Elucidating Demographic History Of Admixed Groups Using IBD Segments with Ancestry Switches

Introduction. Human genetic variation is constantly being shaped by evolutionary processes such as population growth, migrations, bottlenecks, expansions, and admixture (drawing ancestry from multiple source populations). These events create impressions in the genome by which we can surmise a population's size and composition over time. Methods for reconstructing population history have been numerous and involved over the past few decades, most recently using the number and length of runs of homozygosity (ROH) and identity-by-descent (IBD) segments[1]. As these methods have accumulated, so has the amount of whole genome sequence (WGS) data for human populations grown exponentially. As more data is collected, it is important to revisit classical models for demographic inference, many of which are not inclusive towards admixed groups. Most mathematical models in human genetics that are used to infer effective population size have thus far been restricted to populations that contain at most one ancestral source group. **This creates applicability issues and information loss when generalizing to admixed populations**. The effect size of population(s) of interest also influences the power and resolution of following genetics studies, especially where the individual ancestral components must be considered. **I propose jointly modeling the relationship between IBD segments and local ancestry**.

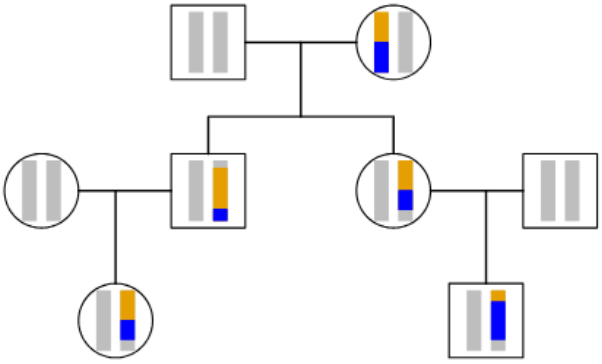

Figure 1. Sample Pedigree of Ancestry Switches within IBD Segments

Background. IBD segments are shared genetic segments among individuals that are inherited from a common ancestor. Notably, these segments may contain one or multiple ancestries. See Figure 1 for a simplistic example, where the colors represent distinct ancestral groups. Previous work has shown that by determining the patterning of these segments we can gain valuable insight into a population's history[1], but has also neglected to incorporate IBD segments with ancestry switches because the mathematical approximations no longer hold[2,3]. I believe this relationship between ancestry switches and IBD **can** be modelled and information concerning the joint distribution of the length of IBD segments and ancestry switches can be used to infer key demographic events in admixed populations. I aim to demonstrate this by: First, simulating shared ancestry switches within IBD segments that are shared across human populations. Second, by developing a novel method to jointly model the length and distribution of IBD segments and ancestry switches to timed admixture events and changes in population size.

Aim 1: Identify ancestry switches within IBD segments in large diverse genomic databases. As part of my preliminary analyses for this proposal, I have generated ancestry switch counts (per individual) within IBD segments (Fig 2) for several admixed populations from 1000 Genomes Phase 3 GRCh38. Counts were obtained by calculating the intersection of IBD segments with local ancestry calls, obtained from Martin et al (2017)[4], using BEDTools[5]. As aforementioned, while previous studies have said the occurrence of these ancestry switches are few and or negligible[2,3], these counts do not reflect this. I will generate ancestry counts for the UKBiobank Afro-British, gnomAD, and PAGE consortium datasets, where we should expect to see similar proportions of ancestry switches. However, the distribution of these observed

ancestry switches will differ per-population, and the per-individual rate of ancestry switches will change due to their respective demographic histories. For example, if admixture were recent, we would expect to see longer segments with fewer ancestry switches, since recombination would not have had a chance to split up the segments.

Aim 2: Simulation Framework Tracking the Relationship Between Local Ancestry and IBD and Post-Admixture Mathematical Modeling of the Joint Distribution of IBD and Local Ancestry Switches. Our goal is to determine the timing of admixture and the strength of migration using the joint distribution of IBD and local ancestry. Currently, we know the rate at which IBD decays per generation, as well as how and why this rate reflects demographic processes.

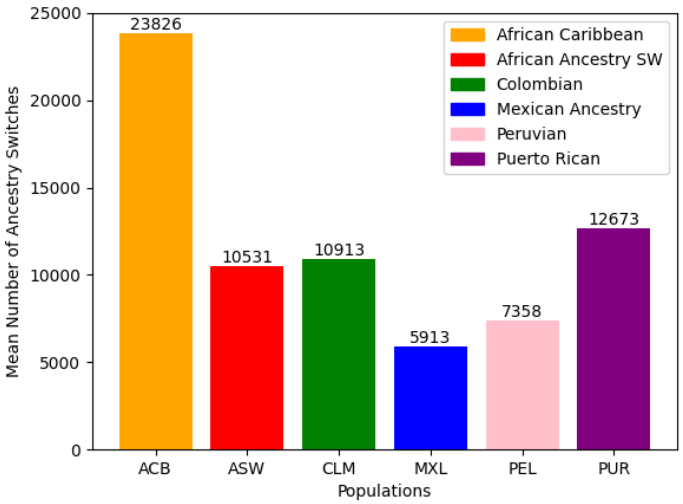

Figure 2. Ancestry Switch Counts Within IBD Segments (Per Individual)

We also know information about the local ancestries. I will base my simulations on the perfect binary tree (PBT) model framework introduced by Liang and Nielsen (2014)[7], re-implementing in msprime[8]. As described by Liang and Nielsen, this framework uses a dyadic interval-based stochastic process useful for inferring IBD segment length in populations with recent admixture. I will then build a generative model that I will validate with the ancestry switch count data generated from various datasets in Aim #1.

Intellectual Merit: In addition to focusing on diverse and understudied populations, this novel method will enable researchers to gain richer and more accurate information about admixed populations than previously thought possible.

Broader Impacts: As someone who identifies as an admixed individual, I am deeply passionate about communicating this research across disciplines and to the public and will continue to do so throughout my career. I will share my findings with the scientific community through publication, software, and presentations at national and international conferences such as SMBE and others. Throughout various research experiences, I have been privileged to have exceptional mentors whose legacy I would like to continue (see Personal Statement). I am particularly eager to mentor and encourage fellow women in science, and this project will give me the opportunity to join Dr. Jazlyn Mooney for introductory research science and coding workshops at GirlsInSTEM Day at the LA County Natural History Museum. These days provide opportunities for girls ages 8-18 to participate in hands-on STEM activities, meet diverse scientists, and encourage them to explore and pursue career paths as scientists. The overall goal is to promote creativity and nurture their scientific interests by providing girls with hands-on activities and most importantly, access to real scientists.

References: [1]Palamara et al 2012 *AJHG*; [2]Gravel et al 2012 *Genetics*; [3]Browning and Browning 2015 *AJHG*; [4]Martin et al 2017 *AJHG*; [5]Quinlan and Hall et al 2010 *Bioinformatics*; [6]Browning and Browning 2012 *Annu. Rev. Genet.*; [7]Liang and Nielsen 2014 *Genetics*; [8]Baumdicker et al 2022 *Genetics*;