

Ancestry Specific Allele Frequency Estimation (ASAFE)

Qian Sophia Zhang

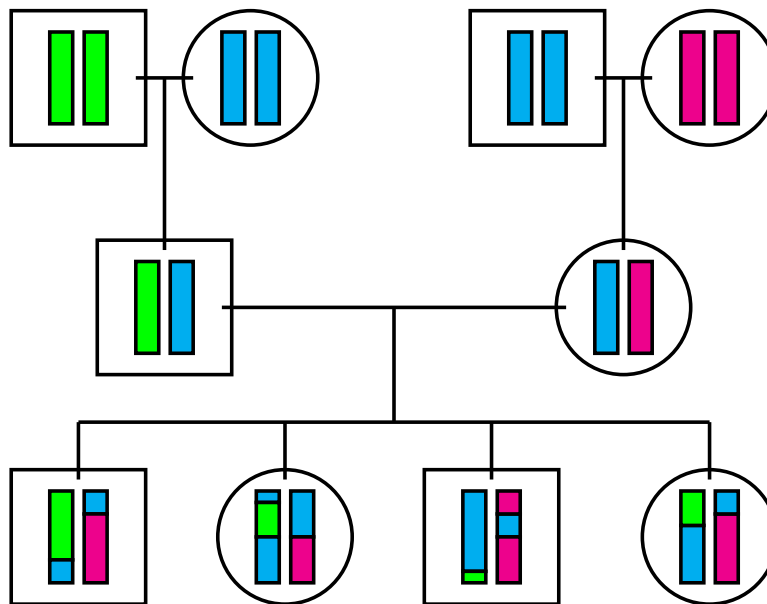
Joint work with Dr. Sharon Browning and Dr. Brian Browning
Department of Biostatistics
University of Washington

Slides @ <http://biostatqian.github.io/ASAFE/>

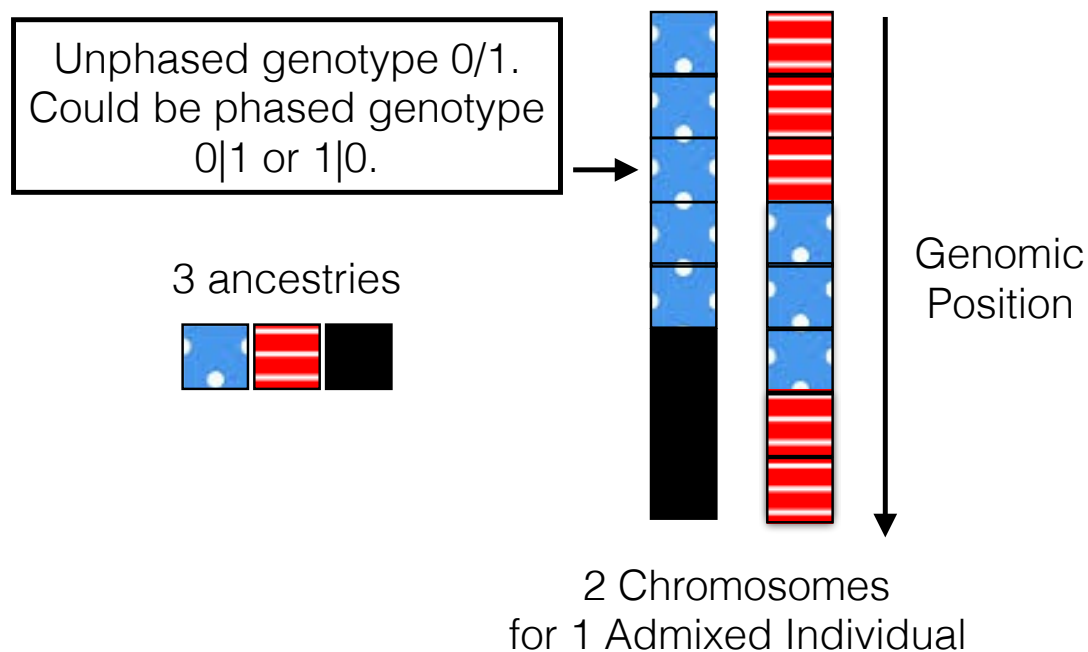
August 1, 2016

ASAFE: EM Algorithm for Estimating Ancestry Specific Allele Frequencies at a SNP

- For a significant SNP (with alleles 0 and 1), want **ancestry-specific allele frequencies** $\coloneqq P(\text{Allele 1} \mid \text{African}), P(\text{Allele 1} \mid \text{European}),$ and $P(\text{Allele 1} \mid \text{Native American})$, i.e. frequencies of allele 1 amongst chromosomes of African, European, or Native American origin at the SNP



Available Data: ASAFE assumes we know bi-allelic unphased genotypes and phased ancestries, but not the genotype order relative to ancestry pair

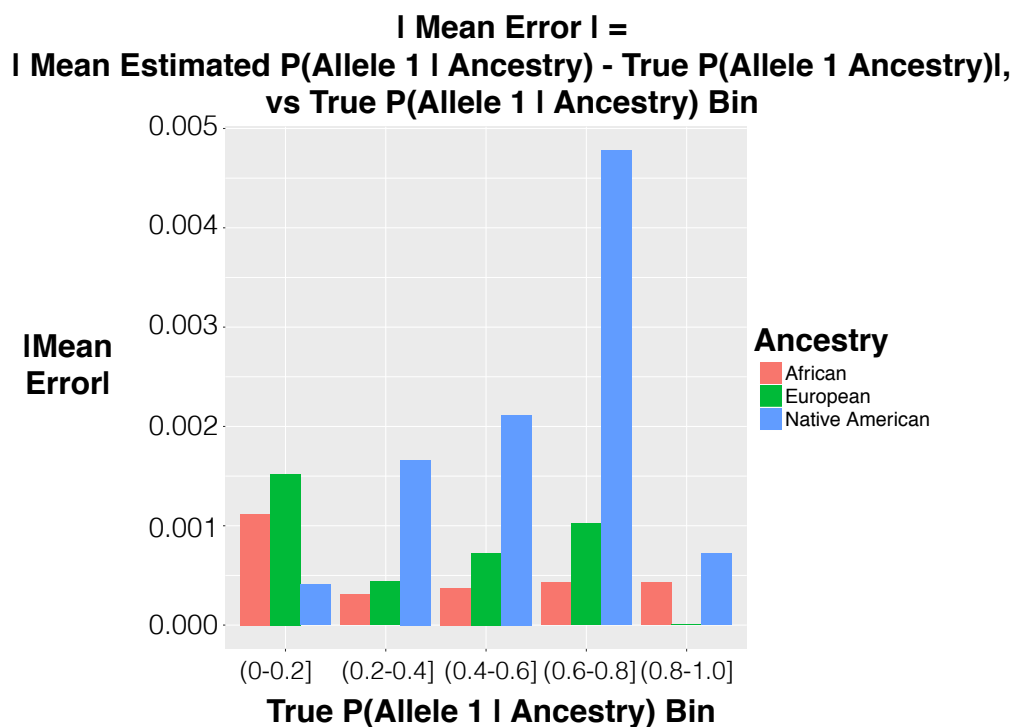


Error Calculation on Simulated Data

- Applied ASAFE to simulated genotypes and ancestries for 56,003 SNPs in a Hispanic sample → Got ancestry-specific allele 1 frequencies for each ancestry (African, European, Native American), at each SNP
- For each SNP and ancestry $a \in \{A, E, N\}$, calculated error = Estimated $p_{1|a}$ - True $p_{1|a}$

Low Error on Simulated Data

- For each $a \in \{A, E, N\}$, grouped SNPs errors by True $P(\text{Allele 1} \mid \text{ancestry})$ bin, and took the mean of errors within the same bin
- Largest $|\text{Mean of Errors}| = 0.005$. Largest SD of Errors = 0.03.



More Info: Paper, Presentation, and Code

- **Poster:** Poster Number 13 at Session Number 213225 at 10:30 AM in CC-Hall F1 West.
- **Paper:** Qian S. Zhang, Brian L. Browning, and Sharon R. Browning. Asafe: ancestry-specific allele frequency estimation. *Bioinformatics*, 32(14):2227–2229, 2016.
- **R package “ASAFE”:** On Bioconductor
- **Presentation slides, code to reproduce analysis:**
<http://biostatqian.github.io/ASAFE/>

