AIPS Software Manual

Jinyoung Byun
Younghun Han
Christopher Amos

Department of Biomedical Data Science
Dartmouth College

Software from
https://morgan2.dartmouth.edu/~jbyun/html/aips.html

July 12, 2017

1. Introduction

AIPS (Ancestry Inference using Principal component analysis and Spatial analysis) is a R-program for inferring genetic ancestry in a distance-based method using Principal Component Analysis and Inverse Distance Weighted Interpolation Method from Spatial Analysis in large-scale SNP genotype data.

2. Quick start

The genotype files must be properly cleaned. Please check the option --MIND in PLINK. The option --MIND 0.05 includes individuals with high genotyping at a least 95% complete. If --MIND is < 95, you may observe an additional population that reflects poor data quality. Also, for ancestry unknown samples (your input file) please replace FID in input.fam with 0.

We provide 952 ancestry known samples as reference data to calculate centroids. Before running R-program, you have to merge the reference data with your data using Plink. Input file in AIPS requires the SNP genotypes recorded in terms of additive components (0, 1, and 2).

For example,
(a) % plink --bfile euro952samples --bmerge input.bed input.bim input.fam --make-bed --out output
      % plink --bfile output --recodeA --out outputA

(b) % plink --bfile euro952samples --bmerge input.bed input.bim input.fam --recodeA --out outputA

Note that you may need to check strand flip. More information on PLINK can be found here:
http://pngu.mgh.harvard.edu/~purcell/plink/

Please first run the function "**AIPS-PCA.R**" to get the scores (principal components). These SNPs in the reference data were chosen from SNPs showing high FST values and differentiating well among intra-European subpopulations. You can choose the method either "eigen" or "svd" to compute principal components. By default, AIPS uses the function "eigen".

The next function "**AIPS-AI.R**" calculates the membership probability of each intra-continental ethnicity definition from scores of the merged data.

If you have any question or encounter problems, please send an email to:
Jinyoung.Byun@dartmouth.edu , Younghun.Han@dartmouth.edu or Christopher.I.Amos@dartmouth.edu