

LD Scans for Ongoing Selective Sweeps

Alan R. Rogers

December 3, 2007

Outline

► Questions

- ▶ Have humans evolved rapidly or slowly during the past 40 kyr?
- ▶ What functional categories of gene have evolved most?

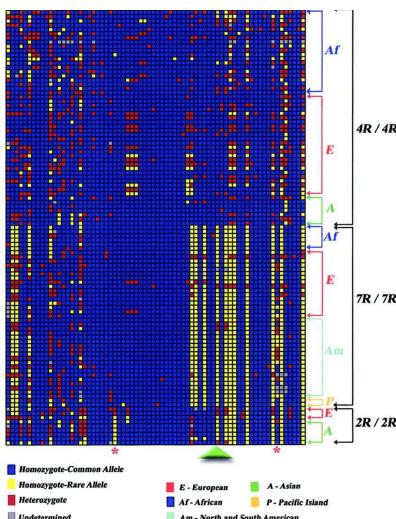
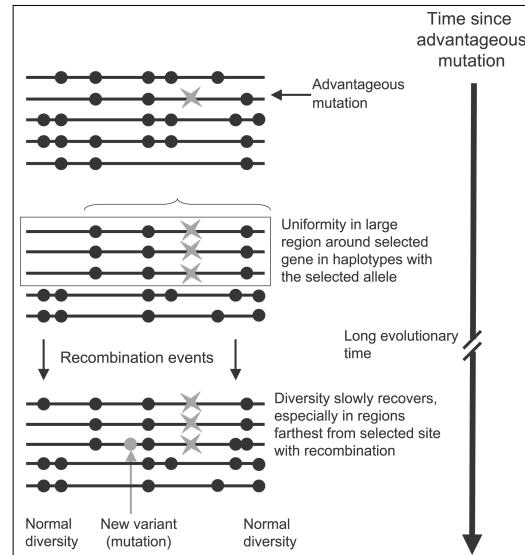
► Selection and recombination

► Data

► Results

Are we still evolving?

- ▶ Since split with chimps, there has been rapid evolution in proteins expressed in brain and in sperm.
- ▶ Recent selection at various loci: lactase, DRD4, etc
- ▶ How common are such loci in the human genome?
- ▶ How can we tell?



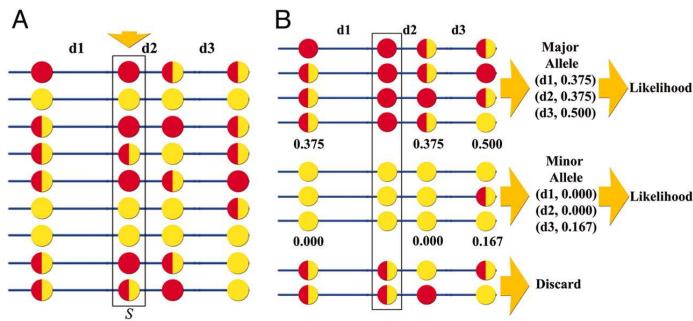
Signature of an ongoing selective sweep at DRD4

- ▶ Sweeping allele is
 - ▶ common
 - ▶ has low diversity over large region
- ▶ High LD over large region

Two methods for finding sweeping alleles

- ▶ Frequency of recombinant chromosomes (FRC): Wang et al 2006
- ▶ Extended haplotype homozygosity (EHH): Voigt et al 2006

Fraction of Recombinant Chromosomes (FRC)



At a given site

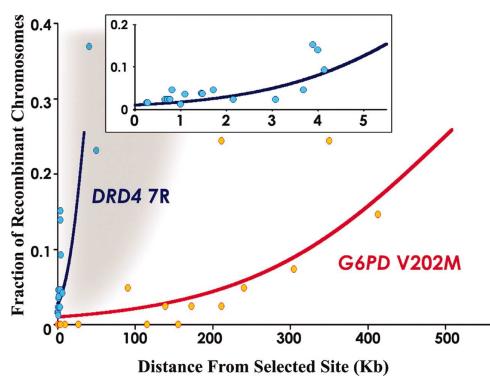
- ▶ FRC increases with time.
- ▶ Rate of increase depends on recombination rate.

At a given time

- ▶ FRC increases with distance along the chromosome.
- ▶ Rate of increase is fast near neutral sites.
(*Neutral* \Rightarrow *old* \Rightarrow *lots of recombination*.)
- ▶ Rate is slow near ongoing selective sweeps.
(*Selected* \Rightarrow *young* \Rightarrow *little recombination*.)

Recipe

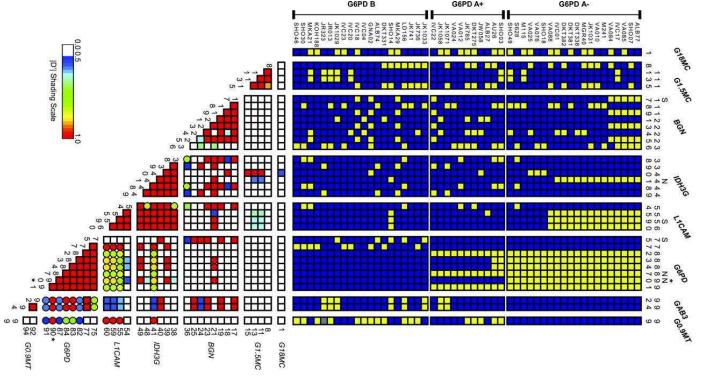
Look for regions where FRC is low in big sections of chromosome.



- ▶ fraction recombinants increases with distance
- ▶ G6PD: recent selection
- ▶ DRD4: older, or in a region with more recombination

Linkage disequilibrium at G6PD

Left: LD plot; right: haplotype plot



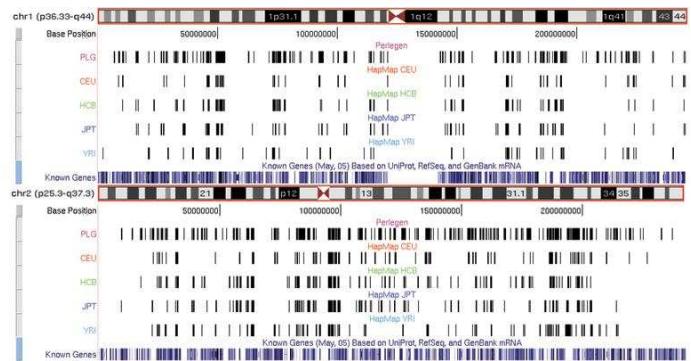
Results for G6PD

- ▶ Favored allele is about 100 generations old.
- ▶ Selective advantage: $s \approx 0.1$.

Study of Wang et al (2006)

- ▶ 1.6 million polymorphic nucleotide sites from throughout the genome
- ▶ most typed in worldwide sample of 71 people

Vertical lines: selected loci (Wang et al 2006)



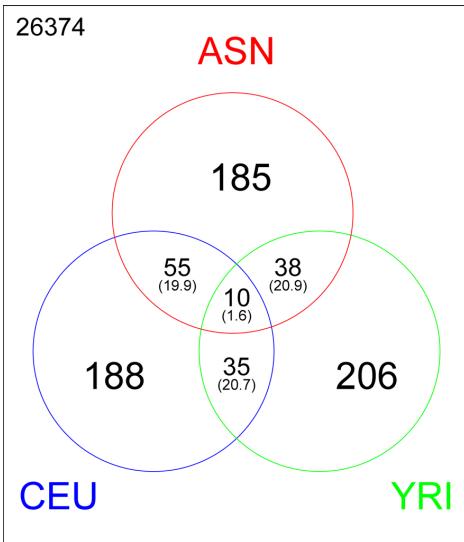
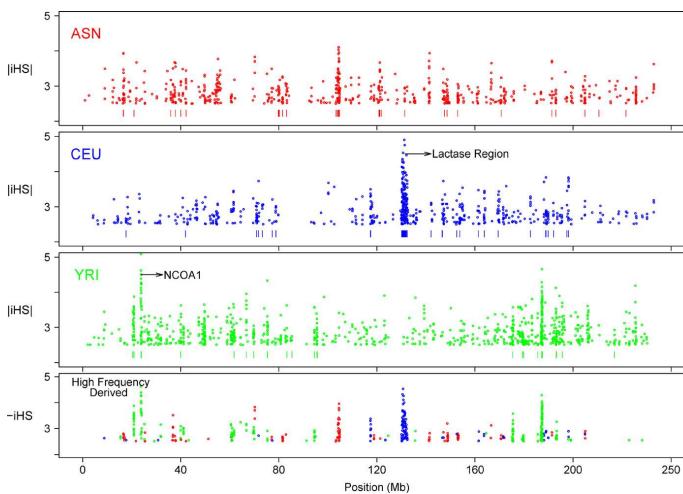
Entire genome (Wang et al 2006)



Study of Voigt et al (2006)

- ▶ 800,000 SNPs in 309 people
- ▶ 431 sweeping loci
- ▶ Most sweeps started w/i past 10,000 years

LD on human chromosome 2 (Voigt et al 2006)



Voight et al (2006):
431 sweeping loci.

ASN: Asia
YRI: Africa
CEU: Europe.

Most are sweeping w/i
only one continent.
Also true of Wang et al
data.

Results of Wang et al

- ▶ At least 1800 genetic loci under positive selection.
- ▶ Most recent estimate: ~ 4000 loci.
- ▶ These mutations arose within past 40 kyr.
- ▶ Reflects a recent burst of adaptive evolution.

- ▶ Evolution did not stop with the origin of modern humans.
- ▶ It accelerated.
- ▶ What kinds of loci have been evolving?

Several categories of gene are over-represented

- ▶ pathogen response
- ▶ neuronal function
- ▶ protein metabolism
- ▶ DNA metabolism
- ▶ cell cycle
- ▶ reproduction

Let us take them one at a time.

Pathogen response

- ▶ Pathogens are easily transmitted in dense populations.
- ▶ Human population became denser after last ice age.
- ▶ Denser again after agriculture.
- ▶ Pathogen load would have increased.
- ▶ Rapid evolution of pathogen response makes sense.

Neuronal function

- ▶ Neuronal function is brain function.
- ▶ Long history of selection for brain function.
- ▶ Has apparently continued, even though brains have not gotten larger.

Protein metabolism

- ▶ With hunting, human diets got more protein, less carbohydrate.
- ▶ With agriculture, the reverse happened.
- ▶ Either change might have required a change in protein metabolism.

DNA metabolism

- ▶ Long-lived animals must repair their DNA.
- ▶ Short-lived animals don't need to.
- ▶ Do these changes reflect a change in human life span?

The other functional categories are mysterious.

Why a recent increase in the rate of evolution?

Cochran & Hawks argue that we have changed our own environment:

- ▶ population size
- ▶ agriculture
- ▶ complex society
- ▶ long-distance trade

To say nothing of the more recent changes.

Summary

- ▶ At least 1800 human genes are under positive selection.
- ▶ Favored alleles originated within past 40 kyr.
- ▶ Reflects a burst of adaptive evolution.
- ▶ Several functional categories.
- ▶ Suggests responses to increased population size, to shifts in diet, and to longevity.