

## Positive Selection in Humans

This lecture provides some examples of how DNA sequence data have been used to identify genes or regions of the genome that have undergone recent positive selection in humans. Some of these examples are strongly supported and are generally accepted by evolutionary biologists. Others are more speculative and controversial. The point of this lecture is not to dwell on specific cases, but rather to explore the kind of information we can get from molecular evolutionary studies and how it can be used, in combination with other data, to infer something about the evolutionary history of humans.

### 1. Unusual patterns of polymorphism

Recent positive selection should lead to strong haplotype structure in the region of the genome surrounding the target of selection. In general, one looks for cases where there is one (or a few) haplotypes that extend over a long stretch of sequence and are in high frequency in the population. Below are some examples of genes that show such a pattern.

Gene/Protein	Selective Force	Conclusion (?)
Beta-globin	Pos = resistance to malaria in heterozygotes; Neg = sickle-cell anemia in homozygotes	Balancing selection
G6PD and CD40L	Pos = resistance to malaria; Neg = ?	Positive selection
CCR5	Pos = Resistance to HIV (AIDS) in heterozygotes, resistance to smallpox, other diseases?; Neg = ?	Balancing selection?
ADH	Pos = acetaldehyde lethal to blood parasites? Neg = inability to drink alcohol?	Positive selection
Lactase	Pos = ability to drink milk as an adult	Positive selection
DRD4	Pos = curiosity, exploration?; Neg = ADHD?	Positive selection?

### 2. Convergent evolution of lactase persistence in European and African populations

LP = lactase persistence, the ability to digest lactose (milk) as an adult.

The frequency of LP varies in human populations. It is highest in northern Europe:

- > 90% in Swedes/Danes
- ≈ 50% in Spain, Middle East
- < 20% in China, west Africa

but LP is also frequent (50-90%) in some African pastoralist (cow-raising) populations.

In Europeans, a single C to T nucleotide change about 14-kb upstream of the lactate gene (LCT) is thought to increase expression in adults and has risen to high frequency due to positive selection for ability of adults to drink milk.

In East African cattle-raising populations (Sudan, Kenya, Tanzania), three different mutations in the LCT upstream region are in high frequency and appear to confer LP. The European mutation is absent from these populations. Thus, LP has evolved independently in the European and African populations. In both cases, there is evidence for positive selection of LP alleles within the past 8,000-9,000 years (Europe) or 3,000-7,000 years (Africa).

### 3. Amylase gene copy number and high-starch diets

Sometimes genes are present in different numbers of copies in different individuals. This is known as copy number variation (CNV).

One gene that has been found to vary in copy number is AMY1, which encodes the salivary amylase enzyme. This enzyme is responsible for starch hydrolysis.

Looking across human populations, there is a positive correlation between AMY1 copy number and amount of starch in the diet, with agricultural populations having high AMY1 copy number.

This suggests that positive selection has favored higher AMY1 copy number (and higher salivary amylase activity) in populations with high-starch diets.

### 4. A common inversion under selection in Europeans

Inversions are a type of chromosomal rearrangement. They are sometimes polymorphic within species.

As an example, imagine a chromosome divided into 10 linear segments A-J:

Chromosome from individual 1: A B C D E F G H I J

Chromosome from individual 2: A B C H G F E D I J

The segment containing DEFGH is “inverted” in individual 2.

There is an inversion on Chromosome 17 that is polymorphic in humans. The inverted region spans 900 kb and contains several genes including MAPT, which is associated with Parkinson’s disease.

The most common form in humans is known as H1. The inverted form is known as H2 and is very rare in most populations, except in Europe where it is in high frequency ( $\approx 20\%$ ). This suggests that selection has favored H2 in European populations.

In Iceland, there are extensive genetic records and records of medical history. Using data from 30,000 Icelanders born between 1925-1965, it was found that women with the H2 inversion had more children (on average 3% more) than women with H1. This supports positive selection. The underlying reason for the reproductive difference is unknown.

Recently, it was found that the H2 inversion is also present in chimpanzee and orangutan (where it is also polymorphic with H1). It is now thought that H2 is the ancestral form and an inversion to H1 occurred independently in human, chimp, and orangutan. Perhaps selection favored H1 in ancestral populations, but later favored H2 in Europe?

### 5. Prion disease and human cannibalism

Prion diseases are caused by misfolded proteins. Examples include Creutzfeldt-Jakob Disease (CJD) in humans, Bovine Spongiform Encephalopathy (BSE) in cows (also known as “mad cow” disease), and Scrapie in sheep.

A prion disease known as Kuru affects the Fore population of Papua New Guinea. It is thought to be caused by cannibalism, as the Fore had a custom of eating their dead relatives. The disease affects mainly women and children (adult men did not participate in the cannibalism).

Australians banned cannibalism in the mid-1950's and there has not been a reported case of Kuru in an individual born after 1950.

Polymorphism in the human protein PRNP is associated with CJD and Kuru. There is a methionine /valine polymorphism at codon 129. Homozygotes are susceptible to the disease, heterozygotes are resistant.

There is a significant excess of heterozygotes at codon 129 among Fore women over 50 years old. Allele frequencies of V and M are 55% and 45%, respectively. The expected genotype frequencies under HWE are: VV=30%, VW=50%, MM=20%. For Fore women over 50, 77% were heterozygotes.

Fore under 50 and Europeans are in Hardy-Weinberg equilibrium (HWE). This suggests strong balancing selection in the Fore population exposed to cannibalism.

Tajima's  $D$  at PRNP is significantly positive (suggesting balancing selection) in several worldwide populations, including Europe, Africa, and Papua New Guinea. This has led to the hypothesis that cannibalism was common in many prehistoric human populations.

#### 6. Selective advantage of FGFR2 mutations in the male germ line

Mutations in the gene FGFR2 cause Apert Syndrome (AS; malformation disease). This disease is dominant and almost always affects children of unaffected parents. Thus, it must be caused by new mutations.

AS is almost always inherited from the father and its frequency increases with the father's age. Most cases are caused by the same C → G mutation at position 755. Because AS affects 1 in 70,000 births, the mutation rate at this position must be ≈500-fold higher than normal.

Analysis of sperm samples indicated that C → G mutant sperm were present in all males tested and the frequency was positively correlated with age. In men over 50, the mutant frequency was 1 in 10,000. This correlates well to the rate of the syndrome.

Why is this mutation present at such high frequency? The mutation may cause spermatogonia (sperm stem cells) to reproduce at a higher rate. Thus, it is positively selected within the testes. Over time, more mutant spermatogonia are present and produce more mutant sperm. This may be why the mutation rate appears to be so high and why it increases with age.

This appears to be a case of evolutionary conflict, where a mutation that is positively selected in the male germ line has a negative effect on organismal fitness.

## References and Additional Reading

News Focus – Seeking the signs of selection. *Science* 298: 1324-1325.

Tishkoff, S. A., *et al.* (2006) Convergent evolution of human lactate persistence in Africa and Europe. *Nature Genetics* 39: 31-40.

Burger, J., *et al.* (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. USA* 104: 3736-3741.

Stefansson, H., *et al.* (2005) A common inversion under selection in Europeans. *Nature Genetics* 37: 129-137.

Zody *et al.* (2008) Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nature Genetics* 40: 1076-1083.

Mead, S., *et al.* (2003) Balancing selection at the prion protein gene consistent with prehistoric Kurulike epidemics. *Science* 300: 640-643.

Goriely, A., *et al.* (2003) Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. *Science* 301: 643-646.

Novembre, J., *et al.* (2007) Adaptive drool in the gene pool. *Nature Genetics* 39: 1188-1190.

Perry, G. H., *et al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 39: 1256-1260.

Sabeti, P. C., *et al.* (2006) Positive natural selection in the human lineage. *Science* 312: 1614-1620.