A sparse group lasso multi-marker mixed model for association mapping with population structure correction

Yingjie GUO, Chenxi WU, Ao LI, Maozu GUO, Alon KEINAN

Abstract: Exploring the genetic basis of heritable traits remains one of the central challenges in biomedical research. In traits with simple Mendelian architectures, single polymorphic loci explain a significant fraction of the phenotypic variability. However, many traits of interest seem to be subject to multifactorial control by groups of genetic loci. Accurate detection of such multivariate associations is non-trivial and often compromised by limited statistical power. At the same time, confounding influences, such as population structure, cause spurious association signals that result in false-positive findings. We propose linear mixed model LMM-SGLasso, a mixed model that allows for both multi-locus mapping and correction for confounding effects. Our approach is simple and free of tuning parameters; it effectively controls for population structure and scales to genome-wide datasets. LMM-SGLasso simultaneously discovers likely causal variants and allows for multi-marker-based phenotype prediction from genotype. We demonstrate the practical use of LMM-SGLasso in genome-wide association studies in Arabidopsis thaliana and linkage mapping in mouse, where our method achieves significantly more accurate phenotype prediction for 91% of the considered phenotypes.

Introduction

Methods

Results

Discussion

Acknowledge

References