Research Accomplishments

Graham Coop, Dept. of Human Genetics, University of Chicago

The current explosion in new data sets and data types offer unparalleled opportunities to explore evolution. These advances naturally require the development of new models and methodologies. I am interested in using careful and inventive analysis to gain intuition about evolutionary processes. To do this, I take a data driven approach, either by direct data analysis or by formulating new models based upon our current understanding of a biological process. I have used a wide variety of methods to do this, including developing new population genetic theory and advanced computational statistical techniques.

My current work focuses on a variety of subjects that can be grouped into three broad areas: recombination, population history, and selection. This grouping is somewhat arbitrary as many of my projects fall in more than one of these areas, but I shall present them as such.

Selection

For my D.Phil. (with my supervisor Prof. Bob Griffiths), I worked on methods to detect positive selection. We developed a method to calculate the likelihood of the selection parameter of a particular mutation using the full information from a non-recombining region (Coop and Griffiths, 2004). To do this, we developed analytic results for population genetic coalescent and diffusion theory and various computationally intensive statistical techniques. We also created a novel method to simulate data under models of positive selection. With Chris Spencer, I incorporated these simulations techniques into a flexible simulation program (SelSim, Spencer and Coop (2004)) that has been used by numerous other researchers to explore models of positive selection.

Much of our current understanding of the effect of positive selection on patterns of diversity is based on a model that assumes that selection acts on a newly introduced co-dominant marker in a constant-sized population. While this model has many features that make it attractive to work with, these assumptions will often be invalid. In collaboration with Molly Przeworski, Jeff Wall and Kosuke Teshima, I have worked to explore how departures from these assumptions might affect the patterns left by selection (Teshima et al., 2006; Przeworski et al., 2005).

In addition, I am interested in developing new methods to allow researchers to identify the action of selection. A powerful test of local adaptation is to examine the geographical distribution of the frequency of an allele. However, there are no methods to judge the significance of such patterns, when allele frequencies have been measured in a large number of populations. I am working with Jonathan Pritchard and Anna Di Rienzo on devising such a test. Using a control set of markers, we estimate a fully Bayesian null model of the covariance of the frequency of an allele between populations. We then use this null model to test for local adaptation at a SNP of interest. We are currently using this method to assess signals of selection across 52 human populations.

Recombination

Recombination is a fundamental component of meiosis, required to help ensure that daughter cells receive the correct complement of chromosomes. In addition to its mechanistic function, recombination also generates new combinations of alleles on which natural selection acts. Surprisingly given the biological importance of recombination, the recombination landscape appears to evolve quickly on all genomic scales. A broad interest of mine is in understanding the causes and consequences of this evolution (Coop, 2005; Coop and Przeworski, 2007).

Recombination does not occur uniformly along the genome and is instead quite spatially heterogeneous on all scales. Recent work has shown that it is highly heterogeneous even at fine scales, with most recombination events concentrated into 1-2 kb hotspots. Interestingly, humans and chimpanzees do not share hotspot locations, despite 99% identity of their genomic sequences. The reason for this relatively fast evolutionary rate is

unclear but could lie with the observation that, as a result of the mechanism of recombination, alleles that disrupt hotspots are overtransmitted to offspring (Boulton et al., 1997; Pineda-Krch and Redfield, 2005). With Simon Myers (now a post doc. at the Broad Inst.), I developed coalescent and diffusion based population genetic models of hotspots incorporating this biased transmission (Coop and Myers, 2007). We found that this biased mechanism can account for rapid hotspot turnover, and in fact makes it very difficult for new hotspots to arise. The fact that we do see hotspots in the human genome leads us to propose that alleles that introduce hotspots must be initially shielded from this biased transmission. We also explored the possibility of detecting alleles that influence the activity of a hotspot; perhaps counter-intuitively, we found that such alleles will be hard to detect from population genetic data.

Population History

Understanding the recent genetic history of populations is critical to a wide number of subject areas, including phylogeography, anthropology and the study of speciation. I have a strong interest in developing methods to help researchers robustly leverage information about historical events from genetic data.

Despite the many data sets available to study the history of recently diverged populations, there are few methods that estimate parameters for models of population divergence. None of the current methods can incorporate intra-locus recombination, which is a serious limitation as it forces researchers to disregard information and can lead to potential biases. I am currently working with Dan Davison (formally a University of Chicago PhD student, now a PostDoc in the Oxford Statistics Dept) to develop a method to use recombining regions to estimate the time at which gene flow ceased between two populations. To do this, we have modifying the recently developed method of Li and Stephens (2003) that uses a Hidden Markov model to approximate the likelihood of data sets under the coalescent. In addition, the Hidden Markov method allows us to colour stretches of haplotypes by the probability that they arose in the ancestral population allowing investigators to visualize patterns of haplotype sharing between populations.

I was involved in a major survey undertaken by Pritchard and Rosenberg

groups of worldwide patterns of human haplotype variation (Conrad et al., 2006). The study furthered our understanding of the relationships between the various human populations, showed that the positions of hotspots are likely conserved across populations and hence over recent time scales, and confirmed the utility of the Hapmap in design of the association studies in many human populations.

References

Boulton, A., R. S. Myers, and R. J. Redfield (1997, Jul). The hotspot conversion paradox and the evolution of meiotic recombination. *Proc Natl Acad Sci U S A 94*(15), 8058–8063.

Conrad, D. F., M. Jakobsson, G. Coop, X. Wen, J. D. Wall, N. A. Rosenberg, and J. K. Pritchard (2006, Nov). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet 38*(11), 1251–1260.

Coop, G. (2005, Dec). Can a genome change its (hot)spots? *Trends Ecol Evol 20*(12), 643–645.

Coop, G. and R. C. Griffiths (2004, Nov). Ancestral inference on gene trees under selection. *Theor Popul Biol 66*(3), 219–232.

Coop, G. and S. Myers (2007). Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet*. Early Online Access.

Coop, G. and M. Przeworski (2007, Jan). An evolutionary view of human recombination. *Nat Rev Genet 8*(1), 23–34.

Li, N. and M. Stephens (2003, Dec). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics 165*(4), 2213–2233.

Pineda-Krch, M. and R. J. Redfield (2005, Apr). Persistence and loss of meiotic recombination hotspots. *Genetics 169*(4), 2319–2333.

Przeworski, M., G. Coop, and J. D. Wall (2005, Nov). The signature of positive selection on standing genetic variation. *Evolution Int J Org Evolution 59*(11), 2312–2323.

Spencer, C. C. A. and G. Coop (2004, Dec). SelSim: a program to simulate population genetic data with natural selection and recombination. *Bioinformatics 20*(18), 3673–3675.

Teshima, K. M., G. Coop, and M. Przeworski (2006, Jun). How reliable are empirical genomic scans for selective sweeps? *Genome Res 16*(6), 702–712.