An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations

Vincent Segura, Bjarni J Vilhjálmsson, Alexander Platt, Arthur Korte, Quan Long, Magnus Nordborg

Combining multiple traits for genome wide association mapping

Arthur Korte[1], Bjarni J. Vilhjálmsson[1,2], Vincent Segura[1,3], Alex Platt[1,2], Quan Long[1], Magnus Nordborg[1,2]

[1] GMI Gregor Mendel Institue of Molecular plant Biology, Austrian Academy of Science, Vienna, Austria

[2] Molecular and Computational Biology, University of Southern California, Los Angeles, California, United States of America

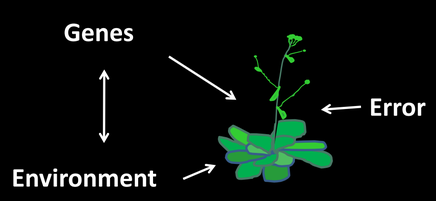
[3] INRA, UR 588 Amélioration génétique et physiologie forestières, Orléans, France

Introduction

Due to decreasing costs of sequencing and genotyping, genome-wide association studies (GWAS) are becoming a standard approach for studying the genetics of natural variation.

A major problem in such studies is that the complicated dependence-structure of the data — between loci as well as between individuals — makes estimating the effect of an individual locus challenging. Mixed models have emerged as a general and flexible approach for dealing with this problem. Here we extend this approach to carry out GWAS of correlated phenotypes. One application is dealing with traits that are biologically related: using human cohort data, we demonstrate greatly increased power to detect pleiotropic loci that effect more than one type of blood lipid.

A second application is dealing with the same trait measured in multiple environments: using Arabidopsis data, we demonstrate the identification of loci whose effect depends on the environment.



Mixed model for correlated traits (MTMM)

$$y = \begin{bmatrix} y1 \\ y2 \end{bmatrix} = \beta_0 + T\beta_1 + G\beta_2 + GT\beta_3 + Zu + \varepsilon$$

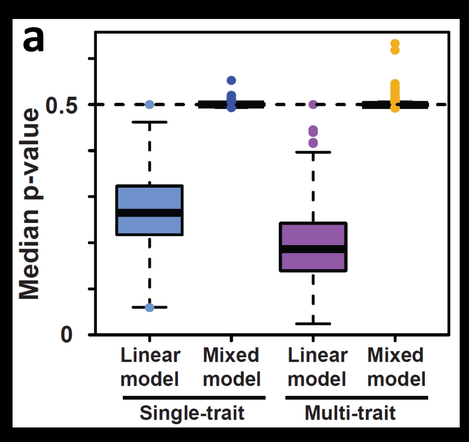
Full test :	H_0: $\beta_2 = 0, \beta_3 = 0$; H_a: full model
Trait specific test :	H_0: $\beta_3 = 0$; H_a: full model
Trait common test :	H_0: $\beta_2 = 0, \beta_3 = 0$; H_a: $\beta_3 = 0$

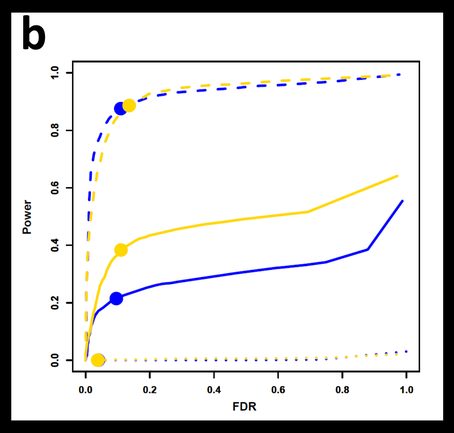
Fig.1

Simulations

Pairs of correlated phenotypes were simulated by adding phenotypic effects to genome-wide SNP data from A. thaliana[1]. A single randomly selected SNP was set to account for up to 2% of the phenotypic variance, but with the possibility different effects in each of the two phenotypes. We compared our ability to identify the focal locus using MTMM and marginal, single-trait analyses.

Three different tests were used: a full test that compares the full model, including the effect of the marker genotype and its interaction, with a model that includes neither; an interaction effect test that compares the full model to one that does not include interaction, and finally; a common effect test that compares a model with a marker genotype to one without (see Fig 1)





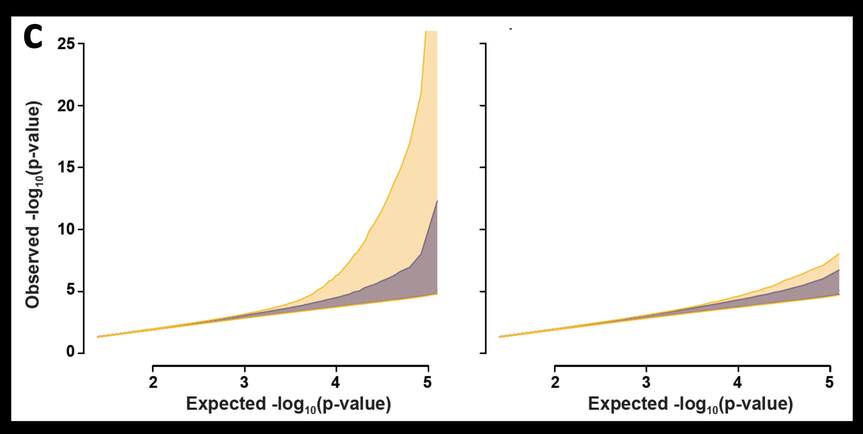


Fig.2

Analysis of human data

To illustrate the utility of MTMM for traits that are correlated because they are part of the same biological system, we reanalyzed data from the Northern Finland Birth Cohort 1966(NFBC1966)[2]. Table 1 shows the correlation estimates of these data.

Table 1 MTMM estimates of correlation and heritability in the NFBC1966 data

	Phenotypic[a]	Genetic	Environ.	Heritability[b]
HDL/TG	-0.37	-0.42	-0.36	0.38/0.18
HDL/LDL	-0.13	-0.19	-0.09	0.39/0.45
HDL/CRP	-0.19	0.24	-0.34	0.39/0.14
TG/LDL	0.32	0.31	0.35	0.19/0.44
TG/CRP	0.21	-0.50	0.34	0.18/0.13
LDL/CRP	0.09	0.08	0.10	0.45/0.13

[a] Direct estimates of the Pearson correlation are identical to the precision given.

[b] Single-trait estimates are: 0.38 (HDL), 0.18 (TG), 0.45 (LDL) and 0.13 (CRP).

The FADS1-FADS2 locus denotes a nice example. This locus was not significant in the marginal analysis of neither TG nor LDL, but became highly significant using MTMM thanks to a very strong interaction effect (Fig. 3). These genes are excellent candidates, and were mentioned in the previous analysis of the NFBC1996 data[2]. Strikingly, they were also identified in a massive meta-analysis involving more than 100,000 individuals[3], which furthermore reported opposite effects on TG and LDL, in agreement with the strong interaction effect we observe (Fig. 3g) using a sample of only 5,000 individuals.

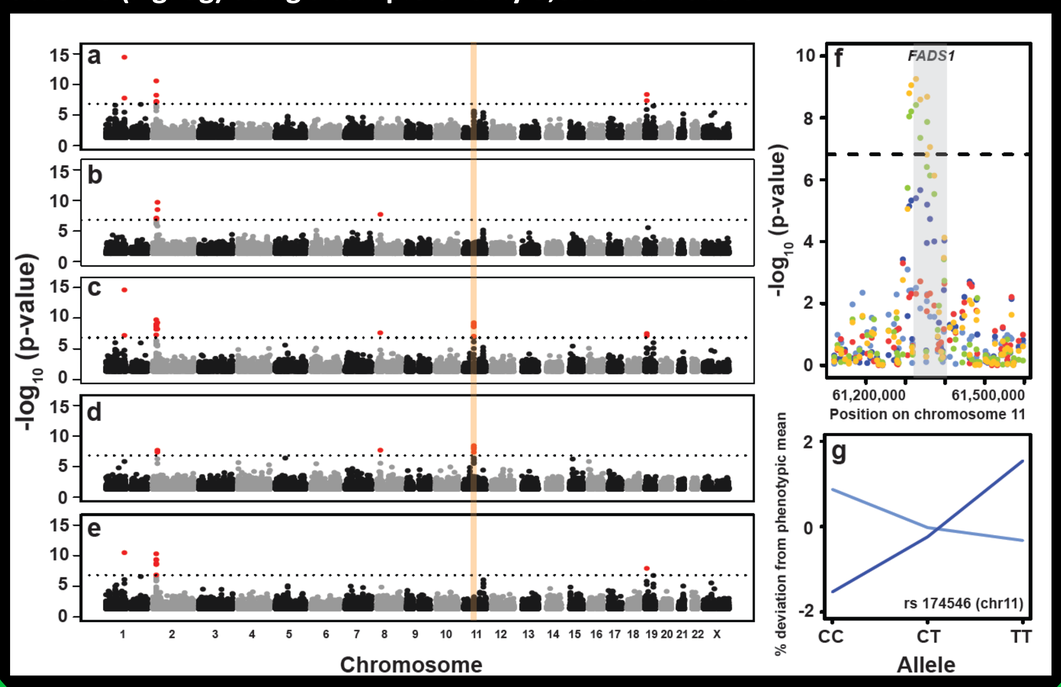


Fig.3

Analysis of *A.thaliana* flowering data

Finally, we analysed published flowering data of *A. thaliana*[4] for which the same phenotype was measured for two simulated growth seasons in two different locations. We performed a combined analysis on all four traits (2x2 factorial setting), highlighting the potential of the MTMM to combine even more than two traits.

This analysis leads to the identification of new associations (Fig.4). Interestingly, none of those newly identified trait specific markers is located near any flowering candidate genes.

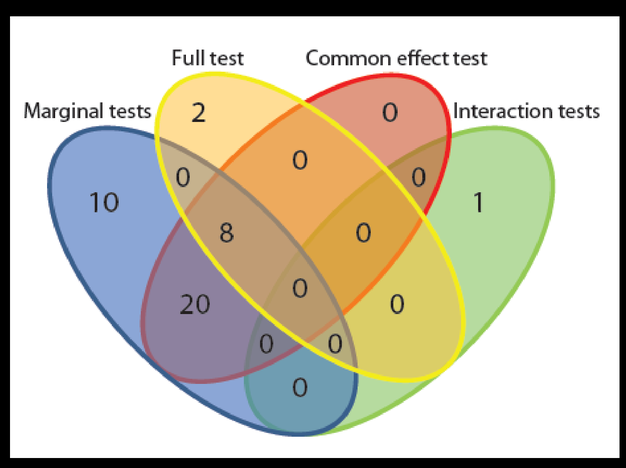


Fig.4

References

(1) Horton *et al.* 2012, *Nat Genet*

(2) Sabatti *et al.* 2009, *Nat Genet*

(3) Teslovich *et al.* 2010, *Nature*

(4) Li *et al.* 2010, *PNAS*

Deutsche Forschungsgemeinschaft DFG

ÖAW Österreichische Akademie der Wissenschaften

http://cynin.gmi.oeaw.ac.at/home/resources/mtmm/poster_complextraits_2012_AK.pdf