# The viroid and viroid-like RNA database

**Daniel A. Lafontaine, Stéphane Mercure, Véronique Poisson and Jean-Pierre Perreault***

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

## ABSTRACT

**The viroid and viroid-like RNA database is a compilation of all natural sequences published in journals or available from the GenBank and EMBL nucleotide sequence libraries. Several information regarding these RNA species such as the position of their self-catalytic domains and the open reading frame of the human hepatitits *delta* virus are provided. The database also includes a determination of the likely ancestral sequence of most species and a prediction of the most stable secondary structures of these sequences. This online database is available on the World Wide Web (http://www.callisto.si.usherb.ca/~jpperra ). It should provide an excellent reference point for further phylogenetic and structure–function studies of these RNA species.**

## DESCRIPTION

We have developed an online database in order to facilitate viroid, viroid-like RNAs and human hepatitis *delta* virus (vHDV) research by presenting a large number of sequence and related data in a comprehensive and user-friendly format. 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. The initial 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 a subsequent update, we included a section focusing on the human hepatitis *delta* virus (vHDV) (2). Unlike the viroid section, the vHDV section comprises all available sequences, irrespective of their completeness, since the majority of the partial sequence are informative (i.e. they correspond to the complete sequence of either the open reading frame or the ribozymes).

In this manuscript, we describe an update of this database which to date includes more than 400 sequences from 36 different species. In addition to the update of the existing database, this new version offers a user-friendly structure which facilitates and accelerates the access to the database. The four sections of the databank are now shown in a frame which is always available and allows easy access to any subdivision:

- •viroids    •satellite RNAs    •HDV    •others

The choice of a section will lead to the second level which is a summary table for each sequence subdivision (Fig. 1). The viroids, satellites RNAs and other RNAs are divided according to the nature of the compiled RNA species. In contrast, the HDV section is subdivided on the basis of sequence completeness (i.e. complete and partial). Both of HDV complete and incomplete sequences were arbitrarily subdivided in two parts in order to accelerate the display.

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 compiled was presented previously (1). For the partial vHDV sequence, a 'p' preceding the specific number is attributed to partial sequence (2). Additional data in each entry includes 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 10 nucleotide blocks. For several sequences, a table lists structural features as the position of the conserved sequences of the self-catalytic domains, completes the entry. 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 (fourth level). The analysis of the viroid and viroid-like RNA compilation (e.g. classification, secondary structure prediction, phylogenetic identification of the likely ancestral variant, etc.) has been published previously (1).

Two sequence alignment files will be also available, in early 1998, through the summary tables (second level). These sequence alignment will be performed using: (i) all complete vHDV nucleotide sequences; and (ii) all amino acid sequences of the vHDV mRNA antigen. These alignments as well as the database itself should provide an excellent reference point for further phylogenetic and structure–function studies of these RNA species.

## AVAILABILITY

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) and in hard copies version are also available upon request for those without electronic access to the database. This database is updated bimonthly. 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 (jp.perre@courrier.usherb.ca).

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

## Compilation of viroids

| RNA species | Abbreviation | Number of sequences | RNA species | Abbreviation | Number of sequences |
|---|---|---|---|---|---|
| **Viroids** | | | | | |
| **ASBVd-type (group A)** | | | | | |
| Avocado sunblotch viroid | ASBVd | 19 | **PSTVd-type (group B)** | | |
| Peach latent mosaic viroid | PLMVd | 6 | ASSVd group (subgroup B2) | | |
| | | | Australian grapevine viroid | AGVd | 1 |
| **PSTVd-type (group B)** | | | Apple dimple fruit viroid | ADFVd | 1 |
| PSTVd group (subgroup B1) | | | Apple scar skin viroid | ASSVd | 4 |
| Coconut cadang-cadang viroid | CCCVd | 7 | Citrus bent leaf viroid | CBLVd | 3 |
| Citrus exocortis viroid | CEVd | 33 | Citrus viroid species III | Cvd-III | 2 |
| Columnea latent viroid | CLVd | 3 | Coleus blumei viroid I | CoVd | 9 |
| Chrysanthemum stunt viroid | CSVd | 4 | Grapevine yellow speckle viroid | GYSVd | 34 |
| Coconut tinangaja viroid | CTiVd | 2 | Grapevine 1B viroid | G1BVd | 1 |
| Citrus viroid species | CVdIV | 1 | Pear blister canker viroid | PBCVd | 1 |
| Hop latent viroid | HLVd | 1 | | | |
| Hop stunt viroid | HSVd | 32 | | | |
| Iresine viroid | IRVd | 1 | | | |
| Mexican papita viroid | MPVd | 9 | | | |
| Potato spindle tuber viroid | PSTVd | 29 | | | |
| Tomato apical stunt viroid | TASVd | 3 | | | |
| Tomato planta macho viroid | TPMVd | 1 | | | |

## Compilation of satellite RNAs

| RNA species | Abbreviation | Number of sequences |
|---|---|---|
| **Satellite RNA** | | |
| **Luteovirus** | | |
| Barley yellow dwarf virus satellite RNA | vBYDV | 1 |
| **Nepovirus** | | |
| Arabis mosaic virus satellite RNA | sARMV | 1 |
| Chicory yellow mottle virus satellite RNA S1 | sCYMV-S1 | 1 |
| Lucerne transient streak virus satellite RNA | vLTSV | 3 |
| Tobacco ringspot virus satellite RNA | sTobRV | 2 |
| **Sobemovirus** | | |
| Subterranean clover mottle virus satellite RNA | vSCMoV | 2 |
| Solanum nodiflorum mottle virus satellite RNA | vSNMV | 1 |
| Velvet tobacco mottle virus satellite RNA | vVTMoV | 2 |

## Compilation of HDV sequences

| RNA species | | Number of sequences |
|---|---|---|
| Complete sequences............................ part I, part II | | 20 |
| Partial sequences................................ part I, part II | | 50 |

## Compilation of other related RNAs

| RNA species | Abbreviation | Number of sequences |
|---|---|---|
| **Other related RNA** | | |
| Carnation stunt associated viroid | rCarSV | 2 |
| Newt satellite 2 transcript | rNS2T | 1 |
| Neurospora VS RNA | VS RNA | 1 |

**Figure 1.** Summary tables of the four sections.

## 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.