ASDB: database of alternatively spliced genes

M. S. Gelfand, I. Dubchak[1],*, I. Dralyuk[1] and M. Zorn[1]

Institute of Protein Research, Russian Academy of Sciences, Pushchino 142292, Russia and [1]National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ABSTRACT

A database of alternatively spliced genes (ASDB) has been constructed based on (i) the results of the analysis of Swiss-Prot entries containing products of these genes and (ii) clustering procedure joining proteins that could arise by alternative splicing of the same gene. ASDB incorporates information about alternatively spliced genes, their products and expression patterns. It can be searched in order to find all products of alternative splicing produced in a particular tissue or a given organism, or all variants generated by a particular transcript. ASDB currently contains about 1700 protein sequences and can be accessed via the Internet at URL http://cbcg.nersc.gov/asdb

INTRODUCTION

Alternative splicing is an important regulatory mechanism in higher eukaryotes (1). By recent estimates, at least 30% of human genes are spliced alternatively (2). Alternative splicing plays a major role in sex determination in *Drosophila*, antibody response in humans and other tissue or developmental stage specific processes (3–6). Alternative splicing can generate up to 64 different mRNA variants from a single transcript (7). The database of alternatively spliced genes can be of potential use for molecular biologists studying splicing, developmental biologists, geneticists and cell biologists.

DESCRIPTION

Version 1.1 of ASDB contains information about protein products of alternatively spliced genes. Selecting all Swiss-Prot (8) entries containing the words 'alternative splicing' has generated 1663 proteins. Then clusters of proteins that could arise by alternative splicing of the same gene were created. Two proteins from the same species belong to a cluster if they have common fragments not shorter than 20 amino acids. Each cluster is represented in the database by the multiple global alignment of its members, allowing for easy identification of regions produced by alternative splicing.

The database contains 241 clusters with more than one member. Distribution of cluster size, representation of species and other relevant statistics are presented in Figures 1 and 2.
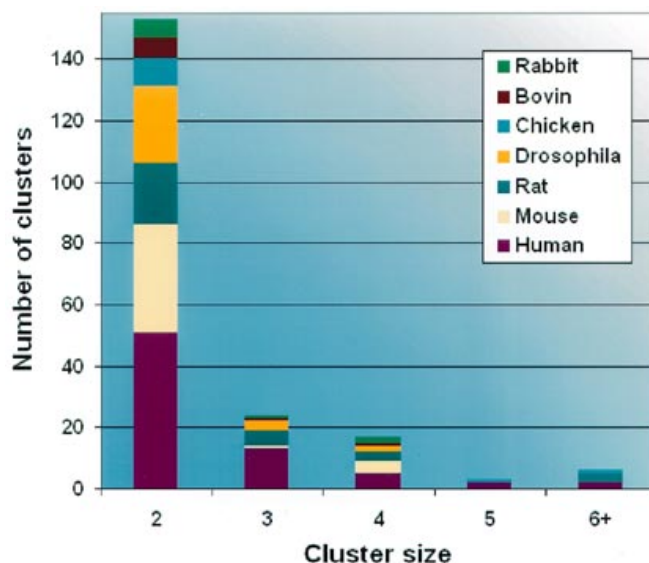

Figure 1. Distribution of cluster size for different species.

The database can be searched using Medline, Swiss-Prot (8), and GenBank (9) identifiers and accession numbers. Standard context search can be performed over Swiss-Prot keywords, description, taxonomy, comment fields and feature tables. ASDB contains internal links between entries and/or clusters, as well as external links to Medline, GenBank and Swiss-Prot entries.

FUTURE DIRECTIONS

The next steps of ASDB development will be incorporation of DNA data, classification of main types of alternative splicing, incorporation of data on aberrant splicing and splicing mutations.

Automated processing of existing databases with minimum manual curation produced the current version of the database. In future we plan to add manual curation of the database, including addition of splicing variants described in the literature but not annotated in GenBank.

---

*To whom correspondence should be addressed. Tel: +1 510 495 2419; Fax: +1 510 486 5548; Email: ildubchak@lbl.gov
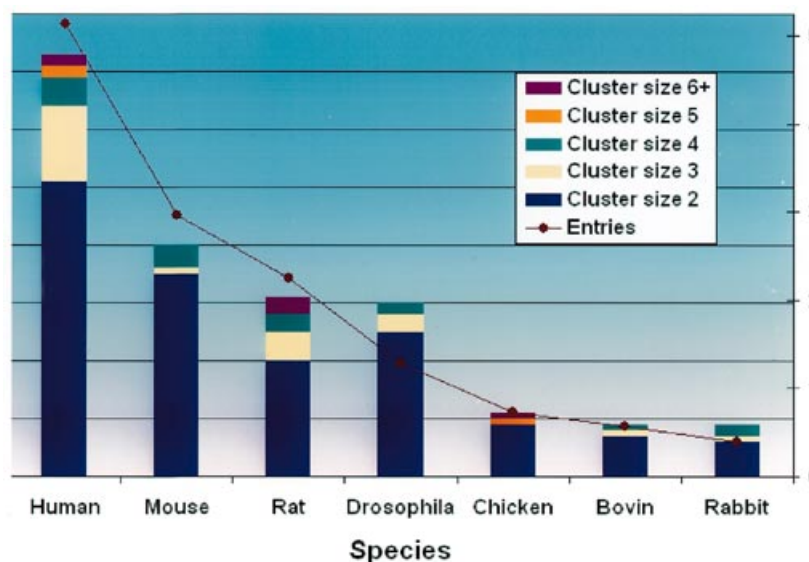
Figure 2. The number of ASDB entries (right vertical axis, dotted line) and the number of clusters (left vertical axis, bars of the histogram) for different species.

AVAILABILITY

ASDB is currently available at the URL http://cbcg.nersc.gov/ asdb . Users of the database are encouraged to provide corrections, comments or new material. The administrator of the database can be contacted by Email at: asdb@lbl.gov.

ACKNOWLEDGEMENTS

REFERENCES

1 Sharp,P.A. (1994) Cell., **77**, 805–8152.
2 Mironov,A.A. and Gelfand,M.S. (1998) *Proc. 1st Int. Conf. on Bioinformatics of Genome Regulation, 1998*. vol. 2, p. 249.
3 Stamm,S., Zhang,M.Q., Marr,T.G. and Helfman,D.M. (1994) *Nucleic Acids Res*., **22**, 1515–1526.
4 Chabot,B. (1996) *Trends Genet*., **12**, 472–478.
5 Breitbart,R.E., Andreadis,A. and Nadal-Ginard,B. (1987) *Annu. Rev. Biochem*., **56**, 467–495.
6 Smith,C.W., Patton,J.G. and Nadal-Ginard,B. (1989) *Annu. Rev. Genet*., **23**, 527–577.
7 Breitbart,R.E. and Nadal-Ginard,N. (1987) *Cell*, **46**, 793–803.
8 Bairoch,A. and Apweiler,R. (1997) *Nucleic Acids Res*., **25**, 31–36.
9 Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F. (1998) *Nucleic Acids Res*., **26**, 1–7.