R packages for kinship and population structure

Stephanie Gogarten

July 18, 2018

GENESIS overview

Author: Matthew Conomos

Iterative process for separating recent familial relatedness from distant genetic ancestry

1. KING-robust[1] to get initial kinship estimates
2. PC-AiR[2] to to select an informative set of unrelated samples, do PCA on unrelated, project into relatives
3. PC-Relate[3] to estimate kinship coefficients adjusted for population structure and admixture using PCs

---

[1]Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K. Sale M., Chen W.M. (2010). *Robust relationship inference in genome-wide association studies*. Bioinformatics, 26(22): 2867-2873.

[2]Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). *Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness*. Genetic Epidemiology 39, 276-293.

[3]Conomos M.P., Reiner A.P., Weir B.S., and Thornton T.A. (2016), *Model-free estimation of recent genetic relatedness*. American Journal of Human Genetics, 98(1): 127-148.

PC-Relate flow chart
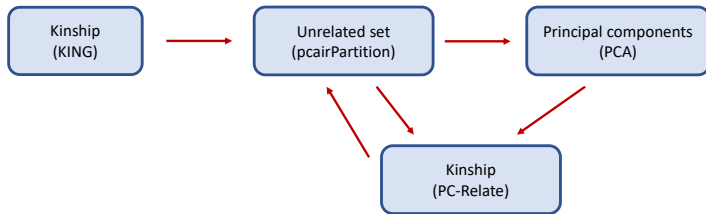
GENESIS key functions

Function	Description
pcairPartition	Select a set of unrelated samples, maximizing ancestral diversity
pcair	Run pcairPartition followed by PCA on unrelated samples and projection of relatives
pcrelate	Estimate kinship coefficients, adjusting for PCs
pcrelateMakeGRM	Return a Genetic Relationship Matrix from pcrelate results
pcrelateReadKinship	Return a table of pairwise kinship from pcrelate results

SNPRelate overview

Author: Xiuwen Zheng

Functions for LD pruning, Principal Component Analysis, Identity by Descent, Genetic Relationship Matrix

TOPMed pipeline uses pcairPartition and SNPRelate functions for PCA in order to prune variants for LD in between selecting unrelated samples and running PCA.

SNPRelate key functions

Function	Description
snpgdsLDpruning	Prune variants in linkage disequilibrium using a sliding window
snpgdsIBDKING	Kinship estimates using KING-robust
snpgdsIBDSelection	Return a table of pairwise kinship
snpgdsPCA	Principal Component Analysis
snpgdsPCASNPLoading	Calculate variant eigenvectors from PCA results
snpgdsPCASampLoading	Project samples onto eigenvectors from snpgdsPCASNPLoading
snpgdsGRM	Genetic Relationship Matrix

Essential R objects

Object	Package	Description
AnnotatedDataFrame	Biobase	R data.frame with metadata describing each column
SeqVarGDSClass	SeqArray	Object providing access to a GDS file
SeqVarData	SeqVarTools	Object linking a GDS file to sample annotation