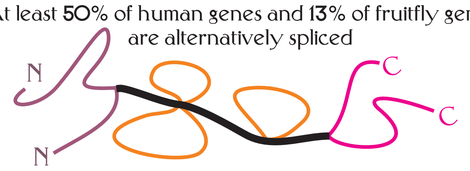
Evolutionary patterns in alternatively spliced coding regions of mammalian and fly genes

Ekaterina O Ermakova ermakova@iitp.ru
[1]Department of Bioengineering and Bioinformatics, Moscow State University, Vorob'evy gory, 1-73, 119992, Moscow, Russia
[2]Research and Training Center "Bioinformatics", Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoi Karetny per. 19, 127994, Moscow, Russia

HGM2006
Helsinki, Finland, 31 May - 3 June 2006

Introduction

At least 50% of human genes and 13% of fruitfly genes are alternatively spliced



Two protein isoforms can have constitutive parts and N-terminal, internal, and/or C-terminal alternatives

Mammals: Homo sapiens — Mus musculus

Flies: Drosophila melanogaster — Drosophila pseudoobscura

In the figures, the value is the median for 2000 bootstrap replications, and the oval height equals 3 times inter-quartile range.

Results: concatenated alignments

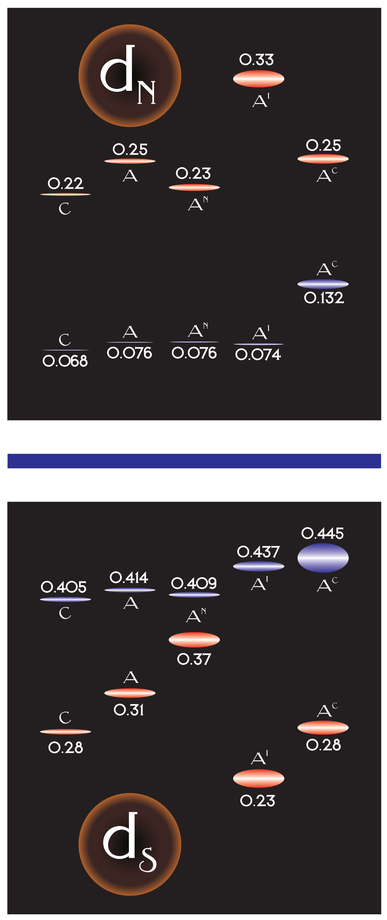
We considered 3029 human and 790 D. melanogaster genes, alternatively spliced in the coding region, and their orthologs in mouse and D. pseudoobscura, respectively. Nucleotide alignments of coding regions were divided into constitutive (C) and alternative (A) fragments; the latter were further sorted into N-terminal (A^N), internal (A^I), and C-terminal (A^C). In these regions, we estimated the nonsynonymous substitution rate (d_N), the synonymous substitution rate (d_S), and $\omega=d_N/d_S$. Concatenated alignments were used (see Methods).

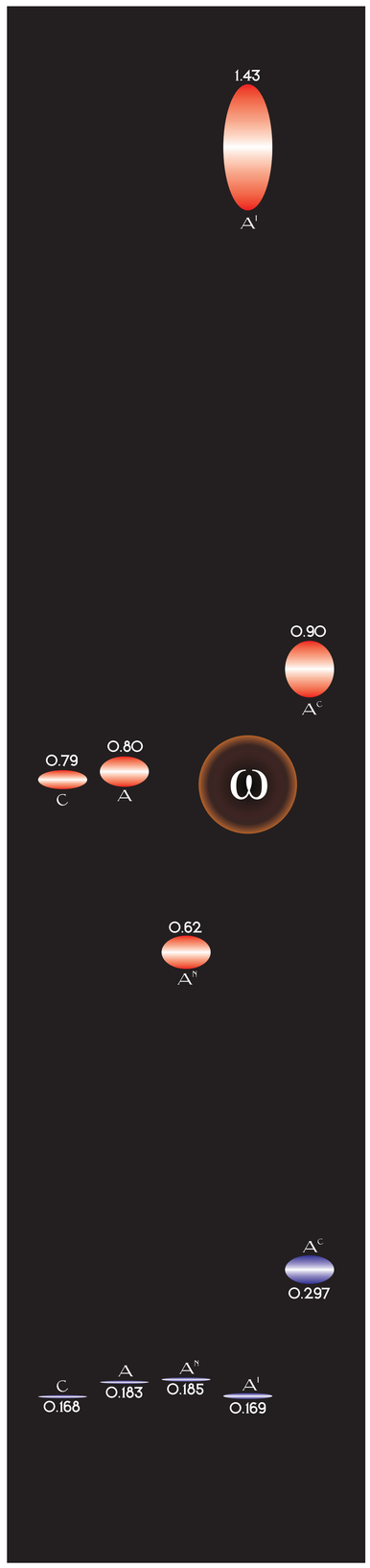
Both in mammals and in flies, amino-acid altering substitutions are more frequent in alternative regions. Moreover, $\omega(A)>\omega(C)$ for alternatively spliced genes irrespective of evolutionary rate. These results show that negative selection is weaker and/or positive selection is stronger in alternative regions.

The pattern of synonymous changes in mammals and flies differ. Fly genes contain more synonymous substitutions in alternative regions than in constitutive ones, whereas in mammalian genes there is no significant difference.

In flies, of all alternative regions, N-terminal alternatives are the most conserved whereas internal ones are the least conserved in terms of nonsynonymous substitutions. The rates of synonymous substitutions are coherent: $d_S(A^C)$ and $d_S(C)$ are close, $d_S(A^I)<d_S(C)$, and $d_S(A^N)>d_S(C)$.

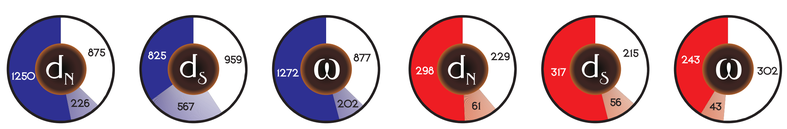
In mammals, the pattern is different. Unexpectedly, d_N and ω are dramatically higher in C-terminal alternative regions. The rate of synonymous substitutions in alternative regions increases in the 5' to 3' direction.



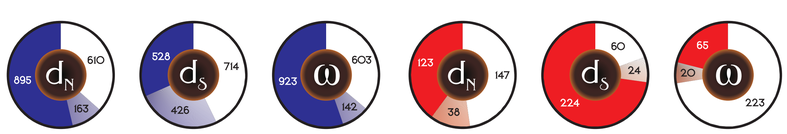


Results: genes with long alternatives

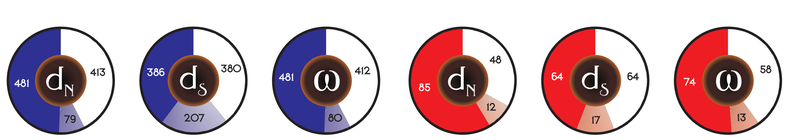
Is the function greater in alternative or in constitutive regions?
2351 mammalian gene pairs and 588 fly gene pairs with long* alternatives vote:



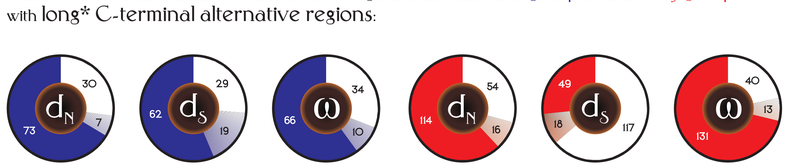
We also compared the same parameters for N-terminal alternative and constitutive regions of 1668 mammalian gene pairs and 308 fly gene pairs with long* N-terminal alternative regions:



... and for internal alternative and constitutive regions of 973 mammalian gene pairs and 145 fly gene pairs with long* internal alternative regions:



... and for C-terminal alternative and constitutive regions of 110 mammalian gene pairs and 184 fly gene pairs with long* C-terminal alternative regions:



* 2351 mammalian gene pairs and 588 fly gene pairs were selected for individual substitution rate analysis. These were the ones with both the total length of the alignment length of the alternative regions and that of the constitutive regions exceeding 80 base pairs. The notatioon is as follows. For the pie chart in the upper left corner: for 875 mammalian gene pairs $d_S(A)<d_S(C)$ [white segment], for 1250 ones $d_S(A)>d_S(C)$ [blue segment], and for 163 ones $d_S(A)$ and $d_S(C)$ are approximately equal [i.e. $|d_S(A)-d_S(C)|/|d_S(A)+d_S(C)|<0.05$].

Notation

C constitutive
A alternative
A^N N-terminal alternative
A^I internal alternative
A^C C-terminal alternative
d_N nonsynonymous substitution rate
d_S synonymous substitution rate
$\omega=d_N/d_S$

Total lengths of concatenated alignments, bp

region type	human vs mouse	D. melanogaster vs D. pseudoobscura
C	2822439	1334760
A	3081642	535074
A^N	2194521	266935
A^I	790026	84558
A^C	97095	183581

Methods

Orthologs Orthologs were identified as described previously in (Jordan et al. 2001) for human and mouse and in (Malko et al. 2006) for D. melanogaster and D. pseudoobscura.

Splicing annotation Human splicing annotation was taken from the EDAS database release 1 [http://www.belozersky.msu.ru/edas/]; we considered only alternative splicing events confirmed at the protein level. D. melanogaster annotation was taken from the Flybase release 3 [ftp://flybase.net/genomes/Drosophila_melanogaster/].

Conservation We considered only conserved splicing. A human splicing event was considered conserved in mouse (or a D. melanogaster splicing event conserved in D. pseudoobscura) if the alignment of the coding regions was good, and the splicing sites were conserved.

Concatenated alignments To estimate the substitution rate in coding regions of a particular type (C, A, A^N, A^I, or A^C), we concatenated all the aligned fragments of this type of all gene pairs for the given organisms into one long alignment. This technique allowed to take into account not long cassette exons only, but short alternative fragments, as ones between alternative donor or acceptor sites, also.

Substitution rates evaluation The transitional to transversional substitution rate ratio R, as well as the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site were estimated by the Ina method I (Ina 1995). For human and mouse R=5.28, for two flies R=2.24.

Precision To evaluate the robustness of the estimates for evolutionary parameters of the concatenates, we used bootstrapping to form 2000 alighments of the same length for reach concatenate and estimated amino-acid

Alternative splicing serves as a testing ground for molecular evolution

- alternatively spliced isoforms are often evolutionary young both in mammals (Modrek and Lee 2003, Nurtdinov et al. 2003) and in insects (Malko et al 2006),
- the rate of nonsynonymous substitutions is higher in alternative regions compared to constitutive ones (this study),
- constitutive exons in genes with genome-specific alternative splicing evolve faster than constitutive regions in genes with conserved structure (Cusack and Wolfe 2005),
- many young (rodent-specific, missing in human and pig as an outgroup) exons are alternatively spliced and tend to have >1 in the mouse-rat comparison (Wang et al. 2005),
- the frequency of nonsynonymous SNPs in human genes is higher in alternative regions than in constitutive regions (Ramensky et al 2005).

References

- Cusack BP, Wolfe KH: Changes in alternative splicing of human and mouse genes are accompanied by faster evolution of constitutive exons. Mol Biol Evol 2005, 22:2198-2208.
- Ermakova EO, Nurtdinov RN, Gelfand MS.Fast rate of evolution in alternatively spliced coding regions of mammalian genes.BMC Genomics. 2006;7:84.
- Ina Y: New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J Mol Evol 1995, 40:190-226.
- Jordan IK, Kondrashov FA, Rogozin IB, Tatusov RL, Wolf YI, Koonin EV: Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. Genome Biol 2001, 2:research0053.1-0053.9.
- Malko DB, Makeev VJ, Mironov AA, Gelfand MS: Evolution of the exon-intron structure and alternative splicing in fruit flies and malarial mosquito genomes. Genome Res, 2006
- Modrek B, Lee CJ: Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 2003, 34:177-180.
- Nurtdinov RN, Artamonova II, Mironov AA, Gelfand MS: Low conservation of alternative splicing patterns in the human and mouse genomes. Hum Mol Genet 2003, 12:1313-1320.
- Ramensky V, Neverov AD, Nurtdinov RN, Mironov AA, Gelfand MS: Human SNPs and alternative splicing. In Proceedings of the Second International Moscow Conference on Computational Molecular Biology: 18-21 July 2005; Moscow. 2005:317-318.
- Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J, Samudrala R, Wang J, Yang H, Yu J, Kristiansen K, Wong GK, Wang J: Origin and evolution of new exons in rodents.Genome Res 2005 15:1258-1264.

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

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