Second Nucleotide

First Nucleotide	T					C					A					G						
			pol	*vol*	*hel*			*pol*	*vol*	*hel*			*pol*	*vol*	*hel*			*pol*	*vol*	*hel*		**mean**
T	TTT	F	5.2	132	0.37	TCT	S	9.2	32	-0.99	TAT	Y	6.2	136	-1.25	TGT	C	5.5	55	-0.53	*pol*	6.6
	TTC	F	5.2	132	0.37	TCC	S	9.2	32	-0.99	TAC	Y	6.2	136	-1.25	TGC	C	5.5	55	-0.53	*vol*	89.69
	TTA	L	4.9	111	0.93	TCA	S	9.2	32	-0.99	TAA	--				TGA	--				*hel*	-0.361
	TTG	L	4.9	111	0.93	TCG	S	9.2	32	-0.99	TAG	--				TGG	W	5.4	170	0.23		
C	CTT	L	4.9	111	0.93	CCT	P	8	32.5	-1.72	CAT	H	10.4	96	0.59	CGT	R	10.5	124	-0.16	*pol*	9.027
	CTC	L	4.9	111	0.93	CCC	P	8	32.5	-1.72	CAC	H	10.4	96	0.59	CGC	R	10.5	124	-0.16	*vol*	90.88
	CTA	L	4.9	111	0.93	CCA	P	8	32.5	-1.72	CAA	Q	10.5	96	0.57	CGA	R	10.5	124	-0.16	*hel*	-0.093
	CTG	L	4.9	111	0.93	CCG	P	8	32.5	-1.72	CAG	Q	10.5	96	0.57	CGG	R	10.5	124	-0.16		
A	ATT	I	5.2	111	0.06	ACT	T	8.6	61	-0.68	AAT	N	11.6	56	-0.97	AGT	S	9.2	32	-0.99	*pol*	9.393
	ATC	I	5.2	111	0.06	ACC	T	8.6	61	-0.68	AAC	N	11.6	56	-0.97	AGC	S	9.2	32	-0.99	*vol*	84
	ATA	I	5.2	111	0.06	ACA	T	8.6	61	-0.68	AAA	K	11.3	119	0.71	AGA	R	10.5	124	-0.16	*hel*	-0.248
	ATG	M	5.7	105	1.39	ACG	T	8.6	61	-0.68	AAG	K	11.3	119	0.71	AGG	R	10.5	124	-0.16		
G	GTT	V	5.9	84	-0.09	GCT	A	8.1	31	1.35	GAT	D	13	54	-0.06	GGT	G	9	3	-1.72	*pol*	9.507
	GTC	V	5.9	84	-0.09	GCC	A	8.1	31	1.35	GAC	D	13	54	-0.06	GGC	G	9	3	-1.72	*vol*	46.63
	GTA	V	5.9	84	-0.09	GCA	A	8.1	31	1.35	GAA	E	12.3	83	1.96	GGA	G	9	3	-1.72	*hel*	0.123
	GTG	V	5.9	84	-0.09	GCG	A	8.1	31	1.35	GAG	E	12.3	83	1.96	GGG	G	9	3	-1.72		
			pol	*vol*	*hel*			*pol*	*vol*	*hel*			*pol*	*vol*	*hel*			*pol*	*vol*	*hel*		
	mean		5.294	106.5	0.471			8.475	39.13	-0.51			10.76	91.43	0.221			8.92	73.33	-0.71		

Figure S4. Genetic code table illustrating the biochemical basis for site-specific base frequency variation. The universal genetic code is shown using DNA sequences for codons, one-letter IUPAC amino acid codes, and three physicochemical characteristics [side chain polarity (*pol*), side chain volume (*vol*), and a measure of propensity to form an α-helix (*hel*)] for each amino acid. The first two properties are described in Grantham (1974) and the helical propensity is property 7 in Table III of Kidera et al. (1985). The weighted averages for each property are shown below (for second codon positions) and to the right (for first codon positions) of the genetic code table. Specific codon positions show very different averages (e.g., codons with T in the second position are non-polar, codons with G in the first position or C in the second are small, and there are low helical propensities for codons with a C or G in the second position). These relationships are imperfect since there is variation within columns and rows, but they are potentially strong enough to drive site-specific preferences in base frequencies if there is selection to preserve any specific amino acid properties at certain positions. This among-sites variation in equilibrium base frequencies is a violation of the GTR model.

REFERENCES

Grantham R. Amino acid difference formula to help explain protein evolution. Science, 1974;185:862-864.

Kidera A., Konishi Y., Oka M., Ooi T., Scheraga H.A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. J Protein Chem, 1985;4:23-55.