# Single-nucleotide variations associated with *Mycobacterium tuberculosis* KwaZulu-Natal strains

Sarbashis Das[1, §], Ragothaman M Yennamalli[1, §], Anchal Vishnoi[1, §], Parul Gupta[1] and Alok Bhattacharya[1, 2*]

[1]*Center for Computational Biology and Bioinformatics, School of Information Technology, Jawaharlal Nehru University, New Delhi 110 067, India*

[2]*School of Life Sciences, Jawaharlal Nehru University, New Delhi 110 067, India*

[§]*These authors contributed equally to this work.*

**Corresponding author (Fax, +91-11-26741586; Email, alok0200@mail.jnu.ac.in)*

The occurrence of drug resistance in *Mycobacterium tuberculosis*, the aetiological agent of tuberculosis (TB), is hampering the management and control of TB in the world. Here we present a computational analysis of recently sequenced drug-sensitive (DS), multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains of *M. tuberculosis*. Single-nucleotide variations (SNVs) were identified in a pair-wise manner using the anchor-based whole genome comparison (ABWGC) tool and its modified version. For this analysis, four fully sequenced genomes of different strains of *M. tuberculosis* were taken along with three KwaZulu-Natal (KZN) strains isolated from South Africa including one XDR and one MDR strain. KZN strains were compared with other fully sequenced strains and also among each other. The variations were analysed with respect to their biological influence as a result of either altered structure or synthesis. The results suggest that the DR phenotype may be due to changes in a number of genes. The database on KZN strains can be accessed through the website *http://mirna.jnu.ac.in/mgdd/*.

[Das S, Yennamalli R M, Vishnoi A, Gupta P and Bhattacharya A 2009 Single-nucleotide variations associated with *Mycobacterium tuberculosis* KwaZulu-Natal strains; *J. Biosci.* **34** 397–404]

## 1. Introduction

Tuberculosis (TB) is a major public health problem throughout the world, particularly in developing countries. A recent estimate suggests that, every year, 8.8 million new cases with 1.6 million deaths can be attributed to TB (WHO 2006). Malnutrition and poor sanitary conditions contribute to the large-scale occurrence of the disease. The HIV pandemic has also played a role in increasing the burden of the disease all over the world, particularly in Africa (Corbett *et al.* 2006). The emergence of drug-resistant (DR) strains of *Mycobacterium tuberculosis* has hampered progress in the treatment and control of the disease. Many of these organisms are resistant to multiple drugs, limiting the choice of drugs for TB control (Lawn and Wilkinson 2006). Multidrug-resistant TB (MDR-TB) has been reported from almost all geographical regions of the world and one estimate suggests that 5.3% of all new cases may be multidrug resistant (Editorial Team 2008). It is suspected that this figure may rise at a rapid rate unless methods are developed to control it. Understanding the molecular basis of drug resistance will help in devising new approaches to

**Keywords.** Drug resistant; KZN; MDR; *Mycobacterium tuberculosis*; SNV; XDR

Abbreviations used: ABWGC, anchor-based whole genome comparison; AFLP, amplified fragment length polymorphism; COG, Clusters of Orthologous Groups; DS, drug sensitive; KZN, KwaZulu-Natal; MDR, multidrug-resistant; SNV, single-nucleotide variation; TB, tuberculosis; XDR, extensively drug-resistant

Supplementary tables pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/sept09/pp397-404-suppl.pdf*

the diagnosis, therapeutics and overall management of the disease burden.

MDR strains show resistance to multiple drugs, such as the front-line drugs isoniazid and rifampicin. The molecular basis of drug resistance has been studied in different bacterial pathogens and a few general principles have emerged. These are essentially based on alterations in drug entry or efflux (Li and Nikaido 2005), drug metabolism and drug targets (Blanchard 1996). Any or a combination of these mechanisms can make any pathogen including mycobacteria resistant to not only a single drug but a number of them. For example, overexpression of efflux pumps can make an organism multidrug resistant (Jiang *et al.* 2008). A number of studies to understand the genetic alterations in drug-resistant organisms are ongoing, particularly in mycobacteria, in order to identify the mechanisms involved in the resistance (Telenti 1998). These studies are essentially based on a few target gene sequences. The other major approach that has been adapted to understand drug resistance in *M. tuberculosis* is the detection of polymorphisms based on methods such as spoligotyping (Sola *et al.* 2000), amplified fragment length polymorphism (AFLP) (Vos *et al.* 1995) and interspersed repeat typing (Edith *et al.* 2001). None of these approaches can help to decipher the global genetic variations among sensitive and resistant strains. Availability of complete genome sequences of different strains and species of mycobacteria including extensively drug-resistant (XDR) strains has given us a way of understanding the global picture of genetic alterations and their relation to phenotypic differences such as drug resistance. The fully sequenced genomic sequences can be analysed to identify all the variations and a detailed study can be conducted to correlate genetic and phenotypic differences. Our group has developed a tool and a database of the genome diversity of *M. tuberculosis* (Vishnoi *et al.* 2007, 2008). In this study, we report further refinement of our tool and analysis of genomic sequences obtained from three strains of *M. tuberculosis,* namely DS, MDR and XDR, obtained from KwaZulu-Natal (KZN) of South Africa and compared these with other DS strains for which genome sequences are already available. These organisms have been sequenced at the Broad Institute, MIT, USA (Broad Institute 2008). The data obtained from our study show involvement of a large number of genes and suggest a diverse range of changes that may have contributed to the altered phenotype.

## 2. Materials and methods

Anchor-based whole genome comparison (ABWGC) was modified for faster and more efficient identification of SNVs (Vishnoi *et al.* 2007). The new algorithm for detection of SNVs is briefly described.

### 2.1 *Identification of SNVs*

2.1.1 Extraction of inter-anchors: Identification of homologous anchors and extraction of inter-anchor regions have already been described in detail (Vishnoi *et al*. 2007). These inter-anchor regions were then classified into three different classes on the basis of the difference in lengths between the two homologous regions.

(i) When the lengths are exactly the same: A method based on dinucleotide frequency was developed to find SNVs in this situation, as such sequences may or may not contain any variation. Nearly 95% of the anchors belong to this category. The fact that the lengths are exactly the same suggests that the two sequences are nearly identical with a few SNVs. A change in the frequency of dinucleotides will reflect an alteration in the nucleotide sequence and this may be able to capture the presence of an SNV. Calculation involves computation of all the possible dinucleotides of both the sequences followed by a comparison between them. If the number of any single–dinucleotide combination in the query inter-anchor differs from that of the target, it suggests the presence of SNVs. It is also likely that the sequence has both insertions and deletions of the same length. This method allows identification of such differences. The exact position is then determined by a suffix tree-based approach. This helps to screen out a large number of inter-anchor regions that do not have any SNVs.

(ii) When the lengths differ by less than 5 nucleotides: In this case, a suffix tree-based method was implemented (Gusfield 1997). The suffix tree was constructed essentially as described earlier. Briefly, the sequences were concatenated and the tree was constructed using the concatenated sequences, and common subsequences longer than 2 nucleotides are obtained. The results are put into a hash table and then parsed to get SNVs.

(iii) When the lengths vary by more than 5 nucleotides: In this case, a global alignment tool such as Needleman and Wunsch was used (Needleman and Wunsch 1970). Our aim was to avoid alignment-based methods as much as possible to save computation time. In the approach described here, less than 1% of the inter-anchor sequences were subjected to global alignment.

### 2.2 *Data*

The whole genome sequences of *M. tuberculosis* H37Rv (Accession number NC_000962.2), *M. tuberculosis* H37Ra (Accession number NC_009525.1), *M. tuberculosis* F11 (Accession number NC_009565.1) and *M. tuberculosis* CDC1551 (Accession number NC_002755.2) were obtained from NCBI (*http://www.ncbi.nlm.nih.gov/genomes/*) and supercontigs of *M. tuberculosis* KZN1435, *M. tuberculosis* KZN605, *M. tuberculosis* KZN4207 were obtained from

the Broad Institute (*http://www.broad.mit.edu/annotation/genome/mycobacterium_tuberculosis_spp/*).

## 3. Results and discussion

### 3.1 *Sequence variations in* M. tuberculosis *KZN strains*

The three strains of *M. tuberculosis* isolated from KwaZulu-Natal, South Africa (KZN strains) were compared with *M. tuberculosis* H37Rv strain in a pair-wise manner to identify SNVs. These were identified using a modified version of ABWGC as described in the Methods section. The SNVs were then compared with the previously identified ones from the four *M. tuberculosis* strains H37Rv, H37Ra, CDC1551 and F11. This helped to locate the unique ones present only in the KZN isolates. The results were compared with the analysis presented in the website of the Broad Institute with respect to the variation observed in comparison with the strain F11 (*http://www.broad.mit.edu/annotation/genome/mycobacterium_tuberculosis_spp/*). Though the two results cannot be directly compared, it is interesting to note some of the major differences. The number of common SNVs (among all KZN strains) observed using ABWGC was found to be 348 whereas the value was much less (25) in the Broad Institute study (figure 1a). Most of the SNVs in the latter study were not shared by KZN strains. Supplementary table S1 lists these unique differences identified in comparison with *M. tuberculosis* strain H37Rv.

SNVs were also identified by comparing drug-resistant (DR) KZN strains with the drug-sensitive (DS) strain 4207 in a pair-wise manner. The XDR strain KZN605 displayed

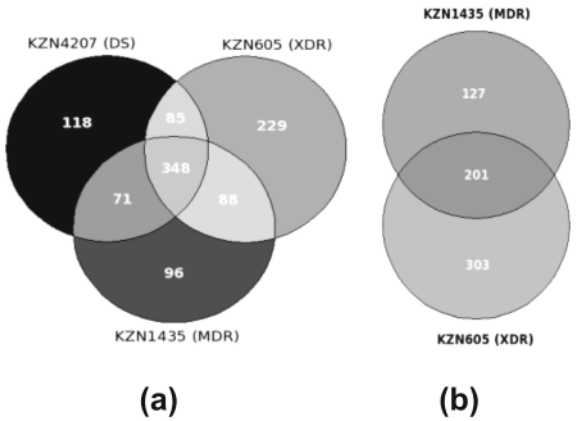


**Figure 1. (a)** Single-nucleotide variations (SNVs) in KZN strains KZN4207 (DS), KZN1435 (MDR) and KZN605 (XDR) in comparison with *M. tuberculosis* H37Rv. The SNVs shared with other *M. tuberculosis* strains (H37Ra, CDC1551, F11) were removed for this analysis. These are only present in KZN strains (unique). **(b)** SNV analysis of the KZN605 and KZN1435 strains in comparison with KZN4207.

750 SNVs, more than that observed for KZN1435 (figure 1b). These values were much lower than those found between strains H37Rv and CDC1551 (1850), suggesting that KZN strains may have evolved recently. The majority of the differences among the strains were due to unique SNVs present only in one strain, i.e. KZN605. A large number of SNVs in KZN605 may have contributed to the extremely drug resistant (XDR) phenotype. The pattern of changes observed in these three KZN strains in comparison with strain H37Rv are likely to be different when the changes are enumerated with respect to other strains, such as F11.

### 3.2 *Distribution of SNVs in different functional categories of genes*

The common SNVs present in all the three KZN strains in comparison with H37Rv, and absent in other strains of *M. tuberculosis,* were identified. These were mapped to 18 different functional classes based on Clusters of Orthologous Groups (COG) (Tatusov *et al.* 2001) (figure 2). The genes and the SNVs are listed in supplementary table S2. The results reflect the variations that may be specific for KZN isolates (figure 2). No SNV was found in N (cell motility) in any of the strains and very few were mapped to F (nucleotide transport and metabolism), suggesting that the genes in these categories may be under evolutionary pressure. Significant numbers of SNVs were found in genes belonging to Q (biosynthesis of secondary metabolites, transport and catabolism) in all the three strains. The results are not surprising as it is known that genes belonging to this functional category may have an important role in adaptation of the organism to its surroundings. The 88 SNVs shared by MDR and XDR strains (*see* figure 1a) are listed in supplementary table S3. There are many genes in the table, for example, *RpoB* (Miller *et al.*1994), *GidB* (Okamoto *et al.* 2007) and *GyrA* (Aubry *et al.* 2004), which are known to be involved in drug resistance. Besides these, there are likely to be a number of other genes that may indirectly influence drug response by altering cellular chemistry. The majority of SNVs belong to the latter class.

### 3.3 *Structural characterization of some of the SNVs*

The effect of nucleotide substitution on a specific gene can sometimes be predicted by analysing the nature of the nucleotide and, consequently, amino acid substitution in the altered sequence. All the unique and common SNVs present in major functional classes were checked for the changes that they are likely to cause in the homologues present in KZN strains and some of the results are shown in supplementary table S1. A significant number of SNVs were found to cause synonymous substitutions with no change in

amino acids (27%). A number of studies in different systems including mycobacteria have shown that the level of protein expression depends upon codon abundance (Kanekiyo *et al*. 2005; Tobias *et al.* 2007). Highly expressed genes can be

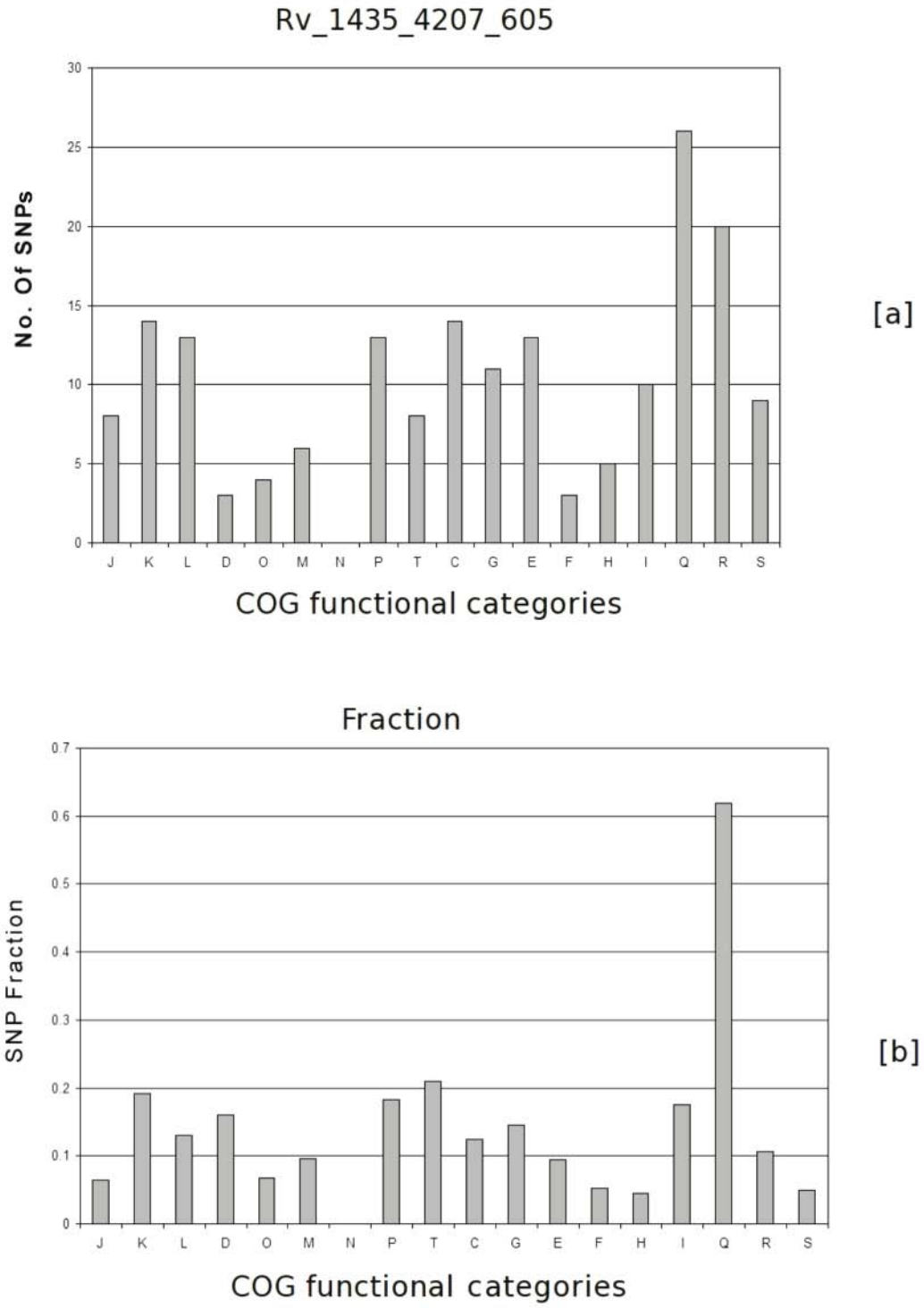


**Figure 2.** Common single-nucleotide variations (SNVs) present in all KZN strains but not in other strains of *M. tuberculosis* mapped to genes as per the COG functional classification. **(a)** Number of SNVs in a given class. **(b)** Number of SNVs normalized with respect to the number of genes in a class.

predicted on the basis of the presence of abundant codons and absence of rare codons (Wu *et al.* 2004). Moreover, codon optimization is an accepted strategy to enhance the expression of heterologous proteins (Puigbo *et al.* 2007). Therefore, the synonymous substitutions were further analysed to check differences in codon abundance. The data on some of the SNVs that showed major differences in codon abundance are included in table 1. In many cases, there was a 5–10-fold change, suggesting that these SNVs may cause a change in the expression of the genes (Kudla *et al*. 2009).

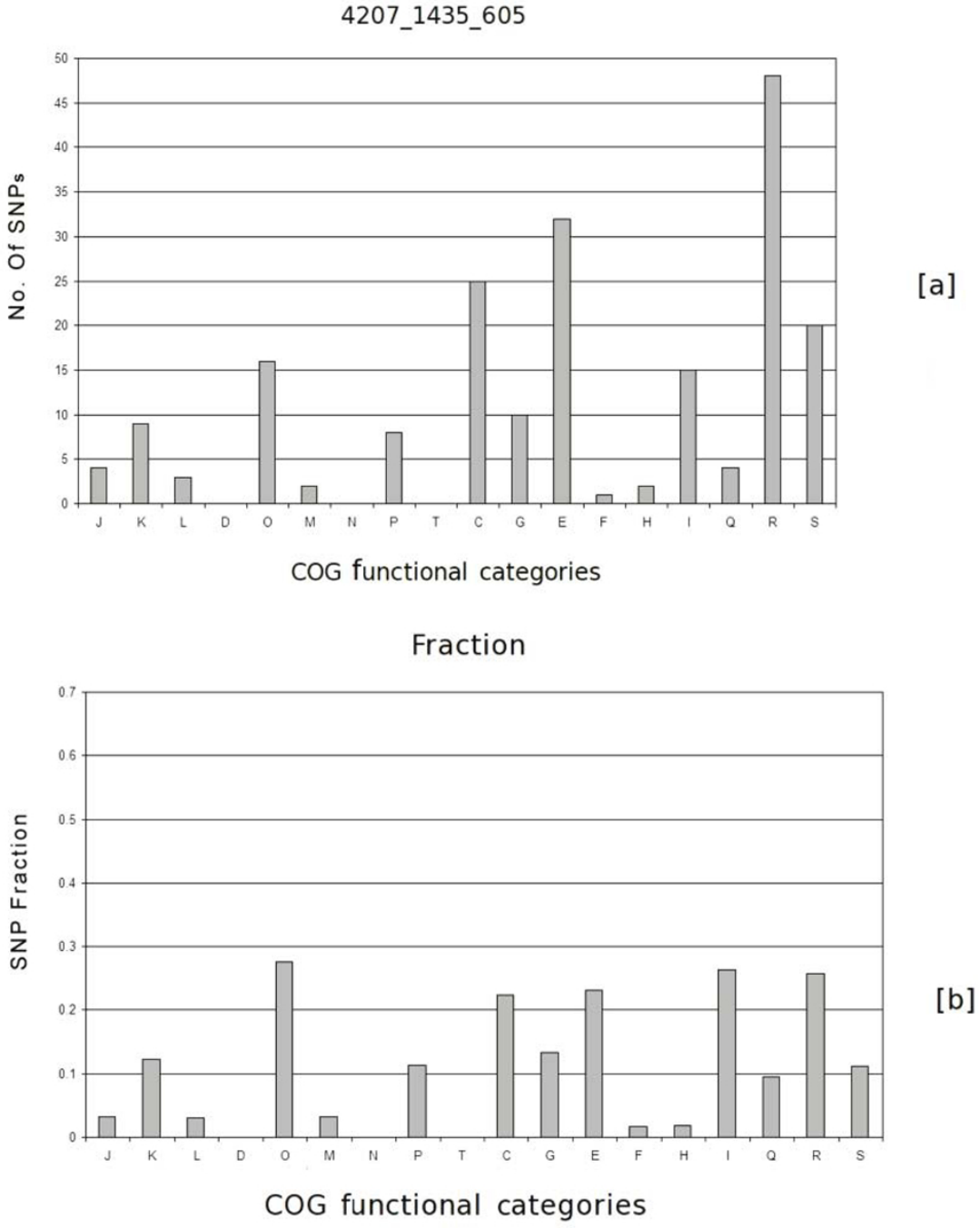


**Figure 3.** Common single-nucleotide variations (SNVs) present in multidrug-resistant (MDR) and extensively drug-resistant (XDR) KZN strains in comparison with drug-sensitive strains mapped to genes as per the COG functional classification. **(a)** Number of SNVs in a given class. **(b)** Number of SNVs normalized with respect to the number of genes in a class.

In case of non-synonymous substitutions, there is likely to be an alteration in the overall structure, if the substitutions involve different classes of amino acids; for example, when charged and uncharged amino acids replace each other. Many SNVs may cause only subtle changes, such as Glu to Asp or vice versa. Such changes have been shown to cause subtle alterations in proteins resulting in functionally or kinetically altered proteins (Marinetti *el al.* 1989; Koster *et al.* 1996; Rao *et al.* 2007). Table 1 lists some of the changes that were observed in DR KZN strains. In this case, it is expected that the some of the altered proteins may be kinetically different from some of the examples cited above. Often, a sense codon is replaced by a non-sense codon resulting in truncation of a protein (table 1).

Prediction of the secondary structure of proteins containing some of the altered amino acids due to non-synonymous SNVs was carried out using Jpred3 (*www.compbio.dundee.ac.uk/www-jpred/*). All the changes, except for Rv1773c, were predicted in the loop region. Only Rv1773c was found to be in the helix state. In general, the loop regions of proteins are involved in functional activity (Wehbi *et al*. 2007). Since it is known that one amino acid alteration can give rise to a change in entropy at the surface-exposed residues (Agnieszka *et al.* 2002), it is possible that some of the SNVs may have an effect on the functionality of proteins.

### 3.4 *Variation within KZN strains*

SNVs were also identified by comparing the two DR strains against DS strain KZN4207 as shown in figure 1b. XDR strain KZN605 displayed many more SNVs, suggesting that this strain has undergone extensive genomic changes. Out of the 201 shared SNVs, 33 were found in the intergenic regions, suggesting that alteration in gene expression between DS and DR strains may be due to these variations. The genes containing the shared SNVs were analysed with respect to the COG classification and normalized to the number of COGs in each category as discussed earlier (figure 3). The majority of SNVs were mapped to four COG classes, namely O (post-translational modification, protein turnover, chaperones), E (amino acid transport and metabolism), C (energy production and conversion) and I (lipid transport and metabolism). This result was not surprising as all these classes of genes can be associated with altered metabolism in organisms of different geographical origin. The genes belonging to these functional categories are likely to affect drug sensitivity directly or indirectly. For example, entry of drugs is an important first step in their action. Changes in the membrane by lipid alteration can alter drug transport processes. Inhibition in the uptake of the drug has been postulated to be a mechanism of drug resistance in Mycobacteria (Raynaud *et al.* 1999).

**Table 1.** Codon usage in some shared single-nucleotide variations (SNVs) among the three KZN strains in comparison with *M. tuberculosis* H37Rv

| Position | SNP | Gene | Amino acid changes | Codon frequency |
|---|---|---|---|---|
| 2007544 | T–G | Rv1773c | GLU–ASP | 16.2–42.2 |
| 3314628 | C–T | Rv2962c | TRP–STOP | 14.6–1.6 |
| 311612 | G–T | Rv0260c | VAL–VAL | 32.7–4.7 |
| 1315190 | A–C | Rv1180 | STOP–SER | 0.5–3.6 |
| 3247315 | C–G | Rv2931 | ASP–GLU | 42.2–30.6 |
| 2216247 | C–G | Rv1971 | PRO–ALA | 31.6–48.7 |
| 489934 | G–C | Rv0405 | ARG–PRO | 24.6–31.6 |
| 1740770 | A–C | Rv1537 | THR–PRO | 35.2–17.0 |
| 3238702 | T–G | Rv2924c | GLU–ASP | 16.2–42.2 |
| 2347441 | G–C | Rv2090 | ASP–HIS | 42.2–15.8 |
| 403979 | G–A | Rv0338c | ALA–ALA | 48.7–10.9 |
| 1759251 | G–T | Rv1552 | SER–SER | 19.4–2.2 |

The list of 201 SNVs along with the annotations is shown in supplementary table 3. SNVs in genes such as *Fgd1* (Bashiri *et al.* 2008) and *pepA* (Gibson *et al.* 1984) have been known to be involved in drug resistance.

The sequencing of MDR and XDR strains of *M. tuberculosis* has given us an unprecedented opportunity to understand the process of drug resistance in this complex organism. As pointed out earlier, there are multiple mechanisms of drug resistance. Some of these can be due to alteration or acquisition of a single or a few genes, such as those encoded by a plasmid or changes in drug efflux pumps (Li and Nikaido 2005). Our comprehensive genomewide analysis has shown that variation in *M. tuberculosis* strains may not be due to loss or gain of genes (Ochman *et al.* 2001). Rather, the changes are many, comprising mostly SNVs, small insertions/deletions, alteration in the number of tandem repeats and insertion/deletion of IS elements and phages. Many of these changes do not cause drastic alterations in gene function. It is believed now that all the components in a cell work by interacting with each other, forming a huge network of interacting partners of genes, proteins and metabolites (Dotan-Cohen *et al.* 2009). Cellular changes can be brought about by changing this network and a number of small alterations in genes due to these SNVs can, in principle, change the network and subsequently the properties of the organism. The larger number of SNVs observed in XDR strains suggest that this can be a potential mechanism for the evolution of drug resistance in mycobacteria. Future work on systems-level analysis of these strains may throw a more definitive light on this important area.

**Acknowledgements**

This work was supported by the Department of Biotechnology and Department of Science and Technology, New Delhi. The authors thank Broad Institute MIT, USA for making available the sequences of the KZN strains.

**References**

Agnieszka M, Yancho D, Daniel K, Kenton L, Zbigniew D, Jacek O and Zygmunt S D 2002 The impact of Glu->Ala and Glu -> Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution; *Acta Crystallogr..D* **58** 1983–1991

Aubry A, Pan X S, Fisher L M, Jarlier V and Cambau E 2004 *Mycobacterium tuberculosis* DNA gyrase: interaction with quinolones and correlation with antimycobacterial drug activity; *Antimicrob. Agents Chemother.* **48** 1281–1288

Bashiri G, Squire C J, Moreland N J, Edward N and Baker 2008 Crystal structures of an f420-dependent glucose-6-phosphate dehydrogenase fgd1 involved in the activation of the anti-TB drug candidate pa-824 reveal the basis of coenzyme and substrate binding; *J. Biol. Chem.* **283** 17531–17541

Blanchard J 1996 Molecular mechanisms of drug resistance in *Mycobacterium tuberculosis*; *Annu. Rev. Biochem.* **65** 215–239

Broad Institute 2008 MIT (*http://www.broad.mit.edu/annotation/genome/mycobacterium_tuberculosis_spp/Downloads.html*) Corbett E, Marston B, Churchyard G and Cock De K 2006 Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment; *Lancet* **367** 926–937

Dotan-Cohen D, Letovsky S, Melkman A A and Kasif S 2009 Biological process linkage networks; *PLos One* **4** e5313

Mazars E, Lesjean S, Banuls A L, Gilbert M, Vincent V, Gicquel B, Tibayrenc M, Locht C and Supply P 2001 High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology; *Proc. Natl. Acad. Sci. USA* **98** 1901–1906

Editorial team 2008 World Health Organization reports highest rates of drug-resistant tuberculosis to date; *Euro Surveill*. 13 pii=8077. Available online: *http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=8077*

Gibson M M, Price M and Higgins C F 1984 Genetic characterization and molecular cloning of the tripeptide permease (tpp) genes of *Salmonella typhimurium*; *J. Bacteriol.* **160** 122–130

Gusfield D 1997 *Algorithms on stings, trees, and sequences: computer science and computational biology* (New York: Cambridge University Press) pp 89–209

Jiang X, Zhang W, Zhang Y, Gao F, Lu C, Zhang X and Wang H 2008 Assessment of efflux pump gene expression in a clinical isolate *Mycobacterium tuberculosis* by real-time reverse transcription PCR; *Microb. Drug. Resist.* **14** 7–11

Kanekiyo M, Matsuo K, Hamatake M, Hamano T, Ohsu T, Matsumoto S, Yamada T, Yamazaki S, Hasegawa A, Yamamoto N and Honda M 2005 Mycobacterial codon optimization enhances antigen expression and virus-specific immune responses in recombinant *Mycobacterium bovis* Bacille Calmette-Guérin expressing human immunodeficiency virus type 1 gag; *J. Virol.* **79** 8716–8723

Koster J C, Blanco G, Mills P B and Mercer R 1996 Substitution of glutamate 781 in the Na,K-ATPase α subunit demonstrate reduced cation selectivity and an increased affinity for ATP; *J. Biol. Chem.* **271** 2413–2421

Kudla G, Murray A W, Tollervery D and Joshua P B 2009 Coding-sequence determinants of gene expression in *Escherichia coli*; *Science* **324** 255–258

Lawn S and Wilkinson R 2006 Extensively drug resistant tuberculosis; *BMJ* **333** 559–560

Li X and Nikaido H 2005 Efflux mediated drug resistance in bacteria; *Drugs* **64** 159–204

Marinetti T, Subramaniam S, Mogi T, Marti T and Khorana H G 1989 Replacement of aspartic residues 85, 96, 115, or 212 affects the quantum yield and kinetics of proton release and uptake by bacteriorhodopsin; *Proc. Natl. Acad. Sci. USA* **86** 529–533

Miller L P, Crawford J T and Shinnick T M 1994 The rpoB gene of *Mycobacterium tuberculosis*; *Antimicrob. Agents Chemother.* **38** 805–811

Needleman S and Wunsch C 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins; *J. Mol. Biol.* **48** 443–453

Ochman H and Moran NA 2001 Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis; *Science* **292** 1096–1099

Okamoto S, Tamaru A, Nakajima C, Nishimura K, Tanaka Y, Tokuyama S, Suzuki Y and Ochi K 2007 Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria; *Mol. Microbiol.* **63** 1096–1106

Puigbo P, Guzma E, Romeu A and Garcia-Vallve S 2007 OPTIMIZER: a web server for optimizing the codon usage of DNA sequences; *Nucleic Acids Res*. *(Suppl. 2)* **35** W126–W31

Raynaud C, Lanéelle M A, Senaratne R H, Draper P, Lanéelle G and Daffé M 1999 Mechanisms of pyrazinamide resistance in mycobaceria: importance of lack of uptake in addition to lack of pyrazinamidase activity; *Microbiology* **145** 1359–1367

Rao S, Fu Z, Albro M, Narayanan, Baddam S, Lee H, Kim J P and Frerman F 2007 The effect of a Glu370Asp mutation in glutaryl-CoA dehydrogenase on proton transfer to the dienolate intermediate; *Biochemistry* **46** 14467–14477

Sola C, Filliol I, Legrand E and Rastogi N 2000 Recent developments of spoligotyping as applied to the study of epidemiology, biodiversity and molecular phylogeny of the *Mycobacterium tuberculosis* complex; *Pathol. Biol.* **48** 921–932

Tatusov R L, Natale D A, Garkavtsev I V, Tatusova T A, Shankavaram U T, Rao B S, Kiryutin B, Galperin M Y, Fedorova N D and Koonin E V 2001 The COG database: a new development in phylogenetic classification of proteins from complete genomes; *Nucleic Acids Res.* **29** 22–28

Telenti A 1998 Genetics of drug resistant tuberculosis; *Thorax* **53** 793–797

Tobias B, Kalin V and Roy K 2007 Evolution and multilevel optimization of the genetic code; *Genome Res.* **17** 401–404

Vishnoi A, Roy R and Bhattacharya A 2007 Comparative analysis of bacterial genomes: identification of divergent regions in mycobacterial strains using an anchor-based approach; *Nucleic Acids Res.* **35** 3654–3667

Vishnoi A, Srivastava A, Roy R and Bhattacharya A 2008 MGDD: *Mycobacterium tuberculosis* Genome Divergence Database; *BMC Genomics* **9** 373

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Muiper M and Zabeau M 1995 AFLP: a new concept for DNA fingerprinting; *Nucleic Acids Res.* **21** 4407–4414

Wehbi H, Rath A, Glibowicka M and Deber M D 2007 Role of the extracellular loop in the folding of a CFTR transmembrane helical hairpin; *Biochemistry* **46** 7099–7106

World Health Organization 2006 *Global tuberculosis control: surveillance, planning, financing* (Geneva: World Health Organization)

Wu X, Jornvall H, Berndt K and Oppermann U 2004 Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance; *Biochem. Biophys. Res. Commun.* **313** 89–96

*MS received 17 February 2009; accepted 28 July 2009*

*e*Publication: 20 August 2009

Corresponding editor: VIDYANAND NANJUNDIAH