186–187 *Nucleic Acids Research, 1999, Vol. 27, No. 1*

The viroid and viroid-like RNA database

Daniel A. Lafontaine, Patrick Deschênes, Frédéric Bussière, Véronique Poisson and Jean-Pierre Perreault*

Département de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec J1H 5N4, Canada

Received August 28, 1998; Accepted September 1, 1998

ABSTRACT

This is an online database to facilitate research on viroid, viroid-like RNAs and human hepatitis *delta* virus (vHDV) by presenting a large number of sequences and related data in a comprehensive and user-friendly format (e.g. position of their self-catalytic domains, open reading frame of the vHDV, prediction of the most stable secondary structures, etc.). Most of these RNA species share a common proposed replication pattern known as a DNA-independent rolling circle mechanism. Together, these species form the 'brotherhood' of the smallest known auto-replicable RNAs. This online database is available on the World Wide Web at http://www.callisto.si.usherb.ca/~jpperra

INTRODUCTION

The initial, 1996, version of this database focused on sequences from viroids, which are small, circular, single-stranded RNA infecting plants, and plant satellite viroid-like RNAs (1). In the 1997 update, we included a section on the human hepatitis *delta* virus (vHDV) (2). The 1998 version was reorganized to be more user-friendly and easier to access the database (3). The four sections (viroids, satellite RNAs, HDV and others) of the databank are shown in a frame which is always available and allows easy access to any subdivision (Table 1). In addition to the update of the existing database, the 1999 version offers several novel features as described below.

DESCRIPTION

The choice of a section (Table 1) will lead to the second level which is a summary table for each sequence subdivision. The viroids, satellites RNAs and other RNAs are divided according to the nature of the compilated RNA species. The choice of either a viroid or viroid-like RNA species in a summary level will lead to the third level. Briefly, each species of RNA appearing in the database is listed by its complete name and number of sequence variants. This is followed, for each species, by a complete listing of the sequence variants and their assigned nomenclature. The identification of species variants is based on its usual acronym followed by a number. The procedure for sequence identification and information compilated was presented previously (1). Additional data of each entry include their accession numbers in sequence library file servers, bank loci (when available), number of nucleotides (total and by type), complete publication information, and the sequence in blocks of 10 nucleotides. For several sequences, a table shows structural features, for example the position of the conserved sequences of the self-catalytic domains (hammerhead, hairpin, delta and VS RNA). In addition, a secondary structure prediction of the most likely ancestral variant of each entry (except for vHDV and VS RNA) was derived using the RNAfold structure prediction package. These predicted structures are appended to the database in connect file format to allow easy manipulation (fourth level).

Unlike the viroid section, the vHDV section comprises all available sequences, irrespective of their completeness, since the majority of the partial sequences are informative (i.e. they correspond to the sequence of either the open reading frame or the ribozymes). Both HDV complete and partial sequences were arbitrarily subdivided into several parts in order to accelerate the

Table 1. Summary of the database sections

VIROIDS
· ASBV-type (group A)
· PSTV-type (group B)
·PSTVd subgroup
·CCCVd subgroup
·HSVd subgroup
·ASSVd subgroup
·CbVd subgroup
SATELLITE RNAs
· Luteovirus
· Nepovirus
· Sobemovirus
· Other satellites
View sequence alignments
vHDV RNA SEQUENCES
· Complete genome sequences
· Partial related RNA sequences
View sequence alignments
OTHERS

*To whom correspondence should be addressed. Tel: +1 819 564 5310; Fax: +1 819 564 5340; Email: jperre01@courrier.usherb.ca

display. For the partial vHDV sequence, a 'p' preceding the specific number is attributed to partial sequence (2). From the 1998 version of the database, sequence alignments of all complete vHDV nucleotide sequences and all amino acid sequences of the vHDV mRNA antigen have been available through the summary tables (second level). This database provides an excellent reference point for further phylogenetic and structure–function studies of these RNA species.

1999 VERSION

Today, more than 450 sequences are now part of this database comprising data from 25 viroid species, nine species of plant satellite viroid-like RNAs, four related species of RNA and vHDV (23 complete and 165 partial). The database is now used daily by several researchers, and all user comments are truly appreciated by the authors. Clearly, the development of the database depends on input from members of the research community. The modifications and additions performed in this new version are in response to constructive comments; for example: (i) in agreement with the recent review of Flores and collaborators (4), the group B viroids were clustered in five subgroups according to the sequence composing their central conserved regions (CCR) (Table 1); (ii) within the vHDV RNA sequence sections, alignment of all sequences for the self-catalytic domains as well as for the transcriptional promoter are in progress and will be added soon; (iii) within the satellite RNAs section, the alignment of hammerhead self-catalytic domains is in progress and will be added soon. Subsequently, a phylogenetic reconstruction of the viroids and plant satellite RNA including hammerhead sequence will soon be available; etc.

COMPLETENESS, ACCURACY AND AVAILABILITY OF THE DATA

This databank is available on the World Wide Web browser (e.g. Netscape Navigator) at the URL http://www.callisto.si.usherb.ca/~jpperra . Floppy disks (readable on microcomputers operating under MS-DOS or Macintosh) and hard copy version are also available upon request only for those without electronic access to the database. This database is updated as soon as sequences become available. Users of the viroid and viroid-like RNA database should cite this publication and are encouraged to provide corrections, new information or other information for inclusion in the database via electronic mail (jperre01@courrier.usherb.ca). The authors would appreciate being informed of any omitted sequences or errors in the data set. We will correct any errors.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Natural Sciences and Engineering Research Council (NSERC), the Medical Research Council (MRC) of Canada, and in part by a scientific team grant from Fonds pour la Formation des Chercheurs et l'Avancement de la Recherche du Québec (FCAR) to J.-P.P. F.B. is a recipient of predoctoral fellowships from NSERC while D.L., P.D. and V.P. are recipients of the Fonds de la Recherche en Santé du Québec (FRSQ). J.-P.P. is an MRC scholar.

REFERENCES

1 Bussière,F., Lafontaine,D. and Perreault,J.-P. (1996) *Nucleic Acids Res.*, **24**, 1793–1798.
2 Lafontaine,D., Mercure,S. and Perreault,J.-P. (1997) *Nucleic Acids Res.*, **25**, 123–125.
3 Lafontaine,D.A., Mercure,S., Poisson,V. and Perreault,J.-P. (1998) *Nucleic Acids Res.*, **26**, 190–191.
4 Flores,R., Di Serio,F. and Hernandez,C. (1997) *Seminars Virol.*, **8**, 65–73.