PROJECT SUMMARY

Overview:

The major goal of the proposed research is to understand how extinction and colonization in interconnected populations (metapopulations) influences the genetics of populations and the evolutionary processes that create and destroy genetic diversity. Evolution in such situations is predicted to be very different from that in single populations, but there have been almost no tests of these predictions in nature. This proposal builds on a now 28-year study of the numerical dynamics of Silene latifolia (white campion) and its pathogen Microbotryum violaceum (anther smut). Recent NSF funding incorporated population genetic studies into these methods, with initial data providing significant insights into the evolution of genetic structure and the genetic basis of colonization success. A long-term population genetic approach is needed to characterize sufficient colonization and extinction events in order to obtain longitudinal data on genetic change in newly colonized populations, and to assess the role of population history on genetic processes. Methods used will include long-term monitoring (eventually more than 800 populations for >30 yrs) and continued sampling (eventually >15yrs). High throughput genotyping of Silene populations will provide unprecedented resolution of changes in genetic structure, including how different regions of the genome may experience variation in Ne and migration, and how genetic structure determines both the source and the success of new colonists.

Intellectual Merit :

This research will investigate how the view of evolution changes when populations are seen as interconnected systems that undergo extinction and colonization, rather than as isolated entities. Theory and prior results from this project suggest these effects could be profound, resulting in a redistribution of genetic variance (opening the potential for inter-demic and context dependent selection), severe inbreeding and extinction (with the possibility of genetic rescue through gene flow), and changes in response to selection. Theoretical work predicts that different regions of the genome will experience very different genetic structuring mechanisms, even at neutral loci. Earlier marker methods have already demonstrated that genetically variable populations give rise to a greater number of new demes, that founder effects among these new demes increase the magnitude of genetic structure, and that subsequent immigration decreases such structure and rescues small populations from inbreeding. Over the next five years, demographic studies and sampling will continue to expand out understanding. These studies will be combined with RADtag sequence methods and assembled genomes of the host to investigate the impact of metapopulation dynamics on different regions of the genome such as the X- and Y- chromosomes, autosomal loci and mitochondrial and chloroplast loci. The demographic census, the banking of seeds, DNA and pathogen cultures will eventually provide nearly four decades of longitudinal data, offering an unprecedented view of the parentage of newly established demes, gene flow during population expansion or contraction, the genetic consequences of seed banks and the importance inbreeding and genetic rescue for population persistence. Also, broad demographic shifts have occurred during the course of this study, opening up the potential to explore the population genetics of non-equilibrium systems.

Broader Impacts :

There is now widespread public concern about conservation, invasions, and global change. Genetic factors have been increasingly posited as crucial for understanding species responses in these situations. The present study will be one of the first to address these issues empirically by direct study of a model field system that shares the characteristics of many other natural populations. The Silene/Microbotryum system has become a model for numerous labs, and this research will provide a publicly available long-term data set of demographic and genomic data, living collections, and DNA samples. This study will provide scientific training in timely important fields of ecology, genomics, bioinformatics, and computational biology. The PIs will continue to encourage international interactions and involvement of groups historically underrepresented in science.

assembled genome for its anther-smut pathogen. Our work has been leveraged by local researchers to obtain funds on conservation and bioremediation in the local regional park where we work. Our work has been reported in newspaper articles in Italy, and has led to new ideas and proposals relating to host-specificity, vector transmission, and the evolution of dioecy.

II. PROJECT DESCRIPTION

1. CONCEPTUAL FRAMEWORK

This research will explore how the process of evolution is affected by the reality that individuals (and their alleles) are often clustered into spatially distributed populations. In spite of a wealth of theoretical insights dating back to Wright (1931), there is a serious paucity of real-world data on the factors that generate population genetic structure, and the evolutionary consequences of that structure.

To directly observe the evolution of genetic divergence as populations are founded, experience gene flow, and undergo demographic and genetic change over time, we will combine long-term demographic data on more than 800 plant populations with a longitudinal population genetic study. Recently derived assignment methods will allow us to study migration directly, to trace the sources of colonists, and follow changing patterns of inbreeding and genetic rescue, all issues of direct relevance to conservation and invasion processes.

The research will evaluate and identify the factors that contribute to the colonization, persistence and extinction of interconnected populations, informing the long-standing debate about the importance of multi-level selection. The research will be relevant to the applied areas of conservation biology, invasion biology and infectious disease dynamics. It will provide training in fields of timely importance to science (e.g. genomics, bioinformatics, statistical genetics, epidemiology). The long-term demographic and genetic data, plant and pathogen isolates and DNA, will be powerful resources for future research.

A. RATIONALE AND SIGNIFICANCE

The theory of interconnected populations (metapopulations) has shown that spatial structure can have profound effects on genetics and evolution. The earliest models of interconnected populations (Wright 1931) quantified how genetic differentiation at neutral loci was generated by drift and diminished by gene flow. In the island model, an equilibrium balance between drift and migration generates the well-known expectation that $F_{ST} \approx 1/(4Nm+1)$, which predicts little genetic differentiation (as measured by F_{ST}) unless migration is very rare.

Many of the simplifying assumptions of the island model are violated in nature (Whitlock and McCauley 1999). In metapopulations, for example, where demes experience extinction and recolonization, founder effects can become a powerful structuring mechanism (Slatkin 1977; Whitlock and McCauley 1999). In this case, population differentiation depends on parameters that influence the frequency and severity of those founder effects, such as extinction and colonization rates as well as the number and source of founding propagules (Slatkin 1977, Wade and McCauley 1988; Whitlock and McCauley 1990). Some insights into the forces influencing population differentiation have been gathered from age-structured populations. For example, elevated population structure (F_{ST}) among newly established demes in metapopulations suggested that founder effect generated the population structure, with gene flow reducing F_{ST} over time (McCauley *et al.* 1995; Giles and Goudet 1997; Ingvarsson and Giles 1999, Fields and Taylor 2014).

Metapopulation dynamics will also influence selection. Most obviously, founder effects reduce gene diversity, increasing inbreeding depression through the exposure of recessive alleles to selection. This may have disastrous consequences unless there is genetic rescue via migration or inter-demic selection (Willi and Fischer 2005, Willi *et al.* 2005, Thrall *et al.* 2008). Alternatively, there may be a more effective purging of deleterious recessives over the longer term (Thrall *et al.* 1998, Whitlock 2002). The balance between these selective forces has received negligible empirical

testing. Spatial structure is also important for selection on more complex traits. In host-pathogen systems for example, dynamics may be asynchronous among demes. Hence, intra-demic selection acting on resistance may interact with inter-demic processes to influence 'finding times' for pathogens and colonization rates of susceptible versus resistant hosts (Frank 1997, O'Keefe and Antonovics 2002, Kerr *et al.* 2006).

Population structure, by redistributing genetic variance among demes, also creates opportunities for inter-demic selection. Multilevel selection in structured populations explains diverse biological phenomena, including the evolution of social behavior (Hamilton 1964; Frank 1998), genetic conflict (Hurst *et al.* 1996, Taylor *et al.* 2002), reduced pathogen virulence (Kerr *et al.* 2006, O'Keefe and Antonovics 2002), sex ratio evolution (McCauley and Taylor 1997, Olson *et al.* 2005), and the origin of multicellularity (Szathmary and Smith 1995, Michod 1997). Fundamental questions remain as to when different forms of multilevel selection drive organic evolution (Manier et al. 2007, Bijma and Wade 2008).

For 28 years, we have studied the regional metapopulation dynamics of the plant, *Silene latifolia*, and its obligate fungal pathogen, *Microbotyrum violaceum*. We have long-term demographic, sex ratio and disease incidence data for more than 800 populations. For the past six years we have used a molecular marker approach to estimate population structure, patterns of inbreeding, gene flow, and the parentage of newly established populations. We have also experimentally examined resistance and virulence structure in this host-pathogen metapopulation.

Our recent genetic data have shown that founder effects are a primary force in generating population structure in this system, and that genetically variable populations contribute disproportionately to the founding of new populations. At a practical level, our initial data prove the concept that the parentage of migrants and colonists can be identified and directly measured in natural populations, and that demes vary in traits that make them contribute differentially to the next generation of migrants and colonists. We have shown that among population variation in disease resistance reflects the past history of pathogen colonization, which in conjunction with theoretical models, shows that genetic structure of host populations contributes to observed regional disease levels.

At a broader level, our initial findings emphasize the fact that most metapopulation genetic models are necessarily simplistic (Pannell and Charlesworth 2000, Pannell and Fields 2014), making assumptions about genetic equilibrium and fixed parameters that are certain to be violated in nature. Rather than focus on how evolutionary forces combine to arrive at some average value of N_e, migration rate or $F_{ST,}$ we propose to focus on the *variation* in structuring mechanisms that occur simultaneously in different populations (Fields and Taylor 2014) and how these will also vary across different regions of the genome. A fundamental objective of the proposed research therefore is to focus on how and why populations, and different regions of the genome, behave differently, and to observe in real time how the interplay of these mechanisms influences population establishment, persistence and local extinction.

B. SPECIFIC AIMS

1) Parameterize the *Silene/Microbotryum* metapopulation ecologically and genetically. Estimate population-specific contributions to population genetic structure, migration rates and the pool of colonists, and use Bayesian statistical methods to associate these with demographic aspects of those populations (age, demographic history, and spatial isolation and interactions among them).
2) Test whether ecological dynamics and population structure interact, perhaps creating opportunities for multi-level selection. Extend the use of genetic data and assignment methods longitudinally to identify whether genetic diversity and spatial structure contribute to the differential growth, reproduction and extinction of populations, and vice-versa.

3) Test whether predicted differences in effective population sizes and rates of migration across the genome (e.g. for autosomes, the X- and Y- chromosomes and organelle genomes) affect how their genetic structure evolves.
4) Estimate migration in both space and time (through seed banks) and investigate how these processes combine to influence metapopulation genetics.
5) Organize and disseminate the demographic, genetic and collections based resources of this project using an integrated and publicly available database.

C. STUDY SPECIES

Silene latifolia (white campion) is a short-lived, dieocious perennial plant, with X-Y sex determination (Baker 1947, Westergaard 1958). The species is native to Europe and is considered invasive in the United States (McNeill 1977, Taylor and Keller 2007). The average life span of a plant that flowers is ~2 yr. The species is easily grown and crossed, producing ~300 seeds per capsule. In the study area, *S. latifolia* is a ruderal species that is largely confined to roadsides. Its roadside distribution allows us to gain rapid access to many populations over a large area (25 km from north to south and 30 km from east to west). Estimates of population genetic structure (McCauley *et al.* 1995; Fields and Taylor 2014) indicate the study region encompasses hundreds of populations, defined as distinct genetic neighborhoods.

Infection of *S. latifolia* by *Microbotryum violaceum* results in anther-smut disease, where anthers produce dark violet fungal spores in place of normal yellow pollen. The pathogen induces female flowers to abort the ovaries and instead produce stamens that bear diseased anthers. The disease is systemic and has large fitness effect, rendering the host effectively sterile. Transmission of the disease is accomplished by pollinators during sexual reproduction by the host. The resulting frequency-dependent transmission and the expression of the disease during reproduction parallels many features of sexually transmitted diseases in other systems (Kaltz 1995, Lockhart *et al.* 1996, Antonovics 2004). The pathogen is easy to culture on petri plates and can be stored indefinitely as sporidial cultures. Crosses are easily made by mixing fungal isolates of opposite mating type on water agar plates.

In Virginia, the host shows within and among population variation for resistance, but there is remarkably little variation in pathogen infectiousness across the species' range. *Microbotryum* does not therefore conform to the classical gene-for-gene model found in several other plant pathogen systems. Instead, resistance behaves as a quantitative trait with a high heritability (Alexander *et al.* 1993). There are also large fitness costs (20-30%) associated with resistance in the absence of the disease; more resistant plants flower later in the season and produce fewer flowers (Alexander 1989; Biere and Antonovics 1996). Populations with a high frequency of resistance are significantly smaller than more susceptible populations (see Initial Data, below).

D: THE *SILENE-MICROBOTRYUM* METAPOPULATION:

The *Silene-Microbotryum* metapopulation has been extensively characterized over the past 28 years, and is now a model for understanding the dynamics of interconnected demes. The system is neither an island model (Wright 1931), nor an idealized metapopulation (Levins 1969); populations vary in size, dispersal is limited, and within population dynamics are important relative to the time scale of the study. Host and pathogen populations are characterized by frequent colonizations and extinctions, on the order of 5-20% per year (Antonovics *et al.* 1994, 1998; Thrall and Antonovics 1995; Antonovics 2004, Section 1.E of this proposal), with smaller populations having a higher probability of extinction (Thrall et al. 1998). Founder effects during colonization enhance genetic differentiation among populations (McCauley *et al.* 1995, Fields and Taylor 2014), with small newly founded populations showing inbreeding depression, especially when isolated from other populations (Richards 2000). The metapopulation dynamics of the *Silene-Microbotryum* system is generalizable to other systems. Hanski (2001) highlighted its similarity to the Glanville fritillary

(*Melitea cinxia*) metapopulation in Finland, and Moody-Weis *et al.* (2008) showed that patterns of colonizations and extinctions scale similarly in sunflowers (*Helianthus annuus*). Resistance costs and genetic specificity are also characteristic of many insect-pathogen systems (Rolff and Siva–Jothy 2004).

The *Silene-Microbotryum* metapopulation has been surveyed in a consistent manner for the past 28 years. Briefly, we count numbers of diseased and healthy plants within contiguous 40m segments of roadsides (Antonovics *et al.* 1994; Fig. 1). The one-dimensional grid segments of 40m include one or two "genetic neighborhoods" (i.e., areas within which genetic exchange is essentially random) as estimated from spore, pollen, and seed dispersal (Alexander 1990) as well as genetic markers (McCauley *et al.* 1995, Fields and Taylor 2014). These neighborhoods may include several distinct patches of *S. latifolia*, or some larger patches may span two or more grid segments. Not all populations of *S. latifolia* occur at roadsides; there are a few so-called "off-road sites" (on average 5.9 % of occupied grid segments) that we include in the annual census because they are potential sources of migrants and colonists. When the field data are pooled across adjacent segments, the patterns of disease incidence are remarkably robust over a grid scale of 40-160m (Antonovics *et al.* 1994). Host and pathogen colonization and extinction rates also scale to patch sizes in a consistent way, indicating that comparative levels of extinction and colonization would not change with segment size (Moody-Weis *et al.* 2008).

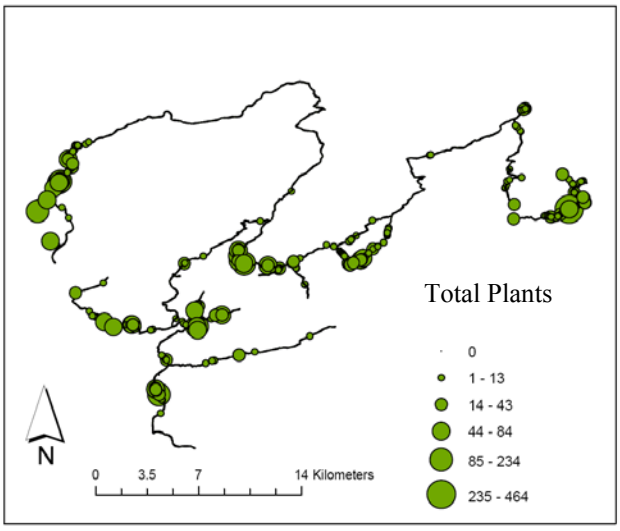

Figure 1. Map of current census route in Giles County, VA. Size of circle indicates population size as of 2006-2007 censuses.

The metapopulation census began in 1988. Data are collected in June, with a re-census carried out in August to verify the extinction/colonization events of the host and/or disease. Our census is rapid; field-work is completed by three crews of three people in a single week. We take pains not to disturb the system by our own activities during the census. Flowers close before midday, necessitating censuses between 5:30 and 11:00 A.M., and again between 7:00 P.M. and dusk, during which time we can determine the disease status visually without touching the plants or trampling on the sites.

The long-term demographic history of the metapopulation emphasizes the non-equilibrium nature of the system. After an initial increase in census population size, the host and disease have declined in abundance over the past two decades, with the disease being close to regional extinction (Figure 2). Historical patterns of patch occupancy suggest that the host decline has been due predominantly extinction rates exceeding colonization rates (especially during the last decade) causing a steady decline in the number of patches occupied by *S. latifolia* (Figure 3).
Patterns of occupancy by the host show that the number of occupied sites asymptotes with an

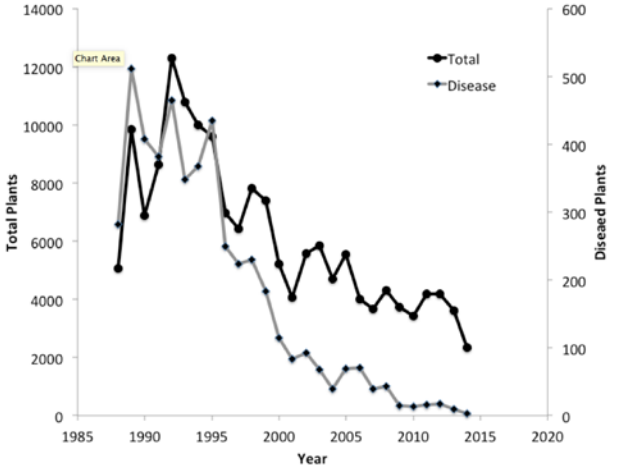

Figure 2. Total number of plants and number of diseased plants in the Silene/Microbotryum metapopulation.

increasing census population size, suggesting there is a maximum number of suitable sites for the host to occupy, with current host populations being well below that level (Figure 4). The data also suggest that the number of patches occupied by the host may be a limiting resource for the pathogen to spread (Figure 5), with the current level of host occupancy (< 300 sites) below what night be necessary for pathogen persistence. Recent demographic trends necessarily influence the future allocation of resources. Specifically we will moderate investment in the future study of pathogen genetics (see Proposed Research). However, this demographic history also highlights the distinct possibility of increased host occupancy leading to rapid host and/or pathogen expansion in the future. The decadal

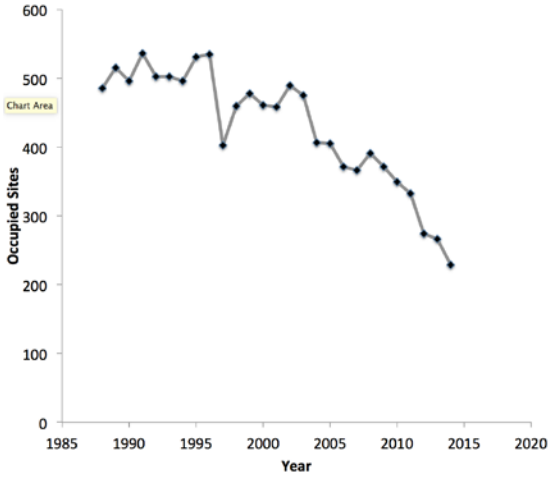

Figure 3. Number of sites occupied by S. latifolia in the Silene/Microbotryum metapopulation.

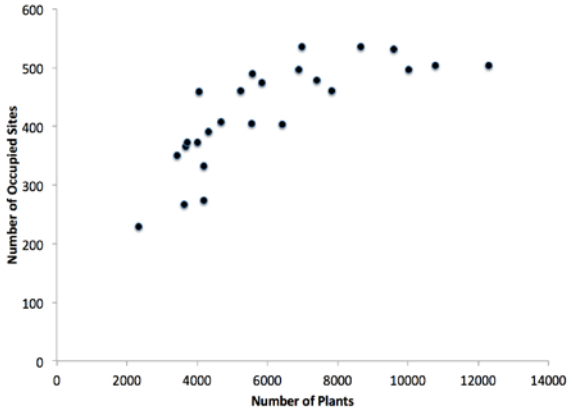

Figure 4. Number of sites occupied versus the total number of plants in the Silene/Microbotryum metapopulation.

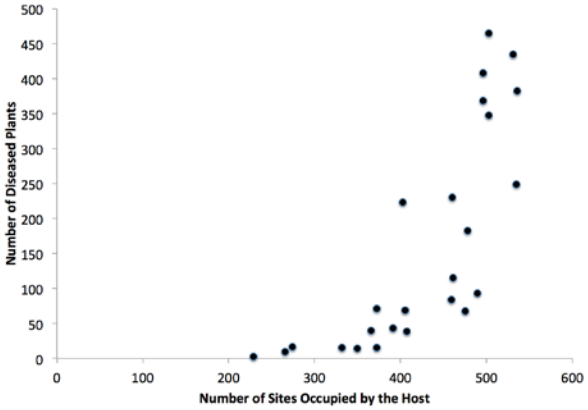

Figure 5. Number of diseased plants versus the total number of sites occupied by S. latifolia in the Silene/Microbotryum metapopulation.

research plan therefore incorporates flexibility to incorporate potentially different demographic scenarios.

The demographic history of the system has important implications for the genetic studies we propose below, and highlights the power of combining demographic and genetic data in our experimental system. For example, the reduction in the number of occupied sites, by definition, increases the distance between the remaining sites. This is predicted to reduce overall migration and enhance genetic differentiation. However, the fact that extinction is non-random, affecting primarily smaller and therefore younger populations (Thrall et al.1998) alters the age structure of the

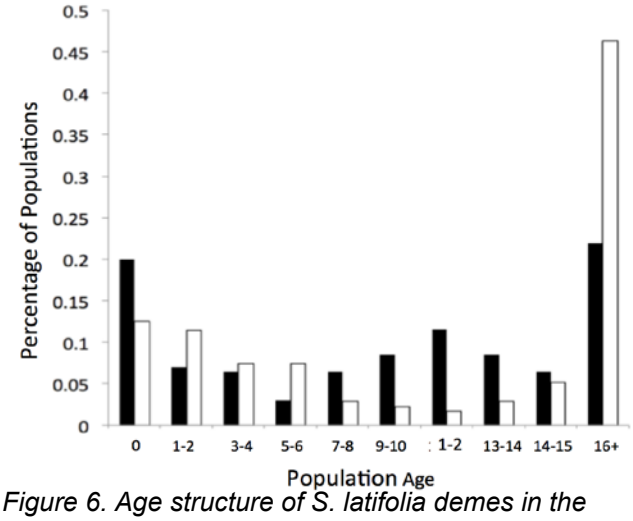

Figure 6. Age structure of S. latifolia demes in the Silene/Microbotryum metapopulation in 2007 (black) and 2014 (white).

metapopulation. The last decade of relatively high extinction rates has shifted the population age structure toward a greater percentage of older populations (Figure 6) that should be less likely to be genetically differentiated (Fields and Taylor 2014). Thus, demographic changes are expected to have different effects on population genetic structure depending on whether founder effects in young populations, or ongoing rates of migration among extant populations, have been the primary determinant of genetic structure.

E. ACHIEVMENTS FROM IN THE FIRST FUNDING PERIOD

For the first 20+ years, the *Silene/Microbotryum* census was conducted mostly in the spirit of volunteerism with expenses primarily out of pocket. The onset of the "modern-era" was precipitated by NSF funding that transformed the annual census from a strictly demographic study to a longitudinal genetic study. Over the past 6 years we have integrated the demographic dataset with population genetic data, intensively sampling leaves, seeds (up to 50 families per population) and spores from all infected plants from three metapopulation sections every two years, including all (>100) newly colonized sites. Plant samples were genotyped at 20 microsatellite loci. All seed collections, fungal samples and DNA samples have been stored in repositories that will be integrated and made public as part of the proposed research. Taken together, these data provided some unique insights into the population genetic consequences of metapopulation structure:

i. Using Hierarchical Bayesian methods (Foll and Gaggiotti 2006), we showed that newly established and/or spatially isolated populations contributed most to population genetic structure. This implicates the overriding importance of metapopulation structure (founder effects and ongoing migration) in driving genetic differentiation (Fields and Taylor 2014)

ii. In collaboration with Oscar Gaggiotti (St. Andrews) and Matthieu Foll (Lausanne), we generalized existing software (GESTE) to allow for variance in molecular marker types and to correct for the effects of unsampled populations on genetic inference (Fields, Gaggiotti and Foll, In Prep.).

iii. By utilizing nuclear and mitochondrial markers from DNA collections made 14 years apart, we showed that contrary to theoretical explorations, cyto-nuclear disequilibrium can be stably maintained. We showed that extinction/colonization dynamics associated with metapopulation structure can continually re-generate LD at neutral loci, suggesting LD-based population genetic signatures previously interpreted as evidence for selection may in fact be the result of neutral, drift-like processes (Fields *et al.* 2014).

iv. We showed that our genetic data can be used to effectively assign parentage to newly established populations. As expected, colonization events depended on distance from their source. R_{ST}/F_{ST} values showed that most dispersal was local (17-43m) but genetic assignments demonstrated substantial long-distance dispersal that would have otherwise gone undetected.

v. By assigning population-level parentage to all new colonists, we were able to estimate, for all potential parental demes, their relative contribution to the pool of colonists. Using Hierarchical Bayesian Methods, we showed that genetically variable populations contributed

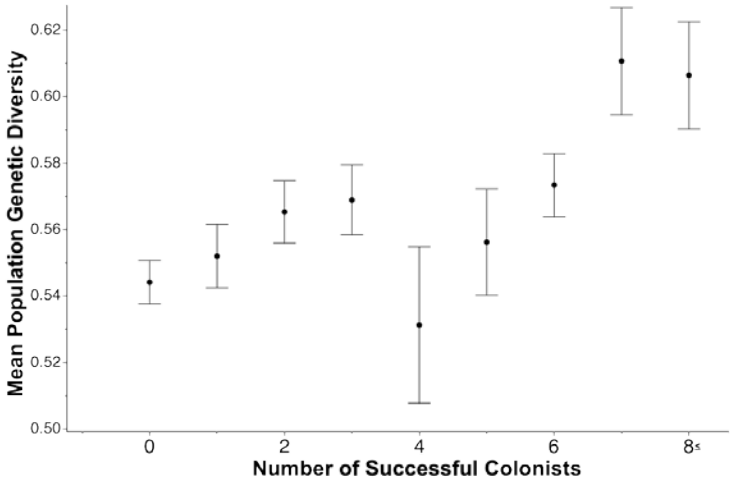

Figure 7. Populations that harbored more genetic diversity contributed disporportionately to the propagules that contributed to the establishment of new populations (colonists).

disproportionately to the founding of new populations (Figure 7). This is the first empirical demonstration of the long held view that genetic diversity contributes to population-level productivity and offers early evidence for the potential importance of interdemic selection.

vi. Our combined genetic and demographic data permit the first full evaluation of the now classic models of metapopulation dynamics (e.g. McCauley and Wade 1988, Whitlock and McCauley 1990), including estimates of extinction, colonization, migration and the number and source of colonizing propagules. We are currently investigating deviations from those models that are likely to be important in nature. For example, because population structure is generated from founder effects, if younger populations also experience higher probabilities of extinction, then F_{ST} should be lower than equilibrium theory predicts. Current theory also assumes that extinction is balanced by colonization so that the number and age structure of demes remains constant. This assumption is clearly violated in this system (Fig. 6), which is predicted to alter the genetic structure as well.

vii. Our combined long-term study of disease ecology and genetic structure showed that the two interact strongly. For example, healthy populations were more genetically differentiated than diseased populations. Moreover, using a large inoculation experiment involving more than 30 populations of known disease and demographic history (2,600 plants) we showed a large cost to resistance, resulting in more resistant populations being on average, less than half the size of susceptible ones (21.6 versus 47.2); since we also showed that size contributed to colonization, this led to an additional "inter-demic" advantage of susceptibility and therefore disease persistence in the metapopulation, as predicted by theory (Antonovics *et al.* 1998, Antonovics 1999). Moreover, host resistance declined significantly with time since the most recent local infection. Hence, the cycle of infection, resistance evolution and the return to susceptibility, is asynchronous among demes, serving to facilitate host pathogen coexistence.

2. DECADAL PLAN RATIONALE

The proposed research will generate 5 years of new data as the first half of a decadal research plan. At the end of decadal plan, we will have 38 years of demographic data and 17 years of intensive sampling, genetic analysis and resource collection and dissemination. It is important to emphasize that this is not an incremental advance. The decadal plan would extend the genetic sampling from 6 to 17 years and allow us to test fundamentally different aspects of the evolution of subdivided populations. Consider extinction and the sources of recolonization - the only way to test for genetic causes of extinction is to accumulate long-term genetic data on populations, then analyze a statistically sufficient number of extinction and recolonization events through time. Thus, although we have observed many local extinction events over the past 28 years, few of those demes were characterized genetically before they perished. Similarly, although we have been successful in identifying the parents of new populations, whether genetics drives establishment success can only be tested by studying new populations over time. Finally, because the metapopulation is not at demographic (or likely, genetic) equilibrium, genetic structure and the underlying causes of population differentiation are expected to shift across time spans that will be covered by our proposed research.

A central objective therefore, and the rationale for a decadal plan, is to accumulate enough observations of population extinction and colonization, including genetic observations prior to extinction and subsequent to colonization, to directly test how metapopulation dynamics affect the genetics of spatially distributed populations (and *vice versa*) through time. Among the specific predictions that we will test are the following: Populations founded by several colonists should have a lower extinction rate than populations founded by fewer colonists; genetic variance should increase as populations age as a result of distance-dependent migration; more genetic differentiation will be seen at sex-linked and cytoplasmic loci due to their lower effective population sizes; seed banks contribute to recolonization and alter the overall magnitude of genetic structure; successful colonization by source populations should increase with the genetic variance of the source.

As with any long-term study, we must be prepared to be opportunistic with respect to emerging technologies and changes in the system itself. For example, if *M. violaceum* were to increase in prevalence, we would take that opportunity to retrospectively study pathogen genetics, or the dynamics of host resistance, during both the decline and subsequent expansion of the disease. With respect to changes in technology, the proposed population genetics research involves the best combination of high throughput and cost-effective genotyping available today, but our research objectives may broaden as genomic technologies allow higher resolution and more cost-effective methods.

3. PROPOSED RESEARCH

A. ONGOING CENSUS AND COLLECTIONS:

The long term metapopulation census and DNA/seed/fungal collection methods have proven to be effective and cost efficient, so we propose to continue them. The demographic census methods are described above.

With respect to fungal collections, we will continue to collect flower buds from all infected plants, germinate the fungal spores within the flowers and maintain the isolates indefinitely as frozen sporidial cultures. With respect to the plant collections, the genetics research has focused on 3 of the 9 census "sections" (sections 2, 6 and 9) where *Silene* is abundant and there is a historically active disease dynamic. We will continue to collect leaves from all plants, and seed capsules from all female plants in these sections every other year. Leaf samples will be collected from so called "colonists" every year. DNA will be isolated using a standardized CTAB extraction (Fields *et al.* 2014) and stored frozen. Seeds will be maintained in dry envelopes at 4 degrees C. Stored this way, seeds retain >90% germination for more than 10 years.

We operationally define "extinction" and "colonization" to maximize the number of samples collected. A colonization event is defined as plants appearing at a site that had been vacant the previous two years. Local extinction is defined as a site that becomes vacant for two years running. We recognize that recolonization could be due to propagules migrating from neighboring sites or emerging from a seed bank (the seed pool has an estimated viability of approximately 4 years (Peroni and Armstrong 2001)). It is also possible plants may be missed from a census because they fail to flower and return after regenerating from rootstock. The longitudinal genetic data will be used to estimate these processes separately by using contemporary populations, as well as plants and populations from the past, as potential sources of colonists.

The collection of seeds and DNA began in 2008. Since that time the three focal sections of the metapopulation averaged 316 occupied sites; each being a potential source population for subsequent colonization and extinction events. Source populations varied considerably in size (~50% had less than 10 plants, ~20% had 10-20 plants, ~30% had >20 plants) with a global average of ~8 plants/site. These numbers change from year to year as colonization events in one year become sources in the next, and as demes go extinct. Since 2008, we observed an average of 32 potential colonizations and 49 extinctions per year (as defined above). A substantial fraction of new colonizations (~50%) go extinct in the first five years. The genetic data accumulated over the period of the decadal plan (and a total study duration of 17 years) will therefore involve well over 500 extinctions and 500 colonizations that will cover a wide range of genetic sources and demographic outcomes. The proposed research involves genotyping all plants in the three focal sections of the metapopulation, every other year, to a limit of 50 individuals per population (~2,000 plants per sample period, plus the genotyping of colonists every year). We will also genotype retrospectively back to 2008. The total for the current funding period therefore will include ~14,000 plants collected over a 12-year period. The decadal plan would extend this to 17 years.

B. FORCES THAT CREATE AND DESTROY POPULATION STRUCTURE AT NEUTRAL LOCI:

A major focus of metapopulation research has been to elucidate the effect of extinction and colonization events on the creation and maintenance of neutral genetic diversity and structure

(Slatkin 1977, Maruyama and Kimura 1980, Whitlock and McCauley 1990, Gilpin 1991, McCauley 1991, 1994). While the precise details of the colonization and gene flow may take a number of forms, the central model parameters are migration (*m*), extinction (*e*) and colonization (*c*) rates, the number of founding propagules (*k*), and the fraction of demes from which the colonists come (Φ) (Wade and McCauley 1988; Whitlock and McCauley 1990). Each parameter has been shown to have a substantive effect on the genetic diversity distributed within and among demes, as well as on the potential outcome of evolution (Pannell and Charlesworth 2000; Pannell, 2003).

Previous work has captured the signature of metapopulation effects through an analysis of F_{ST} in age-structured demes (McCauley *et al.* 1995; Giles and Goudet 1997; Ingvarsson 1998, Fields and Taylor 2014). The first of these studies (McCauley *et al.* 1995) was done in the *Silene* metapopulation, and showed an elevated F_{ST} in newly colonized demes, consistent with the model where founder effect creates population structure, with gene flow eroding that initial genetic structure as demes age (i.e., high *e* and *c*, moderate *m*, but low *k* and Φ).

The set of parameters that combine to drive the evolution of population structure, however, have never been simultaneously estimated, much less in a natural metapopulation using long term ecological data. Some parameters such as extinction/colonization rates (*e* and *c*) are estimated directly from the census data. The remaining parameters make use of the population genetic data. The number of founding propagules (*k*), the fraction of demes from which the colonists come (Φ) and direct estimates of contemporary migration (*m*) will be estimated using assignment tests to trace the deme-specific ancestry of individuals. Estimates will be derived using the software GENECLASS2 (Piry *et al.* 2004) and additional custom pipeline required to deal with larger, genomic scale datasets.

The 30+ year history describing the age structure of demes (Fig 2), overlaid with the genetic data, will provide a powerful analysis of whether overall population structure is generated during colonization and diminished by subsequent gene flow. The 10-year period of funding also provides an unprecedented opportunity to collect predictive data, and observe these processes directly. We will test whether founder effects over successive years generate population structure during colonization, and measure whether and how fast gene flow homogenizes demes over time, as assayed by the increasing number of parental demes that contribute to the genetic composition of demes as they age. The data collected from this project, therefore, will result in a detailed understanding of the metapopulation-level processes that create and destroy population structure in nature.

C. THE EFFECT DEMOGRAPHIC HETEROGENEITY ON POPULATION STRUCTURE:

In addition to estimating regional metapopulation parameters, the proposed research focuses on the *variation* in structuring mechanisms that occur simultaneously in different populations (Fields and Taylor 2014). We will estimate how parameters such as genetic variance and contributions to F_{ST} vary among populations, and relate these to spatial and temporal characteristics of those populations using Bayesian statistical methods. Integrating the demographic and genetic data over time will allow us to examine how founder effects and gene flow are influenced by the size and distance of demes that contribute the colonists and migrants and how the characteristics of populations combine to affect colonization and population persistence

Summary Statistics: We will estimate global summary statistics of population structure, e.g. observed (H_O) and expected (H_E) heterozygosity, rareified allelic richness, $F_{ST,}$ Jost's *D* using programs such as GenoDive (Miermans and van Tienderen 2004), or the hierfstat v. 0.04-10 (Goudet, 2005) package in R (R Core Development Team, 2011). We will estimate population level biparental inbreeding using the software RMES (David *et al.* 2007), and partition inbreeding effects into individual (individual homozygosity) and group (F_{IS} and Identity Disequilibria) components.

Population-specific F_{ST}: Following Fields and Taylor (2014), we will use the hierarchical Bayesian method of Foll and Gaggiotti (2006), implemented in the program GESTE v. 3, to evaluate the effect of spatial and temporal characteristics of populations on the magnitude of genetic differentiation among populations. The method estimates F_{ST} values for each sampled population using the approach first proposed by Balding and Nichols (1995) and relates them to demographic

and environmental factors using a generalized linear model framework (Gaggiotti et al. 2009). We will make use of Bayesian methods implemented in Stan (Stan Development Team 2015) to test *a priori* hypotheses about the relationship between population genetics and other demographic and spatial characteristics of populations. How do population age and degree of connectivity affect population structure? How do genetic variability, demographic history or spatial isolation combine to affect population persistence?

Genetic source analysis: Some of the most important objectives of the proposed research rely on estimating the source of propagules or genes. These methods form the basis of identifying the source and number of colonists as well as direct estimates of migration rates across the genome.

Genetic mixture analysis, or genetic stock identification (GSI), was developed in the context of identification of sources of salmon stocks (Smouse *et al.* 1990). This framework, subsequently extended by Gaggiotti *et al.* (2004) and Faubet and Gaggiotti (2008), will be applied to the multilocus genetic data to identify the population sources of individuals within the newly arisen populations. Specifically, given the multilocus genetic data and *i* possible source populations, we will (1) estimate the proportion x_i that each source *i* contributed to the newly colonized demes and migration between demes, and (2) identify the n (biotic/abiotic) factors, G_1, …, G_n, that contributed to population fecundity via colonization success and migration by including them as elements in the models (Gaggiotti *et al.* 2004; Faubet and Gaggiotti 2008).

A likelihood function relates genetic linkage disequilibrium information present in populations to that in admixed populations and is constructed by defining the probability of observing an individual genotype based upon the genetic characters of potential source populations (Gaggiotti *et al.* 2002; Manel *et al.* 2005; Gaggiotti *et al.* 2004). A Bayesian analysis, using a Markov Chain Monte Carlo (MCMC) technique, is used to directly estimate mixture proportions and the posterior source probabilities for each individual (Manel *et al.* 2005; Excoffier and Heckel 2007).

Gagiotti et al. (2004) extended the basic GSI model by utilizing a Bayesian hierarchical modeling approach to incorporate genetic and environmental data into a single joint analysis. Specifically, hierarchical Bayesian methods model and estimate the effect of a given environmental factor on parameters of interest, e.g. composition of colonizing or migrant group (Foll and Gaggiotti 2005; Faubet and Gaggiotti 2008).

We will use a customized analysis pipeline extending the COLONISE algorithm in Stan to identify the sources of colonists (Foll and Gaggiotti 2005). Migration will be estimated using the software BIMr (Faubet and Gaggiotti 2008; both are available at http://www-leca.ujf-grenoble.fr/logiciels.htm). Population level selection through differential colonization and migration is tested statistically using a Reversible Jump MCMC (Green 1995) based criterion. We will compare the posterior probabilities of competing models that include factors that account for the effects of population structure, i.e. F_{IS} and other measures of biparental inbreeding, demographic data (population size, growth rate, sex ratio, previous disease status), and environmental data (distance from possible source to new population), and their possible interactions. In this way, we will be able to weight our parameter estimates, and therefore our conclusions, based upon both parameter uncertainty (High Density Posterior Intervals) and model uncertainty (Gaggiotti *et al.* 2004; *see* Gaggiotti *et al.* 2002 for the use of Deviance Information Criterion to select best fit models). As genetic data accumulate over the award period, therefore, we will estimate how the introduction of genetic variance (via gene flow) affects the persistence and fecundity of spatially distributed populations (Ingvarsson and Whitlock 2000; Whitlock *et al.* 2000).

D. THE EFFECT OF GENOMIC HETEROGENEITY ON POPULATION STRUCTURE:

In our initial research, we showed that the magnitude of population differentiation varied depending on whether molecular markers were putatively neutral or linked to expressed genes (ESTs), likely reflecting differences in the relative importance of selection, migration and drift across loci (Fields *et al.* In Prep). The promise of population genomics is that saturated marker approaches allow the study genes or genome regions individually, with variance among those regions providing evidence for different evolutionary forces acting in different regions.

We will use second-generation sequencing, combined with advanced analytical techniques, to qualitatively increase simple sequence repeat (SSRs) and single nucleotide polymorphism (SNP) marker coverage (Wang and Hey 2010; Cao *et al.* 2011). Genomic resources for the *Silene* system are rapidly expanding, and the PIs have been directly involved in these efforts. Reference genomes are becoming available for *S. latifolia* and the closely related species, *S. noctiflora* and *S. vulgaris* (through collaboration among D. Filatov (Oxford), D.B. Sloan (Colorado State) and D.R. Taylor (University of Virginia)). The genomes range from ~1 Gb (*S. vulgaris*) to ~3 Gb (*S. latifolia*). Assembled and annotated transcriptomes from 7 *Silene* species (plus an outgroup) have been generated and made public by D.R. Taylor (http://silenegenomics.biology.virginia.edu/), and we have abundant comparative data on the complex organelle genomes in the genus *Silene* (e.g. Sloan *et al.* 2012a, 2012b), also made public on Genbank.

The nuclear genomes are large and complex with currently fragmented assemblies. We are currently combining long-read sequencing platforms (PacBio) and genetic mapping to obtain increasingly reliable positional information for our scaffolds. As expected, new technologies are also emerging. For example, D. Filatov is working with BioNano (http://www.bionanogenomics.com/) to generate high-resolution physical maps that would greatly simplify the assembly of the *S. latifolia* genome (D. Filatov, pers. comm.). As a result, we anticipate continual progress toward increasingly polished reference genomes during the award period.

Given the upcoming availability of SNPs that will be increasingly assigned to chromosome position, we will study (i) how metapopulation structure affects different genome components (sex-linked, Y linked, autosomal regions, organelles) and (ii) begin to generate the data to map focal genotypes to phenotypes (e.g. host resistance, sex ratio distorters, genetic load, Pannell and Fields (2014)). We propose to use double digest RADseq (ddRADseq; Peterson *et al.* 2012). ddRADseq is the most attractive option in this system for several reasons. Whole genome re-sequencing, even pooled sequencing, is still prohibitively expensive for the sample sizes proposed here, and many of the benefits of those methods are diminished by the currently fragmented state of the genome assembly. Sequencing a subset of the genome can generate thousands of markers for thousands of individuals at reasonable costs, and the number of loci (and coverage/locus) can be tuned to the genome of interest. Because we are sampling among local populations of a single species, we anticipate ddRADseq-generated markers to be repeatable through time and across populations. Also, we can generate, with only marginally more effort, a saturated linkage map that would contribute to improving the *S. latifolia* genome assembly.

Based on our genomic data and a recently published RADseq linkage map in *S. latifolia* (Qui *et al.* 2015), we expect to be able to genotype our entire collection at ~10,000 bi-allelic SNPs within a modest budget. When combined with RAD-based linkage maps and an improved reference genome, these should generate excellent resolution for studying population differentiation in different regions of the genome. Briefly, we used the NCBI housed short read archive (SRA) samples provided by Qui *et al.* (2015) to estimate that even for their lowest coverage samples, ~60Mb of data was sufficient to achieve 100x coverage and detect ~6,500 bi-allelic SNPs (this using a standard protocol with an *Sbf1* digest). We expect to detect more SNPs by using higher coverage and mapping reads to an assembled genome. In addition, we propose to explore some alternative digests using a combination of *in silico* experiments using the reference genome and preliminary MiSeq runs. We estimate that ~100Mb per individual will generate ~10,000 bi-allelic SNPs, amounting to ~1.4Tb of data (or ~12 lanes on an Illumina HiSeq 2500 v4 kit). Using Peterson *et al.*'s (2012) estimate of 192 individuals multiplexed per Illumina library, our samples would require 73 Illumina libraries. When possible, libraries will be distributed across lanes to minimize bias generated by lane-to-lane variation. Libraries and sequencing runs can fail, so our budget includes a moderate excess in libraries and sequencing capacity to accomplish our aims. Saturated marker data will substantially advance the objectives of the proposed research, permitting a more detailed analysis of how founder effects, migration and selection combine to act differently across the genome.

In order to obtain high confidence genotypes from ddRADseq data we will follow the methods described in Fields *et al.* (in press). Briefly, reads will be aligned to the *S. latifolia* genome using BWA-MEM (Li 2013), SAM files will be converted to BAM files, sorted, and indexed using

SAMtools (Li *et al.* 2009). SNP polymorphisms will be identified using GATK (McKenna *et al.* 2010) for base quality score recalibration, indel realignment, using standard hard filtering parameters or variant quality score recalibration according to GATK Best Practices recommendations (DePristo *et al*. 2011; Van der Auwera *et al.* 2013). Only SNP polymorphisms with parameters QD (quality-by-depth) ≥ 6 and GQ (genotype quality) ≥ 20 and for which genotype information is available for >95% of genotyped individuals will be retained for downstream analyses. SNPs will be characterized using SnpEff, as well as custom pipelines, in order to generate functional 'bins' (Cingolani *et al.* 2012). These bins will be analyzed using the hierarchical structure of methods developed by Foll and Gaggiotti (2006, 2008). These methods allow an additional hyperprior (e.g. on the $\theta_{i,j}$ parameter in GESTE v.3 (Fields *et al.* in prep)), which can estimate the relative effects of individual markers, or bins of markers, on driving variation in F_{ST}. We used this approach to compare microsatellite markers within coding regions to those in non-coding regions, and showed that microsatellites in non-coding regions had effectively higher migration rates (Fields and Taylor in prep). Bins of SNPs in different regions of the genome can be similarly modeled as to how they are differentially affected by metapopulation processes such as patterns of migration/gene-flow, colonization, and extinction.

E. DISPERSAL THROUGH TIME.

Many organisms, especially plants, reproduce and disperse through time and space via a dormant phase of their life cycle. In plants, seed banks are thought to offer a temporal escape from unpredictable and unfavorable environments (Venable and Brown 1988). This can particularly important in weedy species, where patches may be subject to frequent disturbance, extinction and recolonization events. Although these dynamics are one reason the metapopulation concept is thought to be so important for understanding the ecology and evolution of many plant populations (McCauley *et al.* 1995, Husband and Barrett 1998, Fields and Pannell 2014), the very concept of a metapopulation may not be as relevant for species where recolonization through seed banks is common (Freckleton and Watkinson 2002). Seed banks will also complicate the genetics of metapopulations because migration through time will almost certainly influence the number and genetic source of colonists, and hence population differentiation (McCauley 2014, Falahati-Anbaran *et al.* 2014). The impact of seed banks on metapopulation dynamics and population structure requires a longitudinal genetic data set where migration from neighboring demes can be distinguished from migration from seed banks by including past populations, even extinct ones, as potential source of colonists. For example, Falahati-Anbaran *et al.* (2014) demonstrated recruitment by both routes using SNP genotyping of *Arabidopsis* seeds in plants and soil over a five-year period.

The *Silene/Microbotryum* metapopulation data combines longitudinal-genetic sampling and genetic source identification, with replicated demographic histories of extinction and recolonization. Understanding the relative importance of migration in time versus space, and their effects on

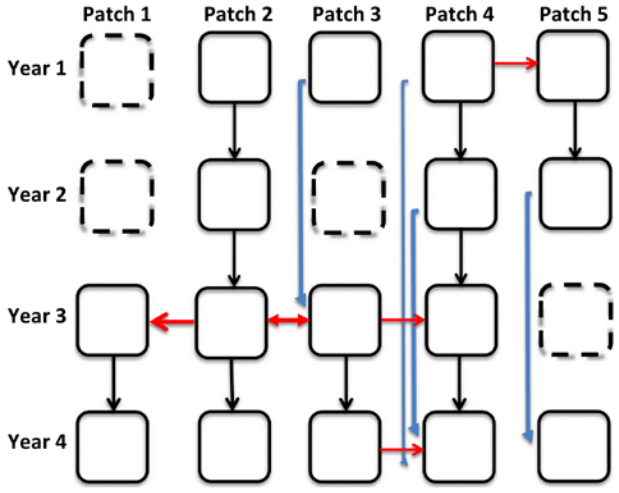

Figure 8. Gene dispersal in time and space. Five patches are either occupied (solid rectangles) or unoccupied (dashed rectangles) over a 4-yr period. Migration or colonization can occur in time, such as being seeds germinating from a seed bank (blue arrows), or in space via seed movement (red arrows). For selectively neutral variation the genetic structure of the five populations extant in Year 4 is a function of recruitment from the seed bank, genetic drift, gene flow, and founder effects associated with colonization events. Figure from McCauley (2014), used with author's permission.

metapopulation genetics is therefore a natural by-product of our ongoing research (see McCauley 2014, Figure 8). We propose to expand our analyses during the course of the study, using genetic assignment tests to extend the pool of source populations increasingly further back in time, and evaluate how different sources of migration contribute to genetic structure. In *Silene*, seeds are known to persist in the soil for up to four years (Peroni and Armstrong 2001), and we might expect the seed bank to contribute more to recolonization events observed shortly after local extinction. Although we know that F_{ST} is higher in newly colonized populations within the *S. latifolia* metapopulation (Fields and Taylor 2014), this does not exclude the possibility of migration from the seed bank, which may also involve a founder effect. What is the importance of this founder effect and is there really a genetic difference between these two routes of colonization? These analyses promise to gather unique insights into the underlying processes determining the nature of this metapopulation and evolution of population genetic structure, and demonstrate the synergistic effects of combining demographic and genetic data over an extended period of time.

F. DISEASE DYNAMICS

For the fungal pathogen, *Microbotryum violaceum*, there is now a completely assembled genome using PacBio technology that has confirmed many of the fascinating genetic features of this pathogen, including a remarkably rapid and complex pattern of structural genome evolution as well as patterns of heterozygosity maintained through its unusual breeding system (Badouin *et al.* 2015). Importantly, this genome assembly will enable the complete characterization of the genomes of not only the ~120 individuals sampled in the past five years, but also individuals sampled from the same valleys 15-20 years previously (spore samples are available from all populations). The major goals could therefore parallel those of the research on the host species, and ask how the genetic structure of different genomic compartments change over time, and as a consequence of colonization and changes in population size.

However, *M. violaceum* has been steadily declining in the study area to the point where it is on the verge of regional extinction. This is important in its own right, and we will continue to monitor this process, maintain collections and study past dynamics. However, we cannot in good conscience propose a specific research plan for the pathogen that extends genetic studies another decade into the future. Therefore, we have not budgeted new experiments into the biology of the pathogen. However, it is important to emphasize that the host has experienced regional-scale demographic fluctuations over several decades, and that pathogen abundance is correlated with host abundance during that time. We will therefore be prepared to shift allocations if there are opportunities to directly observe and analyze a new epidemic of *M. violaceum*.

G. ROLES OF THE PIs:

Antonovics was instrumental in starting and developing the metapopulation census. Taylor joined the effort in 1989, and expanded the census to include sex ratio data in 1992. Peter Fields was instrumental in incorporating population genetics into the metapopulation data set during his dissertation research with Taylor. Over the years, undergraduates, graduate students, post-docs and collaborators of Taylor and Antonovics have helped continue the census and made effective use of the data. For the proposed research, Taylor will be responsible for the plant collections, molecular work and database management while Antonovics will maintain the fungal cultures. Both will contribute to organizing and conducting the ongoing annual census. Peter Fields is currently a post-doc at the University Basel, and though the current award would not fund his salary, he maintains a long-term interest in the development of this system. Dr. Fields will focus on the bioinformatics and statistical genetics analyses.

Students and post-docs will collaborate with the PIs, perpetuating the community of scientists that have made the census a long-term success. Peter Fields (former student with Taylor) is involved with the genotyping efforts and developing the statistical genetics analyses, while Brian Sanderson (student with UVA Colleague, Edmund Brodie III) is developing the relational database. Collaborators such as Stephen Keller (University of Vermont), Michael Hood (Amherst), David McCauley (Vanderbilt U.), Matthew Olson (Alaska, Fairbanks), Daniel Sloan (Colorado State U.) and

Pete Thrall (CSIRO) who have contributed to the census, will benefit from the spatial/historical information in the dataset.

4. BROADER IMPACTS OF THE PROPOSED RESEARCH

A. SCIENTIFIC IMPACTS

The proposed research addresses a fundamental issue in biology of how the process of evolution is affected by the reality that populations are distributed in space. Theoretical treatments of evolution in metapopulations are far beyond experimental studies because of the size and long-term nature of the datasets that are required. These questions can be addressed in the *Silene/Microbotryum* system because we have sufficiently numerous, spatially defined populations with known histories. We have a record of past accomplishments and have developed a distinct and efficient research plan to genetically define the system in the next decade of study.

Understanding evolutionary processes in metapopulations is of practical and applied importance for management professionals in the fields of conservation biology, invasion biology, and for understanding responses to habitat fragmentation, emerging infectious disease and climate change. The *Silene-Microbotryum* system is a model system for studying the impact of disease, including sexually transmitted disease, on natural and human populations. The proposed research will make publicly available a relational database of the entire project, an annotated collection of living tissue and DNA samples through time, links to genomic resources and data downloads, plus new software for genetic data analysis.

B. IMPACTS ON MENTORING AND TRAINING.

The proposed research will educate and train undergraduate students, graduate students and post-doctoral associates in the timely fields of molecular and statistical genetics, genomics and bioinformatics. Post-doctoral mentorship will include participation in grant writing, publication strategies to enhance academic competitiveness, and direct experience supervising students. Post-doctoral associates will be given the opportunity to take aspects of the research program with them to their own faculty position. The PIs have supervised more than 25 post-doctoral associates with more than 20 in tenure-track faculty positions. We will expand undergraduate involvement via REU supplements and the Mountain Lake Biological Station REU program. In addition, we will continue to supervise independent research projects, and to encourage graduate students to spin off their own projects. We regularly incorporate projects related to the *Silene* system into our undergraduate laboratory instruction, with students collecting primary data and preliminary analysis. For example, we developed a lab where students use molecular markers to study a zone of admixture among *Silene* lineages near Mountain Lake Biological Station. Our labs, which seamlessly integrate our research with teaching evolution, are published as curriculum developments in journals such as Chronicle of Higher Education, American Biology Teacher, and Bioscene (Vondrasek *et al.* 2004).

To attract traditionally underrepresented minorities into our research programs, we are working with UVA's office of African American Affairs to instantiate a collaborative relationship with Virginia's Historically Black Colleges and Universities (HBCUs). UVA has the highest graduation rate of minority students of any public university in the country, has well established mentoring programs, and is a participating institution in the Leadership Alliance (http://www.theleadershipalliance.org). The structure of the program includes paid summer research, followed up by research activities at the home institution. It is being incorporated into all our extramural funding proposals, and could serve as a broader institutional model for promoting diversity in the sciences. The advantages of this model are that students at HBCUs can be trained as scientists and become engaged in ongoing projects, gaining the full experience of scientific research (field experiments, molecular biology, bioinformatics, and possibly international travel and scientific collaboration). Students are also likely to build a relationship with UVA, enhancing our ability to recruit minority students that have demonstrated potential into our graduate program.

LITERATURE CITED

*Abbate, J. L., and Antonovics, J. (2014). Elevational disease distribution in a natural plant-pathogen system: insights from changes across host populations and climate. *Oikos* 123: 1126-11136.

Aizen MA, Feinsinger P (1994) Forest Fragmentation, Pollination, and Plant Reproduction in a Chaco Dry Forest, Argentina. *Ecology* 75, 330-351.

Alexander HM (1989) An experimental field-study of anther-smut disease of *Silene alba* caused by *Ustilago-Violacea* - genotypic variation and disease incidence. *Evolution* 43, 835-847.

Alexander HM (1990) Epidemiology of anther-smut infection of *Silene alba* caused by *Ustilago violacea*: patterns of spore deposition and disease incidence. *Journal of Ecology* 78, 166-179.

Alexander HM, Antonovics J, Kelly AW (1993) Genotypic variation in plant disease resistance--physiological resistance in relation to field disease transmission. *Journal of Ecology* 81, 325-333.

*Alexander, H. M., Foster, B. L., Ballantyne, F., Collins, C.D., Antonovics, J., Holt, R.D. (2012). Metapopulations and metacommunities: combining spatial and temporal perspectives in plant ecology. *Journal of Ecology* 100: 88-103.

Antonovics J (1999) Pathogens and plant population dynamics: the effects of resistance genes on numbers and distribution. In: *Ecological effects of pest resistance genes in managed ecosystems* (eds. Traynor PM, Westwood HH), pp. 49-55. Information Systems for Biotechnology, Blacksburg, VA.

Antonovics J (2004) Long term study of a plant-pathogen metapopulation. In: *Ecology, genetics and evolution of metapopulations* (eds. Hanski I, Gaggiotti O), pp. 471-488. Academic Press.

Antonovics J, Abrams JY (2004) Intratetrad mating and the evolution of linkage relationships. *Evolution* 58, 702-709.

Antonovics J, Hood M, Partain J (2002) The ecology and genetics of a host shift: Microbotryum as a model system. *American Naturalist* 160, S40-S53.

Antonovics J, Hood ME, Thrall H, Abrams JY, Duthie GM (2003) Herbarium studies on the distribution of anther-smut fungus (Microbotryum violaceum) and Silene species (Caryophyllaceae) in the eastern United States. *American Journal of Botany* 90, 1522-1531.

Antonovics J, Thrall P, Jarosz A, Stratton D (1994) Ecological genetics of metapopulations: the *Silene-Ustilago* plant-pathogen system. In: *Ecological genetics* (ed. Real L), pp. 146-170. Princeton University Press, NJ.

Antonovics J, Thrall PH, Jarosz AM (1998) Genetics and the spatial ecology of species interactions: the *Silene-Ustilago* system. In: *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions* (eds. Tilman D, Kareiva P), pp. 158-180. Princeton University Press.

*Antonovics J, Boots M, Ebert D, Koskella B, Poss M, Sadd BM. (2012). The origin of specificity by means of natural selection: evolved and non-host resistance in host-pathogen systems. *Evolution* 67: 1-9.

*Antonovics, J., Bergmann, J., Hempel, S., Verbruggen, E., Veresoglou, S., Rillig, M. (2015) The evolution of mutualism from reciprocal parasitism: more ecological clothes for the Prisoner's Dilemma. *Evolutionary Ecology* (in press).

Aspi J, Jäkäläniemi A, Tuomi J, Siikamäki P (2003) Multilevel phenotypic selection on morphological characters in a metapopulation of *Silene tatarica*. *Evolution* 57, 509-517.

Badouin H, Hood ME, Gouzy J*, et al.* (2015) Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. *Genetics*.

Baker HG (1947) Biological flora of the British Isles: *Melandrium* (Roehling em.) Fries. *Journal of Ecology* 35, 271-274.
*Baker, C., and Antonovics, J. (2012). Evolutionary determinants of genetic variation in susceptibility to infectious diseases in humans *PLoS ONE* 7(1) e29089 1-9.
Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96, 3-12.
Barr C, Neiman M, Taylor D (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* 168, 39-50.
Barr CM, Keller SR, Ingvarsson PK, Sloan DB, Taylor DR (2007) Variation in mutation rate and polymorphism among mitochondrial genes of Silene vulgaris. *Molecular Biology and Evolution* 24, 1783-1791.
Bernasconi G, Ashman TL, Birkhead TR*, et al.* (2004) Evolutionary ecology of the prezygotic stage. *Science* 303, 971-975.
*Best, A., Webb, S., Antonovics, J. and Boots, M. (2012). Local transmission processes and disease-driven host extinctions. *Theoretical Ecology* 5: 211-217.
Biere A, Antonovics J (1996) Sex-specific costs of resistance to the fungal pathogen *Ustilago violacea* (*Microbotryum violaceum*) in *Silene alba*. *Evolution* 50, 1098-1110.
Bijma P, Muir WA, Van Arendonk JAM (2007) Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* 175, 277-288.
Bijma P, Wade MJ (2008) The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *Journal of Evolution Biology* 21, 1175-1188.
*Borer, E. T, Antonovics, J., Kinkel, L. L., Hudson, P. J., Daszak, P., Ferrari, M. J., Garrett, K.A., Colin R. Parrish, C. R., Read, A. F., Rizzo, D. M. (2011). Bridging taxonomic and disciplinary divides in infectious disease. *EcoHealth* (DOI: 10.1007/s10393-011-0718-6).
Brodie III ED, Moore AJ, Janzen FJ (1995) Visualizing and quantifying natural selection. *Trends in Ecology & Evolution* 10, 313-318.
Brown JKM (2003) A cost of disease resistance: paradigm or peculiarity? *Trends in Genetics* 19, 667-671.
*Bruns, E., Hood, M.E., Antonovics, J. (2014). Rate of resistance evolution and polymorphism in long- and short-lived hosts. *Evolution* 69: 551-560.
*Bruns, E., Hood, M.E., Antonovics, J. 2015. Demography, dynamics and disease transmission in a population of Dianthus pavonius, an alpine carnation, heavily diseased by anther smut, Microbotryum sp. *Journal of Ecology*. (In revision)
Cafuir L, Antonovics J, Hood ME (2007) Tissue culture and quantification of individual-level resistance to anther-smut disease in Silene vulgaris. *International Journal of Plant Sciences* 168, 415-419.
Cao J, Schneeberger K, Ossowski S*, et al.* (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* 43, 956-U960.
*Chung E, Petit E, Antonovics J, Pedersen A, Hood M E. (2012). Variation in resistance to multiple pathogen species: anther-smuts of Silene uniflora. *Ecology and Evolution* 2: 2304-2314.
Church S, Taylor D (2002) The evolution of reproductive isolation in spatially structured populations. *Evolution* 56, 1859-1862.
Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003) Evidence for multiple pleistocene refugia in the postglacial expansion of the eastern tiger salamander, Ambystoma tigrinum tigrinum. *Evolution* 57, 372-383.
Cingolani P, Platts A, Wang LL*, et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly* 6, 80-92.
Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153, 1989-2000.

David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* 16, 2474-2487.
Delph LF, Knapczyk FN, Taylor DR (2002) Among-population variation and correlations in sexually dimorphic traits of *Silene latifolia*. *Journal of Evolutionary Biology* 15, 1011-1020.
DePristo MA, Banks E, Poplin R*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43, 491-498.
Donohue K (2003) The influence of neighbor relatedness on multilevel selection in the Great Lakes sea rocket. *American Naturalist* 162, 77-92.
Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. In: *Nature Reviews Genetics*, pp. 745-758.
Falahati-Anbaran M, Lundemo S, Stenoien HK (2014) Seed dispersal in time can counteract the effect of gene flow between natural populations of Arabidopsis thaliana. *New Phytologist* 202, 1043-1054.
Faubet P, Gaggiotti OE (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* 178, 1491-1504.
*Fenton A, Antonovics J, Brockhurst MA. 2012. Two-step infection processes can lead to coevolution between functionally independent infection and resistance pathways. *Evolution* 66: 2030-241.
Ferrari MJ, Bjornstad ON, Partain JL, Antonovics J (2006) A gravity model for the spread of a pollinator-borne plant pathogen. *American Naturalist* 168, 294-303.
Fields PD, Taylor DR (2014) Determinants of Genetic Structure in a Nonequilibrium Metapopulation of the Plant *Silene latifolia*. *PLoS ONE* 9, e104575.
Fields PD, McCauley DE, McAssey EV, Taylor DR (2014) Patterns of cyto-nuclear linkage disequilibrium in Silene latifolia: genomic heterogeneity and temporal stability. *Heredity* 112, 99-104.
Fields PD, Reisser C, Dukić M, Haag CH, and Ebert D (2015) Genes mirror geography in *Daphnia magna*. *Molecular Ecology*. (In Press).
Foll M, Gaggiotti OE (2005) COLONISE: a computer program to study colonization processes in metapopulations. *Molecular Ecology Notes* 5, 705-707.
Foll M, Gaggiotti OE (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* 174, 875-891.
Foll M, Gaggiotti OE (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180, 977-993.
Frank SA (1997) Spatial processes in host-parasite genetics. In: *Metapopulation Biology: Ecology, Genetics and Evolution* (eds. Hanski I, Gilpin M), pp. 325-352. Academic Press.
Frank SA (1998) *Foundations of social evolution* Princeton University Press, Princeton.
Freckleton RP, Watkinson AR (2002) Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Journal of Ecology* 90, 419-434.
Freedberg S, Taylor DR (2007) Sex ratio variance and the maintenance of environmental sex determination. *Journal of Evolutionary Biology* 20, 213-220.
Gaggiotti OE, Brooks SP, Amos W, Harwood J (2004) Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Molecular Ecology* 13, 811-825.
Gaggiotti OE, Jones F, Lee WM*, et al.* (2002) Patterns of colonization in a metapopulation of grey seals. *Nature* 416, 424-427.
Gaggiotti OE, Bekkevold D, Jørgensen HBH*, et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic Herring as a case study. *Evolution* 63, 2939-2951.
Gandon S (2002) Local adaptation and the geometry of host-parasite coevolution. *Ecology Letters* 5, 246-256.

Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data. *Genetics* 176, 1635-1651.
Giles B, Goudet J (1997) Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *American Naturalist* 149, 507-526.
Gilpin M (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* 42, 165-175.
Giraud T, Yockteng R, Marthey S*, et al.* (2008) Isolation of 60 polymorphic microsatellite loci in EST libraries of four sibling species of the phytopathogenic fungal complex Microbotryum. *Mol Ecol Resour* 8, 387-392.
Goodnight CJ, Schwartz JM, Stevens L (1992) Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *The American Naturalist* 140, 743-761.
Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5, 184-186.
Green P (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711-732.
Hamilton W (1996) *Narrow Roads of Gene Land: Volume 1: Evolution of Social Behaviour: The Collected Papers of W.D.Hamilton* Freeman/Spectrum, New York.
Hamilton WD (1964) The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7, 1-16.
Hanski I (2001) Spatially realistic models of metapopulation dynamics and their implications for ecological, genetic and evolutionary processes. In: *Integrating ecology and evolution in a spatial context* (eds. Silvertown J, Antonovics J), pp. 139-156. Blackwell Science, Oxford.
Hanski I, Gaggiotti O (2004) Metapopulation biology: past, present and future. In: *Metapopulation Ecology, Genetics and Evolution* (eds. Hanski I, Gaggiotti O), pp. 3-22. San Diego Academic Press.
Hanski I, Simberloff D (1997) The metapopulation approach, its history, conceptual domain and application to conservation. In: *Metapopulation Biology: Ecology, Genetics and Evolution* (eds. Hanski I, Gilpin M), pp. 5-26. San Diego Academic Press.
Hedrick P (2005) A standardized genetic differentiation measure. *Evolution* 59, 1633-1638.
Heisler IL, Damuth J (1987) A method for analyzing selection in hierarchically structured populations. *The American Naturalist* 130, 582-602.
Hood ME (2005) Repetitive DNA in the automictic fungus Microbotryum violaceum. *Genetica* 124, 1-10.
Hood ME, Antonovics J (2004) Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* 166, 1751-1759.
Hood ME, Antonovics J, Koskella B (2004) Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: Microbotryum violaceum and other model organisms. *Genetics* 168, 141-146.
Hood ME, Katawczik M, Giraud T (2005) Repeat-induced point mutation and the population structure of transposable elements in Microbotryum violaceum. *Genetics* 170, 1081-1089.
Hurst LD, Atlan A, Bengtsson BO (1996) Genetic conflicts. *Quarterly Review of Biology* 71, 317-364.
Husband BC, Barrett SCH (1998) Spatial and temporal variation in population size of Eichhornia paniculata in ephemeral habitats: implications for metapopulation dynamics. *Journal of Ecology* 86, 1021-1031.
Ingvarsson P (1998) Kin-structured colonization in Phalacrus substriatus. *Heredity* 80, 456-463.

Ingvarsson PK, Giles BE (1999) Kin-structured colonization and small-scale genetic differentiation in Silene dioica. *Evolution* 53, 605-611.
Ingvarsson PK, Ribstein S, Taylor DR (2003) Molecular evolution of insertions and deletion in the chloroplast genome of Silene. *Molecular Biology and Evolution* 20, 1737-1740.
Ingvarsson PK, Taylor DR (2002) Genealogical evidence for epidemics of selfish genes. *Proceedings of the National Academy of Sciences of the United States of America* 99, 11265-11269.
Ingvarsson PK, Whitlock MC (2000) Heterosis increases the effective migration rate. *Proceedings of the Royal Society B-Biological Sciences* 267, 1321-1326.
Johnson LJ, Antonovics J, Hood ME (2005) The evolution of intratetrad mating rates. *Evolution* 59, 2525-2532.
Jolivet C, Bernasconi G (2007) Molecular and quantitative genetic differentiation in European populations of *Silene latifolia* (Caryophyllaceae). *Annals of Botany* 100, 119-127.
Kaltz O, Schmid B (1995) Plant venereal-disease - a model for integrating genetics, ecology and epidemiology. *Trends in Ecology & Evolution* 10, 221-222.
Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters* 11, 852-866.
Kerr B, Neuhauser C, Bohannan BJM, Dean AM (2006) Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature* 442, 75-78.
Koskella B, Giraud T, Hood ME (2006) Pathogen relatedness affects the prevalence of within-host competition. *American Naturalist* 168, 121-126.
Le Gac M, Hood ME, Fournier E, Giraud T (2007a) Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution* 61, 15-26.
Le Gac M, Hood ME, Giraud T (2007b) Evolution of reproductive isolation within a parasitic fungal species complex. *Evolution* 61, 1781-1787.
Levins R (1969) Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the ESA* 15, 237-240.
Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997*.
Lockhart AB, Thrall PH, Antonovics J (1996) Sexually transmitted diseases in animals: ecological and evolutionary implications. *Biological Reviews* 71, 415-471.
Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution* 20, 136-142.
Manier MK, Seyler CM, Arnold SJ (2007) Adaptive divergence within and between ecotypes of the terrestrial garter snake, Thamnophis elegans, assessed with FST-QST comparisons. *Journal of Evolutionary Biology* 20, 1705-1719.
Maruyama T, Kimura M (1980) Genetic-variability and effective population-size when local extinction and recolonization of sub-populations are frequent. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 77, 6710-6714.
McCauley DE (1991) Genetic consequences of local-population extinction and recolonization. *Trends in Ecology & Evolution* 6, 5-8.
McCauley DE (1994) Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of Silene alba: implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the United States of America* 91, 8127-8131.
McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* 88, 257-263.
McCauley DE (2014) What is the influence of the seed bank on the persistence and genetic

structure of plant populations that experience a high level of disturbance? *New Phytologist* 202, 734-735.
McCauley DE, Olson MS, Taylor DR (2000) An association between chloroplast DNA haplotypeand gender in a plant metapopulation. *Evolutionary Ecology* 14, 181-194.
McCauley DE, Raveill J, Antonovics J (1995) Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* 75, 630-636.
McCauley DE, Stevens J, Peroni P, Raveill J (1996) The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* 83, 727-731.
McCauley DE, Taylor DR (1997) Local population structure and sex ratio: evolution in gynodioecious plants. *American Naturalist* 150, 406-419.
McKenna A, Hanna M, Banks E*, et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20, 1297-1303.
Mcneill J (1977) Biology of Canadian weeds .25. *Silene-alba* (Miller) Krause,EHL. *Canadian Journal of Plant Science* 57, 1103-1114.
Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4, 792-794.
Michod RE (1997) Evolution of the individual. *American Naturalist* 150, S5-S21.
Moody-Weis J, Antonovics J, Alexander H, Pilson D (2008) Predicting local colonization and extinction dynamics from coarser-scale surveys. *Ecography* 31, 61.
†Neiman, M., G. Hehman, J. M. Logsdon Jr., J.T. Miller and D.R. Taylor. (2010). Accelerated mutation accumulation in asexual lineages of a New Zealand snail. *Molecular Biology and Evolution* 27: 954-963.
O'Keefe K, Antonovics J (2002) Playing by different rules: the evolution of virulence in sterilizing pathogens. *American Naturalist* 159, 597-605.
Olson M, McCauley D, Taylor D (2005) Genetics and adaptation in structured populations: sex ratio evolution in *Silene vulgaris*. *Genetica* 123, 49-62.
Pannell JR, Charlesworth B (2000) Effects of Metapopulation Processes on Measures of Genetic Diversity. *Philosophical Transactions: Biological Sciences* 355, 1851-1864.
Pannell JR, Fields PD (2014) Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist* 201, 417-432.
Peroni PA, Armstrong RT (2001) Density, Dispersion and Population Genetics of a Silene latifolia Seed Bank from Southwestern Virginia. *Journal of the Torrey Botanical Society* 128, 400-406.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7, e37135.
R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Richards C (2000) Inbreeding depression and genetic rescue in a plant metapopulation. *American Naturalist* 155, 383-394.
Rolff J, Siva–Jothy MT (2004) Selection on insect immunity in the wild. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271, 2157-2160.
Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* 12, 253-262.
†Sloan DB. 2015. Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist*. 205: 1040-1046.
Sloan DB, Barr CM, Olson MS, Keller SR, Taylor DR (2008) Evolutionary rate variation at multiple levels of biological organization in plant mitochondrial DNA. *Molecular Biology and Evolution* 25, 243-246.
†Sloan, D.B. and D.R. Taylor. (2010). Testing for selection on synonymous sites in plant mitochondrial DNA. *Journal of Molecular Evolution* 5: 479-491.

†Sloan, D.B., A.H. MacQueen, A.J. Alverson, J.D. Palmer, and D.R. Taylor. (2010). Extensive loss of RNA editing sites in rapidly evolving mitochondrial genomes: selection versus retroprocessing as the driving force. *Genetics* 185: 1369-1380.

†Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. (2010). Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. *BMC Evolutionary Biology*. 10:274.

†Sloan DB, Taylor DR. (2012). Evolutionary rate variation in organelle genomes: the role of mutational processes. In Bullerwell, Charles E. (ed). Organelle Genetics. Springer-Verlag. pp 123-146.

†Sloan, DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer, JD and DR Taylor (2012) Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biology* 10(1): e1001241.

†Sloan, DB, Keller, Stephen, S, Berardi, A, Sanderson, B, Karpovich, J, Taylor, D. (2012). De novo transcriptome assembly and polymorphism detection in the flowering plant Silene vulgaris (Caryophyllaceae). *Molecular Ecology Resources* 12: 333-343.

†Sloan, DB, K Müller, DE McCauley, DR. Taylor, and H Štorchová. (2012). Intraspecific variation in mitochondrial genome sequence, structure, and gene content in Silene vulgaris, an angiosperm with pervasive cytoplasmic male sterility. *New Phytologist* 196: 1228-1239.

†Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. (2012). Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus Silene. *Genome Biology and Evolution* 4: 294-306.

†Sloan DB. 2013. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytologist*. 200: 978-985.

†Sloan DB, Triant DA, Forrester NJ*, et al.* (2014) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). *Molecular Phylogenetics and Evolution* 72, 82-89.

†Sloan DB, Triant DA, Wu M, Taylor DR (2014) Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes. *Molecular Biology and Evolution* 31, 673-682.

Stan Development Team (2015) Stan: A C++ Library for Probability and Sampling.URL: http://mc-stan.org/

Smouse PE, Waples RS, Tworek JA (1990) A Genetic Mixture Analysis for use with Incomplete Source Population Data. *Canadian Journal of Fisheries and Aquatic Sciences* 47, 620-634.

Stevens L, Goodnight CJ, Kalisz S (1995) Multilevel selection in natural-populations of Impatiens capensis. *American Naturalist* 145, 513-526.

Storchova H, Hrdlickova R, Chrtek J*, et al.* (2000) An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49, 79-84.

Szathmary E, Smith JM (1995) The major evolutionary transitions. *Nature* 374, 227-232.

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. *Genetica* 117, 27-35.

Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61, 334-345.

Taylor DR, Olson MS, Mccauley DE (2001) A quantitative genetic analysis of nuclear-cytoplasmic male sterility in structured populations of Silene vulgaris. *Genetics* 158, 833-841.

Taylor DR, Zeyl C, Cooke E (2002) Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3690-3694.
Teixeira S, Bernasconi G (2007) High prevalence of multiple paternity within fruits in natural populations of Silene latifolia, as revealed by microsatellite DNA analysis. *Molecular Ecology* 16, 4370-4379.
Thrall P, Richards C, Mccauley DE, Antonovics J (1998) Metapopulation collapse: the consequences of limited gene flow in spatially structured populations. In: *Modeling spatiotemporal dynamics in ecology* (eds. Bascompte J, Sole RV), pp. 83-104. Springer-Verlag Berlin.
Thrall PH, Antonovics J (1995) Theoretical and Empirical Studies of Metapopulations - Population and Genetic Dynamics of the Silene-Ustilago System. *Canadian Journal of Botany-Revue Canadienne De Botanique* 73, S1249-S1258.
Thrall PH, Jarosz AM (1994a) Host-pathogen dynamics in experimental populations of *Silene alba* and *Ustilago-violacea* .1. Ecological and genetic-determinants of disease spread *Journal of Ecology* 82, 549-559.
Thrall PH, Jarosz AM (1994b) Host-pathogen dynamics in experimental populations of *Silene alba* and *Ustilago-violacea* .2. Experimental tests of theoretical-models. *Journal of Ecology* 82, 561-570.
Tsitrone A, Rousset F, David P (2001) Heterosis, marker mutational processes and population inbreeding history. *Genetics* 159, 1845-1859.
Van der Auwera GA, Carneiro MO, Hartl C*, et al.* (2013) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc.
Venable DL, Brown JS (1988) The Selective Interactions of Dispersal, Dormancy, and Seed Size as Adaptations for Reducing Risk in Variable Environments. *The American Naturalist* 131, 360-384.
Vondrasek J, Antonovics J, Taylor D (2004) Evolution Kills! A Web Resource for Instructors of Evolutionary Biology. *Bioscene* 30, 3-8.
Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42, 995-1005.
Wagenius S (2006) Scale dependence of reproductive failure in fragmented Echinacea populations. *Ecology* 87, 931-941.
Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics* 184, 363-U390.
Westergaard M (1958) The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics Incorporating Molecular Genetic Medicine* 9, 217-281.
Whitlock MC (2002) Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160, 1191-1202.
Whitlock MC, Fowler K (1996) The distribution among populations in phenotypic variance with inbreeding. *Evolution* 50, 1919-1926.
Whitlock MC, Ingvarsson PK, Hatfield T (2000) Local drift load and the heterosis of interconnected populations. *Heredity* 84, 452-457.
Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44, 1717-1724.
Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} not equal 1/(4*Nm* + 1). *Heredity* 82, 117-125.
Willi Y, Fischer M (2005) Genetic rescue in interconnected populations of small and large size of the self-incompatible *Ranunculus reptans*. *Heredity* 95, 437-443.
Willi Y, Van Buskirk J, Fischer M (2005) A threefold genetic allee effect: population size affects cross-compatibility, inbreeding depression and drift load in the self-incompatible *Ranunculus reptans*. *Genetics* 169, 2255-2265.

Wilson DS (1997) Altruism and organism: Disentangling the themes of multilevel selection theory. *American Naturalist* 150, S122-S134.
Wolfe LM, Blair AC, Penna BM (2007) Does intraspecific hybridization contribute to the evolution of invasiveness?: an experimental test. *Biological Invasions* 9, 515-521.
Wright S (1931) Evolution in mendelian populations. *Genetics* 16, 97-159.
†Wu Z, Cuthbert JM, Taylor DR, Sloan DB. (2015). The massive mitochondrial genome of the angiosperm Silene noctiflora is evolving by gain or loss of entire chromosomes. Proceedings of the National Academy of Sciences USA. (In press)

PETER D. FIELDS
Post-doctoral Researcher
Zoologisches Institut,
Universität Basel, Basel, Switzerland, CH-4051

(a) Expertise as related to the proposed research

My training is in evolutionary biology, primarily focusing on the development and application of statistical genetic methods for the inference of ecological and evolutionary processes. My dissertation research included the development of high throughput analytical methods focusing on the demographic and genetic dynamics of *Silene latifolia* and *Microbotryum violaceum* metapopulation. I improved the methods of analyzing these genetic data, and similar datasets, by developing of a novel statistical genetic model and simulation framework. The model, developed through our international collaboration with the Université de Grenoble, allows us to use data derived from multiple genomic regions, even if the different regions have very different evolutionary histories and rates of mutation. In addition, my research develops the application of approximate Bayesian computation in order to 1) incorporate the reality that the conclusions drawn may be affected "ghost" populations that are left un-sampled, and 2) reconstruct historical demographic processes during biological invasions. My current research on the genomics of *Daphnia magna* metapopulations has included extensive *de novo* genome sequencing and population genomics which overlaps broadly with the proposed research.

(b) Professional Preparation

Post-doctoral Associate, University of Basel, Basel, Switzerland, 2014-present
University of Virginia, Charlottesville, Virginia, Evolutionary Biology, Ph.D., 2013
Alice Lloyd Collge, Pippa Passes, Kentucky, Biology, B.A., 2006

(c) Five Relevant Publications

Fields, P., D. McCauley, E. McAssey, and D. Taylor (2014). Patterns of cyto-nuclear linkage disequilibrium in Silene latifolia: genomic heterogeneity and temporal stability. Heredity 112: 99-104.

Pannell, J. R. and P. D. Fields (2014). Evolution in subdivided plant populations: concepts, recent advances and future directions. New Phytologist 201: 417-432.

Keller, S., Fields, P., Berardi, A. and D. Taylor (2014) Recent admixture generates heterozygosity-fitness correlations during the range expansion of an invading species. Journal of Evolutionary Biology, 27: 616-627.

Fields, P., S. Keller, P. Ingvarsson, A. Pedersen and D.Taylor (2010). Isolation and characterization of highly polymorphic microsatellite loci in *Silene latifolia* (Caryophyllaceae). Molecular Ecology Resources 10 (1): 232-236.

Keller, S., Gilbert, K., Fields, P., and D. Taylor (2012). Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, Silene latifolia. Molecular Ecology (21): 4721-4734

(d) Five Additional Publications.

Wade, M. J., D. S. Wilson, C. Goodnight, D. Taylor, Y. Bar-Yam, M. A. M. d. Aguiar, B. Stacey, J. Werfel, G. A. Hoelzer, E. D. B. Iii, P. Fields, F. Breden, T. A. Linksvayer, J. A. Fletcher, P. J. Richerson, J. D. Bever, J. D. V. Dyken, and P. Zee (2010) Multi-level and kin selection in a connected world. *Nature*. 463 (E8-E9).

Fields, P. and D. Taylor. 2015. Spatiotemporal effects on F_{ST} in a metapopulation of Silene latifolia. PLoS ONE. 2014;9(9):e104575.
Fields, P., Arnold, G., Kniskern, J. M., and D. Taylor (In Review) Population history and differential consequences of inbreeding and outcrossing in a plant metapopulation. American Naturalist
Fields, P.,Weingartner, L. A., and L. F. Delph (2015) Transcriptome resources for two highly divergent Silene latifolia populations. Molecular Ecology Resources
Fields, P., Reisser, C., Dukic, M., Haag, C., and D. Ebert (In Press) Genes mirror geography in *Daphnia magna*. Molecular Ecology

(e) Synergistic Activities

1. Peer reviewer for the following journals: Evolution, Molecular Ecology Resources, Molecular Ecology, New Phytologist, Genetica, Genetics, Biological Invasions, Heredity
2. Teaching assistant for Introductory Molecular Genetics Laboratory, Introductory Organismal and Evolutionary Biology Laboratory, Genetics and Molecular Biology, and Biology of Infectious Disease at the University of Virginia. Designed and taught a laboratory in an Evolutionary Biology lab where students use molecular markers to study a zone of admixture among Silene lineages near Mountain Lake Biological Station. Co-advisor on an NSF Research Experience for Undergraduates project at he Mountain Lake Biological Station in the Summers of 2009-2010. Past and Current Undergraduate Advisees: Katie Short (UVa), John Soong (UVa), Kimberly Gilbert (UVa; currently a PhD candidate at the University of British Columbia), Brenden Barco (Duke University; currently a laboratory technician at LabCorp, Burlington, North Carolina), Alex Bollinger (University of Paris and Museum National d'Histoire Naturelle; co-advised senior thesis with Frederic Austerlitz), Anthony Ortiz (UVa); Leena M. Abdel-Qader (UVa), and Rachael Hallock (UVa).
3. Working with the University of Virginia Research Computing Lab, I have developed a number of computational projects, including creating a relational database to consolidate and disseminate current and future data on the regional abundance of *Silene* and *Microbotryum* in the Mountain Lake region of Virginia.
4. Assisted in conducting Software Carpentry workshop, titled “Advanced Programming in R”
5. Assisted in the teaching of short course at the University of Basel focusing on Bayesian statistics, titled “Bayesian Data Analyses Using Linear Models with R and WinBUGS”.

FACILITIES, EQUIPMENT, AND OTHER RESOURCES

University of Virginia

Personnel: Sufficient funding for the personnel needed to carry out the proposed experiments has been requested in the budget. In addition, Professors Taylor and Antonovics will be available to supervise lab personnel and manage the overall project. Taylor has a moderate teaching load and administrative duties, leaving sufficient time to take on the responsibilities associated with this proposal. Antonovics' position is entirely research.

Office space: There is a ~140 sq. ft. office space for Dr. Taylor and a ~240 sq. ft. office space for Dr. Antonovics, each located within their main laboratory spaces. Desks for students and post-docs are located in the laboratories with expansion space also available in the wet lab across the hall shared by the UVA evolutionary biology faculty.

Laboratory: The Taylor lab has exclusive use of approximately half of a fully rennovated ~1540 sq. ft. wet laboratory space in Gilmer Hall at UVA. The space is shared in a synergistic way with the lab of Prof. Ben Blackman, who studies plant evolutionary genomics using many of the same techniques. Emergency power is available for all freezers with 24-hour off-site monitoring for -80°C freezers. A Millipore water purification system is available in the central lab space. Dr. Antonovics currently has a ca. 1200 sq. ft. lab for molecular and microbiological work. The lab is currently equipped with a laminar flow hood, electrophoresis rigs, thermalcycler, Eagle-Eye gel visualization system, microcentrifuge, Nikon TMS inverted microscope and a Zeiss 2000 stereoscope

Common use autoclaves and glassware-cleaning equipment are located across the hall. Also across the hall from the main lab spaces are an equipment room for centrifuges, freezers, and incubators; a tissue culture room; a microscope room equipped with vibration-insulated tables; and a cold room. These facilties will be shared with Dr. Taylor, Dr. Butch Brodie, Dr. Laura Galloway, and Dr. Robert Cox. The lab is also in close proximity to members of the Department of Biology, the Department of Biology Genomics Core Facility, and the W.M. Keck Center for Cellular Imaging. In addition, it is a short walk from the UVA School of Medicine Biomolecular Reseach Facility and UVA Bioinformatics Core Facility.

Routine Molecular Biology Equipment: The Taylor lab is well equipped for molecular biology experiments with its own equipment or through equipment shared with Dr. Blackman lab in our joint wet lab suite (thermocyclers (10), water baths, -20ºC freezers, refrigerators, microfuges, microwave, pH meters, balances, vortexes, gel rigs for analysis of nucleic acids, a magnetic rack for bead purifications during sequencing library construction, etc.). We also have access to departmental shared equipment including autoclaves, cold and warm rooms, ultra-, high-, and low-speed centrifuges, spectrophotometers, an ABI 7500/7500 Fast Real-Time PCR system, a film processor, a phosphoimager, microtomes and histological equipment, a nanodrop, a tissue disruptor, a deep-well centrifuge for high throughput isolation of plant DNA, and a gel documentation system.

Core Facilities: The Department of Biology has an in-house stockroom and machine shop. In addition, we have access to two sequencing facilities.

Department of Biology Genomics Core: The departmental core is staffed and right across the hall. Although large format sequencing runs are generally farmed out to larger facilities, the Biology core provides ideal support for the proposed research. The facility has a Beckman BioMek NX robot equipped for automated library construction for high throughput sequencing, and a SageScience BluePippin for size selection. The facility has a MiSeq Personal Sequencer (ideal for initial QC runs), a 454 Life Science / Roche Genome Sequencer FLX sequencer, and a 16-capillary ABI 3100 sequencer for the occasional Sanger sequencing and fragement analysis. A QuBit and a BioAnalyzer for DNA, RNA, and Illumina library quantification and quality control are also available. The core also maintains a Spectromax M3 plate reader that can obtain absorbance, fluorescence, and luminescence data in cuvette, 24-, 96- or 384-well format. The core also features a benchtop Covaris ultrasonicator for nucleic acid shearing, a

UVA School of Medicine Biomolecular Research Facility and Bioinformatics Core: This is fee-for-service facility that we occasionally use. The molecular facility has two high-throughput sequencing platforms: an Illumina GAII sequencer and a MiSeq system. Also available are an ABI Prism 7900HT sequence detection system for quantitative PCR machine, an ABI 3730 DNA analyzer for Sanger sequencing, and a Molecular Devices Gemini fluorescent plate reader. The Bioinfomatics Core provides occasional support and student/post-doc training.

Computers: One PC and three desktop iMac computers, connected to the University of Virginia network, are available in the lab for use by post-docs and students. Software available for analysis of Genotyping-by-Sequencing data, QTL mapping, and other computational and statistical analyses include: R, TASSEL, structure, Bioconductor, bwa, samtools, MSG, Geneious, Perl, Python, C++, JMP, SPSS, and SAS. The Taylor lab also has access to the University of Virginia's Rivanna advanced computed cluster that totals over 6000 cores and a 1.4PB shared filesystem. We also use XSEDE (www.xsede.org) for large assemblies. The Taylor lab also has our own Linux cluster (managed by DRT) for population genetics and routine bioinformatics, including a 32 core (132GB RAM) Dell with 42TB of storage, with a dedicated 42TB file server for off site nightly backups.

Plant Growth Facilities: The plant biologists in the Biology Department (especially Ben Blackman, Laura Galloway and Deborah Roach) combine plant growth chamber resources within Gilmer Hall. For example, the Taylor lab "owns" two Percival AR-75L2 reach-in growth chambers and the Antonovics lab has several smaller reach in chambers for fungal culture and plant innoculations, the community provides an additional seven shared departmental reach-in chambers are accessible during peak demand. The Biology Greenhouses are fully staffed and adjacent to Gilmer Hall. Space is assigned according to need, and there is abundant space for the proposed research. The greenhouse has a full-time staff that takes care of watering, pest-management, and maintenance, and it also has a fully equipped head house.

Mountain Lake Biological Station (MLBS): MLBS is a three hour drive from UVA. There is ample space to establish fenced experimental areas for our plants, and net covered cages for holding diseased plants. There are areas set aside where we can mix soil and pot up plants for experiments. We have access to station vehicles and a 300 gallon tank for watering plants during periods of drought. Laboratory and computer facilities are provided, but can be augmented when necessary by bringing items from our lab at UVA.

In summary, all facilities, equipment, and resources required for successful completion of this project are either already in place or have been budgeted for in the current proposal.

DATA MANAGEMENT AND DISSEMINATION PLAN

Among the most important promises of the 'big data' paradigm are the broad opportunities for interdisciplinary collaboration. However, in order for these collaborations to occur the huge volumes of data that are collected need to not only be shared, but also curated in a way that makes the data approachable to researchers that lack domain-specific expertise. This requires not only the publication of genomic information, as is the current standard in our field, but also the publication of the huge volumes of phenotypic, population, and ecological data that are relevant. The proposed project builds on 30 years of data collection, and we intend to provide these data not just as published papers in scientific journals, but also as useful tools to facilitate future work above and beyond what we envisage here.

Our research group already has a strong history of providing these community resources through our website http://silenegenomics.biology.virginia.edu. Since 2011 we have provided the next-generation data we have generated ahead of publication, as both web-accessible databases that users can query directly on the site as well as downloadable copies of the databases so that users can independently work with the data on their own local computers. As a part of this proposal we will expand the scope of this website to include 1) the additional genomic and transcriptomic data that will be generated, 2) relational databases of the spatial, ecological, phenotypic, and genotypic data that we have recorded over the past 30 years, and 3) a database of our current tissue, DNA, and seed collections that are available to researchers upon request. These database schema will enable researchers to create dynamic queries that draw upon the inherent relationships among these experiments and samples. For example, a researcher designing a new common garden experiment could determine what seeds are available from plants that have known microsatellite genotypes, explicit spatial position, historical ecological measurements, and a history of disease presence/absence.

The website, MySQL databases, and FASTA/FASTQ files will all be served from a dedicated server through the University of Virginia's IT department, which will provide support and regular, offsite backups. We will inform members of the community about the resource via scientific meetings and conferences, and direct postings on public bulletins such as EvolDir. There will be links through related web portals, e.g. the *Microbotryum* portal (http://www.amherst.edu/~mhood/mv/) and the recently proposed portal for *Silene*.

Taken together, we hope that this singular resource will facilitate exciting future collaborations, meta analyses, and extensions of our larger research agenda.

POSTDOCTORAL MENTORING PLAN

Funds are requested for partial support for postdoctoral research associate spanning middle three years of the project where the genomic data collection is expected to be most intense. The early duties will involve DNA preps and working with the UVA genomics core and sequencing services to organize the data collection, but the tasks will broaden with time to include mentorship, data management and laboratory budgeting and management skills. The PIs have supervised more than 25 post-doctoral associates with more than 20 in tenure track faculty positions, and will continue many of their formal and informal mentoring practices:

- The postdoctoral associate will supervise hourly age employees in routine tasks such as maintain the proposed collections and cultures as well as develop new and broader collaborations through (likely) REU supplements and resulting side projects.
- The post-doctoral associate will assume broader responsibilities in the integration of different aspects of the project, gaining valuable experience in the overall project logistics, budgeting, and data management required as a professional scientist.
- The post-doctoral associate will collaboration closely with Antonovics, Fields and Taylor, thus acquiring training in the analysis of the next-generation sequence data, population genetic analyses using large datasets, database management and dissemination, and of course publication of the results.
- The post-doctoral associate will receive individual feedback in weekly meetings with the PIs and group feedback from weekly lab group meetings. Feedback includes issues outside of the project, *per se*, to include people management skills and broader aspects of career development such as grant proposal preparation.
- The PIs will work together with the post-doctoral associate to develop formal mentoring framework, including an Individual Development Plan (IDP). Updated each year, we proposed to follow guidelines from both UVA and the National Postdoctoral Association to cooperatively develop plans for assessment (including self-assessment), establishing goals and carefully considering long-term career paths and interest. These plans will be reviewed and revised, and progress discussed, on an annual basis.