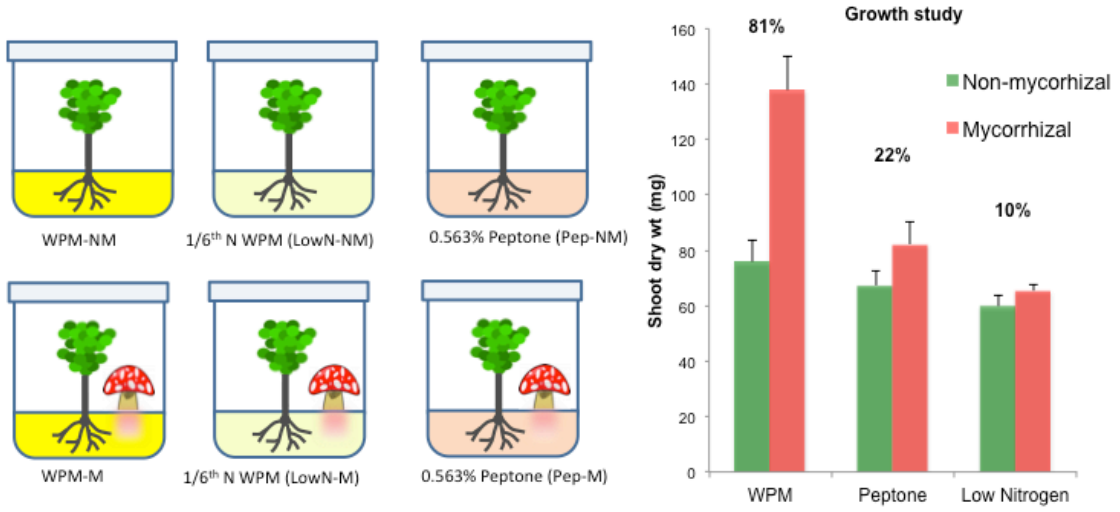
August 30, 2015

# Nutrient cycling for biomass: Interactive proteomic/transcriptomic networks for global carbon management processes within poplar-mycorrhizal interactions




## FINAL REPORT
## Grant Number DE-SC0006652

Leland Cseke, Ph.D., Assistant Professor
302L Shelby Center for Science and Technology
Department of Biological Science
The University of Alabama in Huntsville
Huntsville, AL 35899

## TABLE OF CONTENTS:

## SUMMARY OF PROJECT:

This project addresses the need to develop system-scale models at the symbiotic interface between ectomycorrhizal fungi (*Laccaria bicolor*) and tree species (*Populus tremuloides*) in response to environmental nutrient availability / biochemistry. Using our now well-established laboratory *Laccaria* x poplar system, we address the hypothesis that essential regulatory and metabolic mechanisms can be inferred from genomic, transcriptomic and proteomic-level changes that occur in response to environmental nutrient availability. The project addresses this hypothesis by applying state-of-the-art protein-level analytic approaches to fill the gap in our understanding of how mycorrhizal regulatory and metabolic processes at the transcript-level translate to nutrient uptake, carbon management and ultimate net primary productivity of plants. In most cases, these techniques were not previously optimized for poplar trees or *Laccaria*. Thus, one of the major contributions of this project has been to provide avenues for new research in these species by overcoming the pitfalls that had previously prevented the use of techniques such as ChIP-Seq and SWATH-proteomics.

Since it is the proteins that sense and interact with the environment, participate in signal cascades, activate and regulate gene expression, perform the activities of metabolism and ultimately sequester carbon and generate biomass, an understanding of protein activities during symbiosis-linked nutrient uptake is critical to any systems-level approach that links metabolic processes to the environment. This project uses a team of experts at The University of Alabama in Huntsville (UAH), The University of Alabama at Birmingham (UAB) and Argonne National Laboratory (ANL) to address the above hypothesis using a multiple "omics" approach that combines gene and protein expression as well as protein modifications, and biochemical analyses (performed at Brookhaven National Laboratory (BNL)) in poplar trees under mycorrhizal and free living conditions. Together, the assembled team of experts completed all of the planned milestones set forth in this project. In addition to the planned approaches, several lines of exciting new research have also evolved during the course of this project that involved FTIR Imaging using the National Synchrotron Light Source at BNL. A summary of the approaches used in this project and key highlights are as follows:

***Approach #1: Protein Profiling of Laccaria and Poplar by Targeted and Discovery-based Proteomics:***

Through previous projects with ANL, we have identified ~600 genes that are important in the metabolic and regulatory processes surrounding carbon, nitrogen and phosphorus uptake and management in our system. It was our objective to analyze the protein expression levels of as many of these poplar genes as possible using the mass spectrometry analytic capabilities at the UAB Targeted Metabolomics and Proteomics Laboratory (TMPL). This was originally focused on the development of proteomic techniques specific to poplar trees that would provide quantitative results of specific proteins in each plant tissue. Through the course of this project, significant improvements have been made that allow the efficient quantitation of specific proteins within complex protein extracts. More recently and more significantly, we initiated and optimized new SWATH quantitative proteomic approaches that allow quantitation to be preformed on all proteins within total protein extracts (see ***Discussion of Work Accomplished*** for details). Such methods promise powerful advances in the analysis of protein activity in feedstock crops. To complement these targeted approaches, top-down proteomics approaches using 2D difference gel electrophoresis (DIGE) were also used to explore differences in protein modifications and expression that cannot be assessed using other techniques. The combined efforts represent a significant improvement in the application of proteomics to poplar trees and soil microbes in general.

***Approach #2: ChIP-Seq Analysis of Key Carbon, Nitrogen and Phosphorous Regulators in the Laccaria / Poplar System.***

A number of important carbon and nitrogen metabolism regulatory factors have been identified by their changes in expression in response to nutrient deficiency. However, at the start of this project, nothing was essentially known about the target genes that these proteins control. Consequently, this project worked to identify the target genes of these transcription factors through the development of ChIP-Seq methods specific to poplar species, a previously underdeveloped technology. Here, antibodies to each transcription factor of interest were produced in collaboration with facilities at ANL, and we used these antibodies to collect and analyze ChIP-Seq data to screen potential binding locations throughout the entire poplar genome. The procedures required an extensive amount of optimization to be useful in poplar species, but such procedures are now working well. In addition to the planned identification of potential target genes for each transcription factor, we found that ChIP-Seq provides an excellent opportunity to identify new and novel DNA binding site motifs that are associated with each transcription factor. This provided evidence of new classes of gene promoters that are associated with the target genes, allowing new regulatory networks to be developed. This enhanced understanding of expression, regulation and response was then integrated with ongoing transcriptomic studies to generate potential metabolic network models as part of Approach #4.

***Approach #3: Biochemical Analysis of the Laccaria / Poplar System Under Nutrient Limitation.***

For each of our experimental setups, both phenotypic and specific metabolite factors were assessed from the various plant tissues (leaves, stems and roots) under media conditions with differing amounts of available nutrients. Data on factors such as starch, sucrose, amino acids, nitrate, and phosphate content as well as plant structure and dry weight biomass production were collected to provide critical downstream data that could be associated with gene and protein activity and ultimately with environmental impacts predicted by the systems networks developed in Approach #4. Data collection for this approach was completed fairly early on in the project, including the assessment of all phenotypic and specific metabolite factors from poplar leaves, stems and roots under media conditions with limited nitrogen and phosphorus, compared to control experiments. Four technical replicates for each of four biological replicates of each tissue and each condition were completed by traveling to BNL to make use of the laboratory facilities of Dr. Alistair Rogers. This allowed statistically significant data to be collected on the amounts of starch, sucrose, fructose, glucose, protein, amino acids, nitrates, and phosphates as well as elemental assays, plant structure and dry weight biomass to be assessed. The resulting data is significant to the analysis of the quantified proteins that provided downstream data associated with gene and protein activity. One of the key discoveries is that optimum nitrogen and phosphorus availability are essential to generate trees that direct carbon metabolism toward the biosynthesis of soluble sugars within the mycorrhizal roots. This in turn leads to the establishment of plant / fungal interaction and thus healthier colonization that leads to increased tree biomass. With reduced nitrogen, this series of events breaks down at the expense of colonization, resulting in a greatly reduced capacity to improve biomass.

***Approach #4: Computational Biology to Establish Predictive, Interactive System-Scale Networks Associated with Mycorrhizal Nutrient Cycling.***

A primary goal in our lab is to examine how holistic system design approaches, through the development and utilization of modern molecular techniques, can provide technologies for the deep repair of ecosystems

and the development of solutions for depleted topsoil, broken hydrological cycles, and falling soil fertility. To address this goal, computational models were developed through the implementation of our different approaches to construct gene and regulatory networks that represent how ectomycorrhizal fungi influence rhizosphere symbiotic interactions that assist in tree biomass production. For example, changes in metabolism can be observed using RNA-Seq approaches; however, quantitative SWATH-proteomics revealed that less than 50% of the corresponding protein expression patterns (greater than 2000 proteins) correlate in any way with their RNA expression patterns. Therefore, we have to conclude that care must be taken when making overarching conclusions about plant metabolism based on RNA-Seq data alone. Such approaches lead to several publications, and models that continue to be developed will enhance the understanding of the molecular mechanisms at the mycorrhizal interface where nutrient availability is affected by soil biochemistry and microbial communities. That being the case, further development of SWATH-proteomics, ChIP-Seq, and quantitative imaging techniques will improve our understanding of the true mechanisms of soil community signaling, establishment and maintenance.

It should also be noted that this project supported the work and career development of new scientists that will go on to contribute to the scientific community. This includes the support of two postdoctoral associates, two Ph.D. level graduate students, three Masters level graduate students, six undergraduates, and the partial support of nine other researchers, including the PI and Co-PIs.

## INTRODUCTION AND BACKGROUND:

Healthy soil is an important resource in both natural and agricultural ecosystems. It harbors both chemical and biological sources of nutrients that are required for plant growth, reproduction, insect and pathogen resistance, and response to changing environments. However, current agricultural practices rely heavily on the meticulous management of chemical nutrients. In natural ecosystems, nutrients are taken up by microorganisms and delivered to plants during interactions with plant roots in exchange for fixed carbon. While such symbiotic relationships play an essential role in the regulation of soil nutrient cycling and subsequent carbon management, the complexity and variation of such natural systems has hindered the accurate assessment of the factors that signal, establish and maintain these interactions. Thus, our lab has made use of simplified *Plant-Fungal-Bacterial* laboratory systems and multiple 'omics' approaches to integrate RNA-Seq, SWATH proteomics, ChIP-Seq, biochemical and FTIR imaging analyses to enhance our understanding of how atmospheric carbon is sequestered as plant and/or subsurface fungal biomass during symbiotic interaction and/or nutrient limitation. Our results indicate that molecular events at the protein level are key in drawing conclusions for such processes. We aim to make use of these laboratory systems to study natural ecosystems by providing a mechanism to probe how nature so effectively controls availability of soil nutrients through the activity of beneficial microbes.

This project addresses the need to develop system-scale models at the symbiotic interface between ectomycorrhizal fungi (*Laccaria bicolor*) and tree species (*Populus tremuloides*) in response to environmental nutrient availability / biochemistry. Using our well-established and optimized laboratory *Laccaria* x poplar system, we address the **hypothesis** that essential regulatory and metabolic mechanisms can be inferred from genomic, transcriptomic and proteomic-level changes that occur in response to environmental nutrient availability. It is well known that such tree / fungal interactions greatly benefit tree health and subsequent biomass production, presumably due to the ability of such soil fungi to tap into nutrient stores that the trees would otherwise not be able to access. The molecular mechanisms that regulate such nutrient cycling also play a central role in the accumulation of above and below-ground biomass, and ultimately are important to forest management and atmospheric carbon sequestration.

While modern next generation sequencing (NGS) has greatly expanded our understanding of the transcriptomes of symbiotic relationships, transcript expression levels do not always correlate well with protein expression levels due to many factors, including RNA turnover, differences in translational vs. transcriptional regulatory mechanisms, and posttranslational protein processing and modifications that ultimately control the activity of proteins. It is the proteins that sense and interact with the environment, participate in signal cascades, activate and regulate gene expression, perform the activities of metabolism and ultimately sequester carbon and generate biomass. Thus, an understanding of protein activities during symbiosis-linked nutrient uptake is critical to any systems-level approach that links the internal metabolic processes of such organisms to their environment.

Using a team of experts at the University of Alabama in Huntsville (UAH), The University of Alabama at Birmingham (UAB), Argonne National Laboratory (ANL) and Brookhaven National Laboratory (BNL), we are addressing the above hypothesis using a multiple "omics" approach that combines gene and protein expression as well as protein modifications, and biochemical analyses in poplar trees under mycorrhizal and free living conditions. Specifically, we use targeted mass-spectrometry (and a new technology call SWATH Proteomics), 2-Dimentional discovery-based proteomics, and chromatin immunoprecipitation (ChIP-Seq) characterization of the targets of carbon, nitrogen and phosphorous regulators as well as

biochemical assays and phenotypic changes to identify determinants of keystone systems relationships during symbiotic interactions that control plant carbon management and allocation. The results from such protein-based approaches were then integrated with past and ongoing transcriptomic work at ANL by developing system networks that support the FOA goal to achieve a systems-level dynamic understanding relevant to DOE missions in bioenergy, carbon management, and environmental stewardship.

Each of the chosen approaches makes use of our well-established laboratory *Laccaria* x poplar system that we have optimized in our laboratory over the past 10 years. With this system, we can make controlled alterations to the level and sources of carbon, nitrogen and phosphorus in combinations with and without the presence of interacting *L. bicolor*. The primary objective of this study is to improve the understanding of the molecular mechanisms that orchestrate changes in carbon sequestration and biomass production in response to changes in environmental nutrient availability and/or the presence of beneficial soil fungi. We have addressed this goal through the combined work with our collaborators, including the sharing of UAH post-doc and graduate student personnel to help direct the research at each location. The various aspects of the project can be grouped into several data analysis streams, including phenotypic, genomic, transcriptomic, ChIP-Seq, biochemical, and proteomics data. The research goals and milestones for each of these subdivisions have been completed, and the progress highlights for each is summarized below.

## DISCUSSION OF WORK ACCOMPLISHED:

Dr. Cseke and the assembled research teams at UAH, UAB and ANL have met all of the planned primary milestones for the overall project, including the establishment of several cutting edge technologies that have never successfully been performed before in poplar species. During the course of this project, Dr. Cseke has acted as the research coordinator and facilitated communication with the collaborators at UAB and ANL. Communication avenues consisted of ***(1)*** weekly lab meetings for the UAH research team, ***(2)*** monthly conference calls with experimental team members at UAB or ANL, ***(3)*** Web-Ex platform conference calls to the research groups at ANL and UAB, and ***(4)*** trips to UAB to meet with their research team and discuss progress and troubleshooting approaches. In January 2014 and coming up in January 2016, many of the findings of this work were or will be presented at the PAG meeting in San Diego, CA. In February 2014, coordinated presentations by Dr. Cseke and both the ANL and UAB experimental teams were presented at the DOE Contractor-Grantee Workshop in Arlington, VA. This, combined with talks and presentations at the other DOE Contractor-Grantee meetings (2012, 2013, 2015, 2016) served as one of the primary locations to disseminate conclusions from this research (see ***Papers and Other Products Delivered)***. This meeting was also a primary venue for the annual meeting of UAH, UAB and/or ANL program participants as part of the planned Management Plan activities.

Through other collaboration with ANL, transcriptomic analyses identified ~600 genes that are important in the metabolic and regulatory processes surrounding carbon, nitrogen and phosphorus uptake and management, including ~175 transcription factors that are responsive to changes in N and P concentration (**Figure 1**).

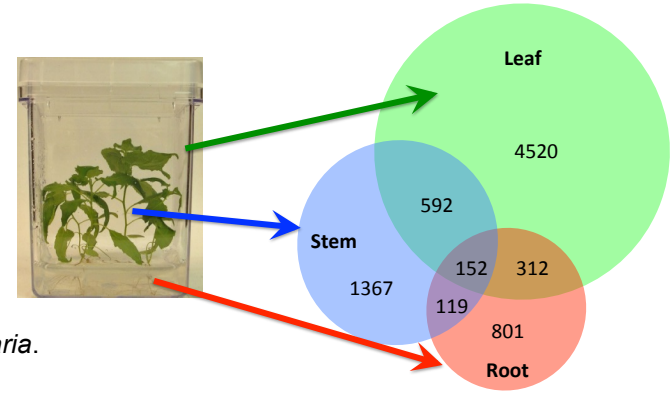



**Figure 1**: Transcriptomics was performed on the leaves, stems and roots of *P. tremuloides* as they interacted with *Laccaria*.

### *Laboratory Laccaria x Poplar System Setup:*

During the course of this project, the *Laccaria* x *Populus tremuloides* laboratory system for both nitrogen limitation and phosphorus limitation experiments with and without ectomycorrhizal interactions was set up in sufficient quantities and biological replicates to allow for all subsequent work. Four technical replicates for each of four biological replicates were used to complete the biochemical analysis of all plant tissues under limited nitrogen and limited phosphorus. In the process, samples were also stored for use with all proteomics work. Thus, the same samples for the biochemical work were used to complete the proteomics approaches, allowing more accurate connection between the data sets. An additional four biological replicates for each tissue and condition, including controls, were set up and stored for use with all ChIP-Seq analyses. Thus, while several changes needed to be made to the originally proposed sample collections, including the use of peptone as a complex organic nitrogen source instead of the proposed alanine, all of the required Nitrogen and Phosphorus limitation tissues were collected, characterized and stored at -80 degrees C for each of the analyses (**Figure 2**).

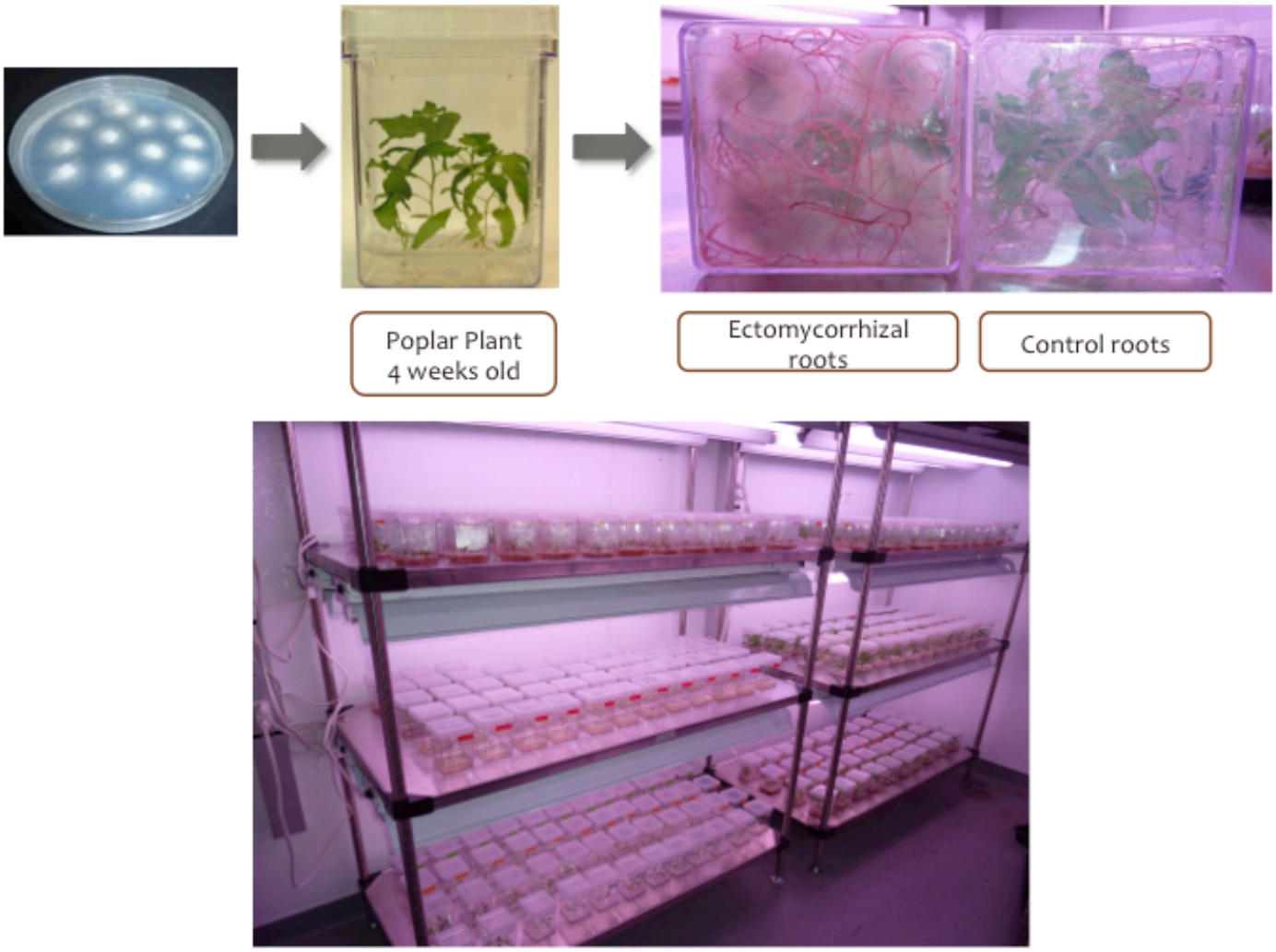



**Figure 2:** The *Laccaria* x poplar laboratory system. Top: *Laccaria* is inoculated into magenta boxes containing poplar seedlings, and the resulting biomass is observed late in development. Even the root system can be seen to increase in biomass in response to fungal interactions. Bottom: an example of the numbers of boxes required for the collection of an appropriate amount of tissue for biochemical analyses (see Approach #3 below).

It should be noted that upon setting up several rounds of biological replicates, proposed changes were made to the original experimental design, based on *Laccaria* developmental results and transcriptomic data collections from early time points. Please refer to the ***Evolving Changes to the Experimental Design*** section of this report for more details of the changes to the experimental design. That being said, the key findings from this section of the research are: ***(1)*** There are significant, phenotypic responses for aspen seedlings when grown on different nitrogen sources, with and without *Laccaria,* and ***(2)*** the availability of nitrogen plays a major role in the ability of *Laccaria* to interact with aspen and contribute to the formation of aspen biomass (Figure 2).

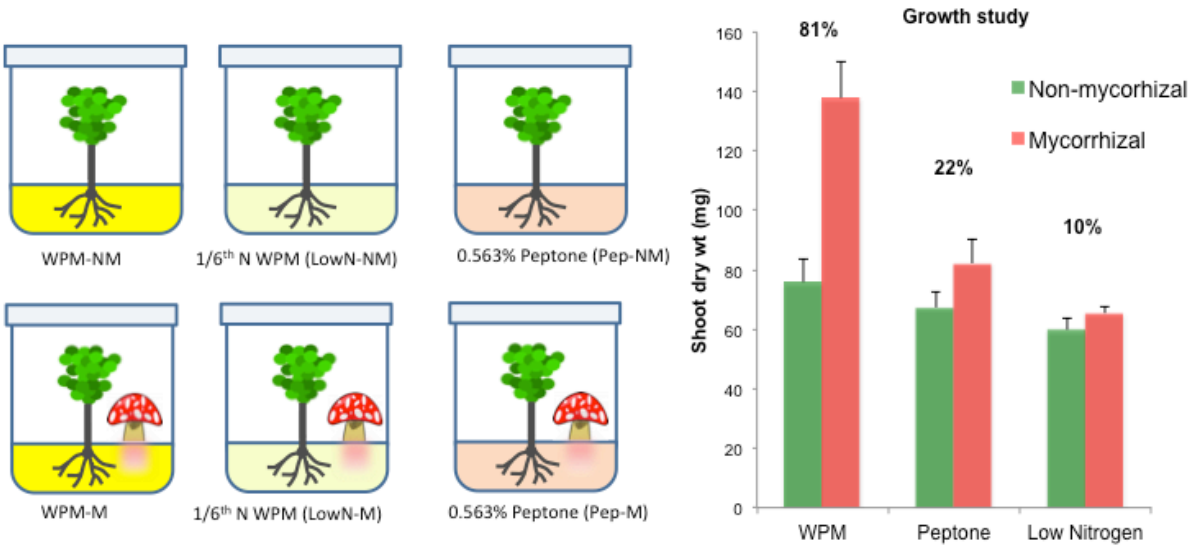



**Figure 3:** System nitrogen conditions and basic biomass response results. Right: WPM refers to Woody Plant Media, a condition with optimal nitrogen levels. Two conditions were used to limit N, one reducing inorganic N to 1/6$^{th}$ optimal and one use Peptone (not accessible by the plant) at a level that mimics optimal. Left: when optimal levels of N are not present, the beneficial impact of the fungus is much reduced. This we later determined is due to the reduced amount of plant / fungal interaction using qPCR approaches.

Phenotypic data was collected for all experimental setups required for the project, including the nitrogen (N) and phosphorus (P) limitation experiments. The primary focus was on fresh and dry weight as an indicator of carbon assimilation and biomass production; however, data on various other parameters were also collected, including leaf count, stem length, leaf shape, and plant coloration. Significant, phenotypic responses for Aspen seedlings were observed in response to nutrient limitation, with and without *Laccaria*. Under N limiting conditions, the presence of *Laccaria* does not significantly impact biomass production compared to conditions where N is plentiful, indicating that N plays a major role in the ability of such soil fungi to contribute to the formation of tree biomass (**Figure 3**). The production of biomass for the P limiting experiments increases in the presence of *Laccaria* independently of the P limiting conditions. The overall results indicate that tree association with ectomycorrhizal fungi can compensate for P limitation more so than it can for N limitation. However, using qPCR approaches, we determined that both forms of limitation also reduce the level of fungal association with the tree roots. In other words, near optimal levels of nutrients are needed to establish and maintain optimal levels of plant / fungal interaction. Once established, however, the mychorrizal interaction provides a continual benefit.

### *Approach #1: Protein Profiling of Laccaria and Poplar by Targeted and Discovery-based Proteomics:*

Because transcript levels do not always correlate with protein levels, proteomics was used to characterize and quantify proteins within the proteome. Initial proteins to be targeted were based on differential expression patterns within the transcriptome and tailored using predicted biological relevance. However, we later decided to target proteins that were clearly involved in the pathways leading to the biochemical and phenotypic changes observed above as a means to better connect phenotype to genotype. Thus, the proteomics work was connected to past transcriptomic and genomic work through the characterization of biochemical product pathways targeted in Approach #3.

The primary goals of this protein profiling approach were:

1. To establish the detectable proteomes of *P. tremuloides* leaves, stems and roots with and without mycorrhizae infection.
2. To determine the relative abundance of proteins identified by transcriptomics to be altered by nutrient conditions and mycorrhizae infection.
3. To determine posttranslational effects of mycorrhizae infection not detected by other methods.

These goals were pursued both at UAH (where protein extractions are made) and at the UAB-TMPL (Targeted Metabolomic and Proteomic Laboratory) facilities, and they evolved into new and powerful techniques for the analysis of poplar proteins using truly cutting-edge proteomics approaches (see SWATH proteomics below).

### *LC-triple quadrupole mass spectrometry:*

The early part of the proteomics analyses encountered difficulties due to the incomplete *P. tremuloides* protein database derived from *P. trichocarpa* DNA sequencing and interference by tree polyphenols on the protein assays, which slowed early progress. However, significant improvements were eventually made in the analysis of poplar proteins detectable by standard proteomics techniques. Here, we developed an updated global protein database for *P. tremuloides* by assembling previously obtained transcriptomic data (>1 billion RNA-Seq reads) using the DOE-National Energy Research Scientific Computing Center (NERSC). This allowed the detection of ~10,500 different proteins in the leaves, ~5,600 in the stems, and ~3,500 in the roots of poplar, a 22.4% increase in successful protein identifications over our previously available databases, and a whopping ~1091% increase from our initial proteomic efforts. It should be noted that these proteins were only detected, not necessarily quantifiable (described below). These proteins were subjected to bioinformatic analysis using the DAVID software, available through NCBI, which identified the major classes and numbers of associated proteins (**Table 1**).

To accomplish these initial proteomic analyses, tissues pulverized at BNL for use in biochemical analyses were also sent to UAB for proteomics analysis. This allowed for accurate comparisons of proteins with specific metabolites. Separate protein databases were established for leaves, stems and roots, and these databases are now available on the UAB-TMPL website (http://www.uab.edu/proteomics/massspec/). While the focus of this site is on Human proteins, the data from this project has now been incorporated into the site's protein prediction and identification tools.

**Table 1:** Identified protein classes found using LC-triple quadrupole mass spectrometry

| Category | Description of Process | Protein Count | % of Total | P-Value |
|---|---|---|---|---|
| KEGG_PATHWAY | Biosynthesis of plant hormones | 76 | 6.4 | 5.70E-11 |
| KEGG_PATHWAY | Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid | 66 | 5.6 | 3.80E-17 |
| KEGG_PATHWAY | Biosynthesis of alkaloids derived from histidine and purine | 64 | 5.4 | 3.20E-13 |
| KEGG_PATHWAY | Biosynthesis of phenylpropanoids | 58 | 4.9 | 1.90E-10 |
| KEGG_PATHWAY | Ribosome | 57 | 4.8 | 1.00E-02 |
| KEGG_PATHWAY | Biosynthesis of alkaloids derived from shikimate pathway | 54 | 4.6 | 5.50E-10 |
| KEGG_PATHWAY | Biosynthesis of alkaloids derived from terpenoid and polyketide | 50 | 4.2 | 1.90E-08 |
| KEGG_PATHWAY | Biosynthesis of terpenoids and steroids | 50 | 4.2 | 4.40E-08 |
| KEGG_PATHWAY | Carbon fixation in photosynthetic organisms | 45 | 3.8 | 4.70E-19 |
| KEGG_PATHWAY | Glycolysis / Gluconeogenesis | 32 | 2.7 | 4.50E-07 |
| KEGG_PATHWAY | Citrate cycle (TCA cycle) | 25 | 2.1 | 1.80E-06 |
| KEGG_PATHWAY | Alanine, aspartate and glutamate metabolism | 24 | 2 | 2.50E-09 |
| COG_ONTOLOGY | Posttranslational modification, protein turnover, chaperones | 23 | 1.9 | 9.70E-03 |
| KEGG_PATHWAY | Pentose phosphate pathway | 22 | 1.9 | 2.90E-06 |
| KEGG_PATHWAY | Pyruvate metabolism | 22 | 1.9 | 9.70E-05 |
| COG_ONTOLOGY | Amino acid transport and metabolism | 20 | 1.7 | 3.40E-07 |
| KEGG_PATHWAY | Arginine and proline metabolism | 20 | 1.7 | 7.60E-05 |
| KEGG_PATHWAY | Glyoxylate and dicarboxylate metabolism | 16 | 1.4 | 4.10E-05 |
| KEGG_PATHWAY | Cysteine and methionine metabolism | 16 | 1.4 | 5.20E-02 |
| KEGG_PATHWAY | Fructose and mannose metabolism | 15 | 1.3 | 1.80E-03 |
| KEGG_PATHWAY | Amino sugar and nucleotide sugar metabolism | 15 | 1.3 | 5.80E-02 |
| COG_ONTOLOGY | Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism | 13 | 1.1 | 3.90E-05 |
| COG_ONTOLOGY | Energy production and conversion | 13 | 1.1 | 8.80E-03 |
| KEGG_PATHWAY | Aminoacyl-tRNA biosynthesis | 13 | 1.1 | 9.20E-02 |
| KEGG_PATHWAY | Nitrogen metabolism | 12 | 1 | 2.50E-02 |
| COG_ONTOLOGY | Carbohydrate transport and metabolism | 10 | 0.8 | 1.70E-03 |

Initially, only 885 proteins were identified using the available *P. trichocarpa* genome protein database (**Figure 4**). This number was not enough to meet the goals of the initial project because very few of them actually correlated with the desired target proteins. Thus, to address this problem, we used a two-pronged approach that included the use of the NERSC facilities to develop new protein databases generated from all available past *P. tremuloides* RNA-Seq data. Here, we developed a *P. tremuloides* protein database with 35,328 unique peptide sequences using our transcriptomics data (>780 million RNA-Seq short reads). This significantly improved the protein identifications for *P. tremuloides* by 48.7% over the *P. trichocarpa* database alone. We also separated the protein extractions into 16 different fractions as a means to reduce the overall peptide background, allowing lower expression proteins to rise above the background (**Figure 5**). This approach, combined with improvements in the protein extractions procedures, then increased the number of individual proteins identified in *P. tremuloides* significantly **(a total of 1091% increase)**, thus resolving the initial complications in all three tissue types (leaves, stems and roots).

Next, in preparation for protein quantification procedures, multiple reaction ion monitoring (MRM) was used to search for proteotypic peptides suitable for quantification on a subset of specific proteins. Multiple reaction monitoring (MRM) has become the method of choice for quantitative analysis of proteotypic peptides for a protein (analogous to oligonucleotides used to quantify DNA and RNA species). The proteotypic peptides are evaluated for their uniqueness using BLAST. However, the BLAST method only examines similarity/uniqueness in letter space. In the MRM method, a selected precursor ion is fragmented by collisional dissociation into product ions, principally b-ions and y-ions. MRMSpace allows users to enter a range of m/z values for their own b- and y-product ions as a means to identify proteotypic peptides that would be quantitative for a given protein of interest. Once fully established, the tool is highly useful for anyone interested in protein quantitation from poplar species. The data from this project has been linked to the *P. trichocarpa* protein annotations as part of the TMPL MRMSpace tools website. http://tmpl.uab.edu/MRMPath/mrmpath.jsp#part3

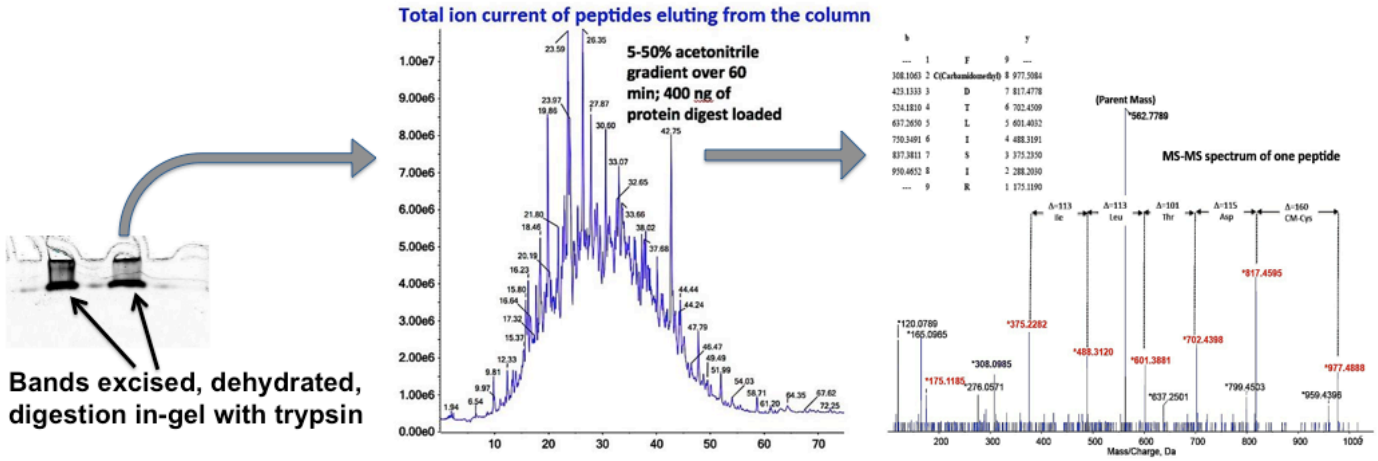
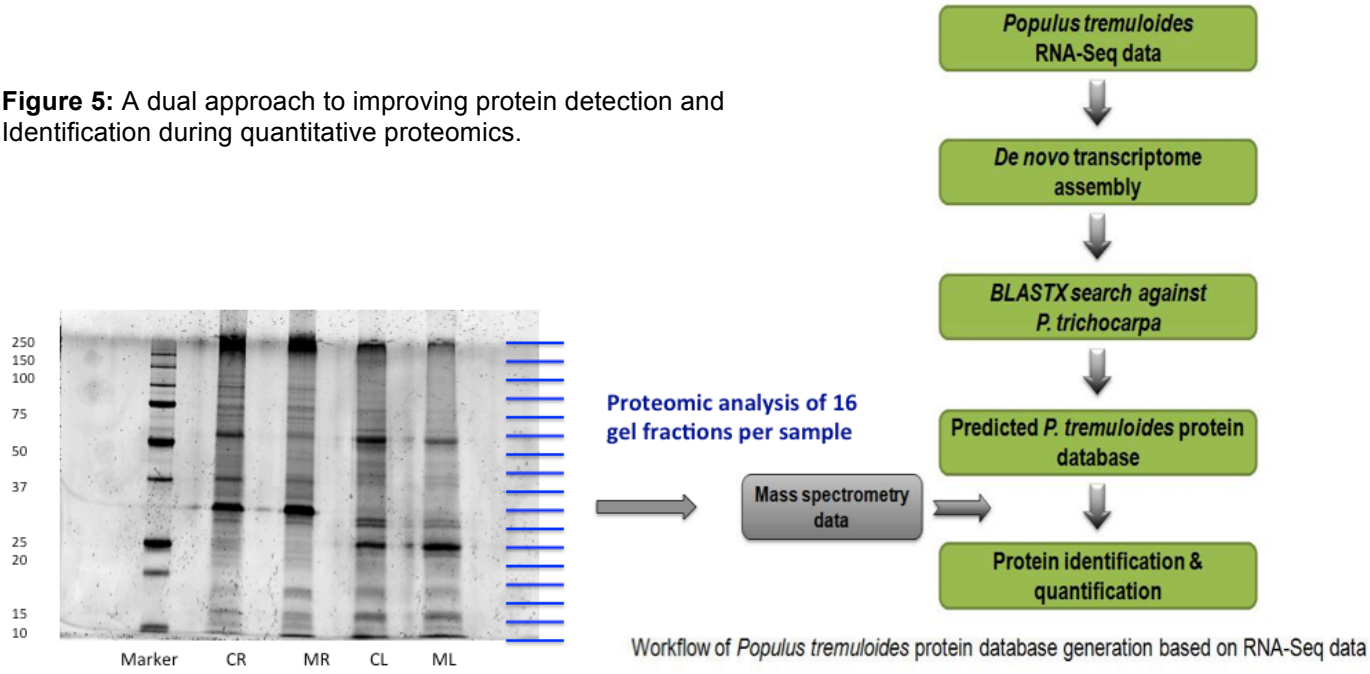



**Figure 4:** Initial targeted proteomics approach using *quantitative LC-triple quadrupole mass spectrometry*.

**Figure 5:** A dual approach to improving protein detection and Identification during quantitative proteomics.




Problems arose when very few of the detected protein had proteotypic peptides that would be usable for typical approaches to quantitative analysis. Even fewer of these proteins were on our lists for proteins of interest. Thus, a major pitfall was encountered for our original quantitative proteomics plans, where we were going to characterize 600 proteins based on differential expression patterns in the transcriptome. Essentially, only a handful of proteins would have fit the criteria, and we had to conclude that this would not do much towards the development of predictive tools that correlate biomass production to omics data. To correct for this, we opted to analyze as many proteins as possible within the plant tissues that change in response to *Laccaria* interaction or environmental nutrient levels. This was done using a promising new proteomics approach that is being developed in animal systems but is still very new in plant systems, called NanoLC-SWATH-MS:

***NanoLC-SWATH-MS:***

To improve upon the initial protein assay results that were based on LC-triple quadrupole mass spectrometry, a new method for the global quantification of proteins in *P. tremuloides* in the individual growth experiments was applied. This is called SWATH-MS (Sequential Windowed data independent Acquisition of the Total High-resolution Mass Spectra). A summary of the SWATH-MS results are as follows:

1. Using nanoLC-SWATH-MS, in the leaf proteome 6,250 tryptic peptides were detected and quantified, representing ~2,000 proteins. 626 proteins were responsive to changes in nitrogen levels and 38 to mycorrhizal infection. Many of these proteins were listed as hypothetical. Of the ones that had identified function, the majority were associated with metabolic processes.
2. In the root, less peptides (784) were observed, but they nonetheless represented 505 proteins. Of these, 91 proteins were responsive to changes in nitrogen levels and 20 to mycorrhizal infection. Again, of those proteins that had identifiable function, the majority were associated with metabolic processes.
3. The proteins in the stems of *P. tremuloides* have been recovered, treated with trypsin and analyzed by nanoLC-SWATH-MS. They are currently being subjected to statistical analysis.

SWATH-MS is a fairly new technology that is driven by the superior speed and high-resolution performance of the TripleTOFTM 5600+ system in both MS and MS/MS mode (**Figure 6A**). As opposed to data-dependent analysis typically used in proteomics which selects which tryptic peptide ions to perform MSMS analysis on, SWATH-MS fragments all tryptic peptides. This technique provides comprehensive quantitative MS/MS data on every component in a sample within a single injection. SWATH Acquisition requires no sample-specific method development, yields better dynamic range than methods that derive quantitation from MS (precursor) data, and delivers limits of detection that were previously only attainable using targeted MRM strategies on triple quadrupole systems. One of the greatest benefits of using MS/MSALL with SWATH Acquisition is that for every sample analyzed, the technique creates a complete digital archive of all possible fragment ions in that sample (**Figure 6B**). This allows re-interrogation of a sample without the need to re-acquire data to test any newly developed hypotheses.

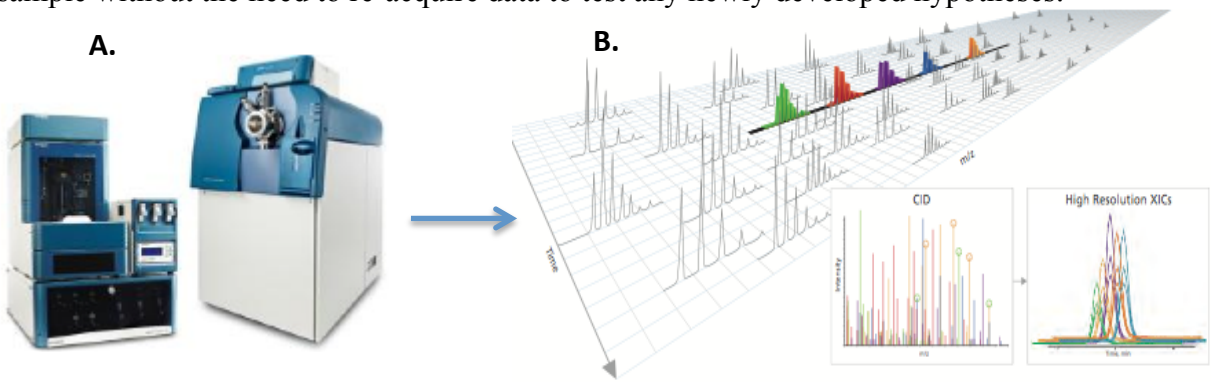



**Figure 6:** A. The TripleTOFTM 5600+ system. B. An example of SWATH data archiving.

SWATH-MS breaks the paradigm of LC-MS proteomics where only the most intense peptide precursors are selected for fragmentation. The precursors are selected with a 1m/z mass window (that's a 2 Da width for a doubly charged peptide). Here, the SWATH-MS mass window is widened to m/z 25 (50 Da) and

fragment ions are collected for 100 msec. All the ions from m/z 400-1200 are scanned in a 32 x 100 msec (total 3.2 sec) cycle. Since peak widths are 20-30 sec wide, enough data points can be collected to accurately measure the area under the peak and hence allow for accurate quantitation. There are two keys to making SWATH-MS work. The first is to create an annotated database where you know the proteins that are present. This we have largely accomplished during our poplar database development. The second is that, unlike analysis on a triple quadrupole instrument where precursor/product ion combinations are pre-selected before the run starts, the ions for the SWATH-MS analysis of each protein peptide can be decided upon AFTER the data are collected, making SWATH a data-independent acquisition (DIA) strategy. A summary of this workflow is shown in **Figure 7**.

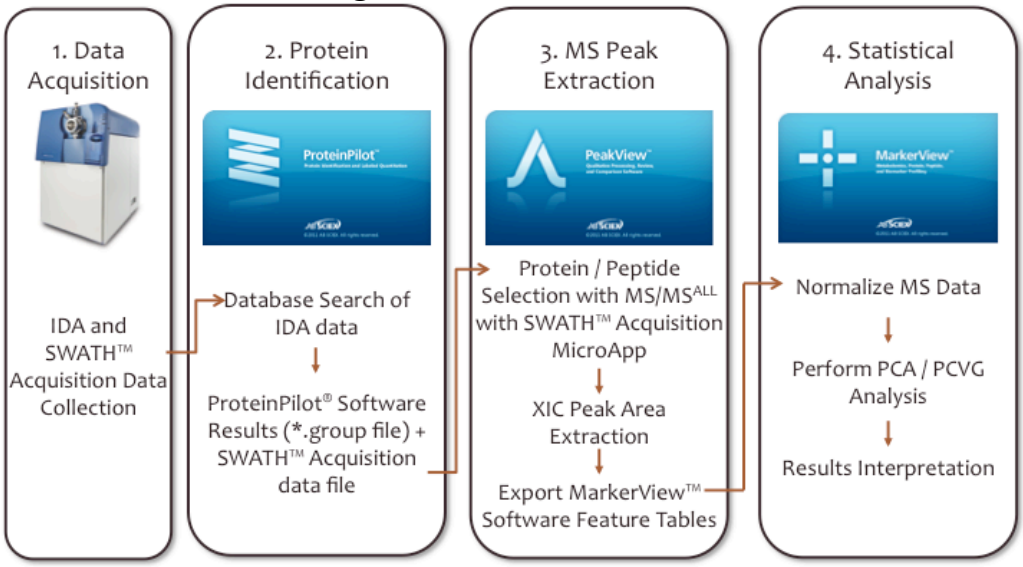
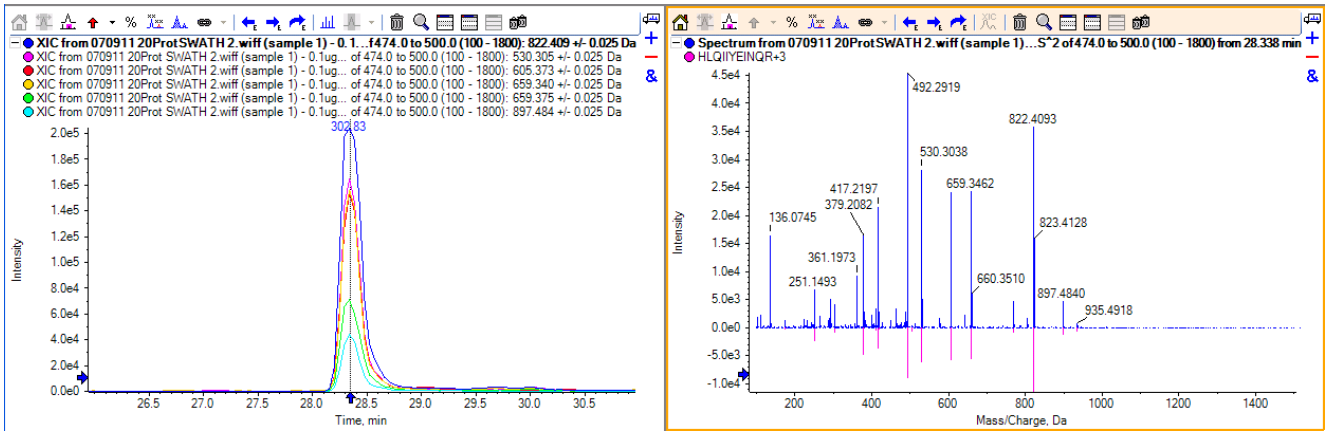



**Figure 7:** An overview of the SWATH-MS data collection and analysis approaches.



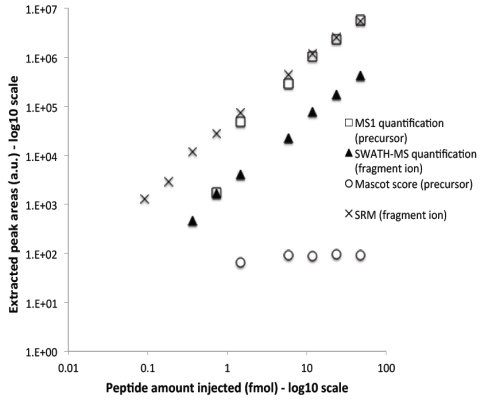




**Figure 8.** Sensitivity assessment for three different methods. The MS1 (old style) quantification goes down to 2 fmol. The SWATH-MS approach goes down to 0.5 fmol but is far more specific since it allows capture of the whole MS/MS spectrum of a peptide.

The power of this approach is that, when we want to know how much of a peptide from a tryptic digestion is present, we simply query the software database to show product ion peaks that

correspond to the expected b- and y-ions for that peptide. Those ions will co-elute since they are derived from the same peptide. They will also co-elute with the peptide identified in the library of SWATH-MS data (**Figure 8**).

Although pre-selected MRM-MS analysis is more sensitive, this approach only detects ONE product ion causing questions about specificity. All in all, SWATH-MS represents a one-stop method for quantifying hundreds to thousands of proteins in a single assay. It is a much more efficient (and effective) way of analyzing the proteins in the *P. tremuloides* samples. **This project represents the first time SWATH-MS has successfully been used to efficiently quantitate *P. tremuloides* proteins.** Here, we used NanoLC-ESI-SWATH-MS to track protein expression patterns during mycorrhizal interaction. Here, we detected and quantified 6,250 peptides, representing the abundance levels of ~2,000 proteins, under each nitrogen limiting and fungal condition (**Figure 9**).

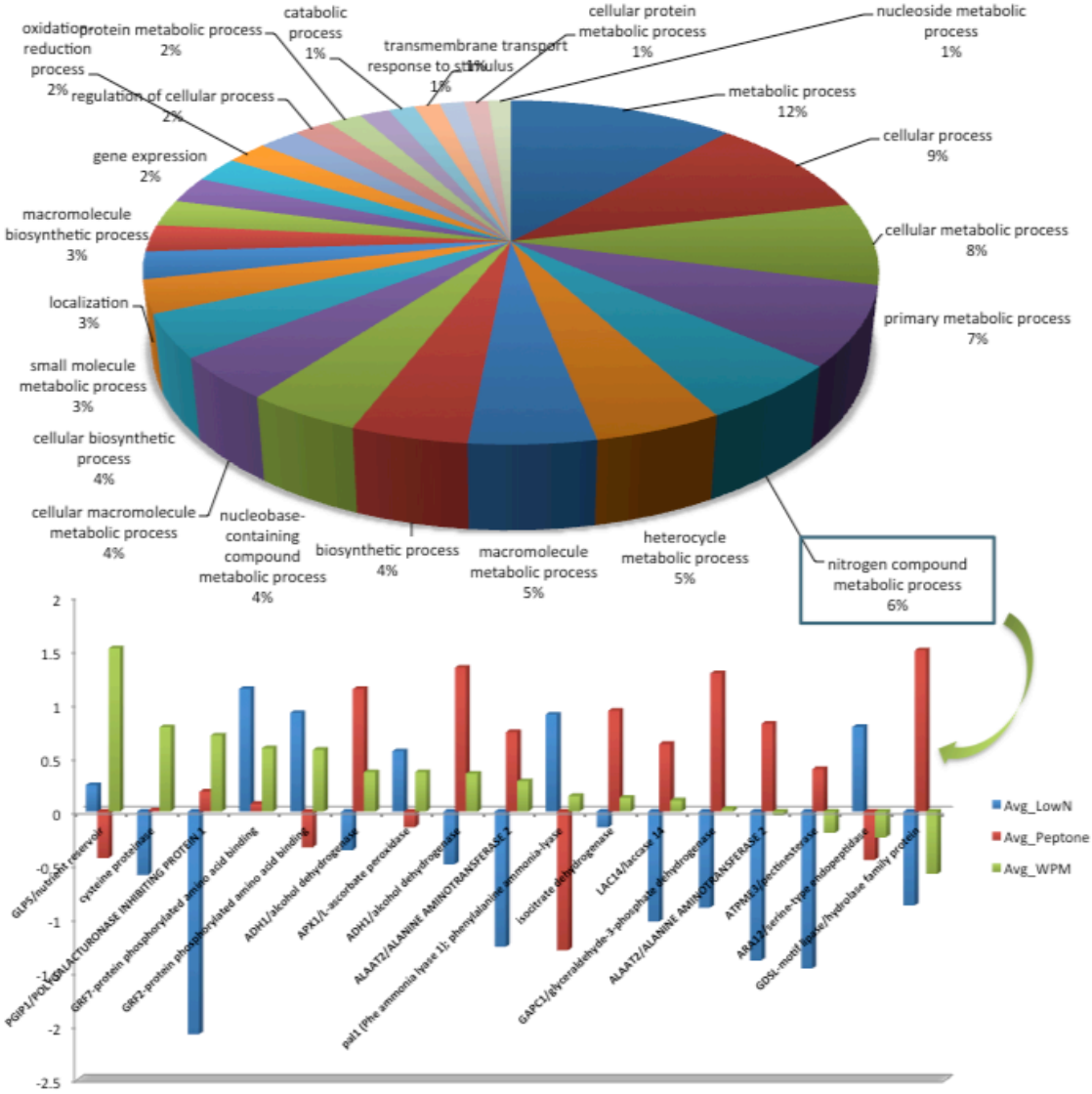



**Figure 9:** Selected results from SWATH-MS Analysis. (Top) Functional classification of identified proteins from aspen root that show differential abundance in response to nitrogen availability. (Bottom) Abundance levels of proteins that belong to functional category 'Nitrogen compound metabolism' on different nitrogen environment. It can

be seen that the expression of individual proteins varies considerably under both inorganic and organic nitrogen limitation.

SWATH-MS of the aspen root proteome successfully detected and quantified 784 peptides, representing the abundance levels of 505 proteins under nitrogen limiting conditions during interaction with *Laccaria*; 91 proteins had differential abundance in response to nitrogen levels (Fig.2a) whereas only 20 proteins showed differential abundance in response to mycorrhizal interaction. Among proteins that were responsive to varying N levels, most proteins were part of metabolic process and were involved in nitrogen compound metabolism and other macromolecule metabolism. Key nitrogen metabolism enzymes, such as Alanine aminotransferase (AlaAT) and Phenylalanine ammonia lyase, were highly responsive to N levels and link primary carbon metabolism with the synthesis of various amino acids (**Figure 9 bottom**). These results delineate metabolic pathways and specific enzymes that play key roles in nitrogen metabolism and biomass development.

***Basic Conclusions from SWATH-MS:***

SWATH-MS is an effective new proteomics approach that brings the processing of quantitative proteomics data into a scenario that is much more similar to how transcriptomics data is currently processed. The resulting SWATH data is stored long-term in a database of quantified peptides, which can be re-interrogated as improved protein databases are developed. As part of our initial proteomics works, we already developed an greatly improved *P. tremuloides* protein database, based on past and ongoing transcriptomics work with ANL. This database, combined with the power of the SWATH-MS approach, allowed us to identify and quantitate over 2000 individual proteins using the SWATH-MS approach compared to our original plan of 600 proteins that were based on prediction from the transcriptomics data.

Among proteins that were responsive to reduced N availability, most were part of metabolic processes and were involved in nitrogen compound metabolism and other macromolecule metabolism. Using *Computational Biology to Establish System-Scale Networks* as part of Approach #4, we also made connections between the expression patterns of the *P. tremuloides* transcripts and their actual proteins. Surprisingly, our analysis revealed that **less than 50% of the quantifiable proteins correlate in any way to their respective RNA expression patterns** (only ~50% correlate with transcripts). Therefore, care must be taken when making overarching conclusions about plant metabolism based on RNA-Seq data alone. That being the case, the further development of SWATH-MS will aide our understanding of the true mechanisms of soil community signaling, establishment and maintenance.

***2D-DIGE analyses:***

High-resolution 2-Dimensional Difference Gel Electrophoresis (2D-DIGE) is a key analytical method in many areas of proteome research. Since its original introduction as a method for resolving complex protein mixtures, the technique has evolved to be capable of generating reproducible and statistically relevant data on cellular proteomes. Here, we used 2D-DIGE as a "discovery-based" approach to identify differences in protein abundance and isoform complexity within the different tissues and treatments. This included the possibility of identifying differences in post-translational modifications that were otherwise undetectable using the above global level proteomics approaches.

Initial protein extractions on control samples were conducted and proteomic analyses were performed at the UAB facilities for selected sets of proteins to be used with 2D-DIGE as a means to look for protein gel

shifts that correlate to protein modifications. Such analyses were performed on free-living poplar and *Laccaria* as well as poplar with interacting *Laccaria* to assess possible technical issues that may have hindered analysis. Assessing the protein content of fractions from the leaves, stems and roots was a consistent problem, and while this delayed the intended progress for some time, these technical issues were eventually resolved using optimized procedures. Thus, differences in protein abundance and/or isoform complexity between leaves from plants that had grown with and without *Laccaria* were analyzed. Protein extracts were labeled with fluorescent dyes, then mixed together, and co-electrophoresed on the same 2D gel, involving isoelectric focusing in the first dimension, and SDS-PAGE in the second (**Figure 10**).

The results indicated that certain proteins could be identified were different in abundance as well as isoform complexity in response to the interaction with *L. bicolor*. However, upon mass spectrometry analysis, these proteins were found to be unknown. Thus, the mass spectrometry is ongoing to identify the proteins indicated by these spots, and such protein identifications should become possible as further improvements are made to the *P. tremuloides* protein databases. This prospect is very likely, as it is suggested by JGI, that the *P. tremuloides* genomic sequence is among those on the priority list for tree species in the near future.

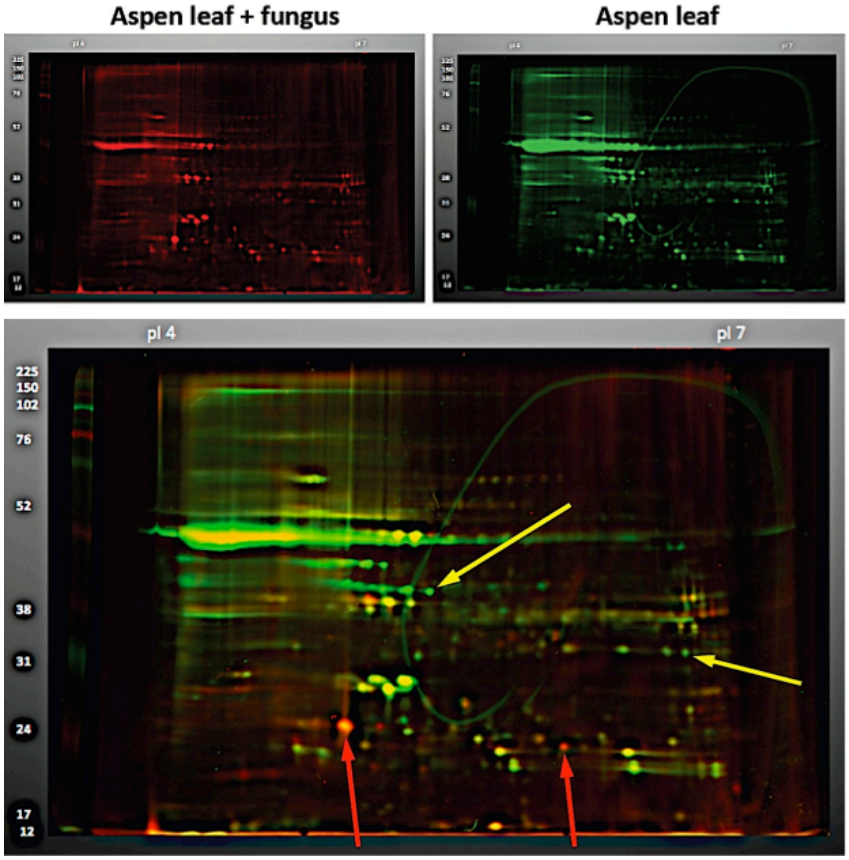



**Figure 10:** An example of our 2D-DIGE gels from *P. tremuloides* leaves grown under normal Nitrogen and Phosphorus condition but co-inoculated either with or without Laccaria. The two top panels show the resulting images from each individual dye, and the low panel shows the merged images where yellow indicate similar results for each dye channel. Red arrows indicate proteins that were both more abundant and structurally shifted in response to *Laccaria* interaction. Yellow arrows indicate proteins that were less abundant in response to *Laccaria*, leaving them more abundant in the control plants.

Nonetheless, to examine some possible common post-translational modifications found in these proteins, differential phosphorylation of individual protein spots was explored observed using ProQ Diamond and Sypro Ruby Red dyes, known to be more specific to phosphorylation modification. Identification of these spots has not yet been pursued, but one exciting possibility is that the SWATH-MS protein database established above has searchable information on posttranslational modifications that can be re-addressed during future research projects.

***Approach #2: ChIP-Seq Analysis of Key Carbon, Nitrogen and Phosphorous Regulators in the Laccaria / Poplar System.***

By combining chromatin immunoprecipitation (ChIP) assays with sequencing, ChIP sequencing (ChIP-Seq) is a powerful method for identifying genome-wide DNA binding sites for transcription factors and other proteins (**Figure 11**). The application of next-generation sequencing (NGS) to ChIP has revealed insights into gene regulation events that play a role in various diseases and biological pathways, such as development and cancer progression. However, the approach has not been used successfully in very many plant species, and as far as we know, this project represents the first time ChIP-Seq as been successfully and reproducibly used in a poplar species. ChIP-Seq has the theoretical advantages of capturing DNA targets for transcription factors or histone modifications across the entire genome of any organism. It can also define transcription factor binding sites, which in turn can reveal gene regulatory networks in combination with RNA sequencing and proteomics. This is one reason we selected ChIP-Seq as a complementary method to examine in Approach #4, where computational biology can identify links between diverse datasets.

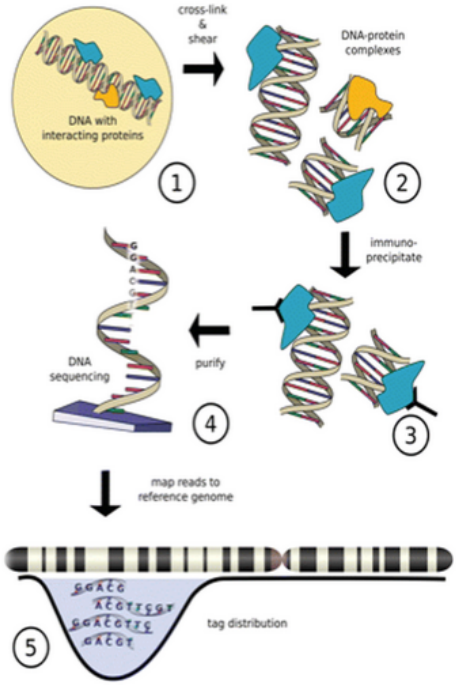



**Figure 11:** Basic ChIP-Seq workflow. Following ChIP protocols, DNA-bound protein, extracted from isolated nuclei is immunoprecipitated using a specific antibody. The bound DNA is then coprecipitated, purified, and sequenced to determine it's identity and location in the genome.

Through a combined effort at UAH and ANL, past transcriptomic data from over 31 biological samples and over 14 billion base pairs of sequence data was examined for transcription factors whose gene expression significantly changes in response to changing nutrient levels (using a bootstrap-style approach to help identify statistically significantly expressed genes). ~175 transcription factors, identified as significant, were carefully scrutinized for gene structure, possible paralogs, and theoretical antigenicity using approaches at both UAH and ANL. This allowed a subset of 16 transcription factors to be selected for protein synthesis and antibody production at the ANL facilities based on biological relevance, predicted structure, and the probability of generating useful antigens (**Figure 12**). The sequences used for cloning were verified using available ESTs as well as short read transcriptomic data. Multiple domains of each protein were produced using the ANL protein production pipeline, which utilizes *Escherichia coli* to mass-produce each protein. Proteins were purified and quantified from extracts using IMAC, and each of the proteins were used for antibody production in each of two rabbits at Pocono

Rabbit Farm and Laboratory, PA. After appropriate bleeds, the antibodies were purified using Protein A columns and evaluated using ELISA assays and affinity to pure protein on Western blots. For more technical detail on how each stage of protein synthesis was conducted, Dr. Collart has submitted a research progress report on behalf of the ANL research team to meet the ANL subcontract requirements.

Each antibody, along with purified protein, was delivered to the UAH lab, where the ChIP portion of the ChIP-Seq procedures was conducted. Each antibody was used to enrich for DNA fragments associated with each transcription factor. In addition, during the development of the methods for this approach, control antibodies were used to collect general ChIP-Seq data from the leaves, stems and roots of poplar trees. This included RNA-polymerase II and H3K9 histone antibodies as well as input DNA (control/non-immunoprecipitated) used to assess the quality of the ChIP process and make correlations to mRNA levels. Other controls included pre-immune serum, no-antibody, and IgG treated samples. In all, over 60 Illumina libraries for 11 of the planned antibodies / proteins were successfully prepared and utilized by the end of the project period. These libraries were sent in batches to ANL for sequencing at the University of Chicago sequencing core. The number of libraries was based primarily on the need to generate appropriate controls that allowed for the identification of true DNA read enrichment over background and non-specific signals.

| TF family | Transcript ID | Description |
|---|---|---|
| MADS-BOX-1 | Potri.002G105600.1 | SVP-like MADS-box TF family protein |
| MADS-BOX-5 | Potri.015G098400.1 | DEFH7-like MADS-box TF |
| MADS-BOX-7 | Potri.012G100200.1 | DEFH7-like MADS-box TF |
| MADS-BOX-10 | Potri.014G074200.1 | PTM5 MADS-box TF |
| MADS-BOX-4 | Potri.002G028400.1 | ALA3-like MADS-box TF family protein |
| MADS-BOX-6 | Potri.001G112400.1 | AGAMOUS-like 19 MADS-box TF |
| MADS-BOX-9 | Potri.017G044200.1 | SVP-like MADS-box TF family protein |
| AP2 | Potri.002G094200.1 | EREBP-like factor |
| bHLH | Potri.002G103300.1 | bHLH DNA-binding family protein |
| AP2 | Potri.005G168700.1 | EREBP-like factor |
| GRAS | Potri.007G063300.1 | Short root-like GRAS family transcription factor |
| GRAS | Potri.017G025900.1 | Short root-like GRAS family transcription factor |
| AP2 | Potri.001G397200.1 | ethylene responsive element binding factor |
| AP2 | Potri.010G183700.3 | DREB-like TF |
| AP2 | Potri.012G134100.1 | EREBP-like factor |
| MADS-BOX-35 | Potri.003G119700.1 | AGAMOUS-like 14 MADS-box TF |

**Figure 12:** Transcription factors targeted for ChIP-Seq analysis

Many of the antibodies used successfully in our optimized *P. tremuloides* ChIP-Seq pipeline were specific to a set of MADS-BOX transcription factors that include those similar to APETALA3 (TF004), AGL19-like (TF006), SVP (TF009), SOC1 (or PTM5, cloned in our lab) (TF010), and AGL14-like (TF035) MADS-BOX proteins. The other successful antibodies were designed for an ethylene response factor (TF017), bHLH (TF018), two AP2's (TF020, TF027.v2), GRAS short root-like (TF025) and DREB (TF028).

As a means of revealing the functional gene regulatory networks that underlie our nutrient limitation conditions, the sequence reads derived from each polyclonal antibody were mapped to the *P. trichocarpa* genome database to identify the target genes regulated by these transcription factors. This made use of the Seq(MACS) Peak-finding Algorithm in combination with annotation of the ChIP peaks (**Figure 13**). Since these steps generate a large number of potential target genes, overlapping peaks between replicate samples were assessed to identify target gene peaks that were not only enriched in each library but were also consistent between replicate libraries that were prepared independently. Once such peaks were established, they were then used to identify the corresponding regions of the genome, allowing promoter regions for each target gene to be extracted and later processed to identify the transcription factor binding motifs that were enriched in each ChIP-Seq library. Thus, genome-wide determination of binding sites for selected

MADS-box TFs has lead to identification of hundreds of potential target genes including GDSL-lipases & Glycosyl hydrolase family 1 (GH1) β-glucosidase, involved in the regulation of plant development, lignification, and cell wall remodeling. Also, many potential target genes regulated by MADS-box TFs, such as MATE efflux family protein and NB-ARC domain protein, play critical roles in the biotic and abiotic stress responses of plants (**Table 2**).

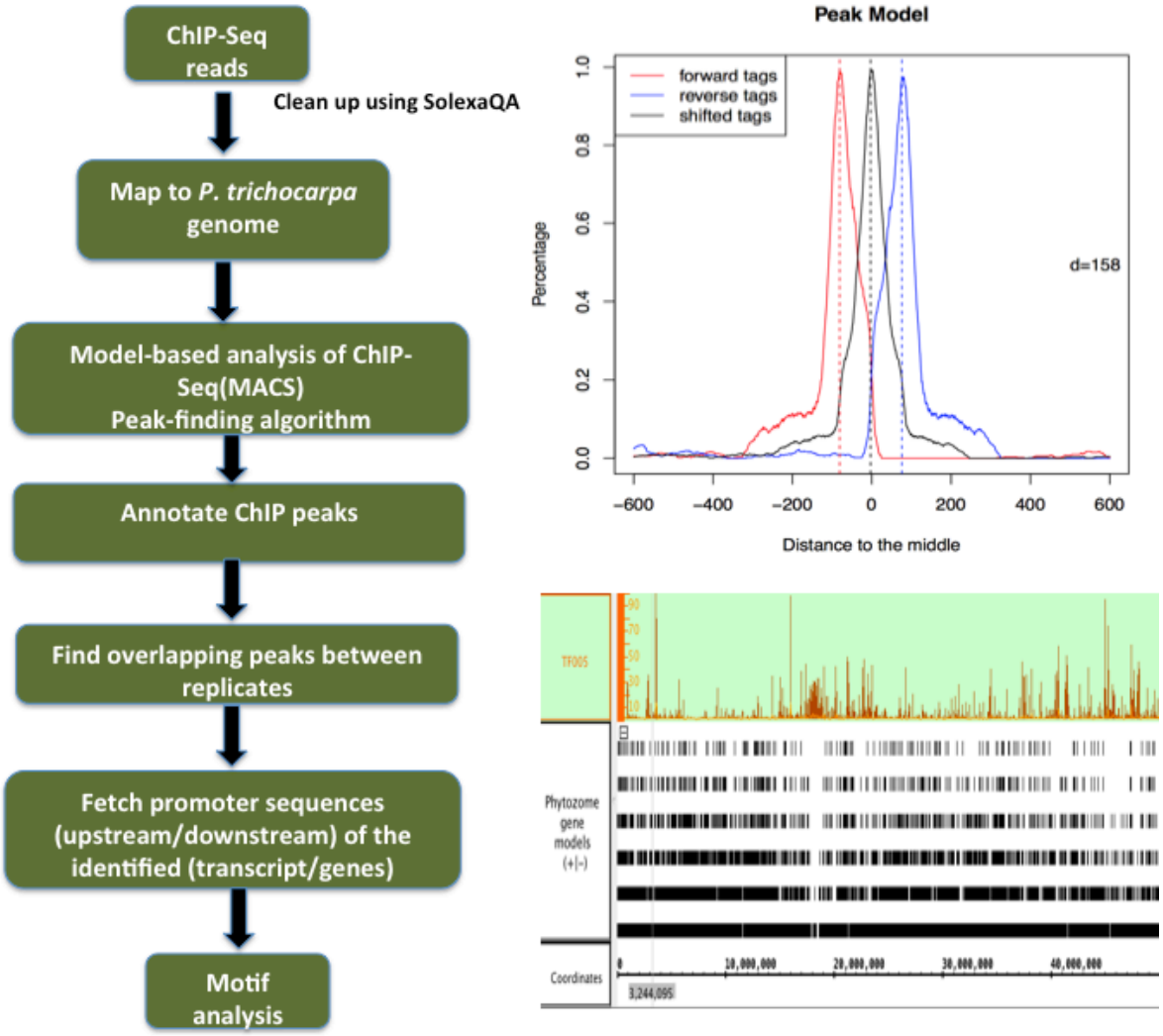



**Figure 13:** Our developed ChIP-Seq pipeline and example results. (Left) The general analysis pipelines used post-sequencing. (Right Top) MACS model for ChIP-Seq tags of MADS-Box TF010 (PTM5) transcription factor using ChIP-Seq data. (Right Bottom) Enriched MACS peaks mapped on *Populus trichocarpa* genome.

**Table 2: List of binding sites and potential genes regulated by MADS-box TFs**

| Space | Nearest Transcript | MACS_Peak | Fold Change | Inside Feature | Start | End | Description |
|---|---|---|---|---|---|---|---|
| Chr01 | Potri.001G126600.1 | MACS_peak_15 | 173.32 | Upstream | 10277814 | 10278171 | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein |
| Chr03 | Potri.003G195900.2 | MACS_peak_92 | 137.88 | five_prime_UTR | 19891775 | 19892130 | NADPH:quinone oxidoreductase |
| Chr04 | Potri.004G054300.1 | MACS_peak_99 | 124.05 | Upstream | 4260494 | 4260845 | GDSL-like Lipase/Acylhydrolase superfamily protein |
| Chr04 | Potri.004G135500.1 | MACS_peak_121 | 222.90 | five_prime_UTR | 15560259 | 15560617 | Leucine-rich repeat transmembrane protein kinase |
| Chr05 | Potri.005G141000.1 | MACS_peak_159 | 505.02 | Upstream | 11885367 | 11885731 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| Chr09 | Potri.009G029500.1 | MACS_peak_256 | 224.19 | five_prime_UTR | 4024130 | 4024503 | OBF-binding protein 3 |
| Chr12 | Potri.012G004000.2 | MACS_peak_305 | 126.28 | Upstream | 348232 | 348602 | Vacuolar iron transporter (VIT) family protein |
| Chr13 | Potri.013G051700.1 | MACS_peak_326 | 194.81 | five_prime_UTR | 3804481 | 3804839 | RmlC-like cupins superfamily protein |
| Chr13 | Potri.013G069300.1 | MACS_peak_328 | 82.80 | five_prime_UTR | 5612587 | 5612959 | MATE efflux family protein |
| Chr14 | Potri.014G010700.1 | MACS_peak_358 | 90.61 | Upstream | 1094104 | 1094460 | NB-ARC domain-containing disease resistance protein |
| Chr14 | Potri.014G033500.1 | MACS_peak_359 | 114.07 | Upstream | 2752731 | 2753085 | phospholipase A 2A |
| Chr14 | Potri.014G150700.1 | MACS_peak_364 | 61.95 | Upstream | 11573601 | 11573806 | alpha/beta-Hydrolases superfamily protein |
| Chr14 | Potri.014G184200.1 | MACS_peak_367 | 51.67 | Upstream | 15930874 | 15931341 | O-Glycosyl hydrolases family 17 protein |
| Chr14 | Potri.014G196400.4 | MACS_peak_405 | 138.05 | Upstream | 18628112 | 18628913 | Maf-like protein |
| Chr19 | Potri.019G071100.1 | MACS_peak_506 | 100.36 | Upstream | 10513864 | 10514220 | O-acyltransferase (WSD1-like) family protein |

In addition to identifying potential gene targets and enriched recognition motifs, our approach utilized and expanded upon the knowledge of known consensus sequences for transcription factors as a means to verify the validity of the results from each protein. Motif analysis using MEME-ChIP discovered distinct motifs for each MADS-box TF (**Figure 14**). Some of these motifs, present in the promoter regions of the target loci, matched known MADS-box motifs, including expected CArG-boxes found in the database of plant TF binding motifs (JASPAR CORE plants). Other well-represented motifs were also pinpointed in the promoter regions of our identified target genes. For example, our tested MADS-BOX transcription factors recognize motifs such as the CArG-box as well as AGL3 and MEF2.

ChIP-Seq is not well established in poplar tree tissues, and thus this project represents the first of its kind to optimize this technology for use in these species. The overall results of our ChIP-Seq analyses have advanced our understanding of biochemical and metabolic pathways associated with cellulose and lignin biosynthesis during plant growth and development. Our results suggest that MADS-box proteins may use the newly identified motifs to specifically recruit additional transcription factors into multi-component regulatory complexes. Moreover, as part of the ChIP-Seq analysis pipeline, Peter Larsen, Co-PI, is also working to develop a novel method for ChIP-Seq data that will allow the analysis of ChIP-data that is applicable to all poplar species and not simply those that have a sequenced genome. This will serve a wide research community working on important model organisms that may face similar hurdles.

| TF Name | Motifs | Sequence | Matched to known motif | Family |
|---|---|---|---|---|
| MADS Box (TF001) | Motif 1 | ATATATAT | | |
| | Motif 2 | AT[TC]CCGA | SOC1 (Other Alpha-Helix) | MADS |
| | Motif 3 | CA[TG]TTGCC | | |
| MADS Box (TF007) | Motif 1 | TATATATA | SEP4 (Other Alpha-Helix) | MADS |
| | Motif 2 | [GT]CCTTTGA | | |
| MADS Box (TF009) | Motif 1 | CCGATT[CA] | | |
| | Motif 2 | ATTGT[TC] | HAT5 (Helix-Turn-Helix ) | Homeo |
| MADS Box (TF010) | Motif 1 | AAT[GAT]CC | FLC (Other Alpha-Helix) | MADS |
| | Motif 2 | TTGAGC[AT] | | |
| | Motif 3 | AA[AT]TCAGT | | |
| MADS Box (TF035) | Motif 1 | GCCGGG[TG] | TCP16, HY5 (Zipper-Type) | Helix-Loop-Helix |
| | Motif 2 | AA[TG]GCAAC | HMG-1 (Other Alpha-Helix) | HMG |
| | Motif 3 | GCACCG[GT]A | RAV1, abi4 (Beta-Hairpin-Ribbon) | AP2 MBD-like |

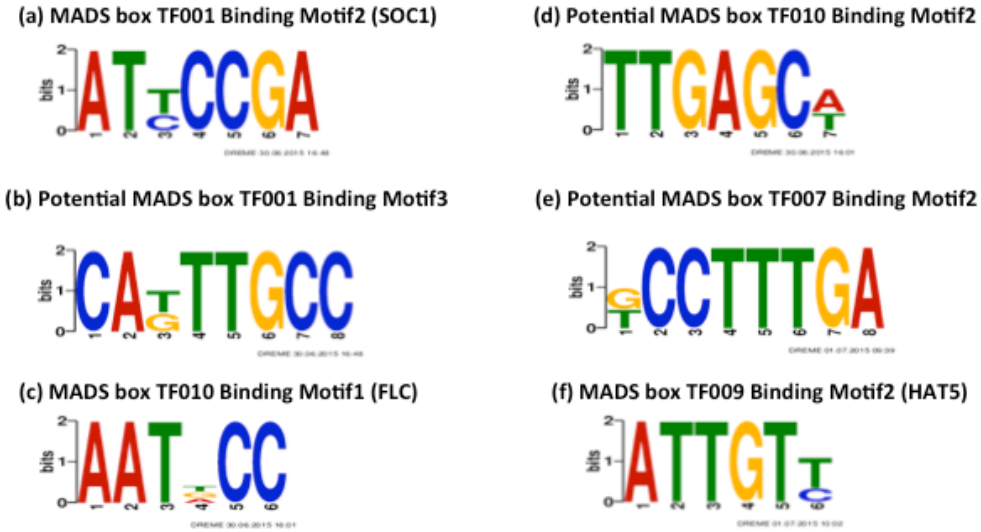



**Figure 14:** Results from motif analysis of enriched MADS-Box transcription factor ChIP-Seq data, revealing new motifs not previously associated with MADS-Box target gene regulation.

### *Approach #3: Biochemical Analysis of the Laccaria / Poplar System Under Nutrient Limitation.*

Biochemical analyses were used to characterize new molecular physiological information that can in turn be used to connect the results of transcriptomics and potentially ChIP-Seq with protein level analyses derived from our SWATH-MS methods for the generation of more streamline predictive tools as part of Approach #4. This, combined with the projects other phenotypic data, allowed us to connect phenotype to genotype through the above analyses of observed protein and RNA expression patterns.

After local and national searches to determine a suitable facility that can address the above goals, arrangements were made to conduct biochemical analyses of carbon, nitrogen and phosphorus compounds to generate differential nutrient allocation data from the various tissues and treatments. The analysis was conducted at Brookhaven National Labs (BNL) in collaboration with Dr. Alistair Rodgers, an expert in the field of biochemical analysis in plants. To ensure statistical significance in the analysis of biochemical compounds, a plan was developed to perform 4 biological replicates each having 4 technical replicates on the leaves, stems and roots of *P. tremuloides* trees from each media type with and without *Laccaria*. Several grams of each tissue biological replicate were cryogenically pulverized, and 26 mg technical replicate samples were characterized for the following parameters: Elemental Analysis (total carbon, hydrogen, and nitrogen), Nitrogen Content (protein, amino acids, and nitrates), Sugar Content (starch,

sucrose, fructose, and glucose), Phosphate Levels, Wet tissue weight, and Dry weight. In addition, the tree tissues were assessed for the following parameters to provide key developmental data for each media type and interaction: Leaf count, Leaf area, Shoot/stem length, Root density, Fresh weight for each tissue, Dry weight for each tissue. While the phenotypic data was collected at UAH, the biochemical analysis was performed at BNL.

Through a series of trips to Dr. Rogers' BNL Lab, two of Dr. Cseke's post-docs and one graduate student completed the biochemical data collection and analyses for both the N and P limitation experiments. The resulting data sets were subjected to various permutations of PCA analyses in order to identify larger trends in the shifting plant metabolism that may result from changes in media composition, interaction with *Laccaria*, or differences in tissue types. Results indicate that tissue type appears to be the parameter having the largest coefficient of variance in the combined data (**Figure 15**).

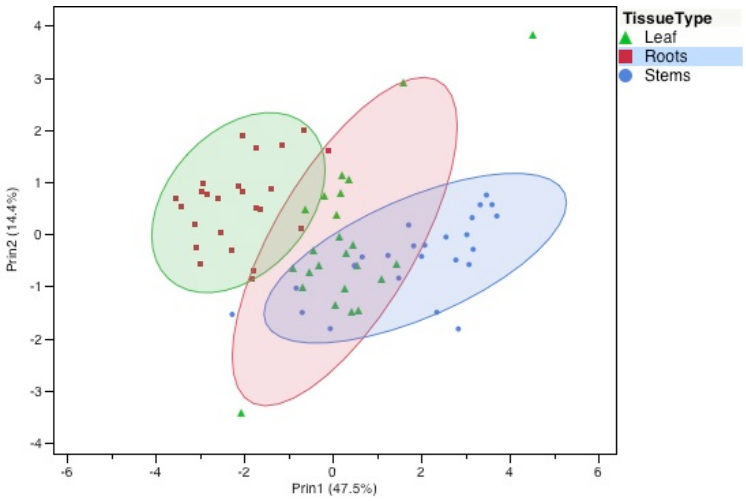



**Figure 15:** PCA of normalized and combined data from Starch, Sucrose, Glucose, Fructose, Proteins, Amino Acids, Nitrates, Phosphates, and Elemental C, N, and H analyses.

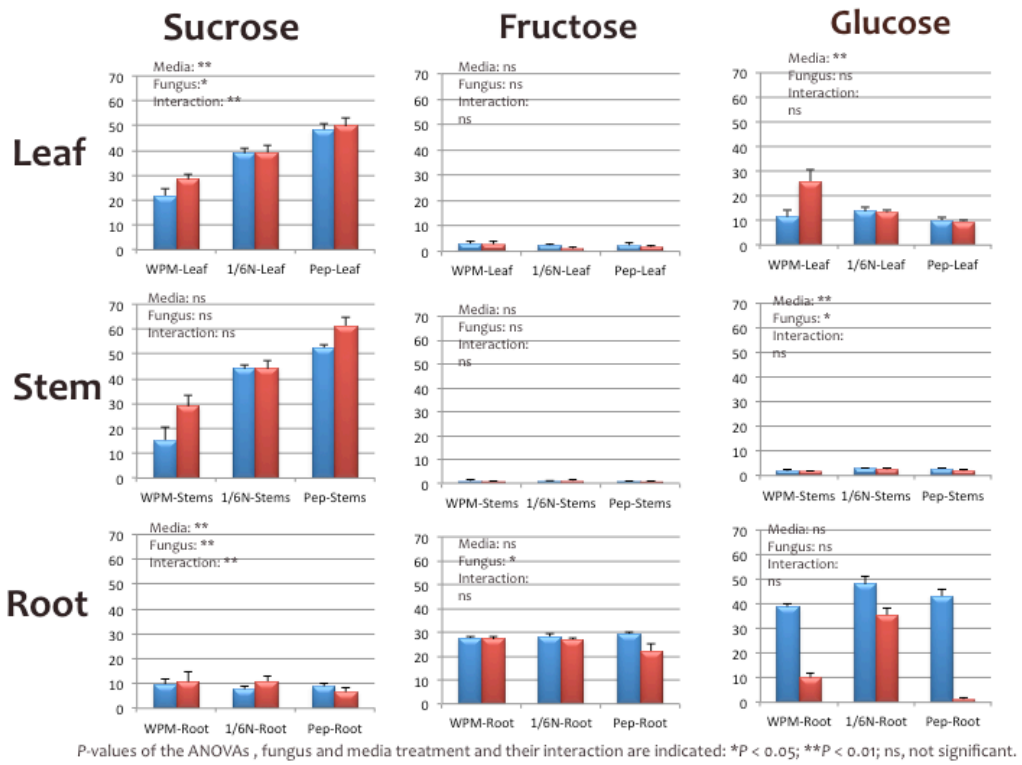



**Figure 16:** Sucrose, Fructose and Glucose data from the leaves, stems and roots of N-limited and peptone treated *P. tremuloides*. Axes were placed on the same scale to allow for comparison between compounds.

As an example of sample numbers, the nitrogen related experiments required a total of 288 data points for each of the above metabolites and elements [ 3 media * 2 interactions * 3 tissues * 4 biological replicates * 4 technical replicates ]. Due to the high throughput nature of the analysis, ~ 600 mg of each tissue from each media and interaction were required [ 20mg * 4 biological replicates * 4 technical replicate + 180 mg for phosphorus ].

The primary finding for the N limitation experiments was that trees under N stress undergo a radical shift in the way that they produce starch. If N is made available in a usable form to the plants, then sugars are made in a mobile form (i.e. sucrose), which is transported to the roots allowing the glucose component to supply carbon to *Laccaria* (**Figure 16**). This in turn causes a large increase in both the above- and below-ground biomass of the trees, presumably by enhancing N uptake to the trees. In N limited conditions, most of the sugar is stored in the non-mobile form of starch, which in turn does not elicit the same increase in biomass (**Figure 17**). The proteomics portion of the project aimed in part to resolve the molecular mechanisms behind this effect. However, protein extractions for the P limitation experiments are currently awaiting SWATH-MS analyses.

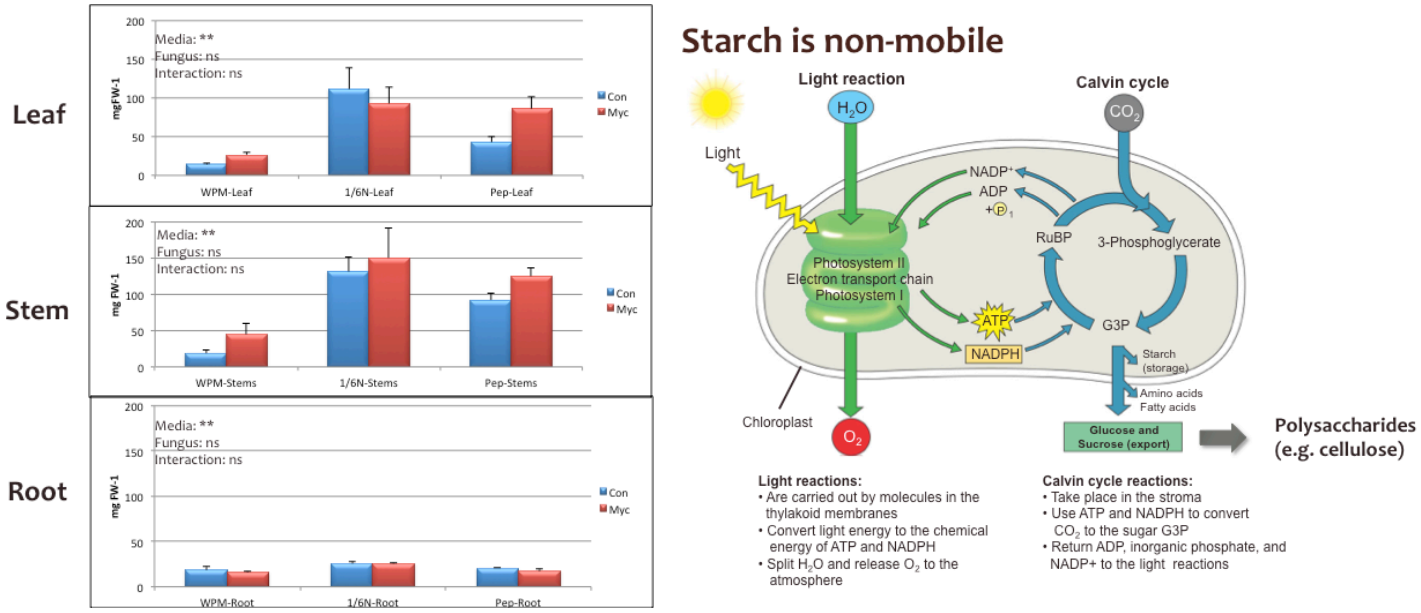



**Figure 17:** There is a fundamental metabolic shift toward starch production under both inorganic and organic N limitation conditions. (Left) Starch data from the leaves, stems and roots of N-limited and peptone treated *P. tremuloides*. (Right) Diagram of the photosynthetic processes involved in starch and sucrose biosynthesis. Starch remains in the plastids and does not travel through the plant's vascular system like sucrose does.

While the results for the N limitation experiments were completed early on in the project, the phosphorus limitation experiments have more recently been completed. After harvesting the leaf, stem and root tissue from 3 month old *P. tremuloides* plants subjected to limiting phosphorus (0.416 mM), with or without association with *Laccaria bicolor,* along with respective non-treated control (1.25 mM Phosphorus) plants, biochemical analyses was conducted as described above. Glucose, fructose and sucrose levels increased in mycorrhizal conditions in leaf and stem tissues, whereas in roots mycorrhizal association decreased the levels of those metabolites (**Figure 18**). This implies that simple sugars are transferred from the plant to its

mycorrhizal partner to establish and maintain a mutual relationship. Starch on the other hand showed similar patterns in both optimal and limiting phosphorus conditions but the amount of starch was more in limiting phosphorus conditions when interacting with *L. bicolor*. Nitrogen containing metabolites (amino acids, nitrates and proteins) showed significantly decreased levels in mycorrhizal conditions in all of the three tissues except for an increase in protein amounts in stem tissues in both of the tested conditions. As was expected, inorganic phosphorus showed reduction in the amount in stem and root tissues whereas in leaf there was slight increase under limiting phosphorus condition. Fungal colonization further dropped the levels of inorganic phosphorus in all tissues in both conditions. Patterns of elemental C, H and N levels remained the same in optimal and limiting phosphorus conditions irrespective of fungal colonization. However, a minor increase in C:N ratio in leaf and stem when in association with fungus was noticed in limiting phosphorus condition.

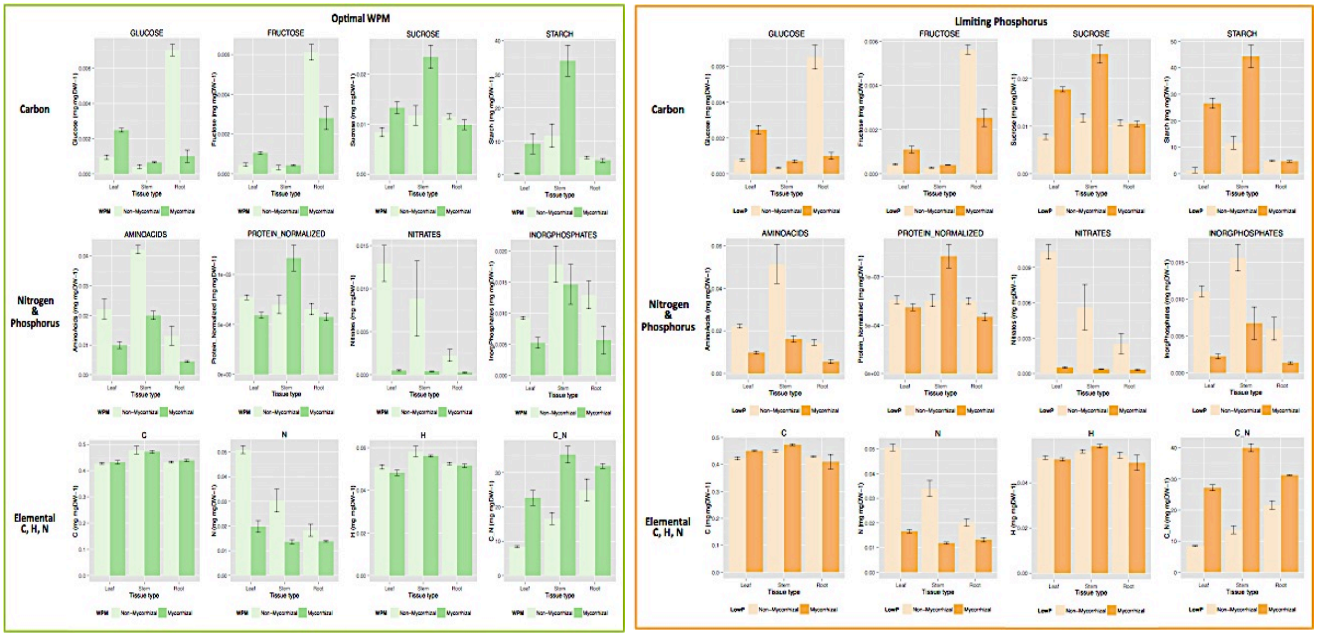
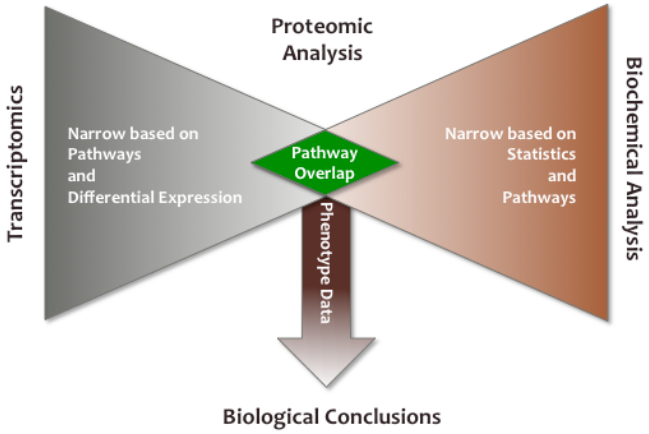



**Figure 18.** Biochemical analyses of *P. tremuloides* leaf, stem and root tissues grown under optimal and limiting phosphorus conditions with (mycorrhiza) or without mycorrhizal colonization (non-mycorrhiza).

In general, nitrogen deficiency is the most common and widespread nutritional deficiency affecting plants worldwide. In both N and P limitation experiments, identification of metabolic pathways in *P. tremuloides* tree roots revealed that carbohydrate and nitrate metabolism was impacted by changing environmental conditions, including interactions with the fungi. These results are being correlated with transcriptomic data as part of Approach #4 to identify metabolic pathways that are involved in articulation of phenotypic and biochemical changes in corresponding biological conditions.

***Approach #4: Computational Biology to Establish Predictive, Interactive System-Scale Networks Associated with Mycorrhizal Nutrient Cycling.***

As described above, computational biology has been used throughout the first three approaches to generate the desired datasets and discoveries specific to each approach. The ultimate goal, however, is the implementation of a multiple "omics" approach that attempts to integrate NGS transcriptomics, proteomics, ChIP-Seq and biochemical analyses to construct regulatory networks and computational modeling approaches to predict how ectomycorrhizal fungi influence rhizosphere symbiotic interactions that assist in tree biomass production. The aim of such models was to identify the specific molecular mechanisms at the mycorrhizal interface where nutrient availability is affected by soil biochemistry, thus generating an important enabling technology for the understanding of a complex system that can provide insight to real-world environmental impacts.




The genomic information used in this project was primarily derived from genome sequences developed by the JGI for both *Populus trichocarpa* and *Laccaria bicolor*. Sequences derived from RNA-Seq or ChIP-Seq were aligned to these genomes for identification and annotation. However, to address the need for species-specific DNA sequences used to identify *P. tremuloides* ChIP-Seq and protein sequences (described above), we also used a *P. tremuloides* resequencing dataset (donated by Jerry Tuskan at ORNL), abundant amounts of *P. tremuloides* transcriptomics data (from our work with ANL) and available EST data (from NCBI) to develop newly annotated datasets using the National Energy Research Scientific Computing Center (NERSC) resources. To assist in the development of such predictive models within the different plant tissues (leaves, stem and roots), some Illumina RNA sequencing was performed on these tissues to set background expression levels of all transcripts in each tissue. This was conducted through the sequencing facilities at the HudsonAlpha Institute for Biotechnology (Huntsville, AL). This approach significantly increase our ability to accurately identify *P. tremuloides* protein sequences (Approach #1), ChIP-Seq enriched target gene motifs (Approach #2), and biochemical pathways associated with changes in specific *P. tremuloides* metabolites (Approach #3). The resulting datasets have been placed into the larger conceptual context of how trees either gain or lose carbon and nutrients either below or above ground (**Figure 19**), and the data was then incorporated into developing the predictive regulatory networks that will help to model plant activity from genes to phenotypes.

This has resulted in publications on the prediction of the ectomycorrhizal metabolome from transcriptomic data as well as multi-omics approaches to identify the molecular mechanisms of plant-fungus mycorrhizal interaction (**see below Papers and Other Products Delivered**). Thus, in many respects this project was successful at meeting the above goal and aim. However, in reality it is much more difficult and time consuming to integrate each of the diverse datasets into one comprehensive model in a way that does not loose clarity and usefulness as apparently uncorrelated data needs to be omitted from the growing regulatory network. In other words, each dataset that is introduced to the network analysis inherently reduces the scale of the network itself. Thus, when we obtain such proteomics results indicating that there less than 50% of all expressed proteins that correlate in any way with RNA expression, the resulting

network experiences a significant blow to its depth and power of prediction. For more technical detail on how each stage of computational network analysis was conducted, Dr. Collart has submitted a research progress report on behalf of the ANL research team. In summary, the network analysis turned out to be much more complex than originally planned, and the analysis of interconnected data is ongoing at ANL. It will likely improve significantly once the full *P. tremuloides* genome database is completed in the next few years.

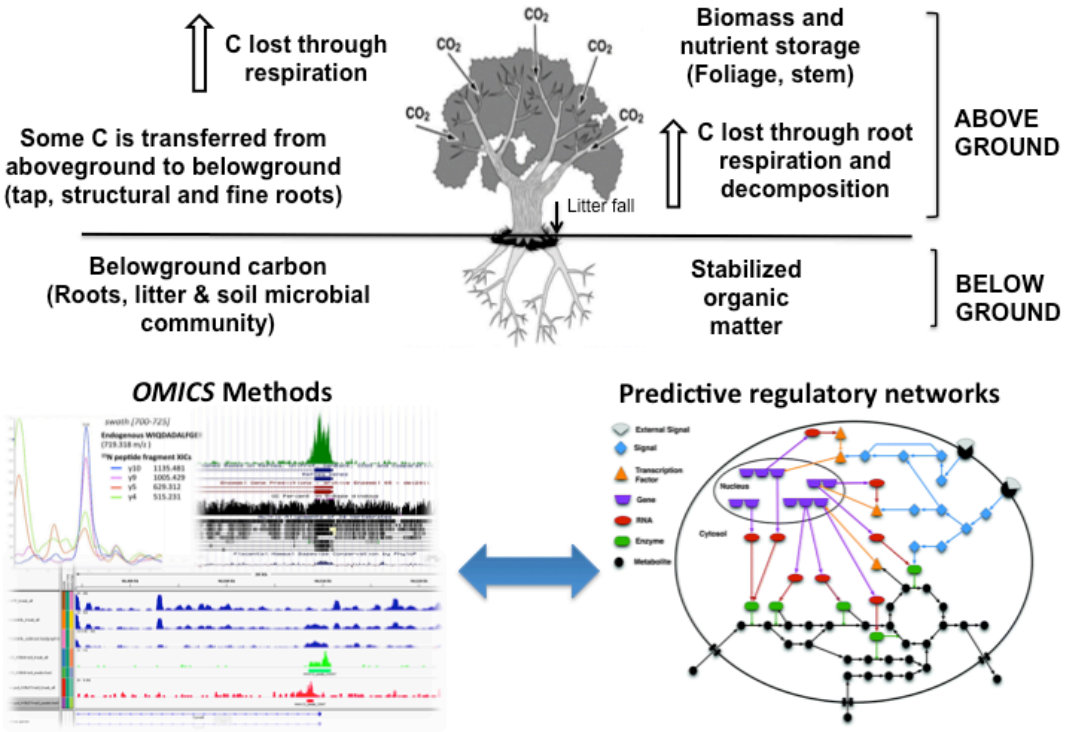



**Figure 19.** The larger conceptual model for how plants gain and lose carbon and nutrients. The aim was to use our omics-level methods to generate predictive regulatory networks that would in turn improve our understanding of the collected data.

That being said, network analysis of a reduced number of datasets is very successful at narrowing in on the molecular events that control a plant's response to nutrients and microbes. For example, we tested the hypothesis that metabolic changes that allow ectomycorrhizal fungi to improve plant health can be identified in transcriptomics data by linking gene expression to biochemical pathway activity in the leaves, stems and roots. Here, the use of biochemical pathway data (derived from the KEGG Pathway Database) to interrogate transcriptomic data is an effective approach to resolving metabolic activities that are significantly impacted by changing environmental / nutrient conditions, including interactions with soil organisms. Using the identification of metabolic pathways in the root tissues to answer the question: How do ectomycorrhizal fungi (ECM) improve plant health under limited nitrogen, it was revealed that pathways associated with nitrate metabolism, glucose production, fructose production and cellulose biosynthesis were up-regulated in aspen roots under organic nitrogen conditions. The narrowed in on 199 metabolic pathways containing 2775 genes compared to the 504 pathways and 19,487 genes that are simply differentially expressed under the same conditions (**Figure 20**). In addition, metabolic changes affected plant health under different nitrogen conditions, but only the metabolic pathways relating to carbohydrate production and nitrate metabolism are down-regulated in the root tissues during limiting nitrogen

conditions. Overall, carbohydrate and nitrate metabolism were impacted by the changing environmental conditions, including fungus colonization.

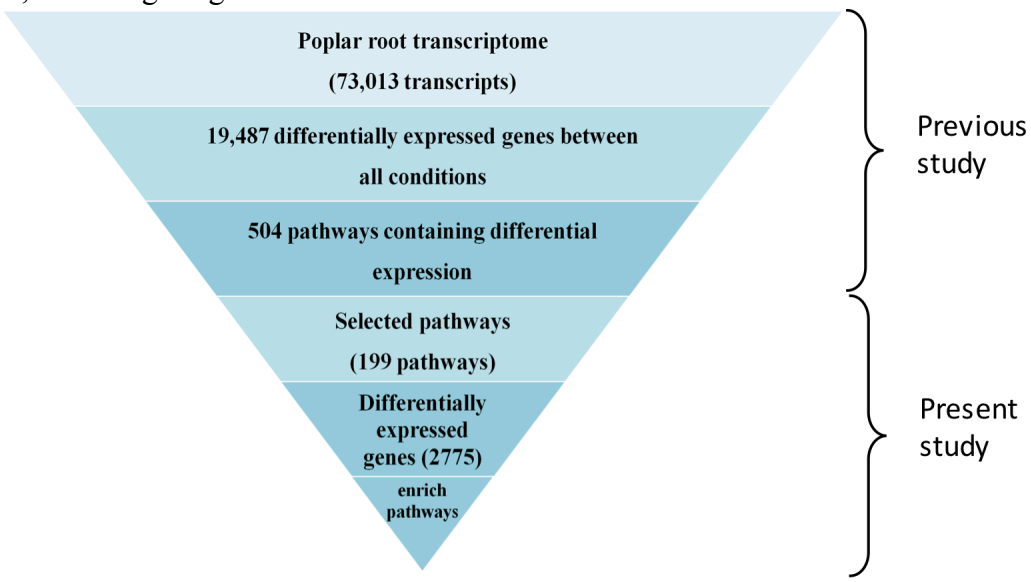



**Figure 20.** Flow chart showing transcriptome analysis when interrogated with the pathway data obtained from biochemical analysis of metabolite abundance in the roots of *P. tremuloides* under limited N conditions.

***Evolving Changes to the Experimental Design:***

As described above, while the basic milestones of the project were completed as compared to the original proposed timetable, a number of technical issues were identified and were eventually overcome. For example, upon setting up several rounds of biological replicates for the *Laccaria* x poplar laboratory system, several changes were made to the proposed experimental design based on *Laccaria* developmental results and growing transcriptomic data collections from other projects. The original experimental plan for our *Laccaria* x poplar system was partitioned into three stages, each dependent on a stage of development of the symbiotic relationship between the tree and the fungus. These developmental stages were ***(1)*** Free living organisms, each on media independent of the other organism. ***(2)*** A seven-day interaction experiment where the fungus and the tree may sense one another but not be in direct contact, and ***(3)*** a 5-6 week experiment where the trees and fungus develop mycorrhizae. In addition, each organism in each of these stages would be grown on Woody Plant Medium (rich in inorganic nitrogen and phosphorus) with ***(a)*** Control amounts of nutrients, ***(b)*** Limiting amounts of inorganic nitrogen sources, ***(c)*** Limiting amounts of phosphorus source, ***(d)*** Limiting amounts of carbon/sucrose, and ***(e)*** Complex/organic nitrogen source (alanine) replacing inorganic nitrogen sources. This would have constituted a total of 20 experimental setups, including free-living *Laccaria*, free-living poplar, interacting *Laccaria* and poplar, and mycorrhizal *Laccaria* and poplar under 5 treatment conditions.

The use of alanine as a complex/organic source of nitrogen was quickly determined to not work for the growth of *Laccaria* in our system. We subsequently found that peptone (a source of various amino acids and small peptides) works quite well as an organic nitrogen source to support the laboratory system. Consequently, experiments using alanine were replaced with experiments using peptone to replace the inorganic nitrogen in the media.

Including an organic source of phosphorus was also proposed to better round out the experimental plan by mimicking the purpose of the organic source of nitrogen (peptone). Good past results have been obtained using Ca-phytate (an organic phosphorus source commonly found in grasslands); however, after much searching, sources in China may prove to be the only reliable source of pure Ca-phytate, and the costs were found to be prohibitive.

Another change to the experimental design resulted from the fact that both observations in the early developmental response of *Laccaria* and resulting transcriptomic-level gene expression changes indicate that the seven day interaction time points are much more variable than either free-living aspen or aspen having active mycorrhizal associations. Since our system relies on the ability to generate highly reproducible results from subsequent biological replicates, we altered the project to focus on free living and mycorrhizal samples from aspen.

Similarly, we did not obtain gene expression-level data from experiments that limit carbon in the form of sucrose due to the fact that the ANL team would like to use such experiments as controls for their proposed SFA research projects on bacterial transporters. Thus, while the proposed experiments in limited sucrose may have generated interesting information for differential gene expression for other projects, they did not really target the goals of this project because changes in sucrose (which is not typically present in soil and is not accessible to *Laccaria*) is not a nutrient limitation that one would typically find in the natural environment. Thus, to improve the overall applicability to natural biomass production, we obtained permission to remove the seven-day and carbon-limitation experiments from the work plan and replaces them with a more detailed analysis of the remaining conditions by assessing the results in the roots, stems and leaves, rather than the proposed single root tissue from each experiment. The more detailed analyses from both above- and below-ground tissues also allowed the available funds to provide a more comprehensive understanding of the processes involved in nutrient cycling that are more applicable to the DOE missions in bioenergy, carbon management, and environmental stewardship.

While technical delays were encountered throughout the project, our assembled team of experts has been able to overcome these challenges by choosing research directions that have actually improved upon the original research plan. Most of these new directions took advantage of cutting edge technologies that have developed over the course of the project, such as high throughput proteomics, protein expression analysis for ChIP-Seq and automated aspects of biochemical analyses.

That being said, two primary sources of unanticipated delays were encountered (1) when the participating ANL research group lab was moved to new facilities on the ANL campus and (2) when an equipment breakdown and subsequent Microsoft software update issues at UAB shut down the primary mass spectrometer that was being used for this project. Since our ChIP-Seq protocols depended on producing purified protein for each of the 16 transcription factors of interest, having the ANL lab shifted to new facilities was an unfortunate delay, although the ANL group did an outstanding job minimizing those delays. In any case, antibody production for each protein takes 4 to 5 months after purified protein is made, including verification of the purified antibodies through ELISA and Western blots. Despite these delays, the procedures for protein and antibody production, poplar-specific ChIP-library preparation, ChIP-sequencing at the University of Chicago, and data analysis was effectively streamlined.

As far as the proteomics aspects of the research go, all of the original problems with protein extractions, quantification, protein databases and the identification of specific proteins from poplar tissues were

overcome at the beginning of 2014. The biggest of these hurdles came from the fact that instead of nucleotide variation in poplar species being (on average) every 300 bases as reported in the literature, such variation is typically as high as one change in every 30 bases. This meant that (on average) every tryptic peptide, which is ~10 amino acids long (derived from ~30 bases), was altered. This in turn greatly reduced the number of peptide "hits" and allowable protein identifications from the available poplar databases. However, the Cseke lab was able to utilize past RNA-Seq data (>billion reads) to solve this problem with the help of the NERSC computer systems to generate new poplar-specific protein databases that will certainly become an important resource for the poplar scientific community (*once published*). More recently, the equipment failure of the AB Sciex 5600 TripleTOF mass spectrometer stopped the generation of proteomic data, and despite the fact that it was subsequently repaired, Microsoft also recently stopped supporting Windows XP. This unexpectedly required alteration of the operating system at UAB. Unfortunately, the new version of the software used to run the Sciex 5600 had several bugs that needed to be resolved to resume operation. This software is also used to collect and analyze the SWATH-MS data, but once these issues were solved, we had no more difficulties.

In addition to the two primary sources of unanticipated delays, there were also two primary sources of technical difficulty: (1) the limited use of the *P. trichocarpa* protein database for the analysis of proteins in other poplar species and (2) the fact that ChIP-Seq is not commonly performed in poplar tree tissues and is inherently technically challenging. Such issues were quickly overcome using the NERSC system to assemble a new protein database for *P. tremuloides* using an abundance of past transcriptomics data. This greatly improved the identification of peptide sequences derived from MS data. Similarly, every step of the ChIP-Seq process is technically challenging and needs to be optimized for each new species. Much of our work during this project was directed at optimizing procedures specifically for poplar species. Optimization appears to be complete, and this will certainly lead to new resources that other researchers can use to address transcription factor targets in other poplar species. Thus, this project represents the development of cutting edge technology for such species.

For proteomics, we were initially going to characterize 600 proteins based on differential expression patterns in the transcriptome. However, we later decided that this would not necessarily do much towards the development of predictive tools that correlate biomass production to omics data other than verifying or disproving that RNA and protein have similar expression patterns. To correct for this, we opted to analyze as many proteins as possible within the phenotypic / biochemical pathways that change in response to *Laccaria* interaction or environmental nutrient levels. This will allowed us to correlate phenotype with both proteomic and transcriptomic data, providing a much more resilient data stream with which we can develop predictive tools. We are in the process of including all proteomics data on a publicly available website at UAB. To date, the data from this project has been linked to the *P. trichocarpa* protein annotations as part of the TMPL MRMSpace tools website. http://tmpl.uab.edu/MRMPath/mrmpath.jsp#part3

## PAPERS AND OTHER PRODUCTS DELIVERED:

A variety of publications, abstracts, presentations and posters have been generated over the course of this reporting period. In addition, a listing of other products associated with the project have been given below, including new technologies that have been optimized for poplar tree species and the personnel that have been trained / educated using the resources supplied by this project. Members of the UAH team are in bold text.

### *Refereed Research and Technical Publications:*

Larsen, P.E., **Sreedasyam, A., Trivedi, G.,** Desai, S., Dai, Y., **Cseke, L.J.**, Collart, F.R. (2016). Multi-Omics Approach Identifies Molecular Mechanisms of Plant-Fungus Mycorrhizal Interaction. *Front. Plant Sci., 6:1061*.

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., Costa de Oliverira, A., **Cseke, L.J.,** Dempewolf, H., et al. (2015). Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnology J., doi: 10.1111/pbi.12467*.

Kole, C., Muthamilarasan, M., Henry, R., Edwards, D., Sharma, R., Abberton, M., Batley, J., Bentley, A., Blakeney, M., Bryant, J., Cai, H., Cakir, M., **Cseke, L.J.,** Cockram, J., et al. (2015). Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects. *Front. Plant Sci., 6:563*.

Larsen, P.E., **Cseke, L.J.**, Miller, M.R., Collart, F.R. (2014). Modeling forest ecosystem responses to elevated carbon dioxide and ozone using artificial neural networks. *Theoretical Biology* 359: 61–71.

Larsen, P.E., **Cseke, L.J.**, Collart, F.R. (2013). Prediction of an Ectomycorrhizal Metabolome from Transcriptomic Data. In *Molecular Microbial Ecology of The Rhizosphere.* Frans J. de Bruijn (Ed.); Wiley-Blackwell Publishers.

**Cseke, L.J.**, Wullschleger, S.D., **Sreedasyam, A., Trivedi, G.,** Larsen, P., Collart, F.R. (2013). Chapter 12. Carbon Sequestration. *Genomics and Breeding for Climate-Resilient Crops.* Kole C. (ed): Springer, Germany.

Larsen, P.E., **Sreedasyam, A., Trivedi, G.,** Podila, G.K., **Cseke, L.J.,** Collart, F.R. (2011). Using next generation transcriptome sequencing to predict an ectomycorrhizal metabolome. *BMC Systems Biology* 2011, 5:70. [HIGHLY ACCESSED]

### *Research Papers Submitted for Publication and Associated with this Project:*

**Agee J., Sreedasyam A., Trivedi G.,** Podila G.K., **Cseke L.J.** (2015) Novel *Laccaria bicolor* Argonaute Homologues Mediate RNAi. *Submitted to Plant Cell and Host.*

### *Manuscripts in Preparation:*

***The following planned publications have actual manuscripts that are in the process of editing for submission.***

**Sreedasyam A., Trivedi G.,** Wilson L.S., Kim H., Barnes S., **Cseke L.J.** Overcoming the challenges in proteome analysis of *Populus tremuloides* using high-throughput RNA sequencing. *To be submitted to the Journal of Proteomics.*

Victor T., Delpratt N., Miller L.M., **Cseke L.J.** Synchrotron imaging of poplar tree interactions with micorrhizal fungi. *To be submit to The Plant Journal.*

**Tran B., Sreedasyam A., Trivedi G.,** Lasota S., Rogers A, **Cseke L.J.** The effect of Nitrogen limitation and ectomycorrhizal fungi interactions on the phenotype and metabolic pathways of *Populus tremuloides*. *To be submitted to New Phytologist.*

**Trivedi G., Sreedasyam A.,** Larsen P.E., Collart F.R. and **Cseke L.J.** Transcriptome analysis of the ectomycorrhizal fungus *Laccaria bicolor* to understand its metabolic flexibility. *(to be submitted to New Phytologist).*
This manuscript is complete and in the final stages of processing before submission.

**Sreedasyam A., Trivedi G.,** Larsen P.E., Collart F.R., **Cseke L.J.** Transcriptional reprogramming in *Populus tremuloides* in response to ectomycorrhizal interaction with two isolates of *Laccaria bicolor*, S238N and H82. *(to be submitted to Mycorrhiza).*
This manuscript is essentially complete and in the final stages of processing before submission. .

**Trivedi G., Sreedasyam A.,** Larsen P.E., Collart F.R. and **Cseke L.J.** Mycorrhizal quantification and regulation of ECM symbiosis under nitrogen limiting conditions. *(Intended to be submitted to New Phytologist).*

Miller L.M., Victor T., Delpratt N., **Kirchner G., Cseke L.J.** The use of infrared synchrotron imaging to characterize environmental nutrients surrounding the roots of aspen trees (*Populus tremuloides*). *(to be submitted to the Journal of Analytical Chemistry).*
This methods paper is in the final stages of writing.

### *Published Abstracts from Oral Presentations at Research Conferences:*

*January, 2016*: "The Vital Importance of Soil Health: from Laboratory Systems to Ecosystems" **Cseke, L.J**. To be presented at the Exploring Phytobiomes Workshop at the 2016 Plant and Animal Genomics Meeting, San Diego, CA.

*January, 2015*: "CO2 Sequestration to Mitigate Effects of Climate Change" **Cseke, L.J**. Climate Change and ICRCGC 2 Workshop at the 2015 Plant and Animal Genomics Meeting, San Diego, CA.

*May, 2014* "The control of plant biomass development mediated by soil microbe nutrient cycling" **Cseke L.J.** Invited speaker presentation for the 2014 NSLS/NSLS-II & CFN Joint Users' Meeting, Upton, NY

*January, 2014*: "Nutrient cycling in the rhizosphere, an important consideration for carbon sequestration in trees" **Cseke L.J**. Climate Change and ICRCGC 2Workshop at the 2014 Plant and Animal Genomics Meeting, San Diego, CA.

*September, 2013:* "Development of *Populus tremuloides* protein database from assembled transcriptome for shotgun and targeted proteomics." **Sreedasyam A., Trivedi G.,** Wilson L.S., Barnes S., **Cseke L.J**. University of Alabama at Birmingham, Birmingham, AL.

*September, 2013:* "A Comprehensive Search for Laterally Transferred Genetic Material in All Sequenced Eukaryote Genomes" **Gray D. and Cseke L.J.** UAH-MTSU Biological Science Joint Symposium. Winchester, TN

*August, 2013:* "De novo transcriptome assembly and characterization of *Populus tremuloides* ectomycorrhizal transcriptome under phosphate starvation." **Sreedasyam A., Trivedi G.,** Larsen P.E., Collart F.R., **Cseke L.J.** Argonne National Laboratory, Lemont, IL.

*August, 2013:* "Development of ChIP-Seq protocols to investigate target genes of MADS-BOX transcription factors in *Populus tremuloides*." **Sreedasyam A., Trivedi G.,** Zerbs S., Larsen P.E., Collart F.R., **Cseke L.J.** Argonne National Lab, Lemont, IL.

*February, 2012:* "Merging Bottom-Up with Top-Down Research in Symbiosis and Nutrient Cycling" **Leland J. Cseke**. Presentation 2012 Department of Energy Genomic Science Awardee Meeting, Bethesda, MD

### *Published Abstracts from Poster Presentations at Research Conferences:*

*July, 2015* "New insights into mycorrhizal interactions enabled by ChIP-Seq and SWATH proteomics." **Trivedi, G., Sreedasyam, A., Tran, B.,** Wilson, L.S., Kim, H., Barnes, S., Zerbs, S., Larsen, P.E., Collart, F.R., **Cseke, L.J.** Poster presentation at the Annual Meeting for The American Society of Plant Biologist, Minneapolis, MN.

*May, 2015* "Quantitative SWATH Proteomics Analysis of Tree-Fungal Interactions Under Nutrient Limiting Conditions" Wilson, L.S., **Trivedi, G., Sreedasyam, A.,** Kim, H., Cui, X., **Cseke, L.J.** and Barnes, S. Poster presentation for the 63rd Conference on Mass Spectrometry and Allied Topics for the American Society for Mass Spectrometry, Baltimore, MD.

*February, 2015* "Effect of Nitrogen and Phosphorus Limitation on the Growth and Morphology of Engineered Microbial-Plant Communities" **Cseke, L.J.,** Desai, S., Zerbs, S., Larsen, P.E., **Trivedi, G., Sreedasyam, A.,** Korajczyk, P., Collart, F.R. Poster presentation for the Genomic Science Contractors–Grantees Meeting XIII, Washington, DC.

*February, 2015* "Nutrient Cycling for Biomass: A multi-omics approach to model ectomycorrhizal regulatory networks" **Trivedi, G., Sreedasyam, A., Tran, B.,** Lasota, S., Rogers, A., Wilson, L.S., Kim, H., Barnes, S., Zerbs, S., Larsen, P.E., Callart, F.R., **Cseke, L.J.** Poster presentation for the Genomic Science Contractors–Grantees Meeting XIII, Washington, DC.

*June, 2014* "Overcoming the challenges in proteome analysis of Populus *tremuloides* using highthroughput RNA sequencing" Wilson, L.S., **Sreedasyam, A., Trivedi, G., Cseke, L.J.**, Kim, H. and Barnes, S. Poster presentation for the 62nd Conference on Mass Spectrometry and Allied Topics for the American Society for Mass Spectrometry, Baltimore, MD.

*May, 2014*: "Optimizing Bioenergy Production by Imaging Nutrient Exchange in the Plant Root Rhizosphere using Infrared Microspectroscopy" Victor T.W., Delpratt N.A., **Cseke L.J.,** Miller L.M. Poster presentation for the 2014 NSLS/NSLS-II & CFN Joint Users' Meeting, Upton, NY

*February, 2014:* "Biochemical and Proteomic Profiling of a Nutrient Limited Poplar x *Laccaria* Ectomycorrhizal System" **Cseke L.J., Sreedasyam A., Trivedi G., Tran B.,** Lasota S., Rogers A., Wilson L.S., Kim H., and Barnes S. Poster for the 2014 Department of Energy Genomic Science Awardee Meeting, Arlington, VA

*February, 2014:* "ChIP Sequencing of Transcription Factors to Infer Gene Regulatory Networks for Carbon and Nitrogen Sequestration During Poplar x *Laccaria* Ectomycorrhizal Interactions" **Sreedasyam A., Trivedi G.,** Zerbs S., Larsen P.E., Collart F.R., **Cseke L.J.** Poster for the 2014 Department of Energy Genomic Science Awardee Meeting, Arlington, VA

*January, 2014:* "Nutrient cycling in the rhizosphere, merging omics-level data to assess carbon sequestration in trees" **Cseke L.J., Sreedasyam A., Trivedi G., Tran B.,** Lasota S., Rogers A., Wilson L., Barnes S., Zerbs S., Larsen P.E. and Collart F.R. Poster for the 2014 Plant and Animal Genomic Meeting (PAG), San Diego, CA

*September, 2013:* "A Comprehensive Search for Laterally Transferred Genetic Material in All Sequenced Eukaryote Genomes" **Gray D., Cseke L.J.** UAH-MTSU Biological Science Joint Symposium. Winchester, TN

*September, 2013:* "The impact of nitrogen limitation and mycorrhizal symbiosis on aspen trees" **Karmacharya C.,** Cebert E., **Cseke L.J.** UAH-MTSU Biological Science Joint Symposium. Winchester, TN

*February, 2013:* "Advances in Proteomic and Biochemical Analysis During Poplar / *Laccaria* Mycorrhizal Interactions" **L.J. Cseke, G. Trivedi, A. Sreedasyam, B. Tran,** H. Kim, G. Robinson, L. Wilson, S. Barnes, S. Lasota, A. Rogers, S. Zerb, P.E. Larsen, and F.R. Collart. Poster for the 2013 Department of Energy Genomic Science Awardee Meeting, Bethesda, MD

*February, 2013:* "Molecular Mechanisms that Mediate Environment Sensing and Response" M. Bourssa, **L.J. Cseke**, S. Desai, A. Dombrowski, P. Laible, V. Landorf, P.E. Larsen, L. Miller, **A. Sreedasyam, G. Trivedi ,** S. Zerbs, and F.R. Collart. Poster for the 2013 Department of Energy Genomic Science Awardee Meeting, Bethesda, MD

*October, 2012:* "Interactive proteomic/transcriptomic networks to understand environmental carbon management processes within poplar-mycorrhizal interactions" **L.J. Cseke, D. Gray, A. Sreedasyam, G. Trivedi,** H. Kim, G. Robinson, L. Wilson, S. Barnes, P.E. Larsen, F.R. Collart. Presentation for the 10th Annual Ecological Genomics Symposium, Kansas City, MO.

*February, 2012:* "Nutrient Cycling for Biomass: Interactive proteomic / transcriptomic networks for global carbon management processes within poplar-mycorrhizal interactions" **Leland J. Cseke,** Frank M. Collart, Peter E. Larsen, Stephen Barnes and Helen Kim. Poster for the 2012 Department of Energy Genomic Science Awardee Meeting, Bethesda, MD

*February, 2012:* "System-Scale Modeling of Mycorrhizal Symbiosis" Peter Larsen, **Geetika Trivedi, Avinash Sreedasyam, Leland J. Cseke,** and Frank Collart. Poster for the 2012 Department of Energy Genomic Science Awardee Meeting, Bethesda, MD

### *Presentations by Invitation with Unpublished Abstracts:*

*October, 2013:* "New 'Omics' Level Plant Systems for International Space Station" **Cseke L.J.** Presentation for the Center of Advancement of Science in Space (CASIS), HudsonAlpha, Huntsville, AL

*October, 2013:* "Healthy Soil and Carbon Sequestration" **Cseke L.J.** Presentation for the Huntsville Botanical Gardens, Huntsville, AL.

*April, 2013:* "Nutrient Cycling for Biomass: Interactive proteomic / transcriptomic networks for global carbon management processes within poplar-mycorrhizal interactions" **Cseke L.J.,** Collart F.M., Larsen P.E., Barnes S. and Kim H. Web Conference for the Department of Energy Program Director.

*November, 2012:* "Understanding and exploiting the molecular interface between plant-microbial biological systems and their environment" **A. Sreedasyam, G. Trivedi,** P.E. Larsen, F.R. Collart, L. Wilson, G. Robinson, H. Kim, S. Barnes, and **L.J. Cseke**. Presentation for Brookhaven National Laboratories, Upton, NY.

*October, 2012:* "Multi-omics approach to study carbon and nutrient cycling within poplar-mycorrhizal interactions" **G. Trivedi, A. Sreedasyam,** P.E. Larsen, F.R. Collart, S. Barnes, H. Kim and **L.J. Cseke**. Presentation for Brookhaven National Laboratories, Upton, NY.

*November, 2011:* "Merging Bottom-Up with Top-Down Research in Plant Molecular Biology" by **Leland J. Cseke**. Research presentation for The HudsonAlpha Institute for Biotechnology Seminar Series, Huntsville, AL.

***Related Publications:***

The following two research publications are not a result of this project's research results. However, they do represent work done in Dr. Cseke's lab and provide key transcriptomics data that was used to help identify the proteins to be analyzed via targeted proteomics by the UAB team and especially the transcription factor proteins to be synthesized by the ANL facilities. These papers also provide initial examples of the types of techniques that will be used to construct computational metabolic and regulatory networks as part of Approach #4.

1. Larsen, P.E., **Sreedasyam, A., Trivedi, G.,** Podila, G.K., **Cseke, L.J**., Collart, F.R. (2011) Using next generation transcriptome sequencing to predict an ectomycorrhizal metabolome. *BMC Systems Biology* 2011, 5:70

   In this publication, we integrated the expression patterns of genes that code for metabolic enzymes and sets of expressed membrane transporters to generate predictive models of the ectomycorrhizal metabolome. The analysis illustrates an approach to generate testable biological hypotheses to investigate the complex molecular interactions that drive ectomycorrhizal symbiosis. These models are consistent with experimental environmental data and provide insight into the molecular exchange processes for organisms in our *Laccaria* x poplar laboratory system.

2. Larsen, P.E., **Sreedasyam, A., Trivedi, G.,** Magle, C., **Cseke, L.J.**, Collart, F.R. (2012) Predicting mechanisms of early *Laccaria*-Aspen interaction signaling from transcriptomic data. (*Submitted to New Phytologist*).

Through computational approaches, deep RNA-Seq data was used to ***(1)*** construct a network of gene regulation, ***(2)*** propose protein networks for signal detection and transcription factor modification, and ***(3)*** identify differentially spliced transcripts. An interactive web-based figure of results can be view as an example at:
http://www.bio.anl.gov/molecular_and_systems_biology/PreMycInteractome/PreMycInteractome.htm

## OTHER PRODUCTS ASSOCIATED WITH THIS PROJECT:

***Post-docs:*** Two post-docs have been funded and trained off of this project at various times (Dr. Avinash Sreedasyam and Dr. Geetika Trivedi). These post-docs are planning to go into careers associated with bioenergy production from plants.

***Graduate Students:*** Two graduate students have been funded and trained off of this project.

Bich Tran has worked on the analysis of biochemical data and its connection with transcriptomic data. She is graduating this summer, 2014.

David Gray has developed strong skills in bioinformatics and has developed a new project based on this work. Here, he seeks to analyze the possibility of Lateral Gene Transfer of both coding and non-coding sequences into the genome of *Populus* species through a comparison of all know genomes with the poplar genome. He will graduate in 2015.

***Undergraduate Students:*** This project has also exposed about 10 undergraduate students to the analysis of plant biomass, soil microbial interactions with plants, qPCR, ChIP-Seq, proteomics and the development of plant feedstocks applicable to DOE program directives.

***Novel New Protocols:*** This project has also established a number of new techniques that are of future importance to the study of poplar plant biomass. Some unexpected yet significant findings have come from these studies that will play an important role in directing future research for projects seeking to build a better understanding of plant biomass production through predictive tools.

***High throughput proteomics*** – the development of SWATH Acquisition technology is a tremendous leap forward in the study of proteins in bioenergy crops.

***ChIP*-Seq analysis** – Since we have optimized the methods usable in poplar trees, this opens the door to the understanding of the function of the many transcription factors that control plant biomass.

***qPCR to quantitate colonization of tree roots by soil fungi*** – to assess the amount of fungus interacting with tree roots under all of our conditions and treatments, we developed a qPCR based method that accurately determines the relative amount of interaction in the roots of trees by assessing the amount of single copy fungal genes relative to single copy tree gene. A major finding from this work is that limiting nitrogen also limits the amount of fungal interaction obtainable. This in turn prevents the fungus from setting up symbiotic interactions that subsequently benefits the tree.

***Synchrotron*-based FTIR imaging at Brookhaven National Laboratory** – During our trips to BNL to work with Dr. Alistair Rogers for the biochemical work, we also started a project with Dr. Lisa Miller at the NSLS-I and NSLS-II facilities (**Figure 21**). This has resulted in the development of techniques that allow us to visualize and quantitate the nutrients within the media that are used during plant / fungal interactions. The basic idea is that plant tissue, such as plant roots, can be visualized under a microscope (**Figure 22**) outfitted with an FTIR camera capable of capturing high-resolution data in the IR spectrum generated by the NSLS IR beamline. This allows for the identification and quantitation of specific compounds within the media (**Figure 23**).

**Figure 21.** The NSLS-II facility at BNL.

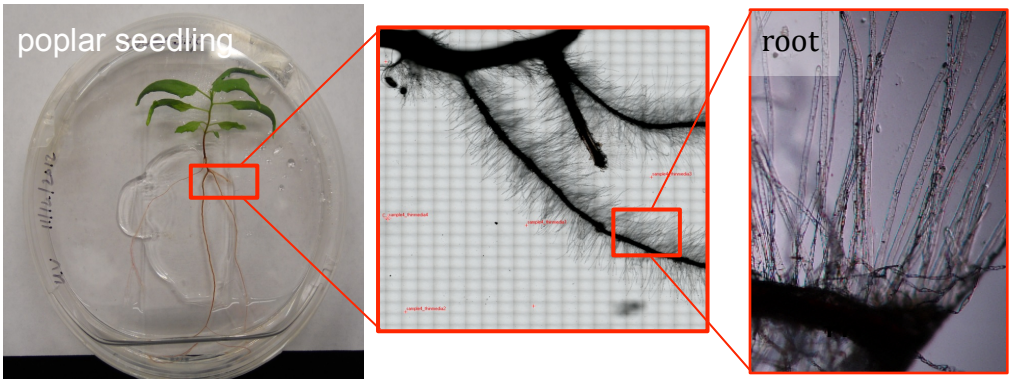



**Figure 22.** Microscopic images of *P. tremuloides* roots and roots hairs within a thin layer of phytagel matrix (WPM + sucrose and MES). When the slide is removed and the matrix dried, the resulting film is 15 µm thick.

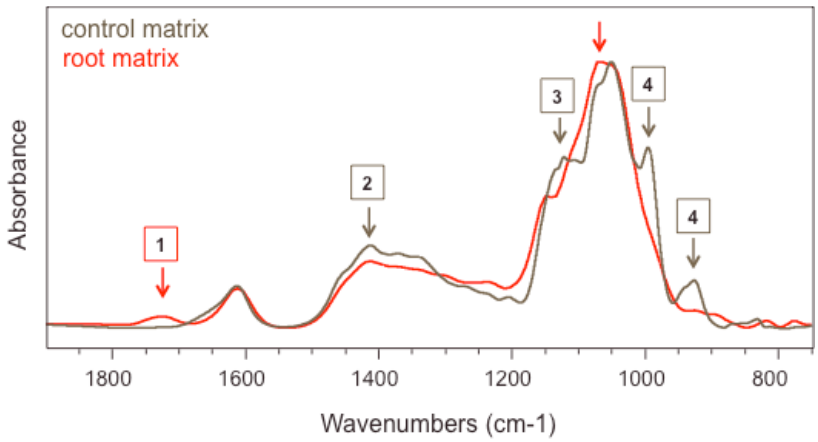



**Figure 23.** Example of the IR spectrum comparison of two samples with and without *P. tremuloides* roots. Differences are specific to carbonyl exudates (1) observed in root matrix at 1725 $cm^{-1}$, nitrate (2), sulfate (3), and sucrose (4), which in this case are all reduced in root matrix.

We have also found that the plant roots excrete different compounds into the media depending on the nutrient levels (**Figure 24**), which may account for why we see reduced tree / fungal interactions in nitrogen limited conditions. While as yet unidentified, such exudates may act as negative signals to interacting fungi.

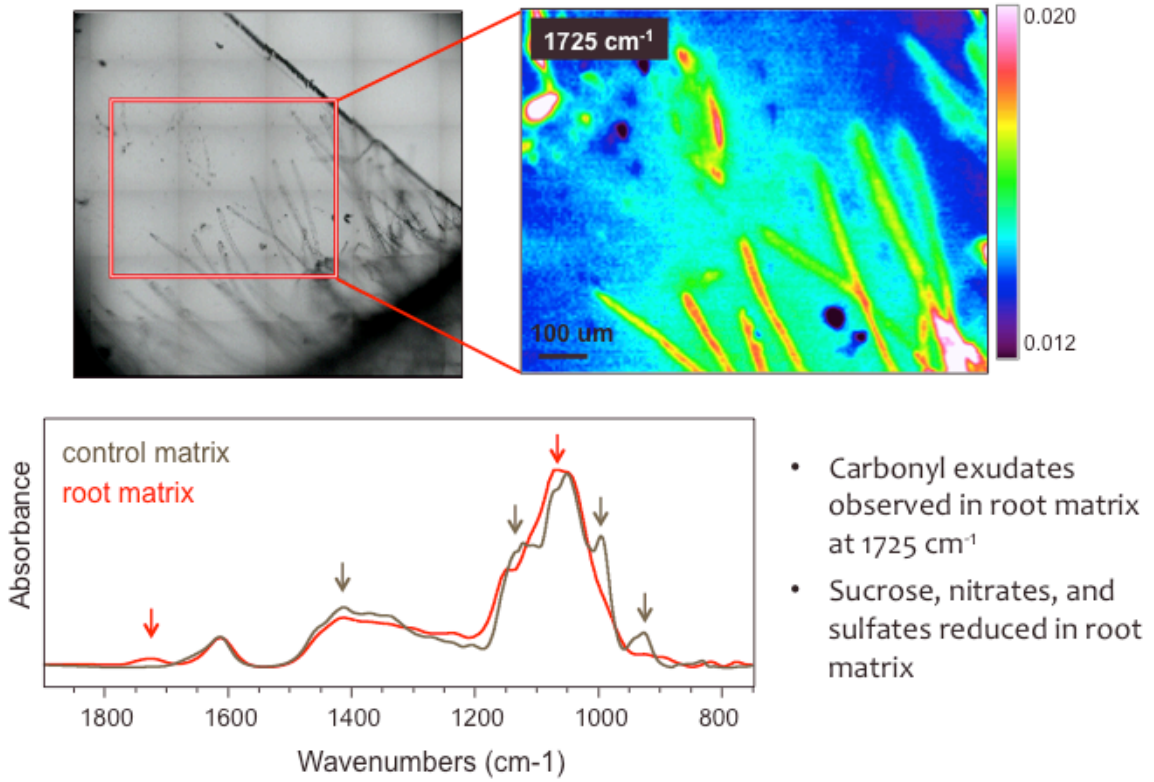



**Figure 24.** FTIR Imaging at NSLS. (Top) The microscopic and associated quantitative FTIR image of *P. tremuloides* roots tuned the to carbonyl exudates indicated by the RED arrows in the IR spectrum (Bottom).

## BASIC CONCLUSIONS:

Our laboratory-based Aspen / *Laccaria* interaction system is an effective model system with which researchers can perform controlled experiments to test the impact of changes in available environmental nutrients. While there are limitations to using LC-pseudoMRM-MS targeted proteomics, we have significantly improved the number of successful protein identifications within complex protein mixtures by generating new protein databases for *P. tremuloides* that are based on past transcriptomic data (a 48.7% increase). Additional improvements in protein identifications are observed when combining the new protein databases with gel-fractionation to reduce background signal within each fraction (a total of 1091% increase). Furthermore, 2D-DIGE discovery-based proteomics allows for the identification of differentially expressed proteins as well as post-translational modifications of specific proteins within complex mixtures of proteins as they are altered by mycorrhizal infection.

A set of transcription factors that are highly responsive to changes in available sources of nitrogen and phosphorus were identified, and antibodies were made at the ANL facilities with which ChIP-Seq experiments were conducted to identify target genes and new protein binding motifs that may correlate to observed expression patterns. In addition, biochemical analyses conducted at the BNL facilities have identified a shift in carbon allocation in nitrogen-starved plants from the synthesis of sucrose to that of starch. Likewise, in the presence of *Laccaria*, more glucose and amino acids are used for growth only when nitrogen is abundant. Indeed, optimum Nitrogen and Phosphorus availability is essential to generate trees that direct carbon metabolism toward the biosynthesis of soluble sugars within mycorrhizal roots. This in turn leads to healthier colonization (as determined visually and via qPCR approaches) and increased tree biomass. Therefore, when attempting to establish plant / fungal interactions in environments suffering from nutrient limitation, land managements procedures will likely need to be developed to provide optimal conditions for the establishment of such interactions prior to simply applying fungal inoculants into the environment.

In addition, changes in metabolism can be observed using RNA-Seq approaches; however, quantitative SWATH-proteomics revealed that less than 50% of the corresponding protein expression patterns correlate in any way with their RNA expression patterns within the same tissues collections. Therefore, care must be taken when making overarching conclusions about plant metabolism based on RNA-Seq data alone. That being the case, the further development of modern techniques like SWATH-MS and ChIP-Seq will aide our understanding of the true mechanisms of soil community signaling, establishment and maintenance.

***The Future?***

This ‘muli-omics’ study has generated layers of differing information that was integrated using computational approaches that led to the discovery of novel biochemical and regulatory pathways. While

integration of data that is fundamentally different in character presents significant hurdles for regulatory network model development, some of the identified pathways make use of specific proteins and post-translational modifications that play important roles in the molecular mechanisms of how plants, including trees, respond to changes in nutrient environment and mycorrhizal interactions. Furthermore, a computational model of these soil community regulatory networks and metabolic pathways continues to be generated to aide in the prediction plant phenotypes with changing soil nutrient environments.

Having the right combination of microbes associated with plants is largely responsible for the plant's ability to mine nutrients from the soil and to develop a strong "immune system". Our current chemically focused and intensive culture tends to forget that plants obtain nutrients in two ways: (1) via water soluble chemical nutrients and (2) via the activity of acquired microbial symbionts. In healthy natural ecosystems, chemical nutrients are always in low abundance because the organisms within that system have already locked such nutrients away within the biological system itself. Thus, in nature it is the biological sources of nutrients and the microbes that have the capacity to mine those nutrients for their plant hosts that actually control the terrestrial nutrient cycles on this planet. Thus, a new push in the future may very well be to use our skills at elucidating complex patterns to strategically guide soil microbe communities to do what we want, essentially allowing nature to do the work of figuring out what is most efficient and effective for human needs. However, the findings of this project and other work in our lab lead to the hypothesis that the specific soil community composition is less important than the emergent properties of those communities. So, additional research into what soil communities are effective and how they are established will be key in developing human understanding of how to manipulate biological systems to meet human needs without causing undue damage to our environment.