Integrative analysis of heterogeneous plant cell wall related data

By Dhivyaa Rajasundaram, University of Potsdam

Abstract

Plant cell walls are complex structures that underpin plant growth and are widely exploited in diverse human activities thus placing them with a central importance in biology. Cell walls have been a prominent area of research for a long time, but the chemical complexity and diversity of cell walls not just between species, but also within plants, between cell-types, and between cell wall micro-domains pose several challenges. Progress accelerated several-fold in cell wall biology owing to advances in sequencing technology, aided soon thereafter by advances in omics and imaging technologies. This development provides additional perspectives of cell walls across a rapidly growing number of species, highlighting a myriad of architectures, compositions, and functions. Furthermore, rather than the component centric view, integrative analysis of the different cell wall components across system-levels help to gain a more in-depth understanding of the structure and biosynthesis of the cell envelope and its interactions with the environment.

To this end, in this work three case studies are detailed, all pertaining to the integrative analysis of heterogeneous cell wall related data arising from different system-levels and analytical techniques. A detailed account of multiblock methods is provided and in particular canonical correlation and regression methods of data integration are discussed. In the first integrative analysis, by employing canonical correlation analysis - a multivariate statistical technique to study the association between two datasets - novel insight to the relationship between glycans and phenotypic traits is gained. In addition, sparse partial least squares regression approach that adapts Lasso penalization and allows for the selection of a subset of variables was employed. The second case study focuses on an integrative analysis of images obtained from different spectroscopic techniques. By employing yet another multiblock approach - multiple co-inertia analysis, in situ biochemical composition of cell walls from different cell-types is studied thereby high-lighting the common and complementary parts of the two hyperspectral imaging techniques. Finally, the third integrative analysis facilitates gene expression analysis of the Arabidopsis root transcriptome and translatome for the identification of cell wall related genes and compare expression patterns of cell wall synthesis genes. The computational analysis considered correlation and variation of expression across cell-types at both system-levels, and also provides insight into the degree of co-regulatory relationships that are preserved between the two processes.

The integrative analysis of glycan data and phenotypic traits in cotton fibers using canonical methods led to the identification of specific polysaccharides which may play a major role during fiber development for the final fiber characteristics. Furthermore, this analysis provides a base for future studies on glycan arrays in case of developing cotton fibers. The integrative analysis of images from infrared and Raman spectroscopic approaches allowed the coupling of different analytical techniques to characterize complex biological material, thereby, representing various facets of their chemical properties. Moreover, the results from the co-inertia analysis demon-strated that the study was well adapted as it is relevant for coupling data tables in a symmetric way. Several indicators are proposed to investigate how the global and block scores are related.

In addition, studying the root cells of Arabidopsis thaliana allowed positing a novel pipeline to systematically investigate and integrate the different levels of information available at the global and single-cell level. The conducted analysis also confirms that previously identified key transcriptional activators of secondary cell wall development display highly conserved patterns of transcription and translation across the investigated cell-types. Moreover, the biological processes that display conserved and divergent patterns based on the cell-type-specific expression and translation levels are identified.