Statement of Research Interests

Graham Coop, Department of Human Genetics, University of Chicago

Recombination

My work on recombination will follow two main directions. The first is to learn about variation between individuals in fine scale rates of crossing over and the determinants of hotspot location. The second is to learn about the impact of recombination on the rate of adaptation.

To date what we know about the fine scale variation in human recombination rates comes from study of a few small regions using sperm typing and genome-wide linkage-disequilibrium studies. While these techniques have revolutionized our view of the recombination landscape in humans, they leave us with a far from complete picture. Sperm typing only informs us about male rates in small regions, while the extent that LD-based methods are affected by selection is unknown. Neither sperm typing nor LD-based methods can inform us about genome-wide variation between individuals or about the heritability of different aspects of recombination. To address these questions, we are undertaking a study of crossing over using a high density SNP chip (500,000 SNPs) in a very large human pedigree. With this density of markers, we can pinpoint the location of crossover to within 10s of kb, allowing fine scale rates of crossing over to be studied. This resolution allows us for the first time to use pedigree data to explore hotspots. In collaboration with Molly Przeworski, I am using this data to learn about how the use of hotspots differs between the sexes and individuals. I am also studying the heritability of both broad and fine scale features of recombination rate in this large inbred pedigree.

One of the main evolutionary consequences of recombination is thought to be that it improves the efficiency of selection, by uncoupling the fates of mutations at linked loci. While this effect has been explored in Drosophila

(Betancourt and Presgraves, 2002; Presgraves, 2005) the relationship between the efficiency of selection and recombination has not been investigated in mammals. I intend to undertake study of this evolutionary important relationship using various mammalian genomes and genetic maps, to explore the correlation between rates of divergence rates at coding sites and recombination.

Population history

Recently I worked as part of a collaboration with the Rubin Group (DOE genome centre, Berkley) on the first Neanderthal genomic DNA (Noonan et al., 2006). I developed novel simulation-based techniques to estimate the time at which the Neanderthal and human populations split (400,000 years) and to explore the possibility of secondary contact between the two species (no evidence but more data is needed). Jonathan Pritchard and I are now working with Svante Paabo's group preparing for the analysis of the Neanderthal genome.

Increasingly large data sets are starting to become available for a range of species, yet most current methods are inadequate to deal with such data sets as they are either computational inefficient or make unreasonable simplifying assumptions. A strong component of my future research will be developing methods to handle inference of selection and population history from large data sets. As a first step, motivated by my work on Neanderthals, I intend to create fast simulation-based techniques to utilize large data sets to learn about speciation and population history.

Selection

I intend to continue my work on broadening our models of positive selection. There are currently no programs available to allow researchers to explore models of positive selection for multiple populations. Given that inter-population comparisons are the basis of one of our most common test statistics of selection (Fst), the absence of a publicly available program to assess the reliability of methods using inter-population comparisons is surprising. I would like to produce a flexible simulation tool to allow researchers to simulate from such models. This will require the development of new

methods to simulate from diffusion based models of population divergence.

Rates of non-synonymous divergence are often used to identify genes that are undergoing adaptation. Recent work has focused attention on discordances between gene trees and species trees (Pollard et al., 2006; Patterson et al., 2006; Hobolth et al., 2007) implying considerable variation in coalescence time in the ancestral populations of closely related species. This variation in tree topology and coalescence time will impact divergence based tests of selection. With the increasing availability of multiple closely related genomes (e.g. species Drosophila and Human, Chimpanzee and Gorilla), there is a need to incorporate coalescent-based models of closely related species into divergence based-tests for selection. Therefore, I intend to create a Hidden Markov method to perform tests of selection from the genomes of closely related species where divergence times and tree topology vary along the sequence.

References

Betancourt, A. J. and D. C. Presgraves (2002, Oct). Linkage limits the power of natural selection in Drosophila. *Proc Natl Acad Sci U S A 99*(21), 13616–13620.

Hobolth, A., O. Christensen, T. Mailund, and M. Schierup (2007). Genomic relationships and speciation times of human, chimpanzee and gorilla inferred from a coalescent Hidden Markov Model. *PLoS Genet*. Early Online Access.

Noonan, J. P., G. Coop, S. Kudaravalli, D. Smith, J. Krause, J. Alessi, F. Chen, D. Platt, S. Paabo, J. K. Pritchard, and E. M. Rubin (2006, Nov). Sequencing and analysis of Neanderthal genomic DNA. *Science 314*(5802), 1113–1118.

Patterson, N., D. J. Richter, S. Gnerre, E. S. Lander, and D. Reich (2006, Jun). Genetic evidence for complex speciation of humans and chimpanzees. *Nature 441*(7097), 1103–1108.

Pollard, D. A., V. N. Iyer, A. M. Moses, and M. B. Eisen (2006, Oct). Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genet 2*(10), e173.

Presgraves, D. C. (2005, Sep). Recombination enhances protein adaptation in Drosophila melanogaster. *Curr Biol 15*(18), 1651–1656.