Plant-Microbe Interfaces: Detection and characterization of signaling polypeptides in plant-microbial systems using high-performance tandem mass spectrometry

Robert Hettich[1*] **(hettichrl@ornl.gov),** Paul E. Abraham,[1] and Richard J. Giannone[1]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the Plant-Microbe Interfaces (PMI) SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.

Small polypeptides are emerging as key signaling molecules that mediate cell-cell communication and various biological processes that occur within and between plants and microbes. These polypeptides can be secreted or non-secreted and are generally categorized into two classes: 1) small polypeptides (~5 to 20 amino acids in length) that are post-translationally cleaved from larger precursor proteins and 2) gene-encoded small polypeptides (i.e., small proteins that comprise of ~40 to 100 amino acids).

The discovery and characterization of these polypeptides is challenging because their post-translational processing is not well understood nor predictable. Many polypeptides can harbor post-translational modifications (PTMs) such as tyrosine sulfation, proline hydroxylation, hydroxyproline, and arabinosylation. In some cases, these polypeptides can form cyclic structures via disulfide or isopeptide bonds. Adding to this complexity, previous studies have shown that small open reading frames (sORFs) embedded in larger ORFs can encode additional sources of functional polypeptides. As such, gene expression levels provide little, if any, value in the characterization of these small polypeptides. Instead, high performance liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is an effective means to identify and characterize these endogenous peptide molecules.

A traditional peptide sequencing measurement (LC-MS/MS) benefits greatly by having prior knowledge of the potential sequences. However, given the extent to which these small polypeptides can be derived and post-translationally modified, traditional bioinformatic approaches have limited applicability for their identification. Because these small polypeptides are a critical signaling component in plant-microbe interactions, there is an increasing need for better methods to facilitate their detection and identification. To this end, we have developed new MS measurement approaches and analysis platforms for *de-novo* sequencing of these small

polypeptides using high-performance LC-MS/MS instrumentation (ThermoFisher Q-Exactive Orbitrap mass spectrometry).

To effectively measure these polypeptides, we optimized a sample preparation protocol (molecular cutoff filtering) for a customized hybrid MS approach that integrates full mass spectral scans (for global measurements of the full set of sample components) with targeted scans that focus on m/z regions encompassing multiple forms of the specific polypeptides of interest. By employing HCD fragmentation methods in a QE-MS instrument, we found that most cyclic peptides usually can be fragmented in sufficient detail to enable sequence tag identifications that can verify suspected amino acid linkages. In particular, we utilized a software approach entitled DirectTag to identify sequence tags in these tandem mass spectra, but are examining more sophisticated approaches that are designed to characterize unusual and cyclic peptide forms.

Recently, it has been shown that plants harbor genes encoding signaling polypeptides that can be expressed in one organ, for instance the shoot, only to be transported to another organ, like the root. Therefore, we evaluated our MS approach to identify novel polypeptide species across several plant sources (i.e., leaf, xylem sap, and root). Overall, using this new approach, we were able to identify 23,572 polypeptide sequence tags (with lengths of at least 7 amino acids) that serve as a foundation to identify the full-length polypeptide species and proteins of origin. Hierarchical clustering revealed distinct groupings of these polypeptides within root, leaf, and stem tissues. We have noted that several of these do not match *de-novo* sequence annotations, indicating the presence of unknown post-translational modifications, and thus have begun more extensive characterizations.

The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research.