A Proposal to Compare the Genomes of MDR and Susceptible *M. africanum* and *M. tuberculosis*

It is estimated that one third of the world’s population is infected with *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), causing 8.7 million new cases and 1.4 million deaths each year (World Health Organization, 2012a). Current treatment is a minimum of six months, and drug resistance is on the rise, with an estimated 440,000 cases of multi-drug resistant-TB (MDR-TB), which is TB resistant to the most effective anti-TB drugs (isoniazid and rifampicin). Thus, improved understanding of this dangerous pathogen is required to control its spread. Despite new advances in next generation sequencing, the number of whole genome *M. tuberculosis* sequences remains low. For example, Mali has an incidence of 62 cases per 100,000 people, with an estimated 140 cases of MDR-TB (World Health Organization, 2012b). However, to our knowledge, no one has sequenced isolates from this country.

Through a collaboration with SEREFO, a research center affiliated with the University of Bamako, we have a collection of 93 clinical isolates from patients from Bamako, Mali. These isolates come from two clinical trials conducted by SEREFO, and have had spoligotyping and drug susceptibility tests performed. Furthermore, between these two studies we have a mix of new cases and treatment failure cases. Of these isolates, we selected all of their MDR isolates and then identified the non-MDR isolates that most closely matched each MDR isolate in terms of spoligotype and patient history. It is of note that extremely drug resistant TB (XDR-TB) has not yet been detected in Mali, though testing for it is generally not done. In addition, we selected all of the non-*M. tuberculosis* strains available, which included a number of *M. africanum* strains, and one *M. bovis* BCG.

One major goal in using whole genome sequencing of our clinical isolates is to further our understanding of drug resistance in TB. First, we are interested in whether there are any mutations outside of the classical resistance mutations that correlate with drug resistance. One hypothesis is that drug resistance mutations cause a fitness defect, and in order for drug resistant strains to compete, there must be a compensatory mutation that overcomes this fitness defect (Gagneux et al., 2006). For example, it has recently been suggested that mutations in *rpoB*, which confer resistance to rifampicin, cause a growth defect that is compensated by mutations in *rpoC* (de Vos et al., 2012). We are interested in whether there are any other such mutations. It is possible that these compensatory mutations could serve as biomarkers of drug resistance, and thus could be used to improve the diagnosis of TB. The fact that the collection contains a number of different TB lineages will help ensure that findings are true across a range of genetic backgrounds. Furthermore, with this spoligotyping data, we know that these *M. tuberculosis* isolates are not clonal populations that originated from a single patient, so we expect to see a variety of mutations.

Secondly, we would like to use these sequences to identify novel resistance-conferring mutations. For some drug resistant isolates, we still do not know the genetic
cause of resistance. Even for rifampicin, for which the mechanism of resistance is relatively well characterized as being caused by mutations in the rpoB gene (which is used by Gene Xpert to diagnose drug resistant TB), there are an estimated 5% of rifampicin-resistant isolates that lack an rpoB mutation, and thus have an unknown mechanism of resistance. For all isolates that are resistant to any drug, we would like to identify any classical resistance mutations present, and for those sequences that lack any of the known mutations, we are interested in using statistical analyses of this large collection to identify any potential previously undiscovered resistance mutations.

Finally, this collection is unique due to the large number of *M. africanum* isolates. *M. africanum* is a close relative of *M. tuberculosis* that causes a disease that is clinically similar to *M. tuberculosis*, but is generally only found in West African patients (de Jong, 2010). *M. africanum* is classified into two lineages, type 1 and type 2. We have isolates from both of these lineages, though the majority of our isolates are type 2, including two MDR type 2 isolates. Another major goal of this project is to explore the genome structure of *M. africanum* and how it compares to *M. tuberculosis*. Very little research has been done on this organism, though the genome has been completed for type 2. The publication of the whole genome revealed many interesting characteristics unique to *M. africanum*, including the deletion of a portion of the genome when compared to *M. tuberculosis* (Bentley et al., 2012). Increasing the number of *M. africanum* genomes available could provide a number of valuable insights into this organism, and may even help explain why it has such a limited host range compared to *M. tuberculosis*. In addition, the similarities between *M. africanum* and *M. tuberculosis* could help identify critical genes to mycobacterial pathogenesis, which could serve as novel drug targets. This would be particularly interesting to study with our collection because we have such a diverse range of TB lineages, as well as both *M. africanum* lineages.

Thus, we hope to use whole genome sequencing of our collection of Malian isolates to address drug resistance and learn more about the different *M. tuberculosis* and *M. africanum* lineages. The variety of lineages, including lineages such as T1 that have not to our knowledge been sequenced, will greatly enhance our ability to answer these questions. Furthermore, the large number of *M. africanum* isolates gives us a rare opportunity to analyze this neglected pathogen. Thus, we have a diverse collection of isolates whose sequencing data could be used to analyze a range of questions.

References

