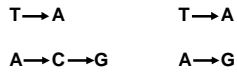


POSITIVE SELECTION

Analysis

$$\Delta \begin{matrix} & \text{A} & \text{S} \\ \text{M}_A & & \text{M}_S \\ \text{N}_A - \text{M}_A & & \text{N}_S - \text{M}_S \\ \text{N}_A & & \text{N}_S \end{matrix} \quad \begin{matrix} \text{M}_A + \text{M}_S \\ \text{L} - \text{M}_A - \text{M}_S \\ \text{L} \end{matrix}$$

Parallelism & Convergence



NUCLEOTIDE SUBSTITUTION PATTERNS

$$P_{ij} = n_{ij} / n_i$$

$$f_{ij} = 100 (P_{ij} / (\sum \sum P_{ij}))$$

Inference

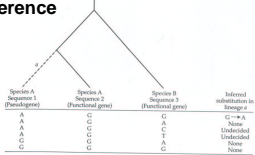


FIGURE 4.9 A tree for inferring the pattern of nucleotide substitution in a pseudofunctional gene sequence. The dashed line indicates "nonfunctional." In cases where the nucleotide occupying homologous sites in sequence 2 and 3 are identical, that different from the nucleotide in sequence 1, the type of substitution in lineage a can be unambiguously inferred.

TABLE 4.5 Pattern of substitution in pseudogene^a

From	To				Row totals
	A	T	C	G	
A	—	3.4 ± 0.7 (3.6 ± 0.70)	4.5 ± 0.8 (4.8 ± 0.9)	12.5 ± 1.1 (13.3 ± 1.1)	20.3 (21.4)
T	3.3 ± 0.6 (3.5 ± 0.6)	—	12.8 ± 1.9 (14.7 ± 2.0)	3.3 ± 0.6 (3.5 ± 0.6)	20.4 (21.7)
C	4.2 ± 0.5 (4.2 ± 0.5)	20.7 ± 1.3 (16.4 ± 1.2)	—	4.8 ± 0.6 (4.4 ± 0.6)	29.5 (22.3)
G	20.4 ± 1.4 (21.9 ± 1.5)	4.4 ± 0.6 (4.6 ± 0.6)	4.9 ± 0.7 (5.2 ± 0.8)	—	29.7 (31.6)
Column totals	27.9 (29.5)	28.5 (24.6)	23.2 (23.2)	20.5 (21.3)	

^aCourtesy of Dr. Alan Cooper.

^bTable entries are the inferred percentages (%) of nucleotide changes from i to j based on 107 presumed pseudogene sequences from humans. Values in parentheses were obtained by excluding all CC dinucleotides from the comparison.

TABLE 4.6 Pattern of substitution in the control region of human mitochondrial DNA^a

From	To				Row totals
	A	T	C	G	
A	—	0.4	1.1	14.3	15.8
T	0.3	—	35.8	0.3	36.4
C	1.1	25.8	—	0.5	27.4
G	20.0	1.3	1.6	—	22.7
Column total	21.4	27.3	38.5	14.9	

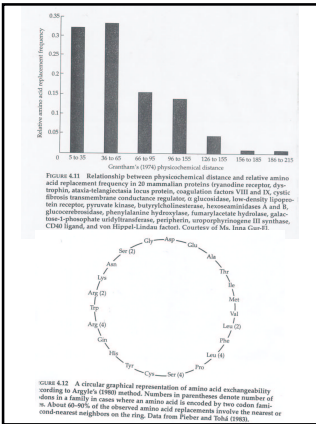
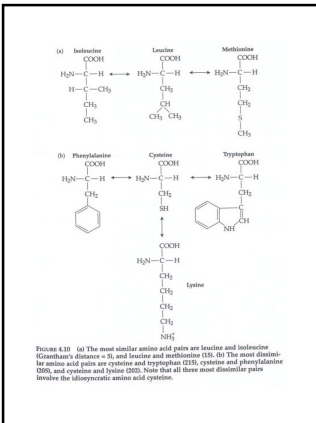
^aFrom Hatters and Nei (1993).

^bTable entries are the inferred percentages (%) of nucleotide changes from i to j based on 95 sequences.

TABLE 4.7. Physicochemical distances between pairs of amino acids*

Arg	Leu	Pro	Thr	Ala	Val	Gly	Ile	Phe	Tyr	Cys	His	Gln	Asn	Lys	Asp	Glu	Met	Ser	
110	145	74	58	99	124	56	142	155	144	112	89	68	66	121	65	80	135	177	Ser
102	103	71	113	94	123	97	97	77	180	29	43	86	26	96	24	91	101	101	Arg
98	92	56	32	138	88	22	36	198	99	113	153	107	172	138	13	41	41	41	Leu
38	27	68	42	95	114	110	149	77	76	91	103	108	93	87	147	76	76	76	Pro
58	49	29	49	103	92	149	47	42	65	79	85	83	83	128	78	78	78	78	Thr
64	69	94	113	112	195	86	91	113	106	125	107	84	148	148	84	148	148	148	Ala
109	29	70	53	102	84	86	113	97	151	121	21	89	76	76	76	76	76	76	Ile
135	153	147	139	98	87	80	127	94	98	127	184	184	184	184	184	184	184	184	Gly
21	23	198	94	109	149	148	134	137	43	43	43	43	43	43	43	43	43	43	Phe
22	205	103	114	136	102	177	140	28	43	43	43	43	43	43	43	43	43	43	Pro
194	83	99	143	105	140	122	36	37	37	37	37	37	37	37	37	37	37	37	Tyr
174	134	139	102	154	170	196	196	196	196	196	196	196	196	196	196	196	196	196	Cys
24	68	32	81	80	87	115	115	115	115	115	115	115	115	115	115	115	115	115	His
46	53	41	29	183	120	120	120	120	120	120	120	120	120	120	120	120	120	120	Gln
94	13	42	142	174	174	174	174	174	174	174	174	174	174	174	174	174	174	174	Asn
101	60	62	119	119	119	119	119	119	119	119	119	119	119	119	119	119	119	119	Lys
45	140	181	181	181	181	181	181	181	181	181	181	181	181	181	181	181	181	181	Asp
152	152	152	152	152	152	152	152	152	152	152	152	152	152	152	152	152	152	152	Glu
67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	67	Met

*From Goodman (1976).
*Mean distance is 100. The largest and smallest distances are emphasized with shading.



BIASED SYNONYMOUS CODON USAGE

$$ENC = 2 + 9F_2^{-1} + F_3^{-1} + 5F_4^{-1} + 3F_6^{-1}$$

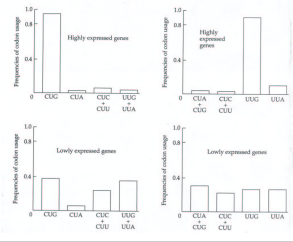


TABLE 3 Codon usage in *Escherichia coli* and *Saccharomyces pombe*

Amino acid	Codon	<i>E. coli</i>		<i>S. cerevisiae</i>	
		High	Low	High	Low
Leu	UUA	0.06	1.24	0.49	1.85
	UUG	0.07	0.87	5.34	1.48
	CUU	0.13	0.72	0.02	4.73
	CUC	0.17	0.65	0.00	0.51
	CUA	0.04	0.31	0.15	0.95
Val	CUG	5.54	2.20	0.02	0.84
	GUU	2.41	1.09	2.07	1.13
	GUC	0.08	0.99	1.95	0.76
	GUA	1.12	0.63	0.00	0.93
	GUG	0.40	1.29	0.02	0.93
Ile	AUU	0.48	1.38	1.26	1.29
	AUC	2.51	1.12	1.74	0.66
	AUA	0.01	0.20	0.00	1.05
Phe	UUU	0.34	1.33	0.19	1.38
	UUC	1.66	0.87	1.81	0.62

From Sharp et al. (1988)

*For each group of synonymous codons, the sum of the relative frequencies equals the number of codons in the group. For example, there are six codons for leucine, and the sum of the relative frequencies for these six codons is 6. Under equal codon usage, the relative frequency of each codon in a group should be 1, and on the degree of deviation from 1 indicates the level of bias in codon usage. "High" and "Low" denote genes with high and low expression levels, respectively.

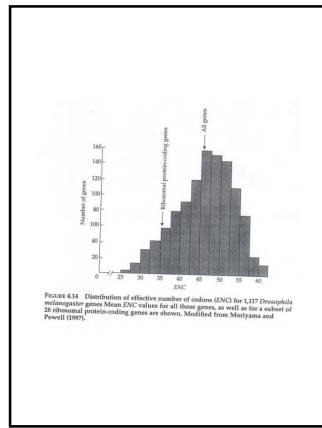


FIGURE 4.14 Distribution of effective number of codons (ENC) for 1,137 *Drosophila melanogaster* genes. Mean ENC values for all these genes, as well as for a subset of 28 ribosomal protein-coding genes are shown. Modified from Moriyama and Powell (1997).