Volume 8 Number 1 1980 Nucleic Acids Research

Codon catalog usage and the genome hypothesis

R.Grantham, C.Gautier, M.Gouy, R.Mercier and A.Pavé

Equipe Evolution Moléculaire, Laboratoire de Biométrie, Université Lyon I, 69622 Villeurbanne
Cedex, France

ABSTRACT

Frequencies for each of the 61 amino acid codons have been determined in every published mRNA sequence of 50 or more codons. The frequencies are shown for each kind of genome and for each individual gene. A surprising consistency of choices exists among genes of the same or similar genomes. Thus each genome, or kind of genome, appears to possess a "system" for choosing between codons. Frameshift genes, however, have widely different choice strategies from normal genes. Our work indicates that the main factors distinguishing between mRNA sequences relate to choices among degenerate bases. These systematic third base choices can therefore be used to establish a new kind of genetic distance, which reflects differences in coding strategy. The choice patterns we find seem compatible with the idea that the genome and not the individual gene is the unit of selection. Each gene in a genome tends to conform to its species' usage of the codon catalog; this is our genome hypothesis.

INTRODUCTION

The genetic code provides options among codons for all 20 amino acids of protein except methionine and tryptophan, which have single codons (see Figure 1 of reference 1). Choices among synonymous codons do not affect the nature of the protein produced but they may relate to expression of a gene. Indeed, mRNA expressivity (rate of synthesis of a protein and the total amount of it made) may be under the control of degenerate base use in the mRNA. This is still largely an untested hypothesis. It is, however, an important one since choices among third bases have often been considered "neutral", that is, of no influence on fitness of the phenotype. Messenger RNA expressivity may well turn out to be an adaptive phenomenon.

In order to better understand codon usage, we have studied all published mRNA sequences of more than about 50 codons. We here report three analyses on these sequences. Table 1 first shows number, name, symbol, length in codons and reference for each mRNA. Then Figures 1, 2 and 3 reveal the analyses. In Figure 1 mRNA are combined according to genome type (RNA or DNA virus, animal, etc.). Frequencies per thousand for each of the 61 amino acid codons appear in Figures 1 and 2 (initiator and terminator codons are excluded). Absolute frequencies are had by multiplying by number of codons (Table 1). Figure 2 presents frequencies in each individual gene (unless it is the only example of a genome type, in which case it is in Figure 1). This allows comparison of codon use not only among genes in the same or different genomes, but also between individual genes and the genome types of Figure 1. Finally, messengers with similar patterns of use of the code's degeneracy are seen as neighboring points in Figure 3. The spacings in Figure 3 are projections into two dimensions of distances calculated by correspondence analysis (2, 3) on codon third base frequencies. Other work has shown that the codon frequencies themselves give correspondence analysis distances that depend mostly on systematic choices between degenerate bases (4, 5).

DISCUSSION

It is not evident that distances between mRNA based on codon frequencies would depend mainly on degenerate base use. The other two codon positions could dominate in establishing these distances if proteins were relatively more different among themselves than are their mRNA. Elsewhere (4, 5) we have found that distances between proteins, determined by correspondence analysis on their amino acid frequencies, do not agree with distances between their mRNA, calculated from frequencies for the 61 codons in each messenger. However, frequencies of duet coded amino acids can particularly affect distances between mRNA (5). In order to eliminate all possible influence of protein composition, therefore, distances in Figure 3 were determined solely from quartet codons (1). Each of the eight quartet sets of codons has a complete choice of bases in codon position III. Figure 3 results from correspondence analysis on these quartet third base frequencies in each mRNA.

The structure in Figure 3 can be considered at several levels. First, animals lie to the right and viruses to the left. Next, there is grouping by genome or genome type: All papova virus (SV40 and BKV) genes lie together (no. 51-56); immunoglobins (IG) fall in a nearby zone (no. 73-78). All mRNA of mammals except immunoglobins (MAM-IG) are grouped to the upper right (no. 79-90). Curiously, however, frameshift genes E of ϕx174 and G4 (no.13 and 23) are near this mammal area. Other genes of these phage are mainly to the upper left of the figure. Finally, most IG have different coding strategies from other mammalian genes (compare, for example, MUSBLP and MUSK2, no. 73 and 80).

What do these groupings and distances mean? We can partly answer this question although the exact biological significance is unknown. The horizontal axis of Figure 3 roughly corresponds to GC content of the third bases (5). The rightmost mRNA often use C and G as third bases. Thus, with MUSBLP 90.5% of all quartet codons have C or G in position III, while the value is only 2.8% in the solitary yeast mitochondrial mRNA (no. 64) and 37.0% in MUSK2. This also explains positioning of the above E genes near mammals: their mRNA show more GC in quartet position III (68.9 and 64.7%, respectively) than do other phage mRNA. Likewise, vertical contrast exists between use of A and U. For example, quartet third bases in MUSK2 contain 42.6% A and 20.4% U while those in FDV5 (no. 33) have 4.1% A and 53.1% U. Consequently, in base choices for quartet codons papova virus mRNA resemble IG mRNA but differ greatly from other mammalian genes, for both GC content and use of A (1). Lastly, correspondence analysis groupings as in Figure 3 are very stable, whether the starting data are frequencies for all 61 codons, the 32 quartet codons, or for third bases of all codons or, as here, third bases of quartet codons. Some rationalization of this stable structure is found in the distribution of the isoacceptor tRNA for the codons (5-7). Coordination of codon usage with tRNA gene amplification and expression may eventually aid in understanding the genome hypothesis.

Our general goal in this work is to identify and understand the biological information in nucleic acid sequences. Our hypothesis, however, is so far mostly only descriptive. We observe that mRNA of the same genome are clustered by correspondence analysis on codon frequencies and that their proteins are not similarly grouped by analysis on the amino acid frequencies (4). Indeed, correspondence analysis on the proteins does not suggest classical systematics as does that on mRNA. In this very limited sense, messengers better reflect evolution than proteins do (4, 5). We do not yet know why.

This work was supported by CNRS LA 243 and GRECO Analyses des systèmes.

Table 1 mRNA Portfolio

No.	Species and gene	Symbol	No.codons	Ref.
1	Phage MS2 gene A	MS2A	392	8
2	Phage MS2 coat	MS2C	129	9
3	Phage MS2 replicase	MS2R	544	10
4	Phage Q beta coat	QBVC	79	11
5	Tobacco mosaic virus gene A	TMVA	55	12
6	Tobacco mosaic virus coat	TMVC	158	13
7	Tobacco mosaic virus 30K protein	TMV30K	105	13
8	Turnip yellow mosaic virus coat	TYMC	188	14
9	Phage ΦX 174 gene A	FIXA	511	15,16
10	Phage ΦX 174 gene B	FIXB	119	15,16
11	Phage ΦX 174 gene C	FIXC	85	15,16
12	Phage ΦX 174 gene D	FIXD	151	15,16
13	Phage ΦX 174 gene E	FIXE	90	15,16
14	Phage ΦX 174 gene F	FIXF	422	15,16
15	Phage ΦX 174 gene G	FIXG	174	15,16
16	Phage ΦX 174 gene H	FIXH	326	15,16
17	Phage ΦX 174 gene J	FIXJ	37	15,16
18	Phage ΦX 174 gene K	FIXK	55	17
19	Phage G4 gene A	VG4A	553	18
20	Phage G4 gene B	VG4B	119	18
21	Phage G4 gene C	VG4C	83	18
22	Phage G4 gene D	VG4D	151	18
23	Phage G4 gene E	VG4E	95	18
24	Phage G4 gene F	VG4F	426	18
25	Phage G4 gene G	VG4G	176	18
26	Phage G4 gene H	VG4H	336	18
27	Phage G4 gene J	VG4J	24	18
28	Phage G4 gene K	VG4K	55	17
29	Phage fd gene 1	FDV1	347	19
30	Phage fd gene 2	FDV2	409	19
31	Phage fd gene 3	FDV3	423	19
32	Phage fd gene 4	FDV4	425	19
33	Phage fd gene 5	FDV5	86	19
34	Phage fd gene 6	FDV6	111	19
35	Phage fd gene 7	FDV7	32	19
36	Phage fd gene 8	FDV8	72	19
37	Phage fd gene 10	FDV10	110	19
38	Phage M13 gene 1	M131	215	20
39	Phage M13 gene 3	M133	423	20
40	Phage M13 gene 4	M134	59	20
41	Phage M13 gene 6	M136	111	20
42	Phage M13 gene 7	M137	32	21
43	Phage M13 gene 9	M139	31	21
44	Phage T7 gene 1	T7VG1	56	22
45	Phage λ gene CI	LAMCI	236	23
46	Phage λ gene CII	LAMCII	96	24
47	Phage λ gene cro	LAMCRO	65	24
48	Phage λ gene O	LAMO	298	25
49	Phage 434 gene CII	434CII	96	26
50	Phage 434 gene cro	434CRO	70	26
51	Simian virus 40 gene T	S40GT	626	27,28

No.	Species and gene	Symbol	No.codons	Ref.
52	Simian virus 40 gene t	S40PT	173	27,28
53	Simian virus 40 gene VP1	S40VP1	361	27,28
54	Simian virus 40 gene VP2	S40VP2	351	27,28
55	Simian virus 40 gene VP3	S40VP3	233	27,28
56	Virus BK gene t	BKVPT	171	29
57	Hepatitis B virus surface antigen	HBVSA	225	30,31
58	Escherichia coli lac 1	ECOLAC	359	32
59	Escherichia coli ribosomal protein L11	ECOL11	141	33
60	Escherichia coli ribosomal protein L1	ECOL1	233	33
61	Escherichia coli ribosomal protein L10	ECOL10	164	33
62	Escherichia coli ribosomal protein L7/L12	ECO712	120	33
63	Salmonella paratyphi ampr gene	SAPAMP	285	34
64	Saccharomyces subunit 9 mitochondrial ATPase	SACMT9	75	35
65	Saccharomyces iso-1 cytochrome C	SACCC1	108	36
66	Psammechinus miliaris H1 histone	PSMH1	85	37
67	Psammechinus miliaris H2A histone	PSMH2A	123	37
68	Psammechinus miliaris H2B histone	PSMH2B	102	37
69	Psammechinus miliaris H3 histone	PSMH3	135	37
70	Strongylocentrotus purpuratus H2A histone	SPUH2A	123	38
71	Strongylocentrotus purpuratus H3 histone	SPUH3	102	38
72	Chicken ovalbumin	GALOVA	385	39
73	Mouse K2 immunoglobulin	MUSK2	117	40
74	Mouse immunoglobulin light chain MOPC 21	MUSLC	107	41
75	Mouse immunoglobulin λ1-type light chain	MUSL1	173	42
76	Mouse VλII immunoglobulin	MUSVL2	165	43
77	Mouse MOPC-41 light chain immunoglobulin	MUSVL41	129	44
78	Mouse immunoglobulin γ-1 constant heavy chain	MUSPH21	153	45
79	Mouse hemoglobin beta	MUSBGL	146	46
80	Mouse beta lipotropin	MUSBLP	48	47
81	Rat growth hormone	RATGH	215	48
82	Rat prolactin hormone	RATPLH	132	49
83	Rat preproinsulin	RATPPI	107	50
84	Rabbit hemoglobin alpha	RABAGL	141	51
85	Rabbit hemoglobin beta	RABBGL	146	52
86	Bovine corticotropin beta lipotropin	BOVCBL	264	53
87	Human alpha chorionic gonadotropin	HUMACG	115	54
88	Human hemoglobin beta	HUMBGL	132	55
89	Human pregrowth hormone	HUMPGH	216	56
90	Human chorionic sommatomammotropin	HUMCSL	168	57

Codon		SS RNA VIRUS			SS DNA	DS DNA VIRUS			
		ALL	PHAGE	PLANT	VIRUS	ALL	PHAGE	PAPOVA	ABVSA
		no 1-8	1-4	5-8	9-43	44-57	45-50	51-56	57
Arg	CGA	8	6	11	5	8	16	0	0
	CGC	10	15	6	15	14	27	0	4
	CGG	5	6	5	3	1	0	1	4
	CGU	20	25	14	21	9	15	2	9
	AGA	12	4	20	5	15	6	28	4
	AGG	6	3	9	2	15	9	24	0
Leu	CUA	11	11	10	4	10	5	12	31
	CUC	17	20	15	15	9	12	3	31
	CUG	8	11	6	18	17	19	13	31
	CUU	13	14	12	30	23	25	22	18
	UUA	16	17	16	24	20	13	28	18
	UUG	12	6	18	17	16	12	21	18
Ser	UCA	19	15	24	18	11	10	9	27
	UCC	12	13	10	13	7	4	7	22
	UCG	19	21	17	8	4	5	0	13
	UCU	20	22	19	36	16	14	18	18
	AGC	13	14	13	4	10	13	7	4
	AGU	11	4	17	8	14	8	20	22
Thr	ACA	17	7	27	11	16	17	14	27
	ACC	26	23	30	12	20	22	18	22
	ACG	9	9	9	8	4	6	0	13
	ACU	32	32	31	25	18	13	25	13
Pro	CCA	10	8	11	7	13	10	13	40
	CCC	16	10	22	5	8	3	10	31
	CCG	11	14	8	10	8	14	0	9
	CCU	14	18	10	17	17	5	29	22
Ala	GCA	23	29	18	10	22	31	14	4
	GCC	19	18	19	12	11	15	8	4
	GCG	17	20	14	15	11	22	0	4
	GCU	22	24	21	35	40	42	41	13
Gly	GGA	17	12	21	8	18	13	23	27
	GGC	11	16	5	22	11	12	11	9
	GGG	10	16	5	5	11	10	12	18
	GGU	22	34	10	37	14	16	12	9
Val	GUA	16	19	12	13	11	7	15	9
	GUC	27	23	31	14	4	6	2	4
	GUG	16	14	17	8	12	9	15	18
	GUU	30	36	24	37	24	26	23	18
Lys	AAA	28	27	29	44	45	48	47	13
	AAG	28	28	28	19	30	39	24	0
Asn	AAC	34	40	29	18	24	35	13	9
	AAU	28	24	32	27	18	11	26	18
Gln	CAA	17	17	16	26	26	24	29	13
	CAG	21	29	13	21	21	22	20	18
His	CAC	3	5	1	4	4	0	7	4
	CAU	3	3	3	9	9	8	12	0
Glu	GAA	18	11	25	23	36	38	39	0
	GAG	24	26	22	18	32	41	26	9
Asp	GAC	25	15	34	23	24	26	24	4
	GAU	25	19	30	25	28	22	38	9
Tyr	UAC	21	26	16	10	12	10	15	13
	UAU	6	5	8	26	15	11	21	13
Cys	UGC	9	8	10	5	11	3	18	22
	UGU	3	4	2	8	9	3	10	40
Phe	UUC	18	20	16	24	16	22	5	40
	UUU	18	11	25	26	25	16	35	31
Ile	AUA	15	10	19	8	9	5	12	18
	AUC	25	25	24	15	17	29	0	31
	AUU	11	12	11	33	26	26	26	22
Met	AUG	13	12	14	16	29	32	27	22
Trp	UGG	12	16	7	13	23	17	24	58

Fig. 1(i)

Codon					DS DNA				
		BACT	MITOCH	YEAST	ALL ANI	ANI-MAM	ALL MAM	IG	MAM-IG
		58-63	64	65	65-90	65-72	73-90	73-78	79-90
Arg	CGA	3	0	0	4	8	2	7	0
	CGC	20	0	0	14	26	9	4	12
	CGG	2	0	0	6	5	6	3	8
	CGU	20	0	0	10	21	7	1	9
	AGA	2	13	28	8	14	6	12	3
	AGG	0	0	0	11	12	11	9	11
Leu	CUA	3	13	9	9	9	8	11	7
	CUC	3	0	0	25	27	24	26	24
	CUG	59	0	0	42	21	51	16	68
	CUU	7	0	9	10	16	7	7	7
	UUA	6	147	9	3	1	4	11	1
	UUG	6	0	46	8	10	7	10	6
Ser	UCA	3	67	9	10	9	11	24	5
	UCC	14	0	0	17	11	20	17	22
	UCG	4	0	9	3	2	4	1	5
	UCU	18	0	19	17	17	17	29	11
	AGC	9	0	0	20	19	21	27	18
	AGU	6	0	0	15	11	17	33	9
Thr	ACA	4	27	19	13	13	13	26	7
	ACC	21	0	28	26	28	25	30	23
	ACG	6	0	0	8	10	7	4	9
	ACU	24	0	28	18	13	19	36	11
Pro	CCA	6	13	28	12	13	12	20	7
	CCC	5	0	0	17	12	18	13	21
	CCG	26	0	0	6	2	8	5	9
	CCU	1	13	9	12	11	13	13	12
Ala	GCA	40	53	0	16	29	11	20	6
	GCC	18	13	37	38	49	34	22	40
	GCG	28	0	0	6	6	5	1	8
	GCU	70	67	28	25	31	23	21	24
Gly	GGA	4	27	0	14	23	10	20	5
	GGC	31	0	19	26	21	28	16	34
	GGG	5	0	19	11	13	11	9	11
	GGU	31	107	74	19	19	18	25	15
Val	GUA	34	67	0	4	8	3	5	1
	GUC	9	0	0	19	25	16	23	13
	GUG	14	0	19	31	22	35	17	44
	GUU	42	13	9	9	12	7	7	7
Lys	AAA	65	27	56	19	29	15	17	14
	AAG	11	0	93	54	74	45	24	56
Asn	AAC	21	13	46	27	21	29	29	29
	AAU	7	13	19	9	6	10	13	9
Gln	CAA	14	13	19	12	17	10	11	9
	CAG	23	0	0	31	29	32	30	32
His	CAC	6	0	19	19	9	22	8	29
	CAU	4	0	19	10	9	10	11	10
Glu	GAA	57	13	46	22	23	22	21	23
	GAG	14	0	19	35	33	36	23	43
Asp	GAC	30	13	28	23	11	27	22	30
	GAU	19	0	9	16	11	18	23	16
Tyr	UAC	10	0	28	21	23	20	20	20
	UAU	4	13	19	12	3	15	17	14
Cys	UGC	4	0	9	9	1	12	6	16
	UGU	2	13	19	9	2	11	18	8
Phe	UUC	12	80	19	29	18	33	30	35
	UUU	12	13	19	14	9	16	13	18
Ile	AUA	3	0	0	4	3	4	7	3
	AUC	28	27	19	25	40	20	21	19
	AUU	18	93	19	11	10	11	18	8
Met	AUG	29	27	19	16	16	16	14	16
Trp	UGG	3	0	9	12	1	16	25	12

Fig 1 (ii)

		MS2 A	MS2 C	MS2 R	ORU C	TMU A	TMU C	TMU 30k	TYM C	FTX A	FTX B	FTX C	FTX D	FTX E	FTX F	FTX G	FTX H	FTX J	FTX K	UG4 A	UG4 B	UG4 C	UG4 D
Arg	CGA	15	0	7	0	18	6	19	0	6	17	0	0	0	0	0	0	27	55	9	17	0	0
	CGC	15	8	24	13	18	0	0	5	22	34	24	26	11	14	23	9	27	0	18	25	12	46
	CGG	8	0	15	0	0	6	10	5	4	8	0	0	44	0	0	3	0	0	5	17	12	0
	CGU	18	23	20	38	36	6	10	5	23	17	12	46	11	45	6	6	108	0	16	17	12	26
	AGA	8	0	7	0	0	32	48	0	12	8	47	0	0	0	0	0	0	0	4	8	0	0
	AGG	10	0	4	0	18	19	0	0	2	0	0	0	0	0	0	3	0	0	2	0	24	0
Leu	CUA	13	16	15	0	18	13	0	11	6	0	37	0	0	0	0	0	0	18	9	0	12	0
	CUC	23	16	28	13	0	0	0	59	10	8	0	13	22	9	17	0	0	36	25	0	48	7
	CUG	5	0	13	25	18	0	0	5	20	0	12	13	78	7	11	3	0	55	20	0	24	13
	CUU	15	16	13	13	18	0	19	11	27	17	35	7	11	40	17	33	27	73	20	8	24	33
	UUA	20	8	15	25	0	44	19	0	14	8	24	13	76	5	11	3	0	55	13	8	48	13
	UUG	15	0	9	0	36	19	10	5	22	0	24	26	44	19	23	15	27	18	11	0	0	7
Ser	UCA	20	16	11	13	76	19	19	21	10	0	35	0	44	14	17	3	0	18	18	8	24	13
	UCC	15	16	22	0	18	0	0	21	10	8	24	7	0	5	0	12	0	0	9	0	12	0
	UCG	26	16	18	25	18	0	29	21	12	0	12	0	22	9	23	6	0	18	7	8	36	0
	UCU	13	23	13	38	0	44	10	21	18	17	35	20	0	40	46	46	54	18	25	8	12	46
	AGC	8	31	17	0	18	19	10	5	6	8	24	0	11	0	0	0	0	0	4	17	0	0
	AGU	10	0	7	0	0	19	48	0	6	8	0	13	11	5	6	6	0	36	4	17	0	0
Thr	ACA	13	0	15	0	23	13	19	5	14	0	0	0	33	2	17	3	0	0	11	8	24	20
	ACC	13	31	22	25	18	6	10	85	6	8	12	20	11	21	17	9	0	0	18	17	12	20
	ACG	15	8	13	0	0	19	0	16	18	0	24	26	33	7	17	12	0	0	9	8	0	0
	ACU	28	31	7	63	18	63	10	32	18	34	12	26	11	50	40	34	27	18	27	25	24	20
Pro	CCA	10	16	6	0	0	19	0	27	2	17	12	0	0	5	11	0	0	18	7	8	0	0
	CCC	13	8	7	13	18	6	10	53	12	0	0	7	0	0	11	0	0	0	14	8	0	13
	CCG	8	8	17	25	0	6	10	16	4	8	0	7	33	26	11	0	27	0	2	8	0	0
	CCU	13	16	18	25	0	19	10	11	18	8	0	20	22	36	23	12	27	18	18	34	0	20
Ala	GCA	18	47	15	38	36	19	10	5	10	34	0	13	0	5	6	9	0	0	14	17	24	7
	GCC	31	16	13	13	0	25	10	43	8	8	0	33	0	9	23	34	0	0	16	0	12	40
	GCG	26	8	22	25	18	32	0	5	10	8	24	13	33	12	0	6	0	36	5	8	0	0
	GCU	15	39	28	13	36	13	10	27	55	59	12	66	11	40	40	89	54	0	29	25	12	79
Gly	GGA	5	16	15	13	18	19	38	11	8	17	12	20	0	5	6	18	0	18	9	34	12	13
	GGC	15	23	13	13	0	0	0	21	16	17	12	13	0	28	6	37	81	0	11	25	12	0
	GGG	13	8	18	25	0	0	19	0	4	17	0	0	0	0	0	0	0	0	2	8	0	7
	GGU	38	23	35	38	0	19	10	11	27	0	24	7	0	33	46	55	135	0	29	0	24	26
Val	GUA	18	23	11	25	18	19	0	11	8	0	0	13	11	5	0	6	27	0	11	0	0	13
	GUC	18	31	18	25	36	13	29	48	0	17	12	20	33	5	23	9	0	18	9	25	0	33
	GUG	20	23	13	0	18	32	19	0	10	8	0	20	22	5	6	6	0	0	18	0	24	0
	GUU	20	31	17	76	18	25	38	16	29	25	35	60	0	38	80	37	0	18	18	25	0	46
Lys	AAA	13	39	17	38	36	6	48	27	39	50	47	20	44	17	23	49	135	55	58	17	84	26
	AAG	23	8	29	51	36	6	57	11	37	8	35	7	44	17	17	18	27	36	24	8	12	0
Asn	AAC	38	47	13	63	73	6	19	16	20	17	0	33	0	19	23	12	0	36	33	34	60	33
	AAU	5	31	20	38	18	57	48	5	31	50	12	7	33	19	34	43	0	36	31	59	24	7
Gln	CAA	23	8	13	25	0	38	0	27	12	67	24	26	0	21	0	43	54	0	24	101	12	26
	CAG	23	39	15	38	18	19	0	16	23	17	0	13	22	24	34	34	54	36	11	8	0	7
His	CAC	8	0	11	0	0	0	0	5	8	17	0	0	0	5	6	0	0	0	11	8	12	7
	CAU	5	0	7	0	0	0	0	11	10	8	35	7	0	26	11	9	0	0	16	25	48	7
Glu	GAA	13	16	17	0	0	25	48	27	14	50	0	60	11	14	0	12	0	55	31	76	36	46
	GAG	31	23	24	25	36	19	29	5	37	34	59	20	11	21	11	37	0	36	11	17	24	26
Asp	GAC	13	23	26	0	55	44	10	27	29	42	47	26	0	36	17	18	0	18	40	50	60	40
	GAU	20	8	35	13	18	6	86	11	37	42	35	26	11	33	46	49	0	18	24	8	0	13
Tyr	UAC	31	31	29	13	36	19	0	11	14	17	12	7	22	7	6	3	0	18	20	8	36	7
	UAU	10	0	9	0	0	6	19	5	25	25	12	20	0	43	29	15	27	0	25	17	0	20
Cys	UGC	8	8	4	13	18	0	0	21	8	0	0	13	22	5	11	6	0	18	7	0	0	7
	UGU	0	8	9	0	0	6	0	0	10	17	0	7	11	5	17	3	0	0	2	17	0	7
Phe	UUC	26	23	29	0	0	32	10	21	14	34	24	20	44	24	6	3	0	0	27	25	36	20
	UUU	15	8	22	0	36	19	38	5	33	17	47	40	0	26	63	21	27	0	7	34	24	40
Ile	AUA	18	0	22	0	18	38	10	11	8	0	12	0	0	5	0	0	0	18	9	0	12	0
	AUC	20	31	24	25	0	13	19	64	14	8	0	7	11	7	0	3	0	36	29	25	0	53
	AUU	3	31	13	0	0	6	29	11	23	25	24	53	11	45	40	49	0	36	18	17	0	26
Met	AUG	18	16	17	0	0	0	38	16	39	0	12	13	22	26	11	31	0	0	29	0	24	13
Trp	UGG	31	16	17	0	0	19	0	11	14	8	24	7	33	12	6	9	27	0	18	0	12	7

Fig. 2 (i)

AA		UG4 E	UG4 F	UG4 G	UG4 H	UG4 I	UG4 K	FDU 1	FDU 2	FDU 3	FDU 4	FDU 5	FDU 6	FDU 7	FDU 8	FDU 10	M13 1	M13 3	M13 4	M13 6	M13 7	M13 9	T7 [illegible]
Arg	CGA	0	0	0	0	0	18	0	10	2	0	0	0	0	0	0	0	2	0	0	0	0	0
	CGC	0	19	17	18	53	0	12	10	2	14	12	0	0	0	9	9	2	17	0	0	0	54
	CGG	0	0	0	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CGU	0	40	11	9	42	18	20	20	9	16	35	4	0	0	55	28	9	0	9	0	65	36
	AGA	0	0	0	3	0	0	9	12	2	2	0	0	31	0	0	14	2	0	0	31	0	0
	AGG	0	0	0	0	0	0	6	0	5	2	0	0	0	0	0	9	5	0	0	0	0	0
Leu	CUA	0	2	11	3	0	18	3	5	2	5	0	0	0	0	0	5	2	0	0	0	0	0
	CUC	42	14	11	21	42	18	6	5	2	9	23	36	0	28	0	14	5	0	36	0	0	0
	CUG	63	19	6	0	0	73	6	20	7	9	47	18	0	42	9	9	7	17	18	0	0	36
	CUU	21	23	11	30	0	73	37	24	14	12	35	81	63	0	27	37	14	0	81	63	32	18
	UUA	52	14	23	3	0	36	49	37	24	35	0	36	0	14	19	42	24	17	36	0	97	12
	UUG	53	0	6	0	0	55	12	24	2	26	12	45	0	0	18	14	0	34	45	0	0	18
Ser	UCA	63	7	23	9	42	18	23	20	9	24	12	18	0	14	0	0	9	85	18	0	32	0
	UCC	21	21	0	21	0	0	9	32	14	16	12	0	31	28	55	9	14	17	0	31	32	0
	UCG	42	2	6	6	0	18	0	5	7	5	12	0	0	14	0	5	5	17	0	0	0	0
	UCU	0	49	51	57	83	36	35	46	47	66	47	45	0	28	45	42	50	51	45	0	65	36
	AGC	0	0	0	3	0	0	12	2	0	9	0	9	0	14	0	9	2	0	9	0	0	0
	AGU	32	0	6	3	0	18	6	5	2	16	0	0	0	0	9	5	2	0	0	0	65	0
Thr	ACA	42	16	23	15	42	0	6	7	9	12	0	0	31	0	9	0	9	0	0	31	0	0
	ACC	0	14	40	24	0	0	12	20	5	7	23	0	0	42	27	5	5	0	0	0	32	0
	ACG	21	9	0	3	0	18	6	5	5	9	0	0	0	0	0	9	5	0	0	0	32	18
	ACU	11	33	34	24	0	36	35	22	43	35	23	18	0	14	9	42	43	51	18	0	32	18
Pro	CCA	0	16	23	6	0	36	3	10	5	9	12	9	0	0	9	0	5	0	9	0	0	0
	CCC	0	2	6	6	0	0	3	5	5	2	0	18	0	14	9	0	7	17	18	0	0	0
	CCG	63	12	11	0	0	0	12	7	5	7	35	18	0	14	9	14	7	0	18	0	0	18
	CCU	32	40	28	3	0	0	20	17	50	21	23	0	0	0	27	23	45	17	0	0	0	0
Ala	GCA	0	9	28	21	0	0	14	24	9	12	0	9	0	42	36	14	9	0	0	0	0	18
	GCC	11	5	28	24	42	0	0	10	7	9	0	0	0	56	9	0	7	0	0	0	32	16
	GCG	21	7	6	6	0	0	6	0	5	5	35	9	94	56	0	5	5	0	9	94	0	0
	GCU	11	38	57	74	0	18	20	34	47	19	12	45	31	42	18	37	50	17	54	31	0	89
Gly	GGA	0	2	6	9	42	18	12	0	0	2	0	0	0	0	0	9	2	0	0	0	0	0
	GGC	0	23	23	48	83	18	12	10	69	19	12	27	0	14	9	28	61	17	18	0	32	0
	GGG	0	0	0	3	0	0	6	5	5	2	0	9	31	0	0	14	7	0	9	31	0	0
	GGU	11	33	11	36	83	0	37	27	87	47	78	36	63	42	27	23	87	17	45	63	32	36
Val	GUA	0	16	11	12	42	0	9	17	5	21	0	18	31	14	27	9	5	85	18	31	0	0
	GUC	21	9	45	12	0	36	9	2	14	7	12	0	31	42	0	9	14	0	0	31	0	0
	GUG	42	2	6	6	0	18	0	0	2	12	23	9	0	0	0	0	2	0	9	0	32	0
	GUU	0	26	45	19	0	0	58	44	31	68	58	54	31	56	64	60	31	85	54	31	65	0
Lys	AAA	42	38	28	42	125	91	61	37	33	26	35	36	0	69	45	51	31	34	36	0	0	0
	AAG	42	2	6	36	42	0	35	12	5	14	35	27	0	42	0	23	7	17	27	0	0	36
Asn	AAC	32	28	45	36	0	36	9	7	21	16	0	9	0	0	9	14	21	0	9	0	0	71
	AAU	0	21	28	30	0	36	37	51	47	59	23	9	0	0	64	33	47	68	9	0	0	0
Gln	CAA	11	14	6	48	42	55	14	15	19	21	23	18	63	14	18	14	21	17	18	63	0	0
	CAG	0	21	17	30	0	18	26	20	17	26	47	18	63	0	9	23	14	17	18	63	0	18
His	CAC	11	16	11	3	0	0	6	7	2	0	0	0	0	0	0	0	2	0	0	0	0	0
	CAU	0	16	23	6	0	0	6	15	5	2	12	0	0	0	9	14	5	0	0	0	0	36
Glu	GAA	21	26	17	36	0	73	14	24	26	21	35	0	0	14	0	14	26	34	0	0	32	54
	GAG	0	14	6	15	0	36	6	20	35	7	12	0	31	14	27	9	35	0	0	31	0	89
Asp	GAC	11	38	17	36	0	0	17	27	28	7	23	9	31	28	27	9	28	0	18	31	0	54
	GAU	0	28	23	24	0	0	49	29	28	56	12	36	31	14	27	65	28	34	27	31	0	0
Tyr	UAC	11	19	6	3	42	18	3	5	17	5	23	0	0	0	0	5	17	0	0	0	0	36
	UAU	0	31	11	21	42	0	58	32	35	12	35	36	31	28	55	37	33	34	36	31	65	0
Cys	UGC	0	2	6	0	0	18	6	10	2	0	0	0	0	0	0	9	2	0	0	0	32	0
	UGU	21	5	0	0	0	0	12	5	17	0	12	9	31	0	18	5	14	0	9	31	0	0
Phe	UUC	42	38	17	9	0	0	12	10	24	14	23	54	63	28	9	5	24	17	45	63	65	54
	UUU	11	14	17	12	0	0	29	51	28	38	12	36	0	28	55	28	31	68	45	0	32	18
Ile	AUA	0	0	0	3	0	0	14	10	5	14	0	9	63	0	0	14	5	0	9	63	0	0
	AUC	32	19	34	21	0	0	3	7	7	7	0	9	63	42	18	0	7	0	9	63	0	54
	AUU	11	26	40	27	42	18	47	37	12	45	47	99	31	14	36	47	12	51	99	31	32	18
Met	AUG	11	38	11	27	0	0	6	15	17	12	12	0	31	28	18	5	17	34	0	31	32	36
Trp	UGG	21	21	11	3	42	0	14	12	9	2	0	9	0	14	18	23	9	17	9	0	32	0

Fig. 2 (ii)

		LAM CT	LAM CII	LAM CRO	LAM O	434 CII	434 CRO	S40 GT	S40 PT	S40 VP1	S40 VP2	S40 VP3	BKV PT	ECO LAC	ECO L11	ECO L1	ECO L10	ECO 712	SAP AMP	PSM H1	PSM H2A	PSM H2B	PSM H3
Arg	CGA	0	31	15	3	31	29	0	0	0	0	0	0	11	0	0	0	0	7	0	8	0	22
	CGC	4	21	31	13	21	43	2	0	0	0	0	0	28	14	21	30	0	25	12	33	10	44
	CGG	0	0	0	3	0	0	2	0	3	0	0	0	6	0	0	0	0	4	0	0	29	0
	CGU	4	21	0	20	21	0	0	0	0	6	9	0	6	14	26	43	8	25	12	33	10	37
	AGA	21	0	0	20	0	0	26	29	19	23	30	41	3	0	0	0	0	7	12	8	0	22
	AGG	4	21	0	3	21	14	10	17	14	34	52	18	0	0	0	0	0	0	0	16	29	7
Leu	CUA	4	10	0	7	10	0	11	17	6	11	9	18	0	0	0	0	0	18	0	8	0	15
	CUC	8	10	15	3	10	14	0	0	3	6	4	6	8	0	4	0	0	7	12	49	29	15
	CUG	4	10	15	30	10	29	13	29	11	9	4	12	70	50	69	79	58	28	24	16	0	37
	CUU	34	42	0	13	42	29	13	35	11	14	17	41	6	0	0	6	0	32	0	24	20	15
	UUA	25	0	15	7	0	29	29	40	19	23	21	35	14	0	0	6	0	18	0	0	0	0
	UUG	4	21	0	3	21	14	34	12	25	20	26	12	14	0	0	0	8	14	12	16	10	7
Ser	UCA	17	10	0	17	10	14	14	0	6	11	13	12	11	0	0	0	0	7	0	16	10	0
	UCC	17	0	0	13	0	0	6	6	6	6	9	12	17	35	9	6	8	11	12	8	20	7
	UCG	0	10	0	7	21	0	0	0	0	0	0	0	14	0	0	6	0	4	12	0	0	0
	UCU	13	10	0	13	10	14	14	17	14	31	26	6	19	14	13	18	33	11	35	24	0	7
	AGC	25	10	31	17	10	0	10	0	8	9	9	6	17	0	13	6	8	7	0	8	49	22
	AGU	21	10	15	0	10	0	27	6	17	28	39	6	14	7	0	0	0	18	0	8	20	7
Thr	ACA	17	10	46	34	10	0	16	0	30	17	13	6	3	0	4	0	0	14	12	8	10	15
	ACC	21	21	31	17	21	43	11	12	25	20	26	12	31	28	30	24	0	11	12	8	69	52
	ACG	4	0	0	3	0	14	0	0	0	0	0	0	11	0	0	0	0	25	35	8	0	7
	ACU	0	10	15	7	10	29	21	17	33	31	26	23	8	35	26	30	25	21	35	8	10	0
Pro	CCA	34	10	0	17	10	0	14	6	30	11	9	6	6	7	4	0	0	21	12	16	10	7
	CCC	0	0	15	3	0	0	10	0	14	9	13	18	17	0	0	0	0	11	0	24	0	22
	CCG	4	10	15	10	10	29	0	0	0	0	0	0	17	57	26	30	17	11	12	0	0	0
	CCU	25	0	0	10	0	0	19	29	33	26	34	35	0	0	0	0	0	7	12	0	10	15
Ala	GCA	38	31	46	30	42	14	8	23	11	23	17	0	14	35	43	49	67	32	35	24	10	30
	GCC	17	0	15	27	0	29	8	6	3	14	13	6	39	21	4	18	0	28	47	49	59	67
	GCG	8	52	31	0	31	29	0	0	0	0	0	0	61	28	17	49	0	14	35	0	0	0
	GCU	17	52	31	13	52	43	40	17	36	91	39	23	8	57	77	85	167	28	24	41	0	37
Gly	GGA	4	10	15	13	21	29	14	40	33	20	13	18	3	7	0	0	0	14	12	33	10	30
	GGC	21	10	31	23	0	0	10	6	6	14	13	18	31	14	47	30	50	14	24	33	20	0
	GGG	21	10	15	20	0	0	13	6	28	14	9	6	11	0	0	0	0	18	35	16	10	0
	GGU	25	0	0	10	10	29	10	12	14	20	17	0	17	64	39	24	17	28	0	41	29	22
Val	GUA	4	0	31	0	0	14	13	12	11	20	26	12	8	43	39	37	58	18	24	0	0	22
	GUC	4	10	0	7	21	0	2	0	3	0	0	6	25	7	0	12	0	7	24	33	39	15
	GUG	8	10	0	30	10	0	19	6	28	23	9	6	39	14	17	0	8	7	12	33	29	0
	GUU	47	52	15	0	42	29	18	17	42	26	17	18	22	28	69	43	67	21	12	8	10	7
Lys	AAA	51	42	62	67	42	71	50	69	53	26	34	53	28	71	86	73	108	21	35	41	29	15
	AAG	21	42	62	27	42	43	32	40	17	6	9	41	3	35	13	0	0	18	118	57	88	81
Asn	AAC	8	31	46	40	31	14	19	12	19	9	13	6	14	28	52	6	8	18	12	49	29	0
	AAU	25	10	0	30	10	0	30	23	30	23	34	18	19	0	0	12	0	11	0	0	0	7
Gln	CAA	8	21	46	17	21	57	21	17	25	48	52	12	36	7	4	12	8	18	24	24	0	15
	CAG	38	21	0	27	21	29	22	6	28	17	21	23	42	43	34	6	0	14	35	8	29	44
His	CAC	0	0	0	3	0	0	8	0	6	9	9	12	11	0	9	6	0	11	12	8	20	7
	CAU	4	0	15	0	0	0	19	17	11	6	9	12	8	0	0	0	0	14	12	8	10	7
Glu	GAA	38	31	31	30	42	43	51	40	28	34	34	47	28	35	39	67	133	39	24	24	10	15
	GAG	55	52	15	20	42	14	27	23	36	17	17	35	14	7	9	24	0	32	24	16	49	44
Asp	GAC	17	42	0	40	31	0	19	17	30	20	21	35	19	35	47	18	42	18	12	16	10	7
	GAU	21	10	46	40	21	14	46	64	25	23	30	41	28	7	13	18	8	39	24	0	10	15
Tyr	UAC	13	0	0	7	0	14	14	29	14	9	13	12	8	7	13	18	0	11	12	16	49	22
	UAU	17	0	46	13	0	0	21	12	17	31	34	12	14	7	0	0	0	4	0	8	0	0
Cys	UGC	8	0	0	13	0	0	10	46	11	0	0	41	3	7	0	6	0	7	0	0	0	7
	UGU	4	0	0	3	0	14	13	17	8	0	0	23	6	0	0	0	0	4	0	0	0	0
Phe	UUC	17	21	15	10	21	14	2	12	3	0	0	12	3	28	4	18	17	4	24	8	10	30
	UUU	34	0	31	17	0	14	53	29	28	34	34	29	8	0	13	18	0	32	0	8	10	0
Ile	AUA	4	0	31	3	0	0	16	12	8	17	9	12	3	0	0	0	0	18	12	0	0	0
	AUC	25	42	15	23	42	0	0	0	0	0	0	0	19	43	39	12	33	25	59	24	59	44
	AUU	13	21	31	20	21	57	30	12	33	34	34	12	28	21	4	18	17	18	0	16	10	7
Met	AUG	34	42	15	27	42	29	27	52	22	9	9	41	25	35	26	30	25	32	12	8	20	15
Trp	UGG	13	31	0	30	31	14	16	35	8	20	26	41	6	0	0	0	0	14	0	0	0	0

Fig. 2 (iii)

		SPU H2A	SPU H3	GAL OVA	MUS K2	MUS LC	MUS L1	MUS VL2	MUS VL4	MUS PH2	MUS BGL	MUS BLP	RAT GH	RAT PLH	RAT PPI	RAB AGL	RAB BGL	ROU CBL	HUM ACG	HUM BGL	HUM PGH	HUM CSL
Arg	CGA	8	20	0	9	19	6	6	0	0	0	0	5	0	0	0	0	0	0	0	0	0
	CGC	33	49	5	0	0	6	6	8	7	0	21	28	0	9	0	0	38	0	0	19	24
	CGG	8	0	0	0	9	0	0	8	0	7	21	5	8	9	7	0	19	0	0	9	12
	CGU	24	29	0	0	0	0	0	8	0	0	21	9	15	37	7	0	15	0	0	5	0
	AGA	8	20	26	17	0	6	24	16	7	0	0	5	8	0	0	0	4	17	0	0	0
	AGG	16	10	8	17	9	0	0	23	7	14	0	5	0	0	7	21	8	17	23	23	18
Leu	CUA	8	29	3	17	0	17	24	8	0	7	0	0	23	0	0	0	0	17	0	19	18
	CUC	57	10	16	9	9	46	48	23	20	27	21	37	8	28	14	14	15	9	23	46	42
	CUG	16	29	26	9	9	23	18	23	13	62	21	88	30	103	99	110	49	17	98	74	65
	CUU	16	20	21	0	0	6	12	16	7	7	0	5	30	19	0	0	8	9	0	9	0
	UUA	0	0	8	0	9	6	18	23	7	0	0	5	0	0	0	0	0	0	0	5	0
	UUG	16	0	10	17	19	0	0	23	0	14	0	9	15	9	7	0	4	17	0	0	0
Ser	UCA	16	0	18	34	28	35	30	8	7	0	0	9	8	0	0	0	0	17	0	14	6
	UCC	8	10	13	9	19	17	18	23	13	14	21	14	15	19	28	21	15	35	15	32	30
	UCG	0	0	3	0	0	6	0	0	0	0	0	9	0	0	0	0	11	9	0	5	24
	UCU	16	10	26	34	9	23	24	62	20	21	0	5	15	0	21	21	4	17	0	14	12
	AGC	8	20	23	26	65	12	12	23	26	21	42	33	8	9	21	0	34	0	0	23	24
	AGU	8	20	16	51	28	12	30	54	20	7	0	5	8	0	7	27	11	9	15	9	6
Thr	ACA	0	29	18	34	19	40	42	8	13	0	0	0	8	9	0	0	0	17	8	23	18
	ACC	16	29	10	17	37	29	30	31	33	21	21	37	8	19	71	14	11	9	23	19	30
	ACG	16	0	0	0	9	6	0	8	0	0	42	0	0	0	0	0	19	26	0	9	12
	ACU	8	20	10	26	37	52	55	16	33	21	0	9	23	0	14	14	4	26	8	5	6
Pro	CCA	16	10	18	17	28	29	12	16	20	0	0	5	0	9	0	7	0	35	15	9	6
	CCC	16	20	3	17	19	6	0	8	26	7	63	28	8	28	35	0	38	9	0	28	12
	CCG	0	0	0	17	0	0	0	8	7	0	21	0	15	19	7	0	30	9	0	5	6
	CCU	8	20	16	17	0	17	12	8	26	14	0	9	38	19	7	21	0	9	30	0	0
Ala	GCA	24	39	42	43	9	17	24	16	13	7	0	9	0	9	0	7	0	17	8	5	6
	GCC	41	59	23	26	19	35	24	16	13	62	21	37	45	37	71	41	64	9	53	28	12
	GCG	8	0	0	0	0	0	0	0	7	0	21	9	0	0	14	7	30	9	0	0	0
	GCU	49	39	26	0	19	35	42	16	13	55	0	47	23	19	7	48	11	26	38	14	0
Gly	GGA	33	29	16	34	9	23	24	23	7	21	0	5	23	9	0	0	4	0	0	0	0
	GGC	33	20	21	9	19	6	12	23	26	48	42	14	0	19	57	41	61	9	61	37	24
	GGG	16	10	5	17	0	12	12	8	7	0	21	14	8	19	0	7	19	9	8	14	18
	GGU	33	0	8	17	9	52	42	31	0	21	21	9	8	28	7	27	11	17	30	0	0
Val	GUA	0	0	13	9	9	6	0	8	0	0	0	0	0	0	0	0	0	17	0	0	0
	GUC	24	20	18	17	37	23	18	16	26	14	0	9	23	19	0	14	8	35	15	14	6
	GUG	33	20	26	26	9	35	6	0	26	68	42	9	15	47	71	82	30	26	91	19	24
	GUU	16	10	23	0	0	17	24	0	0	14	0	0	8	9	0	27	4	9	15	0	0
Lys	AAA	24	29	29	9	19	12	6	23	33	7	0	14	30	9	14	21	4	35	23	5	6
	AAG	73	78	23	9	47	23	6	8	52	68	146	37	30	28	71	62	68	26	53	37	48
Asn	AAC	41	0	18	26	47	35	18	8	39	21	63	14	15	19	21	27	23	35	38	42	36
	AAU	8	0	26	17	28	0	6	0	26	21	0	9	23	0	7	27	4	0	8	0	6
Gln	CAA	24	10	21	26	9	6	6	16	7	0	0	9	30	28	0	0	4	9	0	19	12
	CAG	8	59	18	26	19	23	18	47	46	21	21	60	23	47	7	27	34	35	23	51	42
His	CAC	8	0	8	9	9	0	6	0	26	48	42	14	15	19	71	34	11	26	45	14	12
	CAU	8	10	10	17	9	12	12	0	13	14	0	0	15	0	7	27	4	17	15	5	12
Glu	GAA	33	10	47	26	19	23	12	31	13	7	0	19	68	19	21	27	27	17	15	23	30
	GAG	8	49	39	0	37	23	6	8	65	27	42	51	38	65	28	41	83	17	30	42	48
Asp	GAC	8	10	16	9	37	12	18	31	26	21	21	47	30	9	50	21	30	0	15	37	77
	GAU	8	0	21	17	28	23	24	31	13	41	0	5	30	19	0	7	8	26	38	9	6
Tyr	UAC	16	29	18	26	19	12	24	16	26	21	42	14	15	37	7	14	15	17	8	23	24
	UAU	8	0	8	34	19	12	12	23	0	0	0	19	23	0	14	7	8	43	15	14	24
Cys	UGC	0	0	3	0	9	0	6	0	20	7	0	14	8	37	7	0	15	70	0	14	18
	UGU	0	0	13	26	19	17	24	23	0	7	0	9	0	19	0	7	4	17	15	9	6
Phe	UUC	0	29	23	9	37	29	30	23	52	21	63	33	0	28	57	34	30	35	23	46	48
	UUU	16	0	29	34	0	6	12	23	0	27	0	28	38	9	0	21	4	17	38	19	12
Ile	AUA	0	0	8	0	0	17	18	0	7	7	0	0	23	0	0	0	0	9	0	0	0
	AUC	41	10	42	26	19	12	12	31	26	14	42	23	45	9	21	0	11	9	0	28	30
	AUU	0	20	16	26	19	17	12	23	13	7	0	14	38	9	0	7	0	0	0	9	12
Met	AUG	8	10	42	17	9	0	18	16	26	14	21	28	15	9	7	7	19	26	8	14	30
Trp	UGG	0	0	8	26	19	29	36	16	26	14	21	14	8	19	7	14	15	0	15	9	6

Fig. 2 (iv)

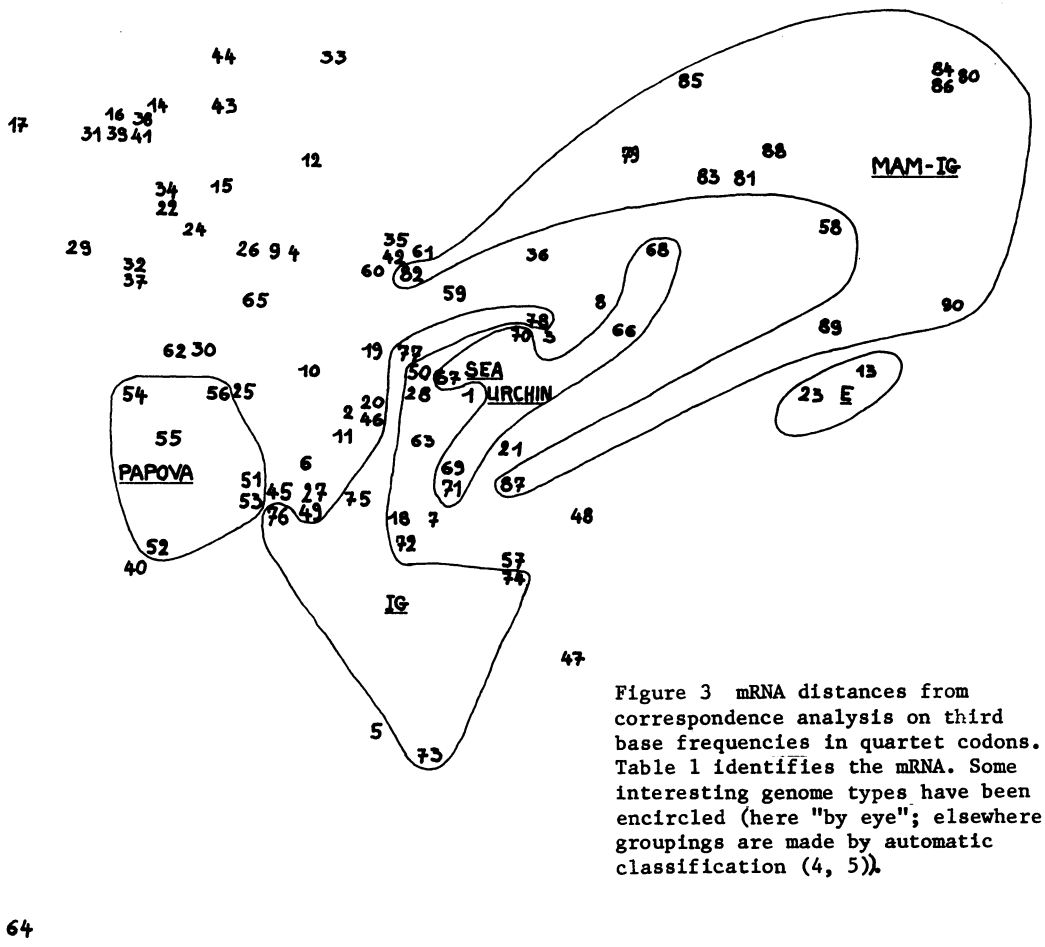

Figure 3 mRNA distances from correspondence analysis on third base frequencies in quartet codons. Table 1 identifies the mRNA. Some interesting genome types have been encircled (here "by eye"; elsewhere groupings are made by automatic classification (4, 5)).

REFERENCES

1. Grantham, R. (1978) FEBS lett. 95, 1-11.
2. Benzécri, J.P. (1973) In L'Analyse des données 2. l'analyse des correspondances, Dunod, Paris.
3. Hill, M.O. (1974) Appl. Statist. 23, 340-354.
4. Grantham, R. and Gautier, C. (1980) Naturwissenschaften, in press.
5. Grantham, R., Gautier, C. and Gouy, M. in preparation.
6. Garel, J.P. (1974) J. Theor. Biol. 43, 211-225.
7. Osterman, L.A. (1979) Biochimie 61, 323-342.
8. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F. and Van Montagu, M. (1975) Nature 256, 273-278.
9. Min Jou, W., Haegeman, W., Ysebaert, M. and Fiers, W. (1972) 247, 82-88.
10. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976) Nature 260, 500-507.
11. Escarmis, C., Sastry, P.A. and Billeter, M.A. (1978) J. Biol. Chem. 253, 8390-8396.
12. Jonard, G., Richards, K., Mohier, E. and Gerlinger, P. (1978) Eur. J. Biochem. 4, 521-529.
13. Guilley, H., Jonard, G., Kukla, B. and Richards, K.E. (1979) Nucleic Acids Res. 6, 1287-1308.
14. Guilley, H. and Briand, J. (1978) Cell 15, 113-122.
15. Sanger, F., Air, G., Barrell, B., Brown, N., Coulson, A., Fiddes, J., Hutchinson III, C., Slocombe, P. and Smith, M. (1977) Nature 265, 687-695.
16. Fiddes, J.C. (1977) Sci. American 237, 55-67.
17. Shaw, D.C., Walker, J.E., Northrop, F.D., Barrell, B.G., Godson, G.N. and Fiddes, J.C. (1978) Nature 272, 510-515.
18. Godson, G., Barrell, B., Staden, R. and Fiddes, J. (1978) Nature 276, 236-247.
19. Sugimoto, K., Sugusaki, H., Okamoto, T. and Takanami, M. (1978) Nucleic Acids Res. 5, 4495-4510.
20. Van Wezenbeek, P. and Schoenmakers, J.G.G. (1979) Nucleic Acids Res. 6, 2799-2818.
21. Huselbos, T. and Schoenmakers, J.G.G. (1978) Nucleic Acids Res. 5, 4677-4698.
22. Mc Connell, D.J. (1979) Nucleic Acids Res. 6, 3491-3503.
23. Sauer, R. (1978) Nature 276, 301-302.
24. Schwarz, E., Scherer, G., Hobom, G. and Kössel, H. (1978) Nature 272, 410-414.
25. Scherer, G. (1978) Nucleic Acids Res. 5, 3141-3156.
26. Grosschedl, R. and Schwarz, E. (1979) Nucleic Acis Res. 6, 867-881.
27. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebaert, M. (1978) Nature 273, 113-120.
28. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978) Science 200, 494-502.
29. Dhar, R., Seif, I. and Khoury, G. (1979) Proc. Nat. Acad. Sci. USA 76, 565-569.
30. Valenzuela, P., Gray, P., Quiroga, M., Zaldivar, J., Goodman, H.M. and Rutter, W.J. (1979) Nature 280, 815-819.

31. Charnay, P., Mandart, E., Hampe, A., Fitoussi, F., Tiollais, P. and Galibert , F. (1979) Nucleic Acids Res. 7, 335-346.
32. Farabaugh, P. (1978) Nature 274, 765-769.
33. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Nat. Acad. Sci. USA 76, 1697-1701.
34. Sutcliffe, J. (1979) Proc. Nat. Acad. Sci. USA 75, 3737-3741.
35. Hensgens, L.A.M., Grivell, L.A., Borst, P. and Bos, J.L. (1979) Proc. Nat. Acad. Sci. USA 76, 1663-1667.
36. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L. and Hall, B.D. (1979) Cell 16, 753-761.
37. Schaffner, W., Keenz, G., Daetwyler, H., Telford, J., Smith, H. and Birnstiel, M. (1978) Cell 14, 655-671.
38. Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033-1044.
39. Mc Reynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723-728.
40. Seidman, J., Leder, A., Nau, M., Norman, B. and Leder, P. (1978) Science 202, 11-17.
41. Hamlyn, P.H., Brownlee, G.G., Cheng, C.C., Gait, M.J. and Milstein, C. (1978) Cell 15, 1067-1075.
42. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15, 1133-1144
43. Tonegawa, S., Maxam, A.M., Tizard, , Bernard, O. and Gilbert, W. (1978) Proc. Nat. Acad. Sci. USA 75, 1485-1489.
44. Seidman, J.G., Max, E.E. and Leder, P. (1979) Nature 280, 370-375.
45. Rogers, J., Clarke, P. and Salser, W. (1979) Nucleic Acids Res. 6, 3305-3321.
46. Konkel, D.A., Tilghman, S.M. and Leder, P. (1978) Cell 15, 1125-1132.
47. Roberts, J.L., Seeburg, P.H., Shine, J., Herbert, E., Baxter, J.D. and Goodman, H.M. (1979) Proc. Nat. Acad. Sci. USA 76, 2153-2157.
48. Seeburg, P.H., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) Nature 270, 486-494.
49. Gubbins, E.J., Maurer, R.A., Hartley, J.L. and Donelson, J.E. (1979) Nucleic Acids Res. 6, 915-930.
50. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977) Science 196, 1313-1319.
51. Heindell, H.C., Liu, A., Paddock, G.V., Studnika, G.M. and Salser, W.A. (1978) Cell 15, 43-54.
52. Efstratiadis A., Kafatos, F.C. and Maniatis, T. (1977) Cell 10, 571-585.
53. Nakanishi, S., Inoue, A., Kita, T., Nakumura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S. (1979) Nature 278, 423-427.
54. Fiddes, J.C. and Goodman, H.M. (1979) Nature 281, 351-356.
55. Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977) J. Biol. Chem. 252, 5040-5050.
56. Martial, J.A., Hallewell, R.A., Baxter, J.D. and Goodman, H.M. (1979) Science 205, 602-607.
57. Shine, J., Seeburg, P., Martial, J., Baxter, J. and Goodman, H.M. '1977) Nature 270, 494-499.