IDENTIFYING FUNCTIONAL ROLES OF SNPS USING METABOLIC NETWORKS FOR IMPROVED PLANT BREEDING

Debolina Sarkar, Pennsylvania State University, USA
debolina@psu.edu
Costas D. Maranas, Pennsylvania State University, USA

Key Words: *GWAS, Arabidopsis, Poplar, functional SNPs, metabolic engineering*

Genetic sources of phenotypic variation have been a major focus of studies in plants aimed at improving agricultural yield and understanding adaptive processes. Genome-wide association studies (GWAS) aim to identify the genetic background behind a trait by examining the associations between specific phenotypes and single-nucleotide polymorphisms (SNPs). Although such studies are now commonly performed, biological interpretation of the results remains a challenge; especially due to the confounding nature of population structure and the systematic biases it introduces. Here, we propose a complementary analysis referred to as SNPeffect that sifts out functional SNPs from the tens of thousands typically identified during a genome sequencing study by integrating biochemical knowledge encoded in metabolic models, superimposed with phenotypic measurements. By design, SNPeffect can handle both monogenic and polygenic traits while offering mechanistic interpretations of the deciphered genotype-to-phenotype relations. SNPeffect was used to explain phenotypic variations such as differential growth rate and metabolite accumulation in *A. thaliana* and *P. trichocarpa* accessions as the outcome of activating and inactivating SNPs present in the enzyme-coding regions of the genotypes. To this end, we also constructed a non-compartmentalized genome-scale metabolic model for *Populus trichocarpa*, the first for a perennial woody tree. As expected, our results indicate that plant growth is a complex polygenic trait which is primarily governed by carbon and energy partitioning. Growth-affecting SNPs in coding regions were found to primarily be in amino-acid metabolism, glycolysis, TCA cycle, and energy metabolism. Faster-growing *Arabidopsis* genotypes were predicted to have higher fluxes through the protein metabolism pathways, indicating that increase in amino acid levels has a positive growth effect. Faster genotypes were also seen to preferentially employ the energy-efficient purine salvage pathway as opposed to *de novo* purine biosynthesis for generating energy metabolites AMP and GMP. We also found putative causal SNPs to be distributed among genes belonging to glycolysis, pyrimidine metabolism, folate biosynthesis, and shikimate metabolism, which can serve as candidate genes for further experimental characterization and/or targeted plant breeding. For both *Arabidopsis* and poplar, a number of deactivating SNPs were predicted to be in genes belonging to the lignin biosynthetic pathway, indicating that the energetics of producing lignin is a major growth determinant. To further decipher the underlying genetic landscape, we calculated all possible epistatic interactions using flux-balance analysis. Interestingly, we detected a significant positive correlation between the number of negative epistatic interactions in a genotype and its replicative fitness, indicating that functional genetic redundancies are beneficial for growth in *Arabidopsis*. This possibly serves to increase robustness to mutational and/or environmental perturbations as these can then be buffered by shuttling metabolic flux through unaffected parts of the network. We anticipate that putative causal roles for many more SNPs can be gleaned if this analysis is repeated with additional genotypes, phenotypes (such as genotype-specific rates of photosynthetic oxygen evolution or nutrient exchange fluxes) and/or omics datasets (such as proteomics or transcriptomics). Hence, as genome sequencing and plant phenotyping technologies are rapidly decreasing in cost, undertaking large-scale studies that incorporate diverse datasets is also becoming more feasible. As more such data is made available, the need for complex analytical tools will also rise. We envision SNPeffect to pave the way for more tools that can mechanistically elucidate the genetic landscape underlying the wide phenotypic variations that is characteristic of plants.