

Project Title: Characterization of *M. tuberculosis* Resistance Through Whole Genome Sequence Analysis to Support Diagnostic Development, Drug Discovery, and Molecular Epidemiology

Background/Public Health Importance

Tuberculosis (TB), the infectious disease caused by the bacillus *Mycobacterium tuberculosis* (MTB), remains a significant global public health threat. TB is the worldwide leading cause of death from a bacterial disease and the second leading cause of death from an infectious disease after Human Immunodeficiency Virus (HIV) infections. Approximately one third of the world's population, 1.8 billion people, is infected with MTB. In the year 2010 alone, there were 8.8 million incident TB cases and nearly 1.5 million deaths associated with this disease as well as an estimated 650,000 cases of multidrug resistant TB (MDRTB). New data suggest that this was likely an underestimate and that the incidence of MDRTB and extensive drug resistant TB (XDRTB) is higher than previously expected and increasing. In addition, only less than one half of one percent of new MDRTB cases are projected to have been treated. Those that are treated have not received the appropriate antibiotics to effectively counter the disease. Finally, drug resistant strains are spreading from human-to-human at alarming rates, suggesting that the 'fitness costs' long thought to hobble drug resistant MTB's transmission are no longer a factor, likely due to acquisition of compensatory mutations. In fact, person-to-person transmission appears to be the dominant route to new cases of MDR and XDR TB.

In view of these new realities, there is a need for an extreme sense of urgency. It is projected that 30 to 40 percent of XDRTB cases are untreatable, and in some regions of South Africa 88 percent of strains are resistant to all TB drugs available in that country. While a cocktail of three to four new antibiotics are required to successfully treat these and other totally drug resistant strains (TDRTB), the reality is that any single new antibiotic may likely be a decade away from regulatory approval. Currently, there are no consistent policies to deal with patients whose TB is untreatable. Proof that disease in these patients is untreatable may take months, during which time they may spread their resistant organisms to family members and others in the community, including health care workers. Thus, there is an urgent need to intensify efforts to develop a true point of care diagnostic that does not require laboratory support and, ideally, can determine the susceptibility to all available TB drugs in each geography. There is also an urgent need to identify new drug targets to speed the development of new antibiotics.

Unfortunately, the MTB genome remains poorly characterized and valuable MTB genomic information that could drive new diagnostics and therapeutics is unavailable. While the whole genome of over 70 TB strains have been sequenced, only 12 of these are reported as MDR/XDR strains. The proposed collaborative sequencing project is designed to close the gap on critical deficiencies in data for the MTB genome with emphasis upon MDR, XDR, and TDR strains.

Project Aim

The overarching goal for this work is to obtain full genome sequence data from well characterized DR, MDR, XDR, and TDR TB strains to understand the genomic patterns of drug resistance in support of (1) predictive analytics platforms and services to support patient treatment, (2) development of new diagnostics, (3) new drug discovery and (4) molecular epidemiology for population control programs including drug supply chain management. Progress in any one of these four areas will have an immediate and positive impact on the global TB burden by providing clinicians with real-time information and better treatment options for new cases of TB.

Key to describing genomic resistance patterns in MTB is the generation of high quality genome sequence from clinical isolates that have undergone rigorous and systematic clinical laboratory drug sensitivity testing (DST). To date, these essential data are not available. A review of several public domain repositories (including GenBank, TBDB and PATRIC as well as the Genome Sequencing Centers at the Broad Institute and the Sanger Center) revealed that only a few of the more than 70 MTB genomes sequenced, assembled and placed into the public domain, are from drug resistant strains (Mardassi, 2005; Ioerger, 2009; Ioerger, 2011; Jassal, 2009; Sandegren, 2011, Zhang, 2011). Furthermore, many publicly available MTB genomes are lacking significant meta-data (both clinical and phenotypic), hampering their usability for developing genomic sequence-based tools for public health applications. As such, this project represents the first truly large scale MTB sequencing effort designed to elucidate sequence variation and novel drug resistant loci.

Strain Selection

We propose to sequence strains from two patient cohorts from the province of KwaZulu-Natal (KZN) in South Africa: (1) PROX, (2) KZSUR (described below). Each strain has been characterized by clinical laboratory DST, confirmed by at least one other laboratory, and is accompanied by other required minimum metadata. Sequence polymorphisms will be identified and compared across genomes derived from both cohorts to identify the suite of polymorphisms corresponding to each drug resistance profile. This analysis will also allow for an assessment of how many genomes are needed to saturate the view of drug resistance polymorphisms from any one geographic locale, like KZN. This information will inform future work to sequence MTB from patient cohorts in this and other locales. In addition, the PROX study will identify polymorphisms from each PROX strain pair (isolated at admission and at eight weeks following treatment as described below) that will be directly compared to identify polymorphisms that arise in MTB strains of antibiotic treated patients over time.

Cohort: PROX

PROX isolates were collected from patients enrolled in the Prospective Study of Extensively Drug Resistant Tuberculosis (XDRTB) study (PROX) of Dr. Max O'Donnell from the Albert Einstein College of Medicine in New York. Patients over the age of 18 admitted to King George V hospital in Durban, South Africa from August 2009 until the present are eligible. All of these patients have clinically confirmed XDRTB (by DST), which is stated in the PROX protocol. For the first 90 enrollees (August 2009 – August 2011), patients were followed monthly for up to two years, including a follow-up questionnaire and sputum sample. For patients admitted after August 2011, study logistics dictate that patients be seen monthly through the first six months and then periodically through up to two years. Clinical data collected in each patient's Case Report Form (provided for reference), including patient outcomes and treatment regimens will be provided when available. **We will sequence 80 XDR MTB strains from this collection, listed as TBR1-TBR80 in the attached metadata spreadsheet, K-RITH_Metadata.xls, 'Sample' tab.**

Among the strains isolated from PROX patient sputum samples are 20 "sequential isolates", representing 42 strains, taken at the time of admission to King George V and then at eight weeks after admission (two patients had one additional sample taken). Data from these strains will allow us to examine how XDRTB therapy influences the genomic evolution of the tubercle bacilli under antibiotic pressure and against key clinical variables. This information can be greatly useful in understanding the mechanisms of resistance and virulence in MTB.

Cohort: KZSUR (KZN Survey)

MTB clinical isolates will also be sourced from two drug-resistance surveillance projects that have been carried out in KZN since 2006. The first screened TB outpatient suspects from sequentially selected hospitals from across the province. To date, the project has screened over 5,000 subjects at 12 sites with an MTB isolation rate of approximately 25%. MDRTB rates from these outpatient sites ranged from 4% to 12% with XDRTB at below 2%. The second project was an inpatient drug resistance survey that involved two rounds of screening all patients admitted to acute medical wards in 20 hospitals KZN. In total, there were 2,964 inpatients on wards where sampling was done and 53% had a current cough and sufficient microbiological and clinical data for inclusion in the study. MTB was isolated from 543 inpatients, 15% of which were found to be MDRTB infected and 3% had XDRTB. **We will sequence 97 strains from the inpatient and outpatient cohorts from the KZN Survey, listed as TBR81-TBR177 in the attached metadata spreadsheet, K-RITH_Metadata.xls, 'Sample' tab.**

Combined, these two studies provide an excellent source of drug resistant strains for sequencing for a number of reasons. The strains were collected with written patient consent for isolate analysis, have been characterized phenotypically for susceptibility to eight drugs, and have patient clinical data obtained through interview and chart review at the time of sample collection. The profile of the strains represents the full spectrum of resistance from mono-resistance and poly-resistance through MDRTB, to pre XDRTB and XDRTB. In addition the wide geographical sampling frame and collection from TB suspects rather than confirmed cases at a single treatment center will increase the chance of capturing the full diversity of drug resistant strains

References:

1. Todar's Online Textbook of Microbiology, <http://textbookofbacteriology.net/tuberculosis.html>
2. CDC, <http://www.cdc.gov/tb/topic/globaltb/default.htm>
3. Global tuberculosis control: WHO report 2011
4. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. Jun. 11;393(6685):537-44.
5. Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs, W.R., Jr, Venter, J.C., and Fraser, C.M. (2002). Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol*. Oct;184(19):5479-90.
6. Jassal, M. and Bishai, W.R. (2009). Extensively drug-resistant tuberculosis. *Lancet Infect Dis*. 9:19-30.
7. Loerger, T.R., Koo, S., No, E-G., Chen, X., Larsen, M.H., Jacobs, W.R. Jr, Pillay, M., Sturm, A.W., and Sacchetti, J.C. (2009). Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One*, 2009 Nov 5;4(11):e7778.

8. Loerger, T.R., Feng, Y., Chen, X., Dobos, K.M., Victor, T.C., Streicher, E.M., Warren, R.M., Gey van Pittius, N.C., Van Helden, P.D. and Sacchettini, J.C. (2010). The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. *BMC Genomics*, Nov 26;11:670.
9. Mardassi, H., Namouchi, A., Haltiti, R., Zarrouk, M., Mhenni, B., Karboul, A., Khabouchi, N., Gey van Pittius, N.C., Streicher, E.M., Rauzier, J., Gicquel, B., and Dellagi, K. (2005). Tuberculosis due to resistant Haarlem strain, Tunisia. *Emerg Infect Dis* 2005 June; 11(6):957-961.
10. Pepperell, C., Hoepfner, V.H., Lipatov, M., Wobeser, W., Schoolnik, G.K., and Feldman, M.W. (2010). Bacterial genetic signatures of human social phenomena among *M. tuberculosis* from an Aboriginal Canadian population. *Mol Biol Evol.* Feb;27(2):427-40.
11. Sandegren, L., Groenheit, R., Koivula, T., Ghebremichael, S., Advani, A., Castro, E., Pennhag, A., Hoffner, S., Mazurek, J., Pawlowski, A., Kan, B., Bruchfeld, J., Melefors, Ö., and Källenius, G. (2011). Genomic stability over 9 years of an isoniazid resistant *Mycobacterium tuberculosis* outbreak strain in Sweden. *PLoS One.* Jan 31;6(1):e16647.
12. Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M., Goguet de la Salmoniere, Y.L, Aman, K., Kato-Maeda, M., and Small, P.M. (2004). Functional and evolutionary genomics of *Mycobacterium tuberculosis*: Insights from genomic deletions in 100 strains. *PNAS* Apr 6;101(14):4865-4870.
13. Zhang, Y., Chen, C., Liu, J., Deng, H., Pan, A., Zhang, L., Zhao, X., Huang, M., Lu, B., Dong, H., Du, P., Chen, W., and Wan, K. (2011). Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family. *J Bacteriol.* Oct;193(19):5591-2.