

Selective Constraints on Amino Acids
Estimated by a Mechanistic Codon Substitution Model
with Multiple Nucleotide Changes

Sanzo Miyazawa

sanzo.miyazawa@gmail.com

Graduate School of Engineering, Gunma University, Japan

presented at

The 13th annual meeting for the Society of Evolutionary Studies, Japan in Kyoto
(July 29-31, 2011)

ABSTRACT

Background Empirical substitution matrices represent the average tendencies of substitutions over various protein families by sacrificing gene-level resolution. We develop a codon-based model, in which mutational tendencies of codon, a genetic code, and the strength of selective constraints against amino acid replacements can be tailored to a given gene. First, selective constraints averaged over proteins for the respective types of amino acid replacements are estimated by maximizing the likelihood of each 1-PAM matrix of empirical amino acid (JTT, WAG, and LG) and codon (KHG) substitution matrices. Then, selective constraints specific to a given protein family are approximated as a linear function of those estimated from the empirical substitution frequency matrices.

Results Akaike information criterion (AIC) values indicate that a model allowing multiple nucleotide changes fits the empirical substitution frequency matrices significantly better. Also, the ML estimates of transition-transversion bias obtained from these empirical matrices are not so large as previously estimated. The selective constraints are characteristic of proteins rather than species. However, their relative strengths among amino acid pairs can be approximated not to depend very much on protein families but amino acid pairs, because the present model, in which selective constraints are approximated to be a linear function of those estimated from the JTT/WAG/LG/KHG matrices, can provide a good fit to other empirical substitution matrices including cpREV for chloroplast proteins and mtREV for vertebrate mitochondrial proteins.

Conclusions/Significance The present codon substitution model with adjustable parameters for codon mutation rates and for selective constraints would be useful as a simple substitution model in ML and Bayesian inferences of molecular phylogenetic trees, and enables us to obtain biologically meaningful information at both nucleotide and amino acid levels from protein-coding sequences.

Reference: doi:10.1371/PLoS One, 6, e17244, 2011.

1. INTRODUCTION

Purpose and distinctive features of the present study:

- To develop a mechanistic codon substitution model for evolutionary analysis of protein-coding sequences.
 - A time-reversible Markov model, which consists of mutations at the nucleotide level and selection at the amino acid level.
 - Codon substitutions due to multiple nucleotide changes in infinitesimal time are allowed.
 - Selective constraints on amino acids specific to a given protein family are approximated as a linear function of the average selection constraints over many proteins.
- To confirm the significance of multiple nucleotide changes in codon substitutions.
- To estimate the average selective constraints for the respective types of amino acid replacements from observed substitution frequency matrices.
- To show that the present codon substitution model is better than nucleotide and amino acid substitution models, because mutational tendencies and the strength of selective constraint can be both tailored to each gene, and the model enables us to estimate their parameters.

2. METHODS

A mechanistic codon substitution model

Transition matrix over time t : $S(t) = \exp(Rt)$ with $f_\mu R_{\mu\nu} = f_\nu R_{\nu\mu}$

Substitution rate matrix: $R_{\mu\nu} = \text{const } M_{\mu\nu} \frac{f_\nu}{f_\nu^{\text{mut}}} e^{w_{\mu\nu}}$ for $\mu \neq \nu$, **normalized** to $\sum_\mu f_\mu R_{\mu\mu} = -1$

Mutation rate matrix: $M_{\mu\nu} = \prod_{i=1}^3 [\delta_{\mu_i\nu_i} + (1 - \delta_{\mu_i\nu_i}) m_{\mu_i\nu_i} f_{i,\nu_i}^{\text{mut}}]$ for $\mu \neq \nu$

Selective constraints: $e^{w_{\mu\nu}} = \begin{cases} 0 & \text{for } \mu \text{ or } \nu \in \{\text{stop codons}\} \\ e^{w_{a_\mu b_\nu}} & \text{otherwise} \end{cases}$ for $\mu \neq \nu$

where $\mu = (\mu_1, \mu_2, \mu_3), \nu = (\nu_1, \nu_2, \nu_3) \in \{\text{codons}\}$, $\mu_i, \nu_j \in \{t, c, a, g\}$, and $a, b, a_\mu, b_\nu \in \{\text{amino acids}\}$.

f_μ Equilibrium frequency of codon μ

f_μ^{mut} Equilibrium frequency of codon μ for the M ; $f_{\mu=(\mu_1, \mu_2, \mu_3)}^{\text{mut}} = f_{\mu_1}^{\text{mut}} f_{\mu_2}^{\text{mut}} f_{\mu_3}^{\text{mut}}$

$f_{\mu_i}^{\text{mut}}$ Equilibrium frequency of nucleotide μ_i for the M

a_μ the type of amino acid coded by codon μ

$w_{ab} = w_{ba}$ Selective constraint against substitutions between amino acids a and b ; $w_{aa} = 0$ and $w_{ab} < 0$ for $a \neq b$

$m_{\mu_i\nu_i}$ Exchangeability between nucleotides μ_i and ν_i ; $m_{\mu_i\nu_i} = m_{\nu_i\mu_i}$

Likelihood of an observed substitution frequency matrix , $(N f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}})$

Log-likelihood:	$\ell(\boldsymbol{\theta}) = N \sum_{\kappa} \sum_{\lambda} f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}} \log(f_{\kappa} \langle S \rangle (\tau, \sigma)_{\kappa\lambda})$	$\kappa, \lambda = \mu, \nu \text{ or } a, b.$
Kullback-Leibler Information:	$\hat{I}_{\text{KL}}(\boldsymbol{\theta}) = -\ell(\boldsymbol{\theta})/N + \sum_{\kappa} \sum_{\lambda} f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}} \log(f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}})$	$\kappa, \lambda = \mu, \nu \text{ or } a, b.$
Mean of $S(t)$ over t or rate:	$\langle S \rangle (\tau, \sigma) = \int_0^{\infty} S(t) \Gamma(t; \tau, \sigma) dt = [(I - \sigma R)^{-1}] \tau$	
	$\mu = (\mu_1, \mu_2, \mu_3), \mu_i \in \{t, c, a, g\}, a, b \in \{\text{amino acids}\}$	

Estimation of parameters:

Amino acid frequencies:	$\hat{f}_a = f_a^{\text{obs}}$
Codon frequencies:	$\hat{f}_{\mu} = f_{\mu}^{\text{obs}}$ or $C_{\mu a} \hat{f}_{\mu} = f_a^{\text{obs}} C_{\mu a} \hat{f}_{\mu}^{\text{usage}} / \sum_{\nu} C_{\nu a} \hat{f}_{\nu}^{\text{usage}}, \hat{f}_{\mu=(\mu_1, \mu_2, \mu_3)}^{\text{usage}} = \hat{f}_{\mu_1}^{\text{usage}} \hat{f}_{\mu_2}^{\text{usage}} \hat{f}_{\mu_3}^{\text{usage}}$
Shape parameter $\hat{\tau}$ of Γ :	$\sum_{\kappa} \hat{f}_{\kappa} \langle S(\hat{\tau}, \sigma) \rangle_{\kappa\kappa} = \sum_{\kappa} f_{\kappa}^{\text{obs}} S_{\kappa\kappa}^{\text{obs}} \quad \kappa, \lambda = \mu, \nu \text{ or } a, b.$
By maximizing a likelihood:	$\hat{\boldsymbol{\theta}} \equiv \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = \arg \min_{\boldsymbol{\theta}} \hat{I}_{\text{KL}}(\boldsymbol{\theta})$
Relative exchangeabilities:	$\hat{m}_{tc ag} / \hat{m}_{[tc][ag]}, \hat{m}_{ag} / \hat{m}_{tc ag}, \hat{m}_{ta} / \hat{m}_{[tc][ag]}, \hat{m}_{tg} / \hat{m}_{[tc][ag]}, \hat{m}_{ca} / \hat{m}_{[tc][ag]}$
Relative ratio of multiple nucleotide changes:	$\hat{m} (\equiv \hat{m}_{[tc][ag]})$
Scale parameter of Γ :	$\hat{\sigma}$ of a Γ distribution for rate variation
Selective constraints:	\hat{w}_{ab} , which is used later as w_{ab}^{estimate} , or $\hat{\beta}$ and \hat{w}_0 in $\hat{w}_{ab} = \hat{\beta} w_{ab}^{\text{estimate}} + \hat{w}_0 (1 - \delta_{ab})$
Codon usage:	$\hat{f}_{\mu_i}^{\text{usage}}$ if codon frequencies are unknown.

Evaluation of model: Smaller AIC or Δ AIC means a better model.

Akaike Information Criterion: $AIC \equiv -2\ell(\hat{\boldsymbol{\theta}}) + 2 \cdot (\text{number of adjustable parameters})$

$$\Delta AIC \equiv AIC + 2N \sum_{\kappa} \sum_{\lambda} f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}} \log(f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}}) = 2N \hat{I}_{\text{KL}}(\boldsymbol{\theta}) + 2 \cdot (\# \text{parameters})$$

Log-odds: $\log-O(\langle S \rangle(t))_{\kappa\lambda} \equiv \frac{10}{\log 10} \log \frac{\langle S \rangle(t)_{\kappa\lambda}}{f_{\lambda}} \quad \kappa, \lambda = \mu, \nu \text{ or } a, b$

Empirical substitution matrices used for model fitting:

The following 1 PAM substitution frequency matrices, $(N f_{\kappa}^{\text{obs}} S_{\kappa\lambda}^{\text{obs}})$, are used.

JTT amino acid: compiled from closely related proteins by Jones et al. (1992).

cpREV amino acid: estimated from chloroplast proteins (Adachi et al., 2000),

mtREV amino acid: estimated from vertebrate mitochondrial proteins (Adachi & Hasegawa, 1996) by maximizing the likelihood of a given phylogenetic tree.

WAG amino acid: estimated from proteins encoded in nuclear DNA (Whelan & Goldman, 2001),

LG amino acid: estimated from proteins encoded in nuclear DNA (Le & Gascuel, 2008),

KHG codon: estimated from protein-coding sequences in nuclear DNA (Kosiol et al., 2007) by maximizing the likelihood of given phylogenetic trees and branch lengths.

Table 1.

Model name	Brief description
ML- n	Selective constraints $\{w_{ab}\}$ are estimated by maximizing the likelihood of an empirical substitution matrix. The suffix n means the number of ML parameters.
ML-87	In the ML-87, multiple nucleotide changes are disallowed, and $\{w_{ab}\}$ for all 75 single-step amino acid pairs are estimated.
ML-91	In the ML-91, multiple nucleotide changes are allowed, and $\{w_{ab}\}$ for all 75 single-step amino acid pairs and for 6 groups of multiple-step amino acid pairs are estimated. Equal codon usage is assumed.
ML-200	In the ML-200 for codon substitution matrices, $\{w_{ab}\}$ for all 190 amino acid pairs are estimated.
ML- $n+$	First, the ML- n is used to estimate parameters, and then $\{w_{ab}\}$ for all multiple-step amino acid pairs are estimated by maximizing the likelihood with fixing all other parameters at the values estimated by the ML- n .
JTT-ML91- n , WAG-ML91- n , LG-ML91- n	$w_{ab} = \beta w_{ab}^{\text{JTT/WAG/LG-ML91}} + w_0(1 - \delta_{ab})$, where $w_{ab}^{\text{JTT/WAG/LG-ML91}}$ is one estimated by maximizing the likelihood of the JTT/WAG/LG _{amino acid} in the ML-91. The suffix n means the number of ML parameters.
JTT-ML91+ $-n$, WAG-ML91+ $-n$, LG-ML91+ $-n$	$w_{ab} = \beta w_{ab}^{\text{JTT/WAG/LG-ML91+}} + w_0(1 - \delta_{ab})$, where $w_{ab}^{\text{JTT/WAG/LG-ML91+}}$ is one estimated from the JTT/WAG/LG _{amino acid} in the ML-91+. The suffix n means the number of ML parameters. The JTT/WAG/LG-ML91+ -0 correspond to the JTT/WAG/LG-F.
KHG-ML200- n	$w_{ab} = \beta w_{ab}^{\text{KHG-ML200}} + w_0(1 - \delta_{ab})$, where $w_{ab}^{\text{KHG-ML200}}$ is one estimated by maximizing the likelihood of the KHG _{codon} in the ML-200. The suffix n means the number of ML parameters. The KHG-ML200- 0 correspond to the KHG-F.

3. RESULTS

The effects of multiple nucleotide changes in the 1-PAM JTT

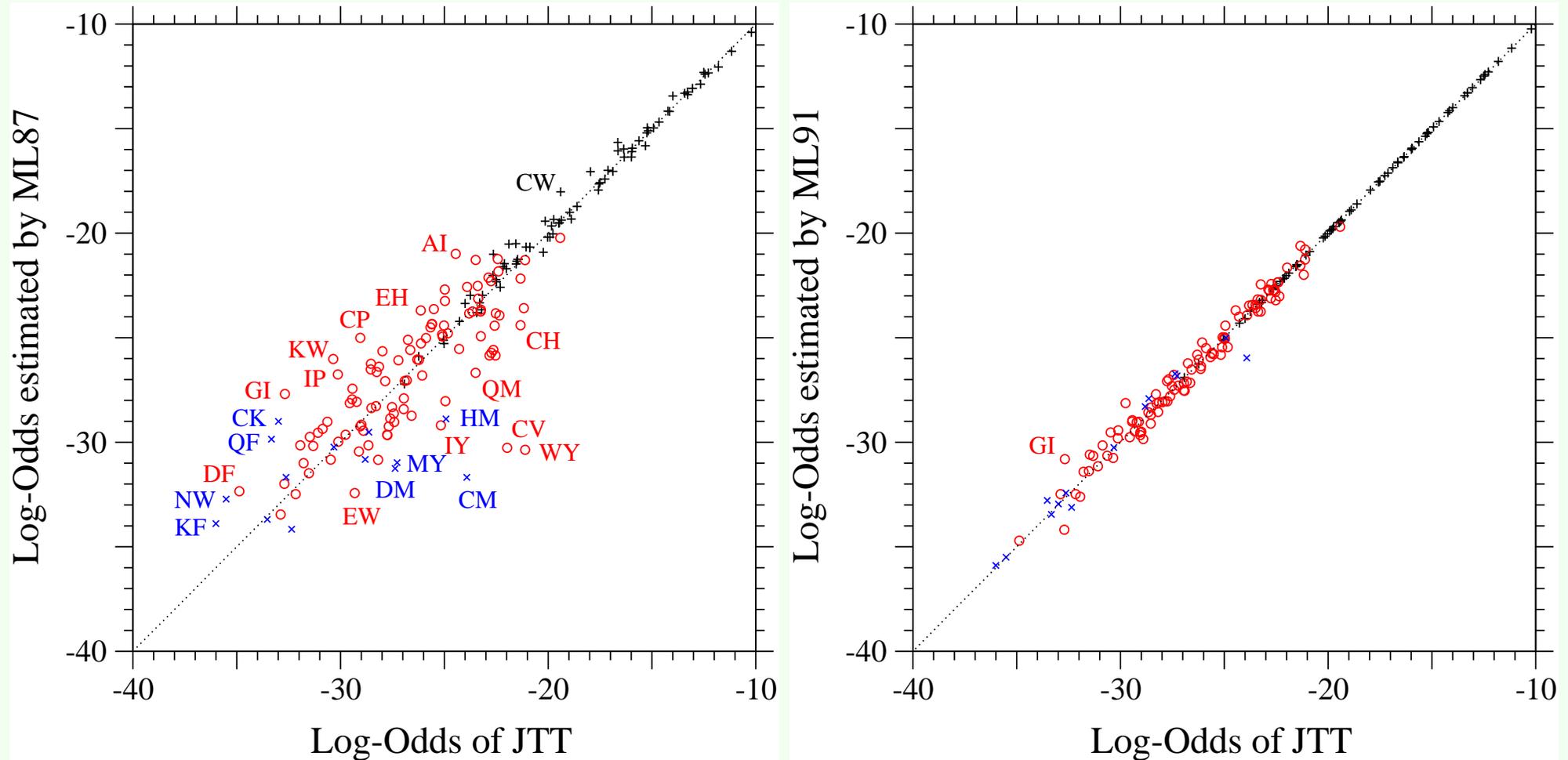
- AIC is significantly improved by taking account of multiple nucleotide changes.

(A) ML-87: Single nucleotide changes only

$$\Delta AIC = 2072.0$$

(B) ML-91: Multiple nucleotide changes allowed

$$\Delta AIC = 297.5$$



+ single nucleotide change

o double nucleotide change

x triple nucleotide change

Fig. 1

Table 2.

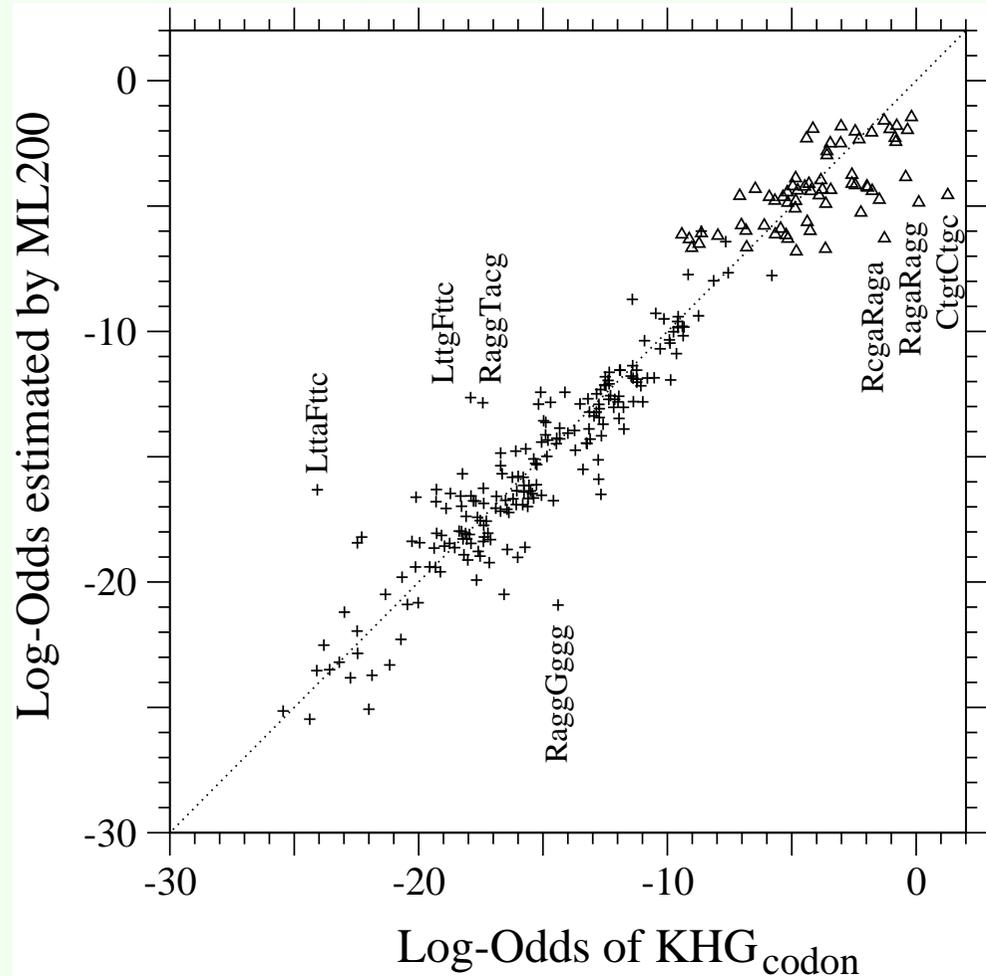
ML estimators in the present models fitted to empirical substitution matrices

id		JTT		WAG		LG	KHG (codon)
		ML-87	ML-91	ML-87	ML-91	ML-91	ML-200
no.	parameter						
0	$-\hat{w}_0$	N/A	N/A	N/A	N/A	N/A	N/A
1	$1/\hat{\beta}$	N/A	N/A	N/A	N/A	N/A	N/A
2	$\hat{m}_{[tc][ag]}$	($\rightarrow 0$)	0.637	($\rightarrow 0$)	1.28	1.08	0.939
3	$\hat{m}_{tc ag}/\hat{m}_{[tc][ag]}$	0.0919	1.57	0.746	1.70	1.85	0.843
4	$\hat{m}_{ag}/\hat{m}_{tc ag}$	1.77	1.14	1.98	1.32	1.23	0.945
5	$\hat{m}_{ta}/\hat{m}_{[tc][ag]}$	0.0293	0.729	0.0477	0.791	0.676	1.52
6	$\hat{m}_{tg}/\hat{m}_{[tc][ag]}$	3.21	0.940	3.64	1.04	1.07	0.554
7	$\hat{m}_{ca}/\hat{m}_{[tc][ag]}$	0.719	1.19	0.110	1.23	1.28	0.573
8	$\hat{f}_{t+a}^{\text{mut}}$	0.408	0.459	0.372	0.367	0.388	0.497
9	$\hat{f}_t^{\text{mut}}/\hat{f}_{t+a}^{\text{mut}}$	0.113	0.501	0.234	0.587	0.450	0.513
10	$\hat{f}_c^{\text{mut}}/\hat{f}_{c+g}^{\text{mut}}$	0.698	0.429	0.425	0.479	0.427	0.470
11	$\hat{f}_{t+a}^{\text{usage}}$	0.0682	(0.5)	0.0669	(0.5)	(0.5)	NA
12	$\hat{f}_t^{\text{usage}}/\hat{f}_{t+a}^{\text{usage}}$	0.461	(0.5)	0.330	(0.5)	(0.5)	NA
13	$\hat{f}_c^{\text{usage}}/\hat{f}_{c+g}^{\text{usage}}$	0.386	(0.5)	0.310	(0.5)	(0.5)	NA
14	$\hat{\sigma}$	27.3	0.738	43.3	0.905	0.415	$\rightarrow 0$
$\hat{\tau}\hat{\sigma}$		0.334	0.0243	0.317	0.0223	0.0246	0.0240
#parameters		107	111	107	111	111	261
$\hat{I}_{KL}(\hat{\theta}) \times 10^8$ ^a		15695	638	35319	1903	2771	269946
ΔAIC ^b		2072.0	297.5	1370.8	284.3	782.5	unknown
Ratio of substitution rates per codon the total base/codon		1.28	1.35	1.38	1.53	1.38	1.29 (1.29) ^c
transition/transversion		0.464	1.08	0.482	0.932	1.18	0.764 (0.765) ^c
nonsynonymous/synonymous ^d		1.13	1.37	1.57	2.07	1.05	0.726 (0.723) ^c
Ratio of substitution rates per codon for $\sigma \rightarrow 0$ total base/codon		1.0	1.22	1.0	1.38	1.31	1.29
transition/transversion		0.101	1.21	0.647	1.11	1.31	0.764
nonsynonymous/synonymous ^d		0.0644	1.04	0.138	1.50	0.853	0.726
Ratio of substitution rates per codon for $w_{ab} = 0$ and $\sigma \rightarrow 0$ total base/codon		1.0	1.45	1.0	1.72	1.67	1.51
transition/transversion		0.0605	0.829	0.499	0.933	0.992	0.427
nonsynonymous/synonymous ^d		11.3	5.58	11.1	8.68	7.45	6.81

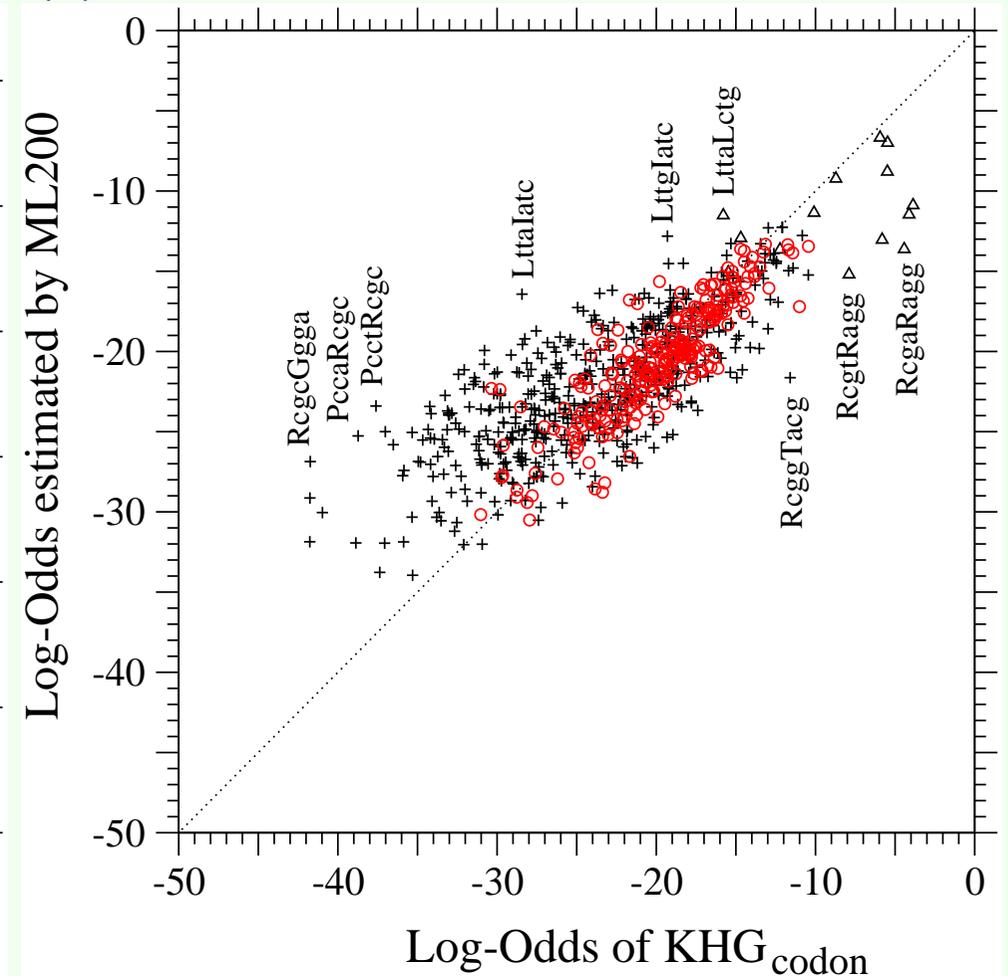
Comparison of the observed and the expected log-odds in the 1-PAM $\text{KHG}_{\text{codon}}$

ML-200 model: The selective constraints (\hat{w}_{ab}) for all 190 amino acid pairs are optimized.

(A) Codon pairs of single nucleotide changes



(B) Codon pairs of double nucleotide changes

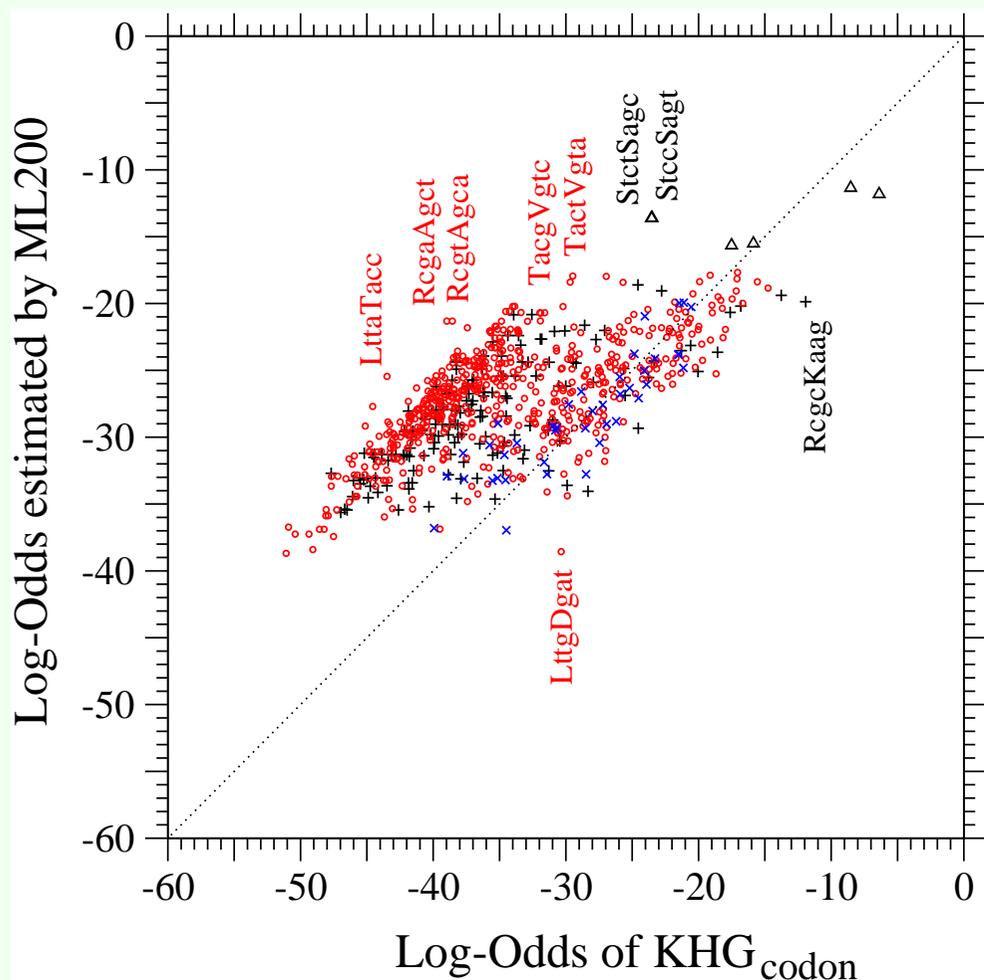


△ synonymous amino acid pair

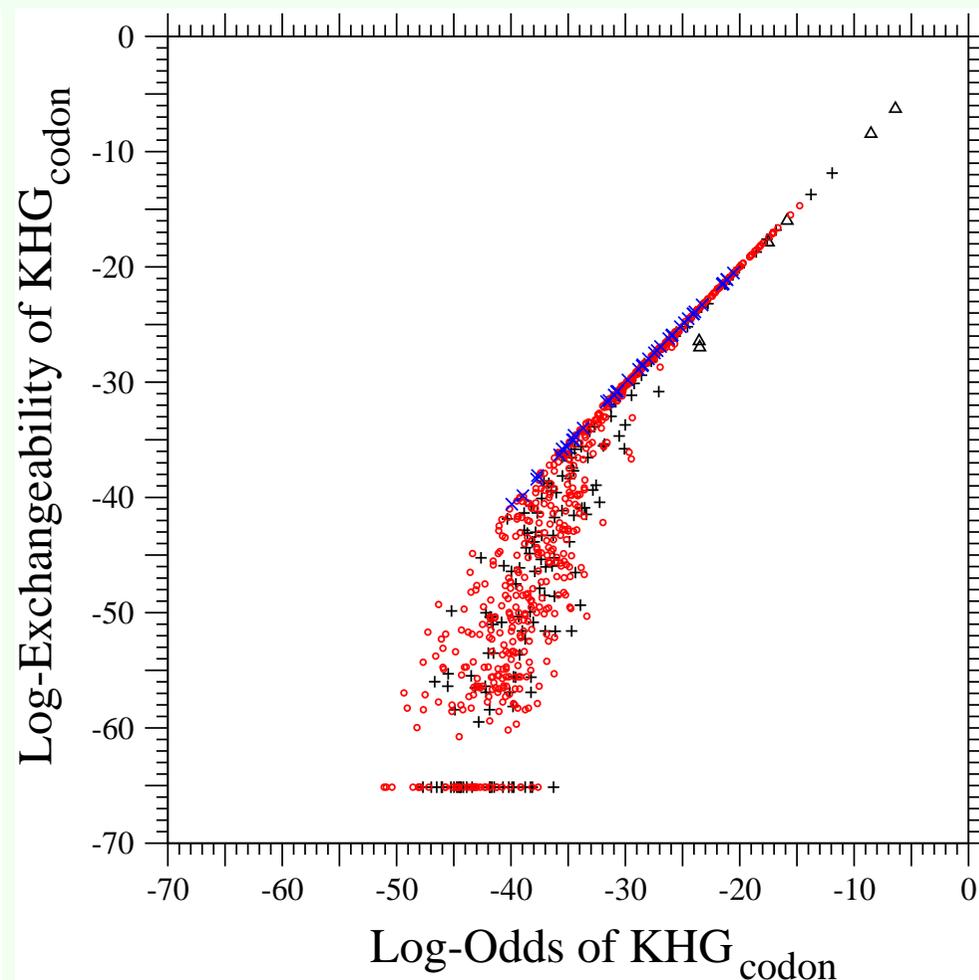
+ one-step amino acid pair

○ two-step amino acid pair

(C) Codon pairs of triple nucleotide changes



(D) Log-exchangeabilities of triple nucleotide changes



\triangle synonymous amino acid pair $+$ one-step amino acid pair \circ two-step amino acid pair \times three-step amino acid pair

Fig. 2

Correlation of selective constraints (\hat{w}_{ab}) between various estimates

Table 3.

	EI	JTT-ML91+	WAG-ML91+	LG-ML91+	KHG-ML200
EI		0.45	0.51	0.59	0.55 (0.65) ^a
JTT-ML91+	0.66		0.80	0.80	0.51
WAG-ML91+	0.68	0.87		0.86	0.55
LG-ML91+	0.71	0.82	0.90		0.58
KHG-ML200	0.71	0.77	0.69	0.74	

Upper half: Correlation coefficients for 75 single-step amino acid pairs.

Lower half: for 86 multi-step amino acid pairs; 29 pairs of the least exchangeable category is excluded.

EI: Physico-chemical estimates of selective constraints based on residue-residue contact energies.

Selective constraints (\hat{w}_{ab}) estimated from JTT and KHG

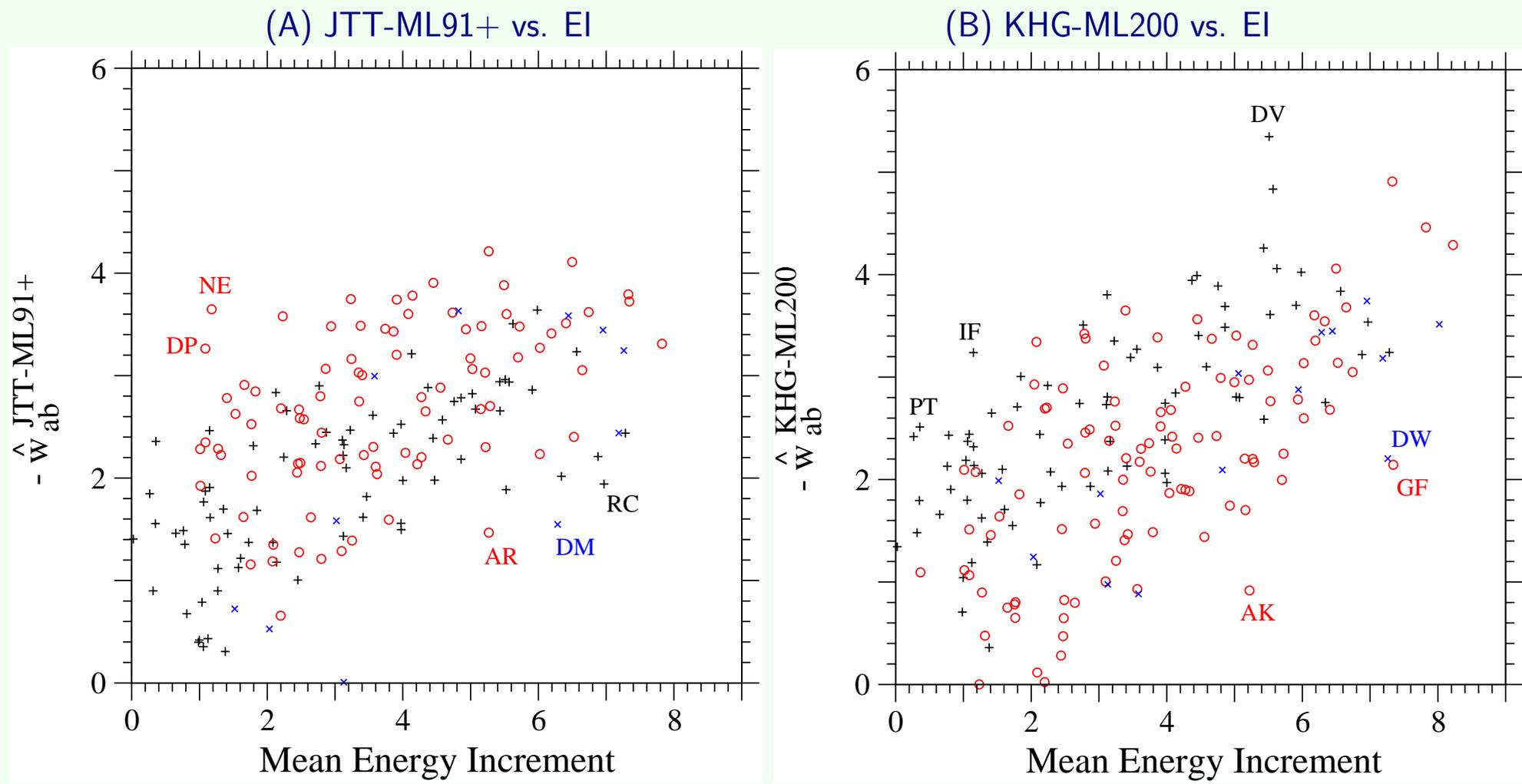


Fig. 3

Performance of various estimates (\hat{w}_{ab}) of selective constraints

- $w_{ab} = \beta w_{ab}^{\text{estimate}} + w_0(1 - \delta_{ab})$, where $w^{\text{estimate}} \equiv \hat{w}^{\text{JTT/WAG/LG-ML91+}}$ or $\hat{w}^{\text{KHG-ML200}}$.
- Parameters including β and w_0 are optimized.

Table 4.

Model No.	#parameters (id no. ^a)	ΔAIC^b					$\hat{I}_{KL}(\hat{\theta}) \times 10^8{}^c$	
		JTT	WAG	LG	cpREV	mtREV	KHG (amino acid)	KHG (codon)
JTT-ML91+								
0	20		2657.5	20807.0	412.1	426.0		
11	31(1-10,14)		1152.9	12140.0	264.5	286.5	40931	
12	32(0-10,14)							473668
WAG-ML91+								
0	20	9095.4		10537.3	283.7	535.1		
11	31(1-10,14)	3299.2		4813.3	235.9	365.1	12789	
12	32(0-10,14)							496804
LG-ML91+								
0	20	13669.8	1806.0		434.5	593.4		
11	31(1-10,14)	3878.5	574.7		243.0	314.9	5732	
12	32(0-10,14)							436557
KHG-ML200								
0	20	15063.5	953.4	12568.9	360.8	593.6		
11	31(1-10,14)	4429.9	518.7	3006.1	229.4	334.1		

^a Parameter id numbers in the parenthesis mean ML parameters in each model and other parameters are fixed to the value of the corresponding parameter listed in the column of the ML-91 or the ML-200 in Table 2; each id number corresponds to the parameter id number listed in Table 2.

^b $\Delta\text{AIC} \equiv 2N\hat{I}_{KL}(\hat{\theta}) + 2 \times \#\text{parameters}$ with $N \simeq 5919000$ for the JTT, $N \approx 1637663$ for the WAG, $N \approx 10114373$ for the LG, $N \approx 149355$ for the cpREV, and $N \approx 137637$ for the mtREV; see text for details.

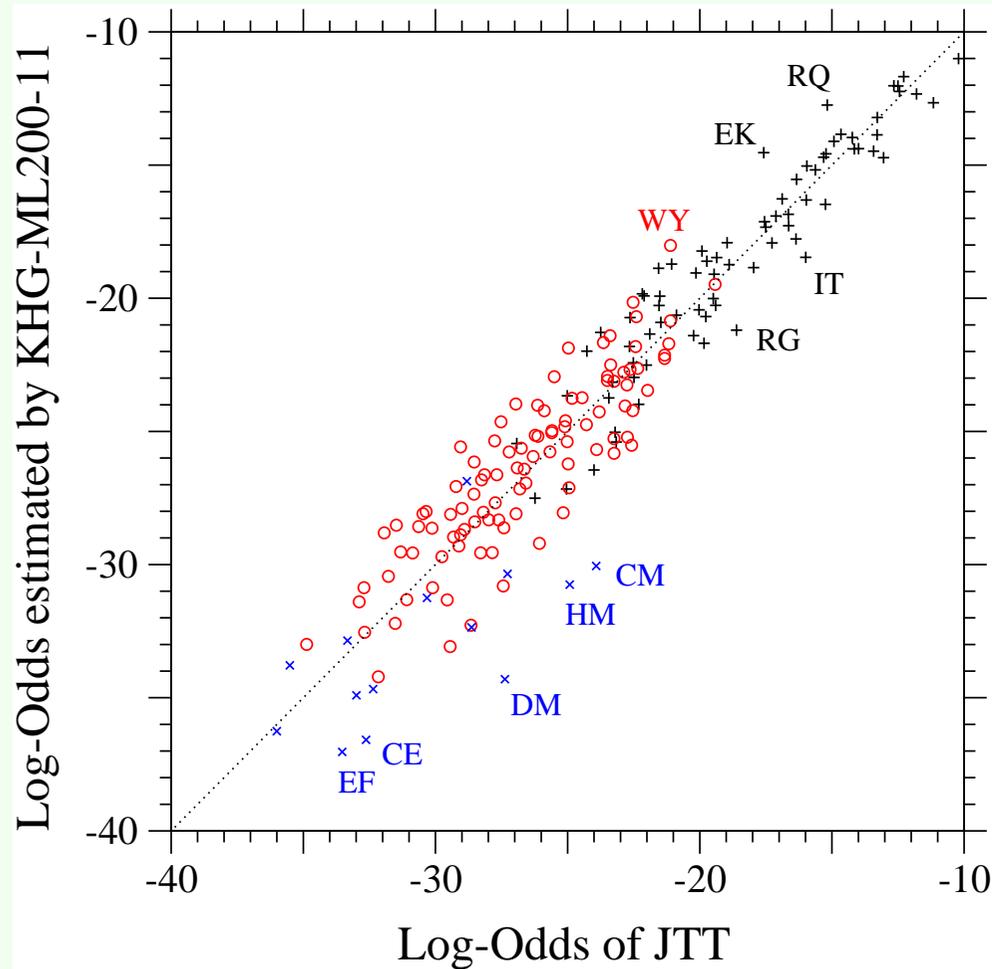
^c $\hat{I}_{KL}(\hat{\theta}) = -(\ell(\hat{\theta})/N + 2.97009788)$ for the KHG-derived amino acid substitution probability matrix, and $-(\ell(\hat{\theta})/N + 4.19073314)$ for the KHG codon substitution probability matrix; see text for details.

Performance of $\hat{w}^{\text{KHG-ML200}}$ estimated from the $\text{KHG}_{\text{codon}}$ in the ML-200

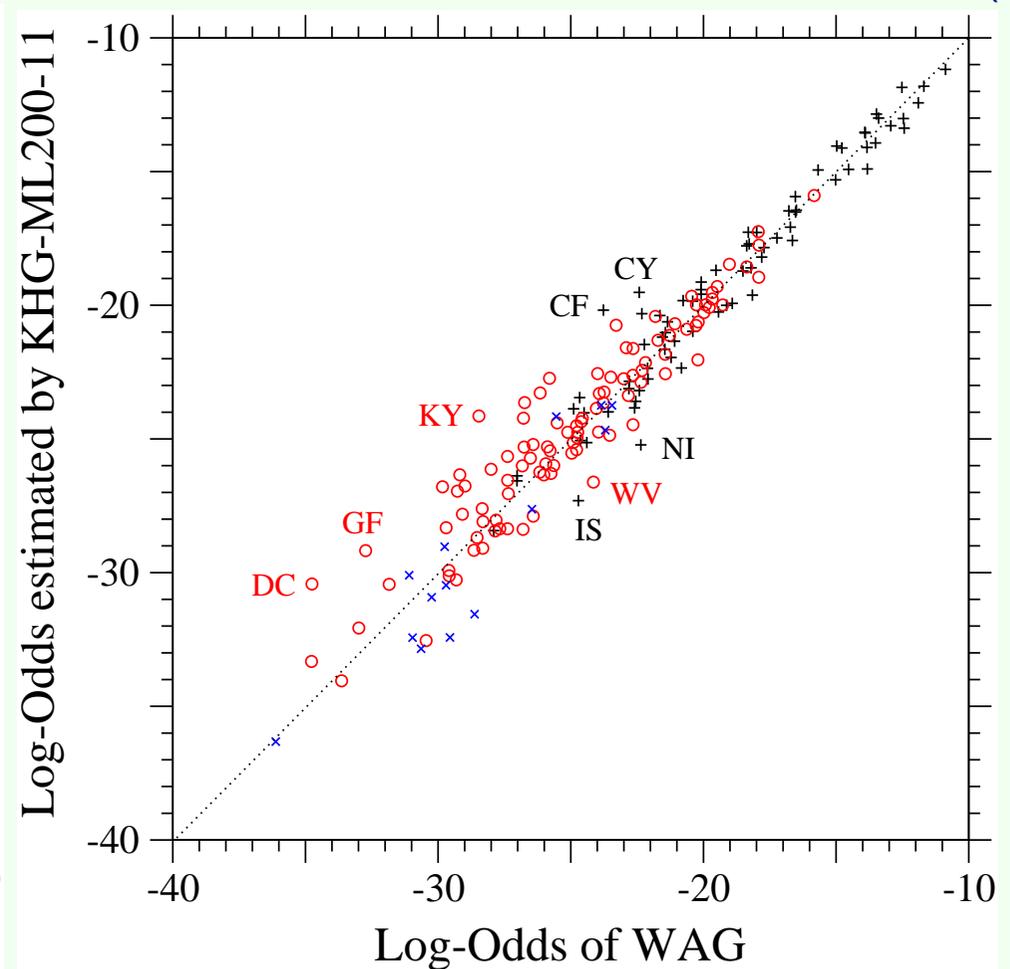
- $w_{ab} = \beta \hat{w}_{ab}^{\text{KHG-ML200}} + w_0(1 - \delta_{ab})$; all parameters including β and w_0 are optimized.

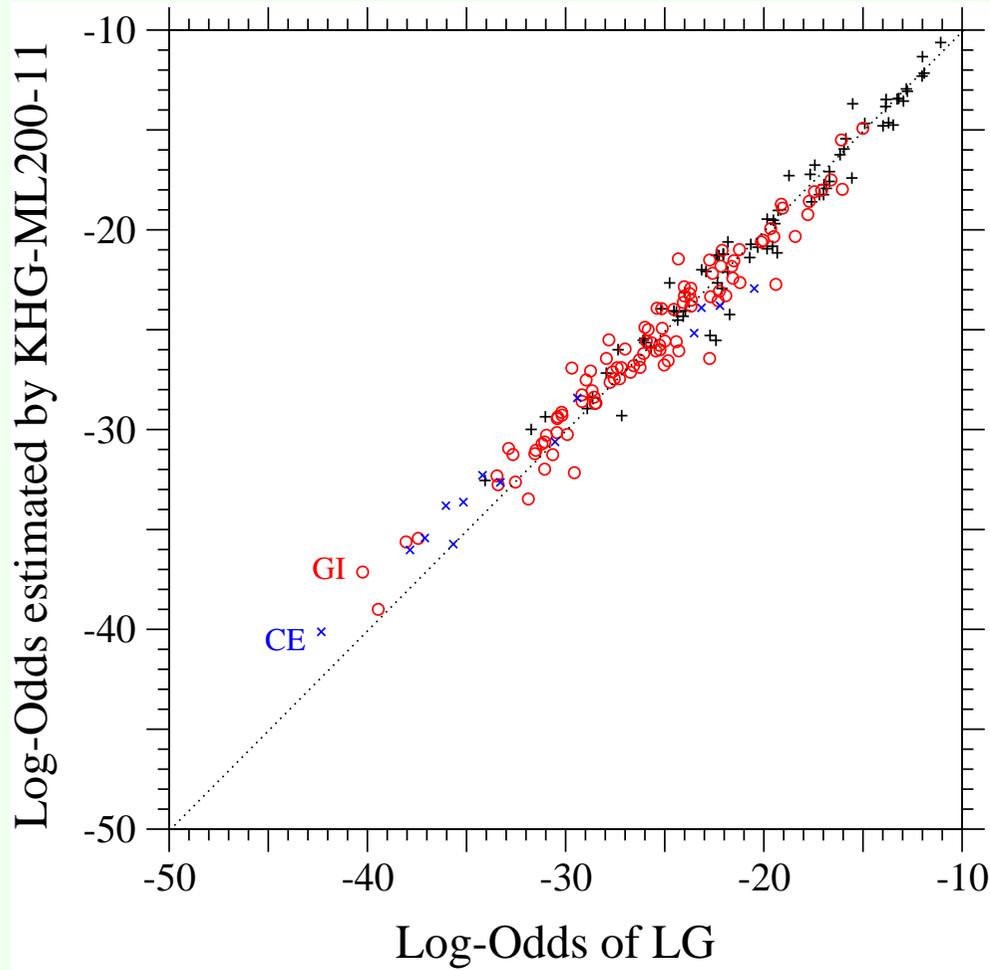
Empirical substitution frequency matrices can be well fitted.

(A)

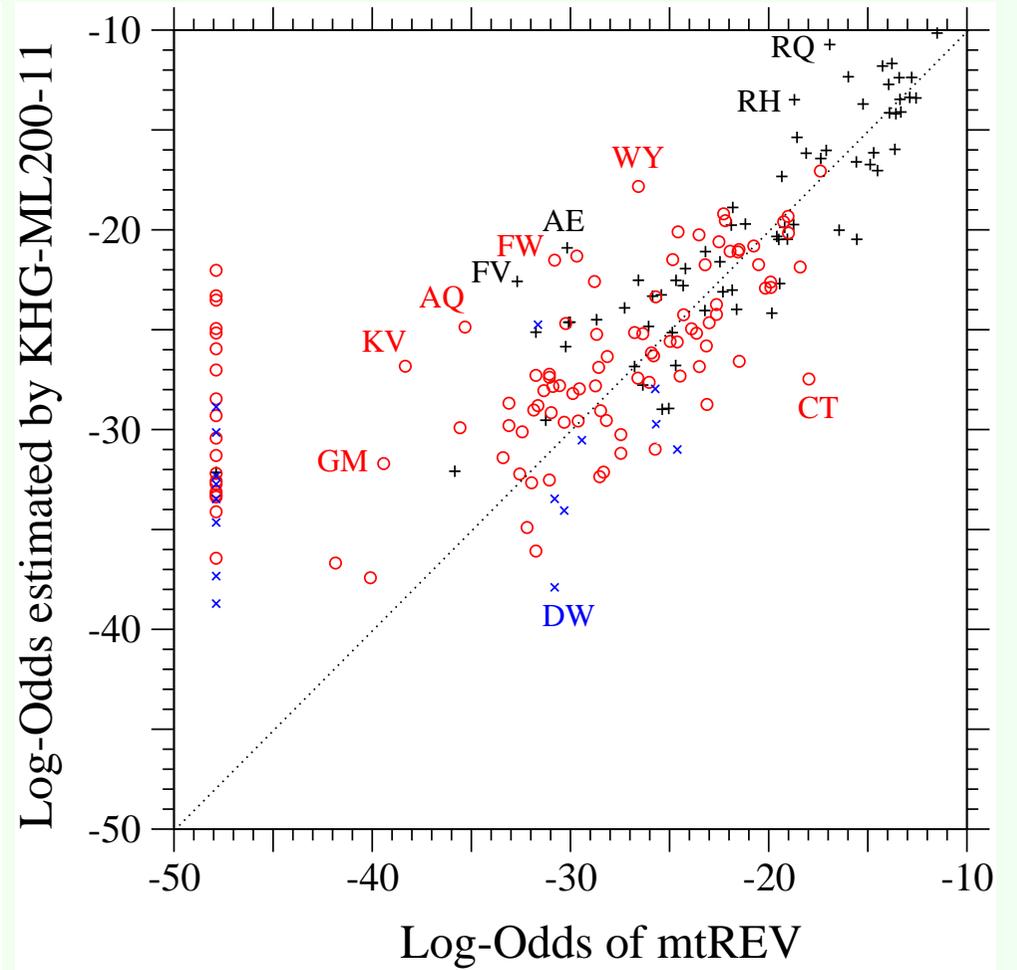


(B)





(C)



(D)

Fig. 4

AIC of a phylogenetic tree of the concatenated coding sequences of 12 proteins in mtDNA from 20 vertebrate species

- The AIC of the tree is significantly improved by the present codon model.

Codon Substitution Model ^a	#p ^b	$\ell+$	AIC–	$\hat{\sigma}$	$\hat{m}(\equiv \hat{m}_{[tc][ag]})$	$\hat{m}_{tc ag}/\hat{m}_{[tc][ag]}$
		116898.6	233917.3			
LG-1-F ^c	60	–1293.8	2587.6			
KHGaa-1-F ^{cd}	60	–1293.0	2586.1			
WAG-1-F ^c	60	–1108.1	2216.1			
JTT-1-F ^c	60	–836.4	1672.8			
mtREV-1-F ^c	60	0.0	0.0			
Equal-Constraints-11-F	70	772.2	–1524.4	0.906	0.273	3.37
EI-12-F	71	1966.6	–3911.2	0.326	0.549	3.60
WAG-ML91+-12-F	71	2268.3	–4514.5	1.84	0.471	4.16
JTT-ML91+-12-F	71	2275.1	–4528.1	3.57	0.506	2.91
KHG-ML200-12-F	71	2355.7	–4689.4	0.469	0.226	2.50
LG-ML91+-12-F	71	2510.0	–4997.9	1.26	0.357	4.32
Equal-Constraints-11-F-dG4	71	2495.4	–4968.9	0.000	0.182	3.62
EI-12-F-dG4	72	3742.4	–7460.7	0.000	0.392	3.95
JTT-ML91+-12-F-dG4	72	4156.9	–8289.8	0.064	0.385	3.11
KHG-ML200-12-F-dG4	72	4190.0	–8356.0	0.000	0.147	2.60
WAG-ML91+-12-F-dG4	72	4196.4	–8368.7	0.042	0.342	4.61
LG-ML91+-12-F-dG4	72	4412.6	–8801.1	0.029	0.253	4.83

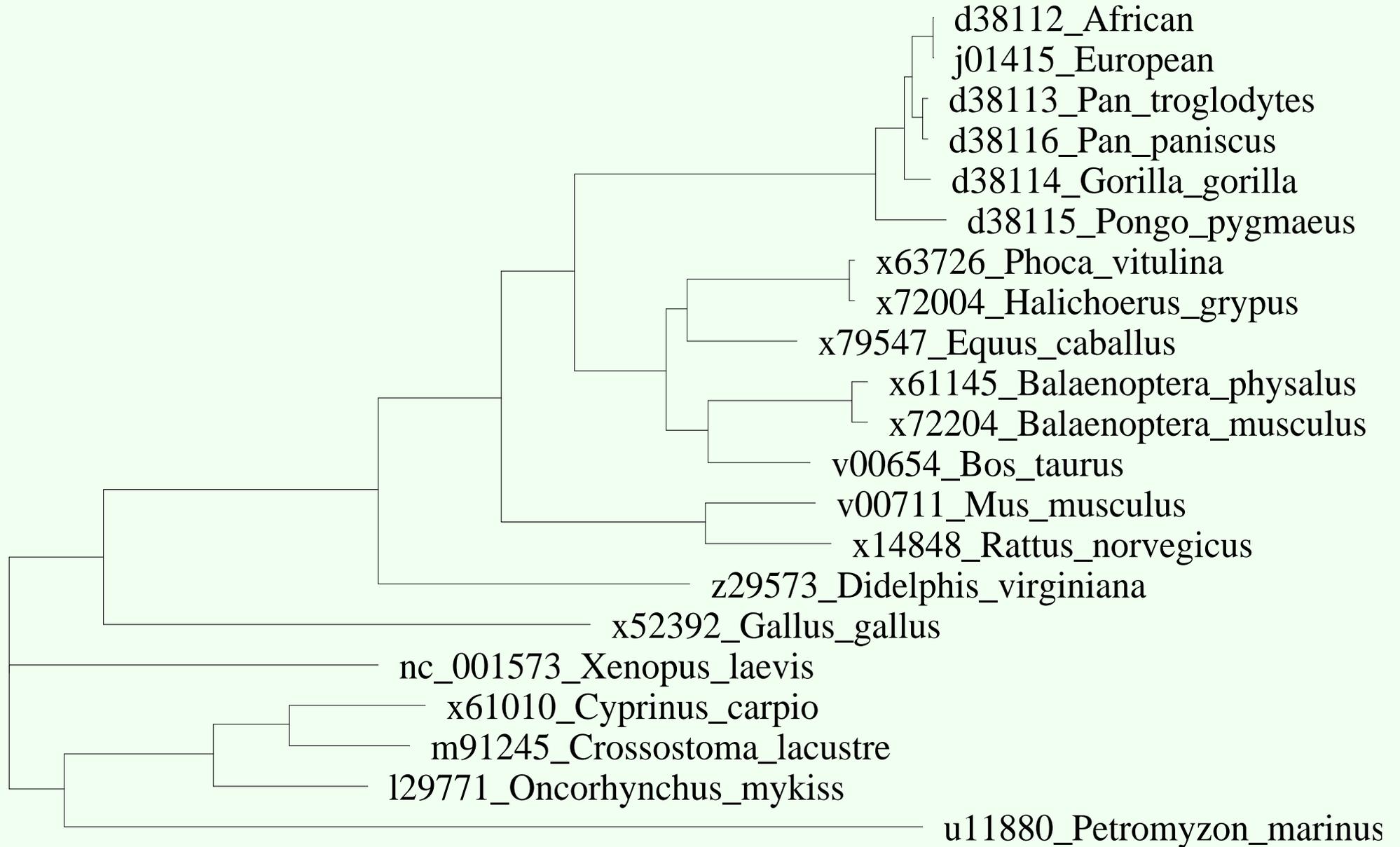
^a In all models named with a suffix "F", equilibrium codon frequencies are assumed to be equal to those in coding sequences. A suffix "dG4" means the discrete approximation of the Γ distribution with 4 categories for rate variation. The parameter w_0 is optimized in all models.

^b The number of parameters; the value for the mtREV-1-F is not quite correct, because mtREV was estimated from the almost same set of protein sequences (Adachi and Hasegawa, 1996).

^c The exchangeabilities of nonsynonymous and synonymous codon pairs are equal to $\exp w_0$ multiplied by those of the corresponding amino acid pairs and all equal to the mean amino acid exchangeability in the empirical amino acid substitution matrix specified, respectively.

^d KHGaa means the amino acid substitution matrix derived from KHG.

Phylogenetic tree of mtDNA from 20 vertebrate species.



4. CONCLUSIONS

- Codon substitutions due to multiple nucleotide changes in infinitesimal time are significant, and must be taken into account to model the substitution process of amino acids.
- The mechanistic codon substitution model is better than nucleotide and amino acid substitution models, because mutational tendencies and the strength of selective constraint can be tailored to each gene.
 - It can well fit a variety of observed substitution frequency matrices such as the JTT, WAG, LG, mtREV, cpREV, and KHG.
 - It can yield better AIC of phylogenetic trees than nucleotide and amino acid substitution models.
 - It allows to estimate mutational tendencies at the nucleotide level and the strength of selective constraint in evolutionary analysis of protein coding sequences; the ratios of transition to transversion exchangeability, non-synonymous to synonymous exchangeability and multiple to single nucleotide exchangeability.