Outline of lectures 20-21

Trees of Genes – The Coalescent.

1. What happens to molecular evolution as it gets down to looking at differences within populations? We can trace a tree of ancestry for the copies (not the individuals, but the copies) within a population.

2. The most highly publicized such tree came in 1987 when Cann, Stoneking and Wilson made a tree of human mitochondria. Mitochondria are effectively haploid, and inherited only from the mother. They saw three patterns

   - All of the 149 mitochondria they looked at were descended from a single female, who has been named Mitochondrial "Eve".
   - She lived about 200,000 years ago (with a large uncertainty about that, maybe $\pm 100,000$ years).
   - She probably lived in Africa. (This is weakly supported by the data).

3. There was a big reaction to their study because in part it resonated with notions of Adam and Eve, and with earth-mother-godesses. Nevertheless it is inevitable that any small region of DNA will be descended from one ancestor, so there will be Cytochrome Herberts and Hemoglobin Friedas as well. Coming from one ancestor is not just a property of mitochondrial DNA.

4. If we trace each gene back to the gene in the previous generation that it was copied from, and continue to trace back generation by generation, the lineages randomly converge. Ultimately there is only one gene copy that is the ancestor of everyone. (We are assuming no recombination – for that see below).

5. The English probabilist (and science administrator and University head) (Sir) J.F.C. Kingman showed in 1982 what the random process of formation of a tree of lineages is expected to look like. He called such a random tree "the n-coalescent" and the name "coalescent" has stuck. The lineages combine at random, it taking longer and longer for them to combine as the go back. The whole process takes an average of about $4N_e$ generations, and the last two lineages (going backwards) take half of that.

6. The process is like "bugs in a box". We have a box full of hyperactive, indiscriminate, voracious, and insatiable bugs. They run around and collide at random. When two bugs collide, one eats the other and then resumes running. This process in fact has exactly the same mathematics as the coalescent: the number of bugs drops rapidly at first, the more slowly as there are fewer bugs to collide with.

7. A sample of 149 lineages is overwhelmingly likely to have the whole population's gene ancestor as its root. So mitochondrial Eve is likely to be the mitochondrial ancestor of everyone.

8. For mitochondria you take off the 4: it takes about N_e generations, where N_e is the effective population size. That means that with a human generation time of 25 years or so, mitochondrial Eve was surprisingly recent unless human population sizes were about 12,000, which is rather small.

9. In the mitochondrial tree the European and Asian sequences are all jumbled together with some of the African sequences. They form a group (clade). The rest of the African sequences split off on both sides of the root. The time at which the European and Asian sequences start diverging is about 100,000 years ago. (These times are based on molecular clock calculations based on the numbers of differences between the sequences).

10. This suggests the Out Of Africa hypothesis: that a random subset of African mitochondrial lineages left Africa about then, entering Europe and Asia (presumably through the Middle East), and becoming the ancestors of the European and Asian sequences.

11. What is very surprising about all this is that there were already *Homo erectus* populations throughout Europe and Asia at that time (as well as African ones). Yet there is no sign that mitochondria from those *Homo erectus* got incorporated into the current human population.

12. The alternative is the Multiregional Hypothesis. All the *H. erectus* populations exchange genes, and the innovations that arise in one area diffuse to all others. No one region is then the place that *Homo sapiens* evolved, and the *H. erectus* populations do not go extinct – they are transformed into modern humans. This is in fact what we would have expected to happen – the Out Of Africa hypothesis is a bit strange compared to it.

13. It is still early days yet, but a number of studies with other pieces of DNA (other loci in the nuclear genome) suggest that Out Of Africa is right. But the evidence is still quite weak and one should not be too ready to jump to conclusions when we have such an important question.

14. The ancestors of different regions of the genome occurred at very different times and at different places. Mitochondrial Eve and Y-chromosome Adam did not know each other.

15. The above picture is true if the gene does not recombine. If it recombines, the lines do not only converge, they split as one goes back, every time there is a recombination within the locus. (Mitochondria don't recombine). More properly, the genes at one end of the sequence then have a slightly different tree from the ones at the other end. As one "walks" along the genome, the tree gradually changes.

16. Calculations of how much recombination will happen on a lineage that is going back to a root $4N_e$ generations ago suggest that the tree will be very different if one goes a distance along the genome such that $4N_e r > 1$, or $r > 1/(4N_e)$. For humans that is a surprisingly small distance. If $N_e = 100,000$ it is the distance one needs to go to get 0.0000025 recombination. That is about 250 nucleotides (if one assumes recombination is evenly spread along the genome. Actually it is somewhat clumped, with "hot spots" and cold spots so that maybe 1,000-2,000 nucleotides is the distance at which one can expect very different new tree. But even that means that there are over 1,000,000 different gene trees for our genome!

17. That means lots of ancestors, of both sexes and in many different generations, contributed to our gene pool.

18. Population growth affects gene trees by making coalescence be fastest when one gets back to periods in which the population size is small. However studies in our lab and others suggest that it will take a lot of loci to get a clear picture of past population sizes.

19. Migration also affects gene trees. The more there is, the less consistency one will see between the present location of populations and their placement on the tree. The amount of migration needed to scramble the placement of individuals on the gene tree is about $4N_em = 1$. That means that $m = 1/(4N_e)$, so only 1/4 of a migrant individual arrives in each population each generation. (Or less gruesomely, one arrives about every 4 generations). Higher rates of migration than this mere trickle will homogenize the genes and lead to the species behaving like one big random-mating population.

20. Recently work by Svante Pääbo and others has resulted in two mitochondrial sequences from Neanderthals. These turn out to be substantially different from modern humans, which is surprising. They imply that much of Neanderthal ancestry diverged from that of modern humans, but they do not rule out that a minority fraction of human ancestry could come from Neanderthals. Coalescent arguments are used to figure out what fraction of our genome could come from Neanderthals and still be consistent with these sequences. As more loci are sequenced from Neanderthals, one would finally expect to see some that would be closer to modern humans, if there is a Neanderthal contribution to our amncestry.

21. The same phenomena occur in all species. The estimation of what all these population properties were is going to require work on lots of loci, and this is barely under way. Stay tuned.