OXFORD

Methodological implementation of mixed linear models in multi-locus genome-wide association studies

Yang-Jun Wen, Hanwen Zhang, Yuan-Li Ni, Bo Huang, Jin Zhang, Jian-Ying Feng, Shi-Bo Wang, Jim M. Dunwell, Yuan-Ming Zhang and Rongling Wu

Corresponding authors: Yuan-Ming Zhang, College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China. Tel.: +086 13505161564; Fax: +086 25 84399091. E-mail: soyzhang@njau.edu.cn; College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China. Tel.: +086 13505161564. E-mail: soyzhang@mail.hzau.edu.cn; Rongling Wu, Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA. Tel.: +001 717 531 2037; Fax: +001 717 531 0480. E-mail: rwu@phs.psu.edu

Abstract

The mixed linear model has been widely used in genome-wide association studies (GWAS), but its application to multi-locus GWAS analysis has not been explored and assessed. Here, we implemented a fast multi-locus random-SNP-effect EMMA (FASTmrEMMA) model for GWAS. The model is built on random single nucleotide polymorphism (SNP) effects and a new algorithm. This algorithm whitens the covariance matrix of the polygenic matrix K and environmental noise, and specifies the number of nonzero eigenvalues as one. The model first chooses all putative quantitative trait nucleotides (QTNs) with ≤ 0.005 P-values and then includes them in a multi-locus model for true QTN detection. Owing to the multi-locus feature, the Bonferroni correction is replaced by a less stringent selection criterion. Results from analyses of both simulated and real data showed that FASTmrEMMA is more powerful in QTN detection and model fit, has less bias in QTN effect estimation and requires a less running time than existing single- and multi-locus methods, such as empirical Bayes, settlement of mixed linear model under progressively exclusive relationship (SUPER), efficient mixed model association (EMMA), compressed MLM (CMLM) and enriched CMLM (ECMLM). FASTmrEMMA provides an alternative for multi-locus GWAS.

Key words: genome-wide association study; mixed linear model; multi-locus model; random effect

Introduction

Genome-wide association studies (GWAS) have been widely used in the genetic dissection of quantitative traits in human, animal and plant genetics, especially in combination with the output of genomic sequencing technologies. The most popular method for GWAS is the mixed linear model (MLM) method [1, 2] because of its demonstrated effectiveness in correcting the inflation from many small genetic effects (polygenic background) and controlling the bias of population stratification [3–7]. Since the MLM of Yu et al. [2]

Yang-Jun Wen is a Ph D candidate in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, China.
Hanwen Zhang is a bachelor student in the Faculty of Applied Science at the University of British Columbia, Canada.
Yuan-Li Ni is a Master student in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, China.
Bo Huang is a Master student in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, China.
Jin Zhang is an associate professor in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, China.
Jian-Ying Feng is a lecturer in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, China.
Shi-Bo Wang is a postdoctoral research fellow in the College of Plant Science and Technology at Huazhong Agricultural University, China.
Jim M. Dunwell is a full professor in the School of Agriculture, Policy and Development at the University of Reading, United Kingdom.
Yuan-Ming Zhang is a full professor in State Key Laboratory of Crop Genetics and Germplasm Enhancement at Nanjing Agricultural University, Nanjing, China and Chutian Scholar Professor of Statistical Genomics in the College of Plant Science and Technology at Huazhong Agricultural University, Wuhan, China.
Rongling Wu is Distinguished Professor of Public Health Sciences and Statistics and the Director of the Center for Statistical Genetics at The Pennsylvania State University, USA. He found the Center for Computational Biology at Beijing Forestry University, China.
Submitted: 24 October 2016; Received (in revised form): 15 December 2016