

Genome admixture components accurately predict quantitative functional traits in plants

L. Gentzbitte^{1*}, C. Ben¹, M. Mazurier¹, M.G. Shin², T. Lorenz³, M. Rickauer¹, P. Marjoram², S.V. Nuzhdin², T.V. Tatarinova³

¹ *EcoLab, Université de Toulouse, CNRS, Toulouse INP, UPS, Toulouse, France*

² *Department of Biological Sciences, University of Southern California, Los Angeles, USA*

³ *Department of Biology, Division of Natural Sciences, University of La Verne, La Verne, CA, USA*

* e-mail: gentz@ensat.fr

Key words: admixture, genomic prediction, prediction accuracy, WhoGEM, genomic selection, molecular ecology, quantitative disease resistance, adaptation, *Medicago truncatula*

Motivation and Aim: Linking plant genomic and phenotypic variation for quantitative characteristics associated with fitness, functional traits or tolerance to biotic and abiotic stress provides tools to understand natural variation and optimize breeding schemes. Being indifferent to the source of genetic similarity between populations of a species (demography and local adaptation or selective breeding), we hypothesize that a large proportion of phenotypic variation between individuals may be best explained by population admixture across several ancestral populations originally separated.

Methods and Algorithms: Here we propose and test a method named “WhoGEM” to explain variation in genetically complex traits using population admixture proportions of plant individuals. The overall goal is to predict quantitative phenotypes rather than identify causative variations or candidate genes, or inferring the relative parts of demography and selection in the evolution of quantitative phenotypes. The WhoGEM analysis program is developed around three key data inputs, data preprocessing (genotypes), ProvenancePredictor (GPS coordinates), and phenotypic characterization (quantitative functional traits). The multicriteria approach is used to determine the optimal number of population subdivisions for the plant species under study, to assign an admixture proportion vector to each accession and to characterize/predict quantitative phenotypes. Geographical coordinates are used as covariates to optimize the definition of number of ancestral populations using the ProvenancePredictor algorithm. The relationships between genome components and phenotypes were estimated using linear models with systematic search for the best minimum model using the leaps R package to cope with dependencies among the predictors (the proportions of genome components must sum to one).

Results: The population structure of the circum-Mediterranean model legume *Medicago truncatula* was assessed by ADMIXTURE, DAPC and ProvenancePredictor. We summarized the genome variations by the eight-dimensional vector of admixture components. The results show that admixture components are significantly correlated with climate and geography. Then, the WhoGEM method was used to assess variation in genome admixture proportions, and revealed that it explains most of the phenotypic variation for quantitative disease resistances and quantitative functional traits. The method’s prediction accuracy outperforms or equates current algorithms for Genomics Selection (GS). WhoGEM analysis of *M. truncatula* produced the first evidence that quantitative phenotypes can be well-predicted using genome-wide patterns of admixture.

Conclusion: The development of WhoGEM demonstrates that population admixture can integrate the effects of demography (i.e., gene flow and genetic drift) and of natural

selection towards adaptation, and thereby explains more phenotypic variation than GS- or Quantitative Trait Loci (QTL)-based approaches. This insight contributes to understanding of phenotypic variation in evolutionary biology, and can accelerate plant and animal breeding and biomedical research programs by linking phenotypic traits.

Acknowledgements: M. Mazurier was supported by a PhD scholarship from the French “Ministère de la Recherche et de l’Enseignement Supérieur” and a “Visiting Student” fellowship from Toulouse INP for a stay at USC. L. Gentzbittel and C. Ben were supported by a fellowship from Toulouse INP for a stay at USC and a “Visiting Scholar” fellowship from the US Feed the Future Innovation Lab “Climate Resilient Chickpea”. T.V. Tatarinova was supported by a “Visiting Scholar” fellowship from Toulouse INP for a stay at Toulouse INP and by the NSF Division of Environmental Biology award No. 145663. We thank the Medicago Hapmap international consortium, mainly funded by the Plant Genome Program of the National Science Foundation (USA), for providing *M. truncatula* genome resequencing data.