Title: Combining Multi-omics with Random Walks to Explore the Function of Candidate Genes

Authors: David Kainer[1]* (kainerd@ornl.gov), Izaak Miller,[1] Mikaela C. McDevitt,[1] Daniel Jacobson,[1] and **Paul E. Abraham[1]**

Institutions: [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Project Goals:

The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

Abstract text:

The SEED project will inoculate a *Populus trichocarpa* (Black Cottonwood) genotype with over one hundred strains of the fungal pathogen *Sphaerulina musiva* (formerly Septoria) in order to measure a phenotype that describes the success of microbial invader establishment. A wide array of omics data will be assayed in the Septoria population, including genomics, transcriptomics and proteomics, which will allow us to explore the factors that influence an invader's ability to become established. GWAS (Genome Wide Association Studies) will be used to associate Septoria genomic variation with variation in establishment phenotypes. However, GWAS for complex traits often provides cryptic results. To unravel the complexity of establishment dynamics we will apply network analysis techniques coupled with Explainable AI (X-AI) algorithms so that all aspects of omics data are incorporated jointly with the GWAS results.

Each available omics layer, including publicly available data on *Sphaerulina musiva* and related species, will be converted to a network format that connects genes to genes based on functional and experimental relationships. Multiple networks will be combined into a multiplex network[1] for use with machine learning algorithms such as Random Walk with Restart (RWR)[2]. The RWR approach enables the exploration of the function of candidate genes using network-topology-association without being limited to their immediate neighborhood. The multiplex network jointly uses the gene-gene connectivity from every omics layer while maintaining the information encoded within the unique topology of each layer. Together this is a powerful multi-omic analysis framework.

References/Publications

1. Battiston, Federico, Vincenzo Nicosia, and Vito Latora. "Structural measures for multiplex networks." *Physical Review E* 89.3 (2014): 032804.

2. Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaëlle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, Anaïs Baudot, Random walk with restart on multiplex and heterogeneous biological networks, *Bioinformatics*, Volume 35, Issue 3, 01 February 2019, Pages 497–505,

Funding statement: *The Secure Ecosystem & Engineering Design Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKPA17*. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.