J Mol Evol (2000) 50:98–102
DOI: 10.1007/s002399910011

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

Testing the Extent of Sequence Similarity Among Viroids, Satellite RNAs, and Hepatitis Delta Virus

Gareth M. Jenkins,[2] Christopher H. Woelk,[2] Andrew Rambaut,[2] Edward C. Holmes[1]

[1]Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received: 30 April 1999 / Accepted: 24 August 1999

Abstract. A Monte Carlo method was used to test the extent of sequence similarity among viroids, satellite RNAs, and hepatitis delta virus. This analysis revealed that there is insufficient sequence similarity among these pathogens to support the hypothesis that they have a common evolutionary origin. Furthermore, while definite patterns of sequence similarity were observed among some viroids, there was a clear lack of overall similarity, indicating that a monophyletic origin for even this group cannot be reliably supported from sequence data alone.

Key words: Viroids — Satellite RNAs — Hepatitis delta virus — Monte Carlo simulation — Sequence similarity

Viroids, satellite RNAs, and hepatitis delta virus (HDV) are subviral pathogens which have a number of common features including single-stranded RNA genomes, a high GC content, RNA-to-RNA rolling circle replication, and, in many cases, ribozyme activity (reviewed by Branch et al. 1990). Viroids and satellite RNAs are small (200–400 nucleotide) plant pathogens, both of which lack open reading frames. However, while viroids are unencapsidated, circular, and replicate autonomously, satellite RNAs are encapsidated, may be circular or linear, and require the assistance of a nonhomologous helper virus. In contrast, HDV is a human pathogen and the only known animal agent to show similarities with viroids and satellite RNAs. Like most viroids, replication in HDV is nuclear and likely to involve host RNA polymerase II, but as with satellite RNAs, HDV is encapsidated and requires the assistance of an unrelated helper virus, in this case hepatitis B virus. An important difference between HDV and the plant viroid and satellite RNAs is that the HDV genome is much larger (1.7 kb) and can be divided into a viroid-like RNA region and a protein coding region which produces the delta antigen protein (HDAg). Interestingly, HDAg may have been acquired through capture of a cellular mRNA transcript by a viroid-like RNA (Branch et al. 1989), a hypothesis supported by the identification of a cellular homolog of HDAg (Brazas and Ganem 1996).

Despite their importance as agents of disease, the evolutionary origins of these pathogens are uncertain (reviewed in Diener 1996). One hypothesis is that HDV, viroids, and satellite RNAs are all ancient relics of precellular evolution (Diener 1989). Alternatively, it may be that they have a much more recent origin, arising from cellular "signal" RNAs (Zimmern 1982). Other suggestions include an evolutionary relationship with introns (Diener 1981; Dinter-Gottlieb 1986) or to retroviruses and transposable elements (Kiefer et al. 1983). Most in-

Correspondence to: Dr. E.C. Holmes; *e-mail:* Edward.Holmes@zoo.ox.ac.uk

Table 1. List of sequences used in the analysis

Group	Agent	Abbreviation[a]	Accession No.
Viroids: ASBVd family	Avocado sunblotch	ASBVd	J02020
	Peach latent mosaic	PLMVd	M83545
	Chrysanthemum chlorotic mottle	CChMVd	Y14700
Viroids: PSTVd family			
PSTVd genus	Citrus exocortis	CEVd	J02053
	Columnea latent	CLVd	X15663
	Chrysanthemum stunt	CSVd	M19505
	Citrus viroid species II	CVd-II	X69519
	Iresine viroid	IRVd	X95734
	Mexican papita	MPVd	L78454
	Potato spindle tuber	PSTVd	J02287
	Tomato apical stunt	TASVd	K00818
	Tomato planta macho	TPMVd	K00817
CCCVd genus	Coconut cadang–cadang	CCCVd	J02049
	Coconut tinangaja	CTiVd	M20731
	Citrus viroid species	CVd-IV	X14638
	Hop latent	HLVd	X07397
ASSVd genus	Australian grapevine	AGVd	X17101
	Apple dimple fruit	ADFVd	X99487
	Apple scar skin	ASSVd	M36646
	Citrus bent leaf	CBLVd	M74065
	Citrus viroid species III	CVd-III	S76452
	Grapevine yellow speckle-1	GYSVd-1	X06904
	Grapevine yellow speckle-2	GYSVd-2	J04348
	Grapevine 1B	G1BVd	325408
	Pear blister canker	PBCVd	S46812
HSVd genus	Hop stunt	HSVd	X00009
CbVd genus	Coleus blumei-1	CbVd-1	X52960
	Coleus blumei-2	CbVd-2	X95365
	Coleus blumei-3	CbVd-3	X95364
Satellite RNAs	Barley yellow dwarf	vBYDV	M63666
	Arabis mosaic	sARMV	M21212
	Chicory yellow mottle S1	sCYMV-S1	221226
	Lucerne transient streak	vLTSV	X01985
	Tobacco ringspot	sTobRV	M14879
	Subterranean clover mottle	vSCMoV	M33001
	Solanum nodiflorum mottle	vSNMV	J02386
	Velvet tobacco mottle	vVTM0V	NA
Hepatitis delta virus	Hepatitis delta virus	HDV	X04451

[a] Viroid abbreviations contain "Vd," whereas circular satellite RNAs begin with "v" and linear satellite RNAs with "s."

triguing is why so many viroid-like agents have been found in plants, but only one among animals.

The most comprehensive phylogenetic study of viroids, satellite RNAs, and HDV proposed that these organisms have a monophyletic precellular origin (Elena et al. 1991). To assess the validity of this proposal and of relationships among these pathogens in general, we used a Monte Carlo randomization procedure to quantify the level of sequence similarity among them. If viroids, satellite RNAs, and HDV are no more similar than might be expected by chance, then no analysis of their evolutionary origin based on a comparison of sequence data is justified.

A total of 29 complete viroid sequences, 8 complete satellite RNA sequences, and the viroid region of HDV were collected from the compilation of viroid-like sequences available at http://www.callisto.si.usherb.ca/~jpperra (Lafontaine et al. 1999). A list of the sequences included in our analysis and their accession numbers is shown in Table 1. Since a preliminary analysis found the extent of similarity among the viroid-like part of different HDVs to be high (data not shown), only one HDV sequence was used for comparison with the viroids and satellite RNAs. The viroids were grouped according to the classification system described by Klotunow and Rezaian (1989) and Flores et al. (1998). Briefly, two viroid families are recognized: those belonging to the PSTVd family, which contain a central conserved region (CCR) in their genomes; and those of the ASBVd family, which lack such a region. The sequence of the CCR can also be used to split the PSTVd family into subfamilies and the phylogenetic analysis of Elena et al. (1991) further divided subfamilies into genera. Members of the PSTVd, CCCVd, and HSVd genera have identical CCR sequences, whereas the ASSVd and CbVd genera have their own distinctive CCRs.

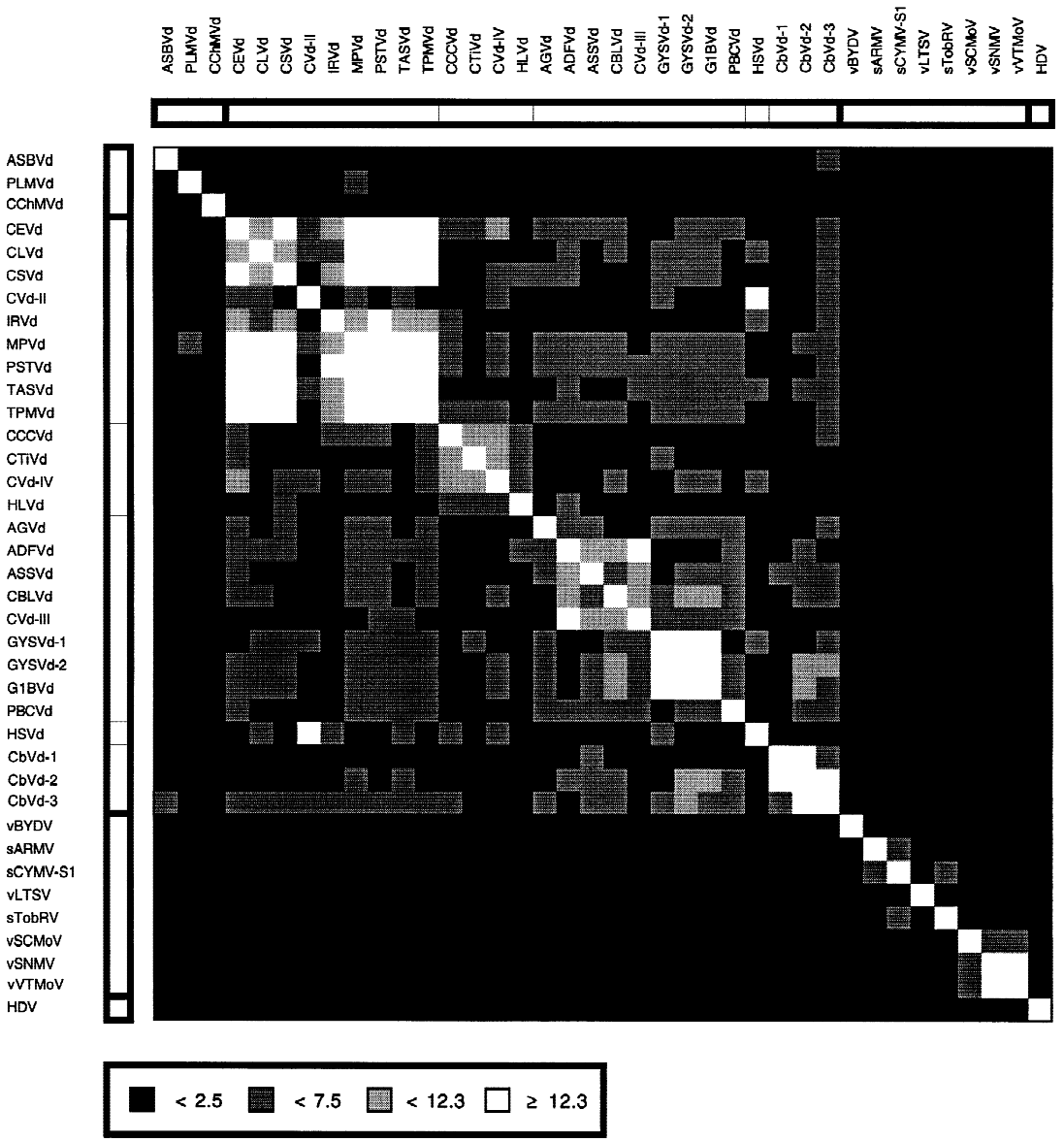

Fig. 1. Results of a Monte Carlo analysis showing the extent of sequence similarity among viroids, satellite RNAs, and HDV. Similarity is quantified as the number of standard deviations (SD) above the random expected similarity score according to the gray scale at the *bottom.* The sequences are listed in the same order as in Table 1. The *thick boxes* divide the sequences into ASBVd-type viroids, PSTVd-type viroids, satellite RNAs, and HDV. Thin discontinuous lines subdivide the PSTVd-type viroids into genera.

The extent of pairwise sequence similarity among viroids, satellite RNAs, and HDV was quantified using the CLUSTALW (version 1.4) sequence alignment program (Thompson et al. 1994). These similarity scores were then compared to those of 200 sets of randomized sequences of the same length and base composition as the reference data. A normal quantile probability analysis confirmed that the pairwise similarity scores among randomized sequences were approximately normally distributed. The similarity among the viroid, satellite RNA, and HDV sequences was then tested against a null hypothesis of random similarity by calculating the number of standard deviations (SD) above the random expected value for each pairwise similarity score. The results of this analysis are shown in a density plot in which different shades of gray depict the range of SD values, with black representing minimal similarity (SD < 2.5) and white maximum similarity (SD 12.3). Given two unrelated sequences, the probability that their similarity score is less than 2.5 SD is 99%. Although very low SD values clearly indicate a lack of support for homology, it is more difficult to assess the exact SD ranges over which there is strong evidence of homology. Also, as the amount of similarity among sequences increases, the dependency among multiple pairwise comparisons may become a problem. For example, if two unrelated sequences are similar by chance, each sequence will also be similar, but nonhomologous, to all close relatives of the other sequence.

The Monte Carlo analysis revealed no significant se-

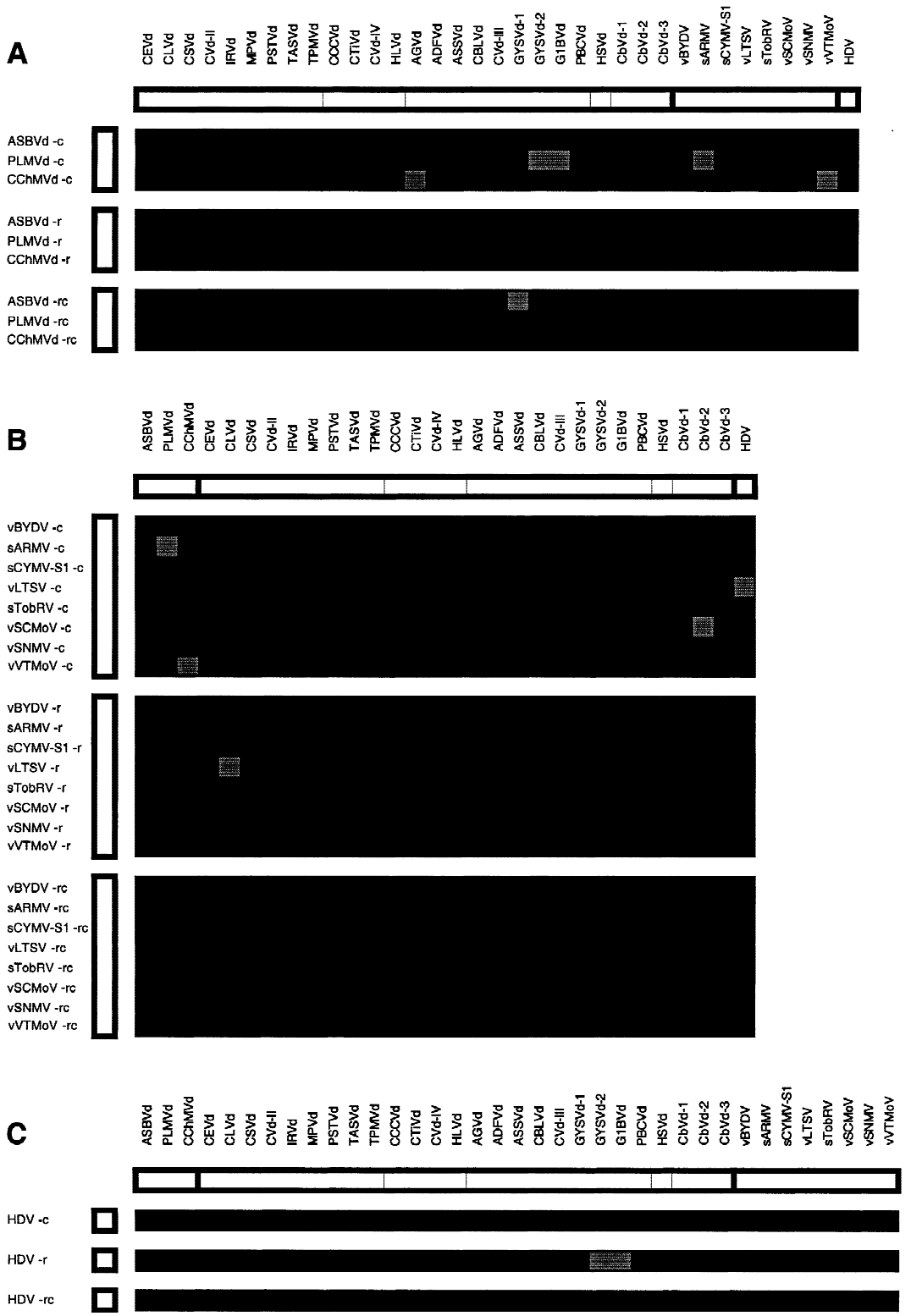

Fig. 2. **A** Sequence similarity among PSTVd-type viroids, satellite RNAs, HDV, and complemented (-c), reversed (-r), and reverse-complemented (-rc) ASBVd-type viroids. **B** Sequence similarity among viroids, HDV, and complemented (-c), reversed (-r), and reverse complemented (-rc) satellite RNAs. **C** Sequence similarity among viroids, satellite RNAs, and complemented (-c), reversed (-r), and reverse-complemented (-rc) HDV. The gray scale is the same as that used as in Fig. 1.

quence similarity among viroids, satellite RNAs, and HDV (Fig. 1). Furthermore, no similarity was found between the two viroid families. In these parts of the density plot, no SD values are in the 7.5 to 12.3 or 12.3 range and only a few values were greater than 2.5. The preponderance of black on the density plot is striking. Nearly all similarity was concentrated within the genera of the PSTVd-type viroids and within the satellite RNAs,

although there are many SD values lower than 2.5 even in these cases. To determine whether any similarity could be found using antigenomic or reversed sequences, the analysis was repeated on sequences which were complemented, reversed, and reverse complemented (Fig. 2). This analysis also failed to detect any significant similarity among HDV, satellite RNAs, and the two viroid families.

While our procedure was sensitive enough to register similarity among most members of the PSTVd family that share a common CCR, which is suggestive of their evolutionary relatedness, there are exceptions. We therefore examined our alignments to verify that the CCRs were correctly aligned. In all cases, CLUSTALW aligned these regions extremely well in the absence of manual editing, and the conserved regions were immediately obvious by visual inspection. The fact that no similarity was registered among certain viroids with identical CCRs therefore more likely reflects the fact that this region is relatively short—two stretches of 15 to 20 nucleotides separated by over 100 nucleotides. This in turn led us to ask whether there were additional highly conserved regions of a similar size in these sequences that had been overlooked by our analysis. However, no such regions were apparent from visual inspection of the sequence alignments. Finally, a BLAST search (http://www.ncbi.nlm.nih.gov/blast/blast.cgi) using the default statistical significance threshold of 0.1 for reporting matching sequences failed to find any viroid or satellite RNA matches to the viroid-like part of HDV (results not shown). A similar search using the eight satellite RNAs likewise did not return any viroid or HDV matches.

Our failure to find any basic sequence similarity among viroids, satellite RNAs, and HDV means that the sequences of these pathogens cannot be aligned in a statistically significant way. This, in turn, undermines any attempt to reconstruct their phylogenetic relationships using sequence data, such as that carried out by Elena et al. (1991). We therefore stress that the phylogenetic analysis of potentially nonhomologous sequences should be avoided as such analysis will produce a tree regardless of whether the sequences actually share common ancestry. In consequence, these sequence data do not support the brotherhood of viroids, satellite RNAs, and HDV.

Acknowledgments. This work was supported by research grants from the Royal Society and the Wellcome Trust. We thank two referees for useful comments on the manuscript.

References

Branch AD, Benenfeld BJ, Baroudy BM, Wells FV, Gerin JL, Robertson HD (1989) An ultraviolet-sensitive RNA structural element in a viroid-like domain of the hepatitis delta virus. Science 243: 649–652

Branch AD, Levine BJ, Robertson HD (1990) The brotherhood of circular RNA pathogens: viroids, circular satellites, and the delta agent. Semin Virol 1:143–152

Brazas R, Ganem D (1996) A cellular homolog of hepatitis delta antigen: Implications for viral replication and evolution. Science 274: 90–94

Diener TO (1981) Are viroids escaped introns? Proc Natl Acad Sci USA 78:5014–5015

Diener TO (1989) Circular RNAs—Relics of precellular evolution. Proc Natl Acad Sci USA 86:9370–9374

Diener TO (1996) Origin and evolution of viroids and viroid-like satellite RNAs. Virus Genes 11:119–131

Dienter-Gottlieb G (1986) Viroids and virusoids are related to group I introns. Proc Natl Acad Sci USA 83:6250–6254

Elena SF Dopazo J, Flores R, Diener TO, Moya A (1991) Phylogeny of viroids, viroidlike satellite RNAs, and viroidlike domain of hepatitis δ virus RNA. Proc Natl Acad Sci USA 88:5631–5634

Flores R, Randles JW, BarJoseph M, Diener TO (1998) A proposed scheme for viroid classification and nomenclature. Arch Virol 143: 623–629

Kiefer MC, Owens RA, Diener TO (1983) Structural similarities between viroids and transposable genetic elements. Proc Natl Acad Sci USA 80:6234–6238

Koltunow AM, Rezaian MA (1989) A scheme for viroid classification. Intervirology 30:194–201

Lafontaine DA, Deschenes P, Bussiere F, Poisson V, Perreault JP (1999) The viroid and viroid-like RNA database. Nucleic Acids Res 27:186–187

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Zimmern D (1982) Do viroids and RNA viruses derive from a system that exchanges genetic information between eukaryotic cells? Trends Biochem Sci 7:205–207