
eQTLs are key players in the integration of genomic and transcriptomic data for phenotype prediction

Abdou Wade*¹, Harold Duruflé¹, Vincent Segura², and Leopoldo Sanchez¹

¹Biologie intégrée pour la valorisation de la diversité des Arbres et de la Forêt – Office National des Forêts, Institut National de Recherche pour l’Agriculture, l’Alimentation et l’Environnement : UMR0588 – France

²Amélioration génétique et adaptation des plantes méditerranéennes et tropicales – Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Centre international d’études supérieures en sciences agronomiques, Institut national d’études supérieures agronomiques de Montpellier : UMR1334, Institut national d’enseignement supérieur pour l’agriculture, l’alimentation et l’environnement, Institut National de Recherche pour l’Agriculture, l’Alimentation et l’Environnement, Institut national d’enseignement supérieur pour l’agriculture, l’alimentation et l’environnement – France

Résumé

Uncertainties posed by climate change, in terms of the adaptability of species, stress the need of developing sound predictions of individual performances. Multi-omics data not only offer new data for backing genomic predictions, but also provide an interesting way to link phenotypes and genome variation that cannot be easily explained with genome sequences alone. Still few studies address the problem of integrating different omics to gain understanding into genetic architecture and predictability of phenotypes. In our study, we used 241 *Populus nigra* genotypes, representative of the species range across western Europe. These poplars were phenotyped (growth, phenology, biochemical traits) in two common garden experiments located on contrasting sites (Orleans, France and Savigliano, Italy). RNA sequencing analysis was also performed on young differentiating xylem and cambium tissues of the entire set of genotypes in the common garden located at Orléans, allowing large genomic and transcriptomic datasets. For each phenotypic trait, prediction models were built with genotypic or transcriptomic data as predictors and these models were compared to those integrating both omics data by concatenation. The relative advantage of combining the two types of omics over models using them separately varied depending on the trait, for some of the most important traits in breeding the concatenation yielded more accurate predictions, while for most others no differences were detected, with still a few showing underperforming concatenation. In order to better understand the reasons underlying such differences, we made use of eQTL analysis to classify the predicting features into CIS and / or TRANS effects and further assessed their relative importance in the different predictive models according to this typology. Over all the traits under study, we found a high and significant correlation between the change in model predictability and the change in importance of predictors for the eQTL with TRANS effects and the transcripts regulated in CIS. Indeed, traits that benefited the most from the concatenation were also those for which the eQTL with TRANS effects and the transcripts regulated in CIS changed most substantially in the importance in the combined predictive model with respect to the respective single omics models. It

*Intervenant

is noteworthy that such relationship was mainly observed for the phenotypes evaluated in the common garden where the transcriptomic data were collected (Orléans). In conclusion, beneficial integration appears to operate through a change in the weighting of important genes and SNPs, notably those showing already mutual associations through eQTLs, and consequently by minimizing the redundancy between predictors. To our knowledge, our work is the first to experimentally clearly show such behavior. These results constitute a promising way to explore data integration and to devise future methodologies for optimizing multi-omics combination through differential weighting of features.

Mots-Clés: Genomic prediction, Multi-omic integration, Poplar, eQTL Analysis, Transcriptomic data