Inferring very recent sex-specific demographic histories from population genomic data

Background. Demographic models often combine information about males and females and do not account for sex-differences in variance in reproductive success and dispersal. However, in many cases sex-specific differences in demographic variables, such as effective population size and migration rate, can provide more accurate and complete pictures of a population's natural history. Much of the early work on sex-specific demography has focused on the mitochondrial genome and the Y-chromosome, which are uni-parentally inherited. However, these markers are non-recombining, and therefore contain limited information about a population's demographic history, as they are single draws of the coalescent process. Comparisons between the X chromosome and autosomes also contain information about sex-specific processes since the X spends more of its time in females than the autosomes. Although extracting this information can be complicated, the large number of independent loci on the X and autosomes makes it possible to determine a more comprehensive view of a population's sex-specific history.[1]

I will investigate sex-differences in effective population sizes and migration rates in humans, using information from (1) rare alleles present in a population and (2) long segments of shared genetic material, by comparing the X to the autosomes. Methodological advances, leveraging information in the frequencies of rare alleles[2,3] and the distribution of block lengths of shared blocks of identical by descent (IBD) ancestry,[4,5,6,7] have facilitated the inference of very recent population demographic history using autosomal data. Despite these advances, little work has been done to elucidate the sex-specific recent demographic history of populations using X-autosomes comparisons. These methods, applied to population genomic data, can move the study of sex-specific demography onto ecologically and anthropologically relevant time-scales.

To gain insights into the different histories of males and females, I can utilize the fact that the X spends two-thirds of its time in females and one-third of its time in males, whereas the autosomes are present equally in both sexes. Simple expressions exist to compare the effective population sizes and migration rates at time τ for the autosomes and X ($N_e^A(\tau)$, $m_A(\tau)$ and $N_e^X(\tau)$, $m_X(\tau)$ respectively).[8] I will develop coalescent and diffusion-based inference approaches, expanding on those available for autosomal analysis to X-autosomal comparisons.

Although my primary aim is to elucidate sex-specific human evolutionary history, using publicly available human population genomic data, I will also work to apply these methods to non-model systems to understand population histories on an ecologically relevant scale. I will carefully quantify the accuracy of these methods using simulated data so that others can choose to collect data from their systems for use with the approach that results in the best estimates.

Methods. *I. Inference from rare allele frequencies*: The distribution of rare alleles in a population, as summarized by the allele frequency spectrum (AFS) is informative about recent population history, as these alleles are usually due to recent mutations.[2,3] For example, recent population expansions are characterized by a relative abundance of rare alleles. Also, rare alleles tend to be geographically clustered, as migration has had little time to spread them. Migration rates can be estimated from a joint frequency spectrum of alleles in two populations.[2]

To estimate sex-specific effective population sizes over time, I will first create a model of the effective population size across time for autosomal loci ($N_e^A(\tau)$) by using current methods that estimate recent population demography from an AFS.[2,3] Then, I will extend this model to test hypotheses of different sex-specific effective population sizes, by generating an expected AFS of the X chromosome under the assumptions of each hypothesis, fitting these models, and seeing if the simulations are consistent with the observed AFS. I will start with the simplest hypothesis and continue increasing in complexity, being careful to account for differences in

mutation rate between males and females. First, I will fit a scaling parameter, f, such that $N_e^X = f(N_e^A)$, using methods similar to those I used to estimate demographic models for the autosomes. I will first test if recent male and female effective population sizes are equal, allowing f = ¾, and whether this model offers a good fit to the observed AFS of the X chromosome. If using a constant rescaling factor is not consistent with the observed data, I will estimate a demographic model for both males and females, allowing the sex-specific population sizes to vary over time.

Likewise to estimate migration rates, I can apply existing models of recent demography and migration to the joint site frequency spectrum of rare alleles in two populations for the autosomes.[2] By developing similar approaches, I will test various models increasing in complexity, this time looking at both sex-specific effective population sizes and migration rates.

II. Inference from IBD block lengths: Segments of genome shared by two individuals inherited IBD from a single common ancestor may be identified in genome-wide genotype and sequence datasets. Because recombination shortens IBD segments, their lengths are inversely correlated to the number of generations since the two individuals last shared a common ancestor. This allows inference of relatedness over tens to hundreds of generations[4] and insight into relatively recent demographic events. A similar model testing approach will be carried out as outlined for the AFS (above), except on distributions of IBD block lengths within and between populations. Again, I will be able to estimate both male and female population sizes and migration rates.

I will carefully compare these two approaches through simulations to help the population and non-model genomics community evaluate the advantages and disadvantages of these different approaches.

Potential extensions. The sex-specific demographic models estimated in the work described above will allow me to study the different roles selection plays in shaping the X and autosomal chromosomes. Due to differences in dominance and population size, selection against rare alleles is expected to differ between the X and the autosomes.[10] My approach will allow me to compare the frequency spectrum of rare deleterious (e.g. non-synonymous) mutations on the X and autosomes, controlling for differences in population sizes.

Broader Impacts. The determination of sex-specific recent demographic histories is relevant to many fields in biology, conservation, and anthropology. I will publicly release the program code and documentation for my methodology so that researchers studying many different systems can utilize these approaches. I will also involve undergraduates in aspects of this project, such as coding small scripts and running simulations, as well as involving them in interpretation of the results. It is important to me to give students opportunities in research to help train the next generation of scientists. Especially as the field of population genetics is flooded with data, exposing undergraduate students to computational research is increasingly important in developing their skills as future scientists. I am well qualified to serve as a mentor to these students (see Personal Statement). I will also share my results as a guest contributor to the blog run by my advisor (http://gcbias.org/category/teaching/popgen-teaching/), Dr. Graham Coop, where he shares lecture notes and teaching tools. Lastly, because my work will provide insight into our own species, the outcome of this study will be of interest to the general public, so I plan to share my research at public outreach events to expose a broad audience to evolutionary biology and genomic methods.

References [1] Schaffner, (2004) *Nat. Rev. Genet*. [2] Gravel, et al., (2011) *PNAS*. [3] Nelson, et al., (2012) *Science*. [4] Palamara and Pe'er, (2013) *Bioinformatics*. [5] Palamara, et al., (2012) *Amer. Jour. Human Genet.* [6] Ralph and Coop, (2013) *PLoS Biol.* [7] Harris and Nielsen (2013) *PLoS Genet.* [8] Charlesworth, (2009) *Nat. Rev. Genet.* [9] Wilder, et al., (2004) *Mol. Biol. Evol.* [10] Vicoso and Charlesworth (2006) *Nat. Rev. Genet.*