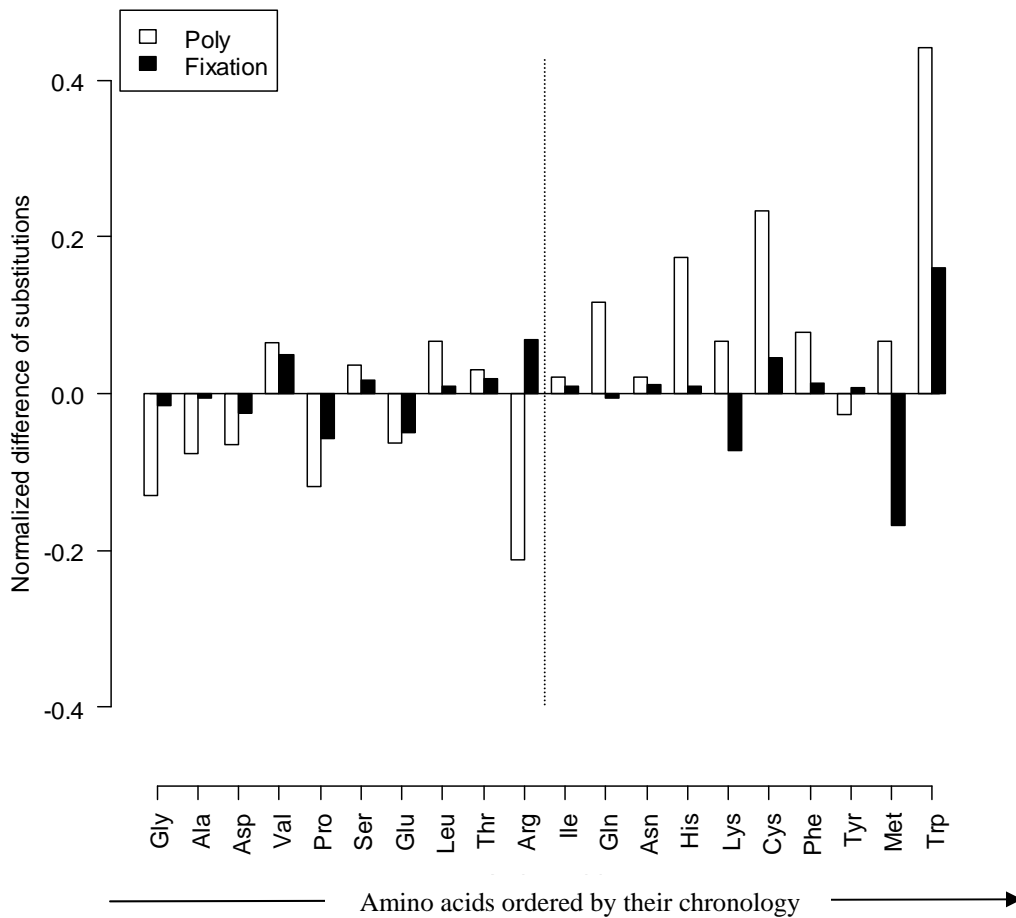


Supplementary data

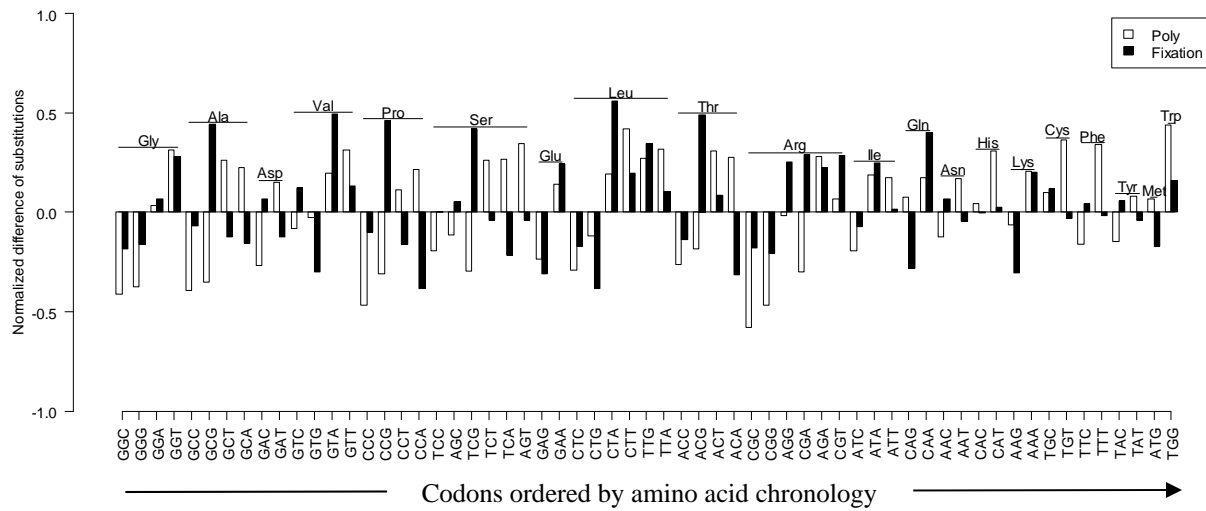
Supplementary Table 1 Sequence composition at each codon position

	Position 1		Position 2		Position 3		Total	
	Poly	Fixed	Poly	Fixed	Poly	Fixed	Poly	Fixed
GC (%)	56.0	55.1	43.3	41.1	57.3	57.6	52.2	51.3
No. of sites	2803	2765	2466	2199	5984	9043	11,253	14,007
A (%)	26.8	27.4	30.8	31.7	20.1	19.5	25.9	26.2
T (%)	17.3	17.5	25.9	27.2	22.6	22.9	21.9	22.5
G (%)	31.0	31.2	19.2	18.5	28.3	28.0	26.2	25.9
C (%)	25.0	23.9	24.0	22.6	29.0	29.6	26.0	25.4

Supplementary Figure 1. Normalized difference of amino acid changes (gain or loss) at the polymorphic and fixed sites in amino-acid coding regions. The normalized difference for an amino acid is defined as the difference of the number of mutations removing the amino acid from creating the amino acid and divided by the total number of mutations in the amino acid. Amino acids are ordered in their temporal order in the genetic code according to Trifonov [1]. The ancient amino acids are at the left while the late ones are at the right.



Supplementary Figure 2. Normalized difference of codon changes (gain or loss) at the polymorphic and fixed sites in amino-acid coding regions. The normalized difference for a codon is defined as the difference of the number of mutations removing the codon from creating the codon and divided by the total number of mutations in the codon. Codons are ordered in the amino acid temporal order in the genetic code according to Trifonov [1]. The ancient amino acids are at the left while the late ones are at the right. Within each amino acid, the early-coming codons are at the left while the late-coming codons are at the right.



Reference

[1] E. N. Trifonov, "The triplet code from first principles," *J. Biomol. Struct. Dyn.*, vol. 22, pp. 1-11, 2004.