

Statement of Research Interests

There are an estimated 100 million species on our planet, and they come in an impressive range of sizes, shapes, forms and physiologies. This great variation in life forms on Earth is a result of diversification at the levels of genes and networks over thousands to millions of years. The question of most interest to me is – what are the mechanisms that contribute to the origination and evolution of complexity in the living world? I believe **plant specialized metabolic pathways** are good models for addressing this question. By definition, specialized metabolites are evolutionary innovations that occur in a phylogenetically restricted fashion. Given that many of these metabolites mediate interactions between plants and their environments, they are important from both an ecological/evolutionary and an agricultural context. Research focused on specialized metabolism can help provide solutions for the significant challenges of tomorrow including enhancing plant resilience to the effects of climate change, reducing the use of synthetic pesticides, boosting beneficial plant interactions as well as making available novel medicinal compounds. Thus, it has the potential to attract funding from funding agencies focused on more basic research (eg: various divisions of NSF, NIH) as well as those with a more translational focus (eg: USDA, Bill and Melinda Gates Foundation).

My research program will be focused on functional, evolutionary and predictive genomics of specialized metabolic networks. The participants in my program and I will use cross-disciplinary approaches including genomics, transcriptomics, metabolomics, bioinformatics as well as hypothesis-driven experimental research to create a multifaceted understanding of plant metabolic networks (Fig. 1). I believe I have been trained in using these approaches through my PhD training in computational evolutionary genomics and my postdoctoral research in biochemical evolution. The outcome of this research program will include experimental knowledge and predictive models of protein function in metabolic pathways as well as a better understanding of the evolution of pathway complexity at different taxonomic levels. The two research areas born out of my interests are described below.

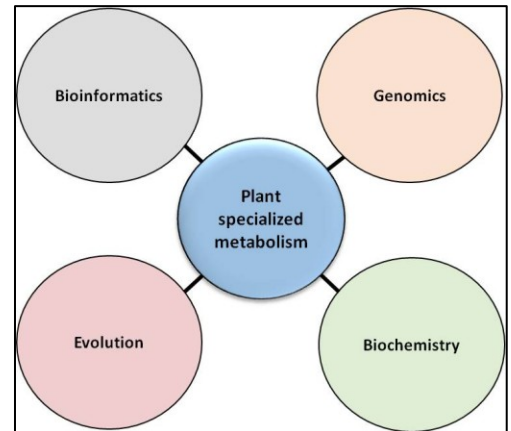


Fig. 1: The proposed research program will develop new knowledge using cross-disciplinary ideas and approaches.

Predictive and evolutionary genomics of metabolic enzyme families

Despite the advent of cheap sequencing and powerful computational tools, genes involved in metabolism remain poorly annotated. One of the reasons for this paucity is the involvement of large gene families comprising of dozens of duplicated members. Functional divergence and/or redundancy between these duplicates creates obstacles in gene function inference, pathway characterization and metabolic engineering of crop plants. I plan to tackle this issue by creating experimental knowledge and computational models for functional inference of metabolic gene family members. To generate such models, we need to identify any rules that allow computational prediction of duplicate gene function.

Previous work: My PhD research involved development of computational strategies to distinguish functional intergenic transcripts from junk (Moghe et al, Plant Phys, 2013), machine learning models for predicting gene fate and function (Moghe et al, Plant Cell, 2014; Lloyd et al, Plant Cell, 2015) as well as helping in the development of the popular gene annotation program MAKER-P (Campbell et al, Plant Phys, 2014). In my postdoctoral research, I focused on functional characterization of a sub-clade of the large BAHD enzyme family (Fig. 2) in multiple non-model Solanaceae species such as *Salpiglossis sinuata*, *Hyoscyamus niger*, *Solanum nigrum* and *Solanum quitoense* (Moghe et al, in preparation; Hurney et al, in preparation), using diverse experimental and computational approaches.

Planned work: While conducting the above research, I frequently grappled with the poor quality of annotation of metabolic enzyme families. To address this persistent problem, as a start, I plan to characterize the functions of various BAHD enzymes using reverse genetics, RNA-seq, mass spectrometry and other methods. **I am currently generating CRISPR-Cas9 deletion lines for four BAHD genes in cultivated tomato. Other genetic resources such as RNAi and introgression lines are also available for this study.** These experiments, in addition to generating valuable insights on individual gene function, will also generate useful data for the computational modeling component of predictive genomics, involving network reconstruction and machine learning (Moghe et al, Plant Cell, 2014). This paradigm can be

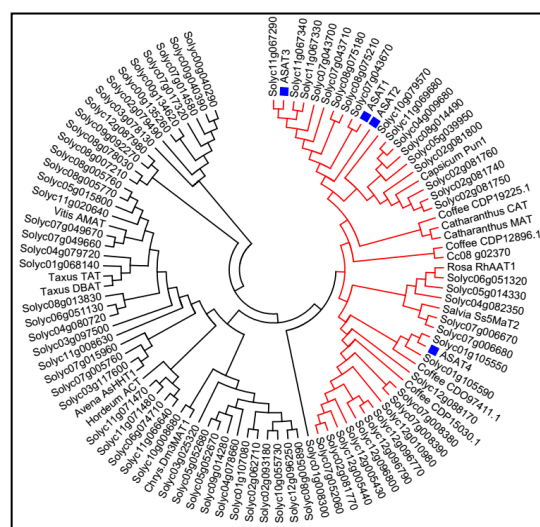


Fig. 2: BAHD family in tomato (Solytc) and other species. Red clade includes all the acylsugar biosynthetic enzymes (blue squares).

extended to address evolutionary questions such as “How do metabolic networks evolve across different time scales?” and “Is there selection on the evolvability of certain enzymes?” (Moghe and Last, Plant Phys, 2015). Towards these goals, I hope to collaborate with chemists, computer scientists, protein structural biologists and experimental biologists to create an integrative research program.

Significance: The proposed research program will significantly contribute to protein function discovery in plants, mitigating the problem of poor annotation quality and aiding metabolic engineering and synthetic biology. **The scope of this study is not tomato BAHD family specific, since the strategy and the rules learnt via the predictive modeling framework can potentially be applied to other enzyme families/species.** In addition, extending this research into other Solanaceae species would provide an integrative understanding of enzyme evolution.

Understanding the origins of pathway complexity

An estimated 200,000 metabolites exist in the plant world. This diversity is a result of evolutionary forces tinkering with metabolic pathways. My goal is to deconstruct the structure and evolution of agriculturally

important specialized metabolic pathways. I have been working on this aspect of biochemical evolution in my postdoctoral research, and I will continue with new projects that have emerged from my studies.

Previous work: During my postdoctoral research, I discovered novel enzyme activities involved with acylsugar biosynthesis in multiple Solanaceae species as well as surveyed related families and orders. Species in the sister Convolvulaceae family produce analogous compounds called resin glycosides (Fig. 3), which may involve the action of enzymes from multiple enzyme families. Over the past few months, I have built a research program to comprehensively probe the biosynthesis, function and evolution of these allelopathic compounds in Convolvulaceae. **With an REU undergraduate student from Cornell, I sampled 13 Convolvulaceae species, selected those with resin glycosides and found that extracts of a morning glory species enriched in resin glycosides have differential effects on germination and root growth of crops and weeds, including an inhibitory effect on root growth of *Echinochloa crus-galli* (barnyard grass), one of world's most successful weeds.**

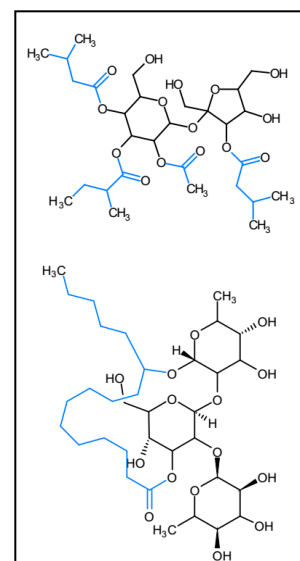


Fig. 3: (Top) Acylsugars, found in Solanaceae. (Bottom) Resin glycosides, found in Convolvulaceae

Planned work: I intend to develop this project to characterize enzymes responsible for resin glycoside biosynthesis as well as their potential agricultural benefits. A complementary objective is to study the origins and evolution of this pathway in Convolvulaceae and understanding its relationship with the acylsugar biosynthetic pathway. I will extend this research to investigate other compounds and species of agricultural and ecological importance, for example, the metabolites involved in **root-microbiome interactions**, using closely related *Solanum* species as models. Students working towards these goals will also have opportunities to bioinformatically study evolution of metabolic pathways using evolutionary genomic approaches across the plant kingdom.

Significance: Resin glycosides are important for sustainable agriculture since they can help reduce the use of synthetic chemicals in the field. They can also provide insights on plant-plant, plant-microbe and plant-nematode interactions. The evolutionary studies can help understand the evolution of the BAHD enzyme family over the past 100 million years, and reconstruct different scenarios for the emergence of pathway complexity.

I believe that “personalized agriculture” is the future of plant science. However, an integrated understanding of plant physiology – from individual gene function to the impact of network evolution on plant interactions – is important for creating “designer plants” suited to specific environments. My research program, focused on plant specialized metabolism, seeks to integrate various omics technologies in a predictive framework and through the “light of evolution”, making it a truly interdisciplinary and collaborative endeavor. Thus, I believe it will be a unique and important step towards addressing the new challenges of tomorrow.