FAQs

Part3 - Evolutionary epidemiology: using phylogenetics to understand DR emergence and Mtb transmission

Q: Are there cases where bacteria are phenotypically resistant but genotypically susceptible after WGS/genotypic diagnostic test (e.g line probe assays) ? If yes, what therefore could be the cause of the phenotypic resistance?

A: Yes, and there are different possible reasons. The most common reason is that **our knowledge regarding the genetic determinants of drug resistance is limited**. We do not know all possible mechanisms and mutations of drug resistance and thus we are not able to detect 100% of them. As more data is generated and analyzed, we expand our knowledge regarding such determinants, and therefore our accuracy in predicting DR increases. We also have to take into account that we have less data regarding second-line treatments with novel drugs, so the search for drug resistance determinants is an active area of research.

Another possible reason is that bacteria can be genotypically susceptible, but survive antibiotic treatment in a process known as **drug tolerance**. For example, bacteria can switch to a dormant metabolic state in which they can survive in the presence of antibiotics, and resume growth once no more antibiotics are present in the environment.

A less probable (yet possible) scenario is that the disagreement is due to **heteroresistance**. Heteroresistance happens when both susceptible and resistant populations coexist within a patient. If this is true, it may happen that the strain that has been phenotypically tested for resistance is not the same that has been sequenced. For example, one could use different cultures of the same isolate for determining phenotypic resistance (e.g in an automated plate-reading

system) and grow bacteria for DNA extraction and sequencing (e.g in solid media).

Q: How Plasmid and horizontal gene transfer make DR prediction difficult for bacteria other than MTB?

Horizontal gene transfer (HGT) does not make DR prediction particularly more difficult, but can complicate studying how DR evolves and spreads. This is because **HGT can confound phylogenetic analysis,** given that some parts of the genome can originate from different bacterial species that evolve under different evolutionary constraints.

Q: How can we explain the fact that a mutation in a drug resistance-associated gene can not cause DR?

When mutations cause DR it is because they alter one or more of the cell components that are involved in the antibiotic mechanism of action. For example, rifampicin binds to the beta subunit of the RNA polymerase, blocking its activity and therefore leading to cell death. Mutations like rpoB S450L, modify the structure of the protein in such a way that rifampicin cannot bind to the beta subunit anymore, *while the RNA polymerase is still working*. In other cases, mutations do not lead to DR. For instance, **synonymous mutations** that do not alter how rifampicin binds to the beta subunit. If the drug target is an essential component, **it can also happen that a mutation alters the protein structure in such a way that is not functional anymore** (and the bacteria dies). Also it is important to bear in mind that **different mutations can alter the binding affinity of the antibiotic to a greater or lower extent**. Different binding affinities can explain why different mutations can confer either high- or low-level of drug resistance.

Q: Not sure how/if sequencing can be used to determine if a mutation is phylogenetic or not. How to tell from analysis whether a mutation is phylogenetic or causes intrinsic resistance. Still not clear to me

We first need to highlight (again) that "phylogenetic mutations" is a confusing term. By phylogenetic mutations we mean mutations that define groups. We are interested in these mutations when studying evolution of DR, because we want

to distinguish mutations that appear as response to antibiotic pressure, as opposed to mutations that are simply inherited from the parent strain to the offspring in the absence of antibiotic. For example, if we detect that a *M. bovis* strain that we are analyzing carries a mutation in pncA conferring pyrazinamide resistance, we can conclude that this strain is, indeed, resistant. However, we know that the reason it is resistant is not due to antibiotic selection, but just to the fact that this mutation appeared long time ago (before the antibiotic era) in the ancestor of *M. bovis*, and therefore all extant *M. bovis* strains carry this mutation and are pyrazinamide resistant. As another example, imagine you are analyzing a large dataset of MTB strains and you found that 60% of them carry the snp 497491 T>A. You might be tempted to think that something special is going on with this mutation for it to be so frequent in your population. However, this is a phylogenetic SNP that all lineage 2 strains carry. At least, now you know that 60% of the strains you are analyzing are lineage 2.

In a regular sequencing experiment focused on diagnosis/surveillance, **we do not aim to identify/define new phylogenetic mutations**. **We rather use pre-existing knowledge** about which mutations are actually phylogenetic. And the way we use sequencing to expand our knowledge about phylogenetic mutations is by performing phylogenetic analysis of diverse datasets of genomes. We study their evolution and which mutations define particular groups and when these mutations appeared. For more information about phylogenetic mutations in MTB (including catalogs defining groups), you can start reading the following references (10.1038/ncomms5812, 10.1186/s13073-020-00726-5, 10.1038/s41598-017-10018-5)

Q: How frequent is it that the exact same DR-mutation occurs independently? I.e. does drug pressure tend to induce mutations on certain loci more frequently?

When the same biological trait occurs several times independently during evolution,we call it **homoplasy**. Homoplasy is normally the result of selective forces that guide the evolution towards particular adaptations. **Antibiotic treatment is a very strong selective pressure.** Within a bacterial population, under antibiotic treatment, cells that acquire DR-conferring mutations will be able to survive, whereas those without such mutations will die. There may be **several mutations that confer DR to the same antibiotic, but the repertoire is**

limited. Because there is no infinite repertoire of mutations that confer resistance to, let's say, rifampicin, we always see the same mutations across the globe in rifampicin-resistant strains. Additionally, some of these mutations confer higher or lower levels of drug resistance, and carry a higher or lower replicative fitness cost. For this reason, some mutations can be clinically more or less frequent. For example, rpoB S450L is by far the most frequent rifampicin resistance-conferring mutation. Its high clinical frequency is probably because it confers a high-level of drug resistance with a low fitness cost (10.1126/science.1124410).

Q: So does it mean there is no standard SNP threshold for defining an epidemiologically linked cluster?

There is indeed no standard SNP threshold in which the whole scientific community agrees. However, we have data that supports the use of some thresholds that work reasonably well (https://doi.org/10.1016/S1473-3099(12)70277-3). Normally these thresholds are 5 to 12 SNPs. There is no "correct" way of choosing thresholds, but there are some considerations that may help in making a decision. For example, the higher the threshold that we use the older the transmission events that we will be including. In low-burden settings normally higher thresholds (10 - 15 SNPs) are used.

Q: Does it mean only SNPS difference can be considered for threshold determination, what about others like indels?

A: Indels or other variants could also be used for determining the degree of genetic relationships but that is not commonly done and there is no benchmarking as for SNPs.