6. Within-patient evolution of *Mycobacterium abscessus*

This work forms the basis of a manuscript in preparation:

J. M. Bryant, D. M. Grogono, U. Hill, C. S. Haworth, J. Foweraker, J. Parkhill and R. A. Floto . Tracking genetic diversity and evolution of *Mycobacterium abscessus* during chronic infection using whole genome sequencing.

Statement of contribution

I carried out all bioinformatic analyses. DMG carried out collection of clinical data, antibiotic susceptibility testing and colony phenotyping. Study was supervised by JP and RAP. All authors contributed to interpretation of the data.

6.1. Introduction

Infections of cystic fibrosis patients are often extremely chronic and difficult to treat, sometimes leaving patients permanently colonised for the rest of their lives. These chronic respiratory infections present an evolutionary scenario where the length of colonisation allows the development of a high level of within-patient diversity and the opportunity to adapt to the cystic fibrosis lung niche. This has been well documented for *P. aeruginosa* and *Burkholderia dolosa*, where processes such as colony morphology switches (Govan and Deretic 1996), hypermutation (Oliver, Canton *et al*. 2000), selective sweeps (Lieberman, Flett *et al*. 2014) and parallel genetic mutations (Lieberman, Michel *et al*. 2011) between patients have been documented. However, how the progression of this diversity over time correlates with clinical phenotype or treatment is yet to be fully quantified and has only been investigated in a small number of patients (Workentine, Sibley *et al*. 2013). In particular, the within-patient diversity of *M. abscessus* is largely unexplored.

Like most cystic fibrosis pathogens, *M. abscessus* is thought to be primarily an environmental bacterium, which means upon entering the human host, genetic and transcriptional changes may be required to allow the bacteria to thrive in the new niche. The most apparent evidence for this process is the spontaneous transformation from a smooth to rough colony morphology, with the latter considered more virulent and able to thrive within the human host. Other than one small scale study (Kreutzfeldt, McAdam *et al*. 2013), the within-host variation of *M. abscessus* infections has been unexplored. Considering that drug resistance is often acquired within patients (Nessar, Cambau *et al*. 2012), and the obscure and irreproducible results of drug susceptibility testing for some antibiotics (Broda, Jebbari *et al*. 2013), it is important that this is fully understood.

Most bacterial whole genome studies to date have been based on colony-purified samples; so have focused on calling consensus SNPs as all variants should be fixed in the sample. This has meant there is little opportunity to investigate within-patient diversity through the detection of unfixed minority variants. This study aimed to capture the within-patient diversity of *M. abscessus*, utilising 151 whole genome sequences from 21 cystic fibrosis patients, described in a previously published data-

set (Chapter 5). To capture as much genetic variation as possible, the samples were not colony-purified, allowing the detection of minority variants and an estimation of the level of diversity present within the patient at each time point. This fine-scale genetic information was correlated with clinical phenotype and treatment, which provided insights into the diversity of *M. abscessus* and chronic bacterial infections in general.

6.2. Methods

6.2.1. Samples

For this study, a subsample of the sequencing data from a previously described dataset was used (Chapter 5). This was limited to patients diagnosed with cystic fibrosis only (n=22, 155 samples). One patient's data (4 samples) were excluded as they were found to have a mixed infection comprising of *M. a. massiliense* and *M. a. abscessus*. For the minority variant analysis, only patients with five or more samples collected at least seven days apart and no evidence of contamination were used resulting in a final dataset of 112 isolates from 10 patients (see appendix 9.5)

6.2.2. Detection of minority variants

Sequencing was carried out on DNA extracted from sweeps of *M. abscessus* maintaining a sample of the within-patient diversity. In order to detect minority variants, where all reads do not agree on a consensus base, additional steps to those used for normal SNP calling are required. As a first step stringent mapping was applied, where in addition to the default SMALT parameters, a minimum nucleotide identity of 0.98 was used, which avoided the mapping of reads with more than one miss-match which could be considered poor quality. The resultant variant data was then filtered for high quality minority variants and using parameters as described in Methods 8.9. Numbers of minority variants were corrected for coverage as described in Methods 8.9.

6.2.3. Phenotyping of colony variants and antibiotic susceptibility

Phenotyping was carried out by Dorothy Grogono at Papworth hospital. Isolates were streaked onto blood agar to examine colony morphology, and were recorded as either

Smooth [S], Rough [R], Rough/ Smooth [R/S] (where an isolate displayed colonies of both morphotypes), or Indeterminate [I].

Individual colonies were picked from solid media and grown up as pure cultures and antibiotic susceptibility testing was performed by serial broth microdilution (using standard CLSI methods (Institute 2011)). Plates were read at 3 - 5 days, and clarithromycin results were also read at 14 days to detect inducible macrolide resistance.

6.2.4. Detection of polymorphisms within genes associated with antibiotic resistance

A literature search was carried out in order to identify possible drug targets and genes associated with resistance to aminoglycosides, cephalosporins, quinolones, macrolides, tetracycline and carbapenems. In addition any genes annotated as beta-lactamases or penicillin binding proteins were also added to the list. The final list can be found in Appendix 9.5. All of the non-synonymous minority variants and consensus SNPs generated within patients (i.e. not shared between all samples in a patient) were crosschecked against this list.

6.2.5. Detection of polymorphisms associated with colony morphology switches

Several different approaches were taken to identify candidate polymorphisms. Firstly, all minority variants or consensus SNPs that differed between isolates of the same patient with a clear differentiation between the rough and smooth morphotypes were identified. Secondly, as heterogenous indels are difficult to call using standard methods, any homopolymers (>4 nt) occurring in the GPL loci were detected using a custom perl script. The length of the homopolymer found in each raw read was then extracted from the mapping data in order to identify minority variants conferring frameshift mutations.

6.3. Results

Overall, 107 consensus SNPs were detected within the patients (variants where over 80% of reads agree) using standard SNP calling methods. An additional 899 minority variants (where 4 or more high quality reads disagree with the consensus) were also detected, of which 630 were unique (only appear in one sample). This demonstrates that significant within-sample variation exists which can be detected using a high depth of coverage (average for this dataset 118 fold).

6.3.1. The number of minority variants can be used as a proxy for population size

When considering consensus SNPs only, the average pairwise diversity between isolates from the same patient was found to correlate positively with the length of their infection (Figure 33) This reflects an increase in diversity over time, which may represent the accumulation of variants being fixed in the population occurring at a rate of approximately 1 SNP per genome per year (derived from the slope of the linear regression model fitted). For minority variants, a positive trend was also observed, however this relationship was not significant (Figure 33).

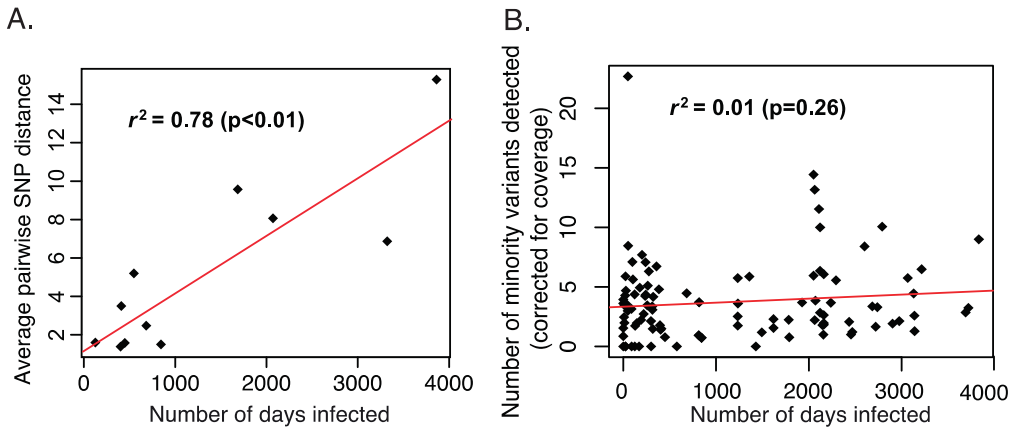

Figure 33 - Relationship between time and diversity. A) Total time since first positive *M. abscessus* sample vs. average pairwise SNP distance for each patient. Only patients with 5 or more sequenced samples available were included. B) Time vs. number of minority variants detected for each isolate. The number of minority variants detected was corrected for the depth of coverage. With the outlier removed (22 minority variants) the P value of the linear regression model reaches significance (P = 0.042).

This lack of significance could be the result of a complex array of factors affecting the bacterial population at any particular time point in the infection. These different factors are further explored in the rest of this chapter.

As evolutionary theory predicts that mutations will occur more frequently in larger populations, the observed changes in the number of minority variants over time could reflect changes in population size. In order to test this, the population size can be measured indirectly using an automated MGIT culture system (BACTEC™ MGIT™ 960), where the time taken to positivity has been shown to correlate negatively with bacterial load (Diacon, Maritz *et al*. 2012). Supporting this, the number of minority variants was found to correlative negatively with time to positivity for individual patients (Figure 34). However, isolates with a rough colony morphology type were excluded from this analysis, as they were observed to excessively clump in culture, which may have interfered with the automated detection system. In addition, when the data was combined any significant negative correlation was lost, which may be due to strain specific differences in their ability to grow in this culture system.

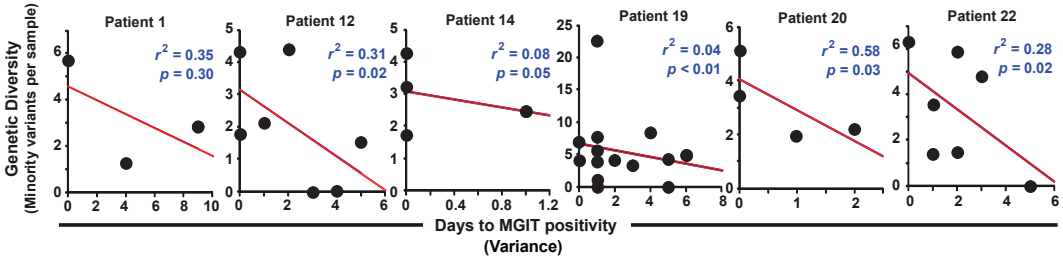

Figure 34 - Days to MGIT positivity vs. genetic diversity. The time taken for the Mycobacterial growth indicator tube to flag positive is known to correlate negatively with bacterial burden, as does the genetic diversity (number of minority variants – corrected) as shown here. Only patient's isolates with a smooth colony morphology were included.

6.3.2. Genetic diversity correlates with infection severity

For many patients, there were large fluctuations in the number of minority variants over the course of their infection. These were plotted against clinical markers of infection (C-reactive protein and Forced Expiratory Volume in 1 second) and antibiotic treatment to see if they indicted factors that might explain these changes in diversity and therefore population size (Figure 35).

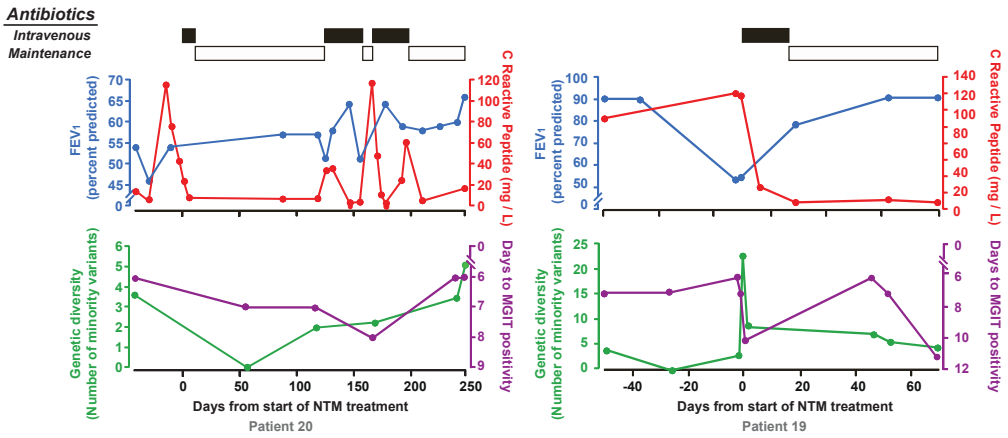

Figure 35 – Changes in genetic diversity and clinical phenotype in two patients over time. Time lines for two patients illustrating how changes in genetic diversity (tracked by the number of minority variants per isolate; green) relate to lung function (measured by FEV1; Blue), systemic inflammation (measured by C Reactive protein; red), bacterial burden (which is inversely related to days to MGIT positivity; purple) as well as maintenance (white bars) and intravenous (black bars) antibiotic therapy.

There were multiple occasions where spikes in the number of minority variants occurred at the time of an exacerbation of infection (substantiated by a rise in C-reactive peptide, fall in FEV1 or both) (Figure 35). Across all patients, CRP levels were found to correlate with the number of minority variants, presumably because both are influenced by changes in population size (Figure 36a). In addition to this correlation, it was found that successful antibiotic treatment (usually associated with administration of intravenous antibiotic therapy) was associated with decreases in the number of minority variants (Figure 36b), suggesting the occurrence of evolutionary bottlenecks. It is worth noting however that this reduction is often not maintained in subsequent samples, suggesting the population size can recover after initial exposure to the drug.

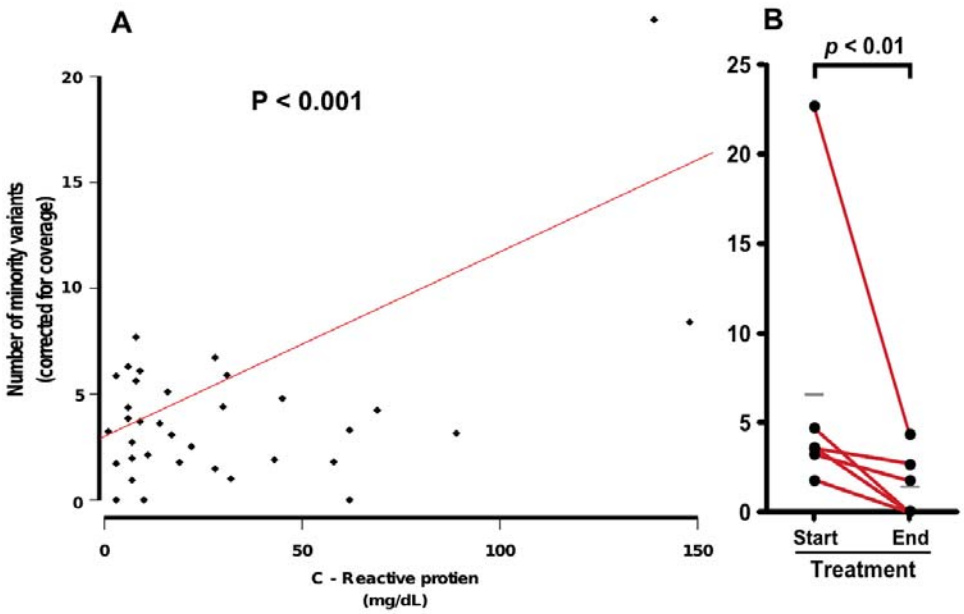

Figure 36 - Correlation between genetic diversity, CRP and treatment. A) Relationship between within sample genetic diversity (number of minority variants) and systemic inflammation of infected patients (monitored by CRP). Calculated using Poisson regression. The correlation was still significant (P < 0.001) with the outlier (MV count greater than 20) removed. B) Effect of successful antibiotic therapy (resulting in improved FEV1 and C Reactive Protein) on within sample genetic diversity. P value calculated using the paired Wilcoxon signed-rank test.

6.3.3. Population bottlenecks occur through patient-patient transmission

As described previously (Chapter 5), there was strong phylogenetic and epidemiological evidence supporting person-person transmission from patient 2 to 28. For patient 2, the number of minority variants increased over time, indicating an increase in population size. For patient 28 however, lower numbers of variants were detected, a pattern indicative of a population bottleneck that may have occurred during transmission (Figure 37).

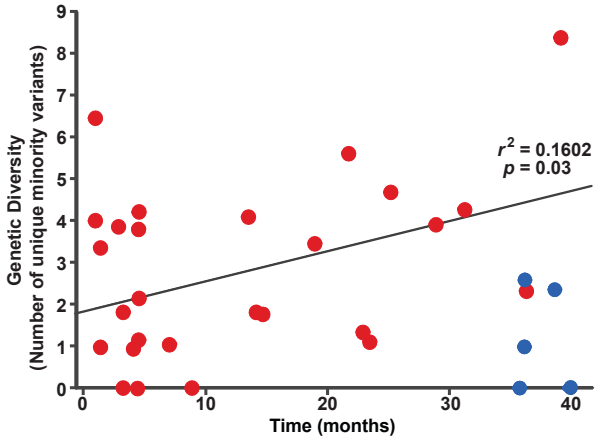

Figure 37 – Number of minority variants over time in patient 2 (red) and patient 28 (blue). Correlation coefficient and P value are only calculated based on patient 2's isolates.

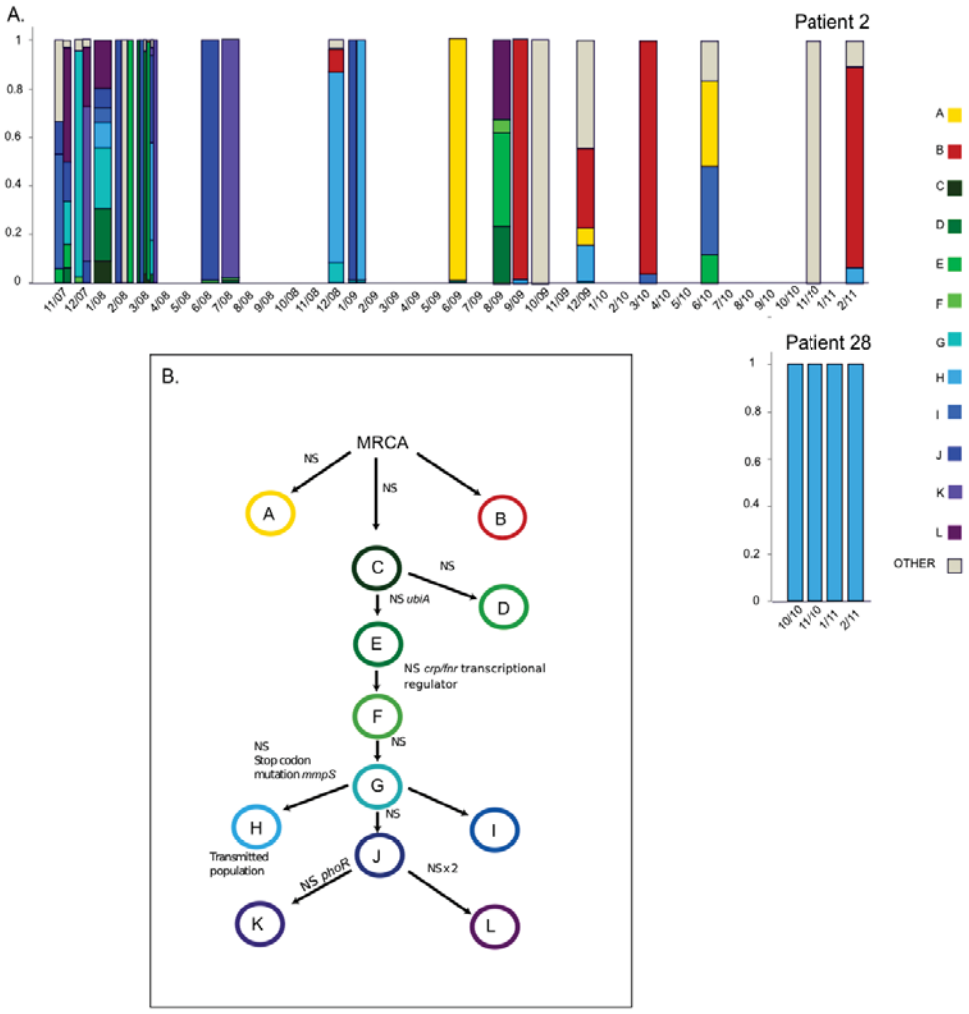

Figure 38 – Sub-clones identified in patient 2 and 28. A) Frequency of each subclone within each isolate analysed during infection of Patient 2 and transmission to Patient 28. This was calculated by determining the frequency of the "terminal" variant marking this subclone as shown in: B) Reconstruction of the lineages found in patient 2 (and transmitted to 28) demonstrates successive acquisition of non-synonymous polymorphisms (NS) by the most recent common ancestral clone (MRCA) in potential virulence genes (*ubiA*, MAB_0173; *crp/fnr*, MAB_0416c; *mmpS*, MAB_0477; *phoR*, MAB_0674).

Most minority variants detected within patients in this study were unique (only detected in one sample). However in patient 2, 17 variants were found to occur multiple times in different samples, and 14 of these co-occurred with one or more other variant at the same frequency (see Appendix 9.5), indicating linkage on the same genetic background. This suggests the presence of several sub-clones within the patient, which could perhaps be attributed to the length of the infection (exceeding 10 years). All of the variants were non-synonymous or intergenic and many were found in genes associated with virulence or drug resistance. Using the observed linkage

between these variants, hypothetical sub-lineages could be reconstructed (Figure 38b). The evolutionary succession of these lineages and the variants associated with them could be inferred if variants sometimes occurred at a lower frequency (but never higher) or at a higher frequency (but never lower) than another linked variant. Using the frequency of the "terminal" variants (frequency of variant at F, indicates frequency of clone C-E-F in absence of variant G), the frequency of the sub-lineages was tracked over time (Figure 38a). Interesting, only one of these clones was found in patient 28 (H-G-F-E-C), which again strongly suggests that a population bottleneck has occurred during transmission.

6.3.4. Parallel adaptation to the lung between patients

Parallel evolution, where traits or mutations occur independently at a frequency higher than could be explained by chance alone, has been observed between patients for other cystic fibrosis pathogens such as *B. dolosa* (Lieberman, Michel *et al*. 2011). Using a similar approach, the data was screened for the occurrence of at least two non-synonymous mutations in the same gene occurring independently in different patients. Out of a total of 107 SNPs, three genes in addition to one gene-pair (Table 7), were identified as having more than one independent non-synonymous SNP. By randomly introducing 107 non-synonymous SNPs *in silico* a 1000 times across the genome, and assuming that all positions across the genome evolve at the same rate, it was found that that this wouldn't be a pattern expected by chance alone (P= 0.001, Figure 39). When considering the 92 synonymous SNPs detected within patients, only two genes were found to have gained a SNP in two different patients, a pattern not significant under the same test. Two of the non-synonymous variants may be implicated in pathogenicity, including the PhoRP global regulator system which has been frequently recognized as important for both drug resistance and virulence in *M. tuberculosis* (Walters, Dubnau *et al*. 2006). Intriguingly, the accumulation of non-synonymous SNPs in the *phoR* gene was also recently observed in a *M. a. bolletii* infection of a cystic fibrosis patient (Kreutzfeldt, McAdam *et al*. 2013). *FolP2* is also of interest as it has been identified as a possible drug target of sulfonamides (Gengenbacher, Xu *et al*. 2008), which are sometimes used to treat *M. abscessus* infections. Mutations were also detected in the global transcriptional regulator *crp/fnr* which may have roles in oxygen or redox sensing and orchestrating bacterial stress

responses (Akhter, Yellaboina *et al*. 2008). It is particularly interesting that a gene orthologous to *crp*, associated with oxygen-related gene regulation, was also found to accumulate SNPs in *B. dolasa* (Lieberman, Michel *et al*. 2011)

Table 7 - Genes with evidence of convergent evolution between patients and clusters. Patients in parenthesis indicate the presence of a non-synonymous minority variant, with the rest being called as consensus SNPs. *One variant is found in both 31 and 22 due to result of transmission so is only counted once.

Gene	Patients	Number of SNPs (minority variants)	Encodes
MAB_0674/MAB_673	20, 15, (*31)*	3 (1)	PhoRP two component system
MAB_3675	20, 24	3	Probable succinate dehydrogenase, flavoprotein subunit SdhA
MAB_1345	7, 11, (*7, 30 x2)*	2 (3)	Probable dihydropteroate synthase 2 FolP2
MAB_0416c	4, 33 x2, 31, 22	4*	cAMP receptor protein

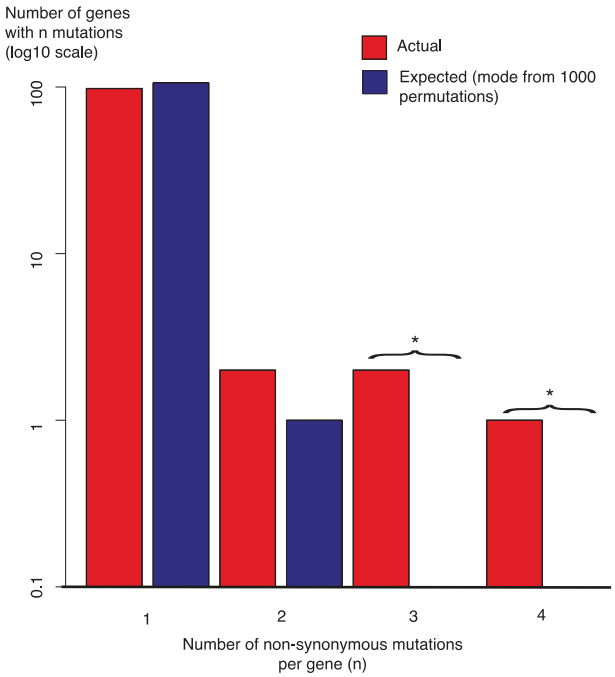

Figure 39 - Expected and actual number of independently acquired mutations per gene. To determine the expected frequency distribution of mutations within a gene, 107 non-synonymous mutations were introduced at random (the total number observed within patients in this study) *in silico* into the *M. abscessus* reference genome (Ripoll, Pasek *et al*. 2009)*;* this process was repeated a 1000 times. The average frequency of mutations per gene was counted (blue) and compared to the actual values (red) and plotted on a log scale. Four genes were found to have 3 or more non-synonymous mutations (Table 4), a pattern not expected by chance alone (stars indicate where the actual frequency

is observed in less than 5% of the 1000 permutations). It is important to note that this test is highly simplified in that it assumes that all positions across the genome mutate at the same rate.

6.3.5. Hypermutation

One patient's isolates were found to have a much higher number of minority variants, with an average of 35 per sample compared to rest of dataset with an average of 3.7 (as a result this patient was excluded from the longitudinal diversity analysis above). The excess number of these variants could not be explained by length of infection or treatment regimen. Both SNPs and minority variants were inspected manually, and a mutation conferring a premature stop codon in a uracil DNA glycosylase (*udg*) gene (MAB_3283c) was identified. *M. abscessus* has only one copy of the *udg* gene, which as part of the base excision repair pathway is responsible for removing miss-incorporated uracil from DNA. A knockout of *udg* in *Mycobacterium smegmatis* has previously been shown to confer a hypermutator phenotype with an excess of G/C->A/T transitions (Wanner, Castor *et al*. 2009). A similarly biased mutation spectrum was found for the minority variants detected in this patient (Figure 40a), supporting the inference that this mutation is having a similar effect in *M. abscessus*.

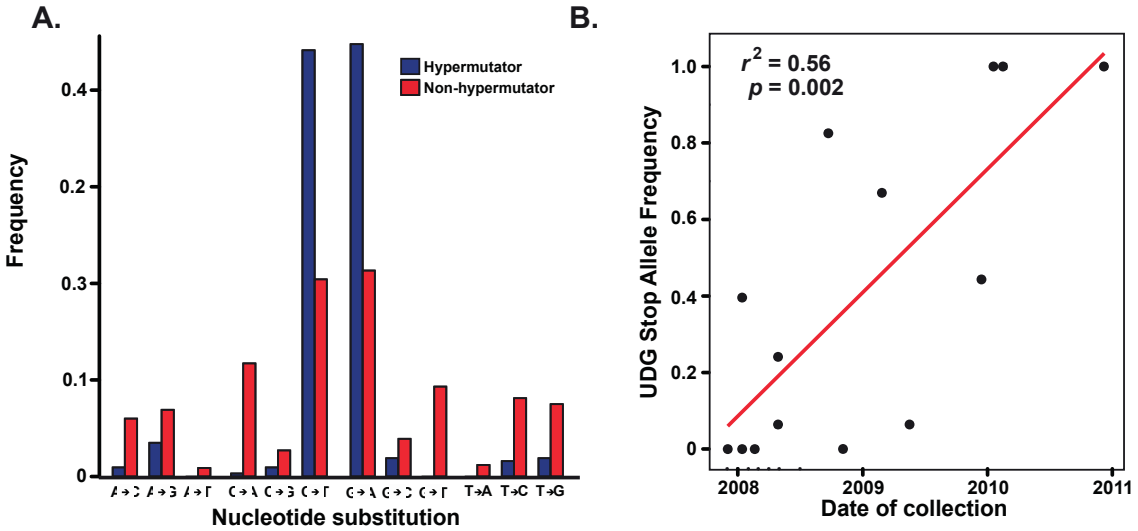

Figure 40 – Mutation frequency and spectra of the hypermutating strain. A) Mutation spectrum of minority variants from hupermutator (blue) and control (red) clones. Control frequency was obtained from patient 2 which has a similar number of isolates. The excess C to T and G to A nucleotide substitutions confirm functional loss of uracil DNA glycosylase (*udg*) in the hypermutator clone. B) Frequency of the hypermutator causing allele in the *udg* gene over time.

By using the depth of coverage that supported the *udg* mutation, the frequency of the hypermutator allele was found to increase over the course of the infection, reaching fixation in the most recent samples (Figure 40b). The evolutionary fitness of the hypermutator could not be ascribed to acquisition of another single gene mutation (as none exactly correlated in frequency with the *udg* mutator allele). However, it was striking that several genes were found to have accumulated non-synonymous SNPs, including the enhanced intracellular survival (*eis*) gene (MAB_4532c), which accumulated five mutations including a premature stop codon. This gene is considered important for virulence in *M. tuberculosis*, through suppression of host immune responses (Kim, An *et al*. 2012) and autophagy (Shin, Jeon *et al*. 2010). In addition, *nrdI* (MAB_3414c) accumulated three mutations, whose functional relevance is less clear; however this protein belongs to class 1 of ribonucleotide reductases, which in *E. coli* has been found to be activated by oxidative stress and iron limited conditions (Cotruvo and Stubbe 2008).

When considering the clinical impact of the hypermutator clone it was found that initially, infection with *M. abscessus* led to a rapid decline in lung function. However, fixation of the *udg* mutation within the patient in 2010 was associated with stabilization of both lung function and CRP with no further requirement for intravenous NTM therapy, suggesting a shift towards chronicity and an attenuation of virulence had occurred.

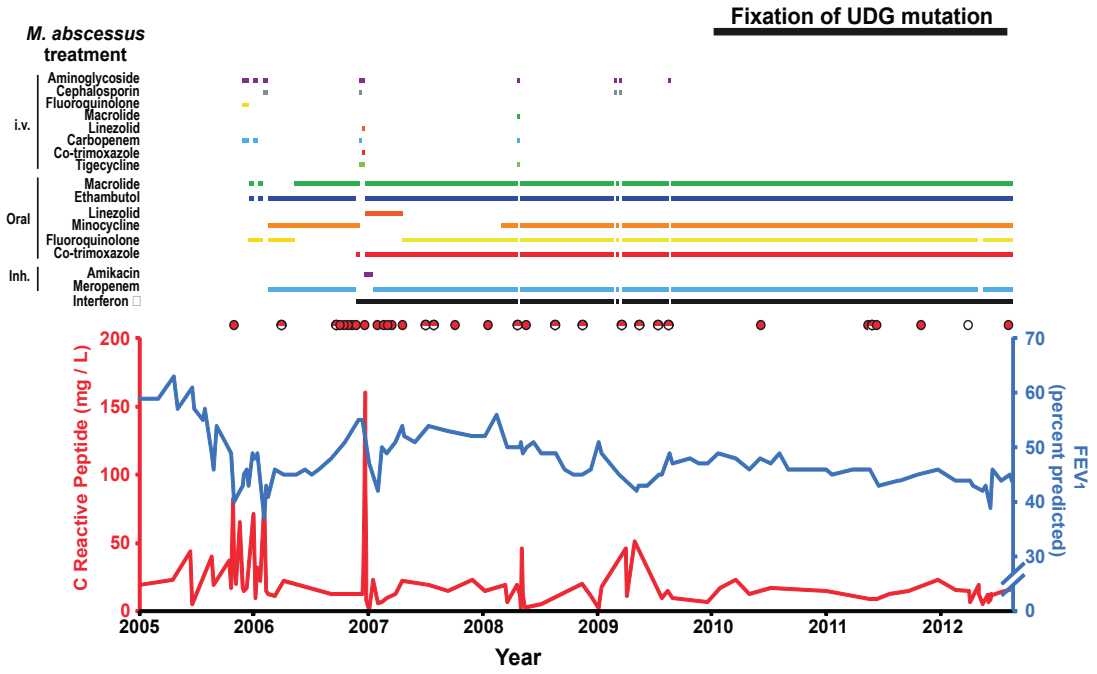

Figure 41 - Clinical trajectory of patient infected with hypermutator clone. Diagram shows changes over time in systemic inflammation (C Reactive protein; *red*), lung function (FEV1; *blue*), sputum cultures (smear negative and culture negative: *white circle*; smear negative and culture positive: *half red circle*; smear positive and culture positive: *red circle*) and intravenous (i.v.), oral and inhaled (*inh*) antibiotic therapy for *M. abscessus*. Fixation of *udg* mutation within the infecting *M. abscessus* population coincides with a more stable clinical course not requiring intravenous antibiotics

6.3.6. Functional consequences of genetic diversity

The presence of genetic heterogeneity in these samples would be expected to lead to functional heterogeneity. By evaluating 10 individual colonies from each *M. abscessus* sample using broth microdilution, considerable heterogeneity was observed in the minimum inhibitory concentration (MIC) of commonly used antibiotics (Figure 42). Corresponding to this phenotypic diversity, several polymorphisms were found in genes associated with antibiotic resistance (Table 8) including 16S rRNA (conferring aminoglycoside resistance (Prammananan, Sander *et al*. 1998)), penicillin binding proteins and beta lactamase (which may confer resistance to carbapenems and aztreonam (Yamachika, Sugihara *et al*.), (Liao and Hancock 1997), and the dihydropteroate synthase *folP2* (a potential target for sulphonamides (Gengenbacher,

Xu *et al*. 2008)). There was also substantial phenotypic heterogeneity in colony morphology (Figure 43), where switches occurred between smooth, rough and mixed morphotypes over the course of longitudinal sampling (examples in Figure 42). In some but not all cases, these switches could be associated with mutations in genes implicated in glycopeptidolipid synthesis, the pathway underlying this morphology (Table 9).

Table 8 - SNPs in antibiotic resistance loci generated within patients. Non synonymous variants in genes associated with antibiotic resistance were identified in patients isolates (the parenthesis indicates the specific sample).

Patient Sample†	Variant type	Gene function	Gene ID
1 (a)	Non-synonymous minority variant	Penicillin binding protein	MAB_3167c
22 (e)	Non-synonymous minority variant	Penicillin binding protein	MAB_3167c
1 (g)	Non-synonymous minority variant	Beta-lactamase	MAB_4947
5 *	A1408G	16s ribosomal RNA	MAB_r5051
9 (c)	A1408G	16s ribosomal RNA	MAB_r5051

†All samples, except 9 (c), were taken from patients established on treatment for *M. abscessus*.
* most patient 5 isolates have a level of heterogeneity at this position

Table 9 - SNPs in GPL loci. Non synonymous or frameshift variants were identified in the GPL loci genes (Ripoll, Deshayes *et al*. 2007) in patients where a colony morphology switch or heterogeneity was observed.

Patient	Variant type	Gene	Phenotype
19 (n)	Non-synonymous minority variant	MAB_0939	S/R
5 (m)	Frameshift insertion	MAB_4098c	R
5 (j)	Heterogeneous (30%) frameshift deletion	MAB_4099c	S/R
1	Frameshift insertion	MAB_4099	S/R

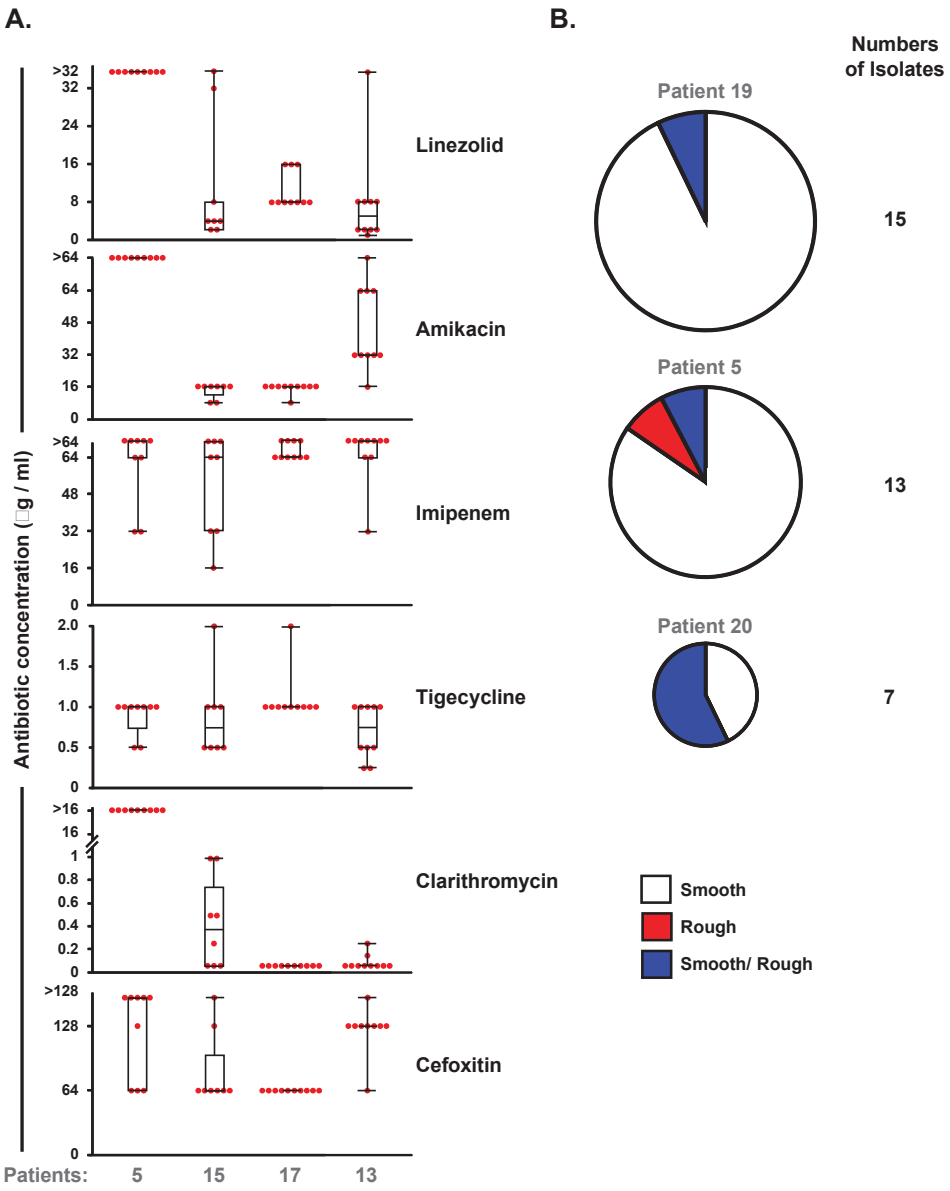

Figure 42 - Within patient phenotypic heterogeneity. A) Within sample variation in antibiotic susceptibility measured by broth microdilution in four patients and 10 colonies. B) Variation in colony morphotype of isolates from four individuals chronically infected with *M. abscessus* was established by growing samples on solid media and inspecting whether colonies were smooth (white), rough (red) or a combination of smooth and rough (blue).

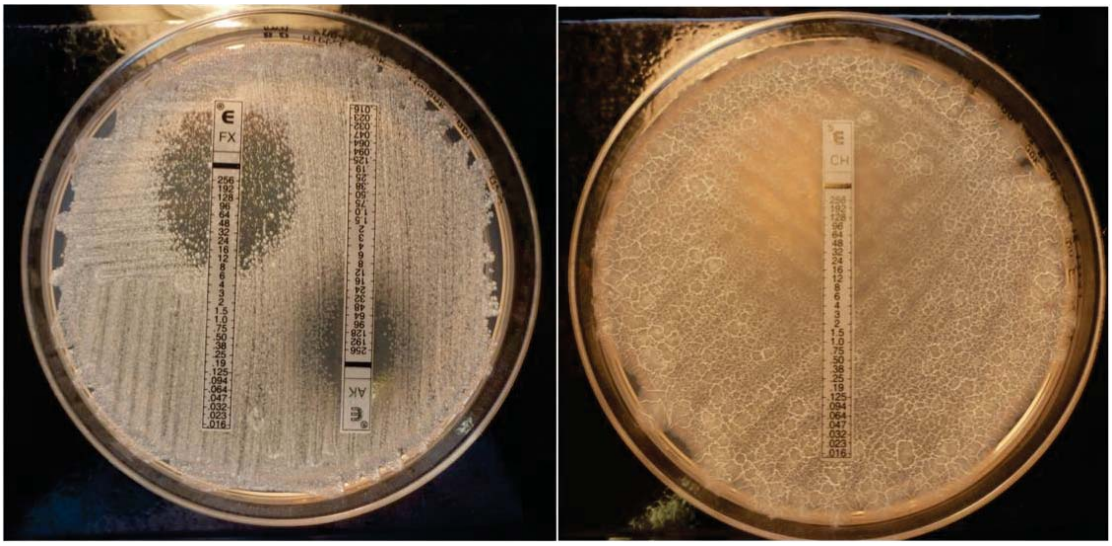

Figure 43 – Smooth (L) and rough (R) colony morphotypes of *M. abscessus*. Picture provided by Dorothy Grogono.

6.4. Discussion

Traditional genotyping and whole genome sequencing techniques of bacteria have often relied on the isolation of a single colony, and this single isolate is assumed to represent the infecting population. It is increasingly being appreciated however, that significant within patient diversity can exist across a wide range of host environments and time scales as shown for *Staphylococcus aureus* (Golubchik, Batty et al. 2013, Harris, Cartwright et al. 2013), *Helicobacter pylori* (Kennemann, Didelot *et al*. 2011) and *M. tuberculosis* (Perez-Lago, Comas *et al*. 2014). This has been more intensively explored in cystic fibrosis pathogens such as *P. aeruginosa* (Smith, Buckley et al. 2006, Chung, Becq et al. 2012) and *B. dolasa* (Lieberman, Michel et al. 2011, Lieberman, Flett et al. 2014), due to the chronicity of their infections. However, how this diversity relates to clinical phenotype, and its impact on how we understand within-patient evolution and transmission is yet to be fully explored.

As predicted by population genetics, this study provides evidence that genetic diversity correlates with population size and that this population size fluctuates over the course of infection. In asymptomatic *S. aureus* carriage, similar fluctuations have also been observed (Golubchik, Batty *et al*. 2013), which are assumed to relate to

cycles in colonization and clearance. In *M. abscessus*, these fluctuations frequently correspond to changes in the clinical phenotype with explosions in genetic diversity coinciding with infective exacerbations and conversely population bottlenecks occurring after intensification of antibiotic treatment. This is the first study, to my knowledge, which has been able to relate changes in within-host diversity of a bacterial pathogen to clinical phenotype and treatment.

A population bottleneck appeared to occur upon transmission of *M. abscessus* from one patient to another, suggesting a limited infectious dose. Although this study is limited to one transmission event, *M. abscessus* appears to be similar to *M. tuberculosis*, where even though significant within-patient diversity is generated, the heterogeneity appears to be lost upon transmission, with only one genotype being found in the recipient patient (Perez-Lago, Comas *et al*. 2014).

This study found that clones picked from a single sample can have different antibiotic profiles. The current microbiology standard is to perform broth microdilution, which can be performed on a clonal culture or a sweep. However both of these are likely to be unrepresentative of the overall population. The former provides results for a single organism, which may not be typical of the rest of the population, whilst the latter is biased towards resistant organisms which will grow even if they are present only as a minority. Antibiotic susceptibility testing results performed in this way are therefore likely to be misleading, and may explain why current results are often inconsistent with patient outcome and can be irreproducible (Broda, Jebbari *et al*. 2013).

Whilst much of this genetic diversity reflects neutral variation, some of it appears to be associated with the observed phenotypic diversity, so could have a selective advantage. In addition, parallel evolution between patients was observed, where several genes were found to accumulate non-synonymous variants independently in different patients; many of which have been associated with virulence or drug resistance in *M. tuberculosis*. In addition to this, several variants in one chronically infected individual were found to change frequency over time and showed linkage to one another. Reconstruction of haplotypes from minority variants has only previously been achievable with longer reads for short viral genomes such as HIV (for example (Zagordi, Klein *et al*. 2010)) and cancer (Fischer, Vazquez-Garcia *et al*. 2014).

However, due to the dense longitudinal sampling, this was able to be achieved with short un-linked reads by correlating them over time, a pattern previously noted for the influenza virus (Watson, Welkers *et al*. 2013). For *M. abscessus* this revealed a pattern of linked lineages that are associated with non-synonymous variants in genes associated with host adaptation or virulence. This suggests that this population structure has been selected for through a series of bottlenecks or selective sweeps. However, the lineages appear to be present across the sampling period, suggesting a lack of replacement by one lineage over another. This has previously been observed in *B. dolosa*, where several adaptive lineages have been observed to coexist over time, with a lack of fixation (Lieberman, Flett *et al*. 2014). This could be explained by geographical separation of different lineages within the lung niche, allowing local selective sweeps. This compartmentalisation has strong parallels with *M. tuberculosis*, which forms multiple discrete lesions in the lung tissue. Although it is currently not known whether *M. abscessus* forms granulomatous lesions in cystic fibrosis patients, the pattern we see in this chronically infected patient is highly suggestive of this kind of geographical structuring.

Finally, this study identified the first clinical hypermutator for a mycobacterium. The hypermutator accumulated many mutations in virulence-associated genes; but some of these resulted in a loss of function, indicating that these mutations are resulting in attenuation of *M. abscessus*. The fixation of the hypermutator was not associated with a clinical decline, indicating that the presence of the hypermutator hadn't worsened the patient's symptoms. These analyses together support a scenario where selection is occurring within patients for variants that favor long term survival and adaptation of *M. abscessus* to the cystic fibrosis lung niche, and a shift away from acute to more chronic infections. Hypermutators provide us with an opportunity to observe an accelerated version of the evolutionary trajectory of bacterial adaptation to the host, and it was found that the same genes accumulated mutations in non-hypermutator patients also. This suggests that the hypermutator's evolutionary trajectory towards attenuation may be shared by other *M. abscessus* infections, and also by other cystic fibrosis species, as this trend has been observed for *P. auriginosa* (Smith, Buckley *et al*. 2006). This, in addition to other parallels such as colony morphology switches (Govan and Deretic 1996) and convergent evolution, suggest that very different cystic

fibrosis pathogens have highly similar ways of adapting to the cystic fibrosis lung niche.