Graham Coop
Statement of Current and Future Research

My research focuses on understanding the evolutionary forces that have shaped genetic differences between individuals, populations, and closely related species utilizing a blend of human genetics, statistical analysis, and theoretical model development. My research falls into the following broadly overlapping areas: inference of recent migration; the genome-wide effects of recent selection; the causes and consequences of variation in recombination rates. My work on these topics is detailed below.

Recent ancestry and migration.

All individuals with a species are related to each other to varying degrees. Population genetics has traditionally concerned itself with small samples of individuals who at most loci are genetically related to each over tens or hundreds of thousands of generations ago. This focus has typically divorced population genetics from the study of the dynamics of contemporary populations. However, this distant genetic relatedness in small sample belies the fact individuals are much closer related in a genealogical sense. Each of us has 2^k ancestors k generations ago, a number that grows so rapidly that all humans are predicted to be at least 10-20th cousins. We are usually unaware of this close level of relatedness as these recent ancestors rarely contribute genetic material to us.

In very large genome-wide datasets we can see rare events where subsets of individuals share regions of their genome identical by descent (IBD) over large regions indicating a very recent common ancestor. This knowledge has the potential to revolutionize our view of natural populations, and allow us to estimate the aspects of the fecundity, demography and the dispersal of individuals over the past tens of generations. A major new direction for my work is building an understanding of the recent genealogy of populations from such data. We are working to use measures of IBD in humans to learn how close relatedness drops off within populations and among geographical separated populations. Our aim is to develop a framework to estimate population sizes and migration rates over the past tens of generations using this information. As genomic tools become available for many non-model systems we will bring these tools to bear on field studies of evolution and ecology.

Natural Selection.

While selection clearly shapes much of the phenotypic diversity within and between species, its role in molecular diversity and divergence has been long been debated. A dominant hypothesis for many years was that much of molecular diversity is neutral (or very weakly deleterious), with levels of diversity representing a balance between mutation and genetic drift. However, this view has crumbled in light of evidence in many species that a large percentage substitutions being driven by selection and that even at putative neutral sites levels of genetic diversity are likely shaped by the effect of selection at linked sites.

A major component of my work since my PhD (e.g. Coop and Griffiths 2004, Spencer and Coop 2004) has focused on what we can hope to learn about natural selection from genome-wide

polymorphism data, and how we should develop our population genetic models of molecular evolution in light of this information.

I led an analysis to learn more about the geographical distribution of selection pressures faced by humans as they moved around the world using genome-wide polymorphism data (Coop et al 2009,Pickrell et al 2009). Despite the evidence that selection shapes patterns of diversity, we found that natural selection has been surprisingly ineffective at fixing variants between regions and that alleles rarely sweep to fixation across broad geographic regions (Coop et al. 2009).

Based on these observations, one of the themes that emerged in my lab's work is the importance geography and the polygenic nature of adaptation in studying population genetic data (Pritchard, Pickrell and Coop 2010).

The role of geography. Most of our simple population genetics models of adaptation, and its effects on linked sites, ignore the complications of population structure. Given the ability to generate large population genomic datasets there is now a pressing need to address this shortfall in order to understand the geography of adaptation.

In collaboration with my postdoc Peter Ralph I have developed theoretical results that suggest geographically separated populations may readily adapt in parallel through new mutations rather than waiting for the spread of the allele by migration. (Ralph and Coop 2010). This suggests it may be rare for a single allele to impact an entire species and that the response to selection may frequently occur through partial selective sweeps. We are pursuing this line of research to understand under what circumstances we expect spatially spread populations to adapt in parallel when selection pressures differ across the species range. This work will be critical to forming a theoretical framework to understand current efforts to map the genetic underpinnings of adaptation to similar ecological conditions across a species range.

Our models of the effect of selective sweeps on linked neutral diversity are also often reliant on an assumption on panmixia. We will extend coalescent models of selective sweeps to account for the spread of a selected allele in a continuous habitat through local dispersal. These models will allow us to develop methods to use patterns of linked diversity to estimate the rate at which selected alleles spread across landscapes and where selected mutations arose.

Finally, we are developing tools to search for loci underlying adaptation to similar environments across a species range. For example, we recently published a Bayesian method to identify selected loci whose frequency correlates more strongly with environmental variables than expect due to migration and shared history (Coop et al 2010). These methods are still in their infancy, and considerably work is needed to fully develop their potential.

The genome-wide impact of selection on linked selection. Despite this revolution in our understanding of molecular variation, our models of population genomic data are still very much rooted in null models of genetic drift or providing simple alternative models to this null model. For example, one of the best-studied alternative models of linked selection is predicated on alleles sweeping from introduction to fixation (the selective sweep model). However, it may be relative rare for advantageous new alleles to sweep rapidly to fixation because fluctuating environments

and changing genetic backgrounds may alter selection pressures (Coop et al 2009, Hernandez et al 2011).

If we are to reach a deeper understanding of the impact of linked selection on genomic patterns of diversity, we need to develop a broader set of models that encompass a fuller range of the effects of linked selection. We are working on coalescent models where alleles recurrently sweep into the population but not necessarily to fixation. These models significantly broaden the qualitative predictions of sweep models further complicating efforts distinguish between models of linked selection. In the face of this difficulty, and with the availability of full population genomic data, I believe that rather than focusing on specific models we should concentrating on moving towards a fuller empirical description of the effect of linked selection on genealogies. To push this idea forward we will work on developing the inference framework to estimate the parameters of a general coalescent model of linked selection. Ultimately this work will provide a flexible framework to facilitate the comparison of the genomic effects of linked selection across species to understand the interplay of genomic environment, and population size with selection regimes.

Recombination

In humans and many other organisms, recombination plays a central role in ensuring the segregation of chromosomes during meiosis, and in generating novel combinations of alleles for natural selection to act upon. Surprisingly recombination rates are highly variable within and between species. On a fine-scale, in many species, recombination events are tightly localized to recombination hotspots, which are often polymorphic between individuals and evolve rapidly between species. As yet, we understand relatively little about the evolutionary forces underlying this variation. I have made a number of contributions to characterizing patterns of recombination, understanding its genetic basis, modeling its evolution, and evaluating the consequences for molecular evolution.

The recent availability of large amounts of genetic variation data from human populations offers an unprecedented opportunity to study the evolutionary genetics of transmission. We have characterized genome-wide fine-scale patterns of crossover in pedigrees (Coop et al. 2008) allowing us to localize recombination events genome-wide, down to the individual hotspot level. We documented extensive heritable variation in recombination rates at all scales, in particular variation in the genome-wide use of recombination hotspots.

I co-lead a study to characterize broad-scale patterns of recombination to better understand the constraints on recombination in humans (Fledel Alon *et al.* 2009). Our analysis showed that a requirement for one crossover (chiasmata) per chromosome may not be absolute, as we found evidence that chromosome 21 seems to be frequently transmitted properly in the absence of a crossovers in females (i.e. achiasmatically). This finding raises the intriguing possibility of an achiasmatic back-up mechanism in mammals aiding in the correct segregation of chromosomes when chromosomes fail to form chiasmata.

The genetic basis of recombination variation. In 2009 I began a collaboration with Bernard de Massy's group at CNRS, France. They work on the molecular biology of mammalian meiosis and hotspot specification and along with others had previously localized a genomic region has harboring a QTL locus for the control of recombination hotspot localization in mouse (Grey et al

2009, Parvanov et al. 2009, Parvanov et al. 2010). The QTL region contained a candidate gene for hotspot control, the zinc finger (ZF) binding protein Prdm9.

Working with them we helped to demonstrated Prdm9's role in genome-wide hotspot specification, and that Prdm9 was predicted to bind a previously identified human genome-wide hotspot motif (Myers *et al.* 2008). Remarkably, Prdm9 was highly polymorphic in humans and we were able to show that this polymorphism was associated with our previously identified heritable variation in genome-wide hotspot usage. This work (Baudat et al 2010) appeared back to back with another paper that provided highly complementary evidence of the role of PRDM9 in hotspot specification (Myers *et al.*, Science 2010). Together this pair of papers demonstrated that the remarkable polymorphism and divergence of Prdm9 across the mammalian phylogeny can explain in part the rapid evolution of fine-scale recombination rates.

This work on Prdm9 dovetails nicely with my earlier theoretical work studying the processes underlying hotspot evolution (Coop and Myers 2007). Simon Myers and I developed models that showed that biased gene conversion in favor of alleles that locally disrupt hotspot motifs (Boulton et al. (1997), Pineda-Krch and Redfield (2005)) would rapidly drive hotspot motifs from mammalian genomes. We predicted that this would force the rapid evolution of hotspot determinants in *trans* (see also Peters 2008) and Prdm9 seems to beautifully fulfill this role. Although it is unclear whether compensating for hotspot loss is the cause or merely a pleiotropic consequence of Prdm9's rapid evolution.

Using much larger pedigree datasets for the Hutterites, the Framingham Heart and the AGRE studies (hundreds of pedigrees typed on 500k markers) we have now localized ~260,000 recombination events in over 7000 meioses (Fledel Alon et al 2010). Using these data we have shown that that a substantial proportion of the additive genetic variance in broad-scale recombination rates is sex specific. We conducted a genome-wide association study to identify the genetic determinants of polymorphism in recombination phenotypes which uncovered a number of novel candidates, and replicated previously reported loci, while casting significant doubt on previous signals.

Sex Specific Recombination rates. One of the most striking patterns in recombination rates across species is the sex-specificity of broad-scale recombination rates. One broad pattern is that across many taxa females tend to recombine more than males, especially close to the centromeres, regardless of the identity of the heterogametic sex. Despite the large body of work devoted to studying the evolutionary advantages of recombination, little attention had been directed towards explaining these patterns. My postdoc, Yaniv Brandvain, and I have proposed that a possible explanation for the sex-specificity of recombination is that changes in female recombination rates can act as potent suppressors/enhancers of female meiotic drive. This follows from the fact that female recombination determines the pairing alleles during meiosis, and their transmission to the egg, and so strongly influences the efficacy of meiotic drivers. We have developed a series of models that articulate the conditions under which meiotic drive can lead to the evolution of high rates of recombination in females, and how the role of recombination in female meiotic drive could explain a number of puzzling features of sex-specific recombination.