FEMS Microbiology Letters, 362, 2015, fnu037

doi: 10.1093/femsle/fnu037
Advance Access Publication Date: 4 December 2014
Research Letter

RESEARCH LETTER – Pathogens & Pathogenicity

A proportion of mutations fixed in the genomes of *in vitro* selected isogenic drug-resistant *Mycobacterium tuberculosis* mutants can be detected as minority variants in the parent culture

Indra Bergval[1,*], Francesc Coll[2], Anja Schuitema[1], Hans de Ronde[1], Kim Mallard[2], Arnab Pain[3], Ruth McNerney[2], Taane G. Clark[2] and Richard M. Anthony[1]

[1]KIT Biomedical Research, Royal Tropical Institute, Meibergdreef 39, 1105 AZ Amsterdam, Netherlands, [2]London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom and [3]King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

*Corresponding author: E-mail: i.bergval@kit.nl
One Sentence Summary: Small subpopulations of genetic mutants that become fixed after selection are visible in the original culture of a *Mycobacterium tuberculosis* laboratory strain.
Editor: David Studholme

ABSTRACT

We studied genomic variation in a previously selected collection of isogenic *Mycobacterium tuberculosis* laboratory strains subjected to one or two rounds of antibiotic selection. Whole genome sequencing analysis identified eleven single, unique mutations (four synonymous, six non-synonymous, one intergenic), in addition to drug resistance-conferring mutations, that were fixed in the genomes of six monoresistant strains. Eight loci, present as minority variants (five non-synonymous, three synonymous) in the genome of the susceptible parent strain, became fixed in the genomes of multiple daughter strains. None of these mutations are known to be involved with drug resistance. Our results confirm previously observed genomic stability for *M. tuberculosis*, although the parent strain had accumulated allelic variants at multiple locations in an antibiotic-free *in vitro* environment. It is therefore likely to assume that these so-called hitchhiking mutations were co-selected and fixed in multiple daughter strains during antibiotic selection. The presence of multiple allelic variations, accumulated under non-selective conditions, which become fixed during subsequent selective steps, deserves attention. The wider availability of 'deep' sequencing methods could help to detect multiple bacterial (sub)populations within patients with high resolution and would therefore be useful in assisting in the detailed investigation of transmission chains.

Key words: tuberculosis; MDR-TB; evolution; whole genome sequencing; mutation; drug resistance

INTRODUCTION

Drug-resistant *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), is an increasing problem that is severely complicating effective treatment of the otherwise curable disease. Multidrug-resistant TB (MDR-TB) is caused by *M. tuberculosis* bacilli that are by definition resistant to first-line drugs rifampicin and isoniazid. Resistance in *M. tuberculosis* is

Received: 7 October 2014; **Accepted:** 19 November 2014

predominantly conferred by genomic mutations and MDR strains typically carry resistance-mediating mutations in *rpoB*, conferring resistance to rifampicin, and in *katG*, conferring resistance to isoniazid (Musser 1995; Ramaswamy and Musser 1998). Resistance mutations can impair bacterial fitness which may affect survival within the host and onward transmission (Cohen, Sommers and Murray 2003; Mariam *et al.*, 2004; Gagneux *et al.*, 2006; Luciani *et al.*, 2009). Some evidence is available that resistant mutants can restore this fitness loss by the acquisition of adaptive mutations (Sherman *et al.*, 1996; Casali *et al.*, 2012, 2014; Comas *et al.*, 2012).

Mycobacterium tuberculosis is a monomorphic organism with no evidence of horizontal transfer of genetic elements and a low recombination rate (Sreevatsan *et al.*, 1997; Brosch *et al.*, 2002; Smith *et al.*, 2006, 2009; Gagneux and Small 2007; Achtman 2008); within the members of the *M. tuberculosis* complex (*M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* and *M. pinnipedii*), there is a genomic resemblance of >99%. The population structure of *M. tuberculosis* furthermore suggests that phylogenetic lineages, which can be distinguished on the basis of few genetic markers (Coll *et al.*, 2014), have co-evolved with and adapted to their respective hosts (Caws *et al.*, 2008).

Typing of *M. tuberculosis* strains is commonly achieved by one or a combination of the following methods: RFLP, CRISPR typing, VNTR/MLVA, MLST or SNP typing. The latter has been gaining popularity, as it is particularly appropriate for the classification of a monomorphic organism such as *M. tuberculosis*, is less prone to homoplasies and is compatible with other sequence-based tools, such as whole genome or deep sequencing. With the cost of next-generation sequencing methods dropping rapidly and their availability increasing, many researchers have used these tools to study the micro-evolution of *M. tuberculosis* isolates. Genomic analysis of *M. tuberculosis* isolates recovered from patients has shown that, even over longer periods of time, the genome of the pathogen remains relatively conserved (Schürch *et al.*, 2010). Likewise, after transmission or selection for antibiotic resistance, minimal genetic variability has been detected (Niemann *et al.*, 2009; Ford *et al.*, 2011, 2012; Saunders *et al.*, 2011; Casali *et al.*, 2012, 2014; Comas *et al.*, 2012; Bryant *et al.*, 2013; Clark *et al.*, 2013;), although in some cases an (initial) increased level of genetic divergence has been observed for drug resistance mutations (Mariam *et al.*, 2011).

The evolution of *M. tuberculosis* is probably most accurately elucidated by genomic analysis of sequential isolates from a single patient or a transmission chain. However, as there may be multiple mycobacterial foci within a single patient, each being subjected to severe sequential bottlenecks, multiple mutants may emerge simultaneously within the patient. Thus, sequencing single isolates may dramatically overestimate the relative prevalence and likelihood of particular mutations and underestimate the dynamics of traits under evolution or very strong selection, such as drug resistance. A recent study focusing on the emergence of quinolone resistance within the host has indeed shown that multiple resistance-associated mutations can occur inside one patient and that the distribution of these mutants can vary significantly between sequential isolates (Streicher *et al.*, 2012).

We investigated the *in vitro* evolution of a collection of well-characterized isogenic drug-resistant mutants after antibiotic selection on solid medium by whole genome sequencing. Our *in vitro* findings corroborate data from previous studies where clinical isolates under evolution show a similar degree of genomic stability (Schürch *et al.*, 2010; Ford *et al.*, 2011; Mariam *et al.*, 2011; Bryant *et al.*, 2013; Clark *et al.*, 2013). Additionally, in the majority of the strains studied we also found evidence of hitchhiking mutations that were present as minority variants in the parent strain, as well as fixation of additional mutations not associated with drug resistance. On the basis of these signature mutations a phylogenetic relationship could be inferred between all strains included in this study. We believe that this illustrates the great potential of whole genome sequencing analysis, in particular deep sequencing identifying minority variants, to aid investigations on patient-to-patient transmission of *M. tuberculosis*, particularly when clinical or epidemiological data is missing or incomplete.

METHODS

Strains

Drug-resistant mutants used in this study were all derived from a cultivar of the *M. tuberculosis* laboratory strain MTB72 (ATCC 35801) which has been propagated in our laboratory for more than 20 years and is susceptible to all commonly used anti-TB drugs. All strains are derived from experiments that have been previously described (Anthony *et al.*, 2005; Bergval *et al.*, 2007; Bergval *et al.*, 2009).

Culture and selection of mutants

Spontaneous mutants were selected in 2003/2004 [R, B and RB strains, Fig. 1 and Anthony *et al.* (2005)] or 2006 [H strains, Fig. 1 and Bergval *et al.* (2009)] from a liquid culture from MTB72, by plating an aliquot from a log-phase MTB72 culture (ca. 3 weeks, 10^8 CFU) onto antibiotic-containing solid medium. Strains R4, R46, R181 and R190 were selected with 8 μg mL^{-1} rifampicin, strains R181 and R190 in a different experiment and with a different starting culture of MTB72, but in the same month. Strain B15 was selected with 8 μg mL^{-1} rifabutin and strain B66 with 0.8 μg mL^{-1} rifabutin in two different experiments using two starting cultures from MTB72 over a period of two months. The rifampicin MIC and the rifabutin MIC was previously determined for all rifampicin-resistant strains (Anthony *et al.*, 2005); R190 showed resistant to rifampicin (32 μg mL^{-1}), but susceptible to rifabutin (<0.8 μg mL^{-1}), which allowed us to subject strain R190 to a second round of selection (Fig. 1). Strains RB14, RB15, RB16 and RB19 are spontaneous, double-resistant, mutants derived from strain R190 selected by plating an aliquot from an R190 liquid culture onto solid medium containing 2 μg mL^{-1} rifabutin.

Strains H15 (March 2006), H71 (August 2006) and H103 (November 2006) are selected by 20 μg mL^{-1} (H15 and H71) and 0.4 μg mL^{-1} (H103) isoniazid; in the case of H71, the parent culture was treated with hydrogen peroxide before an aliquot was plated on selective medium (Bergval *et al.*, 2009). The strains were selected in three different experiments, using three different starting cultures of MTB72.

Genetic analysis

For all mutants, the mutation responsible for resistance to drug(s) in question was determined via Sanger sequencing (Anthony *et al.*, 2005; Bergval *et al.*, 2009); two regions of *rpoB* were analysed for strains resistant to rifampicin and/or rifabutin for the isoniazid-resistant strains parts of the *mabA-inhA* intergenic region and of *katG*, known to contain mutants in clinical isolates, were analysed. Before sequencing the associated genes, the isoniazid-resistant strains were first assessed by an adapted multiplex ligation-dependent probe amplification

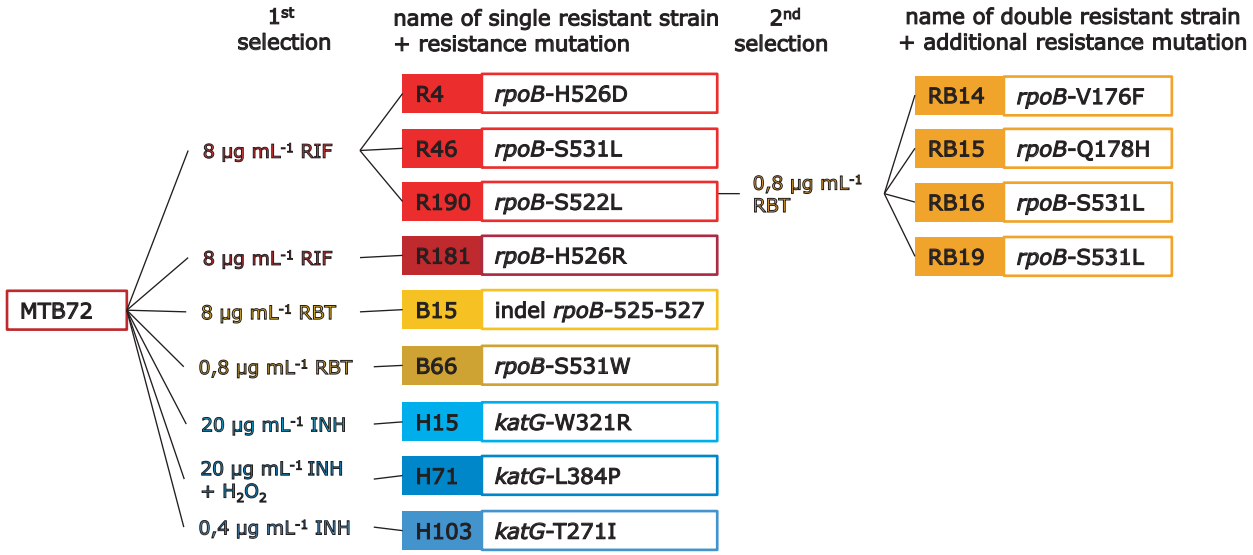

Figure 1. Relationship and selection conditions of strains used in this study. MTB72 is the parent strain to all first-step monoresistant mutants. During seven separate experiments, nine isogenic mutants were selected on solid medium containing the concentrations of antibiotic depicted on the left side of the boxes. Each color indicates a different experiment. The boxes indicate the strain name (solid color) and the identified resistance-conferring mutation it carries (blank box with colored outline). Strain R190 was subjected to a second round of selection, from which the four strains (RB14-RB19) depicted on the right are derived. The mutations indicated in these boxes are in addition to the mutation that was acquired by R190. RIF: rifampicin, RBT: rifabutin, INH: isoniazid, H_2O_2: hydrogen peroxide.

assay, to determine if a deletion in or of *katG* was present (Bergval *et al.*, 2009).

The genomes of strains MTB72 and R190 were previously analysed by comparative genome hybridization (Bergval *et al.*, 2009).

Whole genome sequencing and analysis of genetic variants

DNA isolation and sequencing

For each strain, DNA was extracted via the cetyltrimethylammonium bromide procedure (Soolingen *et al.*, 1991) from 1 mL of a logarithmic phase liquid culture (ca. 3 weeks growth) (Anthony *et al.*, 2005; Bergval *et al.*, 2009). The DNA concentration of the samples used in this study was between 53 and 142 ng mL DNA, as determined with the NanoDrop (Thermo Fisher Scientific Inc., Wilmington, Delaware, USA). The sequencing was carried out at the KAUST Genomics Facility, using Illumina HiSeq 2000 technology with paired-end reads of length 100 base pairs.

Data analysis

Trimmomatic software (Bolger, Lohse and Usadel 2014) was used to remove low-quality reads and trim low-quality 3′ ends of reads. Nucleotide positions in the raw reads with a quality score lower than Q20 were removed. High-quality reads were then mapped to the H37Rv reference genome (Camus *et al.*, 2002) (Genbank accession: AL123456.3) using the BWA-mem algorithm (Li 2014). SAMtools (Li *et al.*, 2009) and GATK (DePristo *et al.*, 2011) were used to call SNPs. SNP sites were then selected as the intersection dataset between those obtained from both programs. Polymorphic sites at non-unique regions of the genome (Derrien *et al.*, 2012) were filtered out. The depth of coverage was calculated from alignment files using a custom Perl script. For each sample, SNP genotypes were called if supported by at least 20-fold total coverage, otherwise they were classified as missing. The read depth and coverage of all variants identified in all strains included in this study are depicted in Fig. S1 (Supporting Information). Multiple genotypes were allowed by adding alleles supported by at least 5% of total number of reads. We excluded positions in the genome with more than 20% missing genotypes across all samples. In the final dataset, we retained 912 SNP sites, 83% (912/1096) of the original SNP sites derived using SAMtools and GATK. Artemis (Rutherford *et al.*, 2000) was used to verify these variants in the BAM files and to determine the read depth. All sequences have been deposited to Sequence Read Archive (http://trace.ncbi.nlm.nih.gov/Traces/sra/), NCBI's repository for high-throughput sequencing data and are publicly available under accession number SRP049364.

RESULTS

Our collection of isogenic mutants included a pansusceptible parent strain (MTB72), nine first-generation (single-step) drug-resistant daughter strains and four second-generation (two-step) daughter strains (Fig. 1). The nine monoresistant strains that were directly derived from MTB72 by a single round of antibiotic selection (R-, B- and H-series, Fig. 1) all acquired at least one additional mutation in a gene not directly associated with drug resistance (Table 1).

A total of 21 non-drug resistance mutations were detected, eight of which were identified in multiple daughter strains. Interestingly, these eight alleles present in multiple daughter strains could already be detected in the genome of the parent strain (indicated as mixed alleles in Table 1). The mutations that could be detected in the parent strain show various linkage patterns: mutations in *ctpV*, Rv1215c, *ceoB* and *Rv2693c* are shared by strains R4, R46, R181, R190 (and offspring) and B66; mutations in *dnaB*, *ispE* and *gcvB* are shared by strains R46 and B66; the mutation in *ilvA* is shared by H15, H71 and H103 (Table 1).

Eleven mutations were uniquely found in six strains and could not be detected in the genome of the parent strain: R4 (*hycQ*-T138T, *zwf1*-R136C, *murF*-A501V), R181 (*lppY*-G97R), R190 (*dnaA*-V270I, *Rv0466*-S93S, shared with offspring: *zmp1*-D637D, Rv1215c-M243I), B15 (Rv2974c-L251L), H103

Table 1. Mutations and codon changes identified in strains analysed in this study. Depicted are only those loci for which at least one of the strains included had acquired a mutation. The genomes of all strains were mapped against the H37Rv reference genome and the genomic location of each mutation therefore refers to this genome. Drug resistance-conferring mutations are depicted in Fig. 1 and are not included in this figure. Sequences for strain MTB72, the susceptible parent strain, are depicted in bold. If multiple alleles were observed above the expected error rate, the sequence of both alleles is given, with the dominant allele first. The sequence in the daughter strains was regarded as a mutant if the sequence differed from the sequence of the dominant allele. : non-synonymous mutation, [a]: the sequence of the dominant allele is different from the H37Rv sequence, whereas the sequence of the minority variant is similar to that of H37Rv, [b]: the mutation was acquired in the parent strain.

Gene/CDS	dnaA	dnaB	hycQ	zmp1	Rv0466	ctpV	ispE	zwf1	Rv1215c	ilvA	Rv1674c	wag22-Rv1760	gcvB	murF	ceoB	Rv2693c	Rv2974c	lppY	Rv3446c	crp	pks2
Genomic position	808	62690	94364	234597	556736	1078856	1130414	1244112	1358716	1763680	1899754	1993090	2077437	2417503	3009506	3011060	3330609	3357890	3863811	4116901	4298071
Codon	270	765	138	637	93	38	75	136	243	85	55	-63	521	501	55	103	251	97	241	142	512
MTB72 (amino acid)	**GTC (V)**	**CGG (R)/ CGA (R)**	**ACG (T)**	**GAC (D)**	**TCC (S)**	**GCA (A)/ GCG (A)[a]**	**GCG (A)/ GTG (V)**	**CGC (R)**	**ATG (M)**	**GGC (G)/ CGC (R)**	**TAC (Y)/ CAC (H)[a]**	**A**	**GCC (A)/ ACC (T)**	**GCA (A)**	**TGA (STOP)/ CGA (R)[a]**	**GCG (A)/ GCC (A)[a]**	**CTG (L)**	**GGG (G)**	**GTC (V)**	**ATC (I)**	**CAC (H)**
R4			ACC (T)			GCG (A)		TGC (C)			CAC (H)			GTA (V)	CGA (R)	GCC (A)					
R46		CGA (R)				GCG (A)	GTG (V)				CAC (H)		ACC (T)		CGA (R)	GCC (A)					
R181						GCG (A)					CAC (H)				CGA (R)	GCC (A)		AGG (R)			
R190	ATC (I)			GAT (D)	TCT (S)	GCG (A)			T (I)		CAC (H)				CGA (R)	GCC (A)					
RB14				GAT (D)[b]		GCG (A)[b]			T (I)[b]		CAC (H)[b]				CGA (R)[b]	GCC (A)[b]					
RB15				GAT (D)[b]		GCG (A)[b]			T (I)[b]		CAC (H)[b]				CGA (R)[b]	GCC (A)[b]					
RB16				GAT (D)[b]		GCG (A)[b]			T (I)[b]		CAC (H)[b]				CGA (R)[b]	GCC (A)[b]					
RB19				GAT (D)[b]		GCG (A)[b]			T (I)[b]		CAC (H)[b]				CGA (R)[b]	GCC (A)[b]					
B15																	CTT (L)				
B66		CGA (R)				GCG (A)	GTG (V)				CAC (H)		ACC (T)		CGA (R)	GCC (A)					
H103										CGC (R)		G									
H15										CGC (R)										ATC (I)/ TTC (F)	CAC (H)/ CGC (R)
H71										CGC (R)									ATC (I)		

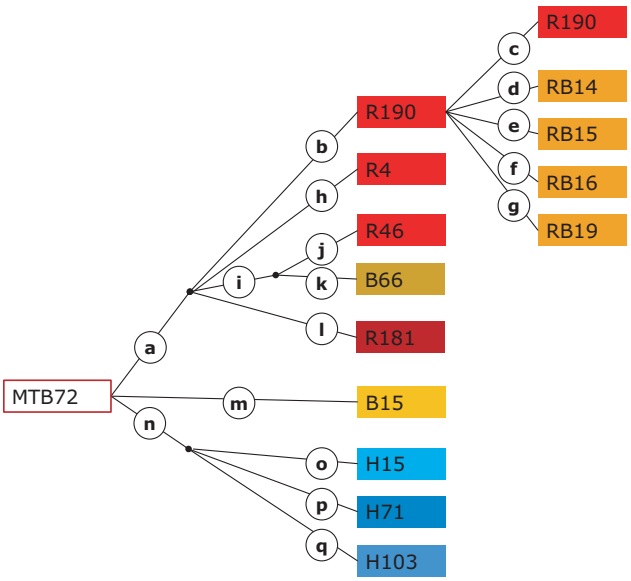

Figure 2. Inferred phylogeny of strains used in this study. Colors of boxes refer to selection procedures explained in Fig. 1 and the section 'Methods'. The phylogenetic tree is based on the following assumptions: (1) mutational events **a**, **i** and **n** occurred only once, (2) they occurred in the culture of the parent strain MTB72 before antibiotic selection and (3) the dominant allele observed in the MTB72 sample is the wild-type (or baseline) sequence. Branches are based on the following mutational events (a more detailed description of each mutation can be found in Table 1): **a.** ctpV-38, Rv1674c-55, ceoB-55, Rv2693c-24; **b.** zmp1–637, Rv1215c-243, rpoB-522; **c.** dnaA-270, Rv0466–93; **d.** rpoB-176; **e.** rpoB-178; **f.** rpoB-531; **g.** rpoB-531; **h.** hycQ-138, zwf1–136, murF-501, rpoB-526; **i.** dnaB-765, ispE-75, gcvB-521; **j.** rpoB-531; **k.** rpoB-531; **l.** lppY-97, rpoB-526; **m.** Rv2974c-251, rpoB-525–527; **n.** ilvA-85; **o.** crp-142, pks2–512, katG-321; **p.** Rv3446c-241, katG-384; **q.** wag22-Rv1760-(-)63, katG-271.

(*wag22*-Rv1760-A(-)63G) and H71 (Rv3446c-V241I) (Table 1). These singletons were fixed in the genomes of the strains carrying them; no other alleles could be detected [see Table S1 (Supporting Information) for coverage and read depth of all variants]. Six out of eleven mutations were non-synonymous (highlighted in blue, Table 1), four were synonymous and one was intergenic. The mutations in *crp* and *pks2* observed in H15 (Table 1) were accompanied by the wild-type allele which indicates that these variants were acquired in H15 post antibiotic selection. Both mutations are non-synonymous and the percentages of both mutant alleles were higher than the corresponding wild-type allele; the allelic distributions detected in H15 were 14/0/0/52 and 0/40/0/9 versus 111/0/0/0 and 0/0/0/59 in MTB72 for *crp* and *pks2*, respectively (order of the bases A/C/G/T).

One of the first-generation mutants, R190, was subjected to a second round of selection with rifabutin and four of its daughter strains (RB14, RB15, RB16 and RB19), with double *rpoB* mutations, were also included in this study (Fig. 1). Strain R190 acquired, in addition to the *rpoB*-S522L mutation associated with rifampicin resistance, four unique mutations (in *dnaA, zmp1*, Rv1215c and Rv0466, Table 1), two of which were absent in its offspring (*dnaA*-V270I and Rv0466-S93S), illustrating that these mutations were acquired and/or selected for after the daughter strains were selected. No additional mutations apart from those associated with rifabutin resistance were identified in the four daughter strains of R190.

Based on the data acquired by whole genome analysis, we inferred a phylogenetic tree (Fig. 2), using the data on drug resistance mutations to resolve the branches in the cases where this was necessary (e.g. the *rpoB*-S531L mutation which was acquired separately by four different strains). This tree is highly similar to the evolution assumed on the basis of our experimental procedure (Fig. 1), with the exception of strain B66 that showed to be more related to the R strains (namely R46, Fig. 2) than we anticipated. Most branches are represented by two or more mutational events, since it was not possible to determine the exact order and timing for each mutation separately.

DISCUSSION

The population structure of *M. tuberculosis* is highly clonal, with all phylogenetic members sharing more than 99% of their genome sequence. Genomic studies have found that epidemiologically linked strains, extracted from (patient-to-patient) transmission chains (Schürch *et al.*, 2010; Bryant *et al.*, 2013; Clark *et al.*, 2013; Torok *et al.*, 2013) or within the patient (Saunders *et al.*, 2011; Sun *et al.*, 2012), are genetically stable with estimated mutation rates of 0.3–0.5 SNPs/year in low-endemic areas (Bryant *et al.*, 2013; Walker *et al.*, 2013).

We have analysed the genomes of an isogenic collection of *in vitro* selected drug-resistant strains to determine if mutations, other than those conferring drug resistance, had accumulated during selection or cultivation. Our results show that during selection, mutations not associated with drug resistance were co-selected with the resistance trait. Some of these hitchhiking mutations were already detectable in the genome of the parent strain as minority variants and were likely accumulated over time in only a discrete amount of subpopulations [Tables 1 and 2 and Fig. S2 (Supporting Information)] (Barrick and Lenski 2013).

At the loci that showed mixed alleles in the MTB72 genome (Table 1), we assumed that the dominant allele was the original sequence and the minority variant represented the mutation (Table 2). However, for *ctpV*, Rv1674c, *ceoB* and Rv2693c, the dominant allele was different to that of the reference genome from H37Rv; instead, the minority allele corresponded to this reference genome. It is possible that the mutant allele had become dominant in the MTB72 culture in these specific locations. Conversely, the mutations at these loci in the MTB72 may have been historical and the minority variants would therefore represent 'back mutations'. From the data we have available, it was not possible to determine which scenario had occurred.

None of the strains acquired a mutation in *rpoA* or *rpoC*, some of which have been suggested to compensate for the fitness loss inflicted by *rpoB* mutations, particularly *rpoB*-S531L (Casali *et al.*, 2012; Comas *et al.*, 2012). However, compared to H37Rv, MTB72, and its entire offspring, carry an *rpoC*-G594E mutation. This mutation is not associated with compensation, but is a genotypic marker, specific for a Euro-American sub-lineage containing X and Haarlem spoligotype families (lineage 4.1 as described in Coll *et al.* (2014). Although an epistatic effect on the fitness cost of the rifampicin resistance mutations acquired of this *rpoC*-G594E mutation cannot be ruled out.

We cannot be sure if the singletons observed in the genomes of R4, R181, B15, H103, H71 and R190 (and its offspring) (Table 1) were either co-selected and therefore pre-existing as rare variants in the MTB72 culture, or acquired post-selection. However, the short cultivation period for these strains between selection and sequence analysis makes it probable that these mutations were co-selected.

The two unique mutations that were observed in strain R190 could only have been acquired after selection with rifabutin, since none of the R190 daughter strains carried either of these mutations. The fixation of these two post-selection mutations

Table 2. Read depth and coverage of loci identified as mixed alleles in the parent strain. Depicted are only those loci where mutations were acquired by at least one daughter strain and mixed alleles were observed in the parent strain MTB72. The order of the bases is A/C/G/T. Data highlighted in bold indicates that the base was detected in >5% of the reads.

Gene/CDS Genomic position	*dnaB* 62690	*ctpV* 1078856	*ispE* 1130414	*ilvA* 1763680	Rv1674c 1899754	*gcvB* 2077437	*ceoB* 3009506	Rv2693c 3011060
MTB72	**18**/0/**174**/0	**87**/0/**22**/0	0/**78**/0/3	0/1/**105**/0	**84**/0/**12**/0	**14**/0/**141**/0	0/5/0/**71**	0/**11**/**119**/0
R4	0/0/**737**/0	0/0/**410**/0	0/**97**/0/0	0/0/**427**/0	0/0/**403**/0	0/0/**608**/0	0/**459**/0/0	0/**490**/0/0
R46	**816**/0/0/0	0/0/**479**/0	0/0/0/**76**	0/0/**489**/0	0/0/**524**/0	**536**/0/0/0	0/**543**/0/1	0/**575**/0/0
R181	0/0/**643**/0	0/0/**353**/0	0/**496**/0/0	0/0/**385**/0	0/0/**404**/0	0/0/**492**/0	0/**460**/0/0	0/**412**/0/0
R190	0/0/**749**/0	0/0/**460**/0	0/**47**/0/0	0/0/**480**/0	0/0/**456**/0	0/0/**626**/0	0/**499**/0/0	0/**533**/0/0
RB14	0/0/**677**/0	0/0/**409**/0	0/**514**/0/0	0/0/**483**/0	0/0/**431**/0	0/0/**553**/0	0/**446**/0/0	0/**497**/0/0
RB15	0/0/**802**/0	0/0/**537**/0	0/**324**/0/0	0/0/**552**/0	0/0/**529**/0	0/0/**656**/0	0/**568**/0/0	0/**589**/0/0
RB16	0/0/**753**/1	0/0/**518**/0	0/**477**/0/0	0/0/**540**/0	0/0/**540**/0	0/0/**613**/0	0/**527**/0/0	0/**647**/0/0
RB19	0/0/**713**/0	1/0/**415**/0	0/**363**/0/0	0/0/**430**/0	0/0/**438**/0	0/0/**499**/0	0/**446**/0/0	1/**475**/0/0
B15	0/0/**250**/0	**133**/0/0/0	0/0/0/**492**	0/0/**166**/0	**128**/0/0/0	0/0/**183**/0	0/0/0/**114**	1/0/**161**/0
B66	**203**/0/0/0	0/0/**96**/0	0/**418**/0/0	0/0/**136**/0	0/0/**120**/0	**183**/0/0/0	0/**111**/0/0	0/**132**/0/0
H103	0/0/**831**/0	**566**/0/0/0	0/**495**/0/0	0/**513**/0/0	**488**/0/0/0	1/0/**654**/0	0/1/0/**541**	0/0/**603**/0
H15	0/0/**137**/0	**66**/0/0/0	0/**500**/0/0	0/**50**/0/0	**78**/0/0/0	0/0/**113**/0	0/0/0/**52**	0/0/**99**/0
H71	0/0/**807**/0	**480**/0/0/0	0/**411**/0/0	0/**467**/0/0	**473**/0/0/0	0/0/**619**/0	0/0/0/**586**	0/0/**593**/0

raises the possibility that one or both confer a benefit and were under selective pressure, possibly restoring a fitness deficit resulting from the *rpoB*-S522L mutation. This appears most likely for the mutation in *dnaA* (Table 1), which is non-synonymous and thus suggestive of a phenotypic effect.

Our findings demonstrate that there can be a detectable amount of minority variants present in a bacterial population, possibly reflecting the genetic diversity in (cultured) patient isolates that are otherwise assumed to be largely homogenous. Studies with *Streptococcus pneumoniae* also show that bacteria are much more likely than previously anticipated to acquire mutations in addition to those that confer the specific trait selected for (Stevens and Sebert 2011). Furthermore, this concept may have consequences for the researcher aiming to investigate single isogenic mutants to assess their associated (*in vitro*) characteristics, as we show here that even during a single selection step, unintended genetic changes may be introduced. Especially against a stable genetic background such as that of *M. tuberculosis* these hitchhiking mutations may be of greater influence on the phenotype than anticipated.

Sequencing technologies that produce greater numbers of higher quality reads of longer length increase the possibility to detect minority variants with low copy numbers, allowing the true genetic dynamics of a bacterial population under evolution to be studied. During patient-to-patient transmission, rigorous bottlenecks can dramatically reduce the effective, net mutation rate by selecting out most of the newly acquired mutations within the bacterial populations transmitted (Sun *et al.*, 2012). Deep sequencing (> 500-fold coverage) therefore has the potential to be particularly appropriate to study the dynamics of *M. tuberculosis* strain(s) within a single patient.

Here, we have shown that a single selection event can result in the one-step fixation of a potentially informative number of pre-existing mutations. Minority variants in clinical isolates should thus not be ignored, as the potentially characteristic signatures of these hitchhiking or co-selected mutations in the genomes of infected strains can be used to help resolve the details of patient-to-patient transmission.

SUPPLEMENTARY DATA

Supplementary data is available at FEMSLE online.

ACKNOWLEDGEMENTS

We would like to thank Mridul Nair for preparation of the Illumina genomic libraries and performance of genome sequencing at the Bioscience Core Lab in KAUST.

FUNDING

FC is funded by a Bloomsbury PhD studentship. TC receives funding from the MRC UK (Grant number MR/K000551/1).

Conflict of interest statement. None declared.

REFERENCES

Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 2008;**62**:53–70.

Anthony RM, Schuitema ARJ, Bergval IL, *et al.* Acquisition of rifabutin resistance by a rifampicin resistant mutant of *Mycobacterium tuberculosis* involves an unusual spectrum of mutations and elevated frequency. *Ann ClinMicrobiol Antimicrobials* 2005;**4**:9.

Barrick JE, Lenski RE. Genome dynamics during experimental evolution. *Nat Rev Genet* 2013;**14**:827–39.

Bergval IL, Klatser PR, Schuitema ARJ, *et al.* Specific mutations in the *Mycobacterium tuberculosis rpoB* gene are associated with increased *dnaE2* expression. *FEMS Microbiol Lett* 2007;**275**:338–43.

Bergval IL, Schuitema ARJ, Klatser PR, *et al.* Resistant mutants of *Mycobacterium tuberculosis* selected *in vitro* do not reflect the *in vivo* mechanism of isoniazid resistance. *J Antimicrob Chemoth* 2009;**64**:515–23.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**:2114–20.

Brosch R, Gordon SV, Marmiesse M, *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *P Natl Acad Sci USA* 2002;**99**:3684–9.

Bryant JM, Schurch AC, van DH, *et al.* Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis* 2013;**13**:110.

Camus JC, Pryor MJ, Medigue C, *et al*. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 2002;**148**:2967–73.

Casali N, Nikolayevskyy V, Balabanova Y, *et al*. Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* 2012;**22**:735–45.

Casali N, Nikolayevskyy V, Balabanova Y, *et al*. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014;**46**:279–86.

Caws M, Thwaites G, Dunstan S, *et al*. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 2008;**4**:e1000034.

Clark TG, Mallard K, Coll F, *et al*. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PLoS One* 2013;**8**:e83012.

Cohen T, Sommers B, Murray M. The effect of drug resistance on the fitness of *Mycobacterium tuberculosis*. *Lancet Infect Dis* 2003;**3**:13–21.

Coll F, McNerney R, Guerra-Assuncao JA, *et al*. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 2014;**5**:4812.

Comas I, Borrell S, Roetzer A, *et al*. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012;**44**:106–10.

DePristo MA, Banks E, Poplin R, *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8.

Derrien T, Estelle J, Marco SS, *et al*. Fast computation and applications of genome mappability. *PLoS One* 2012;**7**:e30377.

Ford C, Yusim K, Ioerger T, *et al*. *Mycobacterium tuberculosis*—heterogeneity revealed through whole genome sequencing. *Tuberculosis* 2012;**92**:194–201.

Ford CB, Lin PL, Chase MR, *et al*. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet* 2011;**43**:482–8.

Gagneux S, Long CD, Small PM, *et al*. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* 2006;**312**:1944–6.

Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 2007;**7**:328–37.

Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 2014.

Li H, Handsaker B, Wysoker A, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9.

Luciani F, Sisson SA, Jiang H, *et al*. The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. *P Natl Acad Sci USA* 2009;**106**:14711–5.

Mariam DH, Mengistu Y, Hoffner SE, *et al*. Effect of rpoB mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob Agents Ch* 2004;**48**:1289–94.

Mariam SH, Werngren J, Aronsson J, *et al*. Dynamics of antibiotic resistant *Mycobacterium tuberculosis* during long-term infection and antibiotic treatment. *PLoS ONE* 2011;**6**:e21147.

Musser JM. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev* 1995;**8**:496–514.

Niemann S, Koser CU, Gagneux S, *et al*. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. *PLoS One* 2009;**4**:e7407.

Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tubercle Lung Dis* 1998;**79**:3–29.

Rutherford K, Parkhill J, Crook J, *et al*. Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)* 2000;**16**:944–5.

Saunders NJ, Trivedi UH, Thomson ML, *et al*. Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. *J Infect* 2011;**62**:212–7.

Schürch AC, Kremer K, Kiers A, *et al*. The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. *Infect Genet Evol* 2010;**10**:108–14.

Sherman DR, Mdluli K, Hickey MJ, *et al*. Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* 1996;**272**:1641–3.

Smith NH, Hewinson RG, Kremer K, *et al*. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 2009;**7**:537–44.

Smith NH, Kremer K, Inwald J, *et al*. Ecotypes of the *Mycobacterium tuberculosis* complex. *J Theor Biol* 2006;**239**:220–5.

Soolingen DV, Hermans PW, de Haas PE, *et al*. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991;**29**:2578–86.

Sreevatsan S, Pan X, Stockbauer KE, *et al*. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *P Natl Acad Sci USA* 1997;**94**:9869–74.

Stevens KE, Sebert ME. Frequent beneficial mutations during single-colony serial transfer of Streptococcus pneumoniae. *PLoS Genet* 2011;**7**:e1002232.

Streicher EM, Bergval I, Dheda K, *et al*. *Mycobacterium tuberculosis* population structure determines the outcome of genetics-based second-line drug resistance testing. *Antimicrob Agents Ch* 2012;**56**:2420–7.

Sun G, Luo T, Yang C, *et al*. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis* 2012;**206**:1724–33.

Torok ME, Reuter S, Bryant J, *et al*. Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak. *J Clin Microbiol* 2013;**51**:611–4.

Walker TM, Ip CL, Harrell RH, *et al*. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013;**13**:137–46.