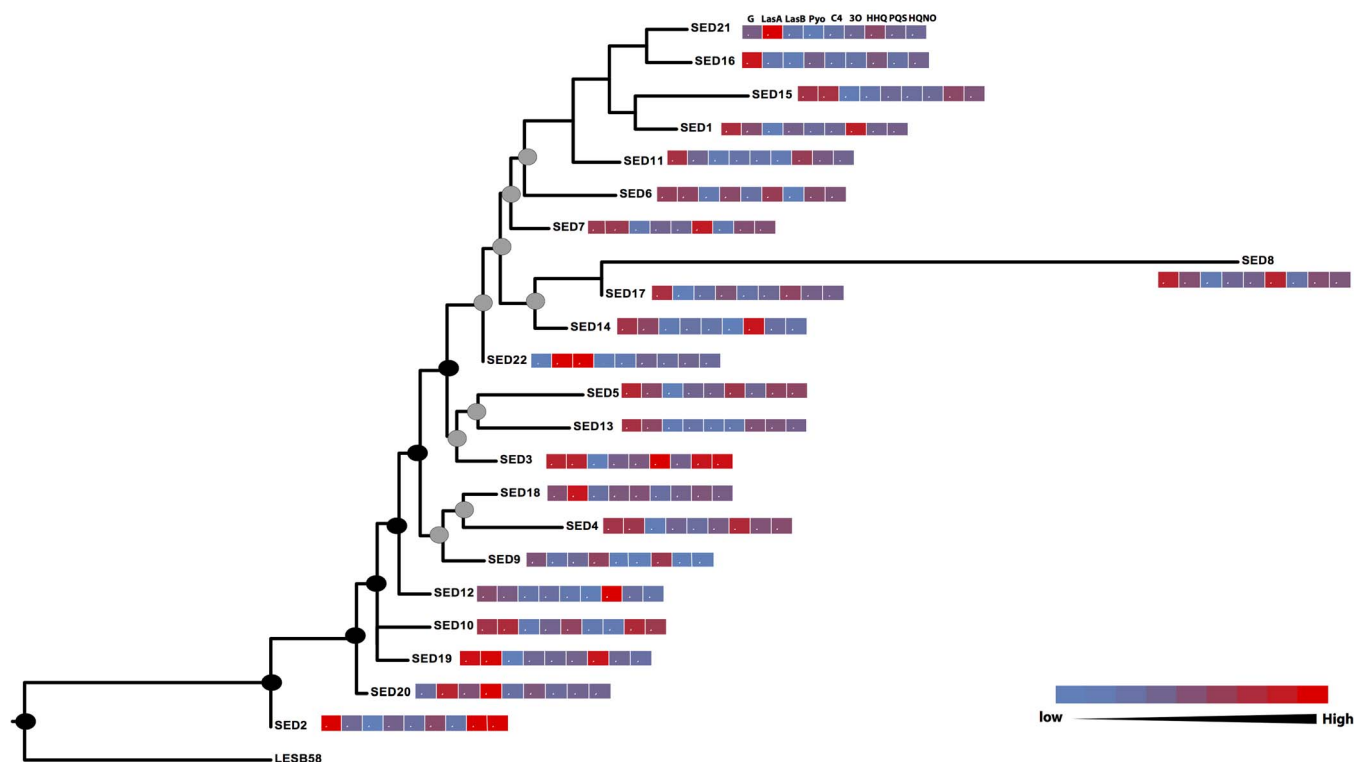


Figure 2 | Maximum likelihood phylogeny based on SNP typing of the 22 sequenced isolates against LESB58. Nodes obtaining $>95\%$ bootstrap support (black circles) and $>75\%$ bootstrap support (grey) are indicated on the tree. The heatmaps accompanying each taxa represent the levels of production of each phenotype relative to the reference PAO1 strain and are indicated above SED21. Abbreviations: G (growth), LasA (LasA protease), LasB (protease), Pyo (pyocyanin), C4 (C4-HSL), 3O (3O-C12-HSL).

just a single observable inversion event in isolate SED5 with the genomic architecture across all 22 variants almost identical by our analysis. Finally we checked for subtle differences, which may result in frameshift mutations. Our previous SNP analysis identified just one mutation in the *mexB* gene present in 2 variants, which introduced a premature stop codon, and two SNPs affecting intergenic regions possibly affecting gene expression. To confirm there were no small indel events causing frame shifts, we systematically compared each of the 22 variant genomes against LESB58 using Artemis Comparison Tool (ACT) looking for truncated ORFs or CDS mergers, as well as indels in intergenic regions in the variants, however no such frameshift mutations were found in the data set.

Given the identification of a number of regions with highly divergent sequence in the variants compared to LESB58, and the earlier identification of homoplasic-like mutations randomly distributed across the population phylogeny, we determined the levels of recombination in the de novo assembled genomes. BRATNextGen analysis on a core genome alignment detected recombination events in all genomes ranging from a single event in SED19 to seven recombination segments in SED18 (Supplemental Fig. S6), with almost all detected recombinant segments shared across some members of the data set, indicative of intra-population recombination. From the alignment of corrected genome assemblies, there were a total of 1436 single nucleotide polymorphisms (SNPs) across the 22 genomes, with 1296 present in regions identified as recombining, giving an r/m value of 9.2571. This indicates that the SED isolates are approximately ten times more likely to have acquired genetic variation by recombination than by spontaneous mutation. More importantly of the 45 high resolution SNPs exhibiting homoplasic-like properties, 40 were located on recombining regions, supporting the likelihood that mutations are being propagated across the extant population by recombination.

Phenotypic diversity in the extant *P. aeruginosa* population does not associate with previously reported and associated mutations.

As mentioned previously, attempts to superimpose the observed phenotypic diversity of the 22 isolates onto the phylogeny failed to show any pattern of acquisition or accrual of phenotypes with variable phenotypes appearing randomly across the tree (Fig. 2), which is consistent with dissemination via recombination. None of the high stringency SNPs identified in the mapping process showed a correlation with the phenotypic diversity we observed as reported in classical bacterial genetics experiments. Given our observation of significant intra-population recombination, we investigated if there was a statistical correlation between recombination events and phenotypic diversity. We utilised the widely accepted technique of performing permutation tests of phenotypes against the whole genome alignments²³ and identified between 5 (LasA production) and 15 (HHQ production) recombining regions significantly associated ($p < 0.05$) with variation observed in a given phenotype (Fig. 4; Supplementary Table S2), including regions enriched for homoplasic mutations in the high-confidence SNP typing data (Table 1).

Each of these recombining regions was mapped onto the LESB58 reference genome, allowing us to identify the CDS present in each region (Fig. 4). This analysis showed a high degree of overlap in the recombining regions associated with different phenotypic changes. Of particular interest are the recombining loci significantly associated with altered levels of production of PQS and HQNO which are completely identical (Supplementary Table S2), suggesting these two phenotypes are intrinsically linked. This observation matches our phenotypic observation that PQS and HQNO production are almost perfectly correlated (Fig. 1d and Supplementary Fig. S1). Though a small number of recombining loci are associated with a change in just one phenotype, a total of 19 recombining CDS are significantly associated with an alteration in 2 or more observed phenotypes. Of



Table 1 | Homoplasic SNPs involved in adaptive evolution

SNP position	CDS	Mutation	No. Haplotypes with mutation	dN/dS
468009	<i>mexB</i>	Stop codon	2	10.658
1211690	PLES_11151 – Carboxyl esterase	Val to Ala	2	9.99
1469230	PLES_13621 – Hypothetical protein	synonymous	5	1.12
1469235		synonymous	5	
1469256		synonymous	9	
1469277		synonymous	13	
1469280		synonymous	12	
1469290		Pro to Ser	16	
1469298		synonymous	15	
1469310		synonymous	17	
1469421		synonymous	21	
1469485	PLES_13631 – hypothetical protein	synonymous	19	0.008
1469498		Leu to Iso	21	
1469581		synonymous	18	
1469584		synonymous	18	
1469593		synonymous	11	
1658081	PLES_15261 – NADH quinone reductase	Glu to Ala	3	1.023
1676171	PLES_15441 – mechanosensitive ion channel	Val to Ala	2	1.063
1986123	PLES_18391 – ATP binding permease	Synonymous	2	0.001
2690373	rRNA-Ser	NA	7	
2690375		NA	5	
3020931	<i>mexT</i>	Val to Ala	2	1.032
3048652	PLES_28331 – putative haemolysin secretion/activation protein	synonymous	19	0.009
3048751		synonymous	17	
3058594	PLES28341 – hypothetical protein	Gly to Ala	9	11.478
3058611		Ala to Thr	5	
3090543	<i>glyA2</i>	Leu to Iso	16	6.06
3090547		Glu to Arg	11	
3090548		Glu to His	8	
3090609		Iso to Leu	4	
3378695	<i>psiK</i>	Synonymous	2	0.001
3378321	<i>pslJ</i>	synonymous	7	0.002
3478510	PLES_31551 - hypothetical	Leu to Phe	2	1.098
3502354	PLES_31741 – trehalose synthase	Leu to Val	2	1.049
3973880	Intergenic PLES_35761 - <i>cysB</i>	NA	5	
4015334	<i>pcrV</i>	Asp to Asn	2	1.104
4026225	<i>pscT</i>	Asn to Lys	4	1.088
4539068	PLES_41121 – NAD deacetylase	Val to Ala	4	1.099
5133783	Intergenic PLES_46606 - <i>sbcD</i>	NA	7	
5172963	<i>pctC</i>	Leu to Pro	5	1.008
6430173	<i>glyA1</i>	Glu to Arg	16	0.305
6430175		synonymous	19	
6557923	PLES_59241 – hypothetical protein	Ala to Gly	14	3.227

equal importance is that none of the loci have been associated with such phenotypic properties by way of classical bacterial genetics and mutagenesis studies.

Phenotypic and genetic diversity results in variation in antibiotic susceptibility between isolates. Previous studies have shown that there is little correlation between antibiotic sensitivity patterns of *P. aeruginosa* isolated during a pulmonary exacerbation and the subsequent clinical response of the patient^{24,25} and reproducibility of the diagnostic test itself has been shown to be low between sample replicates^{26,27}. We determined the resistance of our 44 isolates to 9 commonly used CF therapeutic agents using the BSAC disk diffusion method. We noted large variances in susceptibility between isolates according to the zone of inhibition produced (Fig. 5a). Pairwise correlation analysis revealed several significant positive correlations indicating the tendency for some isolates to have relatively high resistance to multiple antibiotics whilst others had relatively low resistance to multiple antibiotics (Supplementary Fig. S7). Correlations in resistance were also found between antibiotics with similar mechanisms of action (such as the aminoglycosides).

We then mixed all 44 isolates as a population for comparison with individual isolates and repeated the antibiotic susceptibility testing.

We found that variation in the zone of inhibition was reduced in mixed versus single isolates, indicating that picking colonies at random and mixing them together as a population gives a more consistent antibiotic resistant phenotype (Fig. 5b and Supplementary Fig. S8). Furthermore, we found that mixing all 44 isolates together resulted in a level of antibiotic resistance that was higher than the mean of the individual isolates tested ($t_{17,450} = -2.208$, $p = 0.028$), indicating that testing at the individual isolate level may underestimate resistance (Fig. 5b). The complexity of the antibiotic resistance variation observed across all our isolates meant that any attempts to significantly associate recombination events to specific changes in resistance patterns proved unsuccessful. This is almost certainly due to the restricted size of our data set, with recent GWAS studies mapping specific recombination and mutation events to changes in susceptibility profiles requiring thousands of isolates to achieve statistical significance^{28,29}.

Discussion

A number of recent studies have shown extensive phenotypic variation between different *P. aeruginosa* CF strains and isolates of the same strain taken from single patients^{5,8,9,11,12,14,30,31}. Despite all these studies, the reasons for this phenotypic diversity remain poorly

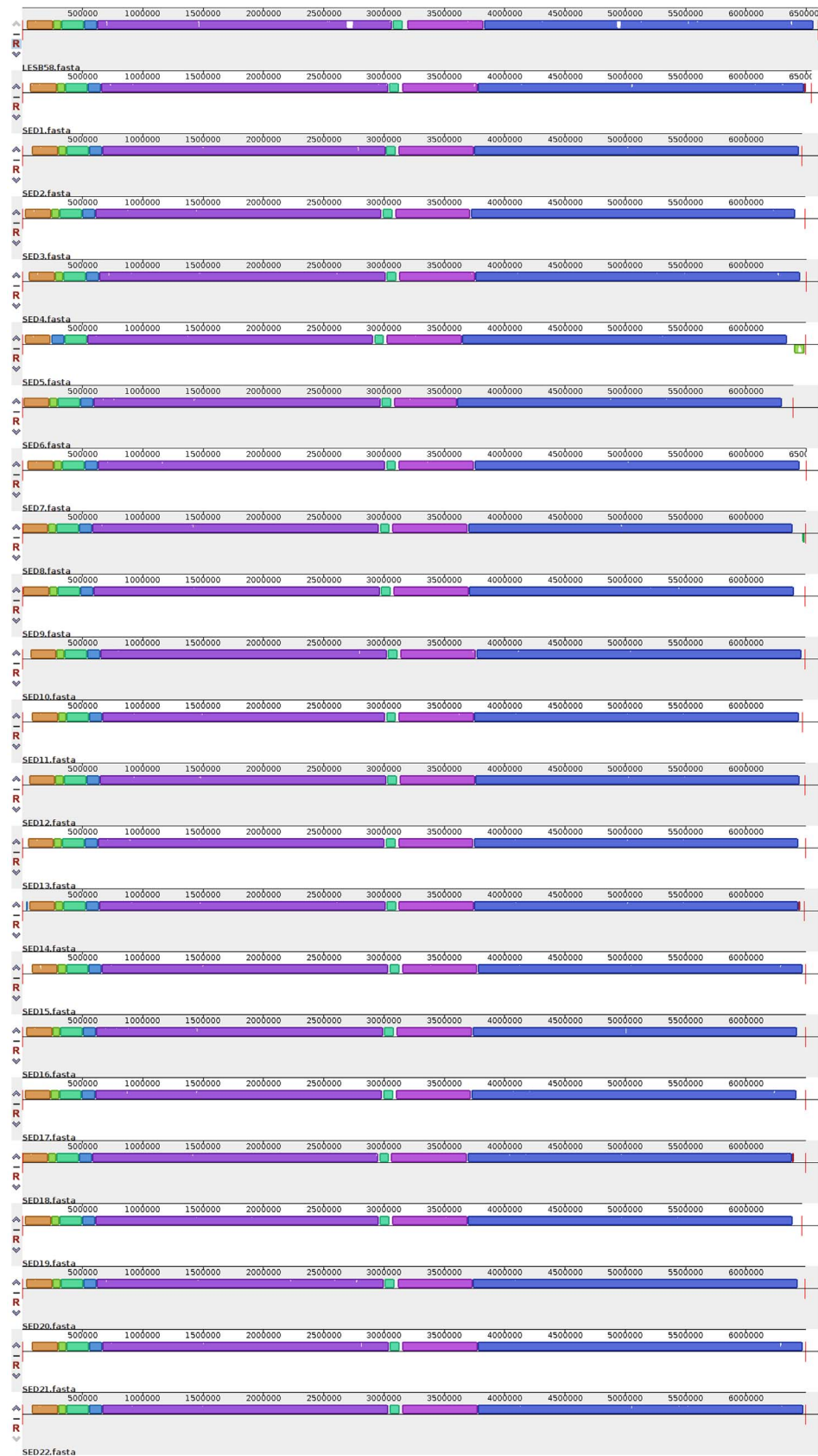


Figure 3 | Whole genome alignments of the de novo assembled genomes of 22 isolates and the reference genome LESB58. Alignment was constructed using Progressive Mauve, and local collinear blocks containing orthologous sequence are colour coded.

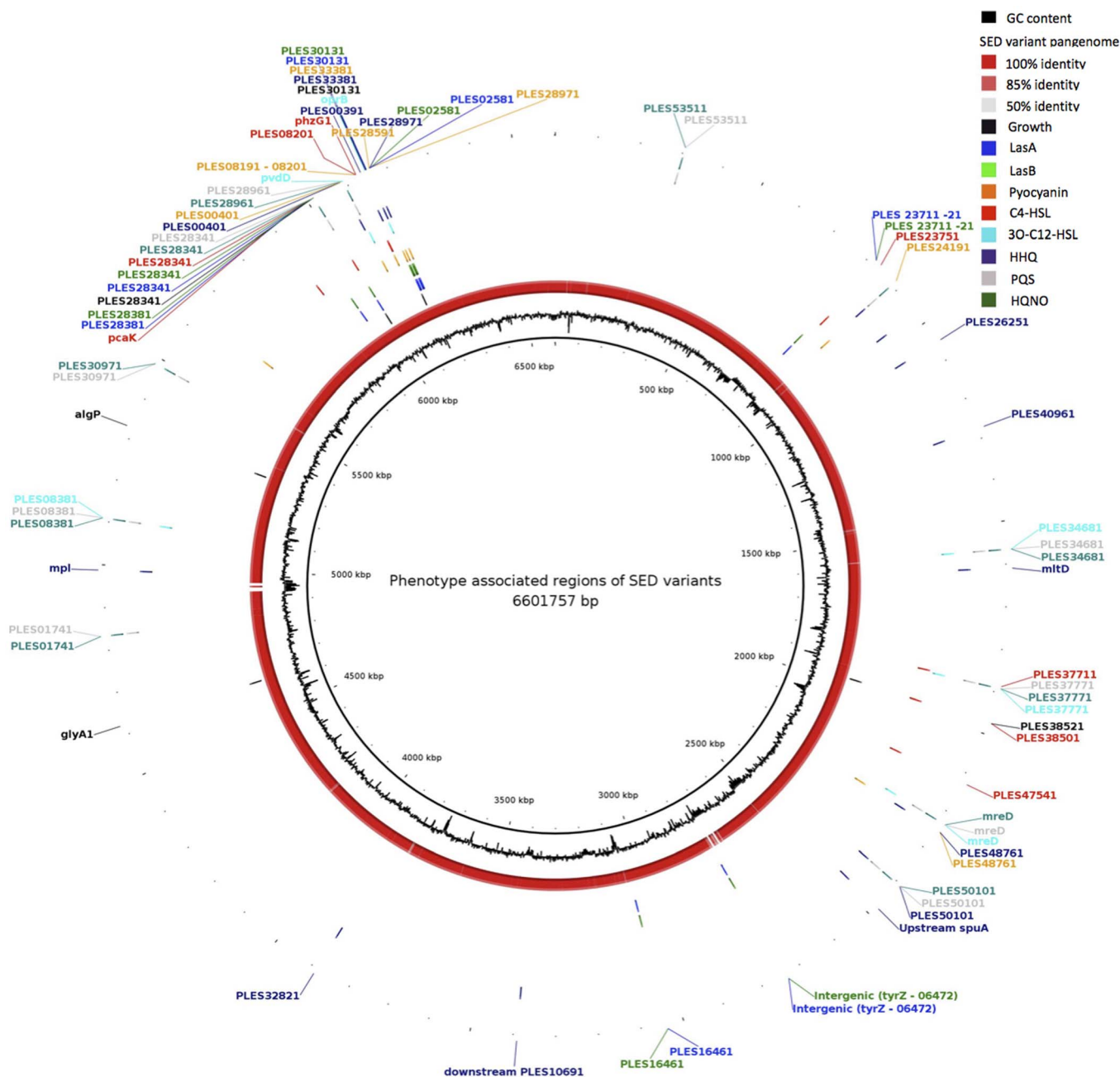


Figure 4 | Circular representation of regions identified as undergoing recombination relative to the LESB58 reference genome. Regions statistically associated with phenotypes are indicated in the concentric ring representing that phenotype (Growth innermost to HQNO outermost). The ORFs encoded in recombining regions are indicated by their gene name as annotated in LESB58. The Gene name-tags are colour coded to indicate clearly the phenotypes they are associated with.

understood. In addition, there is little to no information available about whole genome resolution of the extant diversity of *P. aeruginosa* within a CF lung at any given time point. Here we show that *P. aeruginosa* intra-host diversification is strongly linked to recombination. Our findings have broad implications for (i) classical phenotype/genotype correlations; (ii) informing contact transmission networks and (iii) routine antibiotic susceptibility testing.

We began our study by taking 44 morphologically identical non-nucoid ‘isolates’ from a single sputum sample, obtained from a patient with CF, known to be chronically infected with *P. aeruginosa* for approximately 3 years. At the time of sampling the patient was clinically stable and was prescribed standard maintenance oral and nebulised therapies only. Although all of the isolates appeared morphologically indistinguishable on agar plates, with no small colony

variants, they actually displayed extensive phenotypic variation when screened for growth, pyocyanin, LasA and LasB production and QS signal molecule production. This is in broad agreement with other recently published work. For example, it has been reported that the majority of phenotypic diversity occurs within patients rather than between patients⁹ and that identical colony morphotypes from a given patient demonstrate a large degree of phenotypic variation¹⁴.

We used our phenotypic data set to determine whether there were correlations between (i) phenotypes and (ii) QS signal molecules. Extensive molecular studies over the past 20 years have shown that particular traits are linked by complex genetic networks. For example, the production of different QS signals have been shown to be linked, whilst QS links key phenotypes such as LasB (elastase) and pyocyanin^{32,33}. The majority of these studies have been under-

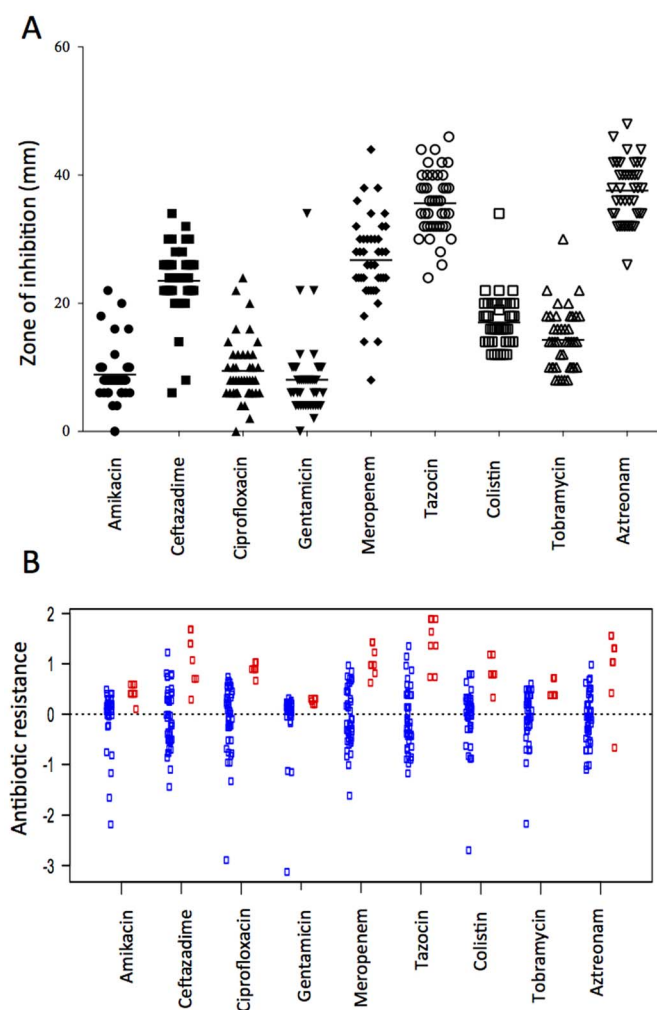


Figure 5 | Antibiotic resistance profiles for 44 isolates measured using the BSAC method. (A) The antibiotic susceptibility of single isolates to common CF therapeutics. The recorded zones of inhibition and mean value for each 44 individual isolates are shown. Each data point represents the mean of three independent biological replicates for each individual isolate. (B) Resistance profiles of each individual isolate to 9 antibiotics compared to a mixed community of all 44 isolates. Each blue point represents the sensitivity profile of a single isolate and red points represent 7 independent measurements of a mixed community containing all 44 isolates. In both (A) and (B), the values for individual clones represent the mean of three independent biological replicates. Individual clones consistently underestimate the resistance of the mixed community.

taken using the laboratory standard *P. aeruginosa* strain PAO1, a wound isolate taken during the 1950's³⁴. Using PCA we found positive correlations between five key QS signal molecules produced by *P. aeruginosa*, with particularly strong correlations between PQS and HQNO and PQS and C4-HSL. These positive correlations broadly fit with what has been described in the literature for PAO1. In contrast we found trade offs between the other phenotypes tested (growth, LasA, LasB & pyocyanin). For example, increased growth resulted in a reduced per cell production of both LasB and LasA (Fig. 1). Our results demonstrate that phenotypic correlations found in well-studied laboratory wild types of *P. aeruginosa* do not necessarily match in *P. aeruginosa* strains recently collected from *in vivo* infections. The reasons for these trade-offs are unknown, but a number of factors, such as within-host competition, lung spatial structure, host immune response, transmission events and host survival, could produce evolutionary trade-offs between phenotype and parasite virulence^{35,36}.

We attempted to contextualize our phenotypic observations by performing genome sequencing on 22 isolates, which were chosen to represent the full spectrum of diversity observed. Examination of differences in gene content showed there were no gene acquisition or loss events across our isolate set. A small cryptic prophage had excised from three of the isolates but this does not correlate with any particular phenotypic pattern. Previous studies examining the diversity of LES strains from across the globe have shown far greater variation in gene content between strains with variation in large genomic islands^{37,38}. In contrast a study of 55 isolates from 21 individuals showed no large scale insertions or deletions, as well as the presence of parallel patho-adaptive mutations³⁹. PFGE based studies examining extant diversity in patients has shown variations in genomic architecture⁴⁰. Similarly frameshifts and small indels have previously been reported to be key in generating phenotypic diversity within clonal *P. aeruginosa* populations⁴¹, but such genetic variation was not detected in our data set. Conversely a recent study, which investigated the experimental evolution of a *P. aeruginosa* population over time⁴², showed very few indels fix within a population over time, and that small indels resulting in frameshift mutations came under extremely strong purifying selection.

Our SNP phylogeny showed a total of 121 variable sites with respect to the LESB58 reference genome, 24 of which were common and so considered ancestral to the original infecting strain. This leaves a total of 97 differentiating SNPs in our set of variant strains. Such levels of diversity within a patient infected for 3 years seem high when one considers a recent genomic study of *P. aeruginosa* strain PA14, which reported an accumulation of only 15 SNPs over a period of 15 years⁴³, and a study of global and temporally distributed LES strains which reported a distance range of 66–156 SNPs³⁷. However our observed levels of extant diversity are not unreasonable if one considers the mutation levels calculated for *S. aureus* of 1 core genome SNP every 6 weeks⁴⁴ which would equate to an expectation that any random isolate variant would differ from its ancestor by approximately 27 SNPs after the 3 year period in which the lung has been infected. This is within the magnitude of values obtained between any pair of isolates in our pairwise SNP distance matrix with the exception of SED8. The possibility also exists that the differences in SNP levels between our study and those conducted in PA14 reflect the differences in mutation accumulation occurring during the initial stages of adaptation compared to those occurring in an established infection where the infecting strain is well adapted to the lung environment.

The SNPs observed at 45 of the reference genome variant sites exhibited characteristics of homoplasy. A number of the homoplastic-like sites also displayed dN/dS ratios far in excess of 1, which is a strong signal for positive selection, suggesting these sites are undergoing adaptive evolution. Our data showing multiple homoplastic-like sites under neutral selection in combination with defined sites under adaptive evolution mirrors that recently shown in a comprehensive study of within patient diversity of *Burkholderia dolosa* within the CF lung⁴⁵. The *B. dolosa* study was performed on multiple isolates accumulated from multiple patients over time and provided a definitive blueprint for pathogen evolution in the CF lung. Our study expands on this by examining extant isolates of a far more prevalent CF pathogen, *P. aeruginosa*, in a single patient, and provides further evidence for the forces underpinning the generation of diversity in a pathogen population in the CF lung. The vast majority of our identified SNPs occurred in hypothetical proteins as opposed to surface associated factors⁴⁵ or drug resistance associated loci⁴⁴.

Our finding that the vast majority of SNPs within our isolate data set are present in regions undergoing recombination also provides further resolution to the mechanisms by which diversity may be generated within the CF lung. It has long been known that genetic recombination occurs in *P. aeruginosa*³⁴, and our data suggests that recombination is an important factor in the short-term generation of



genetic diversity within *P. aeruginosa* isolates within the CF lung with an r/m rate approaching 10. Homologous recombination has been shown to play an important role in the early adaptation of *P. aeruginosa* to the human lung, principally in creating deletions leading to genome reduction¹⁶. That particular study focused on a collection of isolates from multiple patients spanning a 35-year period whereas our study focuses on the diversity sampled at a single time point, and therefore our data suggests for the first time, that recombination is extremely important in the continuous evolution of *P. aeruginosa* within individual patients with CF.

Instances of large scale acquisitions, indels, frame shifts, and chromosomal inversions affecting *P. aeruginosa* phenotypes undoubtedly exist over temporal and geographical scales, as do accumulation of adaptive mutations. However our data, and that from experimentally evolved populations, may suggest that such events occur as a result of enormous selection pressures exerted within fluctuating environments over time, and may be less commonplace in stable populations or in short term evolutionary events. Whilst it cannot be excluded that our observations are highly specific to the strain and/or patient environment we have studied, it must be noted that the populations examined are very different. Our extant diversity study highlights the levels of diversity present in a lung at any given time point. Temporal studies are more likely to identify dominant genotypes within a patient at each time point sampled, therefore identifying long term selective evolutionary events. This is even more likely if sequencing is performed on a sweep of colonies from a growth culture. As such it is possible our data set may identify loci which arise under short term evolutionary pressures as opposed to the longer term evolutionary forces exposed in temporal studies. Recent work examining the evolution of *Salmonella paratyphi*, suggested that the vast majority of mutations which occur in that pathogen population are transient, coming under positive selective pressure for a limited time-frame before being purged by negative selection due to the detrimental effect on metabolic fitness⁴⁶. Our findings suggest that future studies that aim to analyse how a strain evolves within patients by repeatedly sampling patients over time, should consider taking multiple colonies or populations per patient at each time point to better understand how the evolutionary forces currently known to impact pathogen evolution relate to the levels of extant diversity at each time point and how these two interplay. This would provide even greater resolution into how pathogen populations evolve in response to the human host.

The majority of our high-quality SNPs mapped against hypothetical proteins and none of them mapped against genes which would normally be associated with the phenotypic changes observed in our isolates. For example, the phenotypes we measured (LasA, LasB and pyocyanin) have all been shown to be highly QS-dependent^{32,33}. Furthermore, *lasR* mutants have been shown to accumulate in the CF lung¹⁰. Despite this, none of our SNPs mapped against any known genes thought to be involved in QS. When we performed statistical correlations between recombining regions and observed phenotypes, we also found regions significantly associated with phenotypic variation in genetic loci not classically associated with such phenotypes. Our data set highlights the intricate complexity of genetic mutations and their knock on effects on microbial phenotypes. It also suggests that whilst classical bacterial genetic experiments accurately identify loci involved in colonization, persistence and virulence in culture and in model systems, the mutations that underlie phenotypic alteration and adaptation to the human environment may be far more subtle and likely involve genetic loci whose function we do not yet fully understand. Very recent work in *Salmonella typhimurium* has demonstrated that bi-stable expression of virulence factors within an infecting population leads to the formation of other beneficial phenotypic changes in sub-populations, in that instance resistance to antimicrobials⁴⁷. It was also shown that bet-hedging of heterogeneous expression of virulence factors in a population can protect

clonal populations against mutants which subvert labour division. This draws comparisons to our observations in that bet-hedging and labour division can arise through trait variations and subsequent complex interactions between them, exposing degrees of functional complexity of phenotypic variation previously unreported.

Our study also has implications for understanding infection transmission networks. Our SNP distance matrix shows that the patient was likely infected with a clone very similar to LESB58 and that this has diversified in the lung via recombination. This has a potential impact on how to accurately inform contact transmission networks, with many published studies using whole genome sequencing (WGS) to inform outbreak transmission events using a cut-off of <5 SNPs⁴⁸, or in the exemplar case of MRSA, 11 SNPs²¹. Our data suggests that baseline levels of within-patient diversity for pathogens, infectious disease and transmission scenarios may need to be developed for meaningful WGS transmission studies. Otherwise multiple colonies or whole populations may need to be sequenced from every sample during transmission studies, to prevent transmission events from being misclassified.

Finally our work has implications for antibiotic testing. Susceptibility testing of bacterial isolates to antibiotics plays an important role within the diagnostic laboratory and helps to guide antibiotic choice for the treating clinician, and high levels of *P. aeruginosa* diversity within a patient has potential implications for routine antibiotic susceptibility testing. The validity of susceptibility testing has been challenged by several studies. Hogardt and colleagues demonstrated susceptibility results “close to random” when testing quinolones and tobramycin resistance in a selection of *P. aeruginosa* CF isolates⁴⁹, whilst a study conducted in 1994 by Morlin compared individual morphotypes and mixed morphotypes, finding that when predicting resistance to a panel of antibiotics, the assay was far less accurate⁵⁰. In the U.K., the current standard laboratory protocol for the testing of CF sputa relies on the selection of 1 or 2 colonies of different morphotypes in combination with the BSAC disk diffusion method to examine resistance to chosen CF therapeutics. Previous studies have shown that there is little correlation between antibiotic sensitivity patterns of *P. aeruginosa* isolated during a pulmonary exacerbation with the clinical response of the patient^{4,25} and reproducibility of the diagnostic test itself has been shown to be low between sample replicates^{26,27}.

We have built on this previous work and shown that isolates from the same patient display considerable variation in susceptibilities across a broad range of commonly used therapeutics. Pairwise correlation analysis showed that if an individual isolate was resistant to one antibiotic, it was also likely to be resistant to another. Correlations in resistance were also found between antibiotics with similar mechanisms of action. We then mixed all 44 isolates as a population and repeated the antibiotic susceptibility testing in an effort to compare the level of antibiotic resistance between individual isolates and when mixed as a population of isolates. We found that variation in the zone of inhibition was reduced in mixed versus single isolates, indicating that picking colonies at random and mixing them gives a more consistent measure of antibiotic resistance. Furthermore, we found that mixing all 44 isolates together resulted in a level of antibiotic resistance that was higher than any individual isolate tested.

Overall, the literature now highlights the current inadequacies of performing antibiotic susceptibility tests with single isolates or morphotypes to guide clinical decisions regarding antibiotic choice in chronic *P. aeruginosa* infections in CF. It suggests that antibiotic sensitivity testing could be improved by testing multiple colonies or populations of *P. aeruginosa*, and that this may lead to better clinical outcomes in the future. Although ultimately this study provides detail about the mechanisms of *P. aeruginosa* evolution in the CF lung, our findings may also have broader implications for chronic *P. aeruginosa* infections in other clinical settings such as non-CF



bronchiectasis, chronic obstructive pulmonary disease and burns patients.

Methods

Bacterial strains and culture conditions. PAO1 was used as a control strain for phenotypic assays. All clinical isolates were provided in collaboration with Nottingham University Hospitals NHS Trust. Spontaneous sputum samples were obtained from a 19 year old female patient with CF, chronically infected with *P. aeruginosa* for 3 years. At the time of sampling, the patient was clinical stable and was prescribed maintenance oral and nebulised therapies only. Sputum samples were treated with the mucolytic agent Sputasol®. Equal volumes of Sputasol (Oxoid) to sputum were added and incubated at 37°C, 250 rpm for 20–25 minutes until microcolonies were suitably broken down. Serial dilutions in 900 µl of Phosphate Buffered Saline (PBS) (Oxoid) were made from a starting concentration of 900 µl PBS and 100 µl microcolony sample 10⁻¹ through to 10⁻¹⁰. For each dilution, 100 µl was plated onto selective *Pseudomonas* Isolation Agar (PIA) and spread evenly using a sterile plastic spreader. One duplicate of each concentration per isolate was performed. For growth, pyocyanin, LasA and LasB assays and QS signal molecule extractions, pre-cultures grown from a single colony were incubated overnight in liquid Lysogeny Broth (LB) at 37°C, which were used to inoculate fresh media as described below.

Determination of final growth. Overnight pre-cultures were centrifuged at 10000 rpm for 5 min, then washed and resuspended in 5 ml fresh LB. Growth values were taken for each isolate (OD₆₀₀) and used to calculate the inoculation volume for a starting OD₆₀₀ of 0.05 in 50 ml LB. Cultures were then incubated with shaking at 37°C for 18 h. Growth was measured by OD₆₀₀. Cultures were then centrifuged at 10000 rpm for 5 min and 40 ml of supernatant was filter sterilised using 0.2 µm filters and used for phenotypic analysis.

LasA protease assay. LasA protease activity was determined by assessing the ability of *P. aeruginosa* culture supernatants to lyse boiled *Staphylococcus aureus* cells⁵¹. Overnight cultures of *S. aureus* (RN6390B) were boiled for 10 min and then centrifuged at 4000 rpm for 10 min. The resulting pellet was re-suspended in 0.02 M Tris-HCl (pH 8.5) to an optical density of 0.8 at 600 nm. 190 µL of *S. aureus* suspension was added to 8 replicate wells of a 96 well plate (Greiner). 10 µL of filter sterilised *P. aeruginosa* supernatant was then added to each well. Optical density at 600 nm (A₆₀₀) was measured at 10 min intervals for 60 min using a microplate reader (Tecan Infinite® 200). LasA activity was determined as the percentage of the initial A₆₀₀ value remaining after 60 min.

LasB protease assay. The elastolytic (LasB) activity of *P. aeruginosa* supernatants was determined using the elastin Congo red (ECR, Sigma) assay⁵². A 100 µl aliquot of filter-sterilized bacterial supernatant was added to 900 µl ECR buffer (100 mM Tris, 1 mM CaCl₂, pH 7.5) containing 20 mg ECR and incubated with shaking at 37°C for 3 h at 200 rpm. Insoluble ECR was removed by centrifugation and the absorption of the supernatant measured at 495 nm. Buffer containing ECR with and without PAO1 supernatant were used as positive and negative controls respectively. The production of LasB/cell is expressed as the OD₄₉₅ divided by the final growth (OD₆₀₀) previously recorded after 18 h incubation.

Pyocyanin extraction assay. Pyocyanin was extracted and quantified using a previously described assay⁵³. 3 ml of chloroform was added to 5 ml of sterile supernatant and vortexed for 2 min. This was then centrifuged at 10000 rpm for 5 min. The bottom chloroform layer was removed and transferred into a new 50 ml falcon tube to which 2 ml of HCl (0.2 M) was added, mixed and centrifuged. The top 1 ml layer was transferred to a cuvette and the optical density read at 520 nm. Pyocyanin production for each clinical isolate was measured against PAO1.

QS signal molecule extraction. The method used was based on a previously described assay for the extraction of QS signal molecules⁵⁴. A single colony of each isolate and PAO1 was used to inoculate 10 ml LB and incubated at 37°C at 200 rpm for 18 h. After incubation, the OD₆₀₀ was measured for each sample, centrifuged at 13000 rpm for 5 min and 5 ml of supernatant was filter sterilised. 6 ml of acidified ethyl acetate (0.1% glacial acetic acid) was added to the supernatant, vortexed thoroughly and centrifuged at 10000 rpm for 5 min. 5 ml from the separated top layer was transferred into a glass vial. This was repeated a further two times, and samples pooled into the glass vial after each step. Samples were then dried using a rotary evaporator. 1.5 ml of methanol was added to dried fractions and transferred to a 1.5 ml tube. The sample was dried again under nitrogen gas and re-suspended in 250 µl methanol and stored at -80°C until LCMS analysis. LCMS analysis was performed as previously described⁵⁴. The method was performed in triplicate for all 44 isolates, PAO1 and appropriate media controls.

Determination of antibiotic susceptibilities using the BSAC method. Bacterial suspensions were made for each isolate by inoculating 1 ml of sterile distilled water (SDW) with a single colony in a 1.5 ml sterile tube to a 0.5 McFarland standard and vortexed gently. Using a sterile cotton swab, each suspension was streaked onto Isosensitest agar (VWR laboratories) twice to create a bacterial lawn. Antibiotic disks were then placed onto the agar using a standardised applicator (Oxoid) (Amikacin 30 µg, Ceftazidime 30 µg, Ciprofloxacin 1 µg, Gentamicin 10 µg, Meropenem 10 µg, Piperacillin/Tazobactam (Tazocin®) 85 µg, Aztreonam 30 µg, Colistin 25 µg, Tobramycin 10 µg and Chloramphenicol 10 µg (Oxoid)). Plates were then inverted and incubated at 37°C for 24 h. The diameter of the zones of inhibition (mm) were

measured and compared to pre-determined breakpoints provided by BSAC, with each isolate recorded as either sensitive or resistant to each of the chosen antimicrobials. This experiment was performed in triplicate. All 44 isolates were then mixed as a population in a 1 ml suspension with SDW. Individual colonies per isolate were sampled at random from LB agar streaked with glycerol stocks of the chosen isolate. The population was then diluted to a 0.5 McFarland standard, and the experiment repeated as previously described 8 times. The Nottingham University NHS Trust *P. aeruginosa* control strain (NCTC) was used alongside PAO1 as a control throughout.

Antibiotic resistance analysis. The mm diameter measurements of clearance zones were first squared so that values were proportional to the area of the clearance zone. To obtain a score of resistance the squared clearance zones were mean standardised within each antibiotic and the negative mean standardised values were taken to be resistance. In this way we generated antibiotic sensitivity scores that could be compared across isolates and across antibiotics. Pairwise correlations were performed using Spearman's rank correlation revealing 12 significantly positively correlated sensitivity profiles (9 remained significant after correction for multiple testing using false discovery rate, as described above). Multi-drug resistance was calculated as the mean resistance score across all antibiotics.

Genome sequencing and analysis. Genomic DNA was prepared from 14 h cultures of isolates SEDI-22. Multiplexed, 150 bp Paired-end sequencing was performed on the Illumina HiSeq2000 platform to an average of 50× coverage. De novo alignments were performed using Velvet and the assemblies optimised using the PAGIT suite of programmes⁵⁵. Assembled genomes were annotated using Prokka. Comparative genomics were performed by pairwise Blast analysis using BRIG⁵⁶ and by Progressive Mauve genome alignments⁵⁷. A pan-genome of the population was created using LS-BSR⁵⁸ and plotted using ggplot2 in R. Breakpoint analysis using paired end reads was performed to detect inversions and rearrangements using BreakDancer⁵⁹, and the resulting alignments visualised with EasyFig⁶⁰. Core genome alignments were performed using Mugsy as previously described⁶¹. SNP typing was performed using SMALT and Samtools against the LESB58 reference genome, with SNPs in IS, transposons and Phages removed. High fidelity SNPs were then called using a cut off of minimum allele frequency of 0.75, minimum quality score 30, and minimum depth of 8 as described previously⁶². Maximum likelihood phylogenies were created with RaxML implementing the GTRGamma substitution model, and visualised using Figtree. Pairwise SNP distance matrices were created using Mega and visualised by ggplot2 in R. Positive selection on homoplasic SNPs was determined by aligning genes containing SNPs in Mega, and testing for selection using the codon-based Z test for selection. All raw sequence data has been deposited to the European Nucleotide Archive (Accession number: PRJEB5764), and a full set of mapping and assembly statistics are supplied in Supplementary Table S3.

Statistical association of recombination fragments with phenotypic changes. To account for the linkage of SNPs and to avoid overly conservative results due to simplistic multiple testing correction such as the Bonferroni method, we used the standard approach of permutation association tests, where the isolates are retained while the phenotypic labels of the individuals are randomly permuted⁶³. The test procedure for each phenotype was based on 10,000 permutations and the standard ANOVA test statistic where the null hypothesis of no association is compared to a general alternative⁶³ using the functions available in the Statistics Toolbox of MATLAB software (v R2012a). As noted in ref. 23, permutation tests are generally considered the gold standard in multiple testing adjustment in genetic association studies, in order to sacrifice as little power as possible, in contrast to for example Bonferroni type correction. SNPs with the permutation-based p-value < 0.05 were defined as significantly associated with the phenotype under consideration. The entire raw data set of p values for each SNP is presented in Supplementary Table S2.

Recombination analysis. The BratNextGen software⁶⁴ was used to detect recombination events in the 22 genomes, similar to several recent discoveries made using this method^{61,65}. The whole-genome alignment produced with Mugsy was used as input to BratNextGen with the default settings similar to ref. 64 with 20 iterations of the estimation algorithm. The convergence was assessed to be sufficient since changes in the hidden Markov model parameters were negligible over approximately the last 50% of the iterations. Significance of each recombining region was determined as in ref. 64 using a permutation test with 100 permutations executed in parallel on a cluster computer (threshold of 5% was used to determine significance for each recombination).

Principal component analysis. Principal component analysis (PCA) was conducted on QS signal and phenotype data using R 2.14.0 ('R' development core team, 2011) and the *FactoMineR* package. A single PCA on combined signals + virulence factor data was also performed, but this explained less variation than either of the two separate analyses; we therefore present the separate analyses as the more parsimonious approach. Pairwise Spearman's rank correlations were corrected for multiple comparisons using a false discovery rate method⁶⁶.

1. Sadikot, R. T., Blackwell, T. S., Christman, J. W. & Prince, A. S. Pathogen-host interactions in *Pseudomonas aeruginosa* pneumonia. *Am. J. Res. Care Med.* **171**, 1209–1223 (2005).



2. Govan, J. R. & Deretic, V. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* **60**, 539–574 (1996).
3. Hart, C. A. & Winstanley, C. Persistent and aggressive bacteria in the lungs of cystic fibrosis children. *Brit. Med. Bull.* **61**, 81–96 (2002).
4. Goss, C. H., Mayer-Hamblett, N., Kronmal, R. A., Williams, J. & Ramsey, B. W. Laboratory parameter profiles among patients with cystic fibrosis. *J. Cyst. Fibros.* **6**, 117–123 (2007).
5. Fothergill, J. L. *et al.* Widespread pyocyanin over-production among isolates of a cystic fibrosis epidemic strain. *BMC Microbiol.* **7**, 45 (2007).
6. Fothergill, J. L., Walshaw, M. J. & Winstanley, C. Transmissible strains of *Pseudomonas aeruginosa* in cystic fibrosis lung infections. *Euro. Resp. J.* **40**, 227–238 (2012).
7. Jones, A. M. *et al.* Spread of a multiresistant strain of *Pseudomonas aeruginosa* in an adult cystic fibrosis clinic. *Lancet* **358**, 557–558 (2001).
8. Bragonzi, A. *et al.* *Pseudomonas aeruginosa* microevolution during cystic fibrosis lung infection establishes clones with adapted virulence. *Am. J. Res. Crit. Care Med.* **180**, 138–145 (2009).
9. Mowat, E. *et al.* *Pseudomonas aeruginosa* population diversity and turnover in cystic fibrosis chronic infections. *Am. J. Res. Crit. Care Med.* **183**, 1674–1679 (2011).
10. Smith, E. E. *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. USA* **103**, 8487–8492 (2006).
11. Huse, H. K. *et al.* Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations in vivo. *mBio* **1**, no. 4 e00199–10 (2010).
12. Wilder, C. N., Allada, G. & Schuster, M. Instantaneous within-patient diversity of *Pseudomonas aeruginosa* quorum-sensing populations from cystic fibrosis lung infections. *Infect. Immun.* **77**, 5631–5639 (2009).
13. Chung, J. C. *et al.* Genomic variation among contemporary *Pseudomonas aeruginosa* isolates from chronically infected cystic fibrosis patients. *J. Bacteriol.* **194**, 4857–4866 (2012).
14. Workentine, M. L. *et al.* Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a cystic fibrosis patient. *PLoS one* **8**, e60225 (2013).
15. Jiricny, N. *et al.* Loss of social behaviours in populations of *Pseudomonas aeruginosa* infecting lungs of patients with cystic fibrosis. *PLoS one* **9**, e83124 (2014).
16. Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S. & Jelsbak, L. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environ. Microbiol.* **14**, 2200–2211 (2012).
17. Oliver, A., Canton, R., Campo, P., Baquero, F. & Blazquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**, 1251–1254 (2000).
18. Collier, D. N. *et al.* A bacterial cell to cell signal in the lungs of cystic fibrosis patients. *FEMS Microbiol. Lett.* **215**, 41–46 (2002).
19. Middleton, B. *et al.* Direct detection of N-acylhomoserine lactones in cystic fibrosis sputum. *FEMS Microbiol. Lett.* **207**, 1–7 (2002).
20. Salunkhe, P. *et al.* A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *Journal of bacteriology* **187**, 4908–4920 (2005).
21. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Inf. Dis.* **13**, 130–136 (2013).
22. Bryant, J. M. *et al.* Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet* **381**, 1551–1560 (2013).
23. Gao, X., Becker, L. C., Becker, D. M., Starmer, J. D. & Province, M. A. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* **34**, 100–105 (2010).
24. Hurley, M. N., Ariff, A. H., Bertenshaw, C., Bhatt, J. & Smyth, A. R. Results of antibiotic susceptibility testing do not influence clinical outcome in children with cystic fibrosis. *J. Cyst. Fibros.* **11**, 288–292 (2012).
25. Foweraker, J. E. & Govan, J. R. Antibiotic susceptibility testing in early and chronic respiratory infections with *Pseudomonas aeruginosa*. *J. Cyst. Fibros.* **12**, 302 (2013).
26. Foweraker, J. Recent advances in the microbiology of respiratory tract infection in cystic fibrosis. *Brit. Med. Bull.* **89**, 93–110 (2009).
27. Foweraker, J. E., Laughton, C. R., Brown, D. F. & Bilton, D. Phenotypic variability of *Pseudomonas aeruginosa* in sputa from patients with acute infective exacerbation of cystic fibrosis and its impact on the validity of antimicrobial susceptibility testing. *J. Anti. Chemo.* **55** (2005).
28. Chewapreecha, C. *et al.* Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genetics* **10**, e1004547 (2014).
29. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature Genetics* **46**, 279–286, 8 (2014).
30. Kenna, D. T. *et al.* Hypermutability in environmental *Pseudomonas aeruginosa* and in populations causing pulmonary infection in individuals with cystic fibrosis. *Microbiology* **153**, 1852–1859 (2007).
31. 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).
32. Schuster, M., Sexton, D. J., Diggle, S. P. & Greenberg, E. P. Acyl-homoserine lactone quorum sensing: from evolution to application. *Ann. Rev. Microbiol.* **67**, 43–63 (2013).
33. Williams, P. & Camara, M. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. *Curr. Opin. Microbiol.* **12**, 182–191 (2009).
34. Holloway, B. W. Genetic recombination in *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **13**, 572–581 (1955).
35. Alizon, S., Luciani, F. & Regoes, R. R. Epidemiological and clinical consequences of within-host evolution. *Trends Microbiol.* **19**, 24–32 (2011).
36. Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85** (Pt 2), 411–426 (1982).
37. 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).
38. Dettman, J. R., Rodrigue, N., Aaron, S. D. & Kassen, R. Evolutionary genomics of epidemic and nonepidemic strains of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **110** (2013).
39. 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 Genetics* **9**, e1003741 (2013).
40. Warren, A. E. *et al.* Genotypic and phenotypic variation in *Pseudomonas aeruginosa* reveals signatures of secondary infection and mutator activity in certain cystic fibrosis patients with chronic lung infections. *Infect. Immun.* **79**, 4802–4818 (2011).
41. Klockgether, J. *et al.* Intracolon diversity of the *Pseudomonas aeruginosa* cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost identical genome sequence. *Environ. Microbiol.* **15**, 191–210 (2013).
42. Heilbron, K., Toll-Riera, M., Kojadinovic, M. & MacLean, R. C. Fitness Is Strongly Influenced by Rare Mutations of Large Effect in a Microbial Mutation Accumulation Experiment. *Genetics* **197**, 981–990 (2014).
43. 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).
44. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
45. Lieberman, T. D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature Genetics* **46**, 82–87 (2014).
46. Zhou, Z. *et al.* Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc. Natl. Acad. Sci. USA* **111**, 12199–12204 (2014).
47. Arnoldini, M. *et al.* Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation. *PLoS Biology* **12**, e1001928 (2014).
48. Walker, T. M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Inf. Dis.* **13**, 137–146 (2013).
49. Hogardt, M., Ulrich, J., Riehn-Kopp, H. & Tummeler, B. EuroCareCF quality assessment of diagnostic microbiology of cystic fibrosis isolates. *J. Clin. Microbiol.* **47**, 3435–3438 (2009).
50. Morlin, G. L., Hedges, D. L., Smith, A. L. & Burns, J. L. Accuracy and cost of antibiotic susceptibility testing of mixed morphotypes of *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* **32**, 1027–1030 (1994).
51. Kessler, E., Safrin, M., Olson, J. C. & Ohman, D. E. Secreted LasA of *Pseudomonas aeruginosa* is a staphylolytic protease. *J. Biol. Chem.* **268**, 7503–7508 (1993).
52. Ohman, D. E., Cryz, S. J. & Iglewski, B. H. Isolation and characterization of *Pseudomonas aeruginosa* PAO mutant that produces altered elastase. *J. Bacteriol.* **142**, 836–842 (1980).
53. Essar, D. W., Eberly, L., Hadero, A. & Crawford, I. P. Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. *J. Bacteriol.* **172**, 884–900 (1990).
54. Ortori, C. A. *et al.* Comprehensive profiling of N-acylhomoserine lactones produced by *Yersinia pseudotuberculosis* using liquid chromatography coupled to hybrid quadrupole-linear ion trap mass spectrometry. *Anal. Bioanal. Chem.* **387**, 497–511 (2007).
55. Swain, M. T. *et al.* A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nature Protocols* **7**, 1260–1284 (2012).
56. Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402 (2011).
57. Darling, A. E., Mau, B. & Perna, N. T. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS one* **5**, e11147 (2010).
58. 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).



59. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6**, 677–681 (2009).
60. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
61. McNally, A., Cheng, L., Harris, S. R. & Corander, J. The evolutionary path to extraintestinal pathogenic, drug-resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. *Gen. Biol. Evol.* **5**, 699–710 (2013).
62. McNally, A. *et al.* Genomic analysis of extra-intestinal pathogenic *Escherichia coli* urosepsis. *Clin. Microbiol. Inf.* **19**, E328–334 (2013).
63. Balding, D. J. A tutorial on statistical methods for population association studies. *Nat. Rev. Genetics* **7**, 781–791 (2006).
64. Marttinen, P. *et al.* Detection of recombination events in bacterial genomes from large population samples. *Nucleic acids Res.* **40**, e6 (2012).
65. Castillo-Ramirez, S. *et al.* Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Gen. Biol.* **13**, R126 (2012).
66. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Series B* **57**, 289–300 (1995).

Acknowledgments

We thank James Gurney and Catherine Ortori for assistance with extraction of QS signal molecules and LCMS analysis, Mathew Diggle, Katrina Levi and Zoie Aiken for advice and assistance with RAPD PCR and VNTR typing of all isolates. We also acknowledge Roman Popat for help with analysing antibiotic data. We thank the Wellcome Trust Biomedical Informatics Hub, Exeter Sequencing Service for the production of genomic sequencing libraries. Karen Moore prepared the genomics material for sequencing and is funded by the Wellcome Trust Institutional Strategic Support Award (WT097835MF). We would also like to thank the attendees of the Permafrost 2014 workshop for their stimulating discussions and suggestions. SED was funded by an MRC PhD grant. SPD was funded by the Royal

Society, HFSP (RGY0081/2012) and NERC (NE/J007064/1). JC was funded by ERC grant (no. 239784) and AoF grant (no. 251170). We gratefully acknowledge all funders.

Author contributions

Conceived and designed the experiments: S.P.D., S.E.D. and A.M. Performed the experiments: S.E.D. Analysed the data: S.P.D., S.E.D., A.M. and F.H. Contributed reagents/materials/strains/analysis tools: J.C., H.L.B., A.M., K.P., S.A.C., S.H. and A.F. Wrote the paper: S.P.D., S.E.D., A.M. and F.H.

Additional information

Ethics Statement. Spontaneous sputum samples were obtained from a 19 year old female patient with CF, chronically infected with *P. aeruginosa* for 3 years, participating in a concurrent observational study (MRec: 09/H0407/11) and written informed consent was obtained. At the time of sampling, the patient was clinical stable and was prescribed maintenance oral and nebulised therapies only. The board which approved the study was the North Nottinghamshire Ethics Committee. The methods were carried out in accordance with the approved guidelines for this journal.

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: 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; DOI:10.1038/srep07649 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>