



IS: a web-site for intron statistics

Endre Barta^{1,*}, László Kaján^{1,2} and Sándor Pongor²

¹Agricultural Biotechnology Center, 2100 Gödöllő, Hungary and ²Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology, 34012 Trieste, Italy

Received on April 16, 2002; revised on August 5, 2002; accepted on October 25, 2002

ABSTRACT

Summary: A web server has been established for the statistical evaluation of introns in various taxonomic groups and the comparison of taxonomic groups in terms of intron type, length, base composition, etc. The options include the graphic analysis of splice sites and a probability test for exon-shuffling within the selected group.

Availability: introns.abc.hu, <http://www.icgeb.trieste.it/introns>

Contact: barta@abc.hu; kajan@icgeb.trieste.it; pongor@icgeb.trieste.it

The statistical analysis of introns is a time-consuming task which is not included into the repertoire of the currently available intron www-sites (Croft *et al.*, 2000; Damberger and Gutell, 1994; Lopez and Seraphin, 2000; Sakharkar *et al.*, 2001, 2002; Saxonov *et al.*, 2000). Our goal was to establish an automatically updated intron resource for the analysis of experimentally validated and statistically balanced data sets according to various criteria. In addition to sequence retrieval and BLAST similarity search, there are options to compare taxonomic groups based on the NCBI Taxonomy Database, and to perform on-the-fly statistics on intron and exon length, the number of introns per gene, base composition, intron phases, as well as to carry out a graphic comparison of two or more groups in terms of the above variables. In addition, the analysis of

splice sites and testing of the exon shuffling hypothesis (Kriventseva and Gelfand, 1999; Long *et al.*, 1995) are explicitly included. All analyses can be done either on complete or statistically filtered data sets.

REFERENCES

- Croft,L., Schandorff,S., Clark,F., Burrage,K., Arctander,P. and Mattick,J.S. (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nature Genet.*, **24**, 340–341.
- Damberger,S.H. and Gutell,R.R. (1994) A comparative database of group I intron structures. *Nucleic Acids Res.*, **22**, 3508–3510.
- Kriventseva,E.V. and Gelfand,M.S. (1999) Statistical analysis of the exon–intron structure of higher and lower eukaryote genes. *J. Biomol. Struct. Dyn.*, **17**, 281–288.
- Long,M., Rosenberg,C. and Gilbert,W. (1995) Intron phase correlations and the evolution of the intron/exon structure of genes. *Proc. Natl Acad. Sci. USA*, **92**, 12495–12499.
- Lopez,P.J. and Seraphin,B. (2000) YIDB: the Yeast Intron DataBase. *Nucleic Acids Res.*, **28**, 85–86.
- Sakharkar,M.K., Tan,T.W. and de Souza,S.J. (2001) Generation of a database containing discordant intron positions in eukaryotic genes (MIDB). *Bioinformatics*, **17**, 671–675.
- Sakharkar,M., Passetti,F., de Souza,J.E., Long,M. and de Souza,S.J. (2002) ExInt: an Exon Intron Database. *Nucleic Acids Res.*, **30**, 191–194.
- Saxonov,S., Daizadeh,I., Fedorov,A. and Gilbert,W. (2000) EID: the Exon–Intron Database—an exhaustive database of protein-coding intron-containing genes. *Nucleic Acids Res.*, **28**, 185–190.

*To whom correspondence should be addressed.