

Multivariate statistical modelling for QTL detection and marker selection in a bi-parental grapevine population

Charlotte Brault, Marie Perrot-Dockès, Agnes Doligez, Julien Chiquet, Loic Le Cunff, Timothée Flutre

► To cite this version:

Charlotte Brault, Marie Perrot-Dockès, Agnes Doligez, Julien Chiquet, Loic Le Cunff, et al.. Multivariate statistical modelling for QTL detection and marker selection in a bi-parental grapevine population. 17. Meeting of the EUCARPIA Section Biometrics in Plant Breeding, Sep 2018, Gand, Belgium. hal-01868802

HAL Id: hal-01868802 https://hal.inrae.fr/hal-01868802

Submitted on 4 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License







Multivariate statistical modelling for QTL detection and marker selection in a bi-parental grapevine population

Brault.C¹, Perrot-Dockès.M², Doligez.A¹, Chiquet.J², Le Cunff.L³, Flutre.T¹ ¹AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France ²AgroParisTech, UMR MIA, Paris, France ³Institut Français de la Vigne et du Vin, Montpellier, France

Genotyping

We worked on a 191-progeny of Syrah x Grenache. They have been genotyped with 153 SSR markers.

Abstract

In the present study, we perform variable selection with various flavours of the LASSO method (group Lasso, fused Lasso) adapted for multiple responses, extending the model and algorithm from Chiquet et al. (2017). We apply these methods on simulated data and on real data from Coupel-Ledru et al. (2014).

Phenotypic data



The aim of multivariate variable selection is to study multiple responses together in order to take into account the genetic correlation among responses (figure 1).

Figure 1

Genetic correlation table of 3 traits under 2 hydric conditions (WW and WD)

- $\Delta \Psi$: difference of water potential between soil and leaves *TrS* : *specific transpiration rate*
 - *KS* : *hydraulic conductance*

Three types of statistical modelling

 $Y \sim XB + E$

- Univariate with 1 trait in 1 hydric condition

Simulation

We used the design matrix of SSR genotypes from the Syrah x Grenache progeny, and simulated twenty times two responses with the aim of comparing various variable selection methods.

In each simulation, a single marker has a non-null effect. Its magnitude is similar to what was estimated in the real data. But in the simulations, we vary heritability (h2), genetic correlation (rhoB) among responses and add a dominance effect.

As we know the true position of the QTL, we calculate the True Positive Proportion (TPP) and the False Positive and Negative Proportions (FPP, FNP) for each method for these parameters (figure 3). For all variable selection method (except structured elastic net), we used stability selection.

high h2 & high rhoE h2 & low rhoB Figure 3 Results of SIM_QTL Multi_struct_EN Fuse_lasso group_lasso_both group_lasso_trait group_lasso_marker SIM_QTL Multi_struct_EN Fuse_lasso group_lasso_both group_lasso_trait group_lasso_marker *Simulations High* $h^2 = 0.8$ *Low* $h^2 = 0.18$ *Different* $h^2 =$ low h2 & high rhoB low h2 & null rhoB 0.18/0.10 High rhoB = 1SIM_QTL Multi_struct_EN SIM_QTL Multi_struct_EN Low rhoB = 0.5Fuse_lasso group_lasso_both group_lasso_trait group_lasso_marker Fuse_lasso group_lasso_both group_lasso_trait group_lasso_marker Null rhoB = 0*Dominance* : with low h^2 & diff h2 & high rhoB dominance high rhoB SIM : Simple SIM_QTL Multi_struct_EN SIM_QTL Multi_struct_EN Interval Fuse⁻lasso⁻ group lasso both group lasso trait roup lasso both Mapping ároup-lasso-marke

We started by studying each response separately.

- Univariate with 1 trait in 2 hydric conditions We then analysed 1 trait by adding in the statistical model a co-variable for each hydric condition (not shown).

- Multivariate with 6 responses

We finally considered the 3 traits and 2 conditions jointly (figure 2).

We used classical composite interval mapping with R/qtl only in the first modelling. For the last two, we used regularized regression with L1 and /or L2 penalties (classic lasso, group lasso and fuse lasso, structured elastic net). In the multivariate case, the group lasso selects an allele of a marker if there is a non-null estimated effect in all responses. For this, we used the R/glmnet package (Friedman et al., 2010).

The multivariate structured Elastic Net estimates correlations between responses and distinguishes between direct and indirect effects. Moreover, compare to the classical lasso, it can make use of the genetic map. For this, we used the R/spring package (Chiquet et al., 2017).

Conclusion

Given that traits may share the same genetic basis, we expect the multivariate statistical modelling to be appropriate to analyse multiple responses. In this preliminary work, to experiment with various modelling assumptions, we tried several methods, which gave contrasted results on real data. Further work via simulations are required to clarify the impact of these assumptions.

Figure 2



markers ordered by linkage group and genetic position

Main references

Coupel-Ledru, A., Lebon, É., Christophe, A., Doligez, A., Cabrera-Bosquet, L., Péchier, P., Hamard, P., This, P., Simonneau, T. (2014). Genetic variation in a grapevine progeny (Vitis vinifera L. cvs Grenache×Syrah) reveals inconsistencies between maintenance of daytime leaf water potential and response of transpiration rate under drought. Journal of experimental botany, 65(21), 6205-6218. Chiquet, J., Mary-Huard, T., & Robin, S. (2017). Structured regularization for conditional Gaussian graphical models. *Statistics and Computing*, 27(3), 789-804