Nucleic Acids Research, 1997, Vol. 25, No. 1 123–125

Update of the viroid and viroid-like sequence database: addition of a hepatitis delta virus RNA section

Daniel Lafontaine, Stéphane Mercure and Jean-Pierre Perreault*

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

Received September 19, 1996; Accepted September 20, 1996

ABSTRACT

Recently, we developed and made available an online database that includes all the reported (to our knowledge) viroid and viroid-like RNA sequences [Bussire,F., Lafontaine,D. and Perreault,J.-P. (1996) *Nucleic Acids Res.* 24, 1793–1798]. We report here an update of this catalogue which includes the addition of a new section devoted to human hepatitis delta virus (vHDV) sequences. This new section comprises all available vHDV sequences, irrespective of their completeness, which have been either published or were available from nucleic acid libraries. Additional structural characteristics of the vHDV genome, such as the positions of the self-catalytic domains, the antigen open reading frames, etc., are also included. The catalogue is available on the World Wide Web (http://www.callisto.si.usherb.ca/~jpperra) in a user-friendly form. It should provide an excellent reference point for further molecular studies of these small circular pathogenic RNAs.

INTRODUCTION

We previously reported the development of an online catalogue focusing on small circular RNAs (1). The initial version of this catalogue emphasized the small, single-stranded, circular RNA molecules (246–463 nt) known as viroids, whose infection and disease of crop species has resulted in significant economic losses to the agricultural industry. The catalogue also included a sequence compilation of plant satellite viroid-like RNAs and other related RNAs (e.g. Newt satellite 2 transcript and Carnation stunt associated viroid). We present here an update of this database which includes the addition of a section focusing on the human hepatitis delta virus (vHDV). A partial sequence of the viroid-like domain of vHDV appeared in the earlier database under the heading 'Other related RNAs'. However, in an effort to address the comments of several users of the original online catalogue, and taking into consideration the relationship between viroids and vHDV, we have added a section which includes all available sequences related to vHDV. Only naturally occurring vHDV sequences were considered. The resulting more complete compilation will facilitate further phylogenetic and structure–function studies of viroids and related RNAs.

DESCRIPTION

Upon entering the web site compilation section, you will automatically be greeted by a menu which lists the different accessible divisions of the database:

VIROIDS
· ASBV-type (group A)
· PSTV-type (group B)
· PSTV group (subgroup B1)
· ASSV group (subgroup B2)
SATELLITE RNAs
· Luteovirus
· Nepovirus
· Sobemovirus
vHDV RNA SEQUENCES
· Complete genome sequences
· Partial related RNA sequences
OTHER RELATED RNAs

Briefly, each species of RNA appearing in the catalogue is listed by its complete name and number of sequence variants (Fig. 1). This is followed, for each species, by a complete listing of the sequence variants and their assigned nomenclature in the database. The identification of species variants is based on its usual acronym followed by a number. The latter is function of the date of report. For sequences already published or reported in online libraries, priority was given to publication date over library submission date. When more than one sequence was reported simultaneously, we attributed arbitrary numbers to each entry. Exceptionally, for hop stunt viroid (HSVd) variants, a letter corresponding to the host from which the isolate was obtained, precedes the number. For example, HSVd.h1 refers to a hop stunt viroid isolated from the hop. We believe that the proposed nomenclature has facilitated and clarified the identification of viroid sequences (1). Additional data in 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 10 nucleotides blocks. 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 Mulfold structure prediction package (1). These predictions are appended to the catalogue (in connect file format). The analysis of the viroid and viroid-like RNA sections (e.g. classification, secondary structure

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

Hepatitis delta virus RNA (human), 17 variants (complete sequences)

>vHDV.1
accession numbers:X04451(embl), 59963(gi)
Locus HEPHDVXX
1679 nucleotides (334A, 518C, 490G, 337T)
Wang,K.S., Choo,Q.L., Weiner,A.J., Ou,J.H., Najarian,R.C., Thayer,R.M., Mullenbach,G.T., Denniston,K.J., Gerin,J.L. and Houghton,M.
Structure, sequence and expression of the hepatitis delta viral genome. *Nature* **323**, 508-514 (1986)
[published erratum appears in *Nature* 1987 Jul 30-Aug **5**;328,456]
see also: *Nature***323**, 558-560 (1986); *Eur. J. Biochem.***217**, 29-36 (1993)

NOTES:
1) The 'g' missing from the Genbank sequence (erratum appears in *Nature* 1987 Jul 30-Aug **5**;328,456) at position 1113, was added
2) Conflict at position 191, 'g' in Genbank is 't' in *Nature* **323**, 558-560 (1986)
3) Conflict in microheterogeneity, position 589 in original paper is position 590 in Genbank
4) Variations noted between cDNA clones sequenced in order to obtain the complete HDV genome (from Genbank):

● position 264: c is t in variant clone	● position 414: t is c in variant clone
● position 488: c is t in variant clone	● position 553: g is a in variant clone
● position 590: c is t in variant clone	● position 603: t is c in variant clone
● position 653: g is a in variant clone	● position 987: c is t in variant clone
● position 1012: c is t in variant clone	● position 1024: c is t in variant clone
● position 1141: a is t in variant clone	● position 1473: t is c in variant clone
● position 1566: t is c in variant clone	● position 1677: a is g in variant clone

4) Positions 1232-1240: potential glycosylation site
5) Origin of strain: Serial passage in chimpanzee
6) Also known as 'Italy' variant

```
CTTGAGCCAA GTTCCGAGCG AGGAGACGCG GGGGGAGGAT CAGCTCCCGA GAGGGGATGT CACGGTAAAG AGCATTGGAA
CGTCGGAGAA ACTACTCCCA AGAAGCAAAG AGAGGTCTCA GGAAGCGGAC GAGATCCCCA CAACGCCGGA GAATCTCTGG
AAGGGGAAAG AGGAAGGTGG AAGAAAAAGG GGCGGGCCTC CCGATCCGAG GGGCCCAACC TCCAGATCTG ...
```

Structural informations

Genome	**HDV antigen**
● genomic polarity ● complete sequence	● start codon position : 1598 ● stop codon position : 954 ● predicted RNA-edit site : 1012 ● RNA-edit codon : codon tryptophan
(+) polarity-> Delta ribozyme	**(-) polarity-> Delta ribozyme**
● structural domain (686-769) ● cleavage site (685-686)	● structural domain (817-900) ● cleavage site (900-901)

Figure 1. Example of a vHDV sequence entry. For the purpose of the representation, the nucleotide sequence was shortened to allow a complete display of all available information relative to the entry.

prediction, phylogenetic identification of the likely ancestral variant, etc.) has been the object of a previous report (1).

In the present update, several sequence variants of viroids have been added to the catalogue. Furthermore, we have included the sequence of a transcript from a fungal mitochondrial plasmid (*Neurospora* VS RNA) that possesses self-cleavage ability (2) in the 'Other related RNAs' section. For sequences which include self-catalytic domains (hammerhead, hairpin, delta and VS), the localization of the conserved sequences required for cleavage of both plus and minus polarity forms (when present) are indicated. More than 250 sequences are now part of this database (up from 182 in the previous edition) comprising data from 21 viroid species, eight species of plant satellite viroid-like RNAs, three related species of RNA and vHDV (54 sequences). This catalogue comprises all sequences that, to our knowledge, have been published or were available from the sequence library file servers.

vHDV SECTION

The vHDV RNA genome shares several features with that of viroids even though it is unique among animal RNA viruses (3). It is a small (1.7 kb), circular RNA molecule which folds into a stable rod-like secondary structure and replicates via a rolling circle mechanism involving self-catalytic conserved motifs (e.g. delta ribozyme) (3). The vHDV left-terminal domain includes both the genomic and antigenomic self-catalytic motifs, and has been proposed to be a viroid-like domain (Fig. 2). Unlike viroids which do not exhibit any translational capability, vHDV produces two related forms of a single antigen (Fig. 2): a smaller one [HDAg(S)] of 195 aa, and a larger one [HDAg(L)] of 214 aa which arises due to an edition event changing the stop codon of HDAg(S) to a tryptophan codon (4). The translational start site,

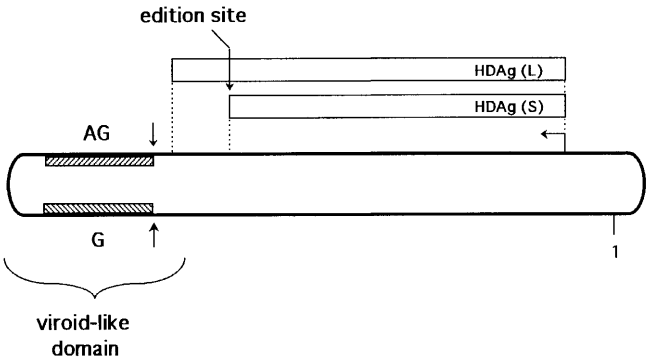

Figure 2. Schematic representation of a vHDV RNA genome. Nucleotide position 1 is based on conventional nomenclature. The viroid-like domain located at the left includes both the genomic (**G**) and the antigenomic (**AG**) self-catalytic motifs (dashed rectangles). Arrows indicate self-cleavage sites. Genomic location of the two open reading frames encoding the large and small antigens are shown as open rectangles.

as well as most of the open reading frame including the edition site, are located outside the viroid-like domain.

The vHDV section includes 17 complete sequences, three of which are new entries created directly from original papers. In contrast to the viroid and viroid-like satellite RNA sections, where only complete sequences were compilated (1), the vHDV section also includes partial sequences. Thirty-seven partial sequences are included in the catalogue, 21 of which were not present at the NCBI site and had to be created directly from the original papers. Among the partial sequences, 35 include either a part or all of the open reading frame. Furthermore, 11 partial sequences include at least either the genomic or antigenomic self-cleavage domain, and four of them include both self-catalytic domains.

For each vHDV sequence variant entry, identification was performed as all other species, except that a 'p' precedes the specific number attributed to partial sequence. This is followed by the usual presentation as shown in Figure 1. A table listing the structural features relative to each sequence completes all entries. This table includes information such as genomic or antigenomic polarity of the sequence and, in the case of partial sequences, the positions of both the start and the stop nucleotides for the entry. To simplify the use or study of partial sequences, we attempted to align each of them with complete vHDV RNA genomes in order to establish their genomic location. In most cases, we looked for proper alignment between partial sequences and representatives of either genotype I (M28267; or vHDV.3), II (X60193; vHDV.9) or III (L22063; vHDV.12) which have been previously described (5). The genome template used for each alignment, and the positions of the partial sequence relative to it, are indicated. We also determined, when appropriate, the positions of the start (first nt of the codon) and stop codons (last nt of the codon) for the large form of the hepatitis delta antigen. Moreover, when possible the predicted RNA edition site was located taking into account the following criteria: (i) the stop codon of HDAg(L) antigen is 57 nt (19 aa) upstream (genomic polarity) of the HDAg(S) stop codon; and (ii) the edition event leads to the synthesis of the HDAg(L). The identity of the codon appearing at the edition site in the sequence entry appears under 'RNA-edit codon' in the table featuring the structural characteristics of vHDV sequences. This table also includes, when present, the position of both the genomic and antigenomic delta self-catalytic motifs and their respective cleavage sites.

A number of the HDV sequences reported in the database were originally obtained from overlapping cDNA clones. In several cases, the authors noted microheterogeneity between the clones required to build a complete sequence. These microheterogeneities and their positions were included in our database under 'Notes' since they are a source of information on sequence variability among HDV populations. In addition, the 'Notes' section includes additional information intended to: (i) help users trace the origin of new sequences (ex. >vHDV.p17; 'this sequence was created from Fig. 3B of an original paper, patient #1'); (ii) notify users of conflict observed between nucleotide sequences found in nucleic acid databases and those used in certain papers for comparison or alignment (vHDV.p3); (iii) emphasize the relationship between sequences present in our catalogue and nomenclature used in the literature for these same isolates (ex. vHDV.3 'is also known as US-1'); and (iv) correct published erratum that were not reported to nucleic acid databases (vHDV.1).

COMPLETENESS, ACCURACY AND AVAILABILITY OF THE DATA

The authors would appreciate being informed of any omitted sequences or errors in the data set. We intend to correct any such errors in the future. This catalogue will be updated as additional sequences become available. In future catalogue updates, no classification priority will be attributed to publication. This compilation is available on the World Wide Web (http://www.callisto.si.usherb.ca/~jpperra). Suggestions for improvement of the database and sequence submission, in a format compatible with the example depicted in Figure 1, are welcome via electronic mail (jp.perre@courrier.usherb.ca). The viroid and viroid-like RNA compilation is also available on floppy disk (readable on microcomputers operating under MS-DOS) and in hard copy (paper). Please refer to this article if your research is assisted by the viroid and viroid-like RNA catalogue.

ACKNOWLEDGEMENTS

The authors would like to thank Dr B. Cousineau, S. Kofalvi, R. Owens, R. Singh, R. Symons and all researchers who reviewed the database and provided us with helpful suggestions. This work was supported in part by a grant from the Medical Research Council (MRC) of Canada to J.-P. P., and in part by a scientific team grant from Fonds pour la Formation des Chercheurs et l'Avancement de la Recherche du Qubec (FCAR). The two first authors have contributed equally. D.L. is recipient of a predoctoral fellowships from FCAR. J.-P. P. holds a scholarship from the Medical Research Council of Canada.

REFERENCES

1 Bussire,F., Lafontaine,D. and Perreault,J.-P. (1996) *Nucleic Acids Res.* **24**, 1793–1798.
2 Saville,B.J. and Collins,R.A. (1990) *Cell* **61**, 685–696.
3 Branch,A.D., Levine,B.J. and Robertson,H.D. (1990) *Semin. Virol.* **1**, 143–152.
4 Lai,M.M.C. (1995) *Annu. Rev. Biochem.* **64**, 259–286.
5 Casey,J.L., Brown,T.L., Colan,E.J., Wignall,F.S. and Gerin,J.L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9016–9020.