

STATEMENT OF RESEARCH INTEREST

As a hypothesis-driven empirical population geneticist, I want to move beyond detecting natural selection toward understanding adaptation in its demographic and ecological context. Population genetics has unrealized potential to provide insights into important biological phenomena, particularly through its integration with ecology. Host-pathogen interaction is an ecology rich in adaptation and broadly relevant, with its own theoretical framework (epidemiology, more broadly population ecology) for integration with population genetics to realize this potential. This potential looms large with the technical advances of genome level approaches, and challenges population genetics in this era not only to keep pace with the data but to make analytical advances practicable to scientists with a range of expertise and perspectives.

My research uses molecular data to study processes on timescales ranging from the present day to millions of years ago. My research formula is to develop and implement molecular markers and model-based analyses to infer demography and population structure, and to characterize the genetics underlying adaptive phenotypic variation. So far I have worked mostly on *Arabidopsis thaliana*, a selfing weedy annual common in the eastern U.S., and a model in molecular genetics, population genetics, and ecology. In addition I have worked on *Ipomopsis*, *Echinacea*, humans and HIV, and on theory general to any organisms.

Demographic processes under study in *A. thaliana* include self-fertilization, dispersal within and between populations, and expansion and admixture of ancestral populations. Likelihood-based coalescent analyses of DNA sequence variation data suggest admixture with unequal contributions from ancestral subpopulations [11]. Polymorphisms at three loci in population samples reveal a low but non-zero outcrossing rate, limited long-distance dispersal and frequent extinction of local populations [7]. And spatial autocorrelation of multi-locus genotypes along local population transects indicate the spatial scale of individual dispersal, on the order of one meter [15]. These results provide a context of *A. thaliana* population structure and history, from the local population to metapopulation and species levels. In species of conservation interest, demographic inference allows assessment of evolutionarily significant units, and provides a context for assessing differentiation of adaptive traits.

Local adaptation is an evolutionary process in which environmental differences interact with limited gene flow to maintain phenotypic variation. *Arabidopsis thaliana* accessions' drought adaptation phenotypes correlate with historical weather data at their sample localities, suggesting that *Arabidopsis* populations are locally adapted to climate [13]. In a functional genomic study of drought adaptation in *A. thaliana*, John McKay (U.C. Davis), Tom Jeunger (U. Texas) and I are proposing to NSF this fall to map and clone the genes underlying drought adaptation through combined ecological and molecular genetic experiments. The genetic architecture (pleiotropy, plasticity and epistasis) of this complex trait, and the molecular population genetics of genes underlying it, will provide complementary approaches to understanding plant adaptation to drought. Molecular population genetic models incorporating natural selection [2], and considering phenotypic allelic effects, will be central to association mapping of complex traits, and to understanding the contributions of molecular variants to phenotypic variation and species evolutionary potential. We expect the ecology and genetics of drought adaptation in *A. thaliana* to be conserved across plants, and therefore applicable to crops as well as endangered species. First on our list for application is *Arabis fecunda*, a close relative of *Arabidopsis thaliana* and an endangered species across a water-limitation gradient on mountain slopes of Montana.

My primary focus is on plant adaptation to disease. Gene-for-gene disease resistance is a burgeoning topic in plant molecular genetics. Resistance (R-) genes show conspicuous signs of adaptive evolution in molecular evolution and patterns of polymorphism [5]. Analyses of nearly 200 R-gene homologs from the complete *Arabidopsis* genome sequence reveal a gold mine of adaptive evolutionary histories that will lead to important discoveries including novel disease resistance specificities [4]. In addition, R-gene homologs are readily attainable genome-wide markers for genomic, evolutionary and ecological genetic studies in almost any plant species. Adaptive, functional variation at the R-gene loci *Rpm1* [6], *Rps5*

[2], *Rps2* [1] and *Rpp8* [12] reveals ubiquitous trade-offs for plant gene-for-gene disease resistance. An ecological genetics approach, akin to linkage disequilibrium mapping, promises to unravel how these trade-offs play out in terms of spatial and temporal variation of disease. In large samples from polymorphic populations, flanking markers show low population differentiation within relative to between *Rpm1* alleles, consistent with temporal variation in selection in asynchronous populations [9]. This study illustrates how linkage disequilibrium mapping is influenced dramatically by population structure and inbreeding, and by the consequences of variation being mapped [3]. Modeling molecular variation under selection has provided a framework for interpreting *Rps5* [2] and *Rps2* [1] data, and will shed light on the spatial structure of the disease interaction involving *Rpm1* [9]. Incorporating genetic exchange between paralogs will help us to understand the origin and maintenance adaptive variation at complex, polymorphic loci such as *Rpp8* [12], with particular ramifications of mating system for the distribution of variation between orthologs versus paralogs, and will allow genomic studies of gene family adaptive radiations. Importantly, analyses of these models under a range of demographic and ecological parameters will suggest what host-pathogen system is ideally suited for this ecological genomics approach, and how to make predictions based on an ecological genomic approach in applied settings.

Coalescent models of identity by descent have a wide range of applications, and are a workhorse for population geneticists in evolutionary biology and conservation. My theoretical work has the ultimate goal of integrating population genetic and ecological approaches. Coalescent models can be formulated quite realistically, to relate molecular data to hypotheses for species demography and for adaptive genetic systems. Computational methods based on Markov chain theory, *e.g.* Markov chain Monte Carlo, allow powerful and efficient likelihood analyses, and in a Bayesian framework can incorporate independent prior information [11].

Theory describing polymorphism in a ‘focal subpopulation’, under the population biological processes colonization, logistic growth, immigration and extinction [14], allows parameters to be studied without assuming a specific model of population structure (which is extremely problematic in practice), under assumptions less restrictive than standard spatially implicit models. Further, model parameters and variables such as environmental quality can be studied in space either empirically (with non-parametric correlation) or in spatial model formulations. I hope that this theory will be widely relevant to high population growth species of conservation and intellectual interest.

Currently I am working toward a synthesis of coalescent models and epidemiological models of population size dynamics allow rigorous inference of ecological parameters using molecular data. These models will allow me to infer the time of origin, growth rate and future extent of the HIV pandemic. This theory is relevant to studies of molecular variation in systems of interacting species generally, and will be a cornerstone of my research in plant-pathogen ecological genetics.

Thus, I look forward to a laboratory where projects involve (1) molecular population genetic inference of demography, and (2) evolutionary and ecological genetic study of adaptive traits. Computational analyses will integrate evolutionary, population genetic and ecological models will bring molecular data to bear on our understanding of demography and adaptation in the field. My ultimate goal is to work in a system of interacting species. I plan to concentrate on plants, including *Arabidopsis* and close relatives (which exhibit a range of life histories and some of which are endangered) but branching out to plants of local conservation interest, while continuing human pathogen and general theoretical work.

I welcome students, postdocs and colleagues from a range of backgrounds including Computer Science, Statistics, and fields within Biology from Genomics to Ecology. Finally, I look forward to developing and sharing the resources for this kind of work, and I hope to ensure their relevance and applicability.