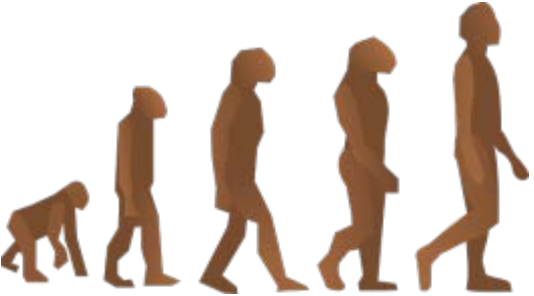
Searching for natural selection on individual genes: McDonald-Kreitman test



Need another test…

- dN/dS can be too conservative for finding adaptive amino-acid changes
  - Many false negatives
- Neutral theory predicts that ratio of nonsynonymous to synonymous changes should be constant through time
  - ∴ Ratio observed among individuals *within species* should be equal to ratio observed *between species*

The neutral theory of molecular evolution

Why NS:S ratio constant through time???

- **Neutral Theory Assumption**: most non-deleterious non-synonymous mutations are neutral
  - Non-deleterious (neutral) nonsynonymous mutations behave just like synonymous mutations
  - Arise at a relatively constant rate (assumption)
  - Fix with same probability as synonymous: 1/(2N)

McDonald-Kreitman test

- Proposed test of selection by seeing if NS:S ratio is constant within and between species
  - Contrasts "present" (within) with "historical" (between)
- Align sets of sequences and identify if **variable** nucleotide sites have:
  - Nonsynonymous differences *within* species
  - Synonymous differences *within* species
  - Nonsynonymous differences *between* species
  - Synonymous differences *between* species

McDonald-Kreitman test procedure

Species 1	CTT ACT TAT ACC CGT CTG ACT TAT ACC CGT CTG ACT TCT ACC CGT CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	T			C			A			G		
T	TTT	Phe	F	TCT	Ser	S	TAT	Tyr	Y	TGT	Cys	C
	TTC	Phe	F	TCC	Ser	S	TAC	Tyr	Y	TGC	Cys	C
	TTA	Leu	L	TCA	Ser	S	TAA	stop	*	TGA	stop	*
	TTG	Leu	L	TCG	Ser	S	TAG	stop	*	TGG	Trp	W
C	CTT	Leu	L	CCT	Pro	P	CAT	His	H	CGT	Arg	R
	CTC	Leu	L	CCC	Pro	P	CAC	His	H	CGC	Arg	R
	CTA	Leu	L	CCA	Pro	P	CAA	Gln	Q	CGA	Arg	R
	CTG	Leu	L	CCG	Pro	P	CAG	Gln	Q	CGG	Arg	R
A	ATT	Ile	I	ACT	Thr	T	AAT	Asn	N	AGT	Ser	S
	ATC	Ile	I	ACC	Thr	T	AAC	Asn	N	AGC	Ser	S
	ATA	Ile	I	ACA	Thr	T	AAA	Lys	K	AGA	Arg	R
	ATG	Met	M	ACG	Thr	T	AAG	Lys	K	AGG	Arg	R
G	GTT	Val	V	GCT	Ala	A	GAT	Asp	D	GGT	Gly	G
	GTC	Val	V	GCC	Ala	A	GAC	Asp	D	GGC	Gly	G
	GTA	Val	V	GCA	Ala	A	GAA	Glu	E	GGA	Gly	G
	GTG	Val	V	GCG	Ala	A	GAG	Glu	E	GGG	Gly	G

Courtesy of M. Rausher

	Fixed Between	Polymorphism within
non-syn		
syn		

McDonald-Kreitman test procedure

CTN = Leucine

ATG = Methionine

Species 1	CTT ACT TAT ACC CGT
	CTG ACT TAT ACC CGT
	CTG ACT TCT ACC CGT
	CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	Fixed Between	Polymorphism within
non-syn	I	
syn		

Courtesy of M. Rausher

McDonald-Kreitman test procedure

CTN = Leucine

Species 1	CTT ACT TAT ACC CGT
	CTG ACT TAT ACC CGT
	CTG ACT TCT ACC CGT
	CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	Fixed Between	Polymorphism within
non-syn	I	
syn		I

Courtesy of M. Rausher

McDonald-Kreitman test procedure

ACN = Threonine

Species 1	CTT ACT TAT ACC CGT
	CTG ACT TAT ACC CGT
	CTG ACT TCT ACC CGT
	CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	Fixed Between	Polymorphism within
non-syn	I	
syn	I	I

Courtesy of M. Rausher

McDonald-Kreitman test procedure

TCT = Serine

TAT = Tyrosine

Species 1	CTT ACT TAT ACC CGT
	CTG ACT TAT ACC CGT
	CTG ACT TCT ACC CGT
	CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	Fixed Between	Polymorphism within
non-syn	I	I
syn	I	I

Courtesy of M. Rausher

McDonald-Kreitman test procedure

ACN = Threonine

Species 1	CTT ACT TAT ACC CGT
	CTG ACT TAT ACC CGT
	CTG ACT TCT ACC CGT
	CTG ACT TCT ACA CGT
Species 2	ATG ACC TCT ACC CGT

	Fixed Between	Polymorphism within
non-syn	I	I
syn	I	II

Courtesy of M. Rausher

McDonald-Kreitman test prediction

	Between Species	Within Species
non-syn	A	B
syn	C	D

If all non-synonymous differences are neutral, expect A/C = B/D

If some non-synonymous differences between species were **advantageous & selected**, expect A/C > B/D

Courtesy of M. Rausher

Examples (human-chimp):

SENP1:
Enables cellular survival during periods of low oxygen, a target for cancer therapy

	Between Human-Chimp	Among Humans
non-syn	4	1
syn	1	5

4:1 >> 1:5

CIAS1:
Affects autoinflammatory response

	Between Human-Chimp	Among Humans
non-syn	5	1
syn	10	2

5:10 = 1:2

Examples (human-chimp):

AGT:
Mutations in it affect hypertension and associate with coronary heart disease

	Between Human-Chimp	Among Humans
non-syn	1	3
syn	13	1

1:13 << 3:1

- What does it mean if there's proportionately ***more*** nonsynonymous variable sites within species than between species?

Examples (human-chimp):

AGT:

Mutations in it affect hypertension and associate with coronary heart disease

	Between Human-Chimp	Among Humans
non-syn	1	3
syn	13	1

1:13 << 3:1

- What does it mean if there's proportionately ***more*** nonsynonymous variable sites within species than between species?
  - What kind of variation may persist for a long time but never fix?

Variation that persists but doesn't fix…

- Hard to eliminate unfit recessive allele
  - Better dominant allele "never quite" gets to fixation

- Diseases tend to be recessive

- *Mutations causing disease stick around but don't fix*

McDonald-Kreitman test prediction

	Between Species	Within Species
non-syn	A	B
syn	C	D

If all non-synonymous differences are neutral, expect A/C = B/D

If some non-synonymous differences between species were **advantageous & selected**, expect A/C > B/D

If **maladaptive non-synonymous differences persist within species**, expect A/C < B/D

McDonald-Kreitman Test

- If A/C = B/D
  - Can't reject neutrality

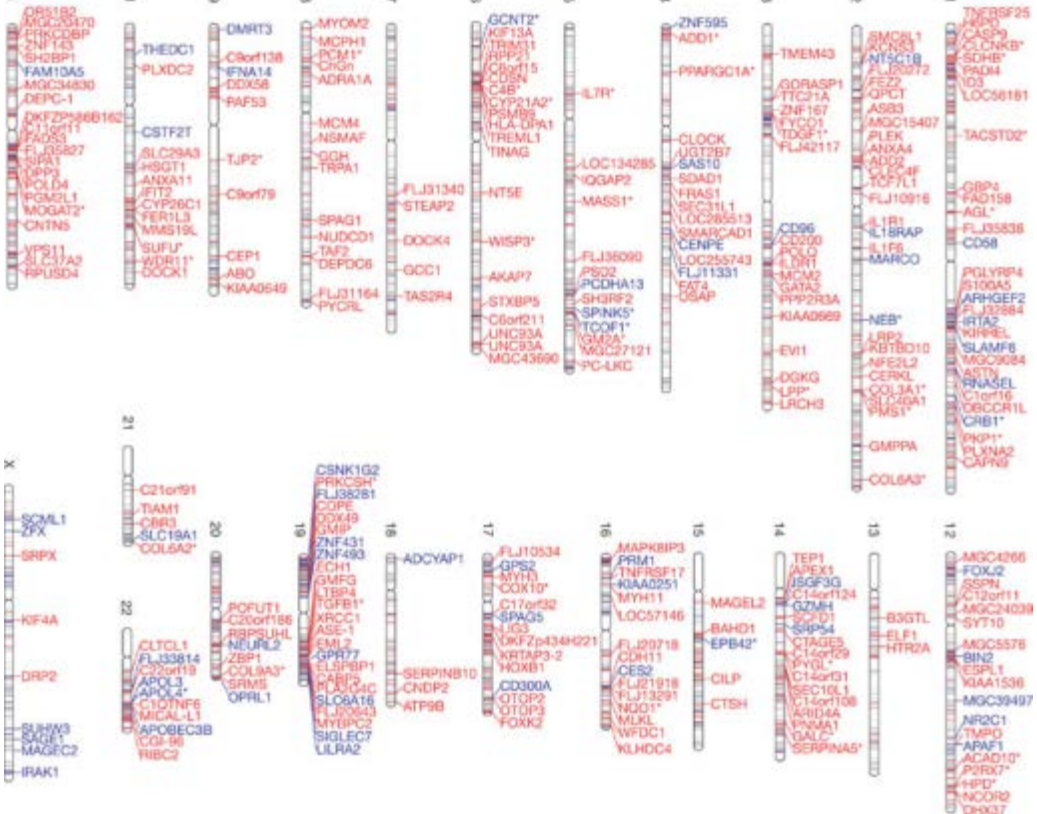
	Between Species	Within Species
non-syn	A	B
syn	C	D

- If A/C > B/D
  - Increased nonsynonymous changes between species- "positive selection"

- If A/C < B/D
  - Decreased nonsynonymous changes between species- "negative selection"

McDonald-Kreitman tests across human-chimp genomes

- Positive selection (adaptive species difference)
- Negative selection (maladaptive variation)



McDonald-Kreitman tests across human-chimp genomes

- Positively selected genes (304):
  - Immunity protein genes, gamete formation genes, sensory perception genes

- Negatively selected genes (813):
  - Many involved in *cytoskeleton* formation: associated with diseases like muscular dystrophy, congenital deafness, cardiovascular disease

One for you to try…

- *arsD* (controls arsenic resistance) 2112

```
Human1      TCT CCT ACA GGG CGT CTA GTT
Human2      TCT CCT ACA GGG CGT CTA GTT
Human3      TCT CCT ACT GGG CGT CTA ATT
Human4      TCT CCT ACT GGG CGT CTA GTT
Human5      TCT CCT ACA GGG CGT CTA GTT
Human6      TCT CAT ACA GGG CGT CTA GTT
-------------------------------------
Chimp1      TAT CCT ACA GGC CGT CTT GTT
```

- Calculate & interpret McDonald-Kreitman test

One for you to try…

- *arsD* (controls arsenic resistance) 2112

```
Human1    TCT CCT ACA GGG CGT CTA GTT
  Human2  TCT CCT ACA GGG CGT CTA GTT
  Human3  TCT CCT ACT GGG CGT CTA ATT
  Human4  TCT CCT ACT GGG CGT CTA GTT
  Human5  TCT CCT ACA GGG CGT CTA GTT
  Human6  TCT CAT ACA GGG CGT CTA GTT
---------------------------------------
Chimp1    TAT CCT ACA GGC CGT CTT GTT
```

One for you to try…

- *arsD* (controls arsenic resistance) 2112

```
Human1    TCT CCT ACA GGG CGT CTA GTT
Human2    TCT CCT ACA GGG CGT CTA GTT
Human3    TCT CCT ACT GGG CGT CTA ATT
Human4    TCT CCT ACT GGG CGT CTA GTT
Human5    TCT CCT ACA GGG CGT CTA GTT
Human6    TCT CAT ACA GGG CGT CTA GTT
---------------------------------------
Chimp1    TAT CCT ACA GGC CGT CTT GTT
```

NS between: 1

S between: 2

NS within: 2

S within: 1

Between Within

1:2 vs 2:1

One for you to try…

- *arsD* (controls arsenic resistance) 2112

```
Human1    TCT CCT ACA GGG CGT CTA GTT
Human2    TCT CCT ACA GGG CGT CTA GTT
Human3    TCT CCT ACT GGG CGT CTA ATT
Human4    TCT CCT ACT GGG CGT CTA GTT
Human5    TCT CCT ACA GGG CGT CTA GTT
Human6    TCT CAT ACA GGG CGT CTA GTT
-------------------------------------
Chimp1    TAT CCT ACA GGC CGT CTT GTT
```

NS between: 1	NS within: 2	Between	Within
S between: 2	S within: 1	1:2 vs	2:1

Result in direction of negative selection

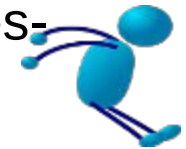
Recap: McDonald Kreitman Test

- If A/C = B/D
  - Can't reject neutrality

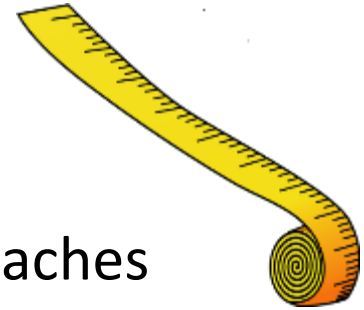
	Between Species	Within Species
non-syn	A	B
syn	C	D

- If A/C > B/D
  - Increased nonsynonymous changes between species- "positive selection"

- If A/C < B/D
  - Decreased nonsynonymous changes between species- "negative selection"



Other metrics



- dN/dS and McDonald-Kreitman are just two approaches

- Many others also exist (we won't use in class)
  - *Tajima's D*: looks at frequencies of alleles to test for recent selection or recent changes in population size

Implementing these tests

- Two general approaches for studying
  1. Scan whole genome using these metrics and look for possible selection, then try to interpret
  2. Look at “candidate genes” (e.g., brain size, speech) and look for signature of selection
- Both are active areas of research
  - Both have given us some clues as to what changes were “important” in our divergence

What makes humans special?

- These DNA sequence-based tests showed:
  - Hundreds of genes underwent selection to change
  - More genes under selection to stay the same (our sequence same as chimpanzee)
  - Some “bad” alleles are sticking around
  - **No single gene change made us what we are today, but we’re not so different from chimpanzees (genetically) as we might like to think we are…**

Image Credits, Unit 15-3

- Selection map, © 2005 Reprinted by permission from Macmillan Publishers Ltd: Nature, Bustamente, C. D. et al. (2005). "Natural selection on protein-coding genes in the human genome." Nature 437, 1153-1157, all rights reserved