

R packages for kinship and population structure

Stephanie Gogarten

GENESIS overview

Author: Matthew Conomos

Iterative process for separating recent familial relatedness from distant genetic ancestry

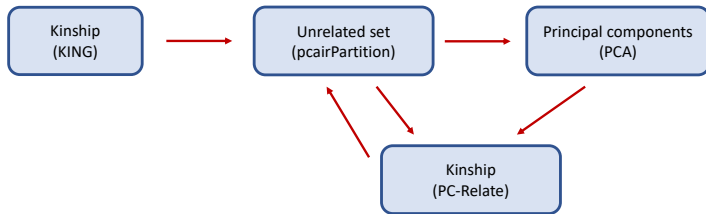
1. KING-robust¹ to get initial kinship estimates
2. PC-AiR² to select an informative set of unrelated samples, do PCA on unrelated, project into relatives
3. PC-Relate³ to estimate kinship coefficients adjusted for population structure and admixture using PCs

¹Manichaikul 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.

²Conomos, 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.

³Conomos 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
pcrelateToMatrix	Return a Genetic Relationship Matrix from pcrelate results

SNPRelate overview

Author: Xiuwen Zheng

- ▶ LD pruning
- ▶ Principal Component Analysis
- ▶ Identity by Descent
- ▶ Genetic Relationship Matrix

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
SeqVarGDSCClass	SeqArray	Object providing access to a GDS file
SeqVarData	SeqVarTools	Object linking a GDS file to sample annotation
SeqVarIterator	SeqVarTools	Object to return next set of variants with each iteration

