**TB Project- Tracking TB Using its Transmission Rate**

Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium tuberculosis. One third of the world’s population is infected with TB and each year almost 2 million people die from it. There is an interest in epidemiological studies of TB, such as estimating the transmission rate of TB, which has led to the creation of deterministic models for analyzing TB. Genetic data must be used for the quantitative studies of TB.

TB genotyping allows Mycobacterium tuberculosis to be analyzed and when combined with epidemiological data, it allows the transmission of TB to be tracked from its chain of transmitters. The main types of genotyping which are used for TB are spoligotyping, IS6110 and MIRU. Spacer oligonucleotide typing (spoligotyping) uses a pattern of sequence spacers in a direct repeat region of the M. tuberculosis genome to identify the M. tuberculosis genotype. IS6110, an RFLP (Restriction Fragment Length Polymorphism) based method, detects variation in the insertion element, IS6110, of M. tuberculosis genome. MIRU uses the difference in numbers of repeating tandems at a loci, specific region. Spoligotyping and MIRU both result in a digital code which allows for convenient analyzing of TB data between laboratories. RFLP is used by special request and results in patterns of 7 or more bands.

Genotype data can be useful in studying the transmission of TB by providing the number of TB genotype clusters. A TB genotype cluster refers to two or more M. tuberculosis clusters which match by genotyping methods such as the same spoligotype pattern. Mutation rate of TB can be key in understanding TB transmission rates. Each cluster of TB represents a strain of TB which is assumed to be recently transmitted. The larger the cluster size of TB, the more recent the transmission and the lower the mutation rate. Thus, considering the mutation rate of TB is essential to understanding the transmission rate of TB.

Our purpose is to use a method for estimating parameters resulting in an estimate of transmission rate of TB. Approximate Bayesian computation is the motivation in creating the model we have used which was developed by M.M. Tanaka et. al (2006). We will apply this method to TB data from the Center for Disease Control and Prevention (CDC), specifically spoligotype, RFLP, and spoligotype-RFLP markers from both New York City and New York State. We will compare our results to those given by Tanaka et. al. (2006), which uses RFLP, and to each different data type. The net transmission rate of TB is estimated using a continuous-time stochastic model, which is similar to a linear birth-death process model. From this model a simulation using Bayesian inference is implemented which takes into consideration the birth, death, and mutation of TB genotypes. The simulation process is implemented using the Markov chain Monte Carlo (MCMC) method.

Table 1 below shows our preliminary results. Each set of data resulted in a different mean and 95% credibility interval. The 95% credibility interval found by M.M Tanaka et. al was (.38, 1.08) and the mean was .69. Our preliminary results follow and describe how we can estimate the transmission rate of TB using different types of genetic markers.

**TABLE 1**

**Posterior estimates of the Net transmission rate**

|  |  |  |
| --- | --- | --- |
| Description  | 95% credibility interval | Mean |
| New York State Spoligotype | (.4564, .4794) | .4679 |
| New York State RFLP | (0.4701, 0.4946) | .4824 |
| New York State Spoligotype and RFLP | (.4556, .4783) | .4669 |
| New York City Spoligotype | (.4768, .5002) | .4885 |
| New York City RFLP | (.4939, .5194) | .5066 |
| New York City Spoligotype and RFLP | (.4659, .4911) | .4785 |