REDIdb: an upgraded bioinformatics resource for organellar RNA editing

Picardi E[1], Regina TMR[2], Verbitskiy D[3], Brennicke A[3], Quagliariello C[2]

Motivation

RNA editing is a post-transcriptional molecular process whereby a genetic message is modified from the corresponding DNA template by means of substitutions, insertions and/or deletions [1]. It occurs mostly in organelles by clade-specific diverse and unrelated biochemical mechanisms [2, 3]. Editing sites are currently annotated in primary databases such as GenBank, even though a specific descriptor for RNA editing is not one of the features of the actual flat-files. As a consequence, editing modifications are annotated under "misc_feature" or in the notes indicating the existence of RNA editing alterations, but omitting important details such as the editing type or the nucleotide positions affected. A systematic study of known and experimentally determined RNA editing events is, thus, time-consuming and sometimes precluded by annotation errors in Genebank flat-files. With the aim to improve the annotation status of RNA editing sites, we developed a specialized database, called REDIdb, devoted to store and give access to editing events in organelles and especially in mitochondria [4]. Here we present an updated and upgraded release of REDIdb with a web-interface refurbished with graphical and computational facilities that improve RNA editing investigations. REDIdb is freely queried at http://biologia.unical.it/py_script/REDIdb/.

Methods

All editing events annotated in REDIdb were extracted from GenBank through a semi-automated methodology based on custom python scripts. Potential inconsistencies mainly due to human errors were corrected by hand to the best of our knowledge taking into account the corresponding literature or contacting the original depositors. REDIdb has been developed on a platform of the Apache web server in Windows 2003 and uses the MySQL4.0 database management system. All facilities to store and query the database are written in Python scripting language (version 2.5.1) making use of two non-standard libraries such as MySQL-python (version 1.2.3) and Biopython (version 1.50). The web application is also in python v2.5 and each script is called through the CGI (common gateway interface) methodology.

[1] Dipartimento di Biochimica e Biologia Molecolare "E. Quagliariello", Università di Bari, Bari [2] Dipartimento di Biologia Cellulare, Università della Calabria, Arcavacata di Rende [3] Molekulare Botanik, Universitat Ulm, Ulm

Results

Actually REDIdb stores 10,600 editing positions distributed over 828 organellar sequences. The 65% of all events are located in the mitochondrion and about the 62% are C-to-U conversions. Editing sites are annotated in their biological context and organized in GenBank-like flat-files. Different search levels can be performed through an appropriate search page according to the user relevant information including the gene or organism name, the cellular location and the molecular source in addition to the editing type and supplementary details. In the upgraded REDIdb version the user can display the content of an individual entry in different ways (Fasta, flat-file plain text, alignment overview and so on) or perform simple analyses such as a Blast search or a multiple-alignment. In annotated protein coding genes, the user can visualize known domains and all editing sites in such domains. Moreover, REDIdb offers unique graphical visualizations that can help the discovery and annotation of new sites, representing an invaluable specialized resource to study the process of RNA editing in organelles.

Contact e-mail

e.picardi@biologia.uniba.it

Supplementary information

References

[1] Gott, J.M. and Emeson, R.B., 2000, Annu Rev Genet 34, 499-531. [2] Gray, M.W., 2003, IUBMB Life 55, 227-33. [3] Brennicke, A., 2009, IUBMB Life. 61, 1105-9. [4] Picardi, E. et al., 2007, Nucleic Acids Res 35, D173-7.