1. Background

All organisms live in environments that vary through time and space. As a consequence of environmental heterogeneity, populations of most species exhibit patterns of local adaptation in life-history, stress tolerance, and morphology. Local adaptation, a form of balancing selection, promotes functional genetic diversity within the species and can sometimes maintain it at intermediate frequencies for long periods of time. The overarching goal of my research program is to understand the genetic and physiological architecture of local adaptation in fitness related traits, to examine the long- and short-term evolutionary history of loci underlying local adaptation, and to further our understanding the nature of standing genetic variation in general.

I study the genetic and physiological architecture of local adaptation in two complimentary systems. The primary system I work on is *Drosophila melanogaster* which has served for decades as an important model for understanding local adaptation[1]. *D. melanogaster* originated in central Africa[2] and colonized the world in the wake of human migration over the last ca. 100-10,000 years[3]. Fly populations arrayed along latitudinal[4] and altitudinal[5] clines exhibit patterns of genetically based phenotypic differentiation consistent with local adaptation to temperate environments. Flies from poleward or higher elevation locales tend to invest resources more heavily in somatic maintenance whereas those from equatorial or lower elevation locales tend to invest more heavily in reproductive output. This trade-off is consistent with the basic resource allocation model of life-history evolution.

Genetically based phenotypic differentiation in fitness related traits also occurs among seasons[4]. Notably, flies sampled in spring - the recent descendants of individuals that survived winter - tend to be more hardy whereas those sampled during fall - the descendants of those individuals that prospered during summer - tend to invest resources more into reproduction. Because of the short generation time of *D. melanogaster*, rapid adaptive change among seasons occurs in fewer than 10-15 generations. We have shown that adaptation over seasonal time scales is driven by dramatic and cyclic changes in allele frequency at hundreds to thousands of loci that have measurable effect on phenotype[6]. These alleles segregate in ancestral populations and are thus likely old, balanced polymorphisms. As I discuss in Recent Progress, understanding the specific genetic architecture of local adaptation is tractable, particularly in organisms like *D. melanogatser*. Moving forward, my lab will pursue a mechanistic, molecular understanding the genetic architecture and physiological processes that underlie adaptation to temperate environments in *Drosophila*.

The second major system that I work on is *Daphnia pulex*. *Daphnia* has long served as a classic model for studying phenotypic plasticity. Individuals in this genus develop defensive morphology (i.e., “neckteeth”) and alter their life-history (growth rates and age-specific fecundity) in response to predator derived chemical cues[7]. *Daphnia* are preyed upon by two main predator groups that impose size selective mortality driving local adaptation and phenotypic plasticity. *Chaoborus* midge larvae tend to prey on small individuals whereas large vertebrate predators such as stickleback fishes tend to prey on large individuals.

Natural ponds vary in predation regime and we have examined adaptive divergence in *Daphnia* collected from ponds with either multiple predators (midge + fish) or a single predator (just midge). We have shown that heterogeneous predation pressure among ponds has driven local adaptation in life-history plasticity. As I discuss in Recent Progress, we have identified hundreds of polymorphisms that predictably vary among ponds with alternate predation regimes. Intriguingly, predation regime varies seasonally and analysis of a limited number of alozymes suggest that the population-genetic structure of *Daphnia* dramatically shifts over seasonal time scales[8]. These observations open the exciting possibility that *Daphnia* exhibits rapid adaptive oscillations in life-histories, similar to what we observe in *Drosophila*. Below, I discuss how we will examine the genetic architecture of local and rapid adaptation by taking advantage of the unique ecological and genetic attributes of *Daphnia*.

Both of the systems that I study are remarkable in that we have found large numbers of loci that show dramatic changes in allele frequency in response to fairly mundane selection pressures (seasons, predation). These, or similar, selection pressures are felt by virtually every species. Flies and *Daphnia* have been exposed to heterogeneity in such selection pressures for long periods of time. This suggests that environmental variation through time and space has maintained a large amount of functional & fitness related genetic variation for eons or has promoted recurrent *de novo* evolution of functional polymorphisms. A strict reading of the Neutral Theory[9] would suggest that such functional variation should be exceedingly rare and, moreover, difficult if not impossible to map to the nucleotide. The empirical fact that we, and others, have had success mapping such polymorphisms to the nucleotide suggests that the Neutral Theory remains incomplete in explaining patterns of genetic variation.

Key Gaps in Understanding and Opportunity for Conceptual Advancement. Recent advances have allowed us to make tremendous strides in understanding the genetic architecture of phenotypic variation and identifying the genomic basis of local adaptation. Many Mendelian traits have now been mapped to the nucleotide[10] and we have developed more nuanced insight into the nature of quantitative traits, even if their exact genetic architecture remains somewhat elusive[11] . We are now able to investigate the genomic basis of local adaptation in virtually every species[12], although our general understanding of how genetic variation maps to phenotypic variation is murky. It has been long suspected that local adaptation in response to environmental heterogeneity through time and space primarily drives phenotypic divergence of highly quantitative fitness traits (e.g., life-histories, morphologies, behaviors, physiology, development). Given the challenges in identifying the genetic architecture of these fitness-related traits there is a key gap in our knowledge of the mechanistic relationship between phenotype and genotype (i.e., 'the architecture') in the context of local adaptation.

The goal of identifying the genetic architecture of adaptation in fitness related phenotypes has recently been criticized[13,14]. For instance, Rockman[13], mirroring earlier sentiments[15], suggests there is an epistemological limit to this pursuit. Notably, he points out empirical support for the 'infinitesimal model' and suggests that measuring the phenotypic effect of common loci with exceedingly small effect size may not be technically feasible. Similar sentiments were recently echoed from a population genetic perspective[16], where a major supposition of the dominant Neutral Theory[9] suggests that the vast majority of common polymorphisms have infinitesimally small effects (i.e., are effectively neutral). Under the infinitesimal model, instances where phenotypic effects of natural alleles are measurable may rather reflect Mendelian loci[13] and thus might not inform our understanding of adaptation of quantitative traits. Yet, as I discuss below the phenotypic effect of loci that contribute to local adaptation in polygenic traits is measurable and strongly affects fitness in the wild. Thus, the dominant models of both evolutionary quantitative genetics and population genetics might not accurately reflect the nature of standing genetic variation.

Because we have a limited knowledge about the genetic architecture of local adaptation, we have a limited ability to address basic questions about their long- and short-term molecular evolutionary history. For instance, with an explicit and mechanistic knowledge of the genetic architecture of local adaptation we will be able to contribute to conceptual advances in evolutionary genetics by addressing the following questions: What fraction of natural polymorphisms contribute to local adaptation - is it a few or many? Is rapid and cyclic polygenic adaptation driven by predictable allele frequency changes at common polymorphisms? How old are ecologically balanced polymorphisms? Does local adaptation promote neutral genetic diversity genome-wide? Rigorously addressing these questions in a systematic and general way requires that we develop an understanding of the genetic architecture of polygenic local adaptation in multiple organisms.

In principal, many of these evolutionary questions can be addressed by identifying loci with significant patterns of differentiation through time and space in the absence of any concrete linkage to phenotypic variation. However, just as in analyses of adaptive evolution from classic molecular evolution and population genetic analysis, such landscape genomic approaches suffer from unknown false positive and negative rates due to historical demography[17]. Thus, without a systematic linkage of genotype to phenotype we have only a limited knowledge of the set of true positive loci with known functional effect. Furthermore, when testing the phenotypic effect of candidate polymorphisms we often focus on those candidates close to well known genes with plausible links to the phenotypes presumed to be under selection. This not only imposes a strong bias[18] but limits opportunities for functional annotation of un-studied genes[19]. One way to systematically link polymorphisms to specific genes in a relatively unbiased fashion is to perform cis-eQTL mapping and layer those functional annotations with estimates of allele frequencies through time and space[20]. Currently we have very little knowledge of the genetic basis of expression variation in *Drosophila* or *Daphnia.* This deficit inhibits us from making clear and testable predictions about the molecular function of polymorphisms we that have already identified from population & ecological genomic inference or addressing the broader questions in evolutionary genetics that I posed above.

2. Recent progress

Drosophila. Classically evidence of adaptation to temperate environments in *D. melanogaster* came from phenotypic, genetic, and genomic analysis along latitudinal gradients, particularly in North America and Australia. Often it was assumed that genetic differentiation along these clinal gradients was generated by spatially varying selection, with demography playing a limited role due to high migration rates among neighboring populations. In contrast, we have now shown that clinal variation in North America and Australia

may have been generated by secondary contact of European and African populations[21]. While adaptation in response to spatially varying selection does occur along these clines, identifying its genetic basis from patterns of allele frequency differentiation through space is difficult because of the confounding effects demography. However, we identified hundreds of loci that repeatedly shift in allele frequency between spring and fall in a single orchard population (Figure 1). Because these short term seasonal fluctuations are independent to the long term demographic history of this species, differentiation at these loci through time is unlikely driven by colonization history and thus likely reflects adaptation[6].

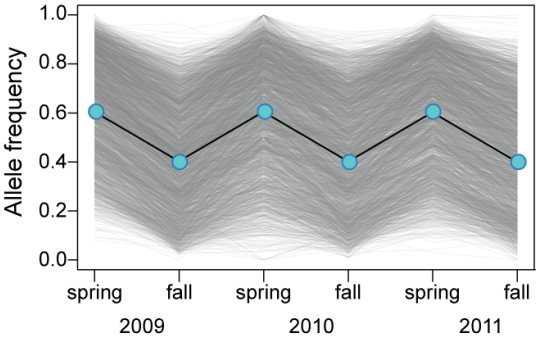

Figure 1. Adaptive oscillations at seasonal SNPs. Each line represents the allele frequency trajectory of a single SNP in a Pennsylvanian orchard population. Bold line shows average change between seasons. Figure taken from Bergland *et al*., 2014.

We sought to test whether adaptive oscillations over seasonal time scales were observable in additional populations throughout North America. To our surprise, large shifts in allele frequency at previously identified 'seasonal SNPs' in these newly sampled populations were only weakly predictable. There are many possible technical reasons for low, but observable, levels of predictability including mis-identification of seasons SNPs due to linkage. However one possibility remains tremendously exciting: **There are many functional polymorphisms that contribute to adaptation among seasons and in any particular phase of the seasonal cycle only some loci change dramatically in allele frequency; under this model, our identified seasonal SNPs only reflect a tiny fraction of functional polymorphisms in the genome that *could* contribute to seasonal adaptation.** Sporadic oscillations of seasonal SNPs might reflect different selective pressures from year to year or population to population. Alternatively, theory we are developing suggests that that weakly predicable changes in allele frequency among independent replicate populations exposed to identical selection pressures is not unexpected expected when populations undergo dramatic polygenic adaptation with per locus $s \sim 10\%$.

Several lines of evidence suggest that seasonal SNPs are likely functional despite only moderate predictability of allele frequency change between spring and fall. First, we find that seasonal SNPs tend to vary among populations in ways that are consistent with adaptation to temperate environments: winter favored alleles are more common in northern locales than summer favored alleles and these patterns are found among our newly collected samples. Second, we have characterized the phenotypic effect of several seasonal SNPs. In one example[22], we examined seasonal SNPs in *Insulin Receptor* (*InR*) and showed that these naturally segregating polymorphisms have measurable affects on highly polygenic life-history and morphological traits. Intriguingly, these SNPs showed large adaptive oscillations in only a subset of years[22] despite adequate power to detect large shifts in allele frequency among seasons. This pattern is reminiscent of the moderate predictability adaptive oscillations among populations. We have begun a more systematic analysis of the phenotypic effect of seasonal SNPs by contrasting patterns of allele frequency change among seasons with SNPs associated with traits known to vary seasonally: generally, winter favored alleles are more likely to be associated with increased stress tolerance[6].

The other line of evidence that many seasonal SNPs are functional comes from population genetic inference. We calculated genetic diversity surrounding seasonal SNPs to test the idea that fluctuating selection pressures (a form of balancing selection) maintains genetic diversity at linked neutral SNPs. Theory suggests that an accumulation of genetic diversity is expected, but its decay rate will depend on the number of functional SNPs in the region. Because of the high recombination rate in Drosophila, we expect that such peaks of diversity should decay by ~500bp around a single balanced polymorphism. Yet, on average across all seasonal SNPs, we see elevated diversity for upwards of 50Kb. One model that can explain our elevated regions of diversity is that there are genes with multiple, independent seasonal SNPs that are maintained for long time-periods as a consequence of environmental heterogeneity through time and space.

To begin to rigorously test this polygenic model of adaptation we need to understand the phenotypic effect of seasonal SNPs identified through population/ecological genetic inference. Thus, as I discuss below, my initial research projects will be focused on identifying regulatory function of seasonal SNPs via cis-eQTL mapping. These projects will utilize resources that I currently have in hand. Such resources include a new set of ~250 inbred and fully resequenced strains of *D. melanogaster* we collected in the fall in Maine or during the spring and fall Pennsylvania. These lines can be used in conjunction with other resequenced populations such as the DGRP population from North Carolina[23]. Analytic methods I will use in the proposed mapping

experiments will be based on a pilot project that uses a similar experimental design but was conducted indoors. In addition, I have generated allele specific mRNA-seq and FAIRE-seq libraries from F1 hybrids between spring and fall inbred lines that were exposed to different environmental conditions; these data can be used to vet and verify the proposed cis-eQTL experiments. Finally, this work will complement the ongoing efforts of a consortium of fly biologists world-wide that I am co-organizing whose goal is to generate genome-wide estimates of allele frequencies through time and space over decadal scales and among multiple continents. Our consortium's work will be an incredible population genomic resource for the community and will enable research in computational and statistical methods to identify loci underlying adaptation to temperate environments. Understanding the functional role of these polymorphisms in a systematic manner will therefore be of great interest and utility to the community and represents a major innovation of this grant proposal.

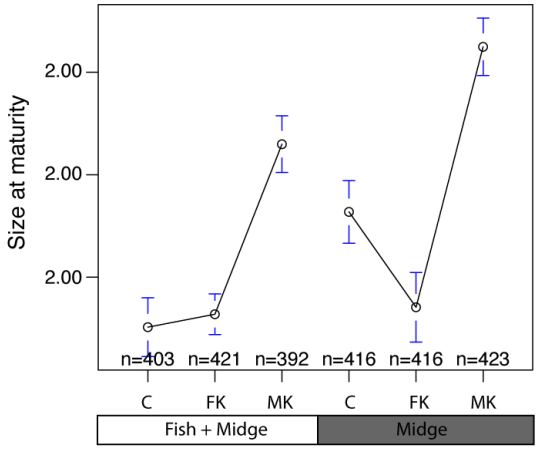

Figure 2. Adaptive reaction norms of Daphnia in response to fish (FK) and midge (MK) kairamones from clones collected in two different predation regimes, either fish and midge or just fish.

Daphnia. Our work on *Daphnia* is at a different stage than our *Drosophila* work but I anticipate it will follow a similar trajectory. Here, I describe the state of our *Daphnia* research which focuses on adaptive differentiation of life-history traits in response to differences in predator regime.

Daphnia, a keystone species group in aquatic ecosystems, are herbivorous grazers that are critical to the prevention of eutrophication and an important prey for larger planktonic predators. *Chaoborus* midge larvae, a gape-limited predator, tend to prey on smaller Daphnia individuals whereas larger vertebrate predators, such as sticklebacks and salamanders, tend to prey on larger individuals. These predation pressures impose strong size selective mortality in *Daphnia* populations.

Daphnia alter their morphology and life-history in response to these predators. In response to midge kairomones, Daphnia grow characteristic neck-teeth which prevent predation. When exposed to these same kairomones, *Daphnia* also tend to sexually mature later and a larger size. Conversely, when exposed to fish cues *Daphnia* tend to mature early and at a smaller size (Figure 1). We have shown that *Daphnia* from ponds with only midge predation (M) or midge and fish predation (MF) have adaptive differences in the plastic life-history response to these predators. These shifts in the age and size of reproduction are predicted by life-history theory[24].

We resequenced pooled samples of *Daphnia* clones from these ponds to estimate allele frequencies genome-wide. These data have identified several hundred SNPs that change in allele frequency among predation regimes. These loci are enriched near endocrine genes with known effects on invertebrate life-histories suggesting that they directly underlie adaptive divergence in life-history traits. Intriguingly, population genomic analyses of these loci suggests that many have undergone recent selective sweeps independently in each pond. The high rate of recurrent adaptation at these loci is possibly consistent with the life-cycle of Daphnia or the colonization history of the focal ponds.

There is much work to be done on this *Daphnia* system and it is ripe for investigation and innovation. The Daphnia genome has been sequenced and annotated and others are actively developing population genomic resources for this model system[25]. Data we generate under this proposed research will therefore be of interest and value to the *Daphnia* community.

3. Research plan.

Herein, I propose a series of projects in *Drosophila* and *Daphnia* to functionally characterize cis-eQTL in the context of ecologically relevant, environmental variation. Identification of environment dependent cis-eQTL will aid in the biological interpretation of polymorphisms identified to underlie local adaptation through time and space. By merging functional data that I will generate with existing and new allele frequency data from natural populations we will be able to make substantial progress in understanding the genetic and physiological architecture of local adaptation. Doing so will enable us to address the broader questions in evolutionary genetics that I highlighted above at the end of *Key Gaps*. By addressing this basic question of the genetic architecture in two species that differ in sexual dynamics (obligate versus facultative out-crossers) living in different ecosystems (terrestrial versus aquatic) and exposed to different classes of

selection pressure (abiotic versus biotic) we will be able to make generalized statements about the long- and short-term evolutionary history of loci that underlie polygenic local adaptation.

Drosophila. My lab will advance our knowledge of the genetic architecture of local adaptation in *Drosophila* by functionally characterizing cis-eQTL and their natural, season specific effects. Identifying cis-eQTL in the context of natural environmental change is important for several reasons. Notably, we know that the laboratory is not the wild and that the physiological pressures of lab life will be different from a fly in nature. For instance, circadian rhythms vary between flies in the lab and those kept in semi-natural conditions; microbiota do as well[26] and these commensals are now understood to dramatically shape organismal physiology[27]. Because physiology directly affects life-history, developmental, and behavioral traits, GWA mapping performed in the lab may identify a limited set of functional loci and possibly reflect idiosyncratic or cryptic associations. Thus, it is crucial to measure phenotypic effects of naturally segregating polymorphisms in response to ecologically relevant, environmental variation.

To further understand the genetic architecture of rapid adaptation to temperate environments over seasonal time scales we will identify season dependent cis-eQTL. To do so, we will utilize four sets of fully resequenced and inbred *D. melanogaster* strains. These four sets of fly strains were derived from collections made in Maine (fall), during the spring and fall in Pennsylvania, and in North Carolina (the DGRP); each set contains 80-200 strains. We will inter-mate strains within population sets for a small number of generations (~4) and use these flies to seed large cages (~9 m^3) maintained outside at an experimental orchard in Charlottesville, VA. We shall rear flies outdoors for one generation, maintaining them on semi-natural food made from a standardized apple-based medium (e.g., cornmeal-molasses medium plus cored, homogenized apples). We will then harvest outdoor reared flies and transfer them directly to RNA-later for down-stream library-prep. We will perform this experiment at four time points from early spring to late fall. At each sampling time-point we will harvest 96 3-5 day old flies per sex per replicate cage (2 reps) per population set (96 flies x 2 sexes x 2 reps x 4 populations x 4 time points = ~6000 flies/year). We will replicate this experiment over the course of three years, thereby generating one of the largest and most highly powered cis-eQTL mapping experiments performed to date in any species. Furthermore, this outbreeding scheme will potentially ameliorate some of the problems of GWA mapping that arise when performed in highly inbred lines[28]. This experimental design is sufficiently flexible and efficient to allow for modifications: e.g., collecting flies at different ages, modifying larval or adult resource level or substrate.

We will subsequently extract DNA and RNA simultaneously from single flies in a high-throughput fashion using an automated liquid-handling robot; individually indexed sequencing libraries will be generated for each fly. We will use a low-cost modified version of the Nextera library preparation method[29] for both gDNA and cDNA; preliminary experiments we have conducted demonstrate that there is sufficient gDNA and mRNA obtained from individual flies to generate these libraries. We will utilize custom synthesized indexing primers[30], enabling a dual-indexing strategy that enables us to multiplex ~6000 individuals per.

In addition to affording us the ability to perform a highly replicated cis-eQTL mapping experiment among seasons and years, the use of fully resequenced inbred strains as the founders of our outbred population cages allows us to infer the full genome of each individual fly with high accuracy even with very shallow coverage per fly (~0.0001-0.001X). We will use existing analytic methods to perform this genome inference[31]. mRNA-seq libraries will be sequenced at higher coverage per fly than gDNA but still relatively low coverage compared to traditional mRNA-seq experiments. Low per-fly mRNA-seq coverage will prohibit accurate estimation of gene expression on an individual basis. But, because we will be sequencing so many individuals we will be able to combine gene expression estimates across individuals with the same allelic state at putative cis-eQTL. *in silico* pooling of individuals will therefore allow us to identify statistically significant cis-eQTL that affect total gene expression and allele-specific expression. Finally, because we will be performing this experiment across seasons we will also assess seasonal changes in gene expression.

One possible complication to our experimental design is that we will be extracting mRNA from whole flies. By averaging signal across so many tissues, signals of differential gene expression in many genes may be attenuated. To ameliorate this potential problem, we will design target capture probes to enrich our samples for genes found to be differentially expression through the season as well as endocrine genes associated with life-histories (e.g., insulin signaling genes, catecholamine synthesis genes, Toll/Imd signaling). In addition, we will design capture probes for the most highly expressed genes (e.g., Actin) and un-enrich our samples for these housekeeping genes. These approaches have been validated[32] and will allow us to

estimate more accurate gene expression levels for those genes that are weakly expressed but may nonetheless contribute to the physiological mechanisms of adaptation to temperate environments.

To follow up on these caged population experiments, we will also collect wild *D. melanogaster* throughout the Virginia area for gene expression & metapopulation structure analysis. The purpose of these experiments with wild flies is to: (1) verify that the seasonal expression measurements we obtain from the caged flies match those of flies in the wild; and (2) assess basic aspects of fine-scale meta-population structure that occurs over meters to kilometers and possibly affects the dynamics of temporally variable selection.

Daphnia. Similar to our experiments with Drosophila, we will identify cis-eQTL in *D. pulex* and asses how these loci are associated with local adaptation to predation pressure. To perform these experiments, we will first establish a new set of Daphnia clones from the ponds in which we based our initial assessment of adaptive differentiation in morphological and life-history plasticity. We will isolate 96 independent clones from each of these eight ponds and fully re-sequence their genomes (8 ponds x 96 clones = 768 genomes). The *Daphnia* genome is relatively small (~200 Mb[33]) and thus we will be able to obtain high quality (~10X) genome sequences for each clone with a modest sequencing effort.

These fully resequenced clones will form the basis of our cis-eQTL and phenotypic mapping efforts. First, we shall re-asses morphological and life-history plasticity in response to a range of either midge and fish kairomones (4 exposure levels/kairomones) following established protocols[7]. Next, we will sample ~100 individuals per clone during the 2nd instar, the developmental period when neckteeth growth begins, for mRNA-seq. Note, this experimental design differs from our Drosophila work in that we will be using multiple individuals per clone/treatment; this approach is justified because *Daphnia* are facultative parthenogens and thus clonal lineages do not suffer the same sort of inbreeding effects as isogenic *Drosophila* strains. We will generate shallow mRNA-seq profiles per clone and kairomones treatment (768 clones x 4 exposure levels x 2 kairomones types = ~6000 samples) using an automated liquid handling robot coupled with an inexpensive Nextera library preparation method and highly multiplexed dual-indexed sequencing adaptors.

Genomic, transcriptomic, and phenotypic data from these clones will allow us to map loci that modulate the plastic response to predation pressure. One possible complication in this experiment is that we will be utilizing clones from distinct populations and thus population structure must be carefully accounted for while performing genome-wide association mapping. However, we can resolve this technical problem by utilizing existing statistical methods designed to perform GWA among structured populations[34]. Furthermore, as I described above for our fly work, one possibility is that we will have limited resolution to identify differentially expressed genes or allele specific expression for weakly expressed transcripts; to ameliorate this problem, we will employ a similar enrichment/un-enrichment strategy using target capture probes.

Future Directions. The work proposed above will provide a solid foundation for addressing my ultimate goal of understanding the genetic architecture of local adaptation. The tools and analytic methods we develop over the course of this proposal will aide in future work which I will briefly describe here.

Drosophila: We are very interested in resolving whether the adaptive oscillations we observe over seasonal time scales are driven exclusively by abiotic factors such as climate or also result from intrinsic cycles that are a consequence of populations overshooting carrying capacity[35]. In flies, parental environments have been shown to modulate offspring physiology[36,37], gene expression[38], and thus likely affects life-history. Thus parental environments plausibly plays a role in cyclic adaptation over seasonal time-scales. The methods and experimental approaches we develop during the course of the proposed research will aide in studying these trans-generational effects.

Daphnia: Adaptive evolution of Daphnia in response to predator pressure likely drives adaptive evolution of phytoplankton that *Daphnia* consume. These phytoplankton (e.g., Chlamydomonas) form multi-cellular clumping phenotypes as a physiological response to *Daphnia* and also show signs of rapid and cyclic adaptive evolution in response to planktonic grazers[39]. Thus, adaptive evolution of *Daphnia* in response to predator heterogeneity through time and space may drive adaptive evolution of the phytoplankton they consume leading to an ‘ Adaptive Cascade’ (similar to a Tropic Cascade[40]). In the future, we shall test this model using sequenced clones that we develop during the course of this proposed research.