Published online 6 November 2006 *Nucleic Acids Research, 2007, Vol. 35, Database issue* ***D141–D144***
doi:10.1093/nar/gkl815

dbRES: a web-oriented database for annotated RNA editing sites

Tao He, Pufeng Du and Yanda Li*

Bioinformatics Division, TNLIST and Department of Automation, Tsinghua University, Beijing 100084, China

Received August 7, 2006; Revised October 3, 2006; Accepted October 4, 2006

ABSTRACT

Although a large amount of experimentally derived information about RNA editing sites currently exists, this information has remained scattered in a variety of sources and in diverse data formats. Availability of standard collections for high-quality experimental data will be by of great help for systematic studying of RNA editing, especially for developing computational algorithm to predict RNA editing site. dbRES (http://bioinfo.au.tsinghua.edu.cn/dbRES) is a public database of known RNA editing sites. All sites are manually curated from literature and GenBank annotations. dbRES version 1.1 contains 5437 RNA editing sites of 251 transcripts, covering 96 organisms across plant, metazoan, protozoa, fungi and virus. dbRES provides comprehensive annotations and data summaries, including (but not limited to) transcript sequences, RNA editing types, editing site locations, amino acid changes, organisms, subcellular organelles (if available), cited references, etc. A user-friendly web interface is developed to facilitate both retrieving data and online display of RNA edit site information.

INTRODUCTION

RNA editing refers to post-transcriptional modifications of RNA molecules, and represents a class of mechanisms that contribute to the complexity of the transcriptome (1). RNA editing occurs in the nucleus, as well as in mitochondria and plastids (2). To date, these modifications have been observed in plants, animals, fungi and virus (2–6). The diversity of this widespread phenomenon includes nucleotide modifications, nucleotide deletions and insertions, either in coding or non-coding region of RNA, which can occur concomitantly with transcription and splicing processes (2–4,7,8).

Two public databases, EdRNA (http://edrna.mbc.nctu.edu.tw 140.113.239.182/%7Emot/index.php) and Editing Sites Database (9), store putative RNA editing sites that have been predicted using computational methods. While these resources have greatly contributed to the study of RNA editing, the ever-increasing availability of experimental RNA editing data remains scattered under a variety of diverse formats and sources. Availability of standard collected high-quality data is important to design novel computational approaches for identifying RNA editing sites on transcripts. We present dbRES, a collection of experimentally verified RNA editing-sites. dbRES is manually curated from primary literature and annotations in GenBank. To our knowledge, dbRES is the first database containing kinds of up-to-date experimentally reported RNA editing sites.

DATABASE CONSTRUCTION

Data collection in dbRES is based on a three-step strategy. First, the latest version of GenBank (Release 154) (10) is downloaded as the data sources, and entries containing the keyword 'RNA editing' are obtained as subset 1. Second, the GenBank accession numbers of the edited transcripts in each bibliographical reference are used to retrieve the sequences from GenBank data sources as subset 2. Finally, for some RNA editing sites with only direct support in the literature, the details are manually curated as subset 3. Once this data has been collected, subsets 1, 2 and 3 (including sites, sequences and annotations) are manually curated to ensure consistency of the data and eliminate redundancy.

Data storage of dbRES is based on MySQL database system. All the data, including editing sites, annotations and sequences, are integrated in a single complex-table with several external links to text files. Because data accumulates continually, data archiving in dbRES is an on-going process that includes automatic, periodic updates to the central MySQL database.

DATABASE CONTENTS

dbRES currently contains 5437 RNA editing events for 251 transcripts covering 96 organisms, including plants, metazoa, protozoa, fungi and virus. Every RNA editing

*To whom correspondence should be addressed: Tel: +86 10 62794295; Fax: +86 10 62794295; Email: daulyd@tsinghua.edu.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

event is categorized as one of 16 different types (e.g. C to U, U to C, A to I). Table 1 shows the distribution of data from dbRES across these different categories of edits and distribution of data across different species is given in Supplementary Table 1.

Neighboring nucleotide preferences for edits of the three main types (C to U, U to C and A to I) are systematically analyzed based on the nucleotide frequency at positions −2, −1, 0, 1 and 2 relative to each editing site. This analysis employs statistical methods that are standard for the RNA editing field (11–13). Figure 1 presents results of this analysis (for the contents of dbRES version 1.1) in the form of sequence logos.

WEB INTERFACE

Data browser

The dbRES database is freely available at http://bioinfo.au.tsinghua.edu.cn/dbRES. The browser function of the dbRES web site supports browsing by species and browsing by editing type. The browser organizes information about species according to a NCBI taxonomy tree (10,14). Users can expand the tree level-by-level. At the deepest level, gene names corresponding to the species are shown as links to the database viewer (described in the next section). The upper part of Figure 2 shows an example of these browsing functions.

Table 1. Data distribution across different RNA editing types

RNA edit type	Count	Percentage (%)
C to U RNA editing	4539	83.48
A to G RNA editing	107	1.97
U to C RNA editing	702	12.91
G insertion RNA editing	1	0.02
C insertion RNA editing	58	1.07
G deletion RNA editing	1	0.02
A to C RNA editing	2	0.04
G to C RNA editing	3	0.06
G to A RNA editing	8	0.15
U insertion RNA editing	2	0.04
A insertion RNA editing	3	0.06
C/U insertion RNA editing	4	0.07
AA insertion RNA editing	3	0.06
A to U RNA editing	1	0.02
GG to AA RNA editing	2	0.04
AA to GG RNA editing	1	0.02

Data viewers

The dbRES web site allows several options for viewing data. One useful viewing mode is by individual RNA editing site. While viewing data for individual sites, information about the corresponding gene is displayed, including related information from GenBank, and links to related external pages (e.g. the associated GenBank page, or MiRBase in the case of MicroRNAs). Information about the original sequence, organism and subcellular organelle (where appropriate) is also displayed. For RNA editing events that have been reported in a PubMed-indexed publication, links to the PubMed abstract are provided. Another useful viewing mode is by transcripts, and in this mode each associated editing event is indicated along with links to display those individual events. The translated amino acid sequences of the edited and their genomic sequences can be viewed when the editing site locates in CDS region. Examples of these data viewing functions can be found in Figure 2.

Search engine

Several options are provided for searching dbRES. The web-interface for the dbRES search engine is displayed in Figure 3. The searching functions of dbRES allow users to retrieve data by querying three fields: gene name, organism and RNA editing type. This search function assists users to retrieve only the data in which they are interested. dbRES can also be searched by using a querying sequence. This function is implemented by running the BLAST program (15) on the dbRES server, with the result automatically parsed to generate a list of dbRES accession numbers. The protocol and the interface for the sequence query function are also shown in Figure 3.

Database download and mirror service

For the convenience of users, we provide dbRES for downloading in two formats: an annotated flat text package and an SQL script package. In addition to these packages, dbRES is also provided the form of a free mirror-service. The mirror-service package contains all web site elements and a set of data processing programs (as PHP scripts) with detailed instructions for setting up the mirror site.

CONCLUSIONS AND FUTURE WORK

The dbRES database has been developed to fill an existing gap in the availability of consistent and high-quality data

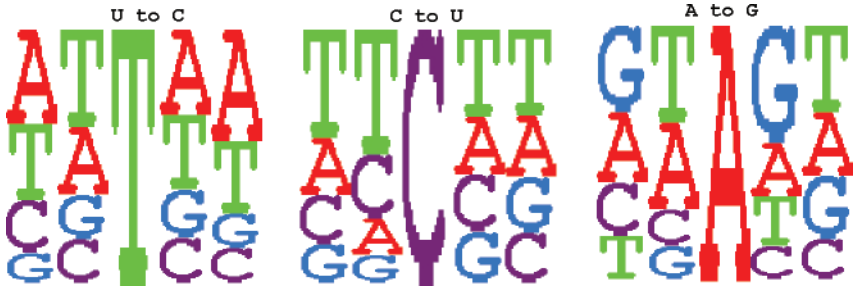

Figure 1. Sequence logo of the nucleotide context of the RNA editing sites and flanking regions.

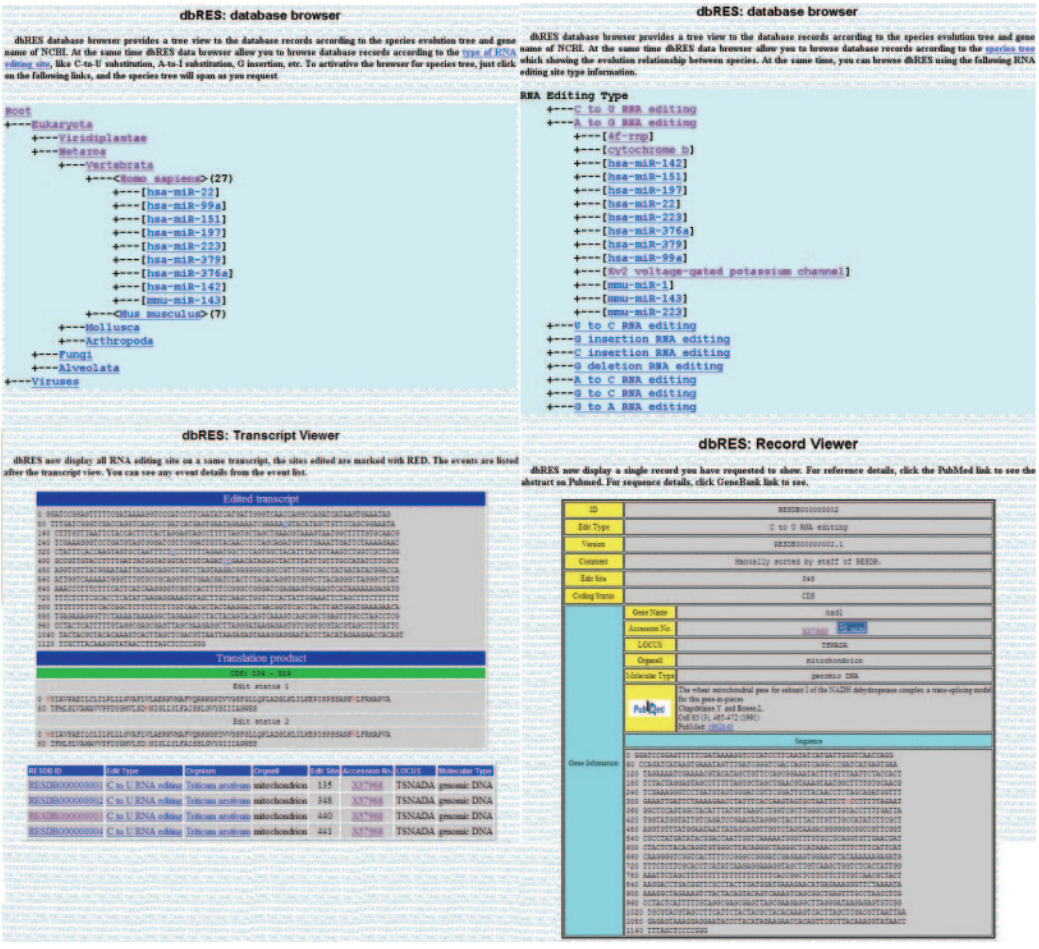

Figure 2. Example of browsing and data viewing functions.

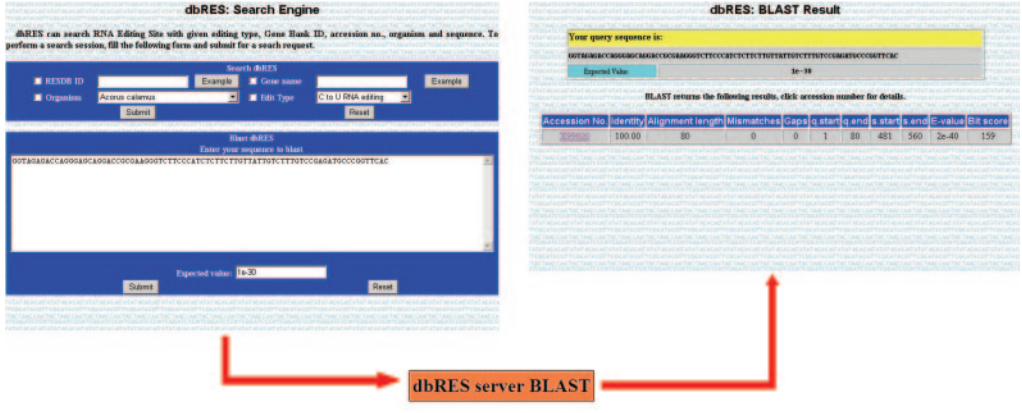

Figure 3. Example of search engine.

collection to use in training and evaluating computational methods for discovering RNA editing sites. The collection described here contains 5437 experimental RNA editing sites identified in plants, animals, fungi and virus. In the next release, we plan to increase the number of A to I editing sites and other types of newly reported editing sites.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Fei Li and Jin Gu for his critical reading and suggestion of the manuscript. We thank Andrew Smith

and Allison Blum for their kind help to carefully copy-edit the manuscript. This work is supported partially by The National Basic Research Program (2004CB518605) and National Science Foundation of China (60405001 and 60572086). Funding to pay the Open Access publication charges for this article was provided by National Science Foundation of China.

Conflict of interest statement. None declared.

REFERENCES

1. Gott,J.M. (2003) Expanding genome capacity via RNA editing. *C. R. Biol.*, **326**, 901–908.
2. Bass,B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.*, **71**, 817–846.
3. Gerber,A.P. and Keller,W. (2001) RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.*, **26**, 376–384.
4. Samuel,C.E. (2003) RNA editing minireview series. *J. Biol. Chem.*, **278**, 1389–1390.
5. Kobayashi,M., Shinohara,M., Sakoh,C., Kataoka,M. and Shimizu,S. (1998) Lactone-ring-cleaving enzyme: genetic analysis, novel RNA editing, and evolutionary implications. *Proc. Natl Acad. Sci. USA*, **95**, 12787–12792.
6. Jayan,G.C. (2004) RNA editing in hepatitis delta virus: unsolved puzzles. *Sci. World J.*, **4**, 628–637.
7. Blanc,V. and Davidson,N.O. (2003) C-to-U RNA editing: mechanisms leading to genetic diversity. *J. Biol. Chem.*, **278**, 1395–1398.
8. Maas,S., Rich,A. and Nishikura,K. (2003) A-to-I RNA editing: recent news and residual mysteries. *J. Biol. Chem.*, **278**, 1391–1394.
9. Levanon,E.Y., Eisenberg,E., Yelin,R., Nemzer,S., Hallegger,M., Shemesh,R., Fligelman,Z.Y., Shoshan,A., Pollock,S.R., Sztybel,D. *et al.* (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, **22**, 1001–1005.
10. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2006) GenBank. *Nucleic Acids Res.*, **34**, D16–20.
11. Liu,T. and Bundschuh,R. (2005) Model for codon position bias in RNA editing. *Phys. Rev. Lett.*, **95**, 088101.
12. Athanasiadis,A., Rich,A. and Maas,S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, **2**, e391.
13. Kim,D.D., Kim,T.T., Walsh,T., Kobayashi,Y., Matise,T.C., Buyske,S. and Gabriel,A. (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.*, **14**, 1719–1725.
14. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **34**, D173–180.
15. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.