Glenn Galvizo Jr.

Efficient Parameter Estimation of a Microsatellite Mutation Model

Variation in DNA sequences allow us to describe the history of humanity. A microsatellite is a form of genetic variation where short sequences are repeated in tandem. Each microsatellite variant differs in how many times the short sequence is repeated. Human history inferences can be made by surveying how the number of repeats change across generations. Given a mutation and evolution model for human microsatellites, the question this research aims to answer is, “What are the best parameters for this model?”.

To find the best parameters given a set of observed microsatellite samples is to maximize a likelihood. For this problem though, the maximum cannot be found numerically in a reasonable amount of time. My methods to circumvent this were as follows: (1) Observations were collected from an existing database. (2) Simulated samples were generated using a coalescent approach, which avoids working on individuals that do not directly contribute to a sample’s evolutionary history. (3) To compare an observed sample and a simulated sample, the angular distance was used. (4) An approximate likelihood calculation was used in lieu of the exact likelihood to account for the low frequency of exact observation – simulation matches. (5) A Markov Chain Monte Carlo (MCMC) method was used to produce samples from the approximate likelihood.

Preliminary results show strong evidence for the existence of optimal parameters for our microsatellite mutation model. With defined parameters, future work entails using the same methodology for parameter determination of population and admixture models.