Evolution of the human genome by natural selection

What you will learn in this lecture
(1) What are the human genome and positive selection?
(2) How do we analyze positive selection?
(3) How is positive selection relevant to human evolution, and the evolution of human health and disease?
(4) Applications and case studies: brains and language, diet, reproduction and disease

The Human Genome: Build 36.2
- ~3 billion nucleotides or basepairs, ~3 million vary among random 2 humans
- ~25,000 genes
- Only 1.5% of genome encodes for proteins
- 43 mammalian genomes are in progress
- Mouse, rat, dog, cow, chimpanzee, macaque, others are complete

The Chimpanzee Genome
- Human and chimps diverged 5-6 mya
- ~99% identical overall to human genome
- ~30,000,000 nucleotide differences
- 29% of genes identical to human homologue (6,250 genes)
- Average divergence per gene: 2 amino acid differences; one per lineage since human/chimp divergence

Gene expression differences in human and chimpanzee cerebral cortex
- Affymetrix oligonucleotide array (~10,000) genes
- 91 show human-specific changes, ~90% increases

Large-scale structural variation between human and chimpanzee genomes

Using genetic variation among and within humans and other primates to understand the presence and form of natural selection on genes
(1) Infer ancestral states, for genes
(2) Infer selection on amino acids in proteins with important functions; relate selection on genes to selection on phenotypes
(3) Infer recent 'selective sweeps', or balancing selection, in human genome

WHAT ARE THE GENETIC AND GENOMIC CHANGES THAT HAVE 'MADE US HUMAN'?
Seeking the 'signatures of selection' in human and primate genomes
(1) Inferring Ancestral states, for genes

<table>
<thead>
<tr>
<th>Gene-of-Interest</th>
<th>human</th>
<th>chimp</th>
<th>macaque</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AGCTGCTGG</td>
<td>AGCTGCTGG</td>
<td>AGCTGCTGG</td>
</tr>
</tbody>
</table>

Inferring Lineage Specific Evolution

<table>
<thead>
<tr>
<th>Gene-of-Interest</th>
<th>G→A</th>
<th>human</th>
<th>chimp</th>
<th>macaque</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AGCTACTGG</td>
<td>AGCTGCTGG</td>
<td>AGCTGCTGG</td>
</tr>
</tbody>
</table>

(2) Inferring adaptive amino acid change in proteins

Measuring selection on protein-coding genes -> selection for particular amino acid changes
Changes are synonymous or non-synonymous

<table>
<thead>
<tr>
<th>Amino Acid Changes</th>
<th>Lys</th>
<th>Lys</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA → AA G</td>
<td>Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAA → CAA</td>
<td></td>
<td></td>
<td>Glu</td>
</tr>
</tbody>
</table>
Ratio of non-synonymous to synonymous changes, controlling for the opportunity for changes to occur: $d_N/d_S$

- $d_N/d_S < 1$ when replacements are deleterious (very few changes in amino acids, along lineage)
- $d_N/d_S = 1$ when replacements are neutral (changes just happen randomly)
- $d_N/d_S > 1$ when replacements are advantageous (lots of changes in amino acids along lineage)

### Measuring protein divergence

<table>
<thead>
<tr>
<th>Species</th>
<th>$d_N/d_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species 1</td>
<td>0.396</td>
</tr>
<tr>
<td>Species 2</td>
<td>0.985</td>
</tr>
<tr>
<td>Species 3</td>
<td>0.591</td>
</tr>
<tr>
<td>Species 4</td>
<td>0.198</td>
</tr>
<tr>
<td>Species 5</td>
<td>2.118</td>
</tr>
<tr>
<td>Species 6</td>
<td>0.985</td>
</tr>
<tr>
<td>Species 7</td>
<td>0.591</td>
</tr>
</tbody>
</table>

- $d_N/d_S < 1$: Purifying Selection
- $d_N/d_S = 1$: Neutral Evolution
- $d_N/d_S > 1$: Positive Selection

### Analogy between phenotype-level and genetic-level selection

**Selection ‘for’ change in one direction**

- Directional selection on phenotype: Ala->Glu, Tyr->Ser

**Selection ‘for’ remaining the same**

- Stabilizing selection on phenotype: Ala, Tyr, retained despite mutations to other amino acids

### (3) Infer recent ‘selective sweeps’, or balancing selection, in human genome

- Alleles and Haplotypes that increase in frequency rapidly due to positive selection will carry lots of “hitch-hiking”, flanking DNA, creating a linkage disequilibrium signature
Results of studies on the signatures of selection in the human genome: brains, food, reproduction and parasites

(1) Genome-wide studies
(2) Studies of brain and language genes
(3) Studies of food genes (lactase, amylase)
(4) Studies of reproduction genes
(5) Studies of disease-related genes

Genome-wide analyses of protein sequence evolution

- Most-significant categories showing positive selection in human lineage include:
  *Immune system: parasites and pathogens
  *Reproduction: genes expressed in reproductive tissues
  *Nervous system genes: expressed in brain
  *Amino-acid metabolism: diet
  *Olfaction: sense of smell
  *Development: such as skeletal
  *Hearing: for speech perception

Another method to (carefully) infer selection: geographic variation in allele frequencies and patterns

<table>
<thead>
<tr>
<th>Gene</th>
<th>Putative selective pressure</th>
<th>Phenotype or disease associations</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT</td>
<td>Climate (temperate)</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>CYP1A1</td>
<td>Climate (cold)</td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>SLC22A4</td>
<td>Climate (hot)</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>FY</td>
<td>Pathogen (Plasmodium species)</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>IL4</td>
<td>Pathogen (unknown)</td>
<td>Asthma</td>
<td>46,48</td>
</tr>
<tr>
<td>IL6</td>
<td>Pathogen (unknown)</td>
<td>Asthma</td>
<td>46</td>
</tr>
<tr>
<td>HLA7</td>
<td>Diet (apricot)</td>
<td>Nutrient tolerance or reliance</td>
<td>28</td>
</tr>
<tr>
<td>LCT</td>
<td>Diet (milk)</td>
<td>Prostate cancer</td>
<td>62</td>
</tr>
<tr>
<td>TFF1</td>
<td>Diet (milk)</td>
<td>Coronary heart disease</td>
<td>62</td>
</tr>
</tbody>
</table>

Link to selective agent
Results of studies on the signatures of selection in the human genome: brains, food, reproduction and parasites

(1) Genome-wide studies

(2) Studies of brain and language genes

(3) Studies of food genes (lactase, amylase)

(4) Studies of reproduction genes

(5) Studies of disease-related genes

Specific genes affecting brain size

Microcephaly genes

- Small (~430 cc vs ~1,400 cc) but otherwise ~normal brain, only mild mental retardation
- Some inherited as autosomal dominant
- Can be due to loss of activity of the ASPM gene

Microcephaly associated with abnormal spindle-like microcephaly, or MCPH1 gene

Were these genes involved in the adaptive evolution of big human brain size?

Positive selection of MCPH1 in primate evolution

Positive selection of ASPM in primate evolution

ASPM is still evolving adaptively in human lineage?!
ASPM and MCPH1 adaptive haplotypes are related to forms of human language, tonal and non-tonal Convergence?

**Results of studies on the signatures of selection in the human genome: brains, food, reproduction and parasites**

(1) Genome-wide studies

(2) Studies of brain and language genes

(3) Studies of food genes (lactase, amylase)

(4) Studies of reproduction genes

(5) Studies of disease-related genes

**Inheritance of a language/speech defect in the KE family in London**

*FOXP2: The "Language Gene"*

- FOXP2 mutations result in an autosomal dominant communication disorder

- Phenotype includes problems with speech articulation and deficits in many aspects of language and grammar

- Intelligence varies among affected individuals but speech/language impairment is always present

- Interestingly, deficits with language are not restricted to speech but influence writing and comprehension/ expression

**FOXP2: Molecular Evolution**

- FOXP2 is highly conserved throughout mammals and beyond but for three nucleotide substitutions that change the FOXP2 protein between humans and the mouse, and two have occurred along the human lineage

- Examination of human genetic variation suggests that the region surrounding the gene underwent a selective sweep in the past 200,000 years

**FOXP2: Neuroimaging**

- Brains of individuals with FOXP2 mutations have reduced grey matter in the frontal gyrus which includes Broca’s area

- Functional abnormalities in Broca’s area during language tasks
**FOXP2**: two genetic variants (SNPs) are associated with risk of some neurodevelopmental disorders involving speech and language, schizophrenia and autism

---

**Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations**

Julie Sarrazin, Amparo Tolosa, José C. González, Eduardo J. Aguilar, Jordi Pèlès-Tur, Carmen Najar, María Dolores Molto and Rosa de Frutos

<table>
<thead>
<tr>
<th></th>
<th>CC</th>
<th>CA</th>
<th>AA</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>48</td>
<td>144</td>
<td>98</td>
<td>0.805</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.002*</td>
</tr>
<tr>
<td>Controls</td>
<td>59</td>
<td>173</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>52</td>
<td>138</td>
<td>75</td>
<td>0.776</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004*</td>
</tr>
<tr>
<td>Controls</td>
<td>59</td>
<td>173</td>
<td>75</td>
<td></td>
</tr>
</tbody>
</table>

---

**RELATING HUMAN ADAPTIVE MOLECULAR EVOLUTION TO HUMAN DISEASE**

Table 3. Associations with disease, for genes inferred as positively selected via selection sweeps in Florentino et al. (2010), Table 19 or Green et al., 2010. Table S5, compared to effect of control genes, for selected genes, only single genes for haplotypes associated with disease were included, to avoid ambiguity regarding which gene was the apparent focus of selection. Disease and phenotype associations were obtained from PubMed searches of 5 May 2010 using gene names. As the goal is to compare selected vs. control genes for frequency and excess of disease association, all associations were included, even reported in a single study. Control genes were ascertained at the gene closest to 4 Mb from the top selected gene, in sense or opposite.

<table>
<thead>
<tr>
<th></th>
<th>Selected</th>
<th>Control</th>
<th>Selected</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No association</td>
<td>35 (2%)</td>
<td>77 (4%)</td>
<td>35 (2%)</td>
<td>77 (4%)</td>
</tr>
<tr>
<td>Association with neurological disease</td>
<td>12 (1%)</td>
<td>31 (2%)</td>
<td>12 (1%)</td>
<td>31 (2%)</td>
</tr>
<tr>
<td>Association with schizophrenia</td>
<td>11 (1%)</td>
<td>17 (1%)</td>
<td>11 (1%)</td>
<td>17 (1%)</td>
</tr>
<tr>
<td>Association with bipolar disorder</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with autism</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with ADHD</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with depression</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with schizophrenia</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with bipolar disorder</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with autism</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with ADHD</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
<tr>
<td>Association with depression</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
<td>10 (1%)</td>
<td>20 (1%)</td>
</tr>
</tbody>
</table>

**Results of studies on the signatures of selection in the human genome: brains, food, reproduction and disease**

1. Genome-wide studies
2. Studies of brain and language genes
3. Studies of food genes (lactase, amylase); changes in human diet during recent evolution
4. Studies of reproduction genes
5. Studies of disease-related genes

---

**The Derived FOXP2 Variant of Modern Humans Was Shared with Neandericals**

Summary

Although animals communicate vocally, no evidence exists for modern humans in language ability. Therefore, aerobic stimuli and under what evolutionary pressures occurred. In speech evolution, we know for sure that the Neanderitals, share with modern humans two evolutionary changes to FOXP2, a gene that has been implicated in the development of speech and language, that previously existed in Neanderital lineages. Other changes in FOXP2 have been described in Neanderital populations, previously shown to have been subject to a selective sweep. These results suggest that these genetic changes and the selective sweep provide the common ancestor with which the more recent evolutionary changes that led to modern human and Neanderital populations.

---

**Conclusions:**

Positive selection related to human brain and language:

Many genes related to primate brain development have been subject to positive selection.

We have identified several positively-selected genes related to brain size and language in humans, but we do not know how they work.

These same genes are also involved in human disorders related to the brain and language.

---

**Lactase persistence**

- All infants have high lactase enzyme activity to digest the sugar lactose in milk.
- In most humans, activity declines after weaning, but in some it persists.

---

**Crespi 2010, Evol. Appl.**

---

---

---

---

---

---

---

---

---

---

---

---

---

---
Molecular basis of lactase persistence

• Lactase level is controlled by a cis-acting element
• Linkage and LD studies show association of lactase persistence with the T allele of a T/C polymorphism 14 kb upstream of the lactase gene

---

2004

Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene

Todd Benfey1, Pardis C. Sabeti1, Nick Patterson1, Toshio Vanderheijden1, Steven S. Shaffer1, Jared A. DiLauro, Matthew Buehler1, David E. Reich1, and Joel N. Huberman1,2

In most human populations, the ability to digest lactose is lost in early childhood, but in European-derived populations, lactase activity frequently persists into adulthood. Several studies have suggested that evolutionary forces have shaped the lactase persistence loci. This genetic variation is linked to several SNPs in the 14-kb region upstream of the lactase gene. In northern European–derived populations, two alleles at tightly associated with lactase persistence (Enattah et al. 2002) suggest a role for the lactase persistence gene.

---

2007

Convergent adaptation of human lactase persistence in Africa and Europe

Sarah A. Tishkoff1,2,3, Pardis C. Sabeti2,3, Alexios Remboutsika4,5, Benjamin F. Voight1,6,7, Courtney C. Baldicht1,6,7, Jostine S. Sturm1,6,7, Holly M. Merchant1,6,7, Bidel B. Biddle1,6,7, Mahan Omran1,6,7, Marion Roach1,6,7,17, and Suleyman Seielstad1,6,7,6,7, Jonathan R. Pritchard2,3, Gregory A. Wray1,6,7,8, and Zhixin Dong1,6,7,8

A SNP in the gene encoding lactase (CTC1; CT1; CT1B) is associated with the ability to digest milk in adults, lactase persistence in Europeans, and the development of human milk teeth. We conducted a genetic association study in 1,067 East Africans and 852 Europeans and identified two SNPs (CTC1; CT1; CT1B) and the genetic variation associated with lactase persistence in the African population. The results support a model of convergent evolution in Africa and Europe.

---

Not just milk - use of starches also increased in human diet

Diet and the evolution of human amylase gene copy number variation

George H. Bray5, Nadhifah I. Demez5, Katrina G. Cowan5, Arthur S. Leu5, Elke Vogel5, Richard Belling5, John Wanner9, Leonardo A Villarosa5,4, Joanna I. Mouton5,4, Bjoern Mino5,4, Nagi F. Cortes5, Charles Leu5,4, and Anne C. Hensel5,4

Celiac disease?
Evidence for selection of suite of genes for meat-eating

**Meat-adaptive genes and the evolution of slower aging in humans.**

Finch CE, Stanford CB.

Andrew Lerman, Stanford University, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA, reich@brown.edu

The chimpanzee life span is shorter than that of humans, which is consistent with a faster schedule of aging. We consider aspects of diet that may have selected for genes that allowed the evolution of longer human life spans with slower aging. Diet has changed remarkably during human evolution. All direct human ancestors are believed to have been largely herbivorous. Chimpanzees eat more meat than other great apes, but in captivity are sensitive to hypercholesterolemia and vascular disease. We argue that this dietary shift to increased regular consumption of fatty animal tissues in the course of hominid evolution was mediated by selection for "meat-adaptive" genes. This selection conferred resistance to disease risks associated with meat eating also increased life expectancy. One candidate gene is apolipoprotein E (apoE), with the e3 allele evolved in the genus Homo that reduces the risks from Alzheimer's and vascular disease, as well as influencing inflammation, infection, and neuronal growth. Other evolution genes mediate lipid metabolism and host defense. The timing of the evolution of apoE and other candidates for meat-adaptive genes is discussed in relation to key events in human evolution.

**Better food, smaller guts, adaptations to meat...**

Conclusions:

Positive selection related to human diet:

Humans have been adapting genetically to a novel diet that includes dairy products, grains, and more meat. The selection involved has been strong.

The molecular adaptations involved in dietary adaptations tend to be local geographically, and still exhibit genetic polymorphisms

Results of studies on the signatures of selection in the human genome: brains, food, reproduction and parasites

1. Genome-wide studies
2. Studies of brain and language genes
3. Studies of food genes (lactase, amylase)
4. Studies of reproduction genes
5. Studies of disease-related genes

**Rapid Evolution of Reproductive Proteins in Mammals**

Greater evolution of reproductive proteins in mammals over time (as shown on the graph).

**Rapid Evolution of Fertilization Proteins**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Importance for Human Fertility</th>
</tr>
</thead>
</table>
| Accessory gland gene functions: “Male-female conflicts”

- Egg laying (increased)
- receptivity to mating (decreased for up to 5 days)
- formation of copulatory plug (matting plug)
- Sperm storage (infertility/no storage in the absence of accessory gland proteins)
- Sperm displacement (accessory gland protein promote displacement of previously stored sperm)

Swanson et al. (2003)
Correlation between SEMG2 Evolution and Primate Sexual Traits

Omega = dN/dS

Results:
CatSper1 - sperm motility
different primate species

Comparative evidence for molecular adaptation related to sperm mobility, with implications for human male fertility

Conclusions:
Positive selection related to reproduction:
Genes involved in primate reproduction (sperm and egg production and properties) exhibit among the strongest signals of positive selection of any category of gene, probably because phenotypic selection is so strong

Several recent studies of primate reproduction genes have linked the presence and strength of positive selection on such genes with aspects of the mating system

Such studies have important implications for human fertility and contraception

Results of studies on the signatures of selection in the human genome: brains, food, reproduction and parasites

(1) Genome-wide studies
(2) Studies of brain and language genes
(3) Studies of food genes (lactase, amylase)
(4) Studies of reproduction genes
(5) Studies of disease-related genes

Balancing selection and MHC

Philip W. Hedrick
Department of Biology, Arizona State University, Tempe, AZ 85287-1501, USA

Abstract
The MHC is highly polymorphic in most vertebrates and the suggested selective mechanisms responsible for the maintenance of this variation are several, including maternal-fetal interactions, parasite resistance, and negative sterilization mating. It is believed these mechanisms are interrelated and estimation of the amount of selection in number of studies are given. Although there is much still to be understood about the mechanism and extent of balancing selection at MHC, recent advances in molecular genetic technology and increasing interest in MHC from many types of biological systems raise the need for further investigation.

molecular signature of balancing selection, multiple alleles actively maintained, selection for heterozygosity within individuals
Strong balancing selection at BLA loci: Evidence from segregation in South American families

Michael J. Bednarek and Philip M. Porteous

ABSTRACT

The genetic polymorphisms for major histocompatibility complex (MHC) class II alleles in South American Indians in which segregation for haplotypes and heterozygotes (inclusive of antigenic types) is observed, combined with strong balancing selection favoring homozygotes. There is an estimate that these differences were accepted with particular alleles by the age of the individual sampled. When these findings were divided into 3 major groups, it is evident that the genetic backgrounds were consistent with the age of the individual sampled. The age of the individual sampled is consistent with the age of the individual sampled. The age of the individual sampled is consistent with the age of the individual sampled. The age of the individual sampled is consistent with the age of the individual sampled.