The distribution of genetic variation within and between populations reflects an interplay between mutation, natural selection, gene flow and genetic drift. While recent decades have brought clarity and concreteness to a century of questions about the origins and maintenance of genetic diversity, we lack a quantitative understanding of the factors involved. For example, what is the relative importance of Mendelian vs polygenic phenotypes for adaptation? How do correlations among phenotypes influence the course of adaptation? How are complex genetic diseases maintained and why does the genetic basis of risk differ among diseases?

Historically, work on these questions was theory rich and data poor, but that script has recently flipped, and taking advantage of new data will require new theoretical and statistical tools. My work leverages insights from theoretical population and quantitative genetics to understanding how the relationship between genotype and phenotype influences the course of adaptation and the maintenance of genetic disease. My work is generally theoretical, statistical, and computational in nature. Here I describe my past and current work in three major research areas, before outlining future objectives.

Past and Current Work

1. Polygenic Adaptation

Typically, adaptive mutations are identified in population genetic scans for unusually large allele frequency divergence (i.e. F_{ST} outliers) or patterns of haplotypic variation suggesting recent and rapid allele frequency change (i.e. selective sweeps). Our understanding of adaptation is therefore substantially biased toward phenotypes with simple genetic architectures and individual alleles with substantial effects on fitness. Although we've known for a century that many phenotypes are highly polygenic, their relative importance in human adaptation remains unclear.

During my PhD, I developed one of the first statistical methods to detect polygenic adaptation by combining data from genome-wide association studies with population genetic allele frequency datasets (Berg and Coop 2014). Applying this method in a study of 37 traits across 187 human populations (Berg et al 2017), I identified signals of polygenic adaptation predominantly in anthropometric traits, including height, infant head circumference, hip and waist circumference, and waist-hip ratio. In spite of genetic correlations among these phenotypes, I showed that patterns of divergence among populations are most consistent with selection targeting at least three separate phenotypes, most likely height, infant head circumference and waist-hip ratio, or some other closely correlated phenotypes.

To understand when and where selection acted to drive modern patterns of differentiation in genetic height among Eurasian populations, I studied patterns of genetic height differentiation in ancient DNA individuals. Here, I've shown that ancient hunter gatherer groups living in western Europe, Scandinavia, and the Eurasian steppe 8,000-15,000 years ago already show evidence of divergence in genetic height, that selection for increased height likely occurred more than once in separate times/places, and that patterns of divergence among modern populations predominantly track variation in the amount of ancestry derived from these ancient groups. I've collaborated with Fernando Racimo (University of Copenhagen) to develop a new method which allows for the explicit assignment of adaptive events to branches in an admixture graph (i.e. a population phylogeny which also allows for admixture among ancestral groups), and allows for more rigorous, model based, and principled approach to testing hypotheses about the history of selection on a quantitative trait.

I've also collaborated with Jeff Ross-Ibarra's group to show that genome size in Maize and teosinte have evolved as an adaptation to altitudinal gradients in order to optimize flowering time, and I have ongoing collaborations with Emily Josephs and Nancy Chen to study the role of polygenic phenotypes for adaptation in Maize and the Florida Scrub Jay respectively.

2. Population Genetics of Complex Disease

A major aim in human evolutionary genetics is to understand the forces responsible for maintaining genetic variation for disease. GWAS have revealed that many diseases are highly polygenic, a reality to which existing theory is not suited. As a post-doc, I have used theoretical population genetic models to study how genetic architecture depends on evolutionary parameters such as the mutational target size, distribution of mutational effects, fitness cost of disease, and the environmental variance for disease risk. Among my primary results so far is the discovery that the selection coefficients experienced by complex disease loci are, under a wide class of plausible mutation-selection-drift models, insensitive to the fitness cost of the disease, and to the environmental variance for disease risk (Berg and Sella, *in prep*). Instead, selection coefficients of individual alleles, and therefore the entire genetic architecture, depend mostly on the total strength of mutational bias toward these disease state, and also on the specific distribution of effects sizes among sites. My results also make predictions about how often we should expect to find that a derived allele at a polymorphic site increases or decreases risk (given its effect size) under polygenic mutation-selection-drift balance, and I am currently working to test whether disease GWAS datasets fit this prediction or not.

3. Selective Sweeps

Selection sweeps provide clear and striking examples of large effect alleles involved in a population's recent adaptive history. A quantitative understanding of the relative frequency of different varieties of selective sweeps is vital to our understanding of the dynamics of adaptation. The footprint left behind by a sweep depends critically on the dynamics of the involved allele(s) during the earliest phases of the sweep, providing a means to determine whether the adaptive allele arose *de novo* (a "hard sweep"), was present in the standard variation or arose multiple times during the sweep (both referred to as "soft sweeps").

As a PhD student, I developed a coalescent model to describe the population genetic footprint of a neutral allele segregating in the population that becomes beneficial and sweeps to fixation after an environmental change (Berg and Coop 2015). This work filled a decade old gap in our understanding of selective sweep footprints, highlighting how subtle differences in patterns of haplotype structure can be used to distinguish sweeps from standing variation from both classical hard sweeps and multiple mutation soft sweeps.

Future Work

My lab's work will center on leveraging genome-wide association study (GWAS) and related data to understand the evolution of complex phenotypes and diseases. Two major foci will be in understanding the role of polygenic phenotypes in local adaptation among human populations (Aim 1), and characterizing the forces which govern genetic disease architecture (Aim 2). Another major focus will be on understanding the process by which large effect alleles contribute to adaptation (Aim 3). Though these will be the major foci, my lab will pursue other directions as well, based on opportunity and on the interests of my students and postdocs.

Aim 1: Assessing the Relative Importance of Simple vs Polygenic Traits for Human Adaptation

While the population genetics of adaptation in genetically simple traits has been a focus since the beginning of the genotyping era, advances in GWAS now allow us to study how natural selection shapes the divergence of quantitative traits. Aggregating information across many loci associated with a trait, we can identify subtle, coordinated shifts in allele frequency that arise in the course of adaptation. Ultimately, I aim to parameterize and estimate evolutionary models of adaptation in both simple and polygenic traits to understand their relative importance to human adaptation. The study of adaptation in quantitative traits comes with extra challenges which need to be overcome, so most of my work will focus on traits of this type.

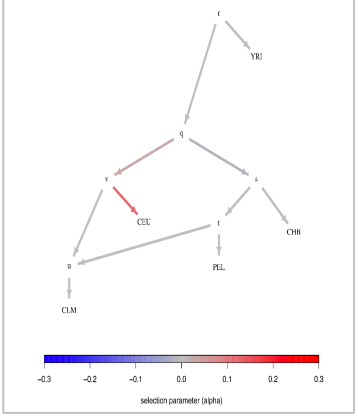
Aim 1A: Identifying Traits Targeted by Polygenic Adaptation

Pervasive pleiotropy and genetic correlation among polygenic traits means that adaptation in any given quantitative trait will likely lead to divergence in many other phenotypes. It is therefore unclear whether phenotypes which show signals of polygenic adaptation (e.g. height, infant head circumference and waist-hip ratio) represent actual targets of selection, or merely one of these correlated responses. This difficulty has been imposed largely by the limited number of well-powered GWAS datasets, but can now be overcome with resources like the UK Biobank. My lab will develop statistical methods to integrate multi-trait GWAS with the study of polygenic adaptation, and to distinguish the phenotypes most closely related to the true targets of selection from those exhibiting correlated responses to selection on other phenotypes.

Aim 1B: Inference of Time and Place of Selective Events

Most studies of polygenic adaptation have focused simply on determining that it has occurred somewhere in the evolutionary history of the populations being examined and on describing the patterns of divergence generated among modern populations. My interest is in obtaining a detailed and quantitative understanding of the process by which these patterns have been created. In which ancestral human population did selection occur? How strong was it, and how long did it persist for? Can we identify repeated, independent adaptations involving a given trait that coincide with significant transitions (e.g. dietary, geographic, climatic) in human population history? In earlier work, I have used ancient DNA to gain a rough understanding of the history divergent selection on height, and have begun developing an inference framework to interpret adaptive events explicitly within the context of human demographic history (Figure 1). In my own lab, I will apply these tools to thousands of ancient and modern genomes together with the latest GWAS data to address the questions laid out above. I will use these data to assign adaptive events to specific branches in the human evolutionary tree. Placing these inferences in a formal modeling context will allow me to estimate the strength of selection on individual quantitative traits. These methods will be general in the sense that they can be applied regardless of genetic architecture, and will therefore facilitate a direct comparison of the relative importance of simple vs polygenic phenotypes for human adaptation.

Figure 1: Admixture graph showing ancient selection on height

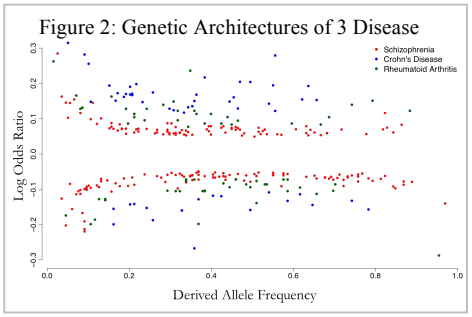


Aim 2: Understand the Evolutionary Determinants of Genetic Disease Architecture and Prevalence

Over the last decade it has become clear that many diseases are genetically complex, i.e. the heritable risk of disease arises from the combined contributions of thousands or tens of thousands of loci. GWAS have begun to partially revealed the genetic architecture of many complex diseases (i.e. number, effect size, and allele frequency of variants). However, we still lack a good understanding of the evolutionary forces that govern genetic architecture. Why do genetic architectures differ among diseases? Can mutational-selection balance alone explain the high prevalence of genetic diseases/disorders with strong fitness costs (e.g. schizophrenia, autism), and if so, are their genetic architectures consistent with this model? Alternatively, are there cryptic costs (or benefits) to low (or high) disease risk which might account for high prevalence, and what is the role of recent environmental change? How does pleiotropy shaping genetic variation for disease? I will address these and related questions through the development of biologically explicit theoretical models with techniques from population and quantitative genetics, and I will develop statistical methods to infer from GWAS the salient evolutionary parameters which govern complex genetic disease architecture.

Aim 2A: Develop generative population genetics models for the architecture of complex diseases

During my postdoc, I've extended the classic mutation-selection-drift model to complex diseases, assuming both risk increasing and decreasing mutations arise continuously in the population, and that the selection acting on them results from their combined impact on a disease with an explicitly specified fitness cost. As described above, I was able to show that selection on disease causing alleles does not depend on this fitness cost, and to relate the strength of mutational bias and distribution of effects on risk to the disease prevalence and genetic architecture. However, other factors are likely to play and important role. Notably, extensive polygenicity again implies the importance of pleiotropy. Such pleiotropic effects will impact the distribution of selection coefficients experienced by disease loci, and therefore play an important role in shaping genetic architecture. While there has long been speculation about the role of pleiotropy in driving disease prevalence, so far we have only verbal or *ad hoc* models. I will extend my modeling work to study the impact of pleiotropic costs or benefits to disease liability, and to understand the impact of pleiotropically related quantitative traits under stabilizing selection.



Aim 2B: Infer the Evolutionary Parameters of Complex Diseases from GWAS data

The theoretical predictions of the aforementioned models can be tested against the results from GWAS to learn about the forces that shape genetic variation for the risk of different diseases. Moreover, these data can be used to infer the basic evolutionary parameters, e.g., the mutational target size, distribution of effects on disease and on pleiotropically related traits. To learn about these parameters and how they shape genetic disease, I will develop composite likelihood statistical methods based on Poisson random field approach. My aim is that these inferences will help us to understand the factors which drive variation among diseases in their genetic architectures, and to deliver conclusions to some old debates about why genetic disease persists. Moreover, knowledge of these parameters will inform more practical questions, e.g. by forecasting the expected number of new disease loci discovered as samples sizes are increased, and in the development of disease risk prediction models for personalized medicine.

Aim 3: Inferring the Dynamics of Adaptation by Large Effect Alleles

A population can adapt to a change in the environment in multiple ways. If standing variation is present at many loci affecting the trait, adaptation may involve small changes to allele frequencies at many loci and thus will be polygenic. Otherwise, adaptation may proceed by large changes in frequency at few loci with large effects (see also Aim 1). I am interested in understanding the dynamics of the process by which large effect alleles contribute to adaptation, and on how these dynamics depend on population size. Do populations adapt from standing variants present at low frequency when the environment changes, or must they wait for a new mutation to arise? Are there typically one or many alleles at a given locus that can fulfill the adaptive requirement? Adaptations involving large effect alleles leave characteristic patterns of genetic variation in their wake, which can be used to identify them. During my Ph.D., I used coalescent theory to model these different scenarios, and showed how, in principle they could be identified based on their effects on haplotype structure (Figure 3). I my own lab, I will develop the statistical machinery to do so in practice, and to obtain genome-wide estimates of the rates of adaptation via these different mechanisms. To understand how these dynamics depend on population size, I will apply these methods in a diverse array of taxa, and learn how the dynamics of adaptation vary across the tree of life.