

Codon usage tabulated from the GenBank genetic sequence data

Ken-nosuke Wada¹, Yoshiko Wada², Hirofumi Doi³, Fumie Ishibashi, Takashi Gojobori and Toshimichi Ikemura

National Institute of Genetics, Mishima, Shizuoka-ken 411, ¹Chukyo University, School of Computer and Cognitive Sciences, ²Niigata University, School of Medicine and ³International Institute for Advanced Study of Social Information Science, Fujitsu Limited, Japan

In 1980 and 1981, Grantham and his colleague (1,2) reported the codon usages in a total of 161 protein genes in this journal, and in 1986, 1988 and 1989, we reported those in all available genes in the days (3–5). Now the codon usages in 15137 genes can be analyzed using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 65.0, Sep., 1990). Because of the growing size of the database, it has become impossible to print the data: It is planned to distribute the electric version of the Sequence Supplement of this journal using a CD ROM, possibly beginning in 1991. This year is a transition year, and thus we will send, upon request, a magnetic tape or a hard copy listing the codon usages in 15137 genes.

To reveal the characteristics of the codon use of a wide range of organisms, as well as viruses and organella, the frequency (per one thousand) of codon use in each organism for which more than 20 genes are available was calculated by summing up numbers of codon use (Table 1). The number of genes summed for each organisms is given in the row designated as No. GENES, and the total codon number thus summed is given at the bottom row. Synonymous codon-choice patterns in different genes of a single unicellular organism are known to be usually similar with each other regardless of gene functions and thus with the pattern listed in Table 1, (dialectal codon-choice pattern found for individual unicellular organisms, see ref. 6). However, codon-choice patterns in one higher vertebrate often differ significantly between different genes (6–9): The diverse codon-choice patterns found among genes of a single higher vertebrate have been pointed out in connection with the evident diversity in the G+C% at the codon third position among the genes (6–9). It should be stressed that the characteristic pattern for the mammals listed in Table 1 is obtained only after summing up the genes with varying functions (3,4). When codon usages of approximately 10 or more genes with varying functions were summed up for each mammal, they usually resulted in a very similar pattern and thus in the pattern listed in Table 1, regardless of differences in the genes used for the summation (3,4). The fact that the pattern roughly common among the mammals does not depend on the genes used for the summation shows that this relates with general characteristics of their genomes: 1) deficiency of CpG (as well as TpA) dinucleotide and the concomitant increase of TpG and CpA (10); 2) paucity of genes in the A+T-rich genome portion, (thus C- and G-ending codons are preferred; see ref. 7,11); 3) gross similarity of tRNA population between different organs of higher vertebrates (our unpublished data); 4) gross similarity of amino acid composition between different proteins, as well as between different mammals.

METHODS

In selecting protein coding sequences we relied on the FEATURES tables of the GenBank, and only complete genes, starting with an initiation codon and ending with one of stop codons, were used in the analysis (see ref. 3 for details). In the GenBank, a group of consecutive genes whose entire region had been sequenced were registered under one LOCUS name. To distinguish the different genes belonging to a single LOCUS, symbol # followed by a number is added after the LOCUS name in our database; the numbers represent the order of the peptides registered in the FEATURES of the GenBank. When introns of a gene have not been completely sequenced, some of its exons are registered in separate entries (LOCUS) in the GenBank. These exons belonging to the same gene but having different LOCUS names were combined, and the LOCUS name of the last exon followed by symbol * was given to the gene thus combined (3,4). The order of the codons in the table is the same as the previous compilation(1–5), and the correspondence of amino acids and codons is followed from the universal codon table.

ACKNOWLEDGMENTS

The authors are very grateful to Dr. S. Osawa (Nagoya Univ.). This work was supported by grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan (No. 01656004).

REFERENCES

1. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pav'e, A. (1980) *Nucl. Acids Res.* **8**, r49–62.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nucl. Acids Res.* **9**, r43–74.
3. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) *Nucl. Acids Res.* **14**, r151–197.
4. Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T., Gojobori, T. and Ikemura, T. (1988) *Nucl. Acids Res.* **16**, r315–402.
5. Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T., and Ikemura, T. (1990) *Nucl. Acids Res.* **18**, Supplement 2367–2441.
6. Ikemura, T. (1985) *Mol. Biol. Evol.* **2**, 13–24.
7. Bernardi, G. and Bernardi, G. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 479–487.
8. Aota, S. and Ikemura, T. (1986) *Nucl. Acids Res.* **14**, 6345–6355 & 8702.
9. Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F. (1988) *Nucl. Acids Res.* **16**, 8207–8211.
10. Nussinov, R. (1984) *J. Mol. Evol.* **20**, 111–119.
11. Ikemura, T. and Aota, S. (1988) *J. Mol. Biol.* **203**, 1–13.

Table 1. Frequency (per one thousand) of codon usages summed up for individual organisms, as well as for virus, phage, plasmid or organella. The name of each species is listed at the top of the column in an abbreviated form, and the abbreviation is explained in Table 3. The number of genes summed for each species is listed at the line specified as No. GENES, and the total codon numbers thus summed is listed at the bottom line.

	HUM	HAM	MUS	RAT	BOV	DOG	PIG	RAB	SHP	CHK	FSB	XEL	BMO	CEL	DDI	DRO	PFA	SOM	SUP	SUS	TRB	ASN	ATH	BLY	MZE	NEU	PEA	PHV	POT	RIC		
No. GENES	1065	56	810	831	284	44	86	133	42	260	94	154	28	41	89	348	52	29	21	42	64	25	44	34	80	47	35	20	29	26		
ARG	6.5	6.4	5.9	6.0	6.4	4.5	6.3	5.2	4.3	6.2	4.5	6.5	2.8	3.5	7.0	1.5	9.0	11.5	6.5	11.5	7.2	8.6	3.3	0.9	2.6	2.1	3.7	2.1	5.5	3.2		
CGA	11.2	12.5	10.7	10.6	12.0	10.6	12.5	12.0	11.6	12.8	14.2	8.6	11.5	11.9	11.9	17.5	0.2	3.0	29.7	7.8	14.5	17.3	4.2	14.1	15.7	14.2	3.4	4.4	2.0	14.7	10.0	
CGC	10.5	9.9	8.0	8.1	11.3	10.4	11.8	10.0	8.5	8.4	4.9	6.2	1.7	2.0	0.1	7.2	0.1	1.6	2.8	2.2	8.7	9.0	2.4	8.9	7.1	4.9	1.0	2.1	0.5	7.7	7.7	
CGG	4.0	7.5	4.6	5.3	3.6	3.1	4.3	4.1	4.6	6.5	8.1	10.0	10.5	22.1	19.4	9.4	3.3	18.6	33.7	18.7	13.7	12.3	14.8	4.1	5.2	13.0	8.3	8.0	9.4	8.7	10.7	
CGU	9.9	9.9	11.0	10.5	9.9	9.2	9.6	9.2	8.2	10.3	15.0	14.9	10.5	16.1	18.7	4.8	16.6	12.3	7.9	7.3	5.6	5.0	15.0	2.1	4.1	3.9	17.4	8.5	7.9	10.7	10.6	
AGA	10.8	9.3	11.0	10.8	11.5	9.4	10.9	11.3	12.5	10.5	13.8	12.2	7.4	1.5	0.6	8.3	2.8	3.4	10.7	9.1	7.5	5.6	13.3	10.2	14.4	6.4	11.0	11.6	10.8	10.6	10.6	
LEU	6.3	6.0	6.7	6.7	4.8	5.1	5.4	4.4	4.1	4.8	6.5	7.5	4.2	2.6	1.5	6.3	3.4	8.5	5.5	5.7	8.3	7.8	6.6	2.5	9.8	2.4	5.4	3.4	7.8	7.9	7.9	
CUC	20.1	18.2	20.0	19.9	21.7	26.8	21.7	23.8	27.8	17.0	22.9	12.7	24.0	25.6	14.1	12.2	2.2	6.7	36.0	21.3	18.5	23.2	18.8	36.0	25.0	38.1	12.5	21.5	10.4	26.8	26.8	
CUG	42.5	35.9	40.3	41.1	47.4	46.0	48.6	51.0	48.5	41.8	48.3	26.0	20.6	6.1	0.3	37.0	0.4	7.1	13.9	15.5	19.2	22.3	5.8	21.8	27.8	15.0	6.2	12.9	5.0	18.5	18.5	
CUU	10.9	13.1	11.3	11.7	10.5	11.2	8.9	9.5	11.3	10.4	9.2	15.2	10.4	29.8	12.6	7.3	9.8	15.8	18.2	15.9	22.7	16.5	24.5	7.3	16.3	13.7	21.6	22.1	26.2	13.8	13.8	
CUA	5.5	4.1	4.5	4.8	4.1	4.0	3.8	4.0	3.3	4.2	4.6	7.8	5.8	2.4	4.7	4.2	5.1	30.2	0.0	3.6	6.4	5.5	4.8	1.0	4.9	0.5	9.7	6.1	16.1	7.8	7.8	
UUA	11.1	10.9	11.6	12.0	10.0	9.8	10.1	9.5	11.0	10.0	10.7	13.1	14.6	17.0	9.8	14.0	8.1	18.1	11.1	10.5	14.7	10.8	21.3	5.4	15.2	11.0	20.4	15.6	21.5	12.2	12.2	
SER	9.8	8.1	10.8	10.8	7.8	7.9	7.7	7.1	6.9	7.9	9.8	11.7	4.9	10.6	32.6	6.1	26.1	21.8	7.5	9.9	11.4	11.1	9.8	2.8	8.7	3.3	15.9	12.0	16.6	8.7	8.7	
UCC	17.6	17.6	18.0	17.0	17.6	18.1	17.7	18.5	21.7	15.9	22.2	14.3	14.4	14.3	14.4	14.3	5.9	19.9	6.0	18.2	16.4	12.0	20.3	11.5	20.1	14.9	23.8	13.9	22.2	8.7	12.1	12.1
UCG	4.0	3.0	4.1	3.7	4.5	4.4	4.1	4.2	5.1	5.1	5.1	3.3	10.0	9.3	0.7	18.6	1.1	5.6	1.2	4.3	8.2	12.6	4.4	8.8	9.6	12.5	2.9	2.8	3.1	8.0	8.0	
UCU	13.2	15.1	15.3	13.8	11.5	9.7	11.1	9.3	8.4	12.4	19.8	17.2	8.2	17.2	18.4	6.1	14.4	10.7	11.9	12.6	10.9	17.9	16.5	9.3	9.2	12.1	20.4	24.3	15.9	11.0	11.0	
ACC	18.6	14.5	18.5	17.1	18.3	17.0	18.4	18.4	23.4	19.9	18.1	17.5	15.5	5.9	1.9	18.0	4.2	7.4	20.2	13.2	13.0	10.2	15.6	16.0	15.3	11.0	18.7	9.3	15.8	15.8	15.8	
ASU	8.7	8.9	10.1	9.8	8.5	8.8	7.8	7.4	7.7	7.7	9.2	12.2	6.9	4.5	12.8	8.6	23.5	17.1	7.1	7.8	13.0	7.2	8.5	4.0	4.7	4.9	14.0	11.0	14.1	10.7	10.7	
THR	14.2	14.4	16.3	14.1	11.2	13.8	11.1	10.6	9.1	14.6	10.8	18.1	8.9	8.3	19.8	9.8	23.9	23.1	9.5	12.8	18.3	12.3	13.9	4.9	7.8	5.6	17.2	11.8	19.8	9.2	9.2	
ACC	22.8	24.1	22.8	20.8	20.8	21.2	24.0	28.9	20.3	25.3	17.4	21.3	19.0	15.5	24.1	6.2	7.2	33.3	27.1	13.5	23.2	21.1	25.8	20.8	30.4	12.0	20.1	11.8	9.2	18.6	18.6	
ACG	6.6	6.4	6.6	6.4	6.6	7.7	7.4	8.1	8.2	7.7	7.4	5.9	3.5	5.6	5.4	0.4	13.9	1.9	5.0	4.8	5.4	15.8	10.8	5.6	13.0	10.8	9.5	3.4	7.7	2.8	8.9	
ACU	12.6	13.7	13.7	13.1	13.1	11.1	10.1	10.1	9.1	12.5	12.1	17.8	15.7	16.0	33.1	8.5	19.8	18.2	7.1	11.2	12.2	15.6	18.7	4.7	8.5	11.5	23.6	12.9	31.5	12.1	12.1	
CCG	14.9	12.9	18.7	13.7	11.8	11.7	11.8	13.5	11.8	12.6	12.1	17.8	5.8	33.9	36.5	12.4	28.0	16.1	10.3	18.8	12.4	11.8	17.7	17.1	17.1	5.2	24.0	11.6	19.5	9.6	9.6	
CCU	6.6	2.9	6.8	6.2	7.9	5.1	7.4	6.0	7.8	6.8	7.8	3.6	5.6	6.2	0.3	16.1	0.8	3.2	1.6	4.0	9.6	12.3	6.9	17.8	18.6	10.2	5.3	2.5	2.9	13.7	13.7	
CCG	6.6	2.9	6.8	6.2	7.9	5.1	7.4	6.0	7.8	6.8	7.8	3.6	5.6	6.2	0.3	16.1	0.8	3.2	1.6	4.0	9.6	12.3	6.9	17.8	18.6	10.2	5.3	2.5	2.9	13.7	13.7	
CCU	15.8	17.4	17.4	14.2	14.4	12.6	14.2	14.4	14.2	14.4	11.8	10.5	15.3	10.9	2.6	6.0	10.8	9.2	9.9	12.3	11.4	14.6	15.7	7.5	13.8	13.9	21.8	15.3	17.0	13.4	13.4	
ALA	14.3	14.8	13.4	14.2	13.2	14.8	12.4	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2	11.2
CCG	29.5	28.6	28.4	28.9	34.9	32.5	32.8	30.8	33.0	27.8	36.5	20.8	27.2	27.5	33.3	38.1	1.3	32.4	7.7	52.3	35.6	19.1	31.1	18.5	51.3	33.4	50.5	10.3	23.3	9.8	28.0	28.0
CCG	7.1	6.3	6.4	6.9	8.1	7.1	8.6	7.7	10.3	8.6	9.5	4.0	10.4	4.1	0.2	13.4	0.7	3.4	2.4	2.4	6.1	18.5	15.4	7.6	28.0	28.1	13.0	4.6	2.8	3.9	16.7	16.7
CCU	19.4	23.7	19.7	21.8	18.4	17.4	18.4	15.7	17.0	24.0	18.1	24.8	24.4	34.5	24.6	15.5	20.7	22.2	33.7	25.6	22.3	36.3	38.3	11.4	28.9	25.5	38.4	27.4	36.9	19.7	19.7	
GLY	17.3	20.2	17.7	20.1	15.3	16.6	14.2	15.5	14.4	15.5	15.0	23.8	27.5	48.7	38.8	19.9	31.4	28.4	35.6	34.4	18.4	12.9	33.1	9.9	9.8	7.6	31.0	46.8	35.1	15.2	15.2	
GCC	25.1	26.1	23.7	23.1	27.0	25.8	32.0	27.1	28.2	23.5	23.2	18.3	35.8	3.4	2.3	30.8	3.4	2.3	30.8	30.8	20.6	19.1	24.5	10.0	54.0	28.5	37.0	9.9	18.4	14.7	33.8	33.8
GCG	17.4	13.8	15.9	15.3	18.1	15.6	18.6	18.6	14.9	14.8	8.9	12.2	6.6	1.7	1.0	4.6	1.9	4.7	6.3	5.8	11.2	10.4	7.3	19.9	13.6	4.4	7.0	15.3	5.9	15.1	15.1	
GGU	11.5	16.3	11.7	11.6	11.1	11.2	10.5	8.4	10.1	11.0	11.0	16.6	41.9	9.1	57.5	16.1	25.4	61.1	25.3	24.9	29.1	20.9	29.8	11.8	15.8	30.1	27.0	48.0	24.3	17.2	17.2	
VAL	6.0	6.9	6.5	6.4	5.1	5.4	4.5	4.8	3.9	4.1	4.8	9.4	4.2	12.8	5.0	21.1	10.8	4.8	8.9	12.0	5.0	4.1	2.8	4.8	2.5	10.3	4.9	8.7	11.5	11.5		
GUC	16.2	16.5	17.1	16.8	18.1	18.6	18.2	18.1	14.3	15.8	23.1	12.5	26.0	19.8	11.0	15.7	3.1	10.7	35.6	25.3	11.0	26.0	19.9	31.4	22.2	36.3	8.4	11.8	8.6	23.8	23.8	
GUG	30.9	32.9	30.8	30.7	32.7	32.4	32.6	33.8	30.8	28.6	26.6	21.2	19.3	8.2	1.2	26.5	2.9	9.5	14.3	12.2	26.5	12.6	18.4	34.8	30.3	9.8	18.1	21.0	14.9	19.8	19.8	
GUU	10.4	12.5	9.9	9.8	9.8	11																										

No.GENES	SOY	SPI	TOB	TOM	WHT	YSC	YSK	YSP	AFA	ANA	ATU	AVI	BME	BPE	BSP	BSS	BST	BSU	BTH	CHT	CLO	ECO	FDI	FPL	HAL	KPN	NGO	PRM	PSE	RCA
ARG	50	31	48	46	23	22	11	65	51	32	117	03	28	14	43	69	55	307	83	64	07	1187	31	27	35	81	22	25	182	24
COA	54	31	48	46	23	22	11	65	51	32	117	03	28	14	43	69	55	307	83	64	07	1187	31	27	35	81	22	25	182	24
COG	77	39	36	35	21	07	06	285	126	241	241	382	115	392	72	48	208	86	28	71	10	220	207	146	268	399	293	30	308	362
COH	20	28	15	21	50	11	10	27	113	74	149	67	04	158	42	24	110	55	20	27	09	46	88	127	130	123	84	09	102	180
COI	78	88	105	100	15	74	27	203	87	174	154	117	50	52	94	79	70	115	112	29	247	214	204	65	90	84	202	98	78	
AGA	145	119	144	163	38	238	250	109	16	79	72	03	34	18	84	148	35	117	219	159	223	20	64	88	17	15	44	37	23	08
AGG	114	139	110	85	66	78	34	47	24	13	80	16	01	37	18	52	29	47	48	45	46	13	12	63	23	16	43	04	29	10
CUA	69	74	77	80	120	119	108	58	24	144	92	08	98	15	79	106	45	52	112	110	78	30	30	27	23	31	12	98	28	00
CUJ	164	122	121	111	179	42	25	73	181	223	201	18	147	68	34	130	92	36	71	25	45	98	70	110	391	184	59	51	188	193
CUK	100	73	59	60	148	86	35	64	447	155	254	680	44	618	136	39	139	207	46	61	45	548	114	562	264	700	238	128	529	590
CUU	242	215	179	284	59	97	102	283	93	37	188	13	229	41	110	120	126	202	164	180	168	89	71	137	41	84	81	111	66	127
CUU	81	87	93	89	17	244	398	235	17	54	2	02	58	11	219	376	186	194	409	288	362	103	259	106	11	44	57	394	26	00
UUA	187	267	267	267	147	320	171	248	111	289	172	42	58	11	91	147	194	129	93	219	98	112	242	67	63	47	317	87	132	37
UCA	137	168	162	171	147	156	159	169	43	88	103	01	186	15	93	172	58	147	159	91	201	83	65	129	22	25	58	145	40	05
UCC	91	171	120	98	94	148	67	138	113	109	137	183	13	156	104	73	79	82	77	128	45	96	148	135	172	107	154	33	123	71
UCG	37	60	60	42	95	67	36	70	220	26	159	108	28	142	69	39	139	61	50	56	20	79	25	52	206	113	77	25	147	232
UCU	164	161	183	178	158	247	289	324	23	182	85	09	143	09	135	129	54	152	203	317	159	104	219	127	17	36	138	186	41	11
AGC	167	117	87	111	124	73	31	96	156	105	172	187	133	272	123	82	128	138	62	100	79	150	217	117	171	257	180	108	235	108
AGU	130	108	122	125	30	115	173	31	96	35	83	88	31	74	17	91	172	56	62	190	101	157	71	113	135	36	23	47	149	78
THR	138	178	168	188	49	154	180	110	51	183	102	03	302	26	196	306	123	234	289	229	252	64	140	113	51	19	78	175	47	08
ACC	139	225	125	108	144	143	77	129	273	244	184	395	24	324	155	78	114	76	88	90	74	246	232	150	282	317	258	179	295	318
ACG	27	39	43	31	42	67	30	46	228	26	174	41	139	219	205	111	269	133	125	64	26	125	13	138	271	159	132	83	128	145
ACU	168	181	214	199	127	221	232	245	53	138	77	28	97	50	122	234	61	97	195	265	250	105	172	100	49	28	114	203	70	10
CCA	318	185	360	225	739	211	149	101	48	89	115	09	174	19	168	198	86	69	187	95	158	81	80	56	43	34	34	125	45	08
CCG	99	102	85	91	106	57	20	105	132	100	115	157	01	167	22	39	32	26	18	42	24	42	106	75	140	113	132	44	125	211
CCU	45	62	41	15	188	42	22	22	243	24	187	290	69	235	148	48	230	141	68	29	39	242	27	204	220	317	153	38	238	341
CCA	228	214	290	290	113	152	152	225	43	338	53	14	1300	27	108	189	64	110	133	184	139	65	172	72	299	39	92	131	59	21
CCA	204	208	181	240	115	150	165	134	157	268	211	33	404	98	239	256	119	217	210	261	256	206	376	212	104	51	282	290	133	15
CCG	167	718	161	141	179	156	58	32	41	423	125	294	210	158	498	145	118	286	185	82	90	71	237	112	304	502	472	457	100	510
CCU	45	70	64	28	131	53	42	32	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165	165
CGU	231	235	343	393	88	280	222	359	109	421	202	68	330	59	218	151	160	194	186	425	259	178	592	162	89	70	153	287	69	141
GCA	231	235	343	393	88	280	222	359	109	421	202	68	330	59	218	151	160	194	186	425	259	178	592	162	89	70	153	287	69	141
GCC	116	104	154	107	253	89	91	33	88	514	173	264	596	186	735	294	103	330	224	70	95	88	307	161	156	461	450	498	206	555
GCG	106	147	82	87	334	51	23	39	61	523	127	53	40	120	98	115	175	96	116	67	45	96	42	133	168	127	47	91	99	153
CGU	227	267	300	265	98	341	284	288	141	209	147	140	341	71	178	205	119	148	158	148	252	280	395	192	88	245	439	164	89	00
GUA	79	112	120	98	43	99	177	115	45	239	75	27	363	61	200	217	93	156	270	192	262	118	199	92	30	54	147	129	68	00
GUC	84	125	124	114	144	147	65	138	207	95	209	320	60	305	142	57	274	164	73	81	39	133	505	213	227	121	261	245	348	00
GUG	281	205	165	187	136	95	53	75	261	90	170	342	109	316	160	76	218	168	108	104	63	253	97	233	268	274	162	131	291	302
GUU	257	314	290	317	113	284	237	234	86	275	181	50	282	36	191	231	132	200	201	268	253	204	209	219	68	87	214	219	86	71
AAA	232	270	198	300	41	382	663	343	54	365	150	65	563	52	32	421	481	563	563	545	563	368	232	327	88	188	570	925	109	115
AAG	342	417	321	353	165	353	213	279	329	155	215	434	117	221	158	144	164	203	82	184	167	119	152	160	177	103	118	86	264	278
AAC	301	215	240	192	123	258	180	214	252	233	174	281	240	226	362	447	173	234	551	288	540	159	200	207	119	119	199	372	81	38
AAD	200	193	283	265	68	315	509	334	100	154	159	50	187	84	368	447	173	234	551	288	540	159	200	207	119	119	199	372	81	38
AAG	286	205	245	220	178	293	189	265	65	383	163	45	367	92	217	285	233	211	375	273	194	130	378	65	43	83	252	377	98	52
CAA	217	110	158	117	835	106																								

J	NA	GENES	PHB	RHL	PHM	SMA	SSP	STA	STM	STR	STY	SYN	TIP	TRN	VIB	YEP	ADB	FLA	FLB	HIV	HPB	HS1	HS2	HS4	HSS	MCV	MEA	MHV	NDV	NPA	PIF	PLY	
	ARG	69	64	43	46	20	33	59	84	69	130	29	82	50	37	25	28	200	68	205	36	128	22	88	54	32	48	22	47	20	28	28	
	CYS	131	102	142	161	61	22	213	34	114	100	129	78	69	85	166	99	85	80	227	188	198	187	181	209	149	103	106	71	63	98	64	
	LEU	304	304	349	704	110	25	661	69	648	382	242	432	432	147	242	341	164	121	153	208	822	467	367	221	179	211	85	160	84	83	126	
	SER	13	45	31	27	349	404	404	404	218	120	48	67	158	208	213	23	73	166	205	169	18	14	49	58	182	133	212	188	74	118	195	
	THR	43	78	37	08	20	77	172	62	153	118	264	169	187	119	211	149	119	171	138	261	86	72	103	156	223	169	179	184	53	129	131	
	PRO	114	105	103	34	74	177	12	187	56	57	114	107	111	145	169	103	187	181	152	229	25	21	100	96	146	126	210	133	111	136	151	
	ALA	47	74	44	51	89	142	22	218	76	90	104	73	123	124	147	147	114	140	153	98	315	49	66	170	110	187	122	181	157	65	84	
	GLY	68	145	87	27	165	180	107	60	114	360	365	248	253	212	203	227	89	81	116	112	327	324	244	245	92	123	194	120	143	67	135	
	VAL	53	298	405	143	133	57	432	124	148	78	304	96	259	224	158	200	193	165	182	153	337	322	233	104	106	236	83	249	23	120	156	
	LYS	108	143	95	143	228	227	17	275	159	181	175	154	200	163	242	110	133	124	73	184	84	81	91	95	275	165	285	138	181	158	173	
	ASN	348	264	267	173	383	163	215	269	118	158	192	247	184	122	210	205	248	250	58	145	186	186	200	197	267	278	289	233	208	183	206	
	GLN	98	139	101	114	241	485	11	392	178	93	173	179	179	317	375	326	244	272	182	111	217	223	182	244	156	168	204	432	204	432	280	
	HIS	234	183	254	342	161	59	235	19	205	123	278	201	176	268	268	132	231	208	270	188	51	51	124	123	154	224	217	425	288	342	499	
	GLU	102	129	124	102	38	58	210	62	107	82	142	94	106	108	71	184	58	85	87	94	210	215	184	187	98	98	84	51	67	105	52	88
	ASP	207	266	289	282	325	602	82	538	407	318	327	341	368	261	250	379	476	411	128	95	85	180	223	229	236	228	207	410	371	369	154	
	TYR	302	319	321	308	158	130	602	198	228	235	198	302	279	202	424	287	208	256	145	365	409	344	317	278	299	198	191	192	159	289	289	
	CYS	164	124	80	184	232	82	231	147	124	193	98	128	171	229	145	89	90	92	78	174	168	369	275	329	285	244	402	381	244	402	381	
	PHE	161	87	85	64	38	21	79	09	57	37	81	57	57	29	114	114	131	100	163	120	134	118	128	75	90	100	106	95	50	80	80	
	ILE	28	31	18	22	40	38	09	19	41	28	42	41	100	31	51	85	104	150	265	56	42	61	107	138	83	151	112	95	113	118	118	
	MET	78	19	70	00	203	268	05	237	205	211	186	236	182	182	207	158	193	193	156	267	182	161	194	207	238	122	327	131	383	140	272	
	TRP	38	17	30	111	249	289	08	311	275	263	182	248	248	261	185	139	205	112	128	188	208	287	173	285	106	206	107	238	153	218	64	
	TER	161	128	118	191	181	182	189	189	214	242	371	182	170	182	218	258	214	242	203	178	170	182	170	182	287	289	160	250	287	270	270	
	UAG	09	09	03	03	01	13	06	07	03	01	15	09	06	03	01	06	05	10	16	14	04	09	05	02	10	10	06	08	07	00	04	02
	TOTAL	6956	7043	19689	6962	4752	19386	28328	38008	45088	8697	10272	10286	87323	25848	58300	9080	72503	9886	44575	21028	5207	25449	9063	27483	6921	14387	10086	6921	14387	10086	6921	

4	PPH	REO	SIV	SND	VAC	VAZ	VSV	LAM	PZ2	PT4	PT7	PZA	EGR	MPO	MZE	PEA	SPI	TOB	YSC	
	NO.GENES												CP	CP	CP	CP	CP	CP	CP	MT
ARG	87	25	72	27	71	72	119	65	20	101	82	23	22	54	36	20	36	64	24	
CGA	87	92	38	28	57	81	118	71	90	57	81	42	92	144	106	110	145	02		
CGC	62	73	39	16	17	11.7	13	166	153	60	145	40	43	20	73	41	50	48	05	
CGG	47	61	24	50	14	10.9	27	100	78	11	30	40	08	11	32	35	41	38	11	
CGU	90	117	15	30	61	147	27	168	138	193	229	113	188	160	167	141	160	194	34	
AGA	202	135	343	267	234	99	204	93	129	87	55	116	263	180	189	129	155	173	199	
AGG	87	91	177	195	40	58	79	47	78	17	34	57	22	09	43	43	63	36	30	
CUA	125	118	161	198	190	110	112	32	73	68	103	146	86	34	136	142	162	133	84	
CUC	41	100	109	158	89	72	127	82	89	41	107	78	03	09	67	50	54	41	11	
CUG	103	185	147	177	71	124	104	363	270	85	218	59	03	08	61	61	64	68	56	
CUU	118	180	108	161	127	174	136	136	189	201	184	134	180	217	190	215	217	191	54	
UUA	329	161	209	181	231	308	164	87	129	265	123	226	450	727	350	385	333	311	969	
UUG	168	171	171	167	192	174	178	57	98	105	102	122	113	63	203	197	219	199	143	
SER	148	287	125	219	133	131	223	133	135	178	90	175	215	137	108	90	118	81	304	
UCA	83	118	78	102	93	145	156	112	66	33	128	45	35	34	151	112	109	104	51	
UCC	36	139	29	66	75	122	44	88	87	40	43	42	32	19	51	36	69	40	26	
UCG	185	175	101	218	244	147	188	68	120	245	179	144	183	255	137	172	206	185	181	
UCU	93	106	97	127	60	90	92	171	147	56	71	97	22	22	40	41	46	56	31	
AGU	163	137	89	110	169	98	139	114	108	109	92	198	116	166	159	126	147	130	100	
THR	384	180	247	270	204	238	248	132	171	173	101	318	336	212	178	157	146	143	210	
ACA	142	99	108	154	87	187	148	211	141	65	184	68	19	28	117	116	120	136	48	
ACC	52	187	38	65	84	173	44	192	90	52	78	87	22	20	58	84	47	52	26	
ACU	218	242	180	181	228	121	188	91	108	273	197	175	229	264	245	263	225	229	163	
CCA	80	186	288	130	184	172	212	88	128	133	81	73	175	161	101	121	125	113	169	
CCC	48	82	122	82	40	178	84	48	48	13	17	23	19	19	100	68	83	74	38	
CCG	48	87	60	61	62	143	80	172	86	57	78	47	30	19	58	46	57	58	20	
CCU	247	119	198	139	130	114	186	74	132	149	148	141	202	216	179	213	198	207	191	
ALA	262	184	281	237	121	208	178	268	339	285	171	158	281	223	230	248	187	218	210	
GCC	82	127	123	123	83	210	83	277	165	55	139	89	38	33	118	114	116	126	59	
GCG	35	153	59	44	78	186	21	265	180	60	147	86	62	25	63	88	82	72	26	
GCU	178	275	190	188	147	181	171	170	249	393	436	222	342	372	374	422	312	370	317	
GGA	171	234	344	172	228	227	365	131	163	303	117	172	383	323	281	276	283	299	128	
GGC	126	90	132	102	35	81	63	214	156	93	145	98	30	41	108	81	99	98	30	
GGG	118	118	164	189	30	154	146	145	99	41	89	99	38	44	147	129	131	133	39	
GGU	217	143	209	136	240	183	108	100	120	189	174	137	291	206	288	392	283	266	420	
GUA	217	143	209	136	240	183	108	100	120	189	174	137	291	206	288	392	283	266	420	
GUC	52	144	87	172	78	90	144	114	102	58	118	73	27	23	67	63	63	66	66	
GUG	170	281	141	192	81	165	160	240	129	52	151	174	32	22	101	78	80	77	44	
GAU	171	209	82	134	229	235	187	189	207	324	187	318	296	289	225	230	203	233	228	
AAA	325	182	378	245	488	246	387	371	438	638	222	427	708	611	377	288	346	342	411	
AAG	148	195	277	272	248	107	280	200	354	169	445	309	59	25	137	84	116	108	136	
AAC	143	186	161	219	189	182	109	213	228	158	330	377	102	67	92	93	137	140	120	
AAU	278	287	289	239	486	217	232	187	210	422	123	307	404	435	233	299	308	293	903	
CAA	271	184	324	162	155	215	184	98	117	235	168	201	242	358	288	288	262	253	218	
CAG	203	249	278	188	62	135	141	331	270	114	198	127	27	22	97	84	73	84	39	
CAC	82	61	78	83	68	108	93	69	72	46	137	56	30	32	61	56	73	73	36	
CAU	184	78	150	102	128	158	181	107	108	129	83	84	184	180	175	210	209	183	177	
GAA	343	175	421	263	422	319	378	372	390	578	287	438	455	456	396	353	382	407	264	
GAG	189	280	261	312	148	188	239	275	330	104	417	224	62	36	110	113	121	131	67	
GAA	232	234	150	210	149	217	269	248	240	149	377	304	81	34	83	73	83	89	110	
GAU	361	403	247	364	473	318	389	363	303	478	247	397	275	297	285	298	301	289	258	
UAC	81	100	127	163	139	130	181	124	177	88	281	175	73	47	67	55	74	92	74	
UAU	283	198	288	202	367	208	239	178	128	319	124	273	259	306	249	241	228	227	348	
UAG	87	44	130	83	34	63	111	86	75	32	60	12	51	20	20	17	26	21	08	
UGC	178	84	148	95	152	149	98	37	48	65	65	52	57	81	61	78	69	68	64	
UGU	48	134	88	173	149	65	231	148	158	110	245	207	100	51	174	197	227	188	265	
UUC	367	189	173	119	347	338	210	182	105	315	108	288	350	543	289	367	374	279	333	
UAU	213	187	214	281	322	180	182	87	120	118	48	163	398	278	221	178	220	190	149	
AUA	244	221	158	182	403	259	238	230	264	507	248	278	353	587	416	438	340	384	988	
AUC	173	287	208	220	282	208	233	227	272	272	272	272	272	272	272	272	272	272	272	
AUU	133	138	271	113	74	108	228	144	159	144	159	144	159	144	159	144	159	144	159	
UAU	33	10	10	07	13	15	18	18	27	28	28	19	48	32	28	17	20	28	34	
UAG	08	02	11	02	03	03	01	04	00	03	04	01	03	03	03	12	05	06	08	
UAA	07	07	11	04	05	05	05	28	39	11	15	14	03	01	11	12	06	08	100	
TOTAL	8582	12774	21880	26116	28163	36314	86213	14149	3357	24028	13749	5798	3715	19786	7427	6048	10694	10694	15768	8099

Table 2. Correspondence between species names and abbreviations in Table 1.

HUM	HUMAN	FPL	F PLASMID
HAM	HAMSTER	HAL	HALOBACTERIUM HALOBIUM
MUS	MOUSE	KPN	KLEBSIELLA PNEUMONIA
RAT	RAT	NGO	NEISSERIA
BOV	BOVINE	PRM	PROTEUS
DOG	DOG	PSE	PSEUDOMONAS
PIG	PIG	RCA	RHODOBACTER CAPSULATUS
RAB	RABBIT	RHB	BRADYRHIZOBIUM JAPONICUM
SHP	SHEEP	RHL	RHIZOBIUM LEGUMINOSARUM
CHK	CHICKEN	RHM	RHIZOBIUM MELILOTI
FSB	FISH (GROUP B)	SMA	SERRATIA MARCESCENS
XEL	XENOPUS LAEVIS	SSP	SULFOLOBUS VIRUS-LIKE PARTICLE
BMO	BOMBYX MORI	STA	STAPHYLOCOCCUS AUREUS
CEL	C. ELEGANS	STM	STREPTOMYCES
DDI	DICTYOSTELIUM DISCOIDEUM	STR	STREPTOCOCCUS
DRO	DROSOPHILA MELANOGASTER	STY	SALMONELLA TYPHIMURIUM
PFA	PLASMODIUM FALCIPARUM	SYN	SYNECHOCOCCUS
SCM	SCHISTOSOMA MANSONI	TIP	TI PLASMID
SUP	SEA URCHIN (P.MILIARIS)	TRN	TRANSPOSON TN
SUS	SEA URCHIN (S.FRANCISCANUS) (S.PURPURATUS)	VIB	VIBRIO
TRB	TRYPANOSOMA BRUCEI	YEP	YERSINIA
ASN	ASPERGILLUS NIDULANS	ADB	ADENOVIRUS TYPE 2
ATH	ARABIDOPSIS THALIANA	FLA	INFLUENZA A
BLY	BARLEY	FLB	INFLUENZA B
MZE	MAIZE	HIV	HUMAN IMMUNODEFICIENCY VIRUS
NEU	NEUROSPORA	HPB	HEPATITIS B VIRUS
PEA	PEA	HS1	HERPES SIMPLEX VIRUS TYPE 1 (HSV-1)
PHV	BEAN (P.VULGARIS)	HS2	HERPES SIMPLEX VIRUS TYPE 2 (HSV-2)
POT	POTATO	HS4	EPSTEIN-BARR VIRUS (EBV)
RIC	RICE	HS5	HUMAN CYTOMEGALOVIRUS
SOY	SOYBEAN	MCV	CUCUMBER MOSAIC VIRUS
SPI	SPINACH	MEA	MEASLES VIRUS
TOB	TOBACCO	MHV	MOUSE HEPATITIS VIRUS
TOM	TOMATO	NDV	NEWCASTLE DISEASE VIRUS
WHT	WHEAT	NPA	A. CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS
YSC	YEAST SACCHAROMYCES CEREVISIAE	PIF	HUMAN PARAINFLUENZA VIRUS
YSK	YEAST KLUYVEROMYCES	PLY	POLYOMA VIRUS
YSP	YEAST SCHIZOSACCHAROMYCES POMBE	PPH	HUMAN PAPILLOMAVIRUS
AFA	ALCALIGENES	REO	REOVIRUS
ANA	ANABAENA	SIV	SIMIAN VIRUS
ATU	AGROBACTERIUM TUMEFACIENS	SND	SENDAI VIRUS
AVI	AZOTOBACTER VINELANDII	VAC	VACCINIA VIRUS
BME	BACILLUS MEGATERIUM	VAZ	VARICELLA-ZOSTER VIRUS
BPE	BORDETELLA PERTUSSIS	VSV	VESICULAR STOMATITIS VIRUS
BSP	BACILLUS SP.	LAM	BACTERIOPHAGE LAMBDA
BSS	BACILLUS SPHAERICUS	P22	BACTERIOPHAGE P22
BST	BACILLUS STEAROTHERMOPHILUS	PT4	BACTERIOPHAGE T4
BSU	BACILLUS SUBTILIS	PT7	BACTERIOPHAGE T7
BTH	BACILLUS THURINGIENSIS	PZA	BACTERIOPHAGE PZA
CHT	CHLAMYDIA TRACHOMATIS	EGR CP	EUGLENA GRACILIS CHLOROPLAST
CLO	CLOSTRIDIUM	MPO CP	MARCHANTIA POLYMORPHA CHLOROPLAST
ECO	ESCHERICHIA COLI	MZE CP	MAIZE CHLOROPLAST
FDI	FREMYELLA DIPLOSIPHON	PEA CP	PEA CHLOROPLAST
		SPI CP	SPINACH CHLOROPLAST
		TOB CP	TOBACCO CHLOROPLAST