Adaptation in the Human Genome

A genome-wide scan for signatures of adaptive evolution using SNP data

Joanna Kelley
GE414
20 Feb 07

Single Nucleotide Polymorphism (SNP)

A nucleotide difference at a given location in the genome

GTAACCTAC
GTACGCCTAC

Discovering SNPs in the Human Genome

Genomic DNA
Random Shotgun Sequencing
mRNA
cDNA Library
Shotgun Overlap
EST Overlap
Sequence Overlap - SNP discovery

~ 10 Million SNPs Available http://www.ncbi.nlm.gov/SNP/
Debbie Nickerson

HapMap

- Genetic resource for developing association maps
  - Compare genotype patterns between individuals and populations
- Populations
  - Nigerian
  - Japanese
  - Chinese
  - European (Individuals from Utah - CEPH)
- Total number of genotyped SNPs released (Jan 07): 3,904,218
  - Approximately 1 SNP per 1000 base pairs

The HapMap is a Resource for Population Genetic Studies

- SNP data can be used to identify natural selection
- Genome-wide scan for regions of adaptive evolution
  - Selective sweeps
  - Balancing selection
Studying natural selection

Look for an excess of rare alleles using Tajima’s D
- Nucleotide polymorphism, θ
- Nucleotide diversity, π

Advantageous Mutation
Neutral Mutation

Selective Sweep
decrease in genetic variation relative to neutral expectations

Affects entire genome, not just one locus

Population demographics

Nucleotide Polymorphism, θ
number of polymorphic nucleotides, normalized for sample size

\[ \theta = S / \left( \sum_{i=1}^{n} (1/i) \right) \]

3 Segregating (polymorphic) Sites

n = 4 (number of sequences)
θ = 0.027

Note: θ does not depend on nucleotide frequency

Nucleotide diversity, π
proportion of nucleotides that differ between two random sequences in a sample

\[ \pi = \frac{n}{(n-1)} \sum_{i=1}^{n} x_i x_j p_{ij} \]

xi, xj - frequencies of i-th and j-th sequences
p_{ij} - proportion of different nucleotides between two sequences

n = 4 (number of sequences)
π = 0.0267

Note: π depends on nucleotide frequencies

Methods
Download SNP information from SNP database (dbSNP)
Create a list of SNPs that map to gene regions
Download genotype data from HapMap
Analyze the data
Calculate #SNPs (sites) per gene, Tajima’s D, etc.
What do you expect Tajima’s D to look like in non-gene regions?

Tajima’s D across Chromosome 1

Genes show dramatic differences in allele frequencies between populations

Distribution of genes by Tajima’s D

- Gene regions have similar demographic history and ascertainment as non-genic regions
- Comparison of genic and simulated non-genic regions show excess of genes with highly negative Tajima’s D
- Identified 385 genes showing evidence of positive selection

Many of the genes in the tails are hypothetical genes

Good targets for comparative sequencing
Verification Using Sequence Data

- Three of the genes have been verified previously by SeattleSNPs and EGP
- Targeted sequencing of 5 genes
  - Genes with:
    - Tajima’s D < -1.0
    - Tajima’s D > 4.0
  - 1.5 kb each, targeted to HapMap polymorphic sites
  - ENAM - dental enamelin
  - PKD1-like - polycystic kidney disease 1-like
  - POLL - Polymerase lambda
  - RAGE - Renal tumor antigen

An in-depth study: 
**enamelin**

- Extreme population differences
- Ecological relevance
- Potential for phenotypic correlation

What is the evolutionary history of **enamelin**?

Tooth enamel properties

- Hardest, most mineralized tissue in human body
- Made up of small mineralized crystallites
- 85% mineral by volume

Tooth enamel thickness

- Tooth enamel thickness used in characterizing hominoid fossils
- Thickness varies from tooth to tooth, individual to individual and species to species
  - Enamel thickness is heritable (Hlusko et al. 2004)
  - Documented human population specific differences in enamel thickness (Harris et al. 2001)
- Diet correlated to enamel thickness
  - Carnivores / hard-object feeders = thick enamel
  - Herbivores / soft-object feeders = thin enamel

Enamel thickness examples

- **Herbivore**
- **Carnivore**

**enamelin** and tooth enamel formation

- Encodes secretory protein
  - Largest protein in enamel matrix
  - Comprises 5% of total enamel matrix protein
- Function
  - Involved in determining enamel crystallite growth and length
  - Critical for proper enamel formation
    - Mutations lead to thin, malformed enamel
  ⇒ Enamel thickness
Evolutionary history of **enamelin**

- Identified by polymorphism study
  - Tajima’s D
- Confounded by population demographics or ascertainment biases
- Need to confirm
  - Direct sequencing
  - Other methods: \(d_N/d_S\)

\[ \text{enamelin primate species tree} \]

- *Papio anubus* (Baboon) - omnivore
- *Macaca nemestrina* (Pig-tailed Macaque) - omnivore
- *Macaca mulatta* (Rhesus Macaque) - omnivore
- *Colobus angolensis polliatus* (Angolan Colobus) - folivore
- *Nasalis larvatus* (Proboscis) - folivore
- *Hylobates syndactylus* (Siamang) - folivore
- *Gorilla gorilla* (Gorilla) - folivore
- *Pan paniscus* (Bonobo) - omnivore
- *Pan troglodytes* (Chimpanzee) - omnivore
- *Homo sapiens* (Human) - omnivore
- *Saimiri sciureus* (Tamarin) - frugivore
- *Leontopithecus rosalia* (Wolff Monkey) - frugivore

**Evolutionary selection between species**

- Synonymous nucleotide substitution (\(d_S\)) - amino acid remains unchanged
- Non-synonymous substitution (\(d_N\)) - encoded amino acid changes

  **No selection (Neutrality):** \(d_N/d_S = 1\)
  **Purifying selection:** \(d_N/d_S < 1\)
  **Positive selection:** \(d_N/d_S > 1\)

\(d_N/d_S\) - measure of selective pressure

\[ d_N/d_S \text{ Model Comparisons} \]

- **Neutral vs. Selection**
  - 35.04** (df = 2)

- **One-ratio** (\(M_0\)) vs. **Discrete** (\(M_3\))
  - 63.22** (df = 4)

- **Beta** (\(M_7\)) vs. **Beta & "** (\(M_8\))
  - 35.06** (df = 2)

**\(d_N/d_S\) sites model analysis**

<table>
<thead>
<tr>
<th>Models Compared</th>
<th>-2 Δ lnL</th>
<th>(d_N/d_S) estimates</th>
<th>% sites under selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral ((M_1)) vs. Selection ((M_2))</td>
<td>35.04** (df = 2)</td>
<td>6.76</td>
<td>4.3</td>
</tr>
<tr>
<td>One-ratio ((M_0)) vs. Discrete ((M_3))</td>
<td>63.22** (df = 4)</td>
<td>6.76</td>
<td>4.3</td>
</tr>
<tr>
<td>Beta ((M_7)) vs. Beta &amp; &quot; ((M_8))</td>
<td>35.06** (df = 2)</td>
<td>6.80</td>
<td>4.3</td>
</tr>
</tbody>
</table>

\(\text{** p} < 0.01\)

**Predicted sites under selection**

- **Charged changes**

- **Cleavage products**
Have specific lineages been subject to positive selection?

- Correlations exist between diet and enamel thickness in primates
- Are specific lineages with dietary changes correlated to bursts of adaptive evolution?

Method to test lineage specific selection

- Reconstruct ancestral diet
- Identify lineages with dietary shift
- Hypothesis: selection on lineages with dietary shift
- Test for selection
  - Neutral model branches $w = 1$
  - Selection model branches $w = \text{free}$

Bursts of adaptive evolution are correlated to lineages with diet change

- Lysozyme and pancreatic RNAse
  - Specific activity in foregut fermenting species (ruminants and colobine monkeys)
  - Adaptive bursts coinciding with dietary changes

enamelin primate evolution

- Primate $d_N/d_S$ analysis indicates adaptive evolution
- enamelin molecular change tracks with dietary change
- What is the pattern of adaptive evolution within humans?
**Distribution of SNPs in *enamelin***

- Derived allele at high frequency in non-African populations

---

**Non-synonymous derived allele and enamel thickness**

<table>
<thead>
<tr>
<th>African American</th>
<th>European American</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele frequency</td>
<td></td>
</tr>
<tr>
<td>Derived allele</td>
<td>Derived allele</td>
</tr>
<tr>
<td>94%</td>
<td>44%</td>
</tr>
<tr>
<td>56%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Documented population specific differences in tooth enamel thickness (Harris et al. 2001)

| Mean enamel thickness | 5.04 mm | 5.56 mm |

---

**Association study: non-synonymous SNP and enamel thickness**

- 220 African and African American patients
- Bite-wing radiographs for enamel thickness measurements
- Cheek swabs for DNA sequencing
- Statistical analysis for association study

Ilona Khosh, dental student, taking a cheek swab

---

**Conclusions**

- Primate analysis indicates adaptive evolution
- Human polymorphism data provide evidence for human population specific adaptation
- Association study between genotype and enamel thickness phenotype

---

**Preliminary Results**

<table>
<thead>
<tr>
<th></th>
<th>GG</th>
<th>GC</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>17</td>
<td>28</td>
<td>4</td>
</tr>
</tbody>
</table>

---

**Acknowledgements**

- Willie Swanson
- Debbie Nickerson
- Josh Akey
- Bob Waterston
- Frank Roberts
- Evan Eichler
- Tracy Popowics
- Frank Roberts
- Ilona Khosh
- Diane Daubert
- Cindy Desmarais
- Kiran Dhillon
- Swanson Lab
- Sigma Xi
- NIH Genome Training Grant