Lignin Management:
Optimizing Yield and Composition in Lignin-modified Plants

2014 GCEP (Y1, partial) Report

Investigators

Clint Chapple (CC), Professor, Department of Biochemistry, Purdue University, West Lafayette, IN USA 47907
Whitney Dolan, graduate student, Department of Biochemistry, Purdue
Wout Boerjan (WB), Professor, VIB Department of Plant Systems Biology, UGent Department of Plant Biotechnology and Bioinformatics, Technologiepark 927, 9052 Gent, Belgium
Kris Morreel, Staff Scientist, VIB, Gent, Belgium
Geert Goeminne, Engineer, VIB, Gent, Belgium
Ruben Vanholme, Post-doctoral Researcher, VIB, Gent, Belgium
Bartel Vanholme, Professor, VIB, Gent, Belgium
Claire Halpin (CH), Professor, Plant Biology and Biotechnology, Division of Plant Sciences, University of Dundee at JHI, Errol Road, Invergowrie, Dundee, Scotland
Chris McClellan, Post-doctoral Researcher, University of Dundee, Scotland
John Ralph (JR), Professor, Department of Biochemistry and the D.O.E. Great Lakes Bioenergy Research Center, The Wisconsin Energy Institute, University of Wisconsin, Madison, Madison, WI USA 53726
Yuki Tobimatsu, Research Scientist, Biochemistry, U Wisconsin, Madison (now an Assistant Professor at Kyoto U.)
Yukiko Tsuji, Research Scientist, Biochemistry, U Wisconsin, Madison
Xu 'Sirius' Li (SL), Professor, Department of Plant Biology and Plants for Human Health Institute, North Carolina State University, N.C. Research Campus, Kannapolis, NC USA
Renee Strauch, Research Specialist, NCSU
Han-Yi Chen, Post-doctoral Researcher, NCSU

Abstract

This project aims to maximize the utility of plant lignocellulosic biomass as an abundant, sustainable, and carbon-neutral energy feedstock by optimizing both its yield and composition to facilitate downstream conversions to fuel and electricity. Working independently with different lignin-deficient mutants, the partners have discovered novel genes that mitigate the growth defects [so-called lignin-modification-induced dwarfism (LMID)] seen in severely lignin-depleted plants. Revealing the mechanism(s) by which this mitigation occurs is critical to fundamental understanding and useful manipulation of how plants partition carbon and may enable biomass manipulation for carbon sequestration in the future.

In this short first partial year, we have identified more than 20 lines that suppress LMID in the highly dwarfed *ref8* (*c3h* gene) lignin mutants, and have started to map the suppressor genes, named as *growth inhibition relieved (gir)*, by next-generation-sequencing-based bulked segregant analysis; we have made good progress in identifying *GIR* genes involved in LMID. Disruption of some of these same genes does not restore

the growth in other mutants. An independent approach has identified the involvement of *REF4* and *RFR1* in LMID. This study has led us to propose a model whereby mutation of *c3h* produces a feedback signal, the interpretation of which requires REF4/RFR1 and that results in widespread transcriptional reprogramming and ultimately dwarfing. We are now working on understanding how these transcriptional changes lead to LMID. We are using dexamethasone-inducible transgene constructs for REF4 and RFR1 (pOpOn REF4/RFR1) that will enable a time-resolved understanding of the transcriptional cascade that leads to LMID in the context of the *c3h* mutation.

For this project, the aim will be to target the orthologs of high saccharification mutants previously identified in *Arabidopsis* for implementation in energy crops (barley, poplar). Effects will be 'stacked' by targeting multiple genes in RNAi constructs. This can be accomplished by down-regulating both lignin biosynthetic genes and genes identified in mutant screens.

To support this work, the next step in semi-automated metabolite profiling has been developed and described. The Candidate Substrate Product Pair (CSPP) algorithm searches for pairs of m/z features (UHPLC-HRMS peaks characterized by a particular retention time and m/z value) for which the mass differences correspond with those expected for well-known enzymatic reactions. The full high-throughput structural characterization pipeline comprising conversion type selection via peak pair generation and the CSPP algorithm including the MS^2 spectral similarity, correlation analysis, and peak grouping algorithm and its application on Arabidopsis leaf extract profiles has been published. At the same time, NMR assignment methods continue to be improved, and were validated in the analysis of triple mutant lines that have their lignins derived almost entirely from the typically minor monolignol, *p*-coumaryl alcohol. The lignins are therefore unlike anything seen before in nature, and the agronomic resurrection of these lines suggests that lignin composition may be pushed beyond previously held compositional bounds.

Background

The Sun is the principal source of energy for our planet, and plant biomass is the major vector by which that energy is captured and stored for human exploitation. This project aims to maximize the utility of plant lignocellulosic biomass as an abundant, sustainable, and carbon-neutral energy feedstock by optimizing both its yield and composition to facilitate downstream conversions to fuel and electricity. Working independently with different lignin-deficient mutants, the partners have discovered novel genes that mitigate the LMID seen in severely lignin-depleted plants. Revealing the mechanism(s) by which this mitigation occurs is critical to fundamental understanding and useful manipulation of how plants partition carbon and may enable biomass manipulation for carbon sequestration in the future. Up to now, much of this work has been performed in the model plant Arabidopsis. However several partners (WB, CC, CH) have developed the capability to make similar genetic changes in poplar and barley, two genuine crop plants with world-wide distribution that are also recognized as ideal models for other trees and grasses, respectively. This project is intended to enable maximal lignin modifications to be generated in plants where normal vascular integrity, strength and disease-resistance are maintained. Deploying these game-changing modifications coupled with the new discoveries arising from this project in crop plants (initially barley) will

open the door for scalability in other agricultural and energy crops. With large-scale deployment and the improved economics enabled by lignin optimization, these advances will underpin the development of cost-efficient biomass-based biofuels and achieve substantial reductions in global greenhouse gas emissions.

Introduction

In the past year, many studies have highlighted the potential for improving biofuel production. As the phenolic polymer lignin coats and is cross-linked to other cell wall components, the reduction or alteration of this compound has long been a target for improving sugar release from cell walls. In the past year, new targets for the reduction of lignin have been identified. The enzyme caffeoyl shikimate esterase (CSE) has been shown to participate in the production of lignin in *Arabidopsis thaliana*, and disrupting this enzyme in *Arabidopsis* improves saccharification.[1] Additionally, defects in the Mediator genes *med5a* and *med5b* are able to rescue the growth defect of the lignin mutant *ref8* while maintaining improved saccharification.[2] In addition to reducing lignin, modifying lignin can also have beneficial effects towards improving saccharification. The expression of a gene from *Angelica sinensis* in poplar is able to modify lignin to make it more amenable for degradation and hence, saccharification.[3] The recent publication of a draft barley genome has allowed for identification of orthologs of genes identified in other species.[4] Recently, an analysis of the *CslF* gene family in barley using the new draft genome identified three members of this family previously unknown in existing databases.[5] Therefore, after identification of genes important for saccharification in other species, we intend to target these genes in crop plants, such as barley, to improve saccharification.

Understanding the molecular mechanism(s) whereby suppressor mutations can restore the biomass yield of reduced-lignin CCR-, C4H- or C3H-deficient plants is of crucial and fundamental importance as it will reveal the basic principles whereby plants partition carbon to lignin and other metabolic compartments. Knowledge of these mechanisms will open up valuable opportunities for greater useful manipulations to lignin either by maximizing yield in saccharification-optimized plants, or increasing carbon allocation and storage capacity in organs such as roots to sequester carbon in soils for longer durations. Fundamental and novel research to illuminate these mechanisms is a major feature of the current work. Although it is possible that LMID may come about via more than one mechanism, a mechanistic understanding of one of them via his recent discovery that the Mediator complex influences carbon allocation to the lignin and phenylpropanoid biosynthetic pathways by transcriptional co-activation or co-suppression of suites of co-regulated genes. Powerful tools, including revolutionary techniques for speeding up the identification of unknown metabolites are being used to investigate these mechanisms.

Progress under the Various Tasks

1. Determine the mechanisms and genes required for LMID

This project aims to discover novel approaches to lignin management for efficient and economical production of sustainable biomass-derived energy. One of the prime objectives is to understand the yield penalties associated with lignin modification. We have focused on elucidating the underlying mechanism of LMID by identifying the

suppressor mutations that rescue the dwarf phenotype of the Arabidopsis *ref8* (*c3h*) mutant by next-generation sequencing.

Because *ref8* exhibits severe LMID and is sterile, a *ref8* mutant line (*ref8*^{pOpON}) carrying a pOpON:C3H construct which enables chemically inducible expression of C3H has been generated.[6] When sprayed with dexamethasone, *ref8*^{pOpON} grew significantly better and was fertile, whereas in the absence of the chemical inducer, *ref8*^{pOpON} behaved the same as the original *ref8* mutant. The bulked seeds for *ref8*^{pOpON} were mutagenized and grown with dexamethasone induction to obtain M2 seeds. M2 plants that had the ability to grow relatively normally in the absence of dexamethasone were identified as *ref8* suppressors.

We have finished screening all M2 plants and identified more than 20 suppressor lines and started to map the suppressor genes, named as *growth inhibition relieved (gir)*, by next-generation-sequencing-based bulked segregant analysis.[7] We backcrossed a recessive suppressor to the original *ref8*^{pOpON} line to generate a F2 mapping population. The F2 plants segregated to two phenotypic groups, *ref8* and suppressor. For each group, we pooled at least fifty individual plants and extract DNA from the bulk. The two DNA samples were subjected to Illumina sequencing. Mutations were identified for each of the two DNA samples by comparing the sequence data with Arabidopsis Col-0 genome. Candidate genes were identified based on the expected allele frequency of the causal mutation in the two bulks (Figure 1). So far, we have obtained candidate genes for six suppressor lines. We have already carried out complementation test for one line and identified one *GIR* gene (Figure 2).

LMID greatly limits the deployment of current lignin engineering strategies for biofuel feedstock improvement. Although its mechanism remains unknown, it is now clear that LMID is a genetically controlled process.[2,6] We have isolated more than twenty suppressor mutants of an Arabidopsis lignin mutant and made good progress in identifying *GIR* genes involved in LMID. It is expected that more *GIR* genes will be identified as the project continues. Further characterization of these *GIR* genes in the near future will provide us more clues to the underlying LMID mechanisms, enabling the development of novel strategies to optimizing yield in lignin-modified plants.

We will continue to map the rest of the suppressors to identify other *GIR* genes. We will also test whether these genes is involved in LMID of other lignin mutants such as *ccr1* and *c4h*. We will also determine the molecular mechanism how these *GIRs* mediate LMID.

2. Identify the signaling and transcriptional changes that lead to LMID in c3h plants

We have previously shown that the REF4 and RFR1 subunits of the Mediator complex are required for LMID in *c3h* plants. Disruption of these genes does not restore growth to a C4H-deficient mutant, and tests with other dwarf lignin mutants is underway. Using RNAseq we showed that more than 30% of the genome is differentially expressed in *c3h* plants compared to wild-type plants. Interestingly, disruption of REF4 and RFR1 in the *c3h* mutant restored the transcription of almost 90% of these genes back to wild-type levels).[2] This has led us to propose a model whereby mutation of *c3h* produces a feedback signal, the interpretation of which requires REF4/RFR1 and that results in widespread transcriptional reprogramming and ultimately dwarfing. We are now working on understanding how these transcriptional changes lead to LMID. The extreme growth

defects of *c3h* mutant plants and the resulting pleiotropic effects of these defects on normal gene expression confounds the identification of immediate target genes that lead to LMID. To dissect the direct REF4/RFR1-dependent targets of Mediator responsible for dwarfing in *c3h* plants from downstream compensatory gene changes, we are using dexamethasone-inducible transgene constructs for REF4 and RFR1 (pOpOn REF4/RFR1) that will enable a time-resolved understanding of the transcriptional cascade that leads to LMID in the context of the *c3h* mutation.

To this end, we have generated the inducible constructs using the pOpOn vector system and transformed them into a wild-type background. Several independent transgenic lines were isolated and then crossed with the *ref4 rfr1 c3h* mutant to generate the triple mutant carrying inducible REF4 or RFR1 (generally referred to as *ref4 rfr1 c3h* pOpOn). This approach was necessitated by the presence of the same selectable marker in both the knockout background and the pOpOn vectors. Once homozygous *ref4 rfr1 c3h* plants carrying the pOpOn REF4 or RFR1 transgenes are isolated, analysis will continue as previously outlined. This will include determining whether the pOpOn REF4 and RFR1 constructs are functional by comparing the growth and metabolic phenotypes of *ref4 rfr1 c3h* plants carrying inducible REF4 or RFR1 in the presence or absence of dexamethasone. If functional, *ref4 rfr1 ch3* pOpOn plants treated with dexamethasone should phenocopy *c3h* plants. Next will define the optimal tissue, developmental stage and post-induction timeframe to use for our gene expression analysis by growing *ref4 rfr1 c3h* pOpOn plants to different developmental stages, applying dexamethasone to induce REF4 or RFR1 expression, and monitoring metabolic and morphological changes over time. Finally, we will use RNAseq to compare genome-wide transcriptional changes in these plants before and after dexamethasone application and in comparison to *ref4 rfr1 c3h* control plants. This analysis will ultimately lead to understanding of transcriptomic and metabolic changes underpinning LMID in *c3h* plants.

3. Reveal the direct targets of Mediator responsible for dwarfing in ref4-3 plants

Similar to *c3h* plants, *ref4-3* plants display widespread changes in RNA transcript abundance when compared to wild-type plants, including decreased transcription of phenylpropanoid biosynthetic genes. Some of these changes may be brought on as a direct result of mutation of the REF4 subunit of the Mediator complex, but many are likely to be indirect effects of perturbed growth and metabolism. We hypothesize that the dwarf phenotype of *ref4-3* mutant plants is the result of gene expression changes induced by inappropriate activation or repression of genes by Mediator complexes containing the *ref4-3* mutant protein. To identify the direct gene targets of these mutant protein complexes, we are using the same dexamethasone inducible-construct approach as outlined above for the *ref4 rfr1 c3h* mutant. In this instance we have transformed wild-type plants with an inducible (pOpOn) *ref4-3* construct and then crossed this transgene into the *ref4 rfr1* background. We are currently working on characterizing these lines for their ability to phenocopy the *ref4-3* mutant upon dexamethasone application. We are also analyzing the effect of pOpOn *ref4-3* on phenylpropanoid accumulation and gene transcription when induced at different developmental stages. When the best conditions for analysis are defined, we will perform RNAseq to look at changes in global transcript levels following dexamethasone application. We hypothesize that a common mechanism underlies LMID in *ref4-3* and *c3h* plants. We will compare gene expression changes in

ref4 rfr1 plants upon induction of *ref4-3* with the gene expression changes seen in *ref4 rfr1 c3h* plants upon induction of RFR1 and any expression changes identified in both experiments will be considered strong candidates for Mediator targets responsible for dwarfing.

In complementary work supported by non-GCEP funds, we performed a suppressor screen for mutations that rescue the growth or metabolism of *ref4-3* plants to identify other proteins and pathways that contribute to the phenotypes of *ref4-3* plants.[2] Interestingly, in the suppressors identified, growth and metabolic restoration do not always correlate with one another, suggesting that these mutants may provide a basis for disentangling these two pathways. To identify the candidate suppressor mutations, we sent the suppressors for whole-genome sequencing. Five of the suppressors identified thus far contain intragenic missense mutations and others are in genes encoding Mediator tail subunits. The functional interaction of these subunits is consistent with the functional and structural interactions identified in yeast and metazoans. The tail subunits of the Mediator complex make many interactions with transcription factors and are the most divergent of the Mediator subunits, presumably as a result of species-specific functionalization. Additionally, we have two suppressors that had no mutations in any of our candidate genes and thus remain to be characterized. Future work will be aimed at further biochemical characterization of these suppressors, including identification of protein-interaction partners by Co-IP followed by mass spectrometry. This project will identify additional genes and proteins required for LMID in *ref4-3* plants.

4. Profiling of metabolites altered by LMID and identification of new pathways

To establish a pipeline for tasks 14-18 ("The profiling of metabolites altered by LMID and identification of new pathways"), we have continued elaborating a strategy for the high-throughput structural characterization of metabolites using reversed phase ultrahigh performance liquid chromatography (UHPLC) coupled to high resolution mass spectrometry (HRMS). Central in this strategy is the **Candidate Substrate Product Pair** (CSPP) algorithm that was mentioned in the proposal of this GCEP project.[8] The CSPP algorithm searches for pairs of m/z features (UHPLC-HRMS peaks characterized by a particular retention time and m/z value) for which the mass differences correspond with those expected for well-known enzymatic reactions. For example, a pair of m/z features that differ by 162.053 Da might be biochemically connected via a hexosylation. Further support is given by the elution order of both m/z features: in the case of a hexosylation, the "product" m/z feature will be less retained on a reversed phase column than the "substrate" m/z feature. Thus, retention time order and m/z difference determine whether a pair of m/z features is classified as a CSPP. Once all CSPPs for diverse enzymatic conversions are generated, a network can be constructed in which nodes and edges represent m/z features and conversions, respectively (Figure 3). High-throughput structural characterization proceeds subsequently via network propagation starting from nodes associated with known compounds. The structures associated with adjacent nodes can then be resolved by considering the enzymatic conversion represented by the connecting edges. Before the start of this project, the CSPP algorithm was tested using metabolite profiles recorded for methanol extracts of Arabidopsis leaves. In total, 3060 m/z features were obtained. The various conversion types that were considered and the number of CSPPs obtained per conversion type are shown in Table I.

Further support that a CSPP might represent a true enzymatic reaction is given by incorporating **MS^2 spectral matching**, i.e. whenever MS^2 fragmentation spectra were available for the "substrate" and "product" m/z features of a CSPP, their similarity was computed. Additional support was derived from considering the **Pearson correlation** between the abundances of both CSPP m/z features across biological replicates. The use of correlations as a filter was statistically explored during the first period of this GCEP project. It was found that the "high-correlated" CSPPs had a higher chance of being associated with a true enzymatic conversion than "low-correlated" CSPPs ($P=1.7 \times 10^{-4}$). In addition, based on solely "high-correlated" CSPPs, the topology (i.e. the number of edges per node or node connectivity) of the CSPP network showed more similarity to that of a metabolic network (i.e. a scale-free distributed node connectivity) than when all CSPPs were incorporated into the network ($P<0.05$). Therefore, at least for reversed phase chromatography (separating mainly secondary metabolites), the abundances of substrates and products are often moderately to highly correlated, hence, advocating the use of correlation analysis as a filter for CSPP determination and for biochemical pathway elucidation.

Textbook enzymatic conversions are not always occurring frequently in a particular tissue type or species. For example, although acetylations occur frequently in secondary metabolism, acetyl moieties are not often observed among Arabidopsis secondary metabolites. Alternatively, certain reaction types might be restricted to only a few species. Therefore, prior to applying the CSPP algorithm, knowledge of the enzymatic conversion types to include is necessary. For this purpose, a so-called "**peak pair generation"** algorithm was developed at the beginning of this project. Via this algorithm, the number of pairs of m/z features (called *# peak pairs*) for each mass difference between 0 and 250 Da (in steps of 0.001 Da) were iteratively computed (here, in contrast with CSPP generation, the retention time order was not taken into account). This lead to a total of 250,000 considered mass differences. The more prominent conversions could then be determined from a Manhattan plot of the *# peak pairs* versus the mass difference (Figure 4).

When eluting from the column and entering the MS, a compound will yield multiple m/z features upon ionization. This is the consequence of adduct formation, in-source fragmentation reactions and the presence of isotopes (e.g., carbon is present as ^{13}C for 1.1%). This yields a lot of redundancy in the CSPP network. Therefore, m/z features belonging to the same compound should be grouped to allow a more efficient CSPP network propagation. Consequently, during the first period of this project, a **"peak grouping"** algorithm has been implemented in the overall CSPP algorithm. Using this algorithm, the 3060 m/z features in the Arabidopsis leaf metabolite profiles were grouped into 229 compounds. Of these, 145 could be structurally characterized including glucosinolates, flavonoids, benzenoids, phenylpropanoids, (neo)lignans/oligolignols, indolics and apocarotenoids. Remarkably, 61 of the structurally characterized compounds have never been described in plants based on the CAS database.

The full high-throughput structural characterization pipeline comprising conversion type selection via **peak pair generation** and the **CSPP algorithm** including the **MS^2 spectral similarity**, **correlation analysis**, and **peak grouping algorithm** and its application on Arabidopsis leaf extract profiles has recently been published.[8] As a side result, the gas phase fragmentation channels that occur upon MS^2 of benzenoids and

phenylpropanoids were explored,[8] (supplementary info) which will allow an improved MS-based structural elucidation of these compound classes in the future.

5. Determine whether LMID is caused by accumulation of phenolic molecules with hormone activity

No reportable activity on this task to date.

6. Understanding lignin structure and discovery of unknown pathways using novel tools

As they become available, mutants and transgenics generated in this project will all be examined by the latest whole-cell-wall NMR methods (as well as the arsenal of available traditional degradative methods available) to determine the impact of LMID and its suppression on lignin composition and structure. These methods have already proven valuable in showing, for example, the novel nature of the *c3h* mutant lines after disruption of REF4 and RFR1 which, like the dwarfed ref8 lines, have their lignins derived almost entirely from the typically minor monolignol, *p*-coumaryl alcohol.[2] The lignins are therefore unlike anything seen before in nature, and the agronomic resurrection of these lines suggests that lignin composition may be pushed beyond previously held compositional bounds.

In parallel, we will employ another powerful method to identify structural details of the lignin polymer by sequencing small molecular weight lignins. This technique, developed by WB's team,[9] revealed that lignin is still likely to contain unidentified, less abundant units and bonds that could provide the basis for new lignin engineering strategies.

Hiring in this task was problematic due to the delayed start of the project and we 'lost' several candidates, but an excellent postdoc has started just this week. In preparation for the onslaught of samples, work continued largely on other funding to improve the ability to identify and authenticate cell wall components, including polysaccharides and the lignins, and are attempting to improve the quantifiability of 2D NMR methods. We have also developed and published on a set of fluorescence-tagged lignin monomers for use in a variety of studies,[10] and are now developing third-generation methods using click chemistry approaches (to be described in the next report).

7. Scalability via translation to energy crops via barley and ultimately poplar

The aim of this task in the project is to generate crop plants suitable for bioenergy use. To this end, genes identified in previous studies as important for saccharification or LMID will be identified in barley and poplar, and plants deficient in these genes will be generated. This will be accomplished through the construction of overexpression or RNAi constructs to up- or down-regulate these genes or through the use of TILLING populations to identify mutants.

The identification of genes important for saccharification in the model plant *Arabidopsis thaliana* has led to the potential for modifying these genes in crop plants like barley. In a recent mutant screen for improved saccharification in a lignin mutant background, several mutants were identified that reduced lignin content further with a corresponding increase in saccharification potential (Figure 5 and Figure 6).

The genes mutated in some of these mutants have been identified, with the aim of targeting their homologs in barley though transgenic and mutant approaches. Barley transformation is routine and highly efficient in Dundee and we have already used it (in

other projects) to generate lines suppressed in each gene of the lignin biosynthesis pathway, often with beneficial impacts on saccharification.

For this project, the aim will be to target the orthologs of the high saccharification mutants previously identified in *Arabidopsis*. Effects will be 'stacked' by targeting multiple genes in RNAi constructs. This can be accomplished by down-regulating both lignin biosynthesis genes and genes identified in mutant screens. These genes will also be used to screen a barley TILLING population, a method to find induced point mutations in targeted genes.[11] Once generated or found, the plants with altered function of the targeted genes will be analysed for saccharification, lignin content, and effect on plant stature.

Summary of Progress on Goals

The progress towards the goals outlined in our proposal is summarized in the Table II.

Conclusions and Outlook

The isolation of *c3h, c4h* and *ccr1* suppressor lines and the phenotypes of the *ref4 rfr1 c3h* triple mutants provide conclusive evidence that we have identified an unexplored domain within the realm of lignin management. From among this set of experimental materials, we have access to metabolites, proteins, and transcriptional mechanisms by which we can dissect, and gain control of, the biochemical mechanisms that underpin LMID. With these tools we will be able to greatly extend the degree to which lignin modification can be manipulated. In addition, work proposed here has the potential to identify novel pathways that may offer new opportunities for lignin manipulation. These advances will ensure that we will be able to increase the amount of sugars released by enzyme-mediated hydrolysis of cell wall polysaccharides and do so at less expense while simultaneously maintaining the structural and vascular integrity of the plants in which lignin deposition has been modified. Further, with the genes required for LMID in hand, we will be able to readily deploy some of the most promising lignin management strategies that we have already discovered (e.g., CCR or C4H suppression) in field crops used for bioenergy production without reducing yield. Finally, if we are correct that the normal function of the mechanisms that shut down lignin deposition in LMID is to control lignin homeostasis in wild-type plants, then gaining control of these mechanisms may permit enhancement of lignin deposition in plants designed for atmospheric carbon sequestration.

Publications and Patents

This project has only just begun. The following are, however, related to it and/or derive from it.

Disruption of the Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. N. D. Bonawitz, J. I. Kim, Y. Tobimatsu, P. N. Ciesielski, N. A. Anderson, E. Ximenes, J. Maeda, J. Ralph, B. S. Donohoe, M. Ladisch and C. Chapple. *Nature*, 2014, in press.

Systematic structural characterization of metabolites in *Arabidopsis* via candidate substrate-product pair networks. K. Morreel, Y. Saeys, O. Dima, F. Lu, Y. Van de Peer, R. Vanholme, J. Ralph, B. Vanholme and W. Boerjan. *Plant Cell*, 2014, in press.

Visualization of plant cell wall lignification using fluorescence-tagged monolignols. Y. Tobimatsu, A. Wagner, L. Donaldson, P. Mitra, C. Niculaes, O. Dima, J.-I. Kim, N. A. Anderson, D. Loque, W. Boerjan, C. Chapple and J. Ralph. *The Plant Journal*, 2013, **76**, 357-366.

In search of clean, affordable energy. R. Nolen-Hoeksema, Z. Bao, S. Benson, J. A. Dionne, K. Maher, W. Boerjan, C. Halpin, R. Nelson, D. Nichols, J. Ralph and T. S. Ramakrishnan. *Oilfield Review*, 2014, **26**, in press.

Patent Applications

-

Contacts

Clint Chapple: chapple@purdue.edu
Wout Boerjan: woboe@psb.vib-ugent.be
Claire Halpin: c.halpin@dundee.ac.uk
John Ralph: jralph@wisc.edu
Xu Li: sirius_li@ncsu.edu
Yuki Tobimatsu: yukitobi@kais.kyoto-u.ac.jp
Yukiko Tsuji: ytsuji@wisc.edu
Renee Strauch: rcstrauc@ncsu.edu
Han-Yi Chen: hchen29@ncsu.edu
Kris Morreel: krmor@psb.vib-ugent.be
Geert Goeminne: gegoe@psb.vib-ugent.be
Ruben Vanholme: ruhol@psb.vib-ugent.be
Bartel Vanholme: bahol@psb.vib-ugent.be
Chris McClellan: Christopher.McClellan@hutton.ac.uk
Whitney Dolan: wsoltau@purdue.edu

References

1. Vanholme, R.; Cesarino, I.; Rataj, K.; Xiao, Y.; Sundin, L.; Goeminne, G.; Kim, H.; Cross, J.; Morreel, K.; Araujo, P.; Welsh, L.; Haustraete, J.; McClellan, C.; Vanholme, B.; Ralph, J.; Simpson, G.; Halpin, C.; Boerjan, W., Caffeoyl shikimate esterase (CSE), a newly discovered gene in the lignin biosynthetic pathway. *Science*, 341, 1103-1106, 2013.
2. Bonawitz, N. D.; Kim, J. I.; Tobimatsu, Y.; Ciesielski, P. N.; Anderson, N. A.; Ximenes, E.; Maeda, J.; Ralph, J.; Donohoe, B. S.; Ladisch, M.; Chapple, C., Disruption of the Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. *Nature*, in press, 2014.
3. Wilkerson, C. G.; Mansfield, S. D.; Lu, F.; Withers, S.; Park, J.; Karlen, S. D.; Gonzales-Vigil, E.; Padmakshan, D.; Unda, F.; Rencoret, J.; Ralph, J., Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science*, 344, 90-93, 2014.
4. Mayer, K. F. X.; Waugh, R.; Langridge, P.; Close, T. J.; Wise, R. P.; Graner, A.; Matsumoto, T.; Sato, K.; Schulman, A.; Muehlbauer, G. J.; Stein, N.; Ariyadasa, R.; Schulte, D.; Poursarebani, N.; Zhou, R. N.; Steuernagel, B.; Mascher, M.; Scholz, U.; Shi, B. J.; Langridge, P.; Madishetty, K.; Svensson, J. T.; Bhat, P.; Moscou, M.; Resnik, J.; Close, T. J.; Muehlbauer, G. J.; Hedley, P.; Liu, H.; Morris, J.; Waugh, R.; Frenkel, Z.; Korol, A.; Berges, H.; Graner, A.; Stein, N.; Steuernagel, B.; Taudien, S.; Groth, M.; Felder, M.; Platzer, M.; Brown, J. W. S.; Schulman, A.; Platzer, M.; Fincher, G. B.; Muehlbauer, G. J.; Sato, K.; Taudien, S.; Sampath, D.; Swarbreck, D.; Scalabrin, S.; Zuccolo, A.; Vendramin, V.; Morgante, M.; Mayer, K. F. X.; Schulman, A.; Conso, I. B. G. S., A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491, 711-717, 2012.
5. Schreiber, M.; Wright, F.; MacKenzie, K.; Hedley, P. E.; Schwerdt, J. G.; Little, A.; Burton, R. A.; Fincher, G. B.; Marshall, D.; Waugh, R.; Halpin, C., The barley genome sequence assembly reveals three additional members of the *CslF* (1,3;1,4)-β-glucan synthase gene family. *Plos One*, 9, 2014.
6. Kim, J. I.; Ciesielski, P. N.; Donohoe, B. S.; Chapple, C.; Li, X., Chemically induced conditional rescue of the Reduced Epidermal Fluorescence8 mutant of Arabidopsis reveals rapid restoration of growth and selective turnover of secondary metabolite pools. *Plant Physiology*, 164, 584-595, 2014.
7. Michelmore, R. W.; Paran, I.; Kesseli, R. V., Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 9828-9832, 1991.
8. Morreel, K.; Saeys, Y.; Dima, O.; Lu, F.; Van de Peer, Y.; Vanholme, R.; Ralph, J.; Vanholme, B.; Boerjan, W., Systematic structural characterization of metabolites in *Arabidopsis* via candidate substrate-product pair networks. *Plant Cell*, in press, 2014.
9. Morreel, K.; Kim, H.; Lu, F.; Dima, O.; Akiyama, T.; Vanholme, R.; Niculaes, C.; Goeminne, G.; Inzé, D.; Messens, E.; Ralph, J.; Boerjan, W., Mass-spectrometry-based fragmentation as an identification tool in lignomics. *Analytical Chemistry*, 82, 8095-8105, 2010.
10. Tobimatsu, Y.; Wagner, A.; Donaldson, L.; Mitra, P.; Niculaes, C.; Dima, O.; Kim, J.-I.; Anderson, N. A.; Loque, D.; Boerjan, W.; Chapple, C.; Ralph, J., Visualization of plant cell wall lignification using fluorescence-tagged monolignols. *The Plant Journal*, 76, 357-366, 2013.
11. McCallum, C. M.; Comai, L.; Greene, E. A.; Henikoff, S., Targeted screening for induced mutations. *Nature Biotechnology*, 18, 455-457, 2000.
12. Chang, X. F.; Chandra, R.; Berleth, T.; Beatson, R. P. Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. *Journal of Agricultural and Food Chemistry*, 56, 6825-6834, 2008.
13. Van Acker, R.; Vanholme, R.; Storme, V.; Mortimer, J. C.; Dupree, P.; Boerjan, W., Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnology for Biofuels*, 6, 1-17, 2013.

Table I. (Bio)Chemical conversions for CSPP network generation.

Nr	Short	Con	m/z dif	Elu[1]	#CSPP	P.C.
1	*Box*	*β-oxidation*	26.016	1	97	
2	*Qui*	*quinate*	174.053	1	102	
3	*Shi*	shikimate	156.042	1	103	
4	*Tar*	*tartarate*	132.006	1	108	
5	*Cul*	*coumaryl alcohol*	116.063	2	142	
6	*Mal*	*malate*	116.011	1	144	
7	*Rha*[5]	deoxyhexose	146.058	1	144	y
8	*Col*	*coniferyl alcohol*	162.068	2	150	
9	*Cat*	**catechol**	136.016	2	151	
10	*Red*	reduction	2.016	1	152	y
11	*Van*	vanillate	150.032	2	154	
12	*Syr*	syringate	180.042	2	156	
13	*Phb*	*hydroxybenzoate*	120.021	2	160	
14	*Caf*	*caffeate*	162.032	2	163	
15	*Dql*	**dimethoxyquinol**	152.047	2	165	
16	*Hql*	**hydroxyquinol**	108.021	2	169	
17	*Cou*	*coumarate*	146.037	2	169	
18	*Sil*	*sinapyl alcohol*	192.079	2	171	
19	*Iso*	*isoprenylation*	68.063	2	173	
20	*Val*	*vanillyl alcohol*	136.052	2	173	
21	*Fer*	ferulate	176.047	2	174	
22	*Pen*	*pentose*	132.042	1	182	y
23	*Pcl*	*protocatechus alcohol*[4]	122.037	2	183	
24	*Pbl*	*hydroxybenzyl alcohol*	106.042	2	185	
25	*Cal*	*caffeyl alcohol*	148.052	2	194	y
26	*Syl*	*syringyl alcohol*	166.063	2	213	
27	Sin	sinapate	206.058	2	217	y
28	*Qul*	**quinol**	92.026	2	225	
29	Sun	syringyl	226.084	2	245	y
30	Hyd	hydration	18.011	1	254	
31	Gun	guaiacyl	196.074	2	290	y
32	Gly	*glycerol*	74.037	3	292	y
33	Oxy	oxygenation	15.995	1	318	y
34	Ace	acetylation	42.011	3	328	
35	Hex	hexose	162.053	1	341	y
36	Mth	malate_hexose[3]	46.042	3	346	y
37	Met	methylation	14.016	2	493	y
38	Mox	methoxylation[2]	30.011	3	532	y

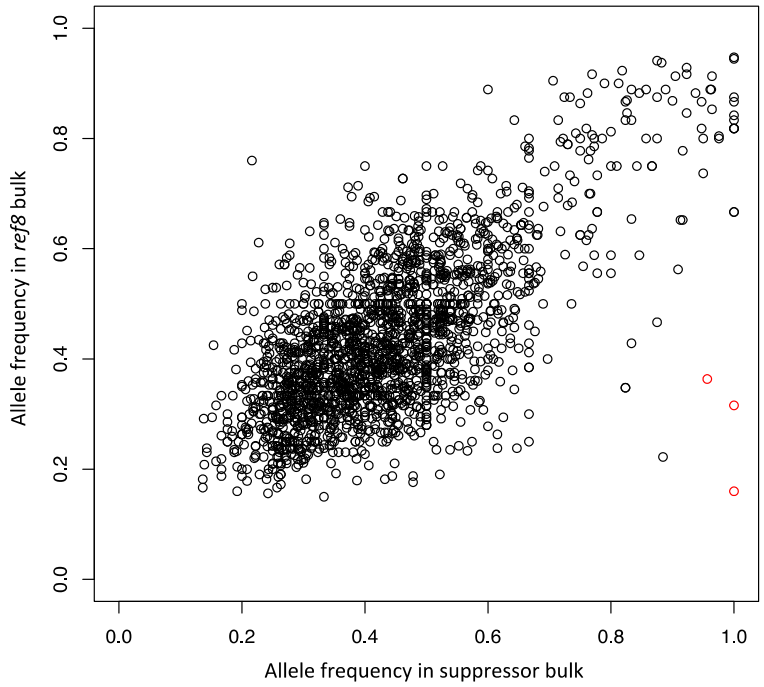
Nr, number; Short, shorthand naming; Con, conversion; m/z dif, m/z difference; Elu, elution behavior of product peak versus substrate peak; #CSPP, number of CSPPs obtained for the conversion; P.C., prominent conversion (based on peak pair generation); y, yes.
[1]Elution behavior: 1, product elutes earlier; 2, product elutes later; 3, not known. *(ctd over)*

[2]Methoxylation is often observed in phenylpropanoid metabolism, yet occurs by a separate oxygenation and methylation enzymatic reaction. [3]Hydroxycinnamoylmalate and hydroxycinnamoylhexose can be transesterified. [4]This conversion can also be the addition of a methoxyquinol. [5]Shorthand name is based on rhamnose. When the shorthand name is underlined or italics, > or ≤ 225 CSPPs were obtained for the conversion, respectively. Conversions written in italics are expected to occur to a lesser extent in *Arabidopsis* secondary metabolism. Conversions written in bold do not represent a true (bio)chemical conversion, but are associated with a structural moiety that is often observed among the profiled metabolites.

Table II. Progress Summary

Task	Description	Who	Prog
1	Use next generation sequencing technologies to identify the *c3h* suppressor mutations	SL	✓
2	Determine whether suppressors of LMID in *ccr1*, *c4h,* and *c3h* are known components of Mediator	CH, SL	❍
3	Investigate whether Mediator mutations can overcome LMID in *ccr1* (task underway) and *c4h* (task complete) plants as is the case for *ref4 rfr1c3h*	CC	✔
4	Depending on the results of the experiments above, the identity of the suppressor genes, and the data from metabolite profiling, develop and test alternative hypotheses regarding other mechanisms of LMID rescue	CH, SL	❍
5	Generate dexamethasone–inducible transgene constructs for *REF4* and *RFR1* using the pOpOn vector system	CC	✔
6	Transform *ref4 rfr1 c3h* mutants with these constructs and isolate several stable transgenic lines for each	CC	✔
7	Confirm that the *REF4* and *RFR1* transgenes are functional by comparison of growth and biochemical phenotypes of T1 plants in the presence and absence of dexamethasone to *c3h* and *ref4 rfr1 c3h* control plants. Plants grown in the presence of dexamethasone should phenocopy *c3h* plants	CC	❍
8	Grow plants (*ref4 rfr1 c3h* mutants containing dexamethasone-inducible REF4/RFR1) to maturity, apply dexamethasone, and monitor metabolic and morphological changes over time, to determine the optimal time point and tissue to use for gene expression analysis	CC	❍
9	Compare global gene expression of *ref4 rfr1 c3h* plants before and after the induction of RFR1 with dexamethasone, as well as non-transgenic *ref4 rfr1 c3h* control plants, using whole-genome transcriptome sequencing technology (RNA seq)	CC	❍
10	Generate dexamethasone-inducible *ref4-3* transgene constructs	CC	✔
11	Transform these transgene constructs into *ref4-* and *ref4 rfr1*-deficient plants	CC	✔
12	Measure gene expression changes upon application of dexamethasone using RNA seq as described above	CC	❍
13	Compare gene expression changes in *ref4 rfr1* plants upon induction of *ref4-3* with the gene expression changes seen in *ref4 rfr1 c3h* plants upon induction of RFR1. Any expression changes identified in both experiments will be considered strong candidates for Mediator targets responsible for dwarfing	CC	❍
14	Develop comparative CSPP networks for mutants with LMID and LMID suppressor mutants	WB	✓
15	Identify metabolites that differ between mutants with LMID and the corresponding suppressors by CSPP and lignomics	WB, JR	❍
16	Purify differentially accumulated compounds	WB	❍
17	Analyze purified compounds by NMR for structural elucidation	JR	❍
18	Chemically synthesize purified compounds for authentication by UPLC-MS	WB, JR	❍
19	Collect aromatic compounds relevant to the consequences of pathway perturbations. Over 200 authentic compounds are already in our ‘library’	WB, JR	✓
20	Assay the purified compounds from 13task 16 and those of task 19 for cell division-promoting activity in the HCS-platform	WB	❍
21	Select molecules with cell division-promoting activity	WB	❍
22	Compare the structures of the cell division-promoting compounds for common chemical entities	WB, JR	❍
23	Apply these molecules to *in vitro*-grown plants transformed with reporter gene constructs that respond to the different plant hormones (P*DR5-GUS* and P*ARR5-GUS*)	WB	❍

Legend: ✓= partially completed; ✔= completed; ❍= in progress.



shown in red.

Figure 2: Complementation of a *ref8* suppressor mutant. The *ref8*^{pOpON} *gir1* plant grows significantly better than *ref8*^{pOpON}. Introducing a wild-type copy of *GIR1* gene back to *ref8*^{pOpON} *gir1* eliminates the growth alleviation.

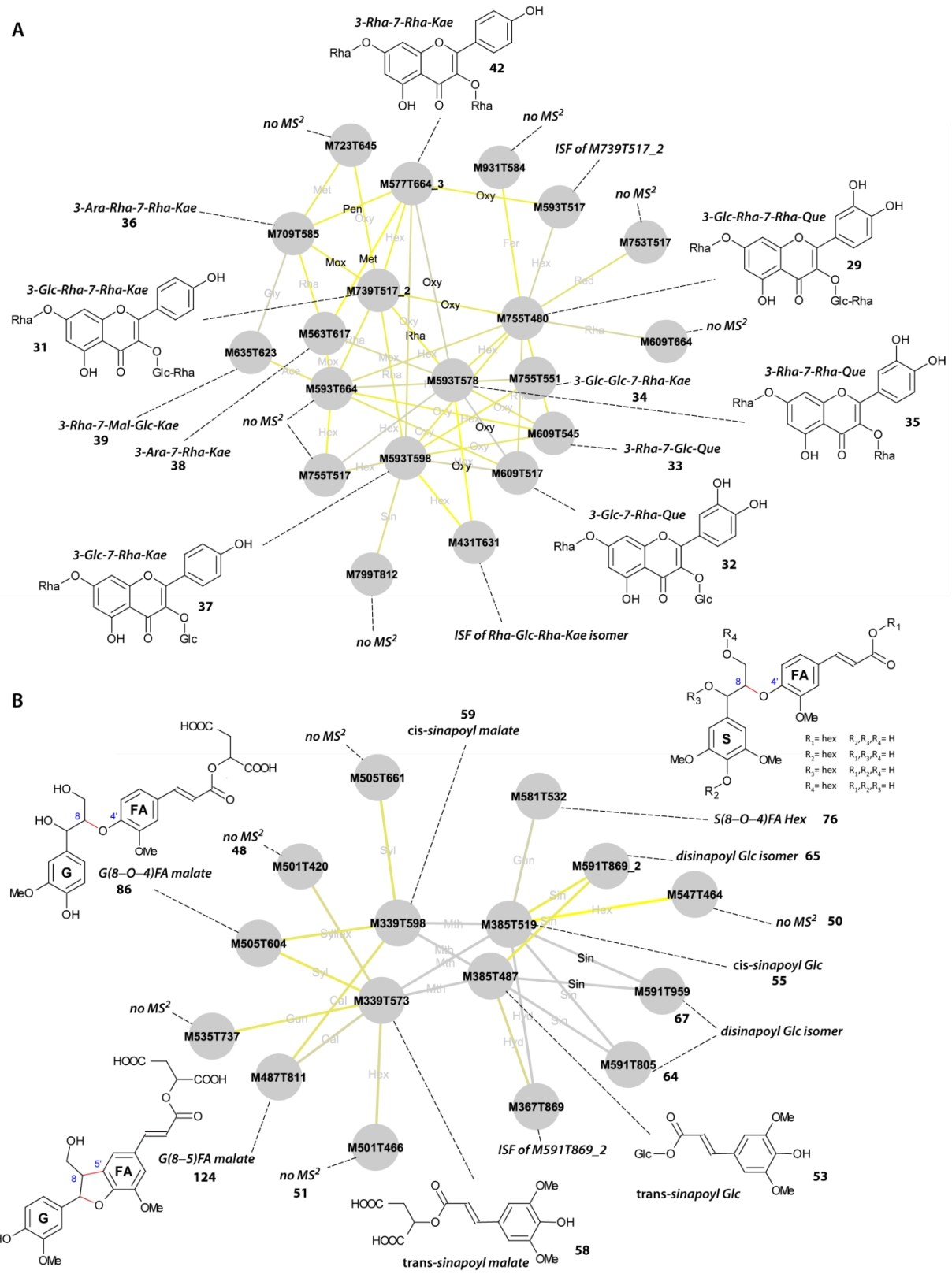

A
3-Rha-7-Rha-Kae
42
no MS2
M723T645
no MS2
M931T584
ISF of M739T517_2
M577T664_3
M593T517
3-Ara-Rha-7-Rha-Kae
36
M709T585
no MS2
M753T517
3-Glc-Rha-7-Rha-Que
29
3-Glc-Rha-7-Rha-Kae
31
M739T517_2
M755T480
M563T617
M635T623
M609T664
no MS2
M593T578
M755T551
3-Glc-Glc-7-Rha-Kae
34
M593T664
3-Rha-7-Rha-Que
35
3-Rha-7-Mal-Glc-Kae
39
no MS2
M609T545
3-Ara-7-Rha-Kae
38
M593T598
3-Rha-7-Glc-Que
33
M755T517
M609T517
3-Glc-7-Rha-Que
32
M431T631
3-Glc-7-Rha-Kae
37
M799T812
no MS2
ISF of Rha-Glc-Rha-Kae isomer
B
59
cis-sinapoyl malate
no MS2
M505T661
M581T532
S(8–O–4)FA Hex
76
R1= hex R2,R3,R4= H
R2= hex R1,R3,R4= H
R3= hex R1,R2,R4= H
R4= hex R1,R2,R3= H
no MS2
48
M501T420
M591T869_2
disinapoyl Glc isomer
65
G(8–O–4)FA malate
86
M547T464
M339T598
M385T519
no MS2
50
M505T604
cis-sinapoyl Glc
55
M385T487
M591T959
67
M339T573
disinapoyl Glc isomer
no MS2
M535T737
M591T805
64
M487T811
M367T869
G(8–5)FA malate
124
ISF of M591T869_2
M501T466
no MS2
51
trans-sinapoyl Glc
53
trans-sinapoyl malate
58

Figure 3: (Previous page) CSPP sub-networks of flavonoids and phenylpropanoids. Nodes and edges represent metabolites and enzymatic conversion types, respectively. Further support that a CSPP might represent a true enzymatic reaction is given by incorporating **MS^2 spectral matching**, i.e. whenever MS^2 fragmentation spectra were available for the "substrate" and "product" m/z features of a CSPP, their similarity was computed. Additional support was derived from considering the **Pearson correlation** between the abundances of both CSPP m/z features across biological replicates. The use of correlations as a filter was statistically explored during the first period of this GCEP project. It was found that the "high-correlated" CSPPs had a higher chance of being associated with a true enzymatic conversion than "low-correlated" CSPPs ($P=1.7 \times 10^{-4}$). In addition, based on solely "high-correlated" CSPPs, the topology (i.e. the number of edges per node or node connectivity) of the CSPP network showed more similarity to that of a metabolic network (i.e. a scale-free distributed node connectivity) than when all CSPPs were incorporated into the network ($P<0.05$). Therefore, at least for reversed phase chromatography (separating mainly secondary metabolites), the abundances of substrates and products are often moderately to highly correlated, hence, advocating the use of correlation analysis as a filter for CSPP determination and for biochemical pathway elucidation.

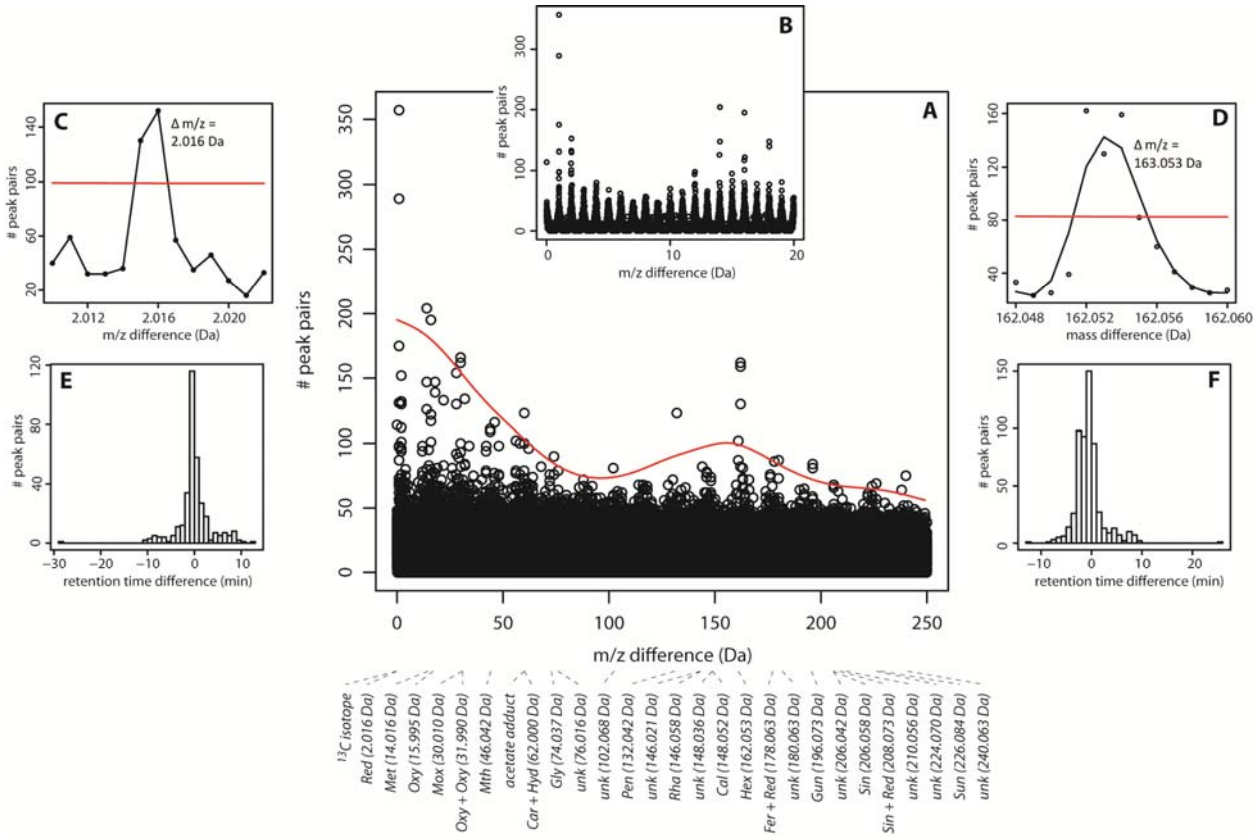


Figure 4: Manhattan plots of the number of pairs of m/z features (*# peak pairs*). The *# peak pairs* for a particular mass difference, up to precisely 3 decimals, was computed. The mass differences between 0.001 and 250 Da and, thus, 250,000 mass differences were considered. (A) Manhattan plot showing the *# peak pairs* (y axis) versus the mass difference (x axis). The smoothed red curve represents the minimum number of peak pairs necessary to consider the mass difference relevant for inclusion as a CSPP conversion type. (B) Manhattan plot with mass differences ranging from 0-20 Da. (C,D) Expansion of the Manhattan plot showing the mass difference region for reduction (C) and for hexosylation (D). (E,F) Distribution of the *# peak pairs* versus retention time difference between both m/z features of the peak pairs obtained for reduction (E) and for hexosylation (F).

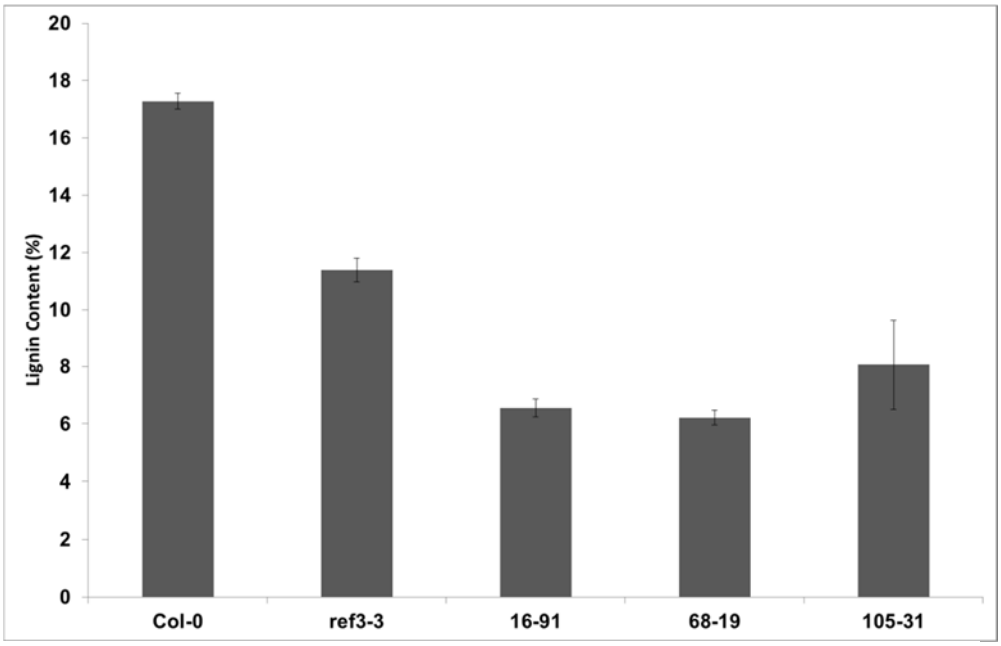


Figure 5: The *ref3-3* enhancer mutants show reductions in lignin content. *Arabidopsis* plants were grown for twelve weeks and stems were ground to powder. Powder was extracted four times with 80% ethanol and once with chloroform:methanol (2:1 v/v). Extracted powder was used in the acetyl bromide assay.[12] Data represents the average of three technical replicates of four biological replicates. Error bars represent standard error of the mean.

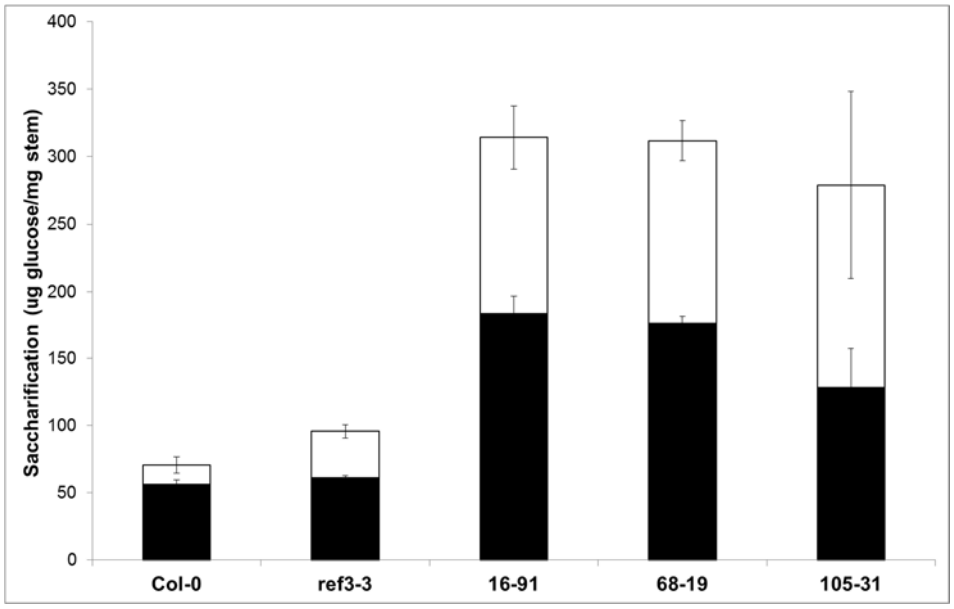


Figure 6: The *ref3-3* enhancer mutants show increases in sugar release upon saccharification. *Arabidopsis* plants were grown for twelve weeks, and stems were collected and ground to powder. Powder was used in saccharification assays for 48 h as in Van Acker *et al*.[13] Black boxes represent no pre-treatment, white boxes represent acid pre-treatment. Data bars represent the average of three technical replicates of four biological replicates. Error bars represent standard error of the mean.