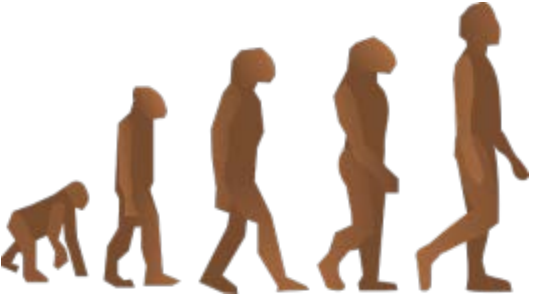
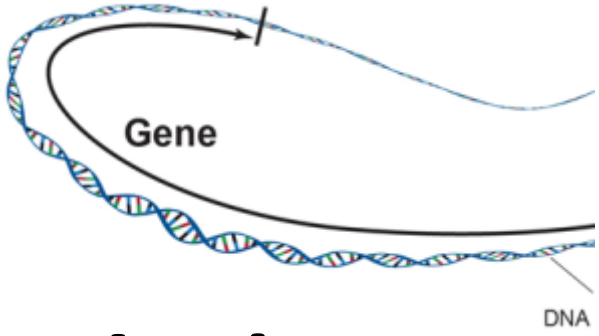
Challenges in searching for signatures of natural selection at individual genes





Today's context:

What gene changes made us "human"?

Nucleotide variation exists within species and between species

Species 1, indiv 1:	AACAGCTGACGTTGTTTAA
Species 1, indiv 2:	AACAGCTGACATTGTTTAA
Species 1, indiv 3:	AACAGCTGACATTGTTTAA
Species 1, indiv 4:	AACAGCTGACGTTGTTTAA
Species 2, indiv 1:	AAGAGCTGACGTTGTTTAA
Species 2, indiv 2:	AAGAGCTGACGTTGTTTAA
Species 2, indiv 3:	AAGAGCTGACGTTGTTTAA
Species 2, indiv 4:	AAGAGCTGACGTTGTTTAG
	1 2 3

Big question…

- How much of the variation observed within and between species is “neutral” (evolving via drift) vs. “selected”

- **Neutralists** – most nucleotide variation within and between species is neutral
- **Selectionists** – very little nucleotide variation is neutral– most variation is selected

Big question for us…

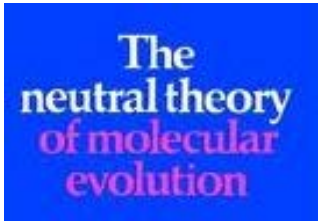
- **What makes humans (or any species) special?**
  - Human genome sequence- 2003
  - Chimpanzee genome sequence- 2005
  - Similarity: **98.77%** in nucleotides

Big question for us…

- **What makes humans (or any species) special?**
  - Human genome sequence- 2003
  - Chimpanzee genome sequence- 2005
  - Similarity: **98.77%** in nucleotides

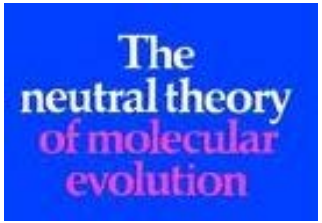
- Evolutionary biologists want to know what specific gene changes were irrelevant (drift) vs. important (selection)
  - Differ at **~47 million bases**…

Neutral Theory of Molecular Evolution



- **Most** mutations that get abundant and eventually get fixed have **no effect on fitness**
  - Arise via **mutation** and spread via **genetic drift**
  - NOT spreading via selection
- Corollary - most nucleotide differences between species spread by drift instead of selection
- Served for many decades as a "null hypothesis"

Neutral Theory of Molecular Evolution



- This theory **DOESN'T** say that all mutations are neutral
  - Acknowledges rare adaptive mutations
  - Acknowledges common bad mutations
  - Just says **MOST** mutations that get abundant and eventually get fixed (e.g., variation within & differences between species) have **no effect on fitness**

Some researchers don’t buy it…

- Hahn (2008) “the patterns apparent from multiple species at multiple loci make the case for rampant nonneutrality.”

- Many (but not all) human pseudogenes appear to evolve neutrally

Big question for us…

- **What makes humans (or any species) special?**
  - Human genome sequence- 2003
  - Chimpanzee genome sequence- 2005
  - Similarity: **98.77%** in nucleotides
  - Differ at **~47 million bases**…

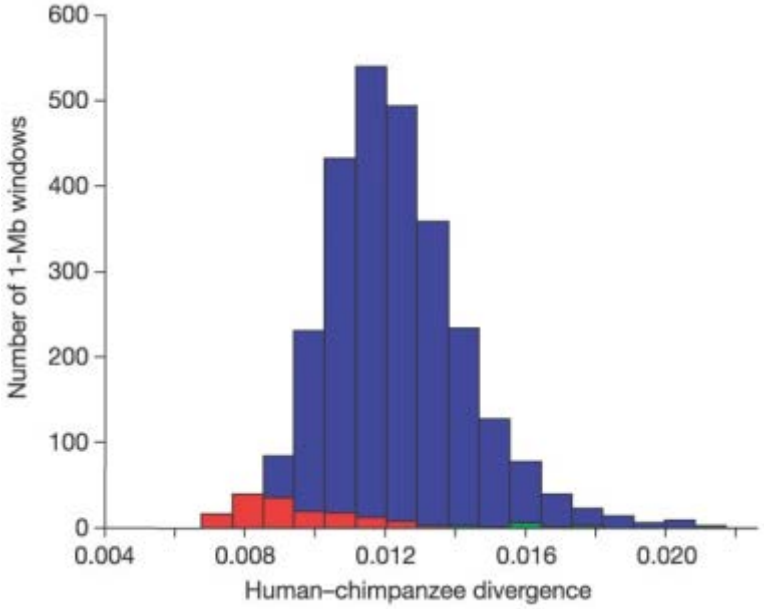
- How do we tell which of these "neutral" that spread by drift vs. "selection-driven"
  and may be important?

Human-chimp:

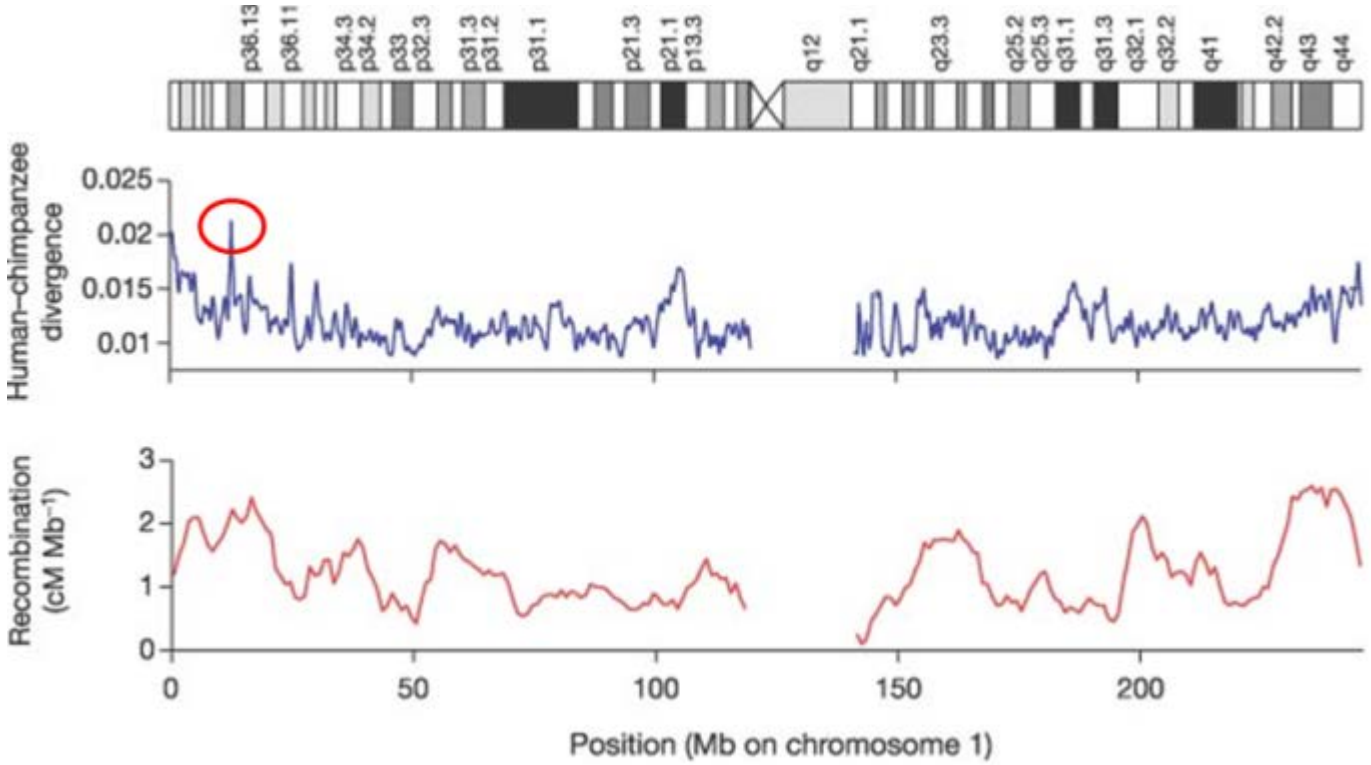
Can we study most different areas?

- Looking across a million bases at a time, some stretches differ at >2%, others at <1%...

- Problem: just because more different, doesn't mean selection involved...



Sequence differences within chromosomes



Not all sequence differences matter…

- Base changes in **pseudogenes** and **introns** (often) have no effect on phenotype
- **Codon third-position** changes often don't change amino acid (synonymous)
  - Second-position always changes (nonsynonymous):

AGA : serine
AGC : serine
AGG : serine
AGT : serine

AAA : phenylalanine
ACA : cysteine
ATA : tyrosine
AGA : serine

Nonsynonymous differences affect the protein

- *Nonsynonymous* differences in coding DNA cause a **different amino acid** to be placed into the protein, **and may affect phenotype**
- These nonsynonymous differences may be:
  - **Detrimental** (if original protein was better),
  - Neutral (if no effect on protein function), or
  - Advantageous (if new protein is better)!
- *Synonymous* differences presumed to be neutral (no effect on phenotype)

We want to “bin” these changes...

- Some gene changes “don’t matter”
  - “Neutral”- not affected by selection
- Many gene changes “bad”
  - Selection prevents them from going to fixation
  - “Negative selection” or “purifying selection”
- Occasional gene changes are “good”
  - Selection makes them likely to go to fixation
  - “Positive selection”

- **How much of each is there within and between species?**

Prediction if all base changes neutral

Mutations lead to differences between species

- Mutation must occur, and new mutant allele must “fix” (get to 100%)
  - Probability of fixation if neutral = 1/(2N)
  - Probability of fixation if “bad” < 1/(2N)

- More coding mutations are nonsynonymous than synonymous
  - All 2nd position, Most 1st position, Some 3rd position changes are nonsynonymous

Codon table

Standard genetic code

1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Still, synonymous sites usually accumulate differences faster…

- Example shown from flu virus over 20 year period
- Probably get more NONsynonymous mutations arising
  - BUT they are selected against right away and never spread

- **Synonymous mutations happen less often but far more likely to spread because not selected against**

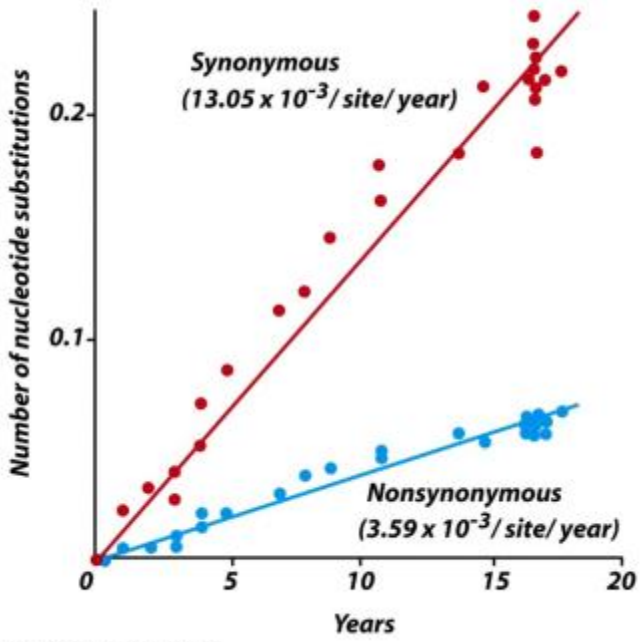


Figure 7-21c Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Can we study genes with many *nonsynonymous* changes?

Can we study genes with many *nonsynonymous* changes?

- But mutation rates not the same in all genes
  - Some genes get more mutations than others, irrespective of selection...
  - If genes get more mutations, then likely to accumulate more differences (both synonymous and nonsynonymous) between species
    - Not telling you much about selection or “importance”

The fix: scale using number of synonymous changes!

- Synonymous differences accumulate neutrally
  - Can use them to scale for mutation rate differences

- RATIO of nonsynonymous to synonymous differences estimates non-neutral changes relative to neutral changes

  - Let's see how this is done with two tests…

Image Credits, Unit 15-1

- Chimps playing, © 2006 Delphine Bruyere, GPL 1.2
- Kids playing, © 2008 Artaxerxes, CC by-SA 3.0
- Human and ape hands, © Eric Isselee, all rights reserved, www.photoxpress.com
- Man rubbing chimp, © Eric Isselee, all rights reserved, www.photoxpress.com
- Human-chimp divergence and Regional variation, © 2005 Republished with permission from Macmillan Publishers Ltd: Nature, The Chimpanzee Sequencing and Analysis Consortium (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome." Nature 437, 69-87, all rights reserved
- Trash bins, © epSos.de, CC by 2.0, www.flickr.com
- DNA mutant, © appler, all rights reserved, www.photoxpress.com
- Synonymous mutations, © 1990 T. Gojobori, E. N. Moriyama and M. Kimura, all rights reserved, "Molecular clock of viral evolution, and the neutral theory", www.pnas.org