ENVIRONMENTAL SCIENCE AND TECHNOLOGY

Seed Money Fund

Metabolic Profiling: A Required Element in Functional Genomics

T. J. Tschaplinski,[1] G. J. Van Berkel,[2] and J. P. Miles[3]
[1]*Environmental Sciences Division*
[2]*Chemical Sciences Division*
[3]*Center College, Danville, Kentucky*

The emerging science of metabolomics or metabolite profiling offers tremendous potential to discover novel genes and assign function to those genes. Metabolic profiling determines the consequences of a targeted change in gene activity and has the potential to provide information on gene function and how it affects the complex biochemical network. The creation of a comprehensive gas chromatography–mass spectrometry (GC-MS) library of trimethylsilyl (TMS) derivatives of organic constituents is essential for the near-term development of metabolic profiling in poplar (*Populus*). A proof-of-principle study was initiated to demonstrate that broad-spectrum metabolite analyses can be achieved, with derivatization protocols effective at reproducibly silylating the 300–450 metabolites, with the number of metabolites analyzed and throughput rates increased by the concurrent development of direct-infusion electrospray mass spectrometry (DI-ES-MS). The research is organized around three tasks that include (1) optimization of GC-MS derivatization and analytical protocols for high throughput; (2) creation of a mass spectral library of known *Populus* metabolites, including the characterization of unknown compounds; and (3) extension of analyses to the liquid chromatography–mass spectrometry (LC-MS) platform to further increase the number of metabolites analyzed and improve characterization of unknowns. The project has resulted in substantial progress towards meeting the goals of all three tasks that places ORNL in a more competitive position to pursue outside funding for metabolomics research.

Introduction

Sequencing of microbial and plant genomes will require ascribing function to the vast amount of draft gene sequence data to establish precise relationships between gene structure and phenotypic variation. Changes in gross phenotypes (e.g., branching, flowering time, growth rate) are the integrated result of changes in the pattern of many interacting genes. In many cases, these complex networks can only be deconvoluted through detailed analysis of changes in metabolites, the precise endpoints of complex biochemical pathways. Assignment of gene function will require broad-spectrum biochemical analyses that be used in high-throughput phenotyping. Key elements in achieving high throughput that we are addressing include establishing a comprehensive metabolite database of mass spectra and minimizing time for sample preparation for GCMS, and maximizing the number to metabolites analyzed by developing DI-MS protocols.

Technical Approach

The research is organized around three tasks that include (1) optimization of gas chromatography–mass spectrometry (GC-MS) derivatization and analytical protocols for high throughput; (2) creation of a mass spectral library of known *Populus* metabolites, including the characterization of unknown compounds and establishment of deconvolution strategies to separate overlapping compounds; and (3) extension of analyses to the liquid chromatography–mass spectrometry (LC-MS) platform to further increase the number of metabolites analyzed and improve characterization of unknowns. Low volatility and thermal instability result in many metabolites not being amenable to analysis by GC/MS. In order to include such difficult-to-derivatize metabolites in a profiling strategy, including those metabolites with higher molecular weight or high polarity, LC-MS protocols were tested. These compounds include sugar phosphates and nucleotides involved in energy transduction. Development of such protocols will obviate the need for derivatization and expand the range of compounds that can be included in metabolite profiling. Phosphorylated metabolites are involved in primary C metabolism, the fixation of carbon dioxide to carbohydrates (e.g., Calvin-Benson-Basham cycle) or the metabolism of carbohydrates to provide energy (e.g., glycolysis and the pentose phosphate pathway). Furthermore, the relative concentrations of phosphorylated metabolites, and especially the nucleotides, are essential to evaluate the metabolic status of an organism. The importance of determining

phosphorylated metabolites by LC-ES-MS is confirmed by a recent publication from the metabolomics research group at the Max Planck Institute.[1]

Though chromatography of phosphorylated metabolites is problematic, the importance of adenylates (the nucleotides, ATP, AMP, etc.) in energy transduction has led several research groups to attempt the protocol development with varying degrees of success. The analysis of such adenylates with reverse-phase, high-performance liquid chromatography (HPLC) is problematic (without using ion-pairing reagents). The concentrations of buffer salts are often 0.1 molar and greater, which would be incompatible with electrospray mass spectrometry (ES-MS). An alternative approach to reverse-phase HPLC is hydrophilic interaction chromatography (HILIC). This approach relies upon partition of components between a somewhat polar mobile phase and a water phase immobilized onto a support. The nature of the stationary water layer is not well known. Elution order is similar to that found in normal-mode chromatography—nonpolar components elute first and more polar (hydrophilic) components are retained longer. Mobile-phase gradients begin with less polar solvent (commonly acetonitrile) and progress to increasing proportions of aqueous buffer. The approach has recently been used by Tolstikov and Fiehn (2002). They evaluated both TSK Gel Amide 80 column from TosoHass and PolyHYDROXYETHYL A column from PolyLC and found the former to be slightly better. Our approach was to test several reverse-phase approaches and focus the most attention on HILIC.

Results and Accomplishments

Task 1. Protocol refinement to minimize time of sample preparation and analysis was conducted by contrasting an aggressive derivatization protocol (stronger TMS donors, higher temperatures, and longer duration) with the lower-temperature methoximation followed by trimethylsilylation, used by the Max Planck researchers. The objective was to find the method that optimizes sample preparation time and spectral output (i.e., maximum metabolites observed). The stability of the low-molecular-weight organic acids was better with methoximation, but there was a tradeoff in less efficient derivatization of higher-molecular-weight phenolic glycosides. Tradeoffs between derivatization efficiency and number of metabolites that can be characterized will have to be made, depending on whether the analyses are targeted to a class of compounds or remain broad spectrum. Included in this task was a major proof of principle, which involved comparing the metabolite profiles of polar compounds of activation-tagged transgenic *Populus tremula* x *alba* with the non-transformed hybrid. The transgenic line (B18) displayed the branchy phenotype. The largest difference between the transgenic and the non-transformed control was a 4-fold increase in a metabolite that matched the predicted fragmentation pattern of ent-7α-hydroxykaurenoic acid, a key precursor of gibberellins (GAs), a class of growth regulators that regulate height growth in plants. The accumulation of a key GA precursor in a branchy mutant suggest GAs play a role in determining apical dominance, previously thought to be controlled by the antagonistic interaction of the plant growth regulators, auxin, and cytokinins. Our first attempt to determine a metabolic phenotype of a transgenic mutant poplar clone was successful.

Task 2. Effort was focused on the establishment of a comprehensive GC-MS library of TMS derivatives of organic constituents, particularly those of *Populus*. Many of the metabolites of *Populus* are shared with willow (*Salix* sp.). Leaf and stem samples of a number of poplar and willow species were first extracted, TMS derivatized, and analyzed by GC-MS. Many metabolites were identified and key fragments and retention time assigned to those compounds that remain unknown for future deconvolution analyses. A substantial database that includes the fragmentation patterns of many compounds was established that now includes ~ 450 metabolites with approximately half of TMS-derivatives of those metabolites not in commercially available databases (Wiley and NIST). The fragmentation pattern of known poplar compounds were predicted to aid in identifying unknowns. Recent metabolic profiles of *Arabidopsis* were used to confirm the presence or absence of positively identified metabolites in *Populus*. The fragmentation patterns of TMS-derivatized metabolites from our previous GC-MS analyses of *Populus* were matched to that of the Max Planck Institute databases.

Task 3. Reverse phase is certainly the most widely used mode in HPLC due to the wide range of analytes and sample matrices addressed and the chromatographic reproducibility offered by modern bonded-phase columns. Our work in this mode acknowledged that retention of the ionic phosphorylated metabolites would be difficult and non-ionic but polar metabolites might be retained irreversibly under the conditions suitable for the anionic components. The results indicate the column operated under two modes. Under conditions of relatively high organic modifier, none of the seven test compounds were retained very long and elution does appear to be in order of decreasing polarity. For instance, glucose phosphate eluted early, and ATP with the presumably more hydrophobic adenine portion of the molecule was more strongly retained. When operated with a mobile phase containing a high proportion of aqueous buffer, the retention order is changed but not simply reversed. Column efficiency was disappointing.

Hydrophilic interaction liquid chromatography (HILIC) has been suggested to offer potential advantages relative to reverse-phase chromatography for mixtures of very polar and ionic substances. The first column that we investigated for HILIC mode was the ThermoHypersil-Keystone BioBasic AX column. This is a bonded phase ion exchange column, with an imine functionality (R=C=N-H). This polar phase allows a water layer to form, which is necessary for the HILIC mode. Of course, for analyte compounds that are anionic, the column continues to show ion exchange behavior. Analyses were conducted using an ammonium acetate mobile phase in which pH was either not controlled or was adjusted to pH 8. Only after we adjusted the buffer to pH 10 did we see retention behaviors that could be acceptable. The separation was better than what we experienced in the reverse-phase mode. Unfortunately silica backbone column packings are well known to be unstable in high pH conditions; since the silica is eventually dissolved, the backbone collapses, and back-pressure increased catastrophically. Further work in this system was discontinued due to this shortcoming.

The use of a "neutral" HILIC column was evaluated to avoid the difficulty of confounding anion exchange and partition behavior, and hopefully exhibit only HILIC mode. The polyhydroxyethyl A stationary phase offered by PolyLC, Inc., has an aspartamide functionality which is claimed to present enough hydrophilic character to immobilize a significant water layer and to not present any anion exchange behavior. In fact, the column is also advertised as suitable for steric exclusion separations. The critical issues for the use of this column then are the proportion of water to acetonitrile and the salt content of the mobile phase that will be necessary to yield acceptable separations. Since a larger-pore-size column will exhibit lower capacity, we conducted some preliminary experiments with a 300A column. Retention of AMP was about a third that for the 100 A pore column, all other conditions identical. The retention of mono-anionic species might well be satisfactory, but di-anionic and presumably tri-anionic substances are retained excessively even at salt concentrations and water/acetonitrile conditions that would lead to poor ES-MS sensitivity.

Ion-pair reverse-phase chromatography was evaluated using two reverse-phase columns and triethylammonium acetate as the ion pair reagent. Peak widths were unacceptably large for serious consideration.

In light of difficulties in identifying a chromatographic system that would give adequate separation under conditions amenable to ES-MS, we chose to approach the larger metabolomics question from a different perspective. All previous work was in pursuit of an analytical system that would reliably identify and quantify a considerable proportion of the metabolites in a sample; a task that demands at least "good" separation of the mixture. An alternative approach to the larger question of metabolomics is to simply identify a technique that can reliably detect differences between samples. Direct-infusion ES-MS would seem to be an ideal approach. The capability to scan a wide mass range yielding semi-quantitative intensity vs mass to charge information would provide a unique sample "fingerprint." The potential criticism of this approach is the variation in sensitivity among species in, for instance, the negative ion mode electrospray mode. This difficulty can actually be turned into an advantage if the sample is examined in both the negative and positive electrospray modes and the negative and positive atmospheric pressure chemical ionization (APCI) modes. Leaf tissue extracts from two samples known to be genotypically distinct poplar clones were injected directly using the four ionization modes identified above. Differences were readily detected by the DI-ES-MS and APCI-MS experiments (Figs. 1 and 2), which lead us to conclude that fingerprinting by mass spectrometry is a potentially fruitful high-throughput avenue to pursue.

Summary and Conclusions

The testing of the different derivatization protocols confirms the need to tailor the protocols to the matrix being analyzed to maximize the amount of metabolite data available in the samples. The expansion of our TMS database gives ORNL a decided edge in being able to identify the metabolic perturbation induced by the introduction of a genetic mutation. Elucidation of many of the currently unidentified metabolites and expansion of that database will continue. Database expansion will serve to further improve our capability to conduct metabolic phenotyping. The completion of this study has brought us closer to being able to conduct such phenotyping at high-throughput rates.

As powerful and rapid as GC-MS is and may become as we migrate our analyses to high-throughput GC time-of-flight MS, further capability development was needed to expand the range of metabolites that can be detected in heterogeneous extracts and expand the spectral (mass) dynamic range of the detectable metabolites. DI-MS protocols were developed as rapid, high-throughput screening tools, where derivatization will not be required and the suite of metabolites profiled is further broadened. A wide array of metabolites is directly injected (infused) into a MS or tandem MS that is operated with multiple ionization modes, including positive and negative mode ES and positive and negative mode APCI. The analyses target metabolites that are otherwise typically difficult to derivatize or analyze by GC-MS because of large size (>1000 Da), too many reactive functional groups, and low volatility. Preliminary DI-MS analyses contrasting extracts

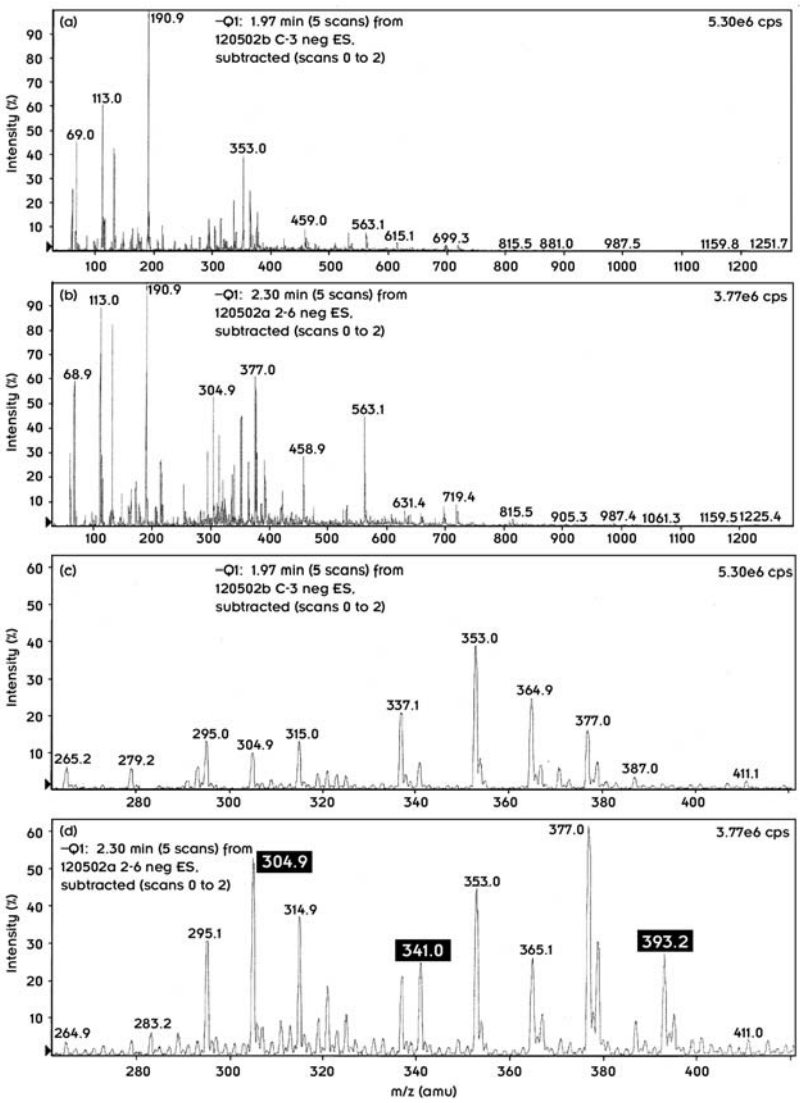

Fig. 1. DI-MS-ES ***negative*** *ion mode of extracts of (A) Populus tremula x alba nontransgenic (control) versus (B) activation-tagged mutant (GA_2ox). The magnified scale m/z scale (~250–450) is shown in (C) and (D), respectively. Note increased abundance of specific isotopomers in the mutant (D).*

from transgenic versus nontransgenic poplar clones have been very promising, with DI-MS readily extending the spectral dynamic range of the analyses from the 1000 m/z Da cutoff in GC-MS analyses up to 3000 Da. To extend this work to very high throughput, we can now apply what has been learned in this study to direct nanoelectrospray infusion analysis, making use of a NanoMate 100 that incorporates ES Chip technology (Advion BioSciences, Ithaca, NY). ORNL is currently one of only a very few labs in the world to have this most cutting-edge sample-handing, nanoelectrospray ion source system. Finally, the studies convinced us that separation and analysis of charged metabolites, such as adenylates, are not likely to be achieved by current reverse-phase HPLC if detection is by ES-MS. This includes the HILIC separations that are currently receiving considerable attention at several labs. As a result of the separations we observed, our approach is now to pursue chemical electrophoresis to resolve or separate charged metabolites, followed by MS or MS/MS analyses for detection and characterization.

The combined activities place ORNL in a more competitive stance as the funding agencies recognize the need to incorporate metabolomics data in functional genomics research programs in order to characterize how the biochemical networks are interrelated and to elucidate the function of unknown genes.

The most likely offices and programs to fund the follow-on research include DOE's Genomes to Life and OBER. The activities conducted over the course of this study allowed the continued development of a robust TMS-based library of mass spectra for GCMS analyses, and the initial development of deconvolution strategies that are essential for near-term metabolomics studies. The database and GCMS profiling approach were immediately (soon after the initiation of the LDRD project) included as a component of a large, multi-institution functional genomics proposal to a Request-for-Proposals solicited by OBER titled "Terrestrial Carbon Sequestration Using the Poplar Program." Our proposal titled "Genome-Enabled Discovery of Carbon Sequestration Genes in Poplar" (G.A. Tuskan, PI) was successfully funded at $1.7 M/yr for 3 yr, with approximately $1M/yr remaining at ORNL. Additionally, the aforementioned GCMS capabilities and the outcome of the LCMS protocol development were used to develop a large $3M (total funds requested) proposal to the Genomes-to-Life Program to further develop high- throughput techniques for metabolite profiling. Unfortunately, this latter proposal was not successfully funded. We will pursue follow-on funding with both NIH and NSF in this fiscal year.

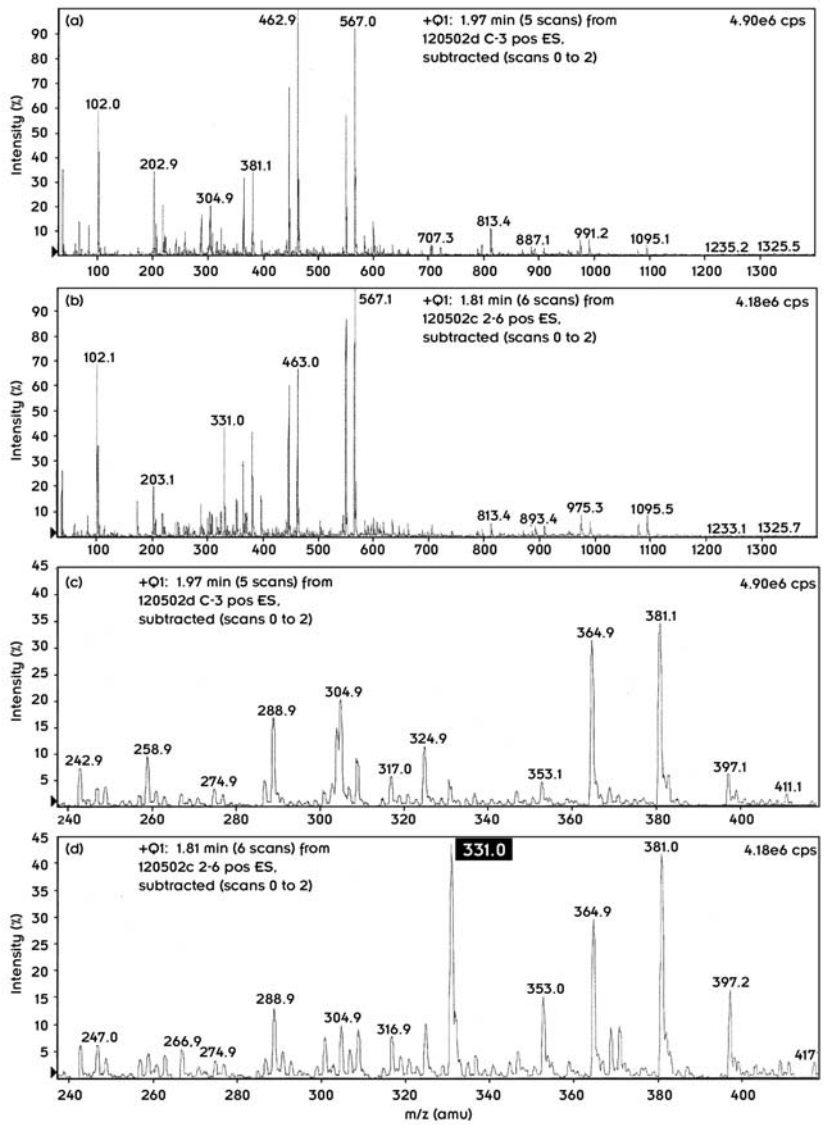

Fig. 2. DI-MS-ES ***positive*** *ion mode of extracts of (A)* Populus tremula *x* alba *nontransgenic (control) versus (B) activation-tagged mutant (GA_2ox). The magnified scale m/z scale (~240–450) is shown in (C) and (D), respectively. Note increased abundance of specific isotopomer 331 m/z in the mutant (D).*

Reference

[1]V. V. Tolstikov and O. Fiehn, "Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry," *Analytical Biochemistry* **301**, 298–307 (2002).

Proof-of-Principle Test for Endophyte-Mediated Effect on the Uptake of Metals by Tall Fescue

A. J. Stewart and M. Z. Martin
Environmental Sciences Division

A preliminary study was conducted to assess the opportunity for modifying a plant-endophyte system with the purpose of enhancing phytoremediation. Leaf material from endophyte-infected (E+) and endophyte-free (E–) tall fescue populations was analyzed by inductively coupled plasma mass spectrometry (ICP-MS) for 11 metals to determine if the presence of a fungal endophyte (*Neotyphodium coenophialum*) increased the ability of the grass to accumulate metals from soil. Of the seven metals with concentrations high enough to be reliably compared, most were higher in the E+ compared with the E– tall fescue. Additionally, the concentrations of Cu, Cr, Ni, and Pb were two to four times greater in the E+ compared to the E– grasses. However, very few of the differences were statistically significant, probably due to the low concentrations of metals in the soils of the test plots, the low number of samples analyzed, and the high analytical detection limits for several of the metals. Although the effect of the fungal endophyte on metal accumulation in tall fescue is uncertain, additional, more rigorous testing to assess its potential role in enhancing phytoremediation is warranted. An alternative analytical technique, laser-induced breakdown spectroscopy (LIBS), was used successfully for the first time to detect metals in grasses. Further research is needed to make this method more quantitative for routine use in the analysis of metals in biological material.

Introduction

The presence of endophytes in plants may enhance their ability to take up and accumulate metals from contaminated soils. Endophytic fungi that live within plant tissues can have large effects on their host, such as increased drought and heat tolerance,[1,2] lower rates of herbivory,[3] and greater tolerance to Al and Zn in soil.[4,5,6] By investigating the metal-accumulating mechanisms and attributes of a plant-endophyte system, it might be possible to eventually enhance the plant's phytoremediation potential. The well-studied model system of a grass (tall fescue, *Festuca arundinacea*) and its endophyte (*N. coenophialum*) was selected to determine if the presence of the fungal endophyte influenced the host's ability to accumulate metals from soil. The proposed study was a proof-of-principle test of this concept, and a necessary step that was required before a credible proposal could be submitted in response to an agency solicitation on phytoremediation.

Technical Approach

Four random samples of tall fescue for metals analysis were collected in December 2002 from each of seven grass seed production plots at the University of Tennessee Knoxville Experiment Station. Two of the plots contained E– forms of the grass and five plots contained E+ forms. The samples were dried, milled, and acid digested. The digestates were centrifuged, neutralized, and diluted (to reduce interference by nitrate) before they were analyzed by ICP-MS. A National Institute of Standards and Technology (NIST) reference material (apple leaf, SRM 1515) was digested and analyzed in the same manner for QA/QC purposes. The instrument was calibrated by analyzing blanks and external standards, both of which were also run during sample analysis to verify the instrument's calibration. Data from the blanks, standards, and NIST reference material were used to select the most appropriate mass for evaluation in cases where more than one mass per element was available. The elements of interest included Mg, Ca, Cr, Fe, Mn, Co, Ni, Cu, Zn, Cd, and Pb.

In addition to ICP-MS, an alternative analytical technique, LIBS,[7] was used to determine if metals of interest, such as Cd, Ni, and Zn, could be resolved in the milled plant material with minimal additional sample processing. To explore this laser application, 400- to 500-mg quantities of milled plant material were pressed into pellets (diameter: 1.3 cm; thickness: 4–7 mm) and analyzed individually with no further preparation. The laser used for LIBS was operated at a pulse of 23 mJ, an excitation wavelength of 532 nm, a delay of 4 µs, and a gate setting of 10 µs. Data were accumulated for 10 laser pulses with a 50-µm spectrometer slit width.

The data obtained by ICP-MS were evaluated using one-way analysis of variance (ANOVA). The plots

contained tall fescues that differed with respect to infection status (E+ or E–) and location (a proxy for soil type). The tests for effects due to plot position provided information on the effect of spatially linked edaphic factors on metal accumulation by the tall fescue-endophyte system.

Results and Accomplishments

Substantial differences were found in the concentrations of metals in E+ and E– tall fescue populations, but most of these were not statistically significant. For example, Cu, Cr, Ni, and Pb were two to four times higher in the tall fescue plots that were infected by the fungal endophyte compared to tall fescue that was endophyte free ($p > 0.05$). In another comparison involving E+ versus E– tall fescue in different seed production plots, E+ tall fescue had higher levels of Ca (18%), Fe (55%), Mn (54%), Ni (39%), Cu (21%), and Zn (13%) than E– fescue. However, in this comparison, only Fe and Mn were statistically significant ($p < 0.05$). Finally, two widely separated plots containing E+ tall fescue on different soil types were tested for spatial effect. In this case, the differences for Ca and Mg were large and significant ($p < 0.02$), but the differences for Mn, Cu, Ni, and Zn were not statistically significant ($p > 0.05$). No investigators have yet measured metals in tall fescue for the purpose of determining how the endophyte affects the host plant's ability to accumulate metals. Indeed, many investigators either do not know that the tall fescue endophyte exists or are unaware that it has a significant impact on the physiology of the host.[8] Although limited in scope, this preliminary study will contribute to the knowledge of endophyte-plant systems and to the importance to plant ecology of fungal-plant mutualisms.

The ICP-MS method was limited in its detection of low concentrations of metals (e.g., Cd) due to the need to digest samples, which diluted them. With LIBS, a sample can be analyzed directly with no dilution. Thus, the study provided an opportunity to assess LIBS as an analytical method especially suited for phytoremediation studies. With minimal sample preparation, LIBS unequivocally detected the presence of Cd, Mg, and Zn (Fig. 1). Other metals detected by LIBS have emission peaks that are outside of the 252- to 286-nm window shown in Fig. 1. Additional research is required to make this method of analysis quantitative enough for routine analysis of metals in plant tissues. However, such efforts should advance phytoremediation significantly, because metals could be analyzed at low cost and at near real time in the field. Such a technique would allow investigators to optimize phytoextraction by providing (1) better accounting of spatial variation in the amount of metal extracted, which is needed for closure on mass-balance calculations; and (2) better monitoring of the time-course of metal accumulation in aboveground plant tissues during the growing season.

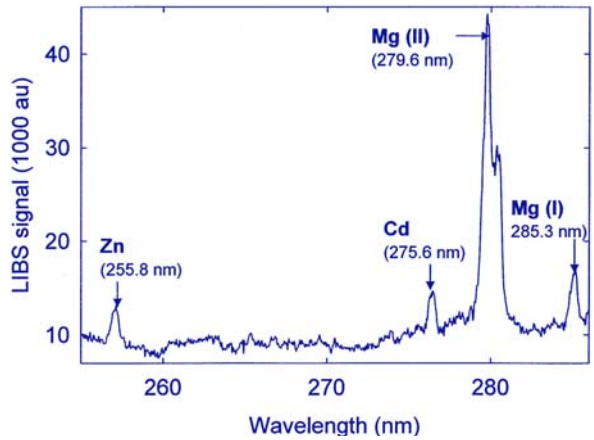

Fig. 1. Laser-induced breakdown spectroscopy (LIBS) emission spectrum for dried, milled tall fescue leaf samples. Peak-height data (Y-axis) are arbitrary units. Gate and delay times, excitation wavelength, and excitation energy can be varied to optimize the detection of metals at other emission wavelengths.

Summary and Conclusions

A preliminary study was conducted to assess the opportunity for modifying a plant-endophyte system with the purpose of enhancing phytoremediation. The study consisted of sampling and analysis to determine if an endophyte-infected grass (tall fescue) had different concentrations of metals than noninfected grass. Leaf material was collected from infected and noninfected tall fescue production plots at the University of Tennessee Knoxville Experiment Station and analyzed for 11 metals using ICP-MS and LIBS, a new experimental technique that had not previously been used to analyze metals in plant material. The tests for effects of infection status on the accumulation of metals in tall fescue leaf tissue were inconclusive. Sampling and analysis of selected metals in infected vs noninfected grasses did not provide strong evidence to support or refute the hypothesis that the presence of *N. coenophialum* affects the metal content of aboveground tall fescue tissues. Although the concentrations of some metals (Cu, Cr, Ni, and Pb) were two to four times higher in the E+ grass, the differences between the E+ and E– plots were not statistically significant. The LIBS technique unequivocally detected the presence of Cd, Zn, and Mg in tall fescue samples, but additional research is needed to make the method more quantitative for routine use in analyzing metals in plant tissues.

Although the results of this preliminary study provided only weak evidence of differences in metal accumulation between tall fescue populations with and

without a fungal endophyte, they emphasize the need for a more definitive and rigorous evaluation of endophyte-mediated effects on metal uptake in grasses. Many federal agencies, including DOE support research on phytoremediation. The use of grasses and other plants to sequester contaminants remains an attractive remediation alternative. It is a relatively low-cost, low-risk cleanup approach with an excellent record of acceptance by the public. By studying metal-accumulating mechanisms and attributes of a plant-endophyte system, it may be possible to develop a system that can specialize in the extraction of certain metals from soil. Even if that goal is not achieved, investigators who work with tall fescue must be alert to the possibility of endophyte-mediated effects, given the increasing use of tall fescue in phytoremediation studies and in the revegetation of disturbed soils.[8]

Efforts to obtain follow-on funding will be focused on federal solicitations for proposals related to phytoremediation. A draft manuscript describing the results of this preliminary study has been completed and will be submitted to the *Journal of Environmental Quality*. Also, this preliminary study has demonstrated the potential value of LIBS, and the data collected in this study will support future efforts to obtain funding to continue the development and application of this innovative technology.

References

[1]D. Malinowski and D. Belesky, "Adaptations of endophyte-infected cool-season grasses to environmental stresses: Mechanisms of drought and mineral stress tolerance," *Crop Sci.* 40, 923–940 (2000).

[2]R. Redman, K. Sheehan, R. Stout, R. Rodriquez, and J. Henson, "Thermotolerance generated by plant/fungal symbiosis," *Science* **298**, 1581 (2002).

[3]K. Clay, S. Marks, and G. Cheplick, "Effects of insect herbivory and fungal endophyte infection on competitive interactions among grasses," *Ecology* **74**, 1767–1777 (1993).

[4]D. Malinowski and D. Belesky, "Tall fescue aluminum tolerance is affected by *Neotyphodium coenophialum* endophyte," *J. Plant Nutrition* **22**, 1335–1349 (1999).

[5]F. Monnet, N. Vaillant, A. Hitmi, A. Coudret, and H. Sallanon, "Endophyte *Neotyphodium lolii* induced tolerance to Zn stress in *Lolium perenne*," *Physiol. Plantarum* **113**, 557–563 (2001).

[6]D. Zaurov, S. Bonos, J. Murphy, M. Richardson, and F. Belanger, "Endophyte infection can contribute to aluminum tolerance in fine fescues," *Crop Sci.* **41**, 1981–1984 (2001).

[7]M. Martin, M. Cheng, and R. Martin, "Aerosol measurement by laser-induced plasma technique: A review," *Aerosol Sci. Technol.* **31**, 409–421 (1999).

[8]S. Humphries, K. Gwinn, and A. Stewart, "Effects of endophyte status of tall fescue tissues on the earthworm (*Eisenia fetida*)," *Environ. Toxicol. Chem.* **20**, 1346–1350 (2001).

Permeable Environmental Leaching Capsules for Nondestructive In Situ Evaluation of Contaminant Immobilization Techniques in Soil

B. P. Spalding and S. C. Brooks
Environmental Sciences Division

A nondestructive technique using permeable environmental leaching capsules (PELCAPs), which encapsulate radioisotope-spiked soil within a water-permeable polyacrylamide matrix, is being tested as an in situ field method to measure immobilization of contaminants. PELCAPs have been retrieved and replaced in ground and surface waters many times during which they have been assayed nondestructively using gamma spectroscopy. As a proof of principle, soils contained in PELCAPs were labeled with two radioisotopes (^{85}Sr and ^{134}Cs) and have been leached in laboratory tests as well as deployed in ground and stream waters at Oak Ridge field sites. In situ retention of both isotopes within PELCAPs, containing thermally treated soil with high degrees of radioisotope immobilization, has been demonstrated up to three months in field tests.

The objective of this investigation is to develop and demonstrate an inexpensive, direct, and effective in situ technique to monitor contaminant immobilization non-destructively in the field using radioisotope-spiked soil contained within PELCAPs. The PELCAP technical approach has potential advantages for the assessment of induced and/or natural contaminant availability within soils, including the following:

- Nondestructive measurement of the amount of immobilized contaminant in a soil over time, thereby avoiding the necessity for repeated, costly, and destructive soil sampling;
- Direct comparison of several immobilization treatments, including a no-treatment control, under identical field conditions within the same well;
- Internal calibration of PELCAP leaching relative to specific reference tracers (^{85}Sr and ^{134}Cs) which have well-characterized environmental leaching behavior;
- Accelerated remedial alternatives evaluations by deploying PELCAPs in uncontaminated regions of a site to avoid possible masking by continued contaminant flux;
- Correlation of PELCAP behavior with widely accepted laboratory selective extraction and isotopic dilution protocols.

This PELCAP soil contaminant availability research should establish five research findings: (1) develop PELCAPs from which ^{85}Sr and ^{134}Cs freely diffuse with coefficients similar to those found in water; (2) develop PELCAPs which retain soil particles greater than 0.1 µm; (3) show that soil available forms of radionuclides in PELCAPs leach to a similar degree as unencapsulated soil in laboratory sequential extractions; (4) show that soil immobilized forms of radionuclides in PELCAPs behave similarly when leached in situ in field groundwater as when leached by laboratory methods; and (5) establish the methodology for PELCAP contaminant availability analysis, PELCAP durability for extended field deployment, and precision of PELCAP performance under field conditions.

A standard formulation for preparing PELCAPs in a polyacrylamide matrix has been established and tested using laboratory sequential extractions with ^{134}Cs and ^{85}Sr spiked soils. The sequential extraction behaviors of these radioisotopes from encapsulated soils and from the same but nonencapsulated soils were quite similar. Two suites of 54 PELCAPs each (spiked with isotopes, with and without soil, and with ambient or thermally stabilized soils) have been deployed in both stream water and groundwater. The natural in situ leaching of the isotopes from PELCAPs by natural waters in the field has taken several months to reach equilibrium but has behaved similarly to the laboratory-extracted PELCAPs. Stability of the immobilized radioisotopes in these field-deployed PELCAPs will continue to be monitored for up to six months, and this performance is a critical parameter to support their planned environmental application.

The PELCAP soil contaminant availability methodology is applicable to DOE's mission to predict the impacts of long-term stewardship of its waste disposal sites and practices. By providing a low-cost methodology to assess long-term performance of contaminated soil, these research findings should enable DOE managers to make judicious choices in selecting remedial actions.

Construction of a Gene-Prediction Algorithm in *Populus*: Adding a New Dimension to Complex Biology

L. E. Gunter[1] and F. W. Larimer[2]
[1]*Environmental Sciences Division*
[2]*Life Sciences Division*

The goal of this project is to create a unique annotation resource for the *Populus* genome sequencing effort. We will develop *Populus*-specific models for the functional analysis of genes using novel information contained in a library of full-length cDNA clones derived from various *Populus* tissues. A collection of accurately characterized, full-length cDNA sequences will allow us to develop a proof of principle for automated gene-prediction algorithms that are critical for recognition of *Populus* genes. The design and implementation of a precise gene-prediction tool for the annotation of the *Populus* genome will open new avenues of functional genomics at ORNL.

The Joint Genone Institute has recently announced its completion of the complete sequence for *Populus*, the first tree and only the third plant to have its genome sequenced. The goal of this proposal is to fully exploit the opportunity afforded by the whole-genome sequencing effort by designing and testing a novel annotation tool that will establish ORNL as a leader in *Populus* functional genomics. The development of this type of informatics tool will furnish us with a resource that will enhance our ability to address future DOE missions in functional genomics of *Populus* and increase the scientific and technical vitality and visibility of ORNL through securing a national and international role in *Populus* genomics. This knowledge could have profound implications for carbon management issues related to long-lived perennials, conversion of biomass to fuels and products, and the deployment of plant-based systems for phytoremediation. The objectives of this research are to (1) provide sequence input of unique *Populus* structural genes to gene-prediction models and (2) design a customized version of GrailEXP to provide unique and consistent exon finding and gene model algorithms for *Populus*.

A large collection (>1,000) of full-length cDNAs from libraries derived from a variety of tissues of *Populus trichocarpa* "Nisqually-1" will be sequenced using dye terminator chemistry, then assembled into full-length reads. Raw sequence data files will be analyzed to assign quality scores to sequenced fragments prior to assembly. Only nucleotides assigned a quality score of Q40 with a 2X or greater coverage will be used to assemble full-length cDNA sequences that will be used to construct the gene-prediction algorithm.

Populus cDNA and EST sequences will be aligned with genomic draft sequence. From these alignments, we will extract a description of exon size and number distribution, coding exon hexamer distribution, intron size distribution, intron hexamers, and splice site structure. These results will be used to train the Perceval (Protein-coding Exon, Repetitive, and CG-Island EVALuator) neural net for exon finding and the Galahad Gene Message Alignment routine for splice junction optimization. If strictly patterned intron structures are recognized, we will configure the Gawain (Gene Assemblies with Alignment Information) gene model assembler accordingly. After the construction of a set of high-quality gene models, we will extend the collection of *Populus* signal motifs using a recently developed motif finding tool, MIRA. This motif finding routine is a background rareness–based iterative multiple-sequence alignment algorithm for regulatory element detection. In addition to motif finding in raw sequence, this tool can utilize clustering information from expression or proteomics data to enhance its search specificity.

To date over 800 full-length poplar cDNAs have been identified from the libraries. Of these, 320 have been completely sequenced and assembled, and another 352 are awaiting internal sequencing for closure. Genome Canada will be contributing another ~1000 full length cDNA sequences to this effort. These cDNAs will be aligned with the poplar genome sequence, which has recently been completed, and for which a draft assembly will be available by late November or December 2003. In addition to the cDNAs, Stefan Jansson from Sweden has submitted 30,000 contributed ESTs to the assembly. This will enhance the gene prediction models and allow more accurate gene calling.

Preliminary hexamer frequency plots have been derived from the first 320 fully assembled cDNAs. Of

these, 122 have high-identity alignments with their Arabidopsis homologs. The sequences are sufficiently similar to use the Arabidopsis gene models to simulate the splicing patterns of poplar. From these analyses, we have generated preliminary profiles of coding exons, coding hexamers, and splice site structure. Exploratory studies have been carried out using an Arabidopsis-trained gene finder against ~1mbp of preliminary poplar assemblies. These results indicate that the poplar intron structure and distribution is sufficiently different from Arabidopsis to require retraining. Alignment of the cDNA library against the full poplar assembly will be necessary to develop the intron profiles for a full training set.

The development of this novel annotation model for poplar will act as a bridge to allow current research to move beyond the sequence and focus on a comprehensive understanding of biological and ecological systems. Our primary purpose in this approach is to generate a capability for engaging in the types of scientific research that DOE considers significant, that is genome-enabled carbon sequestration, phytoremediation, ecosystem science, etc., for which solid annotation of the poplar genome is a prerequisite.

Sounds of Rapids as an Attractant for Migratory Fish

C. C. Coutant,[1] M. S. Bevelhimer,[1] and G. R. Wetherington[2]
[1]*Environmental Sciences Division*
[2]*Engineering Science and Technology Division*

Downstream-migrating salmon and other migratory fish have difficulty finding their way past dams. This project is exploring whether the complex underwater sound of a natural river's rapids serves as an attraction for downstream orientation. If true, artificially generated sound could be used to guide migrants to fish bypasses or through reservoirs where migratory cues are lost. The objectives are to (1) obtain recordings and characterizations of underwater sounds using field measurements and sound-characterization software, and (2) identify the capability of such complex sounds (played back recordings) to orient fish in quiet water in the laboratory and eventually in the field.

The project is aimed at developing effective fish guidance systems that would divert migrating fish away from hydropower turbines (where many are killed) without spilling large amounts of water for fish passage. We are exploring whether fish could be guided by simulating a natural behavior cue. Fish would be guided to alternative passage routes around the dam that successfully pass more fish with less water. Spilling water for salmon passage on the Columbia River, for example, reduces water available for generation and costs the DOE/Bonneville Power Administration over $1 million per day in lost generation. The DOE Hydropower Program's goal of a 10% increase in hydropower generation in the United States could be met by simply converting spill in the Federal Columbia River Power System to generation.

We began this project by setting up three of the classical "U-tank" behavioral-choice experimental tanks in the ORNL Aquatic Ecology Laboratory. We were able to modify existing artificial stream channels, which resulted in considerable cost savings. The tanks consist of a long U, with approximately 10-m-long, 0.3-m-wide, and 0.3-m-deep arms. A speaker connected to a Compact Disk player is placed in one of the two arms (or neither, for a control). The entire U-tank area is enclosed by curtains to avoid visual interference. Experimental fish are introduced at the base of the U, where the arms join. Fish position is recorded with a video multicamera array that provides coverage of the whole U. Video frames are recorded at set intervals (e.g., every 2 seconds) for the duration of an experimental trial. We have explored the use of the experimental equipment (using local stream fish species) but have not yet conducted formal trials.

We have made tentative arrangements for several potential field test sites at hydropower dams, pending completion of successful laboratory tests to be conducted during FY 2004.

This project, if successful, will benefit DOE and other federal agencies [e.g., Corps of Engineers, Bureau of Reclamation, NOAA Fisheries, Fish and Wildlife Service)] concerned with maximizing use of renewable hydropower with minimal detrimental effects on fisheries resources.

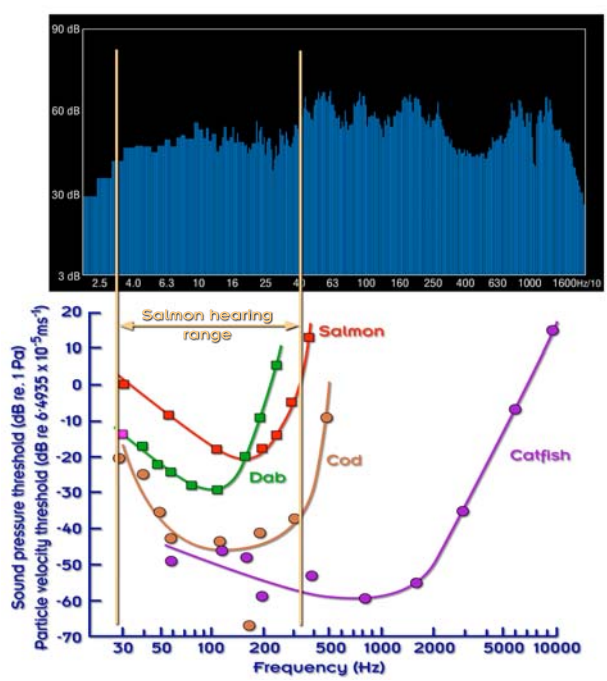

Fig. 1. Underwater sounds of rapids (audiogram in upper panel) occur at frequencies to which many fish species are sensitive (lower panel; the lower the threshold, the greater the sensitivity). The ORNL research will test whether migratory fish are attracted to such sounds, which might be used for guiding fish past obstacles such as dams.