TESTING IF TWO SHOTGUN-SEQUENCING LIBRARIES ARE FROM THE SAME INDIVIDUAL WHEN COVERAGE IS VERY LOW

Samuel Vohr, University of California Santa Cruz

Determining whether two DNA samples originate from the same individual is difficult when the amount of retrievable DNA is limited. In the field of Ancient DNA, DNA preservation and the amount of material available for extraction limit our ability to construct high-coverage shotgun sequencing libraries from historical samples. As a result, in two libraries prepared from samples from the same individual, very few sequenced reads will overlap between libraries and even fewer will fall on polymorphic positions that can be used to verify their origin in a single individual. We describe a method for assessing whether two low-coverage (<0.01x) libraries are consistent with originating in the same individual or two individuals using linkage disequilibrium information obtained from a reference panel of individuals. We compare pairs of closely linked single nucleotide polymorphic sites to determine whether the observation of two alleles at nearby sites in two different libraries is consistent with the patterns of linkage present in the reference panel. In our preliminary work, we perform coalescent simulations to assess the power of our method to distinguish between pairs of simulated low-coverage libraries derived from a single individual or two individuals using a simulated reference panel. We find that our method can distinguish between these two cases and retains power when the reference panel is made up of individuals from a diverged population.