Novel hepatitis D-like agents in vertebrates and invertebrates

Wei-Shan Chang[1], John H.-O. Pettersson[1], Callum Le Lay[1], Mang Shi[1], Nathan Lo[1], Michelle Wille[2], John-Sebastian Eden[1,3], Edward C. Holmes[1]

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia; wcha8651@uni.sydney.edu.au (WSC); mang.shi@sydney.edu.au (MS); edward.holmes@sydney.edu.au (ECH); john.pettersson@gmail.com (JP); nathan.lo@sydney.edu.au (NL)

[2]WHO Collaborating Centre for Reference and Research on Influenza, at The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia; michelle.wille@influenzacentre.org (MW)

[3]Westmead Institute for Medical Research, Centre for Virus Research, Westmead NSW, 2145; Australia; js.eden@sydney.edu.au (JSE);

***** Correspondence: edward.holmes@sydney.edu.au; Tel.: +61 2 9351 5591

Abstract

Hepatitis delta virus (HDV) is the smallest known RNA virus and encodes a single protein. Until recently, HDV had only been identified in humans, where it is strongly associated with co-infection with hepatitis B virus (HBV). However, the recent discovery of HDV-like viruses in metagenomic samples from birds and snakes suggests that this virus has a far longer evolutionary history. Herein, using additional meta-transcriptomic data, we show that highly divergent HDV-like viruses are also present in fish, amphibians and invertebrates. Notably, the novel viruses identified here share HDV-like genomic features such as a small genome size of ~1.7kb in length, circular genomes, and self-complementary, unbranched rod-like structures. Coiled-coil domains, leucine zippers, conserved residues with essential biological functions and isoelectronic points similar to those in the human hepatitis delta virus antigens (HDAgs) were also identified in the putative non-human HDAgs. Notably, none of these novel HDV-like viruses were associated with hepadnavirus infection, supporting the idea that the HDV-HBV association may be specific to humans. Collectively, these data not only broaden our understanding of the diversity and host range of HDV in non-human species, but shed light on its origin and evolutionary history.

Keywords: Hepatitis D virus; evolution; fish; termites; meta-transcriptomics; phylogeny

Introduction

Hepatitis delta virus (HDV), a member in the genus *Deltavirus*, is the smallest known RNA virus and results in chronic or fulminant hepatitis in humans when co-infected with hepatitis B virus (HBV). The HDV genome is a covalently closed, circular, single negative stranded RNA of that forms a viroid-like self-complementary, unbranched rod-like structure of ~1.7kb [1]. An important distinguishing feature of both HDVs and viroids is the use of rolling circle RNA replication [2]. The genome encodes a single protein (Hepatitis Delta Antigen; HDAg) that plays a role in viral packaging.

Notably, human HDV requires an obligatory helper function provided by the HBV envelope protein for assembly, replication and *in vivo* transmission [3]. HDV is estimated to infect 15–20 million people worldwide, and co-infection of HDV and HBV increases the risk of cirrhosis and hepatocellular carcinoma in humans, resulting in higher disease severity and mortality [4] than HBV infection alone. Until recently, HBV-carrier patients were considered the only established hosts for HDV, which has shaped theories of its origin [5-8]. However, this dogma has recently been challenged with the discovery of HDV-like viruses in birds [9] and snakes [10]. Importantly, these viruses were detected in absence of HBV (hepadnavirus) infection. Not only does this raise questions over the relationship between HDV and HBV, but suggests that HDV has a long evolutionary history and originated well before its first appearance in humans [11]. To further explore the origins and evolution of HDVs, we screened for HDV-like circular viruses in ribosomal-RNA (rRNA) depleted cDNA libraries of amphibians, fish, reptiles and termites previously generated [8]. This revealed the presence of four divergent HDV-like circular agents in fish, amphibians and termites, none of which were associated with hepadnavirus infection.

Materials and Methods

RNA library selection, construction and sequencing

Most of the sequence reads used in this study were obtained from previous meta-transcriptomic investigations and are available on the NCBI Sequence Read Archive (SRA) database under the BioProject accessions PRJNA418053, PRJNA314559, PRJ247733 and PRJ355364. Termite libraries available under BioProject accession PRJNA XXXXXnumber (submission pending) were sequenced and constructed as per Shi et al. 2018 [12]. All

libraries were re-screened for HDV-like antigens. The sequence reads of HDV-like agents were uploaded onto SRA under the BioProject submission XXXnumberXXXX (pending submission).

Discovery of hepatitis delta virus-like sequences

Illumina sequencing reads were quality trimmed with Trimmomatic [13] then *de novo* assembled using Trinity version 2.1 [14]. The transcript abundance of all contigs were assessed using the RNA-Seq by the Expectation Maximization (RSEM) method implemented in Trinity version 2.1, and also based on the percentage of raw reads aligned to the virus genome, as described previously [12]. To identify potential HDV-like transcripts while limiting false-positives, the assembled contigs were screened against a custom HDV delta antigen protein database with blast hits (e-values <1E-3) then re-screened against the NCBI non-redundant protein (nr) database with searches using Diamond blastx at an e-value cut-off 1E-5 [15]. Given that the genome size of HDV is expected to be approximately 1.7kb in length, HDV-suspected contigs greater than 1,500 bp and lower than 4,000 bp in length were further examined. According to the blastx results and size selection, putative open reading frames (ORF) of HDV-like contigs were predicted using Geneious R11 (Biomatters, New Zealand). These ORFs were first annotated using a reverse PSI-BLAST [16] search against the conserved domain database (CDD, https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and based on the similarities with previously described HDV genomes. To further validate our HDAg gene predictions, we compared sequences using the protein domain search tools, including HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhblits) [17] (parameter: e-value 1e-3, Minimal coverage of MSA hits 20%) and Phyre2, employing default parameters (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [18]. Lastly, the assembled contigs were checked for circularity by aligning terminal regions to identify any overlap, which were then collapsed to generate a consensus draft genome.

Characterization of novel Hepatitis D-like circular virus and Hepatitis D-like delta antigen

The RNA libraries were mapped against the predicted HDV-associated contigs using BBmap [19] to extract the putative HDV-specific reads and to estimate viral abundance. However, the final genome sequence was obtained by taking the majority consensus from mapping the HDV-specific reads against the circularized contigs in Geneious R11. This mapping tool was

used as it can process circularized reference sequences and align reads that span the termini, thereby confirming circularity.

A diverse set of HDV sequences representing known genotypes were downloaded from GenBank as a reference for comparison. The translated HDAg proteins of these, the recently described bird and snake HDVs, and the putative HDAgs determined here were then aligned using the E-INS-i algorithm in MAFFT v7 [20], after which ambiguously aligned regions were removed using trimAL [21]. This resulted in a final alignment of 203 amino acids. This alignment was also used as the basis for a phylogenetic analysis employing the maximum likelihood (ML) method available within PhyML (version 3.0), assuming the LG substitution model with SPR branch-swapping [22]. Bootstrap resampling (1000 replications) under the same substitution model was used to assess nodal support.

The genomic features of the HDV-like agents were investigated by assessing the GC content, calculated using a sliding window size of 40 nt, as well as the polarity, hydrophobicity and isoelectric point of putative HDAgs estimated in Geneious R11. To determine the circular genome folding into unbranched rod-like structures, circular graphs were constructed using the Mfold webserver [23]. Each coiled-coil region in predicted HDAg ORF was evaluated using MTIDK algorithm on the Marcoil 1.0 webserver [24].

Results and Discussion

Identification of HDV-like agents in meta-transcriptomic libraries from diverse hosts

We investigated a large and diverse set of RNA-Seq libraries generated from previous [12] and on-going studies for the presence of divergent HDV-like agents. The RNA sequencing results of the rRNA depleted libraries from newt, toad, fish and termite libraries resulted in 5,545,902, 4,266,161, 11,064,877, 68,094,815, 366,319,352 and 431,345,357 paired reads, which were assembled into contigs, ranging from 9,687 to 639,393 contigs per library (as listed in Table 1).

By combining meta-transcriptomic and protein homology search approaches, we identified and characterized four highly divergent HDV-like circular agents: (i) from the Subterranean termite (*Schedorhinotermes intermedius*) - termite HDV-like; (ii) from a mixture of fish (from class *Actinopterygii, Chondrichthyes and Agnatha*) - fish HDV-like; and (iii) two from amphibians the Asiatic toad (*Bufo gargarizans*) - toad HDV-like, and the Chinese fire belly

newt (*Cynops orientalis*) - newt HDV-like. In each case the HDV-like genomes were identified from a single contig in each library and by identifying overlapping terminal regions. The full-length circular genomes were confirmed to be between 1,591 and 1,735 nt in length (Table 1), consistent with the genome sizes of other HDVs and HDV-like agents. Remapping of the sequence reads from each library showed the specific coverage for each virus was between 24x and 205x, which corresponds to an abundance of 0.0001% to 0.022% in each library. The GC content in the novel HDV-like agents ranged between 46% to 58%, which is lower than human HDV (approximately 60% GC content) [25] (**Error! Reference source not found.**).

Importantly, and consistent with HDV, the identified HDV-like agents all presented with self-complementary, unbranched rod-like structures (**Error! Reference source not found.**). According to the conserved domain searches, the predicted delta-antigens in the newt (amHDAg), toad (tfHDAg), fish (fiHDAg) and termite (tHDAg) HDV-like viruses encoded proteins of 225, 186, 180 and 184 amino acids (Table 2), respectively, and in each case the HDAg superfamily was the highest scoring match for our protein domain searches. Importantly, none of the contigs matched any known host genes in either the nt or nr databases and the HDAg was again the highest scoring search hit.

We also identified the highly conserved poly(A) signal sequence (5'-AAUAAA-3') upstream of each putative HDAg. In addition, we utilized alternative homology-based tools (HHpred and Phrye2) to define protein domains. Similar to the reverse PSI-BLAST, all the top scoring hits for our putative HDAgs matched with known delta antigen protein. The HHpred results showed a probability over 94% for the HDAg (1A92B, delta antigen; leucine zipper; coiled-coil, oligomerization, hepatitis delta virus). In the case of Phrye2, the putative HDAgs all hit template c12a9B as the best-match, with more than 89% confidence, demonstrating oligomerization domain of hepatitis delta antigen (Table 3). However, the potential new HDAgs are extremely divergent and have amino acid identities to the four human HDV genotypes between 13–26% (Table 4).

Phylogenetic analysis returned a phylogeny broadly congruent with the evolutionary relationships of the hosts, as expected if there were long-term virus-host co-divergence as seen in some other viral families [12,26] (Figure 3). However, the avian and snake HDV-like viruses clearly fell within the diversity of human HDV sequences (although with low node

support), which likely reflects the adverse effects of high levels of sequence divergence and very long branches on phylogenetic accuracy, as well as the short length of the sequences used in the analysis.

Characterization of novel HDAg proteins

Post-translational modifications of HDAgs include lysine acetylation, arginine methylation, serine and threonine phosphorylation important for modulating HDV functions and the viral cycle [27], and these conserved residues are typical HDAg features. For example, arginine residues (R13) using arginine methyltransferase for methylation were proposed to enhance both genomic RNA and mRNA synthesis [28], lysine residues (K72) are acetylated for cell localization and viral RNA synthesis [29], and serine (S176) interacts with the processive RNA pol II, regulates viral antigenomic RNA replication [30]. Further, the leucine residues (red arrow, Figure 4) potentially represent a typical HDV leucine zipper feature, with the exception of the strict heptad repeat (Figure 3). These residues can therefore be considered signatures of putative HDAgs and are all conserved in our viruses, with the exception of the arginine residues in termite HDV that show a potential shift at +6 (R19), and the lysine residues in toad HDV at +2 (K74) (Figure 4). However, the isoprenylation motif, C-X-X-X, required for HDV assembly and release, was not identified in the C-terminal region of the putative delta proteins, including the potential frame-shifted extensions. The amino acid sequence of the carboxyl-terminal extension of HDAgs is conserved within, but not between, HDVs. In our study, the C-terminal sequences of the putative HDAgs lack unique Pro/Gly-rich farnesylated residues, which are important for replication and hypothesized to interact with the HBV envelope proteins for virus assembly. Indeed, the distinct structures observed in our putative HDAgs might imply different packaging properties and virus replication processes in non-human HDV-like agents, although this clearly merits further investigation.

The predicted coiled-coil domains of putative HDAgs, which facilitate multimerization and replication, were found located at N-terminal sites overlapping with the coiled-coil region of other known HDAgs (Figure 4). The putative coiled-coil domains sit at positions 46 aa to 68 aa in afHDAg (newt), 10 aa to 46 aa in tfHDAg (toad), 23 aa to 42 aa in fiHDAg (fish) and 26 aa to 42 aa in tHDAg (termite). The amino acid composition of the sequences were determined, showing that the isoelectric point (pI value) of putative HDAgs were all around 10 (10.4 in amHDAg, 10.9 in tfHDAg, 10.3 in fiHDAg, 10.2 in tHDAg), which is in a similar range comparing with reference HDAgs (pI values range from 10.4 to 10.8)

204

205 **Helper viruses and evolutionary implications**

206 While all the novel HDV-like circular agents identified here are highly divergent compared to
207 existing human HDVs and recently identified bird and snake viruses, they retain important
208 genomic features including size, circular genomes, and unbranched rod-like RNA structures.
209 Similarly, despite their sequence divergence, conserved HDAg domains are readily
210 identifiable and the putative HDAgs also demonstrate similar promoter structures, amino acid
211 properties and conserved post-translational residues. Critically, none of these newly
212 described HDV-like agents were associated with co-infecting hepadnaviruses, which is
213 central to the biology of human HDV. Instead, a number of other viruses were present in the
214 relevant sequencing libraries (Table 5), including Wenling frogfish arenavirus 2, Wenling
215 minipizza batfish hantavirus, Wenling yellow goosefish hantavirus, and Wenling minipizza
216 batfish reovirus 1, 2 & 3, (in the XQTMS library; fish), and Zhejiang chinese fire belly newt
217 astrovirus 1,2 & 3 (in the DFRYC and DFRYG libraries; newts), and an Wuhan asiatic toad
218 influenza virus (in the toadflu library; toad), as determined previously [12].

219

220 This observation is consistent with the bird and snake HDVs described recently that also
221 lacked any evidence of hepadnavirus co-infection, but which were infected with influenza
222 and arenaviruses, respectively [9,10]. It is therefore increasingly likely that deltaviruses
223 might use other helper viruses for generating infectious virion particles or alternative
224 mechanisms for replication and transmission. Interestingly, recent studies suggest that human
225 HDV can exist *in vivo* without HBV replication, or when HBV is suppressed by antivirals
226 [31]. Given the paucity of non-human HDVs the potential diversity of these viruses remains
227 unknown, and their replication mechanisms and possible associated helper viruses awaits
228 exploration.

229

230 Collectively, however, our results suggest that HDV-like agents have perhaps been associated
231 with animal hosts for their entire evolutionary history of the Metazoa. This is stark contrast to
232 the assumption that HDV is only present in humans and may even have evolved as an
233 escaped human gene [7]. A variety of theories have been put forward to explain the origin of
234 HDVs, including their derivation from plant viroids [32], virusoids or retroviroids [1],
235 evolution from host-associated mRNA precursor genes, such as DIPA [33], or directly
236 originated from host transcriptome [7]. Our data challenge these ideas, suggest that HDVs

237 have existed for many millions of years, and imply that more invertebrate and vertebrate
238 deltavirus-like agents will surely be discovered.
239
240 **Funding:** ECH is funded by an ARC Australian Laureate Fellowship (FL170100022). JHOP
241 is supported by the Swedish research council FORMAS (grant nr 2015-710). NL is supported
242 by an ARC Future Fellowship (FT160100463).

243 **Acknowledgments:** The Melbourne WHO Collaborating Centre for Reference and Research
244 on Influenza is supported by the Australian Department of Health. We also acknowledge The
245 University of Sydney HPC service for providing high-performance computing resources that
246 have contributed to the research results reported in this paper.
247
248 **Conflicts of Interest:** The authors declare no conflict of interest.

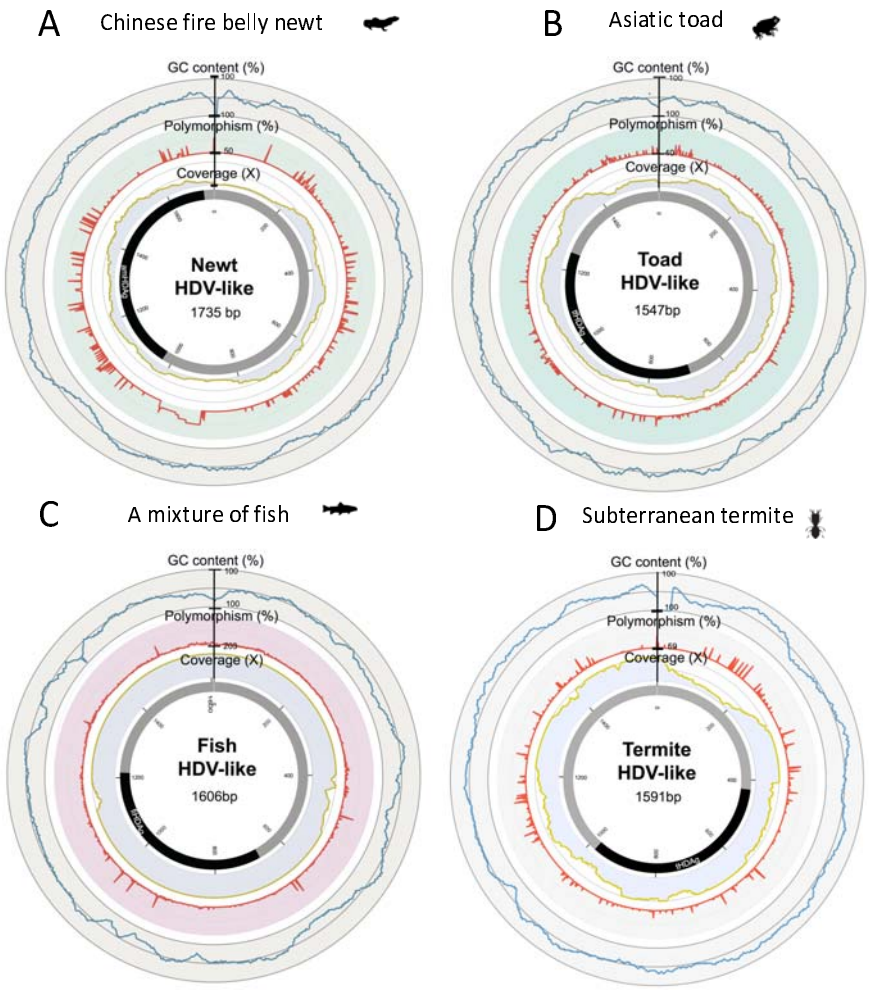
249 **References**

250 1. Taylor, J.; Pelchat, M. Origin of hepatitis delta virus. *Future Microbiol* **2010**, *5*, 393-402,
251 doi:10.2217/fmb.10.15.
252 2. Lasda, E.; Parker, R. Circular RNAs: diversity of form and function. *RNA* **2014**, *20*,
253 1829-1842, doi:10.1261/rna.047126.114.
254 3. Lai, M.M. RNA replication without RNA-dependent RNA polymerase: surprises from
255 hepatitis delta virus. *J Virol* **2005**, *79*, 7951-7958, doi:10.1128/JVI.79.13.7951-
256 7958.2005.
257 4. Suarez-Amaran, L.; Usai, C.; Di Scala, M.; Godoy, C.; Ni, Y.; Hommel, M.; Palomo, L.;
258 Segura, V.; Olague, C.; Vales, A., et al. A new HDV mouse model identifies
259 mitochondrial antiviral signaling protein (MAVS) as a key player in IFN-beta induction.
260 *J Hepatol* **2017**, *67*, 669-679, doi:10.1016/j.jhep.2017.05.010.
261 5. Gerin, J.L. Animal models of hepatitis delta virus infection and disease. *ILAR J* **2001**, *42*,
262 103-106.
263 6. Brazas, R.; Ganem, D. A cellular homolog of hepatitis delta antigen: implications for
264 viral replication and evolution. *Science* **1996**, *274*, 90-94.
265 7. Salehi-Ashtiani, K.; Luptak, A.; Litovchick, A.; Szostak, J.W. A genomewide search for
266 ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **2006**, *313*,
267 1788-1792, doi:10.1126/science.1129308.

8. Taylor, J.M. Host RNA circles and the origin of hepatitis delta virus. *World J Gastroenterol* **2014**, *20*, 2971-2978, doi:10.3748/wjg.v20.i11.2971.
9. Wille, M.; Netter, H.J.; Littlejohn, M.; Yuen, L.; Shi, M.; Eden, J.S.; Klaassen, M.; Holmes, E.C.; Hurt, A.C. A Divergent Hepatitis D-Like Agent in Birds. *Viruses* **2018**, *10*, doi:10.3390/v10120720.
10. Udo Hetzel, L.r.S., Teemu Smura, Barbara Pra□hauser, Olli Vapalahti, Anja Kipar, Jussi Hepojoki. Identification of a novel deltavirus in Boa constrictor. *bioRxiv* **2018**.
11. Huang, C.R.; Lo, S.J. Evolution and diversity of the human hepatitis d virus genome. *Adv Bioinformatics* **2010**, 10.1155/2010/323654, 323654, doi:10.1155/2010/323654.
12. Shi, M.; Lin, X.D.; Chen, X.; Tian, J.H.; Chen, L.J.; Li, K.; Wang, W.; Eden, J.S.; Shen, J.J.; Liu, L., et al. The evolutionary history of vertebrate RNA viruses. *Nature* **2018**, *556*, 197-202, doi:10.1038/s41586-018-0012-7.
13. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114-2120, doi:10.1093/bioinformatics/btu170.
14. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M., et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **2013**, *8*, 1494-1512, doi:10.1038/nprot.2013.084.
15. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **2015**, *12*, 59-60, doi:10.1038/nmeth.3176.
16. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I., et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res* **2015**, *43*, D222-226, doi:10.1093/nar/gku1221.
17. Soding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **2005**, *33*, W244-248, doi:10.1093/nar/gki408.
18. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **2015**, *10*, 845-858, doi:10.1038/nprot.2015.053.
19. B, B. BBMap short-read aligner, and other bioinformatics tools. 2016.
20. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**, *30*, 772-780, doi:10.1093/molbev/mst010.

21. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972-1973, doi:10.1093/bioinformatics/btp348.
22. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **2010**, *59*, 307-321, doi:10.1093/sysbio/syq010.
23. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **2003**, *31*, 3406-3415.
24. Delorenzi, M.; Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **2002**, *18*, 617-625.
25. 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.; Houghton, M. Structure, sequence and expression of the hepatitis delta (delta) viral genome. *Nature* **1986**, *323*, 508-514, doi:10.1038/323508a0.
26. Shi, M.; Lin, X.D.; Tian, J.H.; Chen, L.J.; Chen, X.; Li, C.X.; Qin, X.C.; Li, J.; Cao, J.P.; Eden, J.S., et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**, 10.1038/nature20167, doi:10.1038/nature20167.
27. Huang, W.H.; Chen, C.W.; Wu, H.L.; Chen, P.J. Post-translational modification of delta antigen of hepatitis D virus. *Curr Top Microbiol Immunol* **2006**, *307*, 91-112.
28. Li, Y.J.; Stallcup, M.R.; Lai, M.M. Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication. *J Virol* **2004**, *78*, 13325-13334, doi:10.1128/JVI.78.23.13325-13334.2004.
29. Mu, J.J.; Tsay, Y.G.; Juan, L.J.; Fu, T.F.; Huang, W.H.; Chen, D.S.; Chen, P.J. The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis. *Virology* **2004**, *319*, 60-70, doi:10.1016/j.virol.2003.10.024.
30. Hong, S.Y.; Chen, P.J. Phosphorylation of serine 177 of the small hepatitis delta antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *J Virol* **2010**, *84*, 1430-1438, doi:10.1128/JVI.02083-09.
31. Kabacam, G.; Onder, F.O.; Yakut, M.; Seven, G.; Karatayli, S.C.; Karatayli, E.; Savas, B.; Idilman, R.; Bozdayi, A.M.; Yurdaydin, C. Entecavir treatment of chronic hepatitis D. *Clin Infect Dis* **2012**, *55*, 645-650, doi:10.1093/cid/cis459.

333 32. Rocheleau, L.; Pelchat, M. The Subviral RNA Database: a toolbox for viroids, the
334 hepatitis delta virus and satellite RNAs research. *BMC Microbiol* **2006**, *6*, 24,
335 doi:10.1186/1471-2180-6-24.
336 33. Long, M.; de Souza, S.J.; Gilbert, W. Delta-interacting protein A and the origin of
337 hepatitis delta antigen. *Science* **1997**, *276*, 824-825.
338



339

340 **Figure 1.** Genome organization of the Hepatitis deltavirus-like (HDV) agents in diverse
341 animal taxa. In each metadata ring, the external circles indicate the percentage GC content
342 (blue), percentage nucleotide polymorphism (orange), and read coverage (yellow) of the
343 genomes. The inner gray circle represents the genome, and the black region shows the
344 predicted ORF of hepatitis delta antigen (HDAg).
345

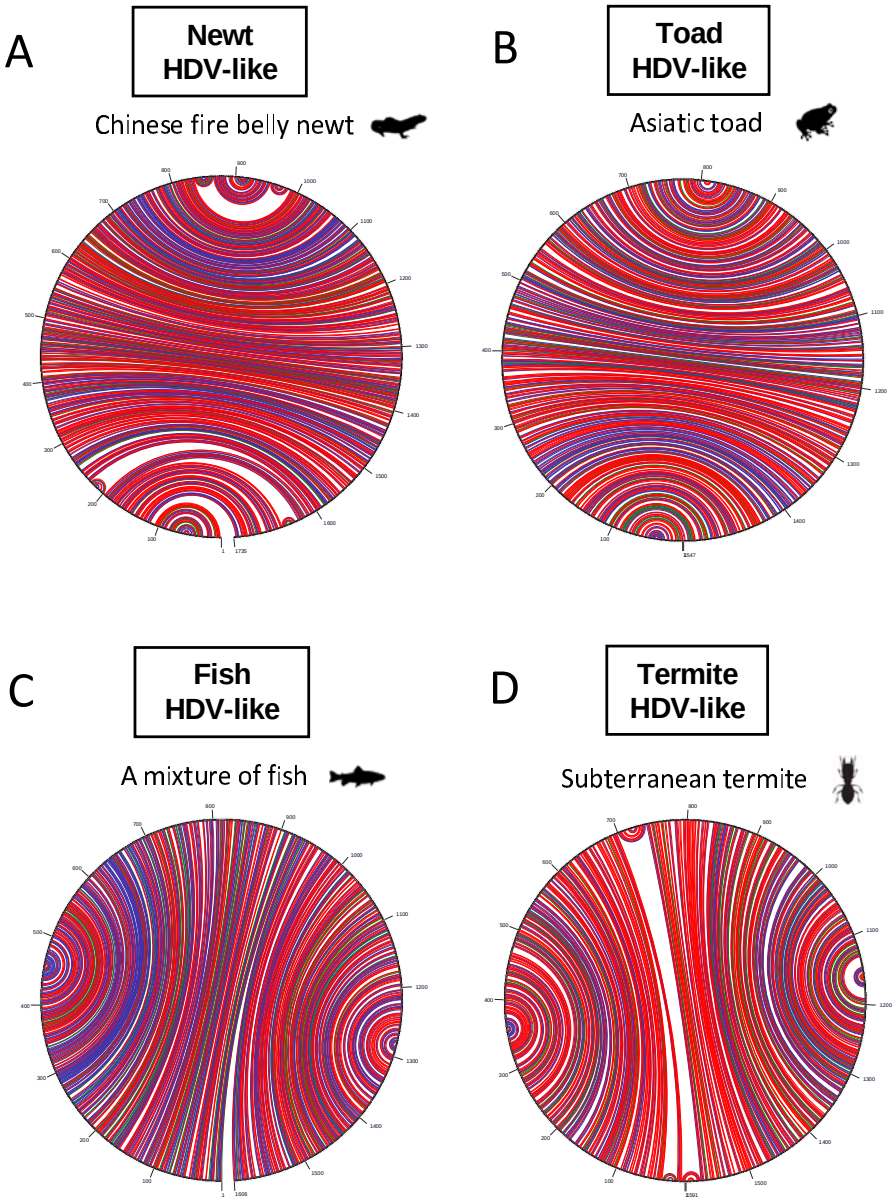


Figure 2. Circle graphs indicating each circular RNA genome structure of HDV-like agents folding into unbranched rod-like structures. The circle circumference represents the genome sequence, and the arcs represent the base pairing. Coloring of arcs are: red for G-C pairing, blue for A-U pairing, green for G-U pairing, and yellow for other types.

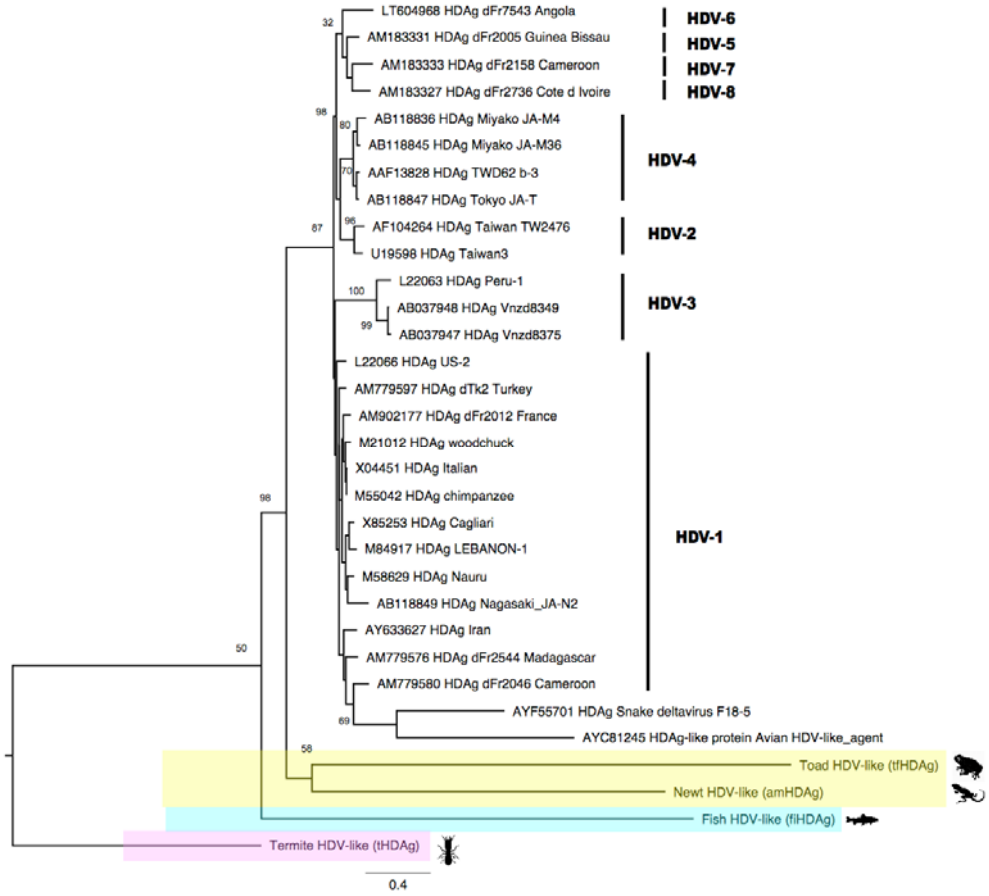


Figure 3. Phylogenetic relationships among the amino acid sequences of HDAg proteins from human HDV and the HDV-like viruses newly determined here. The tree is rooted on the most divergent sequence from the termite. All branch lengths are scaled to the number of amino acid sequences per site. Bootstrap support values are shown for key nodes.

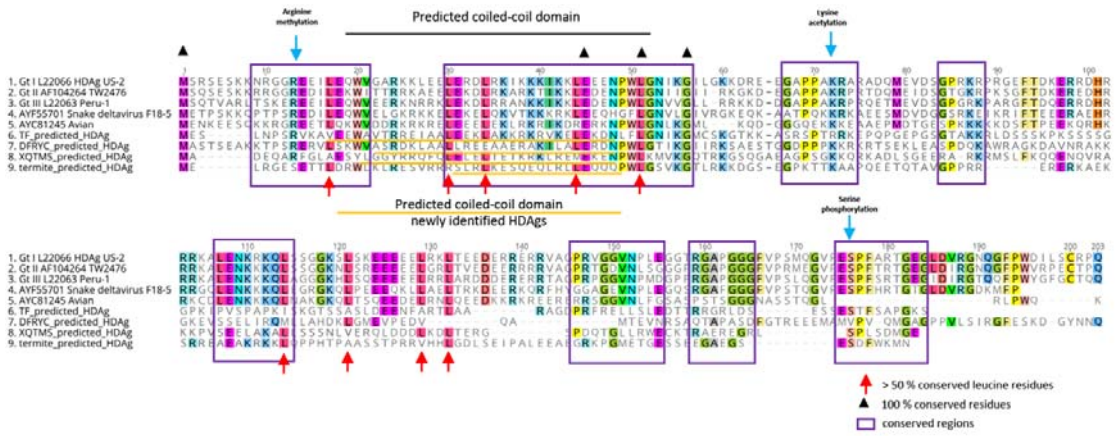


Figure 4. Characterization of the putative HDAg proteins in the HDV-like viruses newly determined here. The translated HDAg genomes of three (human) HDV genotypes were compared with the putative HDAg proteins. The potential coiled-coil region is highlighted, also including the presence of leucine residues in the correct spacing for a leucine zipper (red arrow). Post-translationally modified arginine residues (methylation), lysine residues (acetylation) and serine residues (phosphorylation) that are conserved between different HDV genotypes are indicated with blue asterisks. The conserved regions shared similar signatures between different HDAgs are marked with purple frames.

369 **Table 1**. Information on the RNA sequencing libraries containing HDV-like agents.

Library name	Library accession	Host Class	Host species	Host Organ	Assembly size (nt)	Total contigs	HDV-like agents (nt)	Reads mapped
DFRYC	SRR6291295	Amphibian	*Cynops orientalis*	Gut	5,545,902	12,875	1735	433
DFRYG	SRR6291301	Amphibia	*Cynops orientalis*	Liver	4,266,161	9,687	1735	782
Toadflu (WHHM)	NA	Amphibia	*Bufo gargarizans*	Lung	11,064,877	69,610	1547	371
XQTMS	SRR6291319	Actinopterygii, Chondrichthyes, Agnatha	*Macroramphosus scolopax, Ophidion sp, Eptatretus burgeri, Okamejei acutispina, Proscyllium habereri, Lophius litulon, Eleutheronema tetradactylum, Zeus faber, Antennarius striatus, Halieutaea stellata, Gonorynchus abbreviatus*	Gill	68,094,815	169,140	1547	954
Termite5v	NA	Insecta	*Schedorhinotermes intermedius*	Whole body	366,319,352	560,226	1591	673
Termite6v	NA	Insecta	*Schedorhinotermes intermedius*	Whole body	431,345,357	639,393	1591	579

370

371 **Table 2**. Characterization of the HDV-like agents and their putative HDAgs.
372

Library code	Host common name	HDV-like agents	size (bp)	GC content (%)	putative HDAg	size (aa)
DFRY	Amphibian/Chinese fire belly newt	newt HDV (amHDV)	1735	53.8	amHDAg	225
TF	Amphibian/Asiatic toad	toad HDV (tfHDV)	1547	54.3	tfHDAg	186
XQTMS	Fish/ a pool of fish from class *Actinopterygii*, *Chondrichthyes* and *Agnatha*	fish HDV (fiHDV)	1606	46.3	fiHDAg	180
termite	Termite/Subterranean termite	termite HDV (tHDV)	1591	56.8	tHDAg	184

HDAg protein	**CD-search hit**	**e-value**	**HHpred top hit**	**Probability** (%)	**e-value**	**Phrye2 top hit**	**confidence** (%)	**Identity (%)**
tfHDAg	HDV ag super family	6.64e-5	1A92B Delta antigen; leucine zipper; coiled-coil; oligomerization, hepatitis delta virus	99.83	3.5e-24	c12a9B Oligomerization domain of hepatitis delta antigen	97	44
amHDAg		7.23e-9		99.97	5.7e-22		99.9	54
fiHDAg		1.11e-5		94.78	1.8e-2		89.3	34
tHDAg		3.70e-5		96.1	8.2e-4		92.5	29

Table 3. Protein prediction based on the amino acid sequences of the putative HDAgs using CD-search, HHpred and Phrye2.

376 **Table 4.** Percentage identity among HDAgs of three genotypes and the putative HDAgs
377 identified in this study. Sequence similarity is calculated based on the alignment of amino
378 acid sequences from each complete HDAg, comprising Genotype I: US-2 (AAG26089.1),

HDAgs and Identity (%)	Genotype I	Genotype II	Genotype III	Snake	Avian	tfHDAg	amHDAg	fiHDAg	tHDAg
Genotype I	–	73.76	64.85	51.81	36.63	25.68	23.15	22.95	25.97
Genotype II	73.76	–	65.67	47.67	34.33	24.59	21.18	19.67	20.99
Genotype III	64.85	65.67	–	49.22	35.32	22.4	20.2	20.22	22.65
Snake	51.81	47.67	49.22	–	35.23	19.67	17.53	18.03	20.99
Avian	36.63	34.33	35.32	35.23	–	18.03	16.92	18.78	20.56
tfHDAg	25.68	24.59	22.4	19.67	18.03	–	15.08	17.86	13.53
amHDAg	23.15	21.18	20.2	17.53	16.92	15.08	–	18.08	15.93
fiHDAg	22.95	19.67	20.22	18.03	18.78	17.86	18.08	–	16.67
tHDAg	25.97	20.99	22.65	20.99	20.56	13.53	15.93	16.67	–

379 Genotype II; TW2476 (AAG26088.1), Genotype III: Peru-1 (AAB02596.1), Snake: Snake
380 deltavirus F18-5 (AYF55701.1), and Avian: Avian HDV-like agent (AYC81245).
381

382 **Table 5**. Details of the HDV-like agents and associated viruses discovered previously.

383

Library name	HDV-like agents	Other viruses in library [12]
DFRYC	newt HDV (amHDV)	Zhejiang chinese fire belly newt astrovirus 1
DFRYG	newt HDV (amHDV)	Zhejiang chinese fire belly newt astrovirus 1 Zhejiang chinese fire belly newt astrovirus 2 Zhejiang chinese fire belly newt astrovirus 3
Toadflu (WHHM)	toad HDV (tfHDV)	Wuhan asiatic toad influenza virus Wuhan asiatic toad astrovirus Wenling frogfish arenavirus 2
XQTMS	fish HDV (fiHDV)	Wenling minipizza batfish hantavirus Wenling yellow goosefish hantavirus Wenling minipizza batfish reovirus 1 Wenling minipizza batfish reovirus 2 Wenling minipizza batfish reovirus 3 Wenling fish chu-like virus
Termite5v	termite HDV (tHDV)	NA
Termite6v	termite HDV (tHDV)	NA