



BIO
MATH

Ph.D. Dissertation Defense Seminar

Fast Spatial Ancestry Estimation via Flexible Allele Frequency Surfaces



John Michael Ranola

Doctoral Graduate Student

Department of Biomathematics

UCLA

Monday, November 25, 2013

2:00 PM

Gonda Center 5303

ABSTRACT:

Unique modeling and computational challenges arise in locating the geographic origin of individuals based on their genetic backgrounds. SNPs vary widely in informativeness, allele frequencies change nonlinearly with geography, and reliable localization requires evidence to be integrated across a multitude of SNPs. These problems become even more acute for individuals of mixed ancestry and present substantial computational challenges that have been addressed with approximate models. We attack these problems by borrowing ideas from image processing and optimization theory. Our model discretizes the region of interest into pixels and operates SNP by SNP. We estimate allele frequencies across the landscape by maximizing a product of binomial likelihoods penalized by nearest neighbor interactions. Penalization smooths allele frequency estimates and promotes estimation at pixels with no data. Maximization is accomplished by an MM (minorize-maximize) algorithm. Once allele frequency surfaces are available, one can apply Bayes rule to compute the posterior probability that each location is the origin of a given person. Placement of admixed individuals on the landscape is more complicated and requires estimation of the fractional contribution of each pixel to a person's genome. This estimation problem also succumbs to a penalized MM algorithm. On the POPRES data, the current model gives better localization for both unmixed and admixed individuals than existing methods despite using just a small fraction of the available SNPs. Computing times are comparable to the best competing software.

Doctoral Committee: Kenneth Lange, Ph.D. (chair), Janet Sinsheimer, Ph.D.,
Marc Suchard, M.D., Ph.D., Steve Horvath, Ph.D., Sc.D.