Codon Usage in Bacteria and Mitochondria

Paul Higgs
Wenli Jia
Wenqi Ran

Dept of Physics and Astronomy
McMaster University

Supported by Canada
Research Chairs and NSERC.

The Bottom Line Now (in case we don't reach it…)

Synonymous codons are not used with equal frequency –
Is this translational selection or biased mutation?

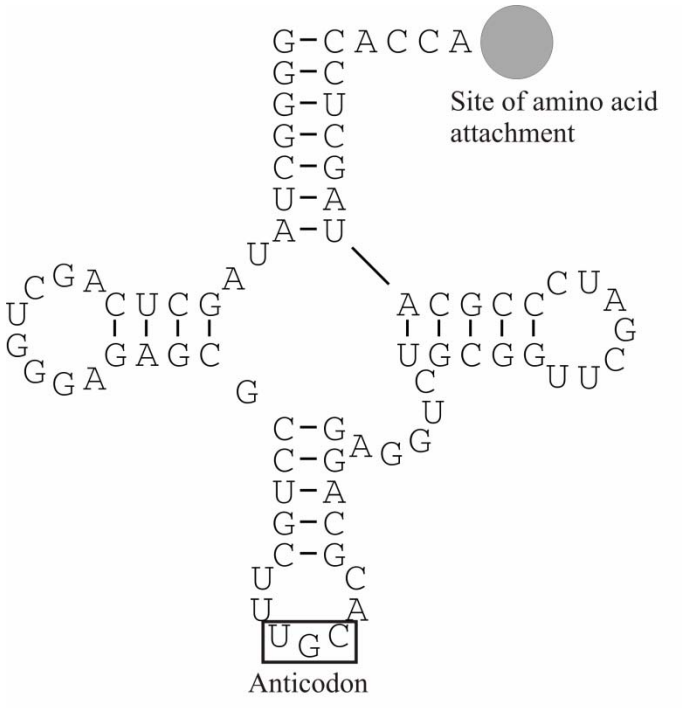
Bacteria - Ran and Higgs (2008) Mol. Biol. Evol.

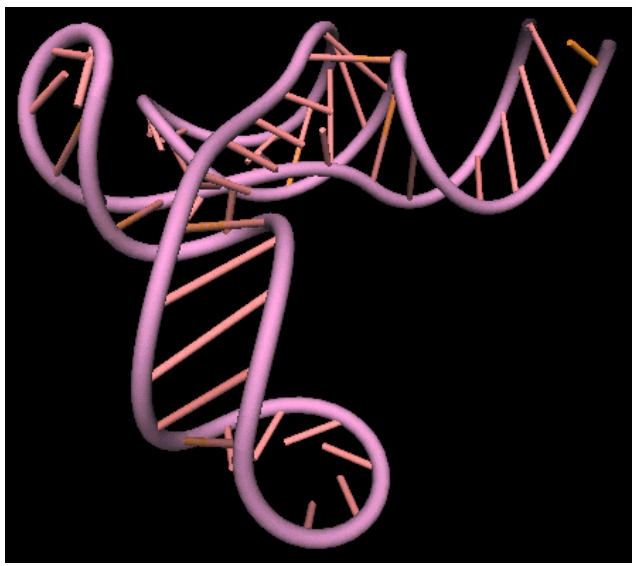
- Comparison of high and low expression genes demonstrates translational selection
- Significant translational selection in most bacteria – varies with growth rate
- Fast multiplying bacteria need fast translation – therefore they have duplicate tRNAs and stronger codon bias.
- Coevolution of tRNAs and codon usage creates multiple stable states in the same organism

Mitochondria - Jia and Higgs (2008) Mol. Biol. Evol.

- Mutation strong enough to cause large amino acid frequency variation as well as synonymous substitutions.
- Mutation is context dependent – leads to dinucleotide correlations
- Mutation varies between strands and along a strand
- Mutational effects dominate translational selection in determining codon usage

tRNA structure

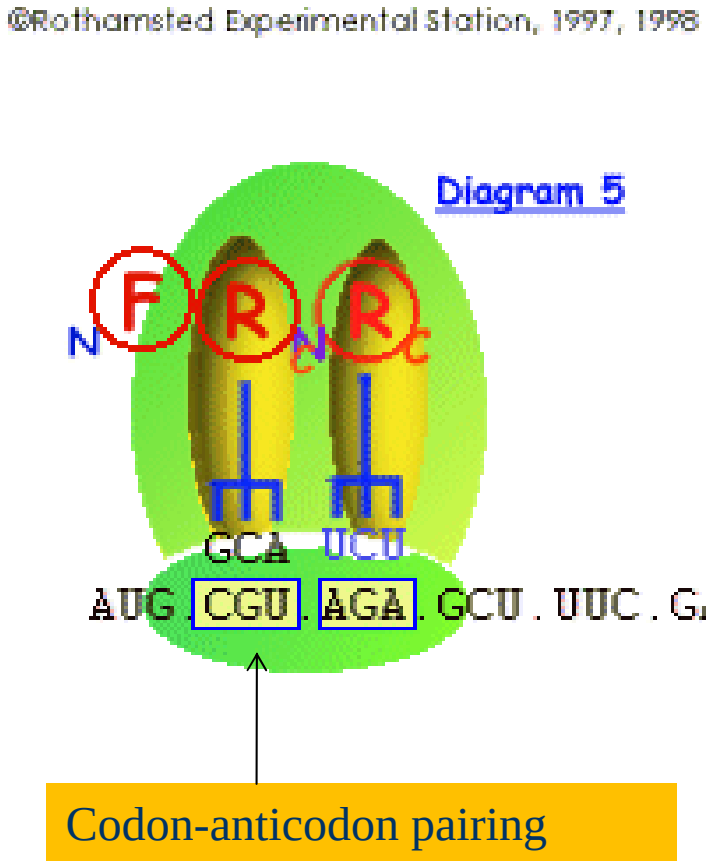




Translation = Protein synthesis

courtesy of 'Molecular Biology Online'

http://www.rothamsted.ac.uk/notebook/index.html



Codon-anticodon pairing

Standard Genetic Code

Phe	UUU
Phe	UUC
Leu	UUA
Leu	UUG

Ser	UCU
Ser	UCC
Ser	UCA
Ser	UCG

Tyr	UAU
Tyr	UAC
*	UAA
*	UAG

Cys	UGU
Cys	UGC
*	UGA
Trp	UGG

Leu	CUU
Leu	CUC
Leu	CUA
Leu	CUG

Pro	CCU
Pro	CCC
Pro	CCA
Pro	CCG

His	CAU
His	CAC
Gln	CAA
Gln	CAG

Arg	CGU
Arg	CGC
Arg	CGA
Arg	CGG

Ile	AUU
Ile	AUC
Ile	AUA
Met	AUG

Thr	ACU
Thr	ACC
Thr	ACA
Thr	ACG

Asn	AAU
Asn	AAC
Lys	AAA
Lys	AAG

Ser	AGU
Ser	AGC
Arg	AGA
Arg	AGG

Val	GUU
Val	GUC
Val	GUA
Val	GUG

Ala	GCU
Ala	GCC
Ala	GCA
Ala	GCG

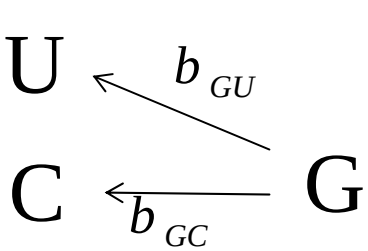
Asp	GAU
Asp	GAC
Glu	GAA
Glu	GAG

Gly	GGU
Gly	GGC
Gly	GGA
Gly	GGG

How many different tRNAs do we need ?

Two codon U+C families

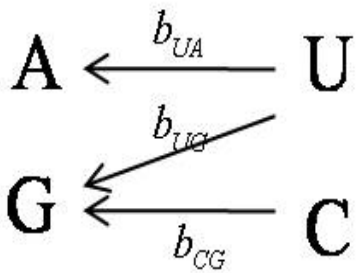
codon	anticodon
Asp GAU	
Asp GAC	GUC



Always a wobble-G tRNA only

Two codon A+G families

codon	anticodon
Glu GAA	UUC
Glu GAG	CUC

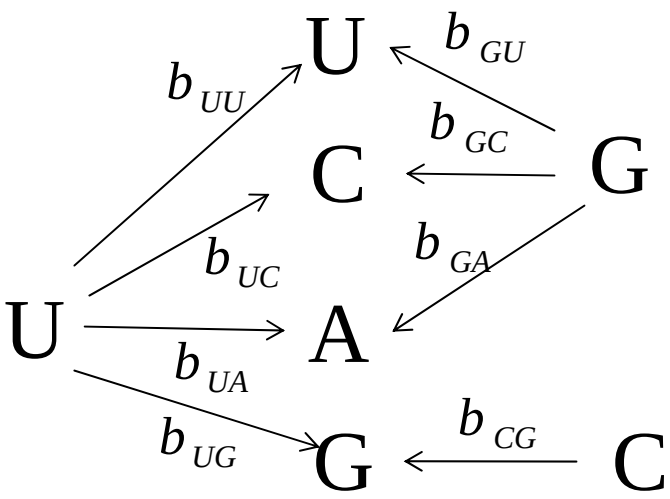


Always a wobble-U tRNA
Sometimes a wobble-C tRNA

How many different tRNAs do we need ?

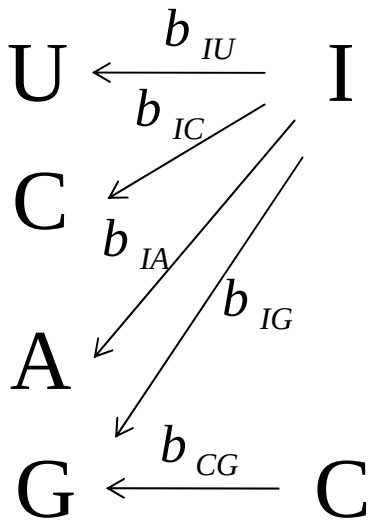
Four-codon families

codon	anticodon
Gly GGU	
Gly GGC	GCC
Gly GGA	UCC
Gly GGG	CCC



Occasionally just wobble-U.
Often wobble-U + wobble-G.
Sometimes wobble-U G and C

Arginine is a special case in bacteria



How many tRNA gene copies are present in genomes?

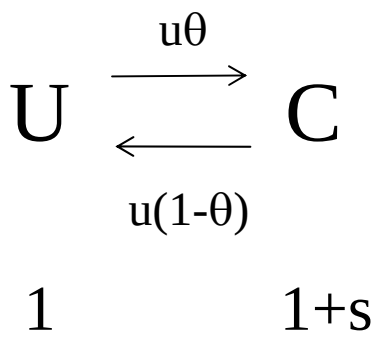
Organism	# gene copies	# different anticodons	minimum doubling time (hrs)
animal mitochondria	22	22	-
Buchneria aphidicola	31	29	36
Agrobacterium tumefaciens	53	39	3
Corynebacterium glutamicum	60	42	1.2
Escherichia coli	86	39	0.35
Vibrio vulnificus	112	32	0.16
Saccharomyces cerevisiae	273	41	-
Caenorhabditis elegans	605	47	-
Human	448	49	-

Evidence for Translational selection #1

More gene copies → More tRNA molecules → More rapidly multiplying bacteria

Selection-Mutation-Drift Theory

Li (1987), Shields (1990), Bulmer (1991)



In absence of selection the relative frequency of C is

$$\frac{n_C}{n_U + n_C} = \theta$$

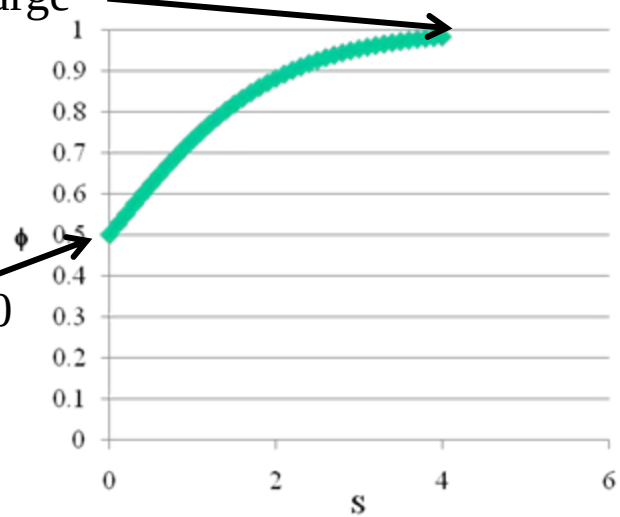
In presence of selection the relative frequency of C is

$$\frac{n_C}{n_C + n_U} = \phi(S) = \frac{\theta \exp(S)}{\theta \exp(S) + 1 - \theta}$$

where $S = 2N_e s$.

$\phi \to 1$ when S is large

$\phi = \theta$ when $S = 0$



$$S = \ln\left(\frac{\phi(S)}{1-\phi(S)}\frac{1-\theta}{\theta}\right)$$

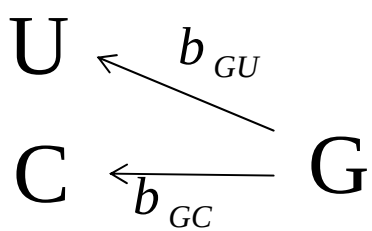
Estimate S from sequence data – U+C families

Assume S is negligible in low expression genes, but significant in high expression genes.

$$\frac{n_C^{low}}{n_U^{low}+n_C^{low}}=\theta \qquad \frac{n_C^{high}}{n_U^{high}+n_C^{high}}=\phi(S) \qquad S=\ln\left(\frac{\phi(S)}{1-\phi(S)}\frac{1-\theta}{\theta}\right)=\ln\left(\frac{n_C^{high}n_U^{low}}{n_U^{high}n_C^{low}}\right)$$

		θ	ϕ	S	NG
Mycoplasma penetrans	Asn	0.236	0.403	0.78	1
	Asp	0.140	0.201	0.43	1
Agrobacterium tumefaciens	Asn	0.561	0.841	1.42	1
	Asp	0.491	0.764	1.21	2
Sinorhizobium meliloti	Asn	0.649	0.817	0.88	1
	Asp	0.642	0.732	0.42	2
Corynebacterium glutamicum	Asn	0.666	0.952	2.29	2
	Asp	0.444	0.722	1.18	2
Escherichia coli	Asn	0.550	0.875	1.74	4
	Asp	0.372	0.657	1.17	3
Bacillus subtilis	Asn	0.435	0.775	1.50	4
	Asp	0.360	0.470	0.45	4
Schizosaccharomyces pombe	Asn	0.343	0.718	1.58	6
	Asp	0.292	0.438	0.64	8
Saccharomyces cerevisiae	Asn	0.410	0.860	2.18	10
	Asp	0.350	0.578	0.93	16
Caenorhabditis elegans	Asn	0.378	0.575	0.80	20
	Asp	0.324	0.559	0.97	27

S is positive in all these examples.
C codon is always preferred.

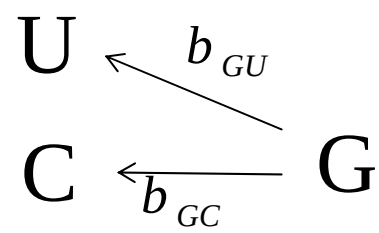


Relate codon usage to translation kinetics

Rate of translation of a codon depends on:

- tRNA concentration $c_0 N_G$
- and a rate constant for anticodon-codon matching $k_0 b_{GC}$

$$r_U = c_0 N_G k_0 b_{GU} \qquad r_C = c_0 N_G k_0 b_{GC}$$



Mean translation times per codon : $t_U = 1/r_U$, $t_C = 1/r_C$

Selection strength proportional to difference in translation times.

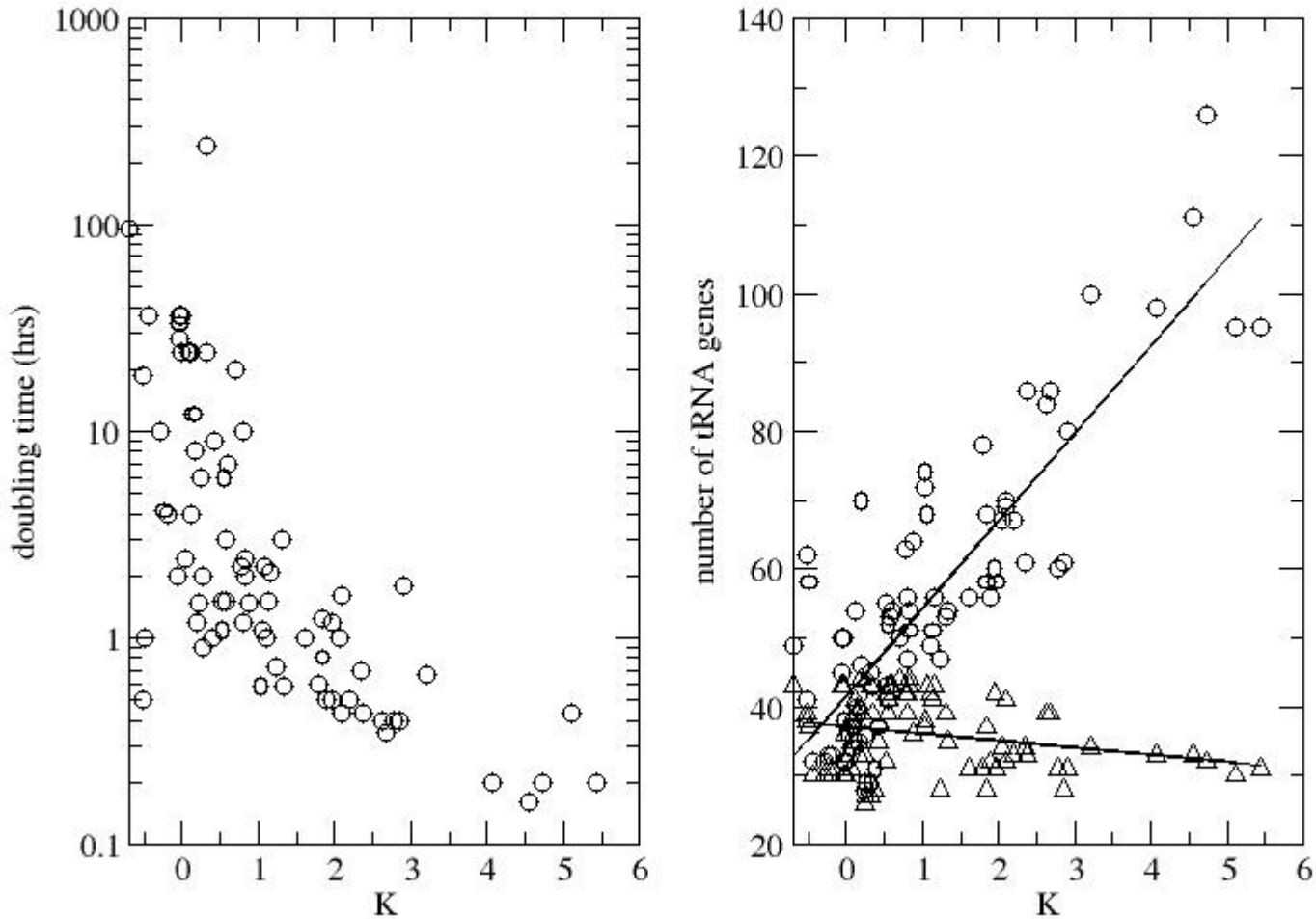
$$S = \sigma(t_U - t_C) = K\left(\frac{1}{N_G b_{GU}} - \frac{1}{N_G b_{GC}}\right)$$

$K \equiv \sigma / k_0 c_0$

Single constant K determines magnitude of selection

$$K = \frac{SN_G}{\left(\frac{1}{b_{GU}} - \frac{1}{b_{GC}}\right)}$$

Test our theory using data on 80 bacterial genomes (Ran & Higgs, 2008)
Estimate K from an average of all the U+C amino acids.

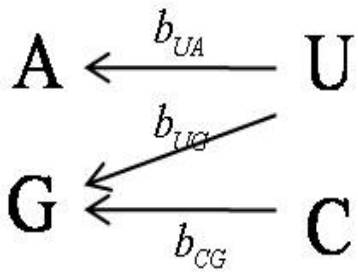


In rapidly multiplying bacteria translational speed is important – K is larger.
Faster translation requires more tRNAs and more biased codon usage.

Estimate S from sequence data – A+G families

		θ	ϕ	S	NU:NC
Mycoplasma penetrans	Gln	0.063	0.017	-1.33	1:0
	Glu	0.094	0.033	-1.11	1:0
Deinococcus radiodurans	Gln	0.828	0.900	0.63	1:1
	Glu	0.561	0.376	-0.75	1:0
Agrobacterium tumefaciens	Gln	0.831	0.982	2.40	1:1
	Glu	0.427	0.273	-0.69	2:0
Clostridium perfringens	Gln	0.138	0.029	-1.67	2:0
	Glu	0.230	0.189	-0.25	3:0
Lactobacillus plantarum	Gln	0.360	0.046	-2.45	2:1
	Glu	0.245	0.030	-2.35	2:1
Sinorhizobium meliloti	Gln	0.818	0.973	2.09	1:1
	Glu	0.579	0.384	-0.79	3:1
Corynebacterium glutamicum	Gln	0.615	0.974	3.16	1:2
	Glu	0.437	0.689	1.04	1:3
Escherichia coli	Gln	0.653	0.813	0.84	2:2
	Glu	0.311	0.244	-0.34	4:0
Bacillus subtilis	Gln	0.488	0.200	-1.34	4:0
	Glu	0.320	0.227	-0.47	6:0
Schizosaccharomyces pombe	Gln	0.285	0.156	-0.77	4:2
	Glu	0.322	0.565	1.01	4:6
Saccharomyces cerevisiae	Gln	0.307	0.009	-3.89	9:1
	Glu	0.300	0.028	-2.70	14:2
Caenorhabditis elegans	Gln	0.343	0.307	-0.16	20:7
	Glu	0.375	0.436	0.25	17:24

$$S = \ln\left(\frac{n_G^{high} n_A^{low}}{n_A^{high} n_G^{low}}\right)$$



Positive *S* means G is preferred.
Negative *S* means A is preferred.

This depends on which tRNA genes are present.

For A+G families, alternative stable states exist in the same organism at the same time

Examples from E. coli

His	CAU
His	CAC
Gln	CAA
Gln	CAG

NU:NC = 2:2 S > 0 G preferred

Asp	GAU
Asp	GAC
Glu	GAA
Glu	GAG

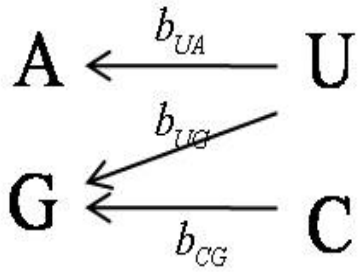
NU:NC = 4:0 S < 0 A preferred

Coevolution of tRNAs and codon usage:

Codon usage adapts to current tRNA genes,
AND tRNA genes must be stable to current codon usage.

Relate codon usage to translation kinetics

$$r_A = k_0 c_0 N_U b_{UA}$$



$$r_G = k_0 c_0 (N_U b_{UG} + N_C b_{CG})$$

$$S(N_U, N_C) = \sigma(t_A - t_G) = K\left(\frac{1}{N_U b_{UA}} - \frac{1}{N_U b_{UG} + N_C b_{CG}}\right)$$

Direction of selection depends on N's and b's

4 alternative states with $N_U + N_C = 4$
$N_U:N_C$ = 4:0 3:1 2:2 1:3
Vary $N_U:N_C$ by anticodon mutations.

$$S(N_U, N_C) = K\left(\frac{1}{N_U b_{UA}} - \frac{1}{N_U b_{UG} + N_C b_{CG}}\right)$$

$$\phi = \frac{\theta \exp(S)}{\theta \exp(S) + 1 - \theta}$$

mean time per codon:

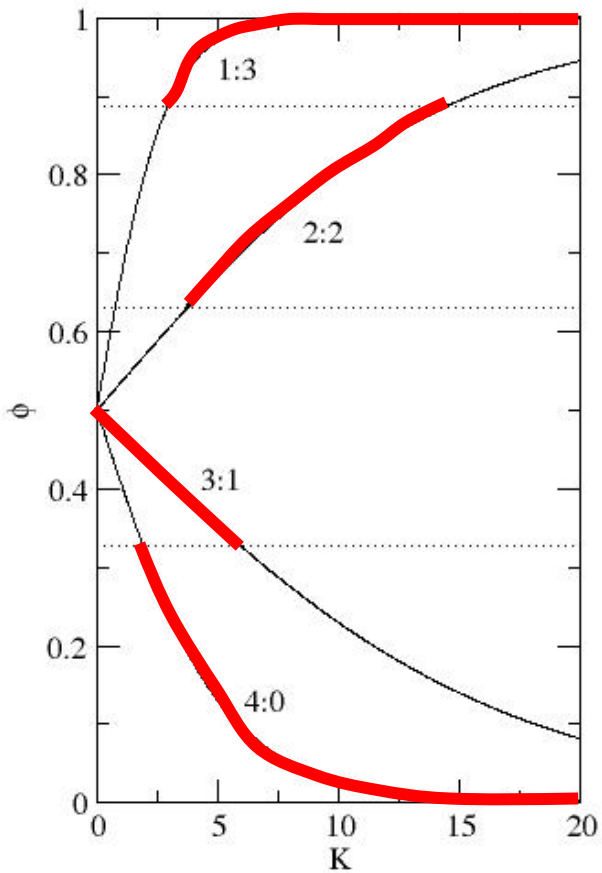
$$\bar{t}(\phi, N_U, N_C) = \phi t_G + (1-\phi) t_A = \frac{1}{k_0 c_0}\left(\frac{\phi}{N_U b_{UG} + N_C b_{CG}} + \frac{1-\phi}{N_U b_{UA}}\right)$$

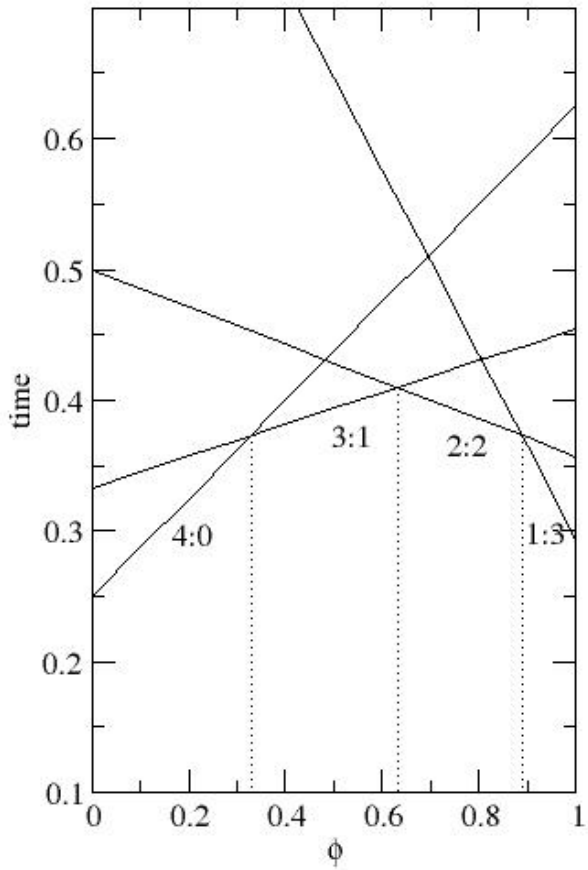
4 alternative states with $N_U + N_C = 4$ as with Gln and Glu in E. coli.

Vary N_U:N_C by anticodon mutations.

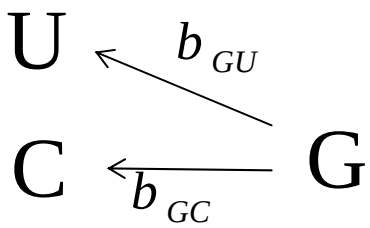
Red lines show stable states.

Assume $b_{UA} = b_{CG} = 1$, $b_{UG} = 0.4$





U + C case with variable tRNA copy number



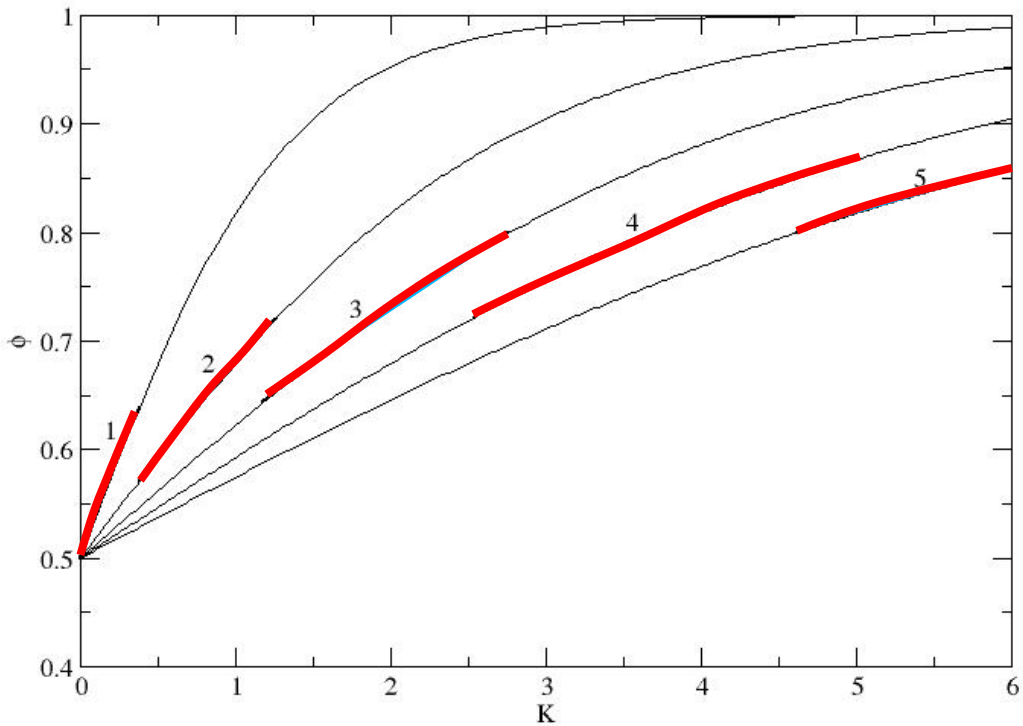
Gene copy number can vary by duplication and deletion.

g = net cost per gene.

f_a = freq of amino acid

T = total translational cost.

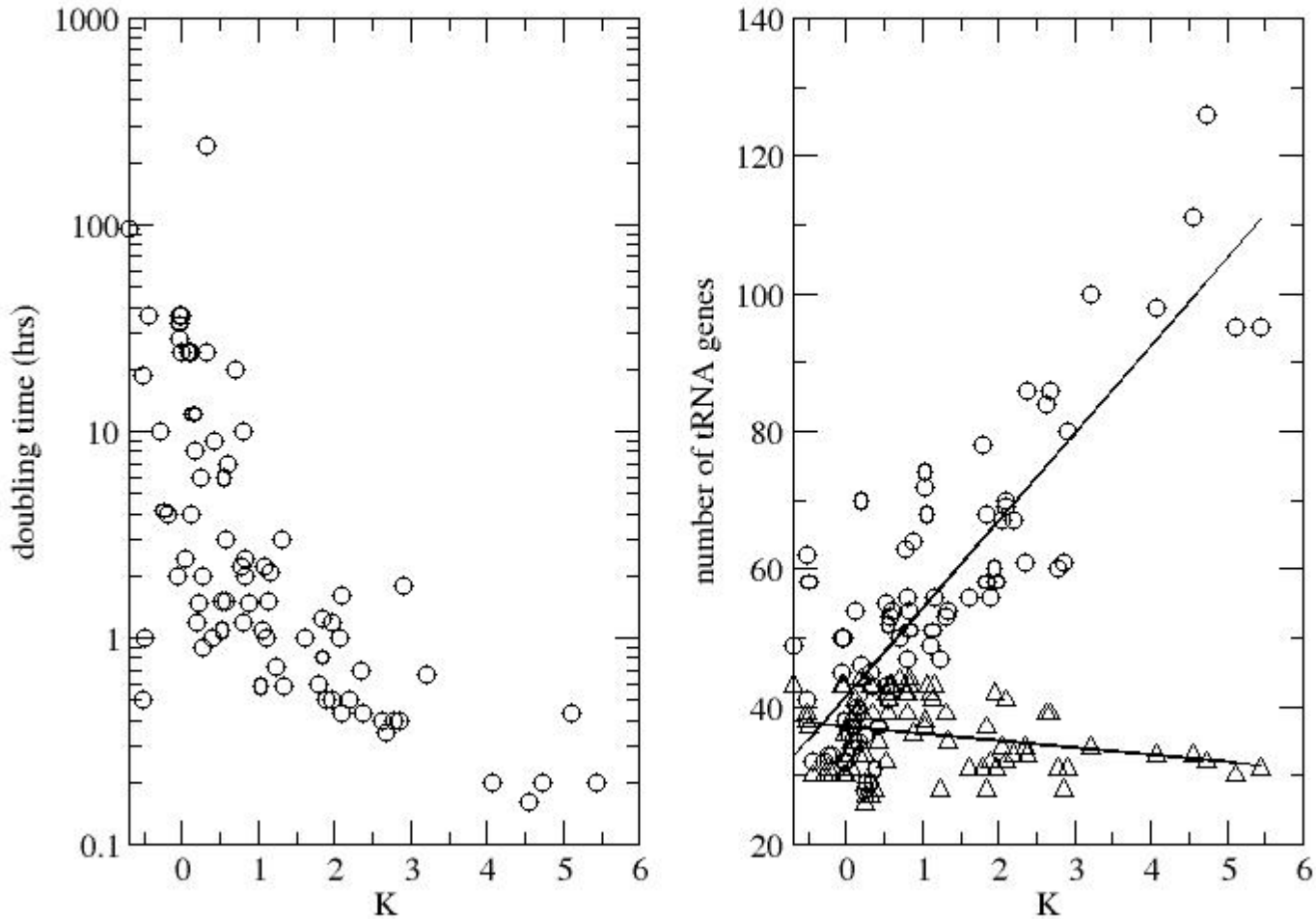
$$T(\phi, N_G) = \frac{f_a K}{N_G}\left(\frac{\phi}{b_{GC}} + \frac{1-\phi}{b_{GU}}\right) + gN_G$$



Gene duplication will be favoured in organisms where K is large (i.e. fast growing organims).

In these organisms we will see
(i)more tRNA genes
(ii)more biased codon usage

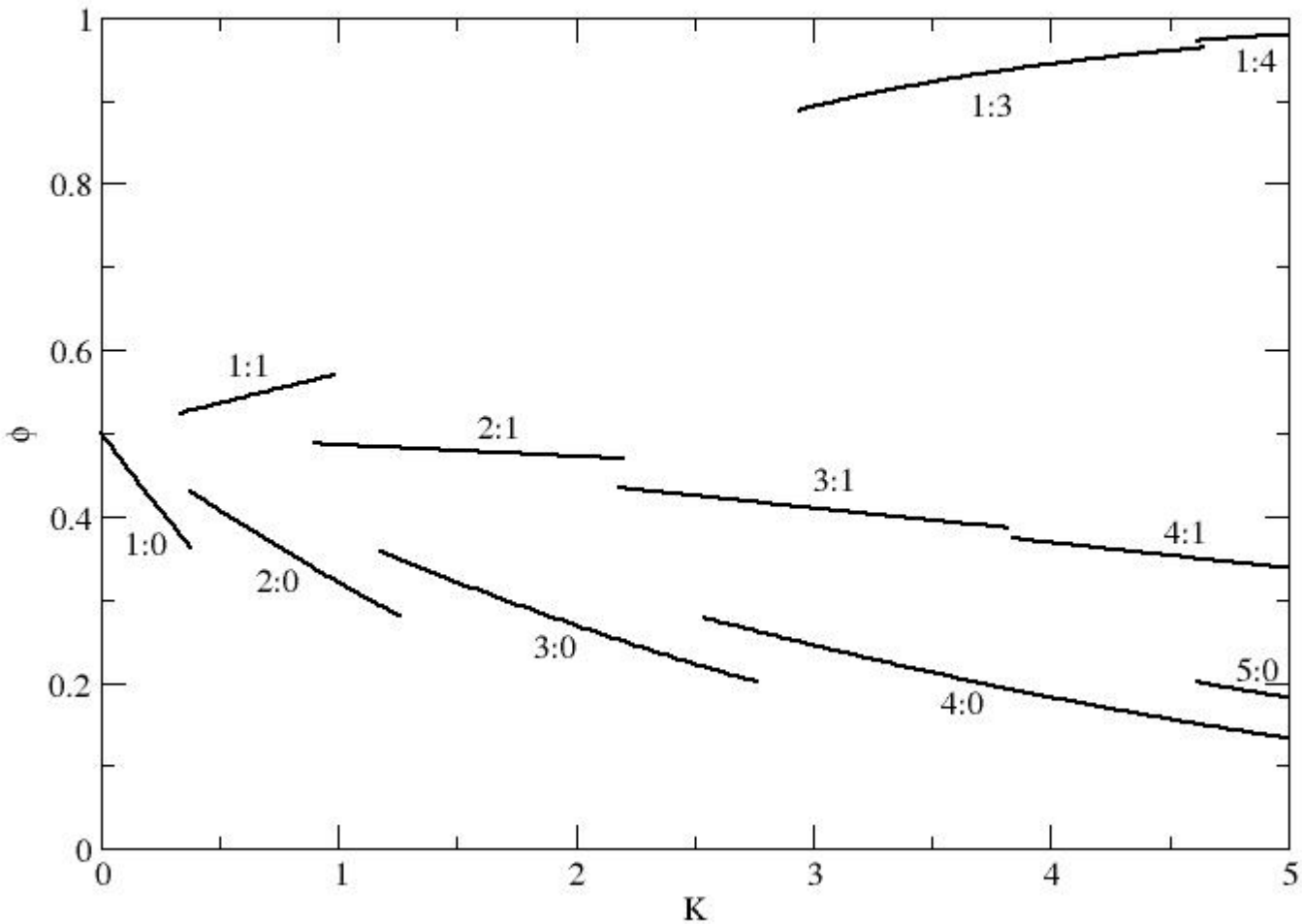
Test our theory using data on 80 bacterial genomes (Sharp et al. 2005)
Estimate K from an average of all the U+C amino acids.



U+C amino acids in 80 bacterial genomes . Is sign of effect correctly predicted?

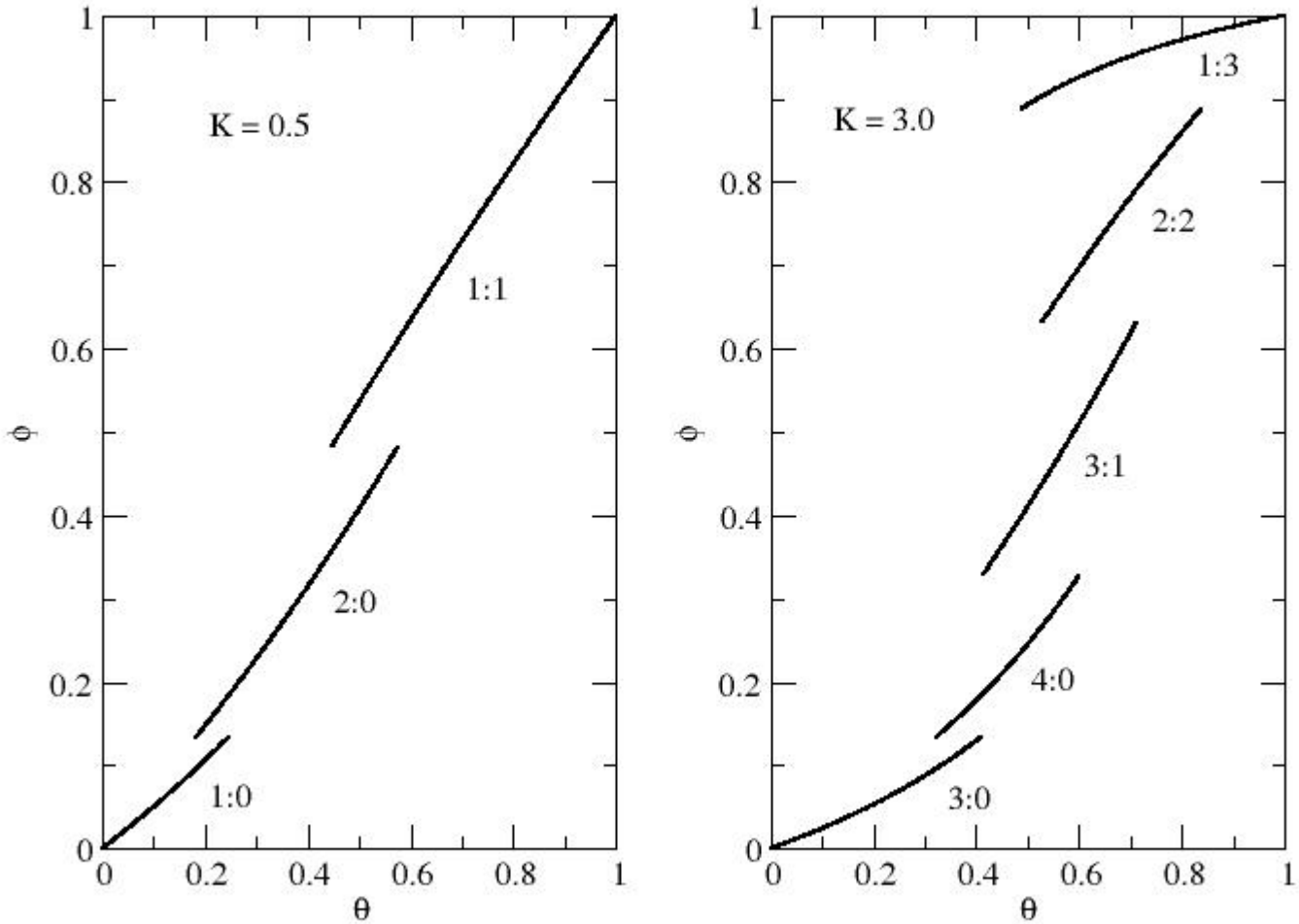
NG	tU	tC	tU-tC	Nobs	mean K	mean S	Nsign
1	2.50	1.00	1.50	207	0.61	0.75	193
2	1.25	0.50	0.75	94	1.82	1.19	94
3	0.83	0.33	0.50	57	2.08	1.03	56
4	0.63	0.25	0.38	23	3.38	1.67	23
5	0.50	0.20	0.30	10	3.00	1.37	9
6	0.42	0.17	0.25	3	4.80	1.07	3
7	0.36	0.14	0.21	1	4.74	3.10	1

A + G families – Allow both anticodon mutations and duplications/deletions of genes



Only stable regions are shown. Not all tRNA combinations have a stable region. Positions of stable regions depend on θ and on g/f_a

A + G families – Allow both anticodon mutations and duplications/deletions of genes.
Vary θ with K fixed.



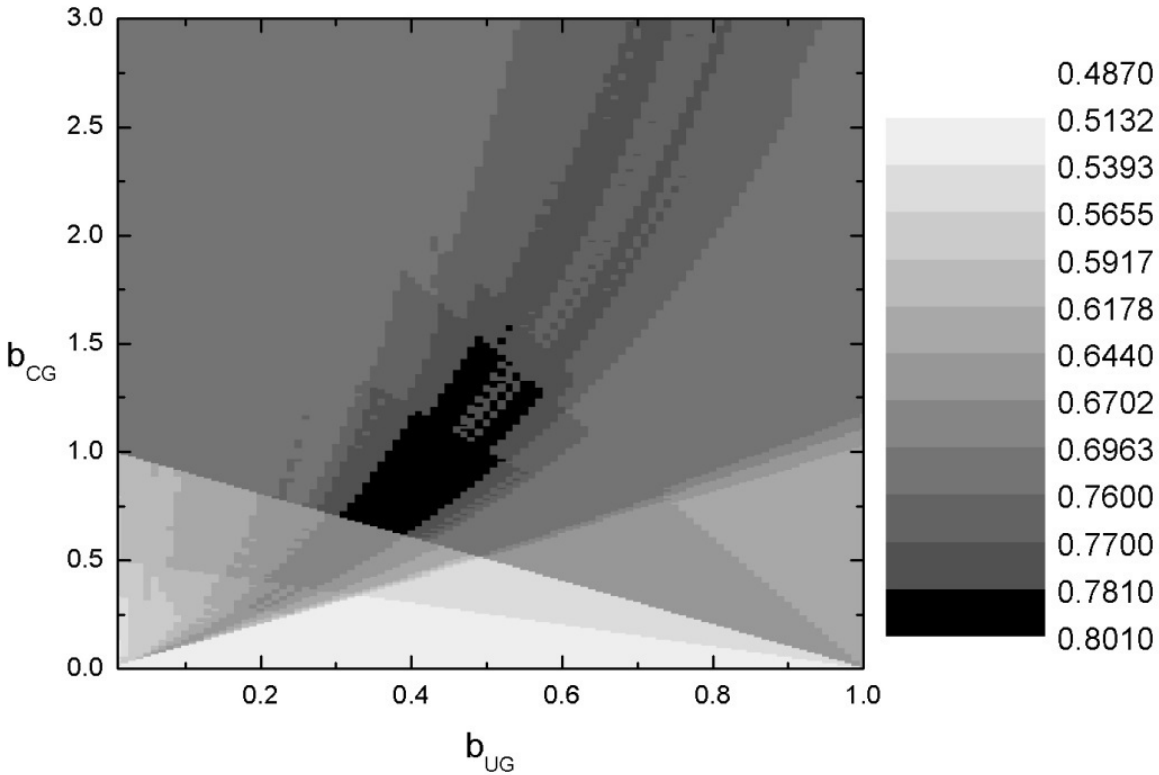
A + G amino acids in 80 bacterial genomes . Is sign of effect correctly predicted?

NU	NC	tA	tG	tA-tG	Nobs	mean K	mean S	Nsign
1	0	1.00	2.50	-1.50	46	0.32	-0.26	31
2	0	0.50	1.25	-0.75	21	1.80	-1.66	20
3	0	0.33	0.83	-0.50	16	2.21	-0.55	12
4	0	0.25	0.63	-0.38	15	2.85	-0.64	14
5	0	0.20	0.50	-0.30	8	3.35	-0.42	6
6	0	0.17	0.42	-0.25	5	3.55	-0.21	4
7	0	0.14	0.36	-0.21	2	4.16	0.07	1
1	1	1.00	0.71	0.29	47	0.64	0.73	37
2	1	0.50	0.56	-0.06	14	1.34	-0.40	8
3	1	0.33	0.46	-0.12	7	1.97	-0.64	5
1	2	1.00	0.42	0.58	5	1.32	2.68	5
2	2	0.50	0.36	0.14	3	1.83	0.48	3
4	2	0.25	0.28	-0.03	1	1.04	-0.23	1
7	2	0.14	0.21	-0.07	1	5.43	-0.19	1

Which values of b's best explain the data?

Current work – Wenqi Ran.

A + G amino acids in 80 bacterial genomes .
Fitness function w = +1 for every correctly predicted sign +1 for every case where observed tRNA configuration is stable to anticodon mutation



Values used in previous examples are close to optimal:
$b_{UA} = 1$ (fixed), $b_{CG} = 1$, $b_{UG} = 0.4$

Experiments: Are preferred codons really translated faster?

Curran and Yarus (1989) – Frameshift method in E. coli.

C codons are more rapidly translated than U codons for Phe, His, Tyr, Cys

G codon is more rapidly translated than A for Gln

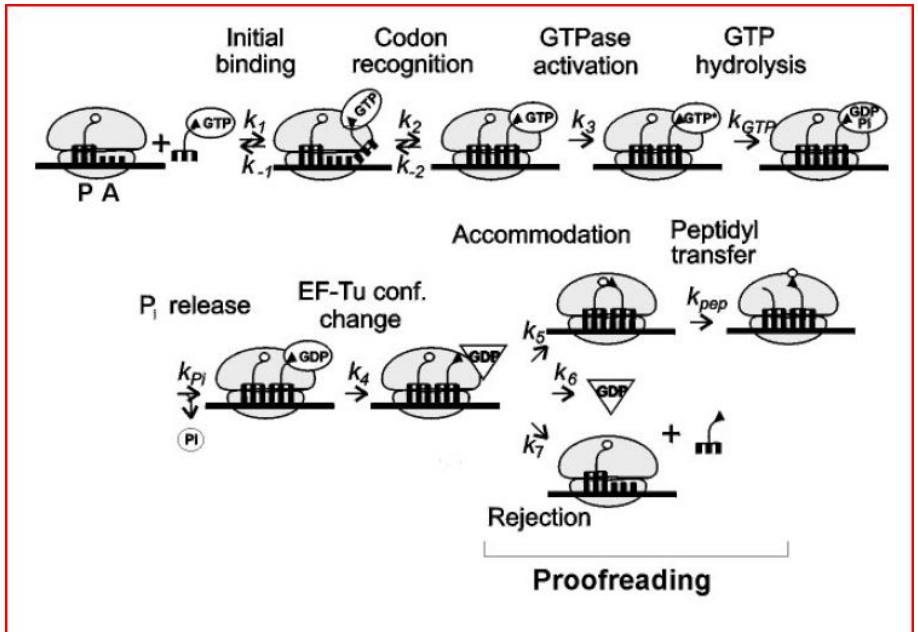
Sorensen and Pedersen (1991)

A codon is translated three times more rapidly than G for Glu

Complications …..

Rodnina et al (2001)

And what about accuracy…

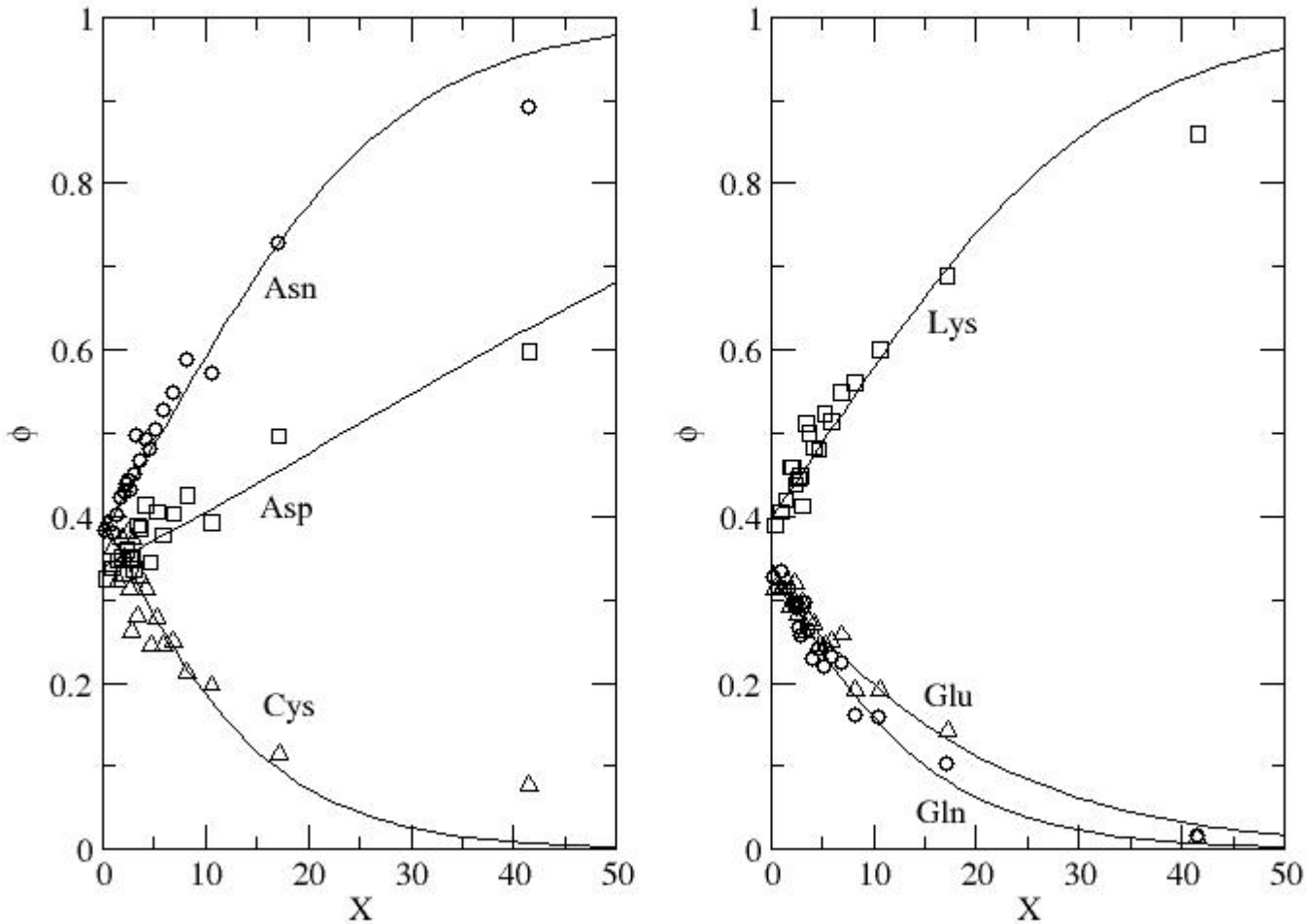


Is selection strength proportional to expression level?

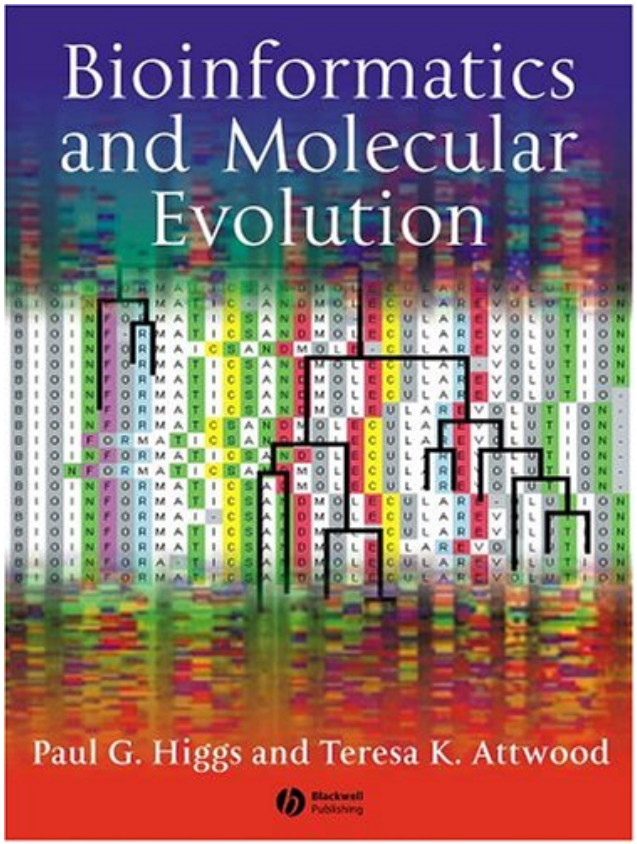
Genes from yeast binned according to expression level X (Akashi).

Assume S = kX.

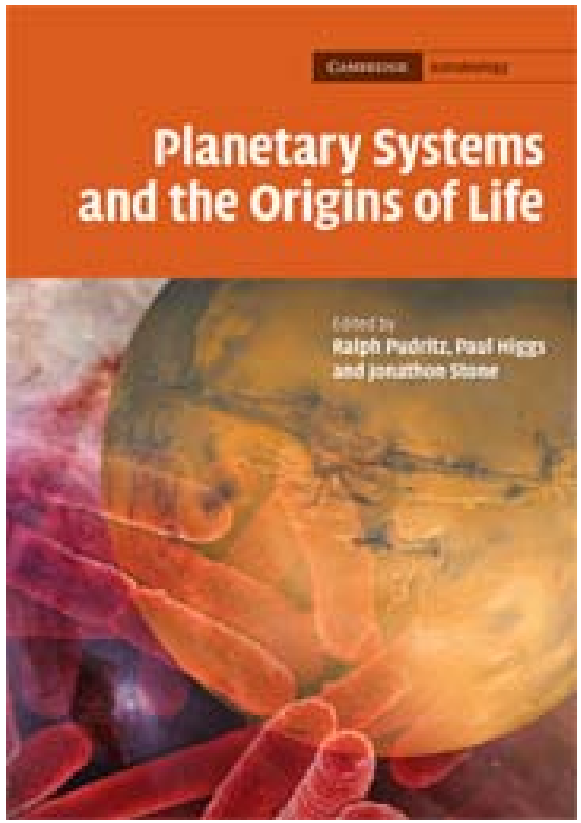
$$\phi(X) = \frac{\theta \exp(kX)}{\theta \exp(kX) + 1 - \theta}$$



End of Part 1 -Commercial Break



Higgs and Attwood
2005



Pudritz, Higgs and Stone
2007

OGRe. search results

Click the box next to an organism to select it. Clicking on the organism's name will bring up a new window with the medline links (if available) and raw data for that particular genome. After selecting genomes choose to view them, view sequences or codon usage using the buttons at the bottom of the screen.

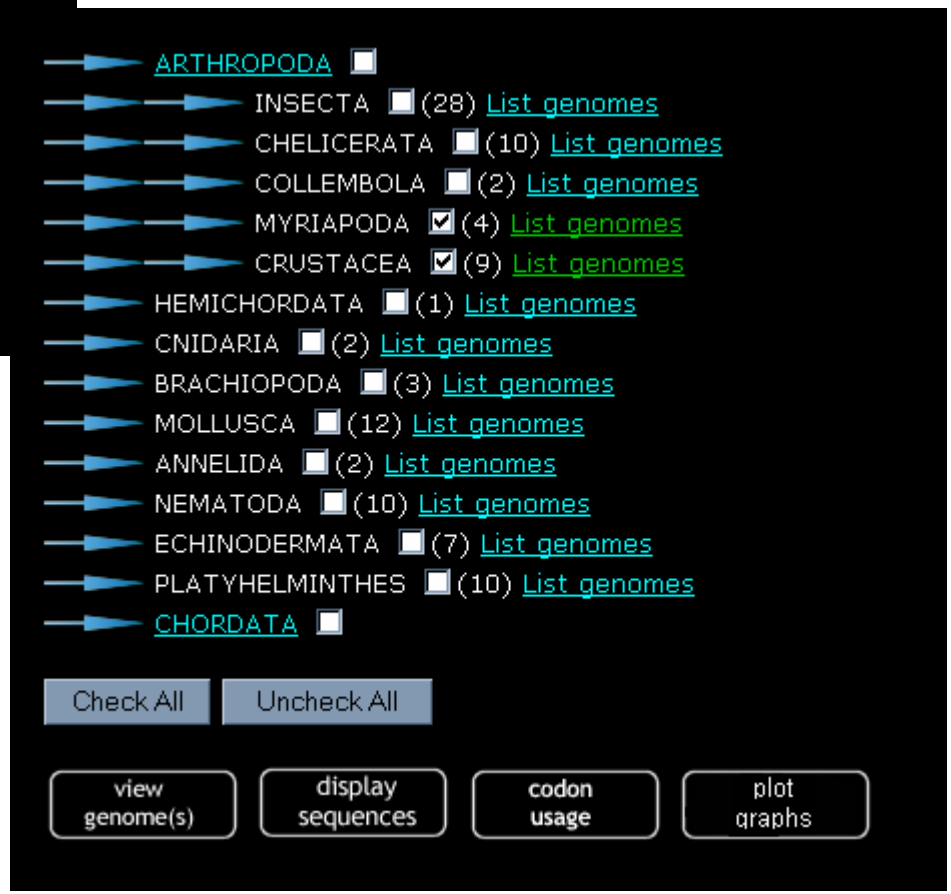
- [x] *Acipenser dabryanus* Yangtze sturgeon - mtDNA
- [x] *Acinonyx jubatus* cheetah - mtDNA
- [ ] *Acipenser stellatus* stellate sturgeon - mtDNA
- [ ] *Acipenser transmontanus* white sturgeon - mtDNA
- [ ] *Acropora tenuis* purple tipped acropora - mtDNA
- [x] *Albinaria caerulea* door snail - mtDNA
- [x] *Albula glossodonta* roundjaw bonefish - mtDNA
- [ ] *Aldrovandia affinis* Gilbert's halosaur - mtDNA
- [x] *Alepocephalus tenebrosus* California slickhead - mtDNA
- [ ] *Alligator mississippiensis* American alligator - mtDNA
- [x] *Allocyttus niger* black oreo - mtDNA
- [ ] *Alligator sinensis* Chinese alligator - mtDNA
- [ ] *Ambystoma mexicanum* axolotl - mtDNA

http://ogre.mcmaster.ca

Currently >1200 animal mitochondrial genomes.

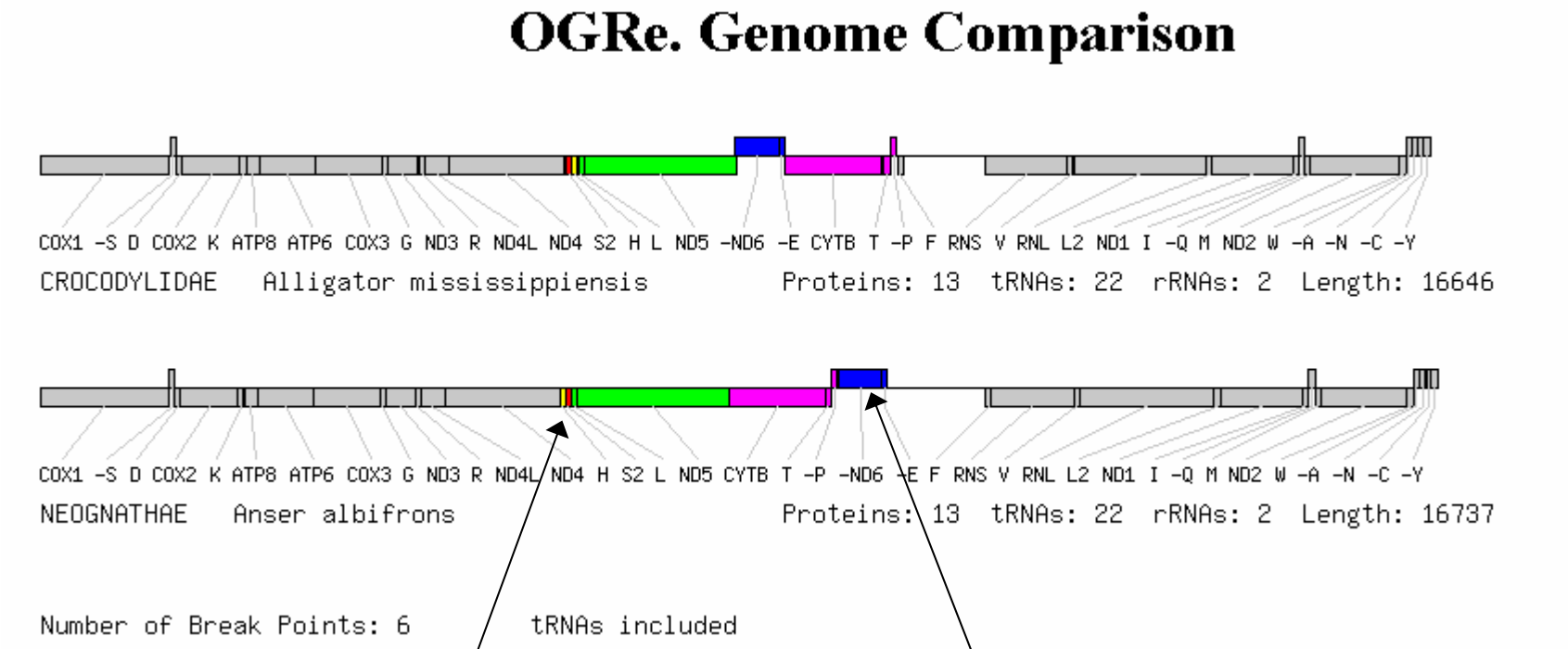
Species may be selected individually from an alphabetical list or by taxa

Information on gene sequences, gene order on genomes, codon usage etc.



Large Scale – Evolution of Gene Order in Whole Genomes

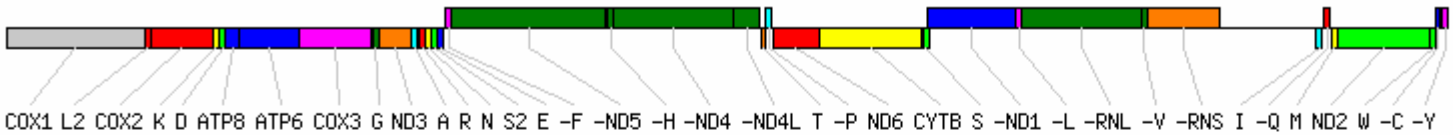
On the ogre web site, a visual comparison can be made of any two selected species. Colour is used to indicate conserved blocks of genes.



Alligator and Bird genomes differ by interchange of two tRNA genes (red and yellow)…

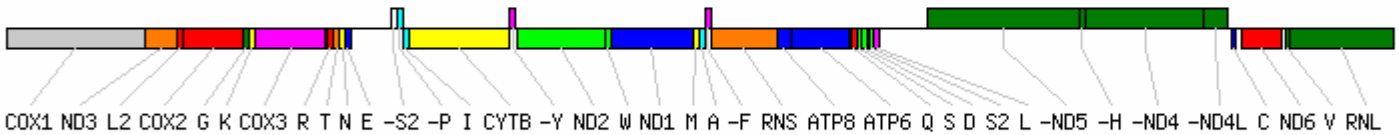
…and by translocation of the two genes in the blue block.

Sometimes things go crazy ….



COX1 L2 COX2 K D ATP8 ATP6 COX3 G ND3 A R N S2 E -F -ND5 -H -ND4 -ND4L T -P ND6 CYTB S -ND1 -L -RNL -V -RNS I -Q M ND2 W -C -Y

INSECTA Drosophila yakuba Proteins: 13 tRNAs: 22 rRNAs: 2 Length: 16019



COX1 ND3 L2 COX2 G K COX3 R T N E -S2 -P I CYTB -Y ND2 W ND1 M A -F RNS ATP8 ATP6 Q S D S2 L -ND5 -H -ND4 -ND4L C ND6 V RNL

INSECTA Thrips imaginis Proteins: 13 tRNAs: 23 rRNAs: 2 Length: 15407

Number of Break Points: 30 tRNAs included

Drosophila and *Thrips* are both insects

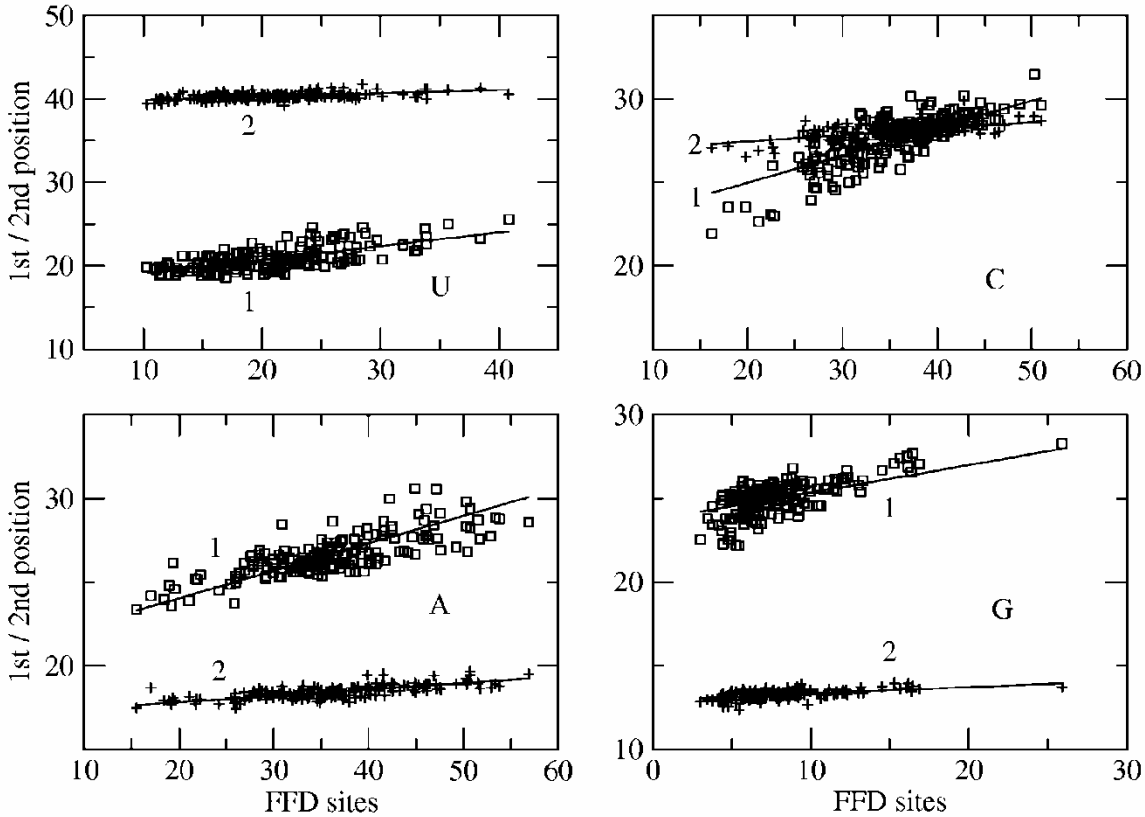
yet there are 30 breakpoints for only 37 genes

i.e. almost nothing in common.

Homo sapiens Strand = + 3624 codons

F	UUU	69	S	UCU	29	Y	UAU	35	C	UGU	5
F	UUC	139	S	UCC	99	Y	UAC	89	C	UGC	17
L	UUA	65	S	UCA	81	*	UAA	4	W	UGA	90
L	UUG	11	S	UCG	7	*	UAG	3	W	UGG	9
L	CUU	65	P	CCU	37	H	CAU	18	R	CGU	6
L	CUC	167	P	CCC	119	H	CAC	79	R	CGC	26
L	CUA	276	P	CCA	52	Q	CAA	82	R	CGA	28
L	CUG	42	P	CCG	7	Q	CAG	8	R	CGG	0
I	AUU	112	T	ACU	50	N	AAU	29	S	AGU	11
I	AUC	196	T	ACC	155	N	AAC	131	S	AGC	37
M	AUA	165	T	ACA	132	K	AAA	84	*	AGA	1
M	AUG	32	T	ACG	10	K	AAG	9	*	AGG	0
V	GUU	22	A	GCU	39	D	GAU	12	G	GGU	16
V	GUC	45	A	GCC	123	D	GAC	51	G	GGC	87
V	GUA	61	A	GCA	79	E	GAA	63	G	GGA	61
V	GUG	8	A	GCG	5	E	GAG	15	G	GGG	19

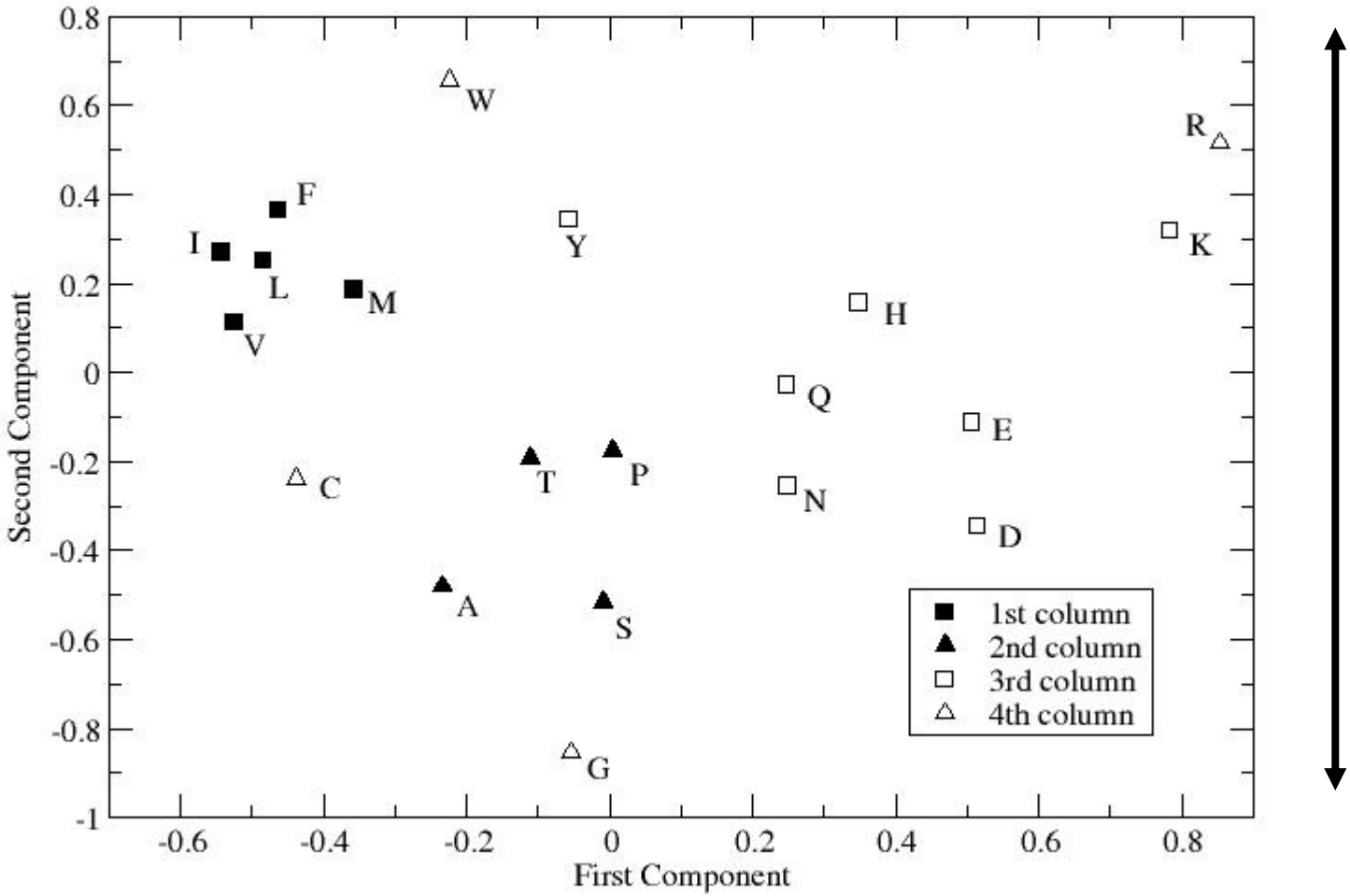
If there is no selection on codon usage, base frequencies at FFD sites are controlled by mutation. Base frequencies at 1st and 2nd positions are influenced by mutation *and* selection



Model fitting (Data from Fish) – assume a fraction of fixed sites and a fraction of neutral sites.

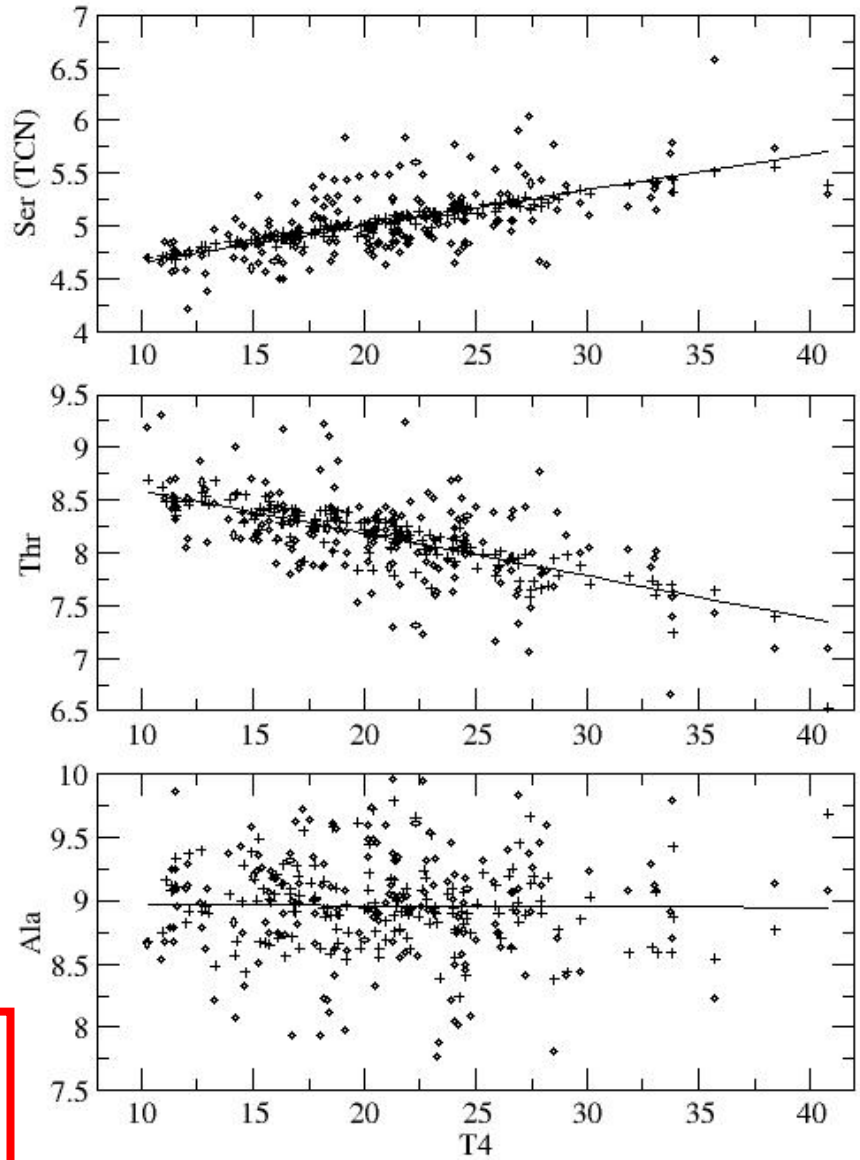
Selection at 1st position is weaker than at 2nd

Principal Component Analysis Projects the 8-d space into the two 'most important' dimensions.



Mutation pressure is sufficient to cause change in amino acid frequencies.

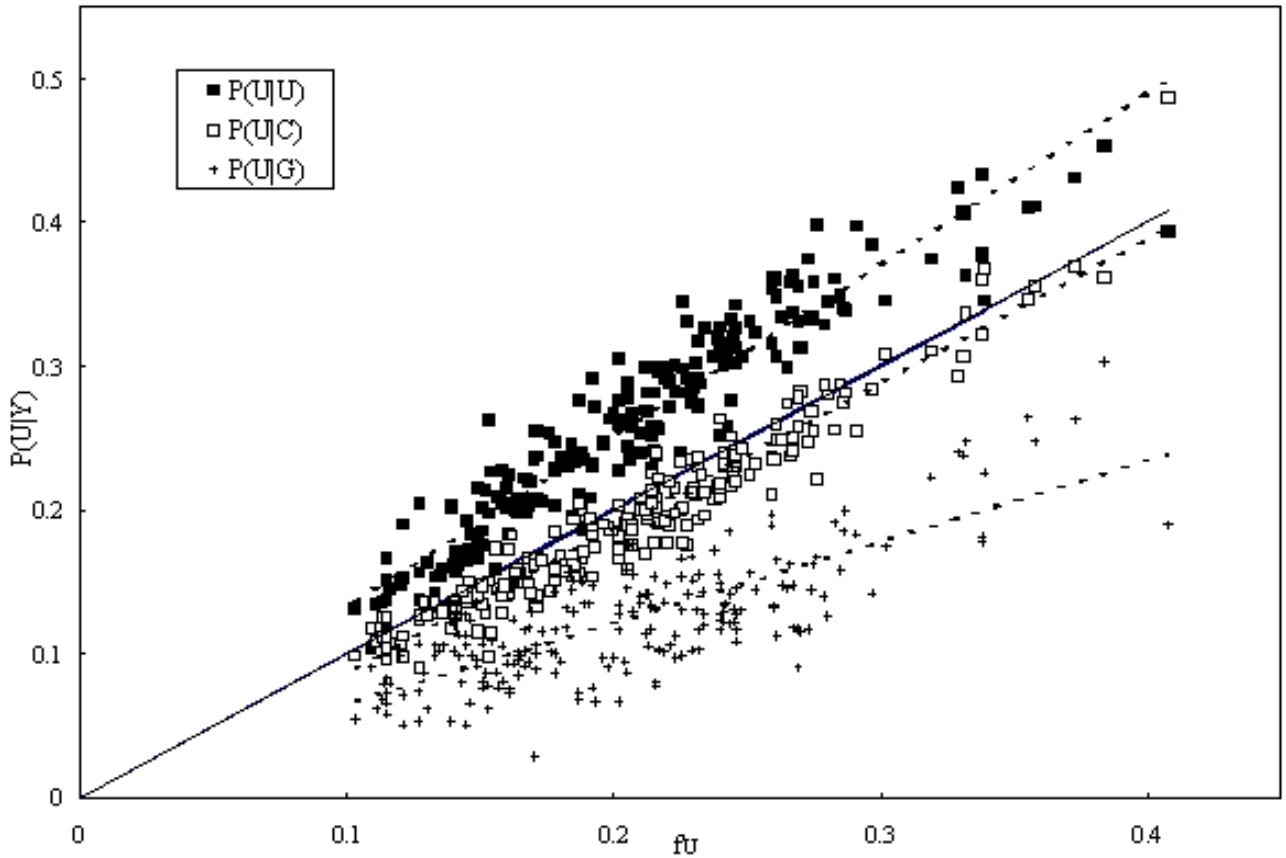
		Second Position				
		T	C	A	G	Third Pos.
First Position	T	F F	S S S S	Y Y	C C	T C A G
		L L L L L L		Stop Stop	W W	
	C		P P P P	H H	R R R R	T C A G
				Q Q		
	A	I I	T T T T	N N	S S	T C A G
		M M		K K	Stop Stop	
	G	V V V V	A A A A	D D	G G G G	T C A G
				E E		



Urbina et al (2006) J. Mol. Evol.
Can predict which amino acid frequencies vary most in terms of distance matrix.

Context-dependent mutation

Frequency of U at FFD sites following a U, C or G



Frequency of U at FFD sites

Context-dependent mutation causes correlations between neighbouring bases (Jia and Higgs, 2008)

Frequency ratios

$$r(X_2Y_3) = \frac{p(X_2Y_3)}{q(X_2)q(Y_3)}$$

Fish - 23					
UU	1.250	CU	0.939	GU	0.605
UC	0.756	CC	1.205	GC	0.878
UA	1.030	CA	0.938	GA	1.145
UG	1.274	**CG**	**0.554**	**GG**	**1.891**
Mammals - 23					
UU	0.939	CU	1.101	GU	0.763
UC	0.743	CC	1.163	GC	1.005
UA	1.136	CA	0.906	GA	1.027
UG	1.433	**CG**	**0.552**	**GG**	**1.654**

Fish - 31							
UU	0.933	CU	1.162	AU	0.907	GU	0.911
UC	0.918	CC	1.371	AC	0.739	GC	0.839
UA	1.096	CA	0.849	AA	1.135	GA	0.758
UG	1.049	**CG**	**0.609**	AG	1.228	**GG**	**1.499**
Mammals - 31							
UU	0.855	CU	1.082	AU	0.996	GU	1.115
UC	0.994	CC	1.363	AC	0.797	GC	0.873
UA	1.206	CA	0.945	AA	0.974	GA	0.776
UG	0.856	**CG**	**0.546**	AG	1.293	**GG**	**1.369**

Mitochondrial genome replication

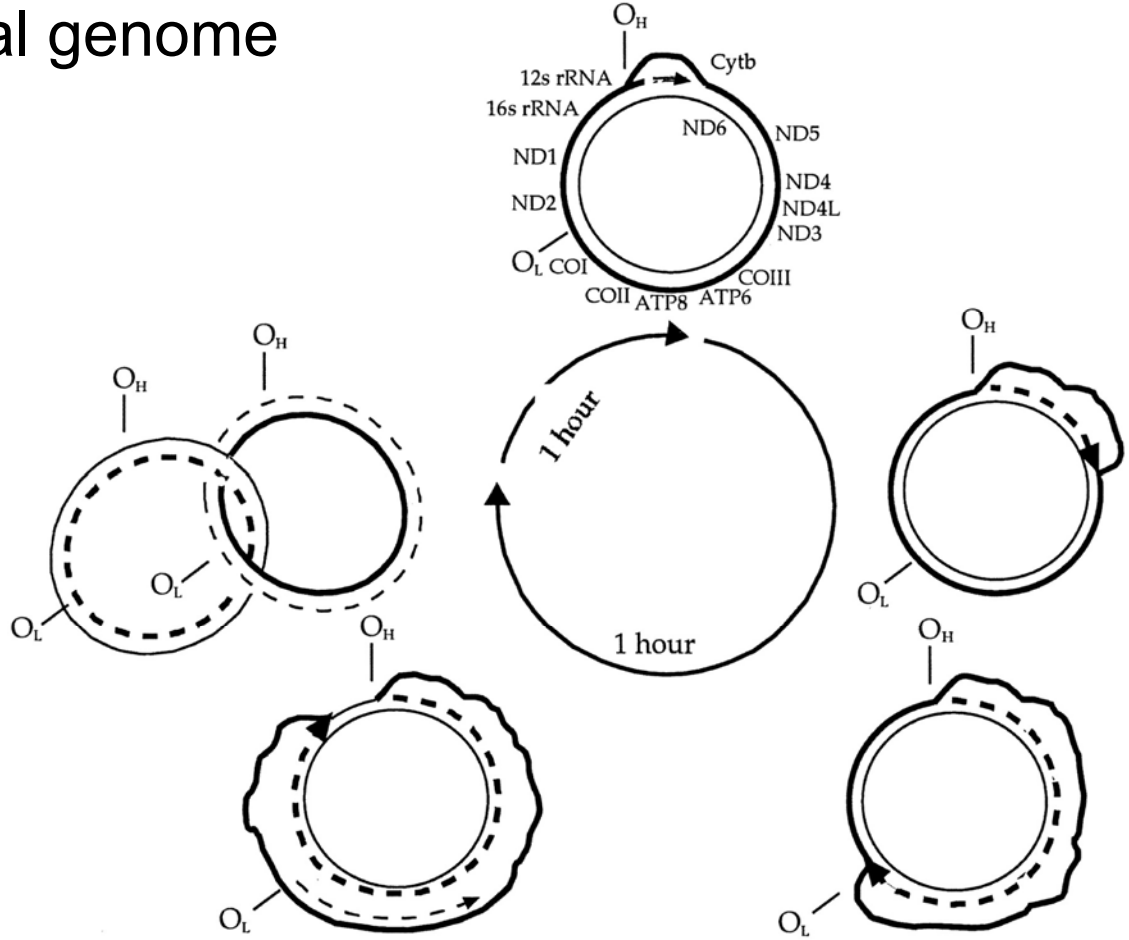


Figure from

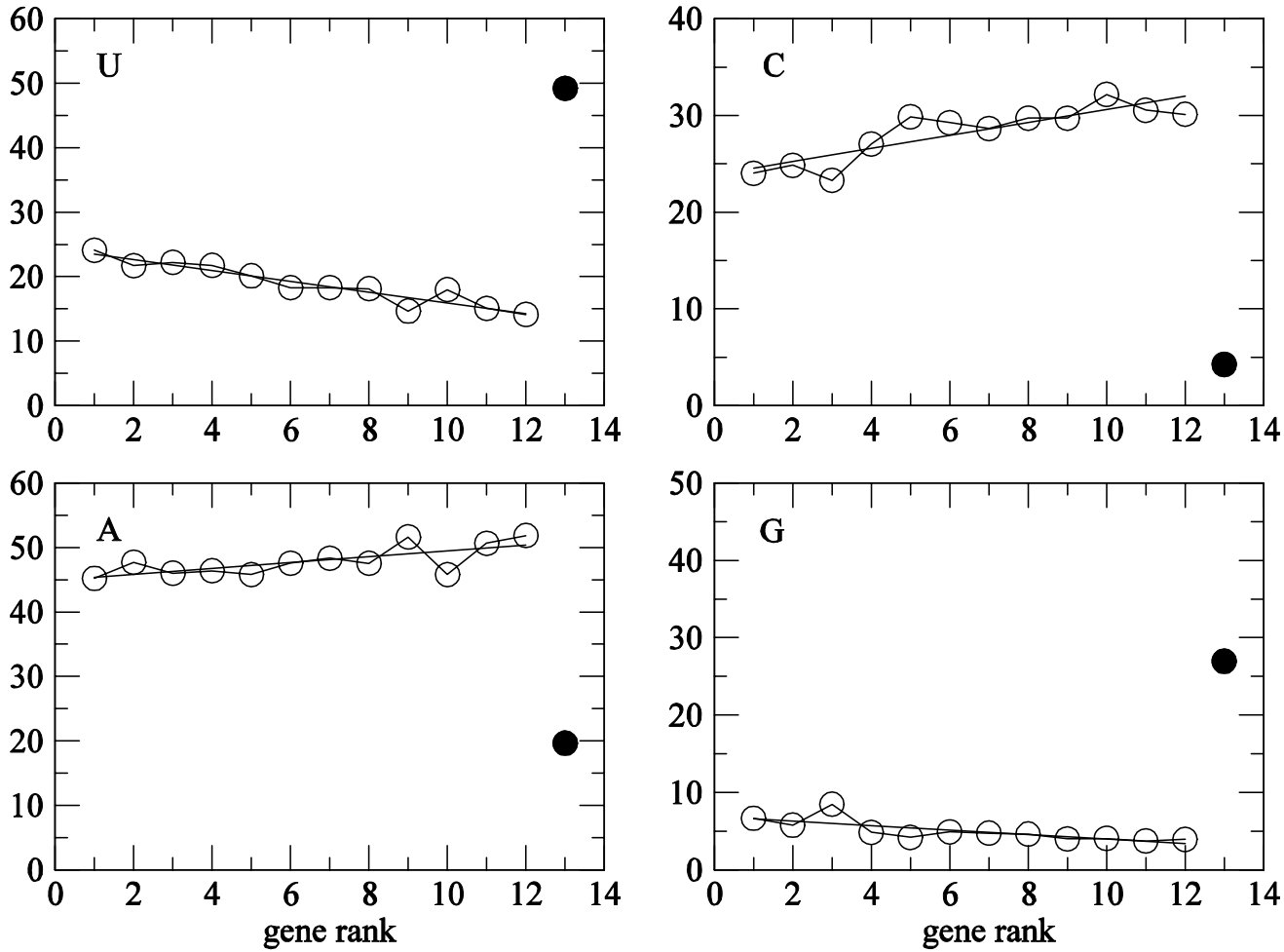
Faith & Pollock (2003) *Genetics*

Rank genes in order of increasing time spent single stranded

COI < COII < ATP8 < ATP6 < COIII < ND3 < ND4L < ND4 < ND1 < ND5 <ND2 < Cytb

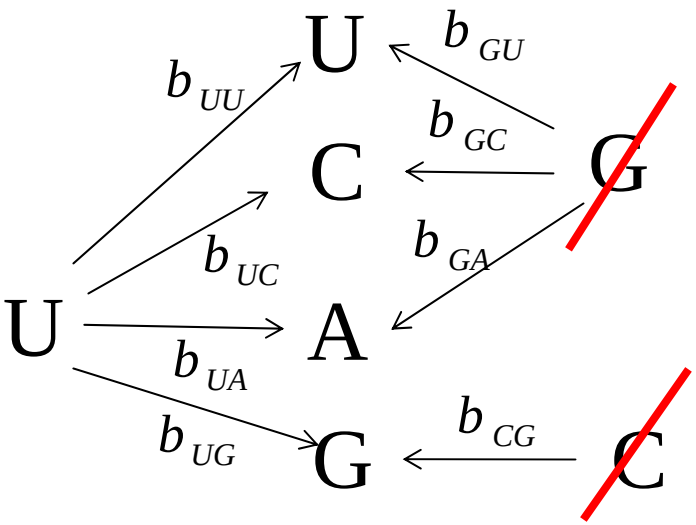
ND6 is on the other strand

Base frequencies at FFD sites in each gene (averaged over mammals)



Deamination: C to U and A to G on the heavy strand

Can there be translational selection as well as mutational effects?



For all four codon families in mitochondria there is only 1 tRNA gene with wobble position U.

We expect minimal numbers of tRNA genes when translational selection is negligible.

This tRNA must be sufficiently versatile to translate all codons, but in priciple the U could prefer one codon over another.

Probably little selection, for Translational Speed, but if there were selection it should be the same in all families

Cannot use high/low expression comparison to detect selection for Speed.

But selection for Translational Accuracy can be detected by comparing conserved and variable sites (Akashi).

Homo sapiens Strand = + 3624 codons

F	UUU	69	S	UCU	29	Y	UAU	35	C	UGU	5
F	UUC	139	S	UCC	99	Y	UAC	89	C	UGC	17
L	UUA	65	S	UCA	81	*	UAA	4	W	UGA	90
L	UUG	11	S	UCG	7	*	UAG	3	W	UGG	9
L	CUU	65	P	CCU	37	H	CAU	18	R	CGU	6
L	CUC	167	P	CCC	119	H	CAC	79	R	CGC	26
L	CUA	276	P	CCA	52	Q	CAA	82	R	CGA	28
L	CUG	42	P	CCG	7	Q	CAG	8	R	CGG	0
I	AUU	112	T	ACU	50	N	AAU	29	S	AGU	11
I	AUC	196	T	ACC	155	N	AAC	131	S	AGC	37
M	AUA	165	T	ACA	132	K	AAA	84	*	AGA	1
M	AUG	32	T	ACG	10	K	AAG	9	*	AGG	0
V	GUU	22	A	GCU	39	D	GAU	12	G	GGU	16
V	GUC	45	A	GCC	123	D	GAC	51	G	GGC	87
V	GUA	61	A	GCA	79	E	GAA	63	G	GGA	61
V	GUG	8	A	GCG	5	E	GAG	15	G	GGG	19

Chi squared tests for frequencies of bases at FFD sites

	Not significant $p > 0.05$	Significant $0.001 < p \leq 0.05$	Highly significant $p \leq 0.001$
Total Fish Species 214			
Expected No.	203.3	10.5	0.2
UN/CN/GN	0	0	214
UCN/CCN/ACN/GCN (SPTA blocks)	58	75	81

The FFD base is not independent of the second position base in every single species.

There are many species for which there is a difference between four-codon families with the same second base.

Likelihood-based tests for factors influencing codon usage

Z is the base at the fourfold degenerate site. It may depend on:

A – tRNA differences between amino acids
B – translational accuracy (conserved v. variable)
D – gene position (time spent single stranded)
X – context dependent mutation (following 1st pos)

data = numbers of codons

probabilities from a theoretical model

$$\ln L = \sum_A \sum_B \sum_D \sum_X \sum_Z \left(n_{ABDX}(Z) \ln P_{ABDX}(Z) \right)$$

Use Akaike's Information Criterion as a means of Model Selection

$$AIC = 2(-\ln L + \#params)$$

Want to choose a model that has sufficient parameters to explain the trends in the data but not overfit.
Minimize AIC → Maximize likelihood subject to a penalty on too many parameters

Comparison of models using AIC

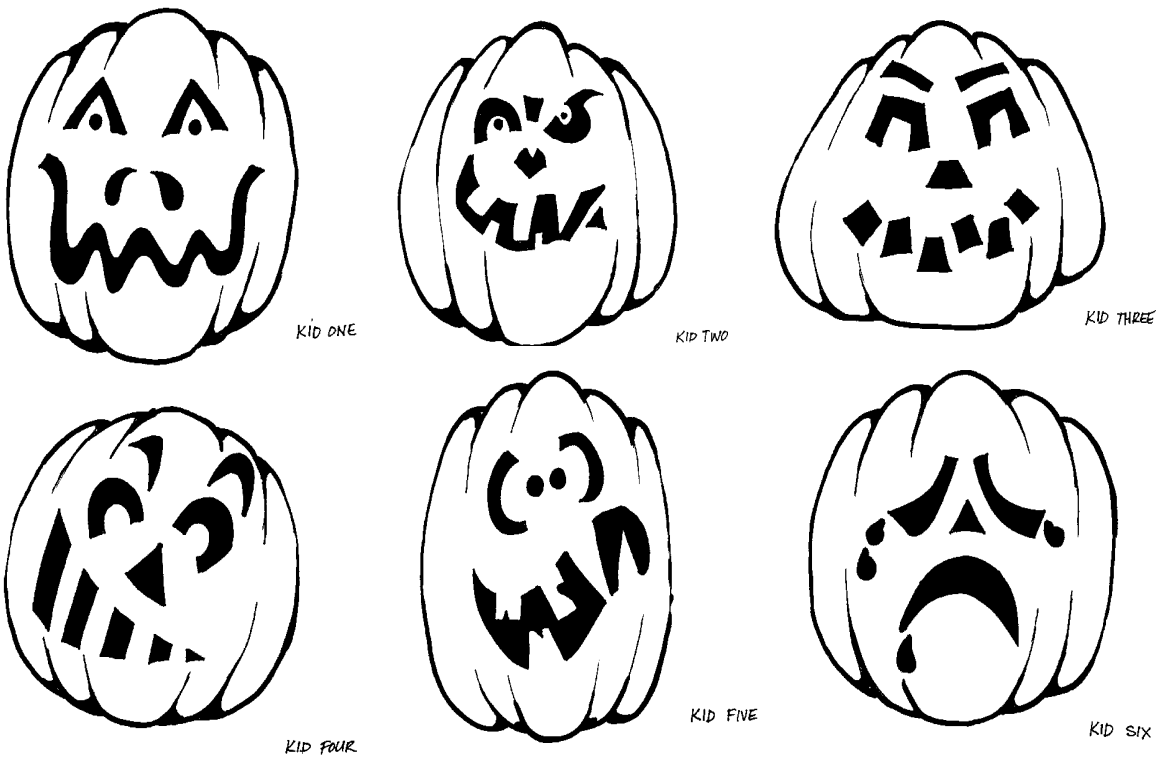
Data for SPTA blocks - Jia & Higgs (2008)

Model	ΔAIC for *Homo sapiens*	Average ΔAIC	# of species for which ΔAIC < 0	# of species for which this model is the best of 0 and single-factor models	# of species for which this model is the best of all models	# of species for which this factor is included in the best model
0	0.00	0.00	----	0	0	----
A	-2.02	-6.96	28	6	3	12
B	5.26	2.78	6	0	0	2
D	-3.06	-2.26	27	2	0	8
X	-19.08	-29.61	38	**32**	**21**	**34**
AB	12.45	3.97	15	----	1	----
AD	4.30	-1.62	22	----	2	----
AX	-6.49	-19.12	33	----	6	----
BD	6.85	2.77	14	----	0	----
BX	-2.24	-18.15	32	----	1	----
DX	-17.50	-24.09	38	----	6	----

Consider 40 unrelated independent species.
Factor X has the largest effect (context dependent mutation) in almost all species.
The model chosen by AIC is either X alone or X in combination with another factor.

Conclusion – translational selection dominated by context-dependent mutation in mitochondria

Polymorphisms in Cucurbita



Pumpkin pumpkin
big and round,
I'm glad you grow
upon the ground.
I'm glad you don't
grow in a tree
for then you might
fall down on me.

Human Genetics and Heredity

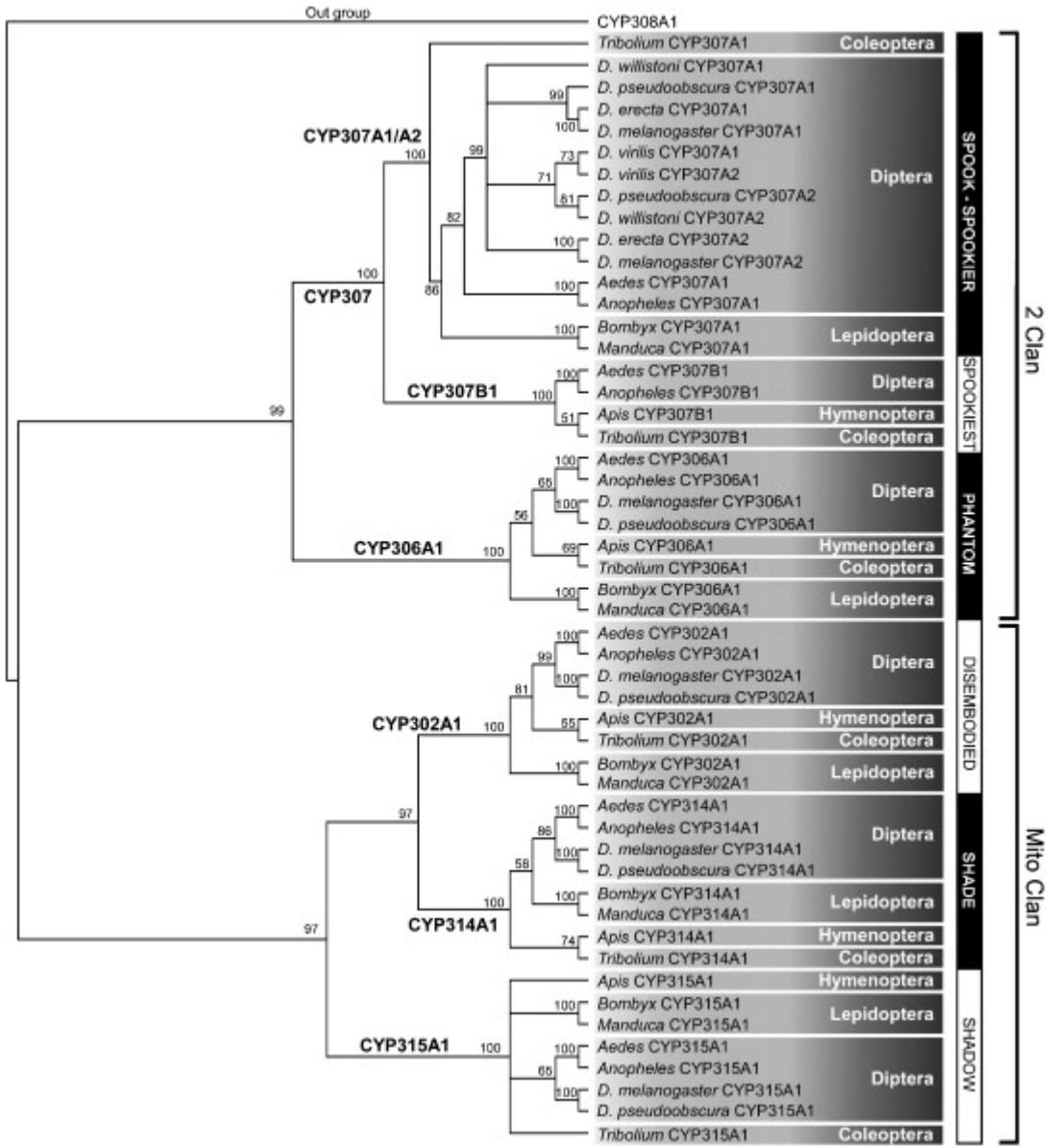
Hereditary Witch? Are you Born a Witch?
June 28, 2005 9:40 AM

Can you be a Witch because your ancestry was one? I guess that my answer my seem odd to many since my own Mother and Great Grandmother both were practicing witches and the communities knew it. But NO, I don't personally think that you can be a witch simply because your mother, or father or Grandmother or heck your great great great grandmother was a Witch. I don't think it works that way.

Harry Potter: The study looked at how wizarding genes could be passed down the generations.

Now an analysis of wizardry, published in the British Medical Journal, has concluded that there is indeed good evidence that magical abilities are passed down the generations. Based on an analysis of the Harry Potter novels, Sreeram Ramagopalan, Dr Marian Knight, Prof George Ebers, and Dr Julian Knight of Oxford's Wellcome Trust Centre for Human Genetics, conclude that "magic shows strong evidence of heritability."

There really are Halloween Genes in Drosophila!



Phylogenetic analysis of the insect Halloween family of P450s.

Rewitz et al (2007)
Insect Biochem and Mol. Biol.