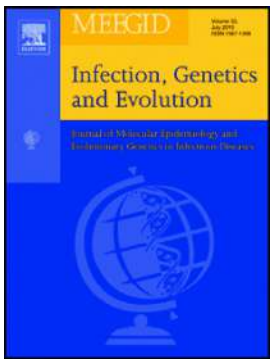


Intra-host genetic population diversity: Role in emergence and persistence of drug resistance among Mycobacterium tuberculosis complex minor variants

Carlos Arturo Vázquez-Chacón, Felipe de Jesús Rodríguez-Gaxiola, Alejandro Sánchez-Flores, Sarita Montaño, Ciresthel Bello-Rios, Salvador Fonseca-Coronado, Cruz Fernando López-Carrera, Armando Martínez-Guarneros, Ricardo Parra-Unda, Noemí García-Magallanes, Eliakym Arámbula-Meraz, Alejandro Escobar-Gutiérrez, Mayra Cruz-Rivera, Paúl Alexis López-Durán

PII: S1567-1348(22)00085-5

DOI: https://doi.org/10.1016/j.meegid.2022.105288

Reference: MEEGID 105288

To appear in: *Infection, Genetics and Evolution*

Received date: 20 July 2021

Revised date: 2 April 2022

Accepted date: 22 April 2022

Please cite this article as: C.A. Vázquez-Chacón, F. de Jesús Rodríguez-Gaxiola, A. Sánchez-Flores, et al., Intra-host genetic population diversity: Role in emergence and persistence of drug resistance among Mycobacterium tuberculosis complex minor variants, *Infection, Genetics and Evolution* (2021), https://doi.org/10.1016/j.meegid.2022.105288

Intra-host genetic population diversity: Role in emergence and persistence of drug resistance among *Mycobacterium tuberculosis* complex minor variants

Carlos Arturo Vázquez-Chacón[a,b,1], Felipe de Jesús Rodríguez-Gaxiola[c,1], Alejandro Sánchez-Flores[d], Sarita Montaño[c], Ciresthel Bello-Rios[e], Salvador Fonseca-Coronado[f], Cruz Fernando López-Carrera[g], Armando Martínez-Guarneros[b], Ricardo Parra-Unda[c], Noemí García-Magallanes[h], Eliakym Arámbula-Meraz[c], Alejandro Escobar-Gutiérrez[b], Mayra Cruz-Rivera[i] and Paúl Alexis López-Durán[j,*]

[a]*Facultad de Medicina y Cirugía, Universidad Autónoma Benito Juárez de Oaxaca, Oaxaca, México*

[b]*Instituto de Diagnóstico y Referencia Epidemiológicos, Ciudad de México, México*

[c]*Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Sinaloa, México*

[d]*Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico*

[e]*Laboratorio de Biomedicina Molecular, Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Guerrero, Chilpancingo, México*

[f]*Facultad de Estudios Superiores Cuautitlán, Universidad Nacional Autónoma de México, Estado de México, México*

[g]*Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Ciudad de México, México*

[h]*Laboratorio de Biomedicina y Biología Molecular, Universidad Politécnica de Sinaloa, Sinaloa, México*

[i]*Departamento de Microbiología y Parasitología. Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México*

[j]*Facultad de Ciencias de la Salud, Universidad Anáhuac, Campus Norte, Estado de México, México*

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis complex
Intra-host
Drug-resistance
Whole genome sequencing

ABSTRACT

Drug resistant tuberculosis (DR-TB) is an important public health issue in different parts of the world. *Mycobacterium tuberculosis* complex variants (MTBC vars) preferentially infect certain hosts, limiting their distribution to different ecosystems. However, MTBC vars can infect other hosts beyond their preferred target potentially contributing to persistence of drug resistance (DR) in other niches. Here, we performed a comprehensive intra-host genetic analysis for the identification of DR-related mutations among all MTBC minor vars whole genome sequences (8,095 strains) publicly available worldwide. High confidence drug-resistance mutations in *kat*G (isoniazid), *rps*L (streptomycin), *pnc*A (pyrazinamide), *rpo*B (rifampicin) and *gyr*A (fluoroquinolones) genes were identified among intrahost minor sub-populations in 197 different strains (2.43%) belonging to vars africanum, bovis, caprae, microti, orygis and pinnipedii. In addition, a three-dimensional structure modeling analysis to assess the role of novel mutations was also performed. Our findings highlight the importance of detecting discrete intra-host populations carrying DR mutations.

1. Introduction

Tuberculosis (TB) is a major public health problem in different geographical regions [44]. Drug resistant TB (DRTB) is an increasingly growing issue that significantly impacts therapy and overall patient's management [13]. DR-TB is often the result of improper treatment and/or regimen adherence. Occurrence of multi-drug (MDR) and extensively drug resistant (XDR) strains further jeopardizes the success of tuberculosis control programs aiming to limit spread of disease [41]. The World Health Organization has recently updated the definitions for DR-TB. Thus, MDR-TB is defined as strains exhibiting resistance to both isoniazid and rifampicin. In turn, pre-XDR-TB includes MDR strains which are rifampicin (RR-TB) and fluoroquinolone resistant [30]. XDR-TB is defined as MDR/RR-TB strains which are resistant to fluoroquinolones in addition to at least one group A drug [30]. Group A includes some of the most potent antibiotics for TB treatment such as levofloxacin, moxifloxacin, bedaquiline and linezolid [15, 24].

*Corresponding author paul.lopezdu@anahuac.mx (P.A. López-Durán)

ORCID(s): 0000-0002-2577-6599 (P.A. López-Durán) [1]These authors contributed equally to the work

MDR and XDR mechanisms of resistance are majority attributed to chromosomal mutations. Isoniazid is a prodrug that inhibits mycobacterial mycolic acid synthesis and the main gene involved is *kat*G. Other anti-tuberculosis drugs block mycobacterial protein synthesis by several means; rifampin binds to the beta subunit of RNA polymerase (coded by *rpo*B gene), fluoroquinolones bind to DNA gyrase (coded by *gyr*A gene), and streptomycin binds to 30s ribosomal protein (coded by *rps*L gene), whereas pyrazinamide mechanism of resistance is mediated by pyrazinamidase enzyme (coded by *pnc*A gene) and is not yet fully understood but requires an acidic pH with to kill mycobacteria [26].

TB is caused by a group of closely related but genetically distinct mycobacteria also known as the *Mycobacterium tuberculosis* complex (MTBC) [37]. MTBC groups 12 variants (vars) including *M. tuberculosis* vars tuberculosis (*sensu stricto*), africanum, bovis, caprae, microti, canettii, pinnipedii, mungi, orygis, suricattae, "dassie bacillus" and "chimpanzee

bacillus". MTBC host range is preferentially limited to certain hosts. As a result, MTBC vars tuberculosis *sensu stricto*, and africanum are primarily associated with human infections while bovis mostly infects bovines and deer [7]. MTBC var microti is generally observed among rodents. Other MTBC minor vars infect a wide range of hosts [1, 11, 17, 31, 37]. However, MTBC vars can and do infect other mammals beyond their preferred hosts. This is of importance since DR-TB strains could potentially be transmitted among hosts in different ecological niches; resulting in reservoir pockets that could participate in maintaining DR strains in a given geographical region.

Mycobacterial sub-populations within a single host are often clonal or displayed low genetic diversity [27, 36]. Thus, it is well accepted that consensus whole genome sequence (WGS) might be sufficient to identify DR patterns among isolates [28]. Consequently, mycobacteria are frequently sequenced at relative low depth limiting the detection of discrete traits among minor sub-populations [18]. Analyzing the architecture of the intra-host genetic population could potentially shed light on the micro-evolution [21, 38], emergence and persistence of DR strains. However, intrahost genetic diversity among minor MTBC vars (all members of the complex except for var *sensu stricto*) has been poorly investigated [4]. In TB, resistance-related mutations commonly emerge *via* intra-host evolution as a consequence of inadequate therapy or lack of adherence to treatment [4]. As a result, mycobacterial sub-populations within a given host have the potential to exhibit some degree of diversity due to the occurrence of the *novo* mutations or mixed infections [8, 23]. Heteroresistance (HR) is defined as the cooccurrence of sub-populations with increased antibiotic resistance in the presence of a susceptible main population in a clinical isolate [2]. HR is frequent among different bacteria, and it is particularly associated with treatment failure due to enrichment of low frequencies resistant sub-populations during the treatment course [14]. Heteroresistant subpopulations can both, revert to a susceptible phenotype or expand in the absence of the corresponding antibiotic [14]. HR has been previously reported among MTBC var *sensu stricto* [29]. However, the extent of HR among other MTBC vars is largely unknown.

WGS is rapidly becoming the standard method for advanced genetic characterization of TB strains [42]. Importantly, sequence depth has a significant impact in our ability to identify discrete traits among intra-host sub-populations [39]. Bacteria isolates are commonly sequenced at low depth (<100×), particularly when reference genomes are used for read mapping. This significantly limits the identification of discrete traits [18]. It is noteworthy that most algorithms overlook the intra-host minor sub-populations by exclusively reporting consensus sequences [27]. Additionally, state-of-the-art assembly algorithms tend to exclude genomic regions where variants occur, due to a low k-mer frequency, effectively preventing the identification of discrete traits. This is a major disadvantage that effectively prevents the identification of sub-populations carrying discrete traits such as DR mutations. Therefore, the development of frameworks for the detection of minor intra-host mycobacteria sub-populations harboring mutations conferring resistance is of importance. Such frameworks could potentially lead to the identification of novel DR mutations.

Thus, DR-TB among MTBC minor vars might play a critical role in the maintenance of a resistant phenotype in a given region. Therefore, identification of mutations conferring resistance to the first- and second-line drugs among intra-host sub-populations in MTBC minor vars could provide relevant information about the genetic mechanisms driving acquisition and fixation of drug resistant mutations among different MTBC lineages. Moreover, persistence of a resistant phenotype among non-human hosts in different ecological niches could potentially interfere with success of local TB control programs. Here, we performed a comprehensive intra-host genetic analysis including all minor MTBC WGS reported in public databases. Additionally, we carried out a three-dimensional structural analysis for all novel (not previously reported) mutations localized at the characteristic residue positions responsible for high confidence DR mutations. Our findings highlight the importance of implementing high resolution laboratory and bioinformatics approaches for the detection of discrete traits among intra-host mycobacteria sub-populations in MTBC vars.

2. Methods

2.1. MTBC WGS

All WGS (8,074 sequences) available from the Sequence Read Archive, National Center for Biotechnology Information were included in the study with the exception of strains belonging to MTBC var *sensu stricto*. Raw reads were processed as previously described [43]. Briefly, FASTQ files were obtained using the SRA Toolkit. Sequence read sequence quality control was assessed with FastQC v0.11.5 [3]. Fastp was used to remove adapters, short and poor-quality reads (Phred<25) [9]. Reads were scanned for the specific codons or nucleotide for high confidence mutations conferring resistance. If a given resistance-related mutation was found, the individual quality of the adjacent nucleotides was assessed. Thus, only high-quality sequence reads were analyzed. All strains meeting the inclusion criteria are shown on Suppl. Table 1. Post-processed samples were also analyzed with Res-finder [6], Point-finder [45] and TB-Profiler [33] tools. WGS were categorized accordingly to the infecting MTBC var. Only sequences with a minimum of 40X depth of coverage were included in the final analysis (8,047). BAM tools were used to generate the corresponding alignment files.

2.2. DR mutations

Three different strategies were used to identified DR mutations: a) TB-profiler [33]; b) Point-Finder [45]; and c) a custom approach to identify individual sequence reads carrying DR mutations. TB-profiler and Point-Finder were used to identify strains exhibiting major sub-populations (consensus) carrying DR mutations. Our custom approach uses the raw sequences of whole genomes obtained by massive sequencing to extract the regions in which mutations associated with high-confidence MTB resistance are found. These sequences also contain the PHRED scores to determine the probability of error for each nucleotide. Sequences are first filtered regarding the average PHRED quality greater than 20 and then according to the resistance determining codon. This script filters and discards those that do not meet the length and quality criteria. Furthermore, translates the codons to the respective reading frame and generates a table that allows us to determine the variability of changes in each categorical and numerical region, resulting in detecting the smallest variations at low sequencing depth. This method provides more information than Point-Finder or Res-finder since the identification of any amino acid change is related to the count number of changes of the original NGS reads in addition to the assemblies of the others which delete minor changes.

2.3. DR prediction analysis

Effect on amino acid variation for non-previously reported mutations was assessed with PROVEAN, SUSPECTRIF and SUSPECT-PZA. PROVEAN [10] uses an alignment-based scoring system to predict the effects of mutations, deletions, and insertions on proteins. SUSPECT-RIF [34] and SUSPECT-PZA [20] are predictive binary classifiers to interpret genomic variants in *rpo*B and *pnc*A genes respectively. Changes in alignment scores denote the impact of a mutation on protein function; results are then reported as deleterious or neutral or as a predicted resistance. All novel mutations (161) were analyzed using the consensus amino acid sequence as background for each protein (*gyr*A, *kat*G, *pnc*A, *rpo*B and *rps*L). Scores are depicted in Supp. table 7.

3. Results

3.1. MTBC minor vars

We defined MTBC minor vars as all members of the complex with the exception of var TB *sensu stricto*. Therefore, our analysis selectively included all MTBC WGS minor vars reported in the Sequence Read Archive database. The final data set comprises all 8,095 MTBC minor vars WGS publicly available until Feb 2021 including *M. tuberculosis* vars africanum, bovis, caprae, microti, canettii, pinnipedii, mungi, orygis, suricattae, "dassie bacillus" and "chimpanzee bacillus" (Suppl. Table 1, deposited on Mendeley at http://dx.doi.org/10.17632/zv2gx42cs7.1). Sequences were categorized by variant and host. The summary of all strains included in the study are shown in Table 1.

All MTBC var africanum strains reporting host information were recovered from human cases (751). MTBC var bovis isolates were recovered from badgers (199), bison (4), cattle (2,844), coyote (1), goat (2), deer (321), elephant (1), cat (7), hedgehog (1), human (134), capybara (1), llama (3), wild dogs (5), bobcat (1), weasels (118), opossum (11), sheep (1), chimpanzee (2), panthers (4), pig (74), racoon (14), rhinoceros (1), wild boar (11), buffalo (7), antelope (1), brushtail possum (113). Host information was not provided for 2,781 isolates and 7 isolates were reported as recovered from primates. Host information was not available for strains belonging to MTBC var microti except for one human isolate. The only MTBC var mungi strain included in the study was recovered from a mongoose. MTBC var caprae isolates were recovered from cattle (61), deer (4), Asian elephant (2), human (1), and rhinoceros (1). Three strains were reported as recovered from primates and one from monkey. Four MTBC var pinnipedii isolates were recovered from sea lions and eight lack host information. MTBC var orygis were obtained from gaur (1), cattle (14), deer (5), and humans (10). Eight isolates did not report host data. MTBC var d. bacillus were recovered from rock hyrax (3) and suricate (1). All MTBC var suricattae strains were obtained from suricate (4). All MTBC var canettii isolates (25) reporting host information were recovered from human cases; in addition to twenty two strains with no host information. The only isolate from MTBC var c. bacillus was obtained from a chimpanzee.

3.2. DR profile characterization

All WGS were assembled, and the corresponding DR pattern was inferred using our custom approach as well as TB-Profiler [33] and PointFinder [45]. We initially identify the consensus DR profile for all strains using all three strategies. Mutation in residue 57 (H⇒D) in the *pnc*A gene was not taken into consideration for the DR analysis for MTBC var *bovis* strains.

3.3. Consensus DR detection

DR mutation detection using our custom approach are listed in Suppl. Table 2. A total of 176 different DR strains belonging to MTBC vars africanum and bovis were found to exhibit at least one DR mutation, including 26 MDR isolates, 64 poli-resistant and 133 mono-resistance strains. MDR was exclusively limited to africanum strains.

3.3.1. Isoniazid DR-TB

Mutations associated with DR to isoniazid located in the *kat*G gene were exclusively observed among MTBC var africanum (69), with the exception of two bovis isolates. All 71 isolates carried the characteristic substitution in nucleotide position 944, corresponding to residue 315 in the catalase-peroxidase enzyme. The point mutation located in S315N, S315I and S315T, were observed in two, nine and sixty strain, respectively. We observed 11 MDR isolates among DR MTBC var africanum strains, 6 recovered from human cases and 5 with no host information.

3.3.2. Rifampicin DR-TB

Rifampicin DR was also primarily observed among africanum isolates (21) and one bovis strain. Mutations in the *rpo*B gene corresponded to amino acid changes in the protein located at D435V and S450L, with 8 and 14 strains, respectively. Eleven isolates exhibiting a MDR phenotype belonged to MTBC var africanum with 6 strains reporting human as host.

3.3.3. Fluoroquinolone DR-TB

Twenty-three fluoroquinolone DR strains (10 africanum and 14 bovis) showed mutations in the *gyr*A gene. These substitutions corresponded to residue A90, and D94, nine and fourteen mutations, respectively (Suppl. Table 2). All MTBC var africanum strains were recovered from human cases and showed the characteristic high confidence mutation in residue A90V with the exception of one strain that exhibited a mutation in residue D94G. MTBC var bovis strains showed mutations exclusively in D94G. None of the isolates exhibited a MDR phenotype.

3.3.4. Streptomycin DR-TB

Streptomycin resistance associated with mutations in the *rps*L gene was observed among 98 strains, 64 africanum and 34 bovis isolates. In the protein these mutations correspond to K43R and K88R, with 77 and 21 isolates, respectively. Three africanum strains were MDR, two of them recovered from human cases.

3.3.5. Pyrazinamide DR-TB

Isolates with mutations conferring resistance to pyrazinamide in gene *pnc*A belonged exclusively to MTBC var africanum. In the protein these mutations (Q10P, H57D and L172P) were recovered from two, five and one strain, respectively. No host information was available from any of these isolates. However, none of the strains exhibited a MDR phenotype. The characteristic mutation in H57D was absent in 14 bovis strains, resulting in a TB-like phenotype.

We compared the performance of our custom approach *versus* two state-of-the-art approaches, TB-Profiler and PointFinder, for the detection of DR-Tb. Our custom approach exclusively focused on high confidence mutation while both TB-Profiler and PointFinder are capable to identify more DR mutations (medium and low confidence). Only high confidence DR mutations were included in the comparative study since our custom approach selectively excludes all medium and low confidence DR mutations. High confidence mutations identified by TB-Profiler and PointFinder are listed in Suppl. Tables 3a and 4a, respectively. Figure 1 depicts the overlapping between all three methods.

Figure 1. TB-Profiler, Point-Finder and custom algorithm comparison. All three strategies are color coded (yellow; Custom algorithm, red; TB-Profiler and blue; Pointfinder). The overlapping regions represent the isolates that were classified identically by different methods.

Overall, our custom approach exhibited a good correlation with the data obtained with TB-Profiler and PointFinder. Discrepancies between our custom approach (10) and TB-Profiler and PointFinder were confirmed by inspecting all original files individually. All medium and low confidence mutations identified by TB-Profiler and Point-Finder are listed is Suppl. Tables 3b and 4b, respectively.

DR isolates were identified by TB-Profiler and PointFinder recovered from non-human hosts included cattle, elephants, and deer. Our custom approach identified DR mutations in intrahost sub-population in a wide range of non-human hosts (possum, cattle, deer, pig, sheep, raccoon, llama, and badger).

3.4 DR detection among minor sub-populations

We observed minor intra-host sub-populations carrying substitutions occurring in residues associated with DRTB. In total, 197 strains displaying minor sub-populations bearing, at least one, high confidence DR mutations were identified (Suppl. Table 5).

3.4.1. Isoniazid DR-TB

Twenty-one different strains exhibited minor sub-population with DR mutations conferring resistance to isoniazid. Three different mutation were observed in residue 315 (S315I, S315N and S315T). Both MTBC vars africanum and bovis isolates were recovered from human and non-human hosts (possum and cattle). *3.4.2. Rifampicin DR-TB*

Twenty-four isolates displaying intra-host sub-populations harboring mutations in the *rpo*B gene were identified. In the protein, four different residues mutations conferring high confidence resistance to rifampicin were observed (D516V, H526D, H526Y, and S531L). Strains belonged to MTBC vars africanum (2); bovis (18), caprae (1), orygis (2), and pinnipedii (1). Both africanum isolates were recovered from human cases while 12 bovis strains were obtained from cattle. The only caprae strain was also obtained from cattle. Three bovis strains were recovered from different non-human hosts [possum (1), deer (1) and raccoon (1)]. The orygis strains were recovered from human and cattle.

3.4.3. Fluoroquinolone DR-TB

Eighty-six isolates displaying resistance to fluoroquinolones in the *gyr*A gene were detected in minor sub-populations. Amino acid substitutions in the protein were located at A90V, S91P, S91P, D94Y, D94A, D94G, D94H and D94N. Mutations in residues A90, S91 and D94 were observed in 24, 14 and 47 strains, respectively. These strains primarily belonged to MTBC vars africanum (36), and bovis (44), however, other vars such as caprae (4) microti (1) and pinnipedii (1) were also identified. Human isolates were recovered from 19 africanum strains and one microti isolate. MTBC var bovis isolates were recovered from cattle (31), badger (1), possum (1), pig (1), sheep (1), raccoon (1), deer (1), and llama (1).

3.4.4. Streptomycin DR-TB

Thirty-five strains exhibiting DR mutation to streptomycin in the *rps*L gene among intra-host minor subpopulations were also observed. In the protein, mutations in residue K43 and K88 showed the characteristic K⇒R substitution in 16 and 19 isolates, respectively. All strains belonged to MTBC var bovis with the exception of four africanum isolates. All four africanum strains were recovered from human cases while 19 bovis isolates were obtained from cattle. Two bovis isolates were recovered from deer and one from a ferret.

3.4.5. Pyrazinamide DR-TB

MTBC var bovis strains carrying mutation H57D located in the *pnc*A gene were not considered for this analysis due to the intrinsic occurrence of this substitution in this lineage. Drug resistance to pyrazinamide was observed in 48 different strains. In the protein these mutations were located in residues Q10P, D12A, D12N, Y34D, D57H, H57D, and L172P. Strains carrying such mutations belonged to vars africanum (10), bovis (33), and caprae (4). Human isolates belonged to MTBC vars africanum (7) and bovis (1). Bovine strains were obtained from MTBC vars bovis (16) and caprae (4). Additionally, two bovis isolates were recovered from deer, one from a badger and one from a possum.

3.5. Novel mutations

Our custom approach also identified a large set of not previously reported mutations located in residues associated with DR-TB (Suppl. Table 6). In total, 707 different strains carrying these novel mutations were identified. Most novel mutations were observed among intra-host subpopulations. However, in some isolates, the corresponding mutation represented the predominant population. Importantly, none of the novel substitution identified by our custom approach were simultaneously identified by TB-Profiler or PointFinder. The PROVEAN index for each substitution was used as a

proxy for DR (protein function disruption) for the prediction of isoniazid, fluoroquinolone and streptomycin DR. Suspect Rif and PZA were used to predict rifampicin and pyrazinamide DR.

3.5.1. Genetic diversity in residues participating in DR-TB 3.5.1.1. Catalase-peroxidase

The *kat*G gene showed the lowest degree of genetic variability among all five DR-related genes analyzed. In the catalase-peroxidase KatG protein, six different substitutions in residue 315 among 11 different strains were observed in the protein. The characteristic serine found in wild type strains was replaced by three hydrophobic residues (A, C and M); in two, five and one strains respectively. Two hydrophilic (D and K), as well as one neutral (P) residue were observed in one strain each. Isoniazid drug resistance mediated by substitutions in residue 315 is highly susceptible to modifications reducing the size of the heme access pore as result of replacements with larger amino acids [16]. Amino acid volume showed a good correlation with PROVEAN scores [10]. Five substitutions resulted in larger amino acids likely reducing the pore size and effectively preventing drug access to the oxidizing site (Figure 2). On the other hand, the two strains carrying the smaller A at this position were not likely to reduce the size of the pore (sensitive). Seven isolates were recovered from non-human hosts; two deer strains and five from cattle. These finding could potentially represent an increase likelihood of DR-Tb in different ecological niches.

Figure 2. Novel amino acid substitutions in the *kat*G gene. Amino acid substitutions located in residue 315. PROVEAN indexes and hydropathy (Kyte-Doolitle scale) are depicted in the Y and X axis, while the bubble size represents the amino acid volume. Hydropathy scores are represented by colors: hydrophilic (blue), hydrophobic (yellow), neutral (red). All substitutions displayed deleterious PROVEAN indexes and larger volumes than serine (wild type). The dotted line depicts the PROVEAN threshold. Solid lines delimit the hydropathy zones.

3.5.1.2. RNA polymerase (β subunit)

The *rpo*B gene was the second most variable (Suppl. Table 6). The protein DNA-directed RNA polymerase subunit alpha presented a total of 34 different substitutions occurring in three residues (D435, H445 and S450) associated with high confidence DR were observed among 146 different strains. Residues H445 and S450 interact directly with rifampicin by establishing hydrogen bridges at the drug binding site (Figure 3). On the other hand, D435 participates in the binding process *via* van der Waals bonds [22]. Consequently, changes in these positions are likely to affect drug binding, resulting in DR. All identified novel amino acid substitutions rendered deleterious PROVEAN scores. SUSPECT-RIF also categorized all substitutions as probable DR mutations. Substitutions in these residues included a wide array of amino acids with very different physical-chemical properties (Figure 3, Suppl. Table 7). This suggest that minor changes in these three residues effectively alter the protein function, resulting in resistance to isoniazid. MTBC vars carrying these substitutions belonged to africanum (41) bovis (74), caprae (18), microti (1), orygis(1), and pinnipedii (11).

Figure 3. Genetic variability along the rifampicin binding site in the RNA polymerase β subunit (PDB: 5UHC). Amino acid substitutions located in residues 435, 445 and 450 are depicted. Dotted lines represent the distance in Angstroms. Rifampicin is shown in cyan.

3.5.1.3. DNA gyrase

Thirty-two different substitutions were identified along the DNA gyrase protein in residues A90, S91 and D94 (Suppl. Table 6). No specific DR predictive tools are available DNA gyrase for fluoroquinolones. The PROVEAN index was used as a proxy for DR for mutations along the *gyr*A gene. Nineteen substitutions were classified as deleterious by PROVEAN (Suppl. Table 7). Two deleterious substitutions (I and L), in A90 resulted in replacement for significantly more hydrophobic amino acids. Neutral or hydrophilic substitutions had no deleterious effect on position 90. Eight deleterious amino acid changes in S91 were identified. The deleterious effect observed in residue S91 was directly associated with amino acid size. Deleterious substitutions were related to larger amino acids (D, H, I, L, M, N, R, and Y) than the serine found in sensitive strains. This likely increase the stereochemistry effect resulting in DR. Interestingly, S91C rendered a neutral mutation likely due to the smaller size of cysteine which does not seem to prevent drug binding. All nine substitutions in D94 (C, E, I, L, P, Q, R, S, and T) were deleterious suggesting a strong structural restriction in this residue.

3.5.1.4. 30S ribosomal subunit

PROVEAN scores were used as proxy for DR since no specific algorithms for the prediction of DR associated with the *rps*L gen are available. All sixteen substitutions identified in Y43 (7) and Y88 (9) conferred a deleterious phenotype (Suppl.

Table 7). The physical-chemical properties of all 16 substitutions were significantly different. Therefore, the variability of the amino acids replacing the characteristic wild type K suggest that discrete changes have a significant impact in DR [12].

3.5.1.5. Pyrazinamidase

The *pnc*A gene showed the highest genetic variability. Seventy-one different novel substitution in six residues were detected in the protein pyrazinamidase (Figure 4A). Substitutions occurring in position 57 have been shown to play a critical role in DR [40, 32], since they directly affect the interaction with the Fe^{+2} ion. It is noteworthy that all resistance predicted substitutions occurred in residues not associated with drug or co-factor binding site. Most substitutions in residues 34 and 172 were not predicted as DR mutations. This suggests that these positions do not play a critical role in DR. All amino acid changes in positions 10, 12 and 14 resulted in DR (Suppl. Table 7). These residues seem to contribute to DR indirectly via interactions with amino acids associated with the co-factor (iron) binding site (Figure 4B). Indeed, both Q10 and D12 are intimately ligated to the catalytic triad [19].

Figure 4. Pirazinamidase novel mutations. Amino acid substitutions located in residues 10, 12, 14, 34, 57 and 172 are depicted. **A)** Non-deleterious substitutions are shown in red fonts. Hydropathy scores are color coded as in figure 2. Font sizes represent amino acid volumes **B)** 3D structural analysis (PDB: 3PL1) shows residues Q10, D12 and C14 (in orange) contributing to the deleterious effect observed in the pyrazinamidase. All three residues are distant from the characteristic H57 residue associated with DR. Likewise, all residues are not part of the drug binding site. Dotted lines represent the distance in Angstroms, iron ion are shown in green.

4. Discussion

Here we have shown the occurrence of DR mutations among intra-host sub-population in MTBC minor vars. Our results suggest the possible circulation of pre-emergent DR-Tb strains in different non-human hosts and niches. Identification of drug resistant MTBC vars other than *tuberculosis sensu stricto* is of importance since they can represent an alternative niche for the emergence and persistence of DR-TB in a given geographical region [7]. These findings highlight the importance of characterizing the intra-host mycobacteria population at sufficient depth to accurately detect minor sub-populations harboring DR mutations in a given host [21]. Additionally, we also report the occurrence of other novel mutations with possible implication in DR-TB among different members of the MTBC.

Our framework specifically identifies minor variants carrying discrete traits among mycobacteria intra-host populations. It has been shown recently that sequencing MTBC isolates at a higher depth (500–1000x) results in a much higher resolution allowing to link cases to the corresponding source [21]. This emphasizes the need to account for within host diversity into outbreak investigations as well as drug resistance detection and identification of other discrete traits. The implications of analyzing the intra-host genetic variability for other bacterial infectious agents of human interest, besides TB, is key to understand their molecular evolution. Detection of discrete variations within bacterial populations is determined by factors such as sequencing depth, pre-selection and isolation methods [27]. This is of importance since adequate representation of the bacteria population in a given sample could be misrepresented due to any of the aforementioned elements [39]. The possibility that other MTBC vars can contribute to maintain the circulation of tuberculosis lineages and resistance in other hosts and ecological niches warrants further research.

Our custom approach and analyzes exclusively focused on high confidence, "*bonafide*" DR mutations [25]. We specifically selected high confidence DR mutations to increase the possibility of identifying strains with high potential for DR. While not all relevant DR mutations (medium and low confidence mutations) were considered in our custom approach, the most relevant issue is still the ample circulation of different MTBC vars strains from non-human hosts with a discrete DR-TB phenotype. Our approach produced comparable results to TB-Profiler and PointFinder for the high confidence mutations included in the algorithm (Figure 1). Using all three algorithms provided the highest resolution for DR detection. Our approach was also much more astringent for the detection of DR mutations. This was purposely hard coded to warrant the identification of true DR strains. In consequence, we cannot rule out the presence of other strains carrying actual DR mutations. Possible errors in our pipeline cannot be ruled out either due to the occurrence of sequencing errors. However, attributing all variant calls to these phenomena is unlikely. Likewise, the occurrence of sequencing errors at the residues related to DR-TB; leading to the characteristic amino acid associated with the corresponding DR phenotype is highly unlikely. The development of a more sensitive and comprehensive tool for the identification of hetero-resistance among low frequency intra-host sub-populations is needed.

MDR strains identified by TB-Profiler and PointFinder belonged primarily to africanum with few MDR bovis isolates (Suppl. Tables 3a and 4a). However, the possibility that some of these strains also exhibit minor intra-host subpopulations carrying other DR mutations associated with MDR is worrisome. Importantly, no XDR strains were detected by TB-Profiler or PointFinder among these isolates. However, and similarly to MDR, the risk of having strains bearing intra-host sub-populations with a DR associated to XDR TB remains a concern.

WGS is frequently performed at insufficient depth to properly identify discrete traits. Under these circumstances, the consensus DNA sequence is composed of high frequency variant calls generated by the assembler. This effectively "masks" low frequency traits harbored by minor intra-host sub-populations. As a consequence, well known genome assembly strategies fail to report the occurrence of DR mutations in low frequency sub-populations. Therefore, alternative molecular approaches such as real time PCR-based technologies still remain valuable tools in the absence of adequate strategies for the identification of discrete traits. The advent of NGS platforms with unprecedented depth such as the NovaSeq platform are likely to facilitate the identification of discrete traits in the foreseeable future.

In our analysis, it is of critical importance to discriminate several non-human hosts, as it might represent a niche for the emergence; and consequently, persistence of DR MTBC variants which in turn could potentially impact local efforts to control TB. Indeed, drug resistance MTBC minor variants can infect humans despite the limitations imposed by the ecological niche in which some MTBC variants evolve
[5, 35].

Our results suggest that DR mutations occurring in low frequencies sub-populations can be maintain in different non-human hosts and niches. Discrete DR-Tb in low frequency populations could be expected in the absence of the drug due to the lack of an evolutionary pressure. However, and upon possible transmission to humans, low frequencies carrying DR mutations can rapidly become the most predominant variant upon start of treatment. The degree of transmissibility of these drug resistance minor variants remains to be assessed. However, the potential to become an alternative niche for emergence and persistence of drug resistance TB is worrisome. This is of key importance for TB control since these niches might potentially act as reservoirs for the emergence and persistence of DR, MDR or XDR-TB.

A large set of novel mutations were identified in this work. Our analysis showed that mutations occurring in residue 315 of the catalase-peroxidase enzyme are particularly important. Five of the substitutions detected here resulted in much larger amino acids in the access pore. The size of the pore is critical for drug access and binding; thus, is highly likely that most substitution will affect it since serine is a very small amino acid. We also identified a mutant carrying alanine instead. We hypothesized that this mutation does not confer DR due to the smaller size of alanine, which also produced a borderline deleterious PROVEAN index (Figure 2). All novel substitutions occurring in the RNA polymerase exhibited a possible DR phenotype (predicted). This suggest of high degree of structural restrictions where small modifications could significantly alter drug binding. On the other hand, substitutions in the DNA gyrase in residues A90 and S91 were size or hydropathy dependable, respectively. However, residue D94 showed a much higher degree of structural restrictions with minor modifications affecting drug binding and in consequence resulting in a DR phenotype. Similarly, the 30S ribosomal subunit also showed important restrictions that lead to a suspected deleterious phenotype in all amino acid substitutions found. The pyrazinamidase showed a high degree of genetic variability. However, several of these changes were neutral and did not have an effect on DR. Interestingly, most predicted DR substitutions were mapped to residues not directly involved with drug or co-factor binding. Interestingly, these modifications showed significant interaction among them and with other residues directly associated with co-factor binding (Figure 4). Whether these substitutions are the result of amino acid co-variation related to stabilizing the protein is unknown. The confirmation of these substitutions as *bonafide* DR mutation warrants further research.

An important limitation in our study is the number of MTBC vars other than africanum and bovis. We analyzed all WGS available in the Sequence Read Database. There is a need for a larger number of sequences belonging to all other MTBC vars. It is challenging to derived solid conclusions about those other MTBC vars with the available data. However, with the decreasing cost of sequencing, this limitation will be overcome with the availability of more data and the proposed framework will allow to explore the evolution of MTBC vars in more detailed.

In conclusion, our results highlight the importance of implementing robust molecular surveillance of MTBC minor vars. Frameworks aiming to identify discrete traits are critical for proper identification of drug resistance among different MTBC vars. Monitoring the emerging of DR strains recovered from non-human hosts among different niches is likely to shed light on the evolution of mycobacteria. This information is likely to provide information regarding maintenance of DR among different MTBC vars. Implementation of advanced characterization of mycobacteria isolates at adequate depth and analyzed

with sensitive tools for heteroresistance detection will aid in the understanding of the transmission dynamics exploited by members of the MTBC between animals and humans.

Acknowledgments

This work was supported by Fondo de Investigación de la Universidad Anáhuac México 2020 project #201935 and FESC-UNAM-PIAPI-2020-2018. Paul Alexis López-Durán and Cruz Fernando López-Carrera were supported by the Consejo Nacional de Ciencia y Tecnología awards 1007810 and 1076257, respectively.

References

[1] Alexander,K.A.,Laver,P.N.,Michel,A.L.,Williams,M.,vanHelden, P.D., Warren, R.M., Gey van Pittius, N.C., . Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi* 16, 1296–1299.
URL: http://wwwnc.cdc.gov/eid/article/16/8/10-0314_article.htm, doi:10.3201/eid1608.100314.

[2] Andersson, D.I., Nicoloff, H., Hjort, K., . Mechanisms and clinical relevance of bacterial heteroresistance 17, 479–496. Number: 1 Reporter: Nature Reviews Microbiology.

[3] Andrews, S., . FastQC: A quality control tool for high throughput sequence data URL: http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/.

[4] Bespyatykh, J., Shitikov, E., Bespiatykh, D., Guliaev, A., Klimina,
K., Veselovsky, V., Arapidi, G., Dogonadze, M., Zhuravlev, V., Ilina, E., Govorun, V., . Metabolic changes of mycobacterium tuberculosis during the anti-tuberculosis therapy 9, 131. URL: https://www.mdpi. com/2076-0817/9/2/131, doi:10.3390/pathogens9020131.

[5] Bhembe, N.L., Green, E., . Molecular epidemiological study of multidrug-resistant tuberculosis isolated from sputum samples in eastern cape, south africa 80, 104182. URL: https://linkinghub. elsevier.com/retrieve/pii/S1567134820300149, doi:10.1016/j.meegid.
2020.104182.

[6] Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B.B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett,
N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K., Amaro,
A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M., . ResFinder 4.0 for predictions of phenotypes from genotypes , dkaa345URL: https://academic.oup.com/jac/ advance-article/doi/10.1093/jac/dkaa345/5870997, doi:10.1093/jac/ dkaa345.

[7] Brites, D., Loiseau, C., Menardo, F., Borrell, S., Boniotti, M.B., Warren, R., Dippenaar, A., Parsons, S.D.C., Beisel, C., Behr, M.A., et al., 2018. A new phylogenetic framework for the animal-adapted mycobacterium tuberculosis complex. Frontiers in microbiology 9, 2820.

[8] Casali, N., Broda, A., Harris, S.R., Parkhill, J., Brown, T., Drobniewski, F., . Whole genome sequence analysis of a large isoniazidresistant tuberculosis outbreak in london: A retrospective observational study 13, e1002137. URL: https://dx.plos.org/10.1371/
journal.pmed.1002137, doi:10.1371/journal.pmed.1002137.

[9] Chen, S., Zhou, Y., Chen, Y., Gu, J., . fastp: an ultra-fast all-in-one FASTQ preprocessor 34, i884–i890. URL: https:
//academic.oup.com/bioinformatics/article/34/17/i884/5093234, doi:10.1093/bioinformatics/bty560.

[10] Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the functional effect of amino acid substitutions and indels. PloS one 7, e46688.

[11] Coscolla, M., Lewin, A., Metzger, S., Maetz-Rensing, K., Calvignac-Spencer, S., Nitsche, A., Dabrowski, P.W., Radonic, A., Niemann, S., Parkhill, J., . Novel mycobacterium tuberculosis complex isolate from a wild chimpanzee 19. URL: http:
//dx.doi.org/10.3201/eid1906.121012, doi:10.3201/eid1906.121012. reporter: Emerg Infect Dis.

[12] Demirci, H., Murphy, F., Murphy, E., Gregory, S.T., Dahlberg, A.E., Jogl, G., 2013. A structural basis for streptomycin-induced misreading of the genetic code. Nature communications 4, 1–.

[13] Dheda, K., Gumbo, T., Maartens, G., Dooley, K.E., Murray, M., Furin, J., Nardell, E.A., Warren, R.M., Esmail, A., Nardell, E., et al., 2019. The lancet respiratory medicine commission: 2019 update: epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable tuberculosis. The Lancet Respiratory Medicine 7, 820–826.

[14] El-Halfawy, O.M., Valvano, M.A., 2015. Antimicrobial heteroresistance: an emerging field in need of clarity. Clinical microbiology reviews 28, 191–207.

[15] Ginsburg, A.S., Grosset, J.H., Bishai, W.R., 2003. Fluoroquinolones, tuberculosis, and resistance. The Lancet infectious diseases 3, 432– 442.

[16] Hazlehurst, L., Hacker, M., 2009. Chapter 15 - drug resistance, in: Hacker, M., Messer, W., Bachmann, K. (Eds.), Pharmacology. Academic Press, San Diego, pp. 371–385. URL: https:
//www.sciencedirect.com/science/article/pii/B9780123695215000154, doi:https://doi.org/10.1016/B978-0-12-369521-5.00015-4.

[17] Huard, R.C., Fabre, M., De Haas, P., Lazzarini, L.C.O., Van Soolingen, D., Cousins, D., Ho, J.L., 2006. Novel genetic polymorphisms that further delineate the phylogeny of the mycobacterium tuberculosis complex. Journal of bacteriology 188, 4271–4287.

[18] Jajou, R., Kohl, T.A., Walker, T., Norman, A., Cirillo, D.M., Tagliani,
E., Niemann, S., de Neeling, A., Lillebaek, T., Anthony, R.M., et al., 2019. Towards standardisation: comparison of five whole genome sequencing (wgs) analysis pipelines for detection of epidemiologically linked tuberculosis cases. Eurosurveillance 24, 1900130.

[19] Junaid, M., Li, C.D., Li, J., Khan, A., Ali, S.S., Jamal, S.B., Saud, S., Ali, A., Wei, D.Q., . Structural insights of catalytic mechanism in mutant pyrazinamidase of mycobacterium tuberculosis 12, 1–4. Number: 1 Reporter: Journal of Biomolecular Structure and Dynamics.

[20] Karmakar, M., Rodrigues, C.H., Horan, K., Denholm, J.T., Ascher, D.B., 2020. Structure guided prediction of pyrazinamide resistance mutations in pnca. Scientific reports 10, 1–10.

[21] Lee, R.S., Proulx, J.F., McIntosh, F., Behr, M.A., Hanage, W.P., . Previously undetected super-spreading of mycobacterium tuberculosis revealedbydeepsequencing9,e53245. URL:https://elifesciences. org/articles/53245, doi:10.7554/eLife.53245.

[22] Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y.W., Li, S., Feng, Y., Zhang, Y., Mandal, S., Jiang, Y., Liu, S., Gigliotti, M., Talaue, M., Connell, N., Das, K., Arnold, E., Ebright, R.H., 2017. Structural basis of mycobacterium tuberculosis transcription and transcription inhibition. Molecular Cell 66, 169– 179.e8. URL: https://www.sciencedirect.com/science/article/ pii/S1097276517301685, doi:https://doi.org/10.1016/j.molcel.2017.03.001.

[23] Liu, Q., Via, L.E., Luo, T., Liang, L., Liu, X., Wu, S., Shen, Q., Wei, W., Ruan, X., Yuan, X., et al., 2015. Within patient microevolution of mycobacterium tuberculosis correlates with heterogeneous responses to treatment. Scientific reports 5, 1–8.

[24] Lopez, B., Siqueira de Oliveira, R., Pinhata, J.M., Chimara, E., Pacheco Ascencio, E., Puyén Guerra, Z.M., Wainmayer, I., Simboli, N., Del Granado, M., Palomino, J.C., et al., 2019. Bedaquiline and linezolid mic distributions and epidemiological cut-off values for mycobacterium tuberculosis in the latin american region. Journal of Antimicrobial Chemotherapy 74, 373–379.

[25] Miotto, P., Tessema, B., Tagliani, E., Chindelevitch, L., Starks, A.M., Emerson, C., Hanna, D., Kim, P.S., Liwski, R., Zignol, M., et al., 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in mycobacterium tuberculosis. European Respiratory Journal 50.

[26] Miotto,P.,Zhang,Y.,Cirillo,D.M.,Yam,W.C.,2018. Drugresistance mechanisms and drug susceptibility testing for tuberculosis. Respirology 23, 1098–1113.

[27] Morales-Arce, A.Y., Sabin, S.J., Stone, A.C., Jensen, J.D., . The population genomics of within-host mycobacterium tuberculosis URL: http://www.nature.com/articles/s41437-020-00377-7, doi:10.1038/s41437-020-00377-7.

[28] Okumura, K., Kato, M., Kirikae, T., Kayano, M., Miyoshi-Akiyama, T., 2015. Construction of a virtual mycobacterium tuberculosis consensus genome and its application to data from a next generation sequencer. BMC genomics 16, 1–17.

[29] Operario, D.J., Koeppel, A.F., Turner, S.D., Bao, Y., Pholwat, S., Banu, S., Foongladda, S., Mpagama, S., Gratz, J., Ogarkov, O., et al., 2017. Prevalence and extent of heteroresistance by next generation sequencing of multidrug-resistant tuberculosis. PLoS One 12, e0176522.

[30] Organization, W.H., et al., 2021. Meeting report of the who expert consultation on the definition of extensively drug-resistant tuberculosis, 27-29 october 2020 .

[31] Parsons, S., Drewe, J.A., van Pittius, N.C., Warren, R.M., van Helden, P.D., . Novel cause of tuberculosis in meerkats, south africa 19, 2004–2007. URL: http://dx.doi.org/10.3201/eid1912.130268, doi:10.3201/ eid1912.130268. number: 12 Reporter: Emerging Infectious Diseases.

[32] Petrella, S., Gelus-Zietal, N., Maudry, A., Laurans, C., Boudjelloul, R., Sougakoff, W., 2011. Crystal structure of the pyrazinamidase of mycobacterium tuberculosis: insights into natural and acquired resistance to pyrazinamide. PloS one 6, e15785.

[33] Phelan, J.E., O'Sullivan, D.M., Machado, D., Ramos, J., Oppong, Y.E.A., Campino, S., O'Grady, J., McNerney, R., Hibberd, M.L., Viveiros, M., Huggett, J.F., Clark, T.G., . Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs 11, 41. URL: https://genomemedicine.biomedcentral.com/articles/10.1186/ s13073-019-0650-x, doi:10.1186/s13073-019-0650-x.

[34] Portelli, S., Myung, Y., Furnham, N., Vedithi, S.C., Pires, D.E., Ascher, D.B., 2020. Prediction of rifampicin resistance beyond the rrdr using structure-based machine learning approaches. Scientific reports 10, 1–13.

[35] Prodinger, W.M., Indra, A., Koksalan, O.K., Kilicaslan, Z., Richter, E., . *Mycobacterium caprae* infection in humans 12, 1501–1513. URL: http://www.tandfonline.com/doi/full/10.1586/14787210.2014.974560, doi:10.1586/14787210.2014.974560.

[36] Pérez-Lago, L., Comas, I., Navarro, Y., González-Candelas, F., Herranz, M., Bouza, E., García-de Viedma, D., . Whole genome sequencing analysis of intrapatient microevolution in mycobacterium tuberculosis: Potential impact on the inference of tuberculosis transmission 209, 98–108. URL: https://academic.oup.com/jid/article-lookup/ doi/10.1093/infdis/jit439, doi:10.1093/infdis/jit439.

[37] Riojas, M.A., McGough, K.J., Rider-Riojas, C.J., Rastogi, N., Hazbón, M.H., . Phylogenomic analysis of the species of the mycobacterium tuberculosis complex demonstrates that mycobacterium africanum, mycobacterium bovis, mycobacterium caprae, mycobacterium microti and mycobacterium pinnipedii are normal heterotypic synonyms of mycobacterium tuberculosis 68, 324–332. URL: http://ijs.microbiologyresearch.org/content/journal/ijsem/ 10.1099/ijsem.0.002507, doi:10.1099/ijsem.0.002507. number: 1 Reporter: International Journal of Systematic and Evolutionary Microbiology.

[38] Roger, V., Luca, F., MMaximillian, M., Elaine, E.L., MMelissa, S., Irina, O., David, D., Michael, S., Max, S., Reda, F.M., . In-host population dynamics of mycobacterium tuberculosis complex during active disease 1, 10:e61805. Number: 1 Reporter: eLife.

[39] Russell, S.L., Cavanaugh, C.M., . Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes 34, 2747–2761. URL: http://academic.oup.com/mbe/article/34/11/2747/3925276, doi:10.1093/molbev/msx188.

[40] Sreevatsan, S., Pan, X., Zhang, Y., Kreiswirth, B.N., Musser, J.M., 1997. Mutations associated with pyrazinamide resistance in pnca of mycobacterium tuberculosis complex organisms. Antimicrobial agents and chemotherapy 41, 636–640.

[41] Tiberi, S., Zumla, A., Migliori, G.B., 2019. Multidrug and extensively drug-resistant tuberculosis: epidemiology, clinical features, management and treatment. Infectious Disease Clinics 33, 1063–1085.

[42] Van Der Werf, M.J., Ködmön, C., 2019. Whole-genome sequencing as tool for investigating international tuberculosis outbreaks: a systematic review. Frontiers in public health 7, 87.

[43] Vázquez-Chacón, C.A., Rodríguez-Gaxiola, F.d.J., López-Carrera, C.F., Cruz-Rivera, M., Martínez-Guarneros, A., Parra-Unda, R., Arámbula-Meraz, E., Fonseca-Coronado, S., Vaughan, G., LópezDurán, P.A., 2021. Identification of drug resistance mutations among mycobacterium bovis lineages in the americas. PLOS Neglected Tropical Diseases 15, e0009145.

[44] WHO, . Global tuberculosis report, 2020, geneva: World health organization; 2020 .

[45] Zankari, E., Allesøe, R., Joensen, K.G., Cavaco, L.M., Lund, O., Aarestrup, F.M., 2017. Pointfinder: a novel web tool for wgs-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. Journal of Antimicrobial Chemotherapy 72, 2764–2768.

Table 1. MTBC minor vars isolates characteristics.

Var	Host	Total number
africanum	H	1,232
bovis	H (134); NH (6,535)	6,669
caprae	H (1); NH (72)	73
microti	H (1); NH (13)	14
canettii	H	47
pinnipedii	NH	12
mungis	NH	1
orygis	H (10); NH (28)	38
suricattae	NH	4
d. bacillus	NH	4
c. bacillus	NH	1

H, human host; **NH**, non-human host.

Conflict of Interest and Authorship Conformation Form

Please check the following as appropriate:

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript:

Author's name	Affiliation
Carlos A. Vázquez-Chacón	Facultad de Medicina y Cirugía, Universidad Autónoma Benito Juárez de Oaxaca, Oaxaca, México
Felipe de J. Rodríguez-Gaxiola	Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Sinaloa, México
Alejandro Sánchez-Flores	*Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico*
Sarita Montaño	Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Sinaloa, México
Ciresthel Bello-Rios	Laboratorio de Biomedicina Molecular, Facultad de Ciencias Químico-Biológicas, Universidad Autonóma de Guerrero, Chilpancingo, México
Salvador Fonseca-Coronado	Facultad de Estudios Superiores Cuautitlán, Universidad Nacional Autónoma de México, Estado de México, México
Cruz Fernando López-Carrera	Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Ciudad de México, México
Armando Martínez-Guarneros	Instituto de Diagnóstico y Referencia Epidemiológicos, Ciudad de México, México
Ricardo Parra-Unda	Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Sinaloa, México
Noemí García-Magallanes	Laboratorio de Biomedicina y Biología Molecular, Universidad Politécnica de Sinaloa, Sinaloa, México
Eliakym Arámbula-Meraz	Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Sinaloa, México
Alejandro Escobar-Gutiérrez	Instituto de Diagnóstico y Referencia Epidemiológicos, Ciudad de México, México
Mayra Cruz-Rivera	Departamento de Microbiología y Parasitología. Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México
Paúl Alexis López-	Facultad de Ciencias de la Salud, Universidad Anáhuac,

Durán	Campus Norte, Estado de México, México

Carlos Arturo Vázquez-Chacón: Conceptualization, Methodology, Investigation and Writing - Original Draft, **Felipe de Jesús Rodríguez-Gaxiola**: Methodology, Software and Data Curation, **Alejandro Sánchez-Flores**: Supervision, **Sarita Montaño**: Visualization, **Ciresthel Bello-Rios**: Visualization, **Salvador Fonseca-Coronado**: Writing - Review & Editing, **Cruz Fernando López-Carrera**: Methodology, Software, Data Curation and Visualization, **Armando Martínez-Guarneros**: Supervision and Investigation, **Ricardo Parra-Unda**: Investigation, **Noemí García-Magallanes**: Investigation, Formal analysis and Methodology, **Eliakym Arámbula-Meraz**: Investigation , **Alejandro Escobar-Gutiérrez**: Investigation and Writing - Review & Editing, **Mayra Cruz-Rivera**: Project administration, **Paúl Alexis López-Durán**: Project administration, Funding acquisition, Visualization, Formal analysis and Writing - Review & Editing.

Highlights

Intrahost genetic population diversity: Role in emergence and persistence of drug resistance among *Mycobacterium tuberculosis* complex minor variants

Carlos Arturo Vázquez-Chacón, Felipe de Jesús Rodríguez-Gaxiola, Alejandro Sánchez-Flores, Sarita Montaño, Ciresthel Bello-Rios, Salvador Fonseca-Coronado, Cruz Fernando López-Carrera, Armando Martínez-Guarneros, Ricardo Parra-Unda, Noemí García-Magallanes, Eliakym Arámbula-Meraz, Alejandro Escobar-Gutiérrez, Mayra Cruz-Rivera, Paúl Alexis López-Durán

- MTBC minor variants infecting non-human hosts harbor drug resistance mutations
- Intrahost bacterial populations aid drug resistance mutation emergence and persistence
- Insufficient sequence depth and inadequate algorithms contribute to neglect the existence of drug resistance mutations

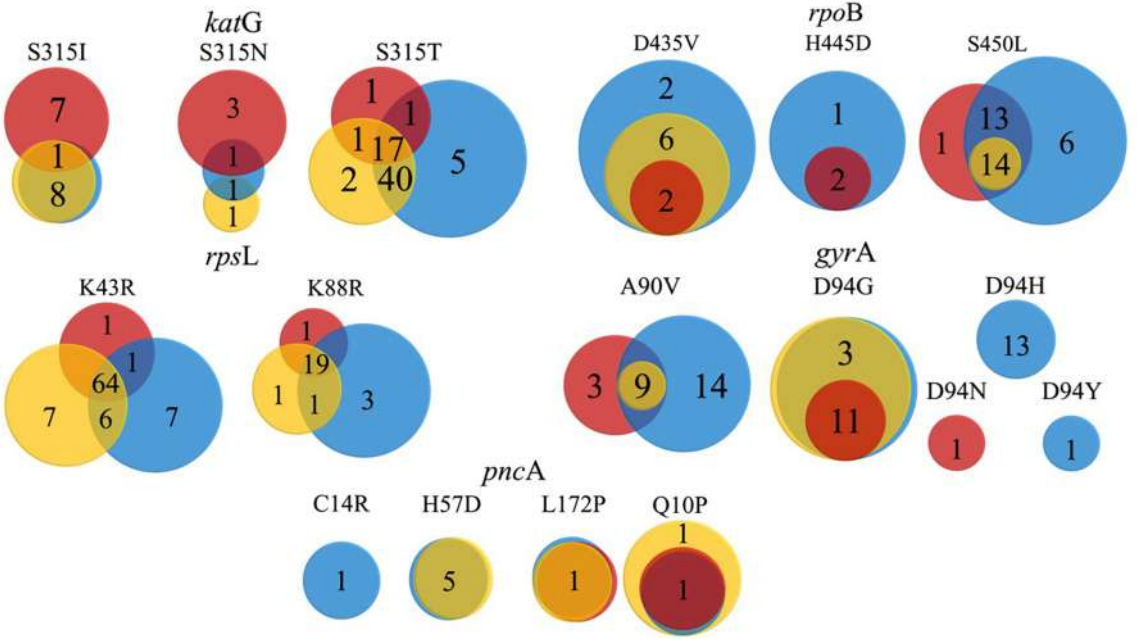

katG
S315I
S315N
S315T
7
1
8
3
1
1
1
1
17
2
40
5
rpoB
D435V
H445D
S450L
2
6
2
1
2
1
13
14
6
rpsL
K43R
K88R
1
1
64
7
6
7
1
19
1
1
3
gyrA
A90V
D94G
D94H
3
9
14
3
11
13
D94N
D94Y
1
1
pncA
C14R
H57D
L172P
Q10P
1
5
1
1
1

Figure 1

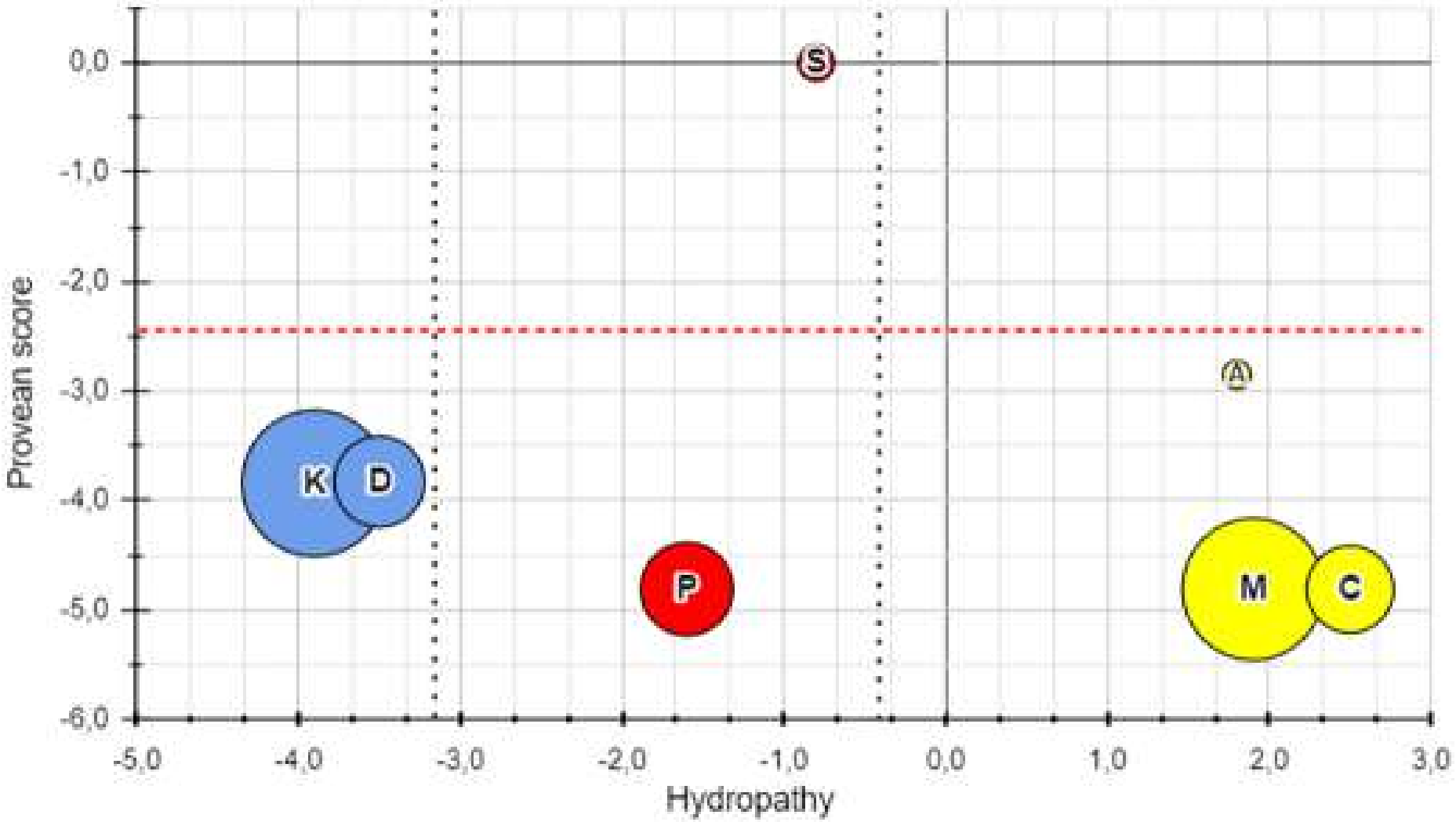

0,0
-1,0
-2,0
-3,0
-4,0
-5,0
-6,0
Prowean score
S
A
K
D
P
M
C
-5,0
-4,0
-3,0
-2,0
-1,0
0,0
1,0
2,0
3,0
Hydropathy

Figure 2

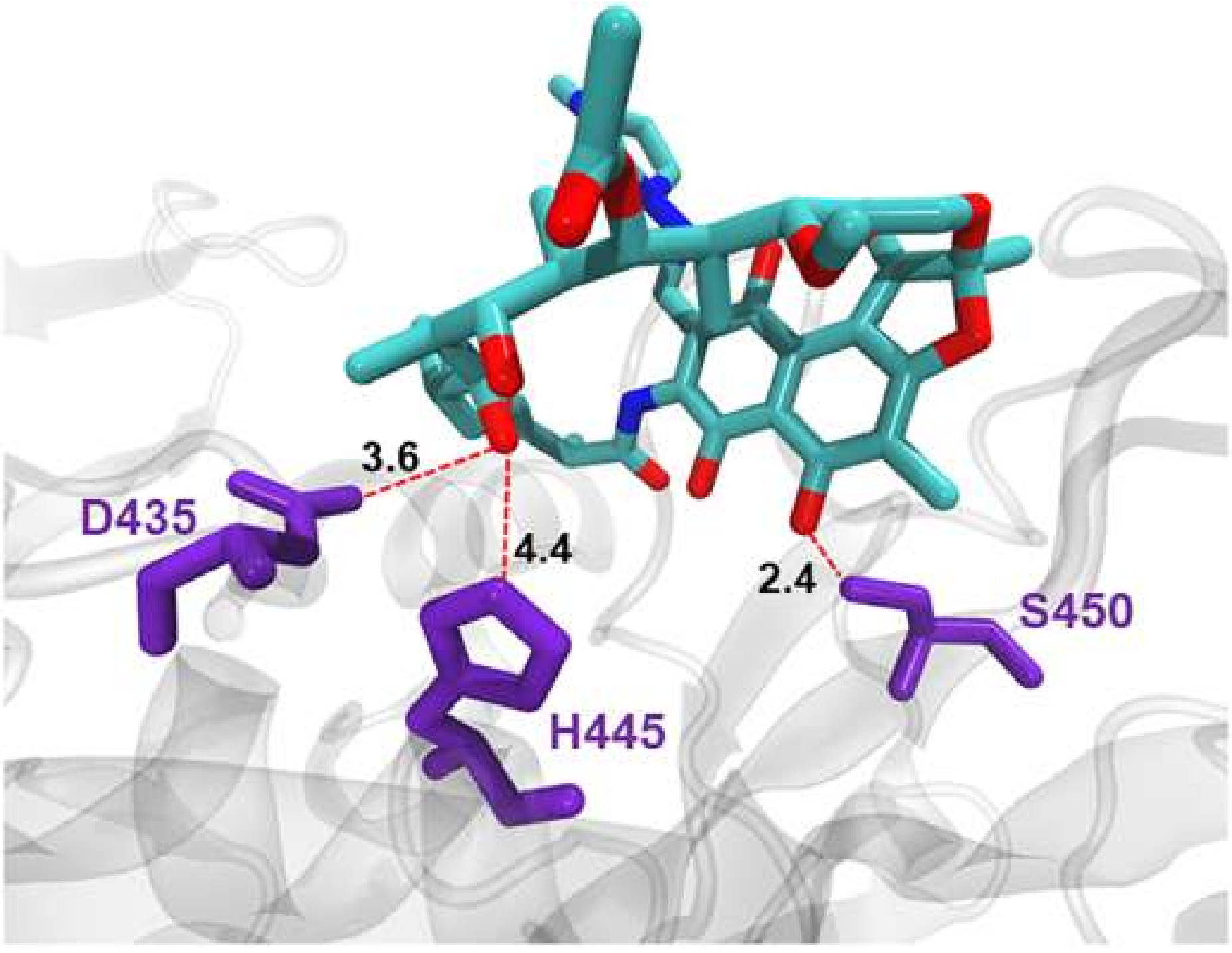
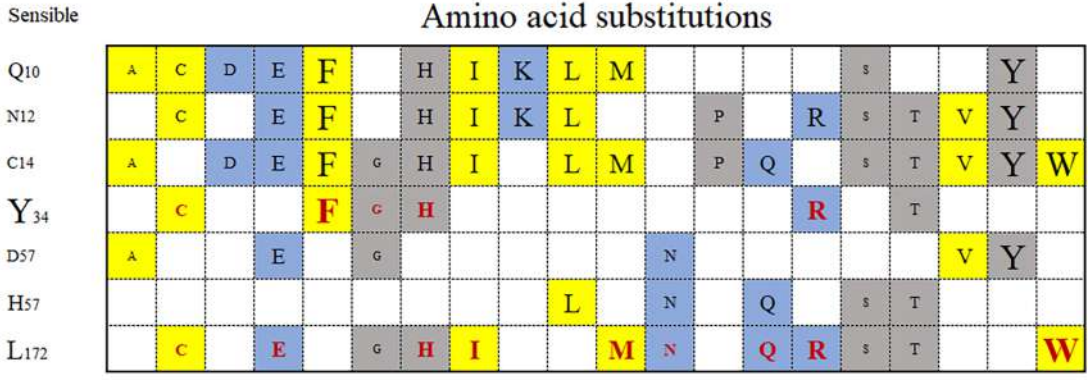


Figure 3

A


Sensible
Amino acid substitutions
Q10
N12
C14
Y34
D57
H57
L172

B

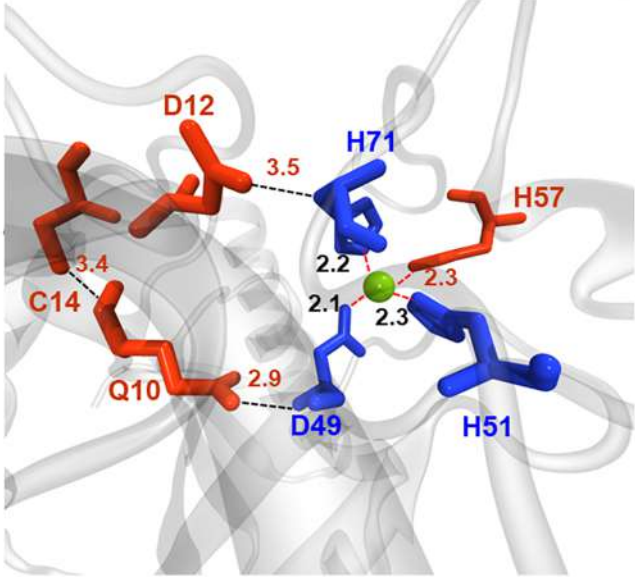

D12
H71
3.5
H57
2.2
2.3
3.4
C14
2.1
2.3
Q10
2.9
D49
H51

Figure 4