Systematic structural characterization of secondary metabolites and their biosynthetic pathways in poplar

Desmet, S.[1], Morreel, K.[1] and Boerjan, W.[1]
[1]Dept. Plant Biotechnology and Bioinformatics, Ghent University, Belgium
E-mail: sasme@psb.vib-ugent.be, krmor@ psb.vib-ugent.be, woboe@ psb.vib-ugent.be

1. Introduction

The current global climate change demands a transition from a fossil-based economy towards a bio-based economy. Poplar is a promising crop in regions with a temperate climate for bio-ethanol production but also for the purification of industrially interesting chemicals[1,2]. However, a main bottleneck in search for valuable metabolites is the identification of metabolites. Currently, 90% is still unknown.

2. Approach

The 'Candidate Substrate Product Pair' (CSPP) algorithm is an innovative structural characterization technique that simplifies and accelerates the identification of unknown metabolites by searching liquid chromatography – mass spectrometry profiles for peak pairs, called CSPPs, that have a mass and retention time difference corresponding to those of the substrates and products of enzymatic or chemical conversions[3]. These CSPPs are then visualized in a network with nodes, the LC-MS peaks and edges, the putative (bio)synthetic conversions. Starting from LC-MS peaks with a known structure, network propagation via the edges enables the structural characterization of neighbouring peaks.

False positive CSPPs in which the candidate substrate and product are not biochemically related, have been shown to occur and are partially removed by considering the MS/MS similarity of the spectra of the CSPP candidate substrate and product. Here, we further optimize the CSPP algorithm by combining it with feeding experiments with ^{13}C isotope labelled precursors.

3. Results

30 poplars (*Populus tremula* x *alba*) were harvested after four months of growth. Methanol extracts of bark, stem and leaf were analysed with reversed phase ultrahigh pressure liquid chromatography (UHPLC) coupled to mass spectrometry (MS). Each sample was analysed four times, i.e. on both a Fourier Transform-Ion Cyclotron Resonance-MS (FT-ICR-MS) and a Quadrupole-Time-of-Flight-MS (Q-ToF-MS) using negative and positive ionization mode. Following data processing of the FTMS chromatograms in negative ionization mode, approximately 3110, 2562 and 3762 m/z features were profiled for bark, stem and leaf respectively. These m/z features were then used for the construction of a CSPP network covering the different tissues.

An in-house database, called DynLib, containing all UHPLC-MS data was built and a package in R, called RDynLib was written to connect all MS data across the different chromatograms that belonged to the same compound, allowing the visualization of all MS data together (Desmet et al., in preparation).

To further improve the reliability that a CSPP reflects a biochemical conversion and to trace the biochemical network within the CSPP network. Feeding studies were performed with the aromatic amino acids [$^{13}C_9$]-labelled phenylalanine, [$^{13}C_9$] labelled tyrosine, [$^{13}C_{11}$] labelled tryptophan and with [$^{13}C_6$] labelled benzoic acid. An R package, called RDynSIL, was written to (1) trace all $^{12}C/^{13}C$ peak doublets, (2) visualize the ^{12}C and ^{13}C spectra together and (3) link this data to the CSPP network and the DynLib database (Desmet et al., in preparation). Via CSPP network propagation starting from the nodes that are already annotated and using the information from the feeding experiments, 323 compounds have been structurally characterized in poplar up to now. These structurally characterized compounds covered metabolites from different metabolite classes, i.e. aromatic amino acids, small organic acids, benzoic acids, phenylpropanoids, flavonoids, oligolignols and (neo)lignans.

4. Discussion

This research will result in the structural characterization of hundreds of unknown secondary metabolites of poplar, and will potentially lead to the annotation of compounds with a potential added value for the bio refinery. Moreover, the use of the feeding experiments will elucidate the putative biosynthetic routes towards these high value compounds. The identification of plant metabolites and their biosynthetic pathways could vastly accelerate the transition towards a bio-based economy. In addition, this research has the potential to greatly advance the value of metabolomics in systems biology.

References

1. Littlewood et al. (2014). Bioethanol from poplar: a commercially viable alternative to fossil fuel in the European Union. Biotechnology for Biofuels. **7**:113
2. Devappa et al. (2015). Forest biorefinery: Potential of poplar phytochemicals as value-added co-products. Biotechnology Advances. 33(6 Pt 1):681-716
3. Morreel et al. (2014). Systematic structural characterization of metabolites in Arabidopsis via candidate substrate-product pair networks. Plant Cell. Mar;26(3):929-45.