Comparative Genomics: Human versus chimpanzee

1. Introduction

The chimpanzee is the closest living relative to humans. The two species are nearly identical in DNA sequence (>98% identity), yet vastly different in phenotype. For comparison, *Drosophila melanogaster* and *D. simulans*, two fruit fly species that are nearly indistinguishable in phenotype, have DNA sequence divergence of around 5%.

If human and chimp are so similar at the DNA level, why are they so different in phenotype?
- Are there genes specific to human or chimp that cause differences?
- Are structural changes (i.e. amino acid replacements) responsible for the difference?
- Are gene regulatory changes responsible for the difference?
- Do changes in a few genes have a large effect on phenotype?

In recent years it has become possible to investigate such questions through comparative genomics.

2. How divergent are we?

A draft version of the chimpanzee genome was published in 2005 and allowed whole genome comparisons between human and chimp.

See: The Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69-87.

Nucleotide divergence is 1.23% between 1 human and 1 chimp genome (≈35 million single nucleotide changes). However, this number is inflated because some nucleotide variants are polymorphic within humans or chimps. The estimated fraction of fixed differences is ≈1%.

Note, however, that the above divergence is calculated from single nucleotide differences in aligned sequences, excluding gaps. These gaps in the alignment are caused by insertions or deletions of DNA bases (also known as “indels”).

There are ≈5 million small (1-15 bp) indel differences between human and chimp. If these are counted as mismatches, the divergence between species is around 5%.

There are also much larger insertions/deletions, differences in gene or genome region copy number, differences in transposable elements, and differences in repetitive DNA that are not considered in the calculation of divergence.

On average, the Y chromosome is the most divergent, while the X chromosome is the least. This may be due to a higher mutation rate in the male germline, because male reproductive cells undergo more cell divisions than female. The Y chromosome spends its entire evolutionary history in males. Autosomes spend 1/2 of their evolutionary history in males. The X chromosome spends 1/3 of its evolutionary history in males. Thus, if there is a higher mutation rate in males, we would expect divergence to follow the pattern: Y > Autosomes > X.

At the protein level, the average protein shows 2 amino acid differences between human and chimp. 30% of all proteins are identical between the two species.

Some variants known to cause human diseases are found in the chimp genome. In many of these cases, the "disease" variant appears to be fixed in the chimp population and in many cases it appears to be the ancestral variant. This suggests that there may be epistatic interactions between sites that lead to disease. There may be compensatory mutations in the chimp that render the mutations "harmless" in the chimp genetic background.

3. Olfactory receptor (OR) genes

OR genes are the largest gene family in mammalian genomes. They are involved in sense of smell. There are over 1000 OR genes in the human genome, but only 40% have an intact open reading frame (ORF). The other 60% are pseudogenes

In other great apes ≈70% of OR genes have an intact ORF.

Thus, it appears OR genes are being lost rapidly in humans.

Why?
Humans are the only primates who consume cooked food. "One might speculate that cooking leads to a reduced need to identify toxins in foods" (since these are denatured by cooking).

However, some of the OR genes that remain in the human genome show evidence for positive selection. So it appears that many unnecessary OR genes are lost and become pseudogenes, but others have evolved adaptively in humans.

See: Gilad *et al*, 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. American Journal of Human Genetics 73: 489-501.

4. Genomic searches for positively selected proteins

The type of selection acting on a protein-encoding gene can be estimated from the Ka/Ks ratio.

Ka = the number of nonsynonymous differences per nonsynonymous site
Ks = the number of synonymous differences per synonymous site

Ka/Ks < 1: negative (purifying) selection
Ka/Ks = 1: no selection; completely neutral evolution
Ka/Ks > 1: positive selection

Note that Ka/Ks is sometimes known as dN/dS or ω (omega)

To search for positively selected proteins/genes, one can compare all protein-encoding genes between human and chimp and look for those with Ka/Ks > 1. If an additional outgroup species is used, such as macaque (monkey) or mouse, one can determine if selection occurred on the human lineage or on the chimp lineage. Some newer statistical methods can detect positive selection even when Ka/Ks < 1.

Since these studies examine thousands of genes at once, they are not very powerful for detecting individual genes because of the problem of multiple testing. However, they can

reveal functional groups of genes that are enriched for positive selection. In general, the following types of genes tend to show up in these analyses:

- tumor supression and apoptosis
- spermatogenesis
- sensory perception
- immune defense
- testes expressed genes
- genes on the X chromosome

Interestingly, these same groups of genes tend to show up as positively selected in comparisons of other species, such as mouse-rat or *D. melanogaster-D. simulans*. In this respect, they do not appear to be special to humans. Contrary to what might be expected for humans, there is little evidence for positive selection on genes expressed specifically in brain.

See:
Clark *et al*., 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios Science. 302: 1960-1963.

Nielsen *et al*. 2005, A scan for positively selected genes in the genomes of humans and chimpanzees. PloS Biology 3:e170.

Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316: 222-234.

5. Individual gene studies

In recent years, a number of individual genes have been proposed to play an important role in the phenotypic differentiation of humans and chimps. In general, these are genes that were identified because mutations in them cause human defects. Some examples:

FOXP2 – humans with mutations in this have impaired speech and language skills.

microcephalin – humans with mutations in this gene have primary microcephaly = small brain (≈400 cm^3). Normal ≈ 1400 cm^3.

ASPM – mutations also cause primary microcephaly in humans.

These genes show evidence for accelerated evolution (more amino acid changes) in lineages leading to humans. This suggests that natural selection may have favored larger brains and the ability to use language in humans.

There is also population genetic evidence that suggests that the two microcephaly genes (*microcephalin* and *ASPM*) continue to evolve adaptively in the human population, but this is very controversial. It has also been suggested that a *microcephalin* allele was introgressed into modern humans from a now extinct *Homo* lineage (perhaps Neanderthals).

See:
Enard *et al*., 2002. Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418: 869-872.

Wang and Su, 2004. Molecular evolution of microcephalin, a gene determining human brain size. Human Molecular Genetics 13: 1131-1137.

Evans *et al*., 2004. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Human Molecular Genetics. 13: 489-494.

Evans *et al*., 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science 309: 1720-1722.

Mekel-Bobrov *et al*., 2006. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science 309: 1720-1722.

Evans *et al*. 2006. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into Homo sapiens from an archaic Homo lineage. Proc Natl Acad Sci USA 103: 18178-18183.

6. Gene expression

Because humans and chimps are so similar in DNA and protein sequence, it has been suggested that many of the phenotypic differences between species may be caused by differences in gene expression. That is: when, where, and how much the genes are expressed.

The first large-scale comparison of human and chimp gene expression was performed in 2002.
See: Enard *et al*., 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296: 340-343.

These authors compared the transcriptomes of blood, liver, and brain from humans, chimps, orangutans, and macaques using two different types of microarrays. They also compared human and chimp proteomes using 2D-PAGE.

Microarray 1: Affymetrix human oligonucleotide GeneChips (≈12,000 genes). RNA was from from brain and liver of 3 humans, 3 chimps, and 1 orangutan. All RNA was extracted from dead males. A similar experiment was performed using 3 mouse species of nearly equal divergence as the primate species and Affymetrix mouse chips.

Microarray 2: cDNA microarrays (human unigene set, ≈18,000 genes) were used to compare blood, liver, and brain expression among humans, chimps, and Rhesus macaques.

Result: the brain transcriptome appeared to evolve faster along the human lineage. This is consistent with the hypothesis that rapid evolution of the human brain was caused by changes in gene expression.

2D-PAGE comparison of proteomes found a large excess of “quantitative” changes in human brain, relative to “qualitative”. This is also consistent with many changes in brain gene expression.

Conclusion: Biggest difference between human and chimp is gene expression in the brain.

Newer Results:
See: Khaitovich *et al*., 2004. A neutral model of transcriptome evolution. PLoS Biology 2: e132.

This study used arrays of 28,000 cDNAs to compare expression in brain and liver of human, chimp, orangutan, and macaque.

They found a linear change in both brain and liver expression divergence with time – no increase of expression changes in the human brain. This is analogous to a "molecular clock" of gene expression, which would be expected under neutral evolution.

How do they explain the contradiction between the two studies?
The first study was based on only about 5% of brain-expressed genes

a) fewer genes on chips
b) results based only on genes with expression differences between chimp and human

Even newer results:
See: Khaitovich et al., 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309: 1850-1854.

With the completion of the chimp genome, it became possible to compare both rates of expression divergence and rates of protein divergence of genes expressed in different tissues.

In a comparison of genes expressed in brain, heart, liver, kidney, and testis, it was found that genes expressed in brain evolve the slowest at both the expression and protein level. Genes expressed in testis evolve the fastest.

Still newer results:
See: Khaitovich *et al*., 2006. Positive selection on gene expression in the human brain. Current Biology 16: R356.

Although the brain may show fewer gene expression/protein changes between human and chimp when compared to other tissues, when the changes are mapped onto the human or chimp lineage, there is an excess of changes in the lineage leading to humans. This suggests that either:

1) there is a reduction of purifying selection in the human brain, allowing more changes to accumulate

   or

2) natural selection has favored gene expression and amino acid changes in the evolution of the human brain

There is some evidence supporting the second possibility: The regions of the human genome containing these genes that changed in expression show higher linkage disequilibrium (LD) than regions with genes that did not change. This is consistent with "selective sweeps", where positive selection reduces variation and increases LD in regions of the genome containing a target of selection.

7. The Metabolome

At the level of metabolites (small molecules of <1,500 Daltons that are extracted from various tissues), there are many differences between human and chimp, with the majority of the changes occurring on the human lineage. The greatest changes appear to have occurred in muscle, followed by the brain. It has been hypothesized that the energy requirements of the human brain have led to a reduction in energy consumption by muscles. This could explain why humans have larger brains with greater cognitive abilities, but have waeker muscles than chimps and other primates.

See:

Bozek et al. (2014) Exceptional Evolutionary Divergence of Human Muscle and Brain Metabolomes Parallels Human Cognitive and Physical Uniqueness. PLoS Biol 12(5): e1001871.