Alternatively Spliced Regions Evolve Faster

Ekaterina Ermakova *ermakova8@yandex.ru*

[1] *Department of Bioengineering and Bioinformatics, Moscow State University, Vorob'evy gory, 1-73, Moscow, 119992, Russia*

[2] *Research and Training Center "Bioinformatics", Institute for Information Transmission Problems RAS, Bolshoi Karetny per. 19, Moscow, 127994, Russia*

For each gene pair, the alignments of constitutive and alternative regions were extracted and concatenated forming two meta-alignments consisting of constitutive and alternative fragments respectively. Then N-terminal, internal & C-terminal alternative regions were considered separately.

In an alternatively spliced gene, constitutive regions are defined as the ones that are always exonic and coding, and alternative regions as the ones that are either coding or spliced out.

The data flow through the analysis pipeline

12356 human mRNA sequences from RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) aligned to mouse orthologs

7283 genes present in the EDAS database (http://www.belosersky.msu.ru/edas/)

5754 genes with more than one protein isoform in EDAS

3079 genes with protein-derived alternatives read in a single frame

3029 alignments with >70% nucleotide identity

2358 genes with the total length of the constitutive regions and the total length of the alternative regions each exceeding 80 nt

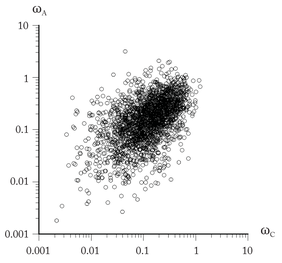
Nucleotide substitutions

Gene evolution can be considered as a story of fixation of nucleotide substitutions. When substitutions in coding regions are considered, the number of nonsynonymous substitutions is usually estimated per nonsynonymous site and the number of synonymous susbstitutions per synonymous site.

The second position of the TCT *codon is nonsynonymous*		*The third position of the* TCT *codon is synonymous*	
TTT	*Phe*	TCT	*Ser*
TCT	*Ser*	TCC	*Ser*
TAT	*Tyr*	TCA	*Ser*
TGT	*Cys*	TCG	*Ser*
A nonsynonymous substitution		*A synonymous substitution*	
GTA	*Val*	GTA	*Val*
↓		↓	
GCA	*Ala*	GTG	*Val*

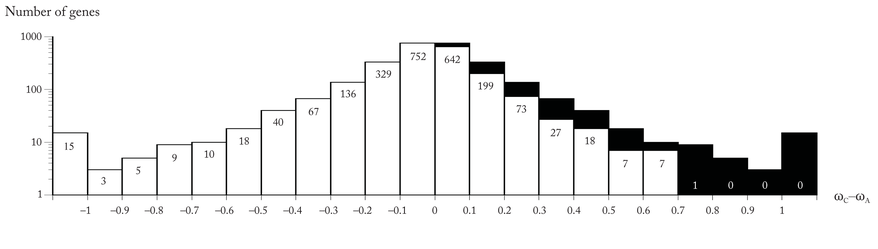
The Ina method I (Ina 1995) gives good estimates of the number of nonsynonymous substitutions per nonsynonymous site d_N and of synonymous substitutions per synonymous site d_S in a coding gene segment since its divergence with an ortholog. $\omega=d_N/d_S$ is used as a measure of selective pressure.

This study was supported by grants from HHMI, LICR, RFBR, RSSF.



For most genes

$$\omega_A > \omega_C$$



d_N/d_S ratio ω was estimated for constitutive (ω_C) and alternative (ω_A) regions of 2358 human genes. The distribution of ω_C—ω_A is asymmetrical: $\omega_A>\omega_C$ for 1384 genes, $\omega_A<\omega_C$ for 974 genes.

genes	*regions*	*total alignment length, bp*	*amino-acid identity*	d_N	d_S	ω
all	*C*	2822439	0.891	0.068	0.408	0.166
	A	3081642	0.879	0.077	0.411	0.187
slow	*C*	897471	0.964	0.020	0.323	0.063
	A	920970	0.960	0.023	0.320	0.072
medium-speed	*C*	978984	0.913	0.052	0.418	0.124
	A	1092459	0.903	0.059	0.419	0.140
fast	*C*	945984	0.800	0.134	0.485	0.277
	A	1068213	0.785	0.147	0.489	0.300

Substitution rates in constitutive and alternative regions of the human-mouse concatenated alignments

genes	*regions*	*total alignment length, bp*	*amino-acid identity*	d_N	d_S	ω
all	*C*	2822439	0.891	0.068	0.408	0.166
	AN	2194521	0.880	0.075	0.405	0.186
	AI	790026	0.884	0.074	0.421	0.176
	AC	97095	0.813	0.132	0.452	0.291
slow	*C*	897471	0.964	0.020	0.323	0.063
	AN	670623	0.961	0.022	0.319	0.069
	AI	230754	0.957	0.025	0.322	0.076
	AC	19593	0.940	0.036	0.327	0.111
medium-speed	*C*	978984	0.913	0.052	0.418	0.124
	AN	751620	0.903	0.059	0.413	0.142
	AI	313896	0.905	0.057	0.431	0.133
	AC	26943	0.887	0.073	0.450	0.162
fast	*C*	945984	0.800	0.134	0.485	0.277
	AN	772278	0.788	0.143	0.481	0.296
	AI	245376	0.787	0.147	0.511	0.288
	AC	50559	0.724	0.207	0.509	0.407

Substitution rates in the constitutive (C), N-terminal alternative (AN), internal alternative (AI), and C-terminal alternative (AC) regions

On average, the alternative coding regions evolve faster than the constitutive ones. In alternative regions, purifying selection is weaker and/or positive selection is stronger. C-terminal alternatives make the main contribution to the observed difference. The effects become even more pronounced in a subset of fastly evolving genes.

Alternative splicing does serve as a testing ground for molecular evolution:

(i) alternatively spliced isoforms are often evolutionary young both in mammals (Modrek and Lee 2003; Nurtdinov et al. 2003) and in insects (Malko, this conference)

(ii) the rate of nonsynonymous substitutions is higher in alternative regions compared to constitutive ones (this study)

(iii) the frequency of nonsynonymous SNPs in human genes is higher in alternative regions than in constitutive regions (Ramensky, this conference)

This is joint work with Mikhail Gelfand. We are grateful to I.K. Jordan, R.N. Nurtdinov, D.B. Malko, V.E. Ramensky, G.A. Bazykin, A.A. Mironov, and D.A. Petrov.