ViroidDB: a database of viroids and viroid-like circular RNAs

Benjamin D. Lee [1,2,*], Uri Neri[3], Caleb J. Oh[4], Peter Simmonds[2] and Eugene V. Koonin [1,*]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes Health, Bethesda, MD 20894, USA, [2]Nuffield Department of Medicine, University of Oxford, Oxford OX1, UK, [3]The Shmunis School of Biomedicine and Cancer Research, Tel Aviv University, Tel Aviv 6997801, Israel and [4]Unaffiliated, Gambrills, MD 21054, USA

Received August 15, 2021; Revised October 01, 2021; Editorial Decision October 04, 2021; Accepted October 06, 2021

ABSTRACT

We introduce ViroidDB, a value-added database that attempts to collect all known viroid and viroid-like circular RNA sequences into a single resource. Spanning about 10 000 unique sequences, ViroidDB includes viroids, retroviroid-like elements, small circular satellite RNAs, ribozyviruses, and retrozymes. Each sequence's secondary structure, ribozyme content, and cluster membership are predicted via a custom pipeline optimized for handling circular RNAs. The data can be explored via a purpose-built user interface that features visualizations, multiple sequence alignments, and a portal for downloading bulk data. Users can browse the data by sequence type, taxon, or typo-tolerant search of metadata fields. The database is freely accessible at https://viroids.org.

INTRODUCTION

Viroids are a unique class of plant pathogens, some of which cause economically important diseases of agricultural plants (1). Comprising small circular RNA molecules, between 220 and 450 nucleotides in size, viroids encode no proteins and are the smallest known infectious agents. Unlike viruses, which are parasites of the host's translational system, viroids are parasites of the host's transcriptional system. By hijacking host DNA-dependent RNA polymerase (DdRP) to transcribe RNA, viroids replicate and can cause pathology purely by means of RNA (2). Employing the rolling circle mechanism, viroids replicate by producing multimeric intermediates, which are then cleaved to unit length either by ribozymes formed from both polarities of the viroid genomic RNA or through coopting host RNAses (3). The monomers are then ligated using the host's DNA ligase (4,5). Since their discovery in 1971, knowledge of the diversity of viroids has steadily increased, with two families, *Avsunviroidae* and *Pospiviroidae*, currently classified. Members of the former family rely on encoded autocatalytic ribozymes for cleavage to unit length during replication, whereas members of the latter family use host enzymes. Specifically, members of the family *Avsunviroidae* encode the hammerhead ribozyme (HHR), a defining trait of the family (6). Members of the *Pospiviroidae* family, in contrast, do not use ribozymes during replication, instead relying upon conserved sequence regions that are not catalytic but do serve as a cleavage site for host RNase III (7). Structurally, viroids of these families may also differ as well, with some members of the family *Avsunviroidae* adopting a branched conformation (*i.e.* those classified in the genus *Pelamoviroid*), whereas members of the family *Pospiviroidae* adopt a rod-shaped or quasi-rod-shaped conformation (8).

In addition to viroids proper, several groups of RNA agents share similar key features. One such agent is the retroviroid-like element (hereafter retroviroid), carnation small viroid-like RNA (CarSV) and its DNA form (9–11). CarSV, the only retroviroid so far described, is similar to viroids in size, circularity and the presence of hammerhead ribozymes in both polarities. However, it differs from viroids in that there exists a homologous extrachromosomal DNA sequence that can be integrated into the plant genome by means of a helper pararetrovirus. Unlike viroids, CarSV does not appear to transmit horizontally from plant to plant.

Small circular satellite RNAs (also sometimes called virusoids and hereinafter abbreviated satRNAs) are another type of infectious circular RNA that resemble viroids in many respects, such as rolling circle replication and size (about 300 nucleotides), but differ in that they are encapsidated into the capsid of the respective helper viruses and replicate using the helper virus RdRP (12). Effectively, these agents can be considered encapsidated viroids. Unlike viroids, satRNAs are known to encode both HHRs and hairpin ribozymes, representing a different autocatalytic motif. The taxonomy of these agents, though for-

*To whom correspondence should be addressed. Tel: +1 602 614 0850; Email: benjamindlee@me.com
Correspondence may also be addressed to Eugene V. Koonin. Tel: +1 301 435 5913; Email: koonin@ncbi.nlm.nih.gov

Published by Oxford University Press on behalf of Nucleic Acids Research 2021.

mally under the purview of the International Committee on the Taxonomy of Viruses (ICTV), has not yet been established.

Another group of viroid-like agents are the members of *Ribozyviria*, a recently described viral realm (13) that includes deltaviruses, such as hepatitis delta virus (HDV), a human pathogen. Similar to some viroids, ribozyviruses possess rod-shaped circular RNA genomes that replicate via the rolling circle mechanism and encode ribozymes, albeit distinct from those in viroids, that autocatalytically process replication intermediates. However, ribozyviruses have substantially larger genomes than viroids, encode their own nucleocapsid protein, and their reproduction depends on a helper virus (hepatitis B virus in the case of HDV) which provides the envelope protein for ribozyvirus virions, Until recently, HDV was the sole member of the *Deltavirus* genus, but the discovery of distantly related viruses in a variety of vertebrates and invertebrates, including rodents (14), snakes (15), termites (16), fish (16), birds (17) and bats (18), suggests a considerable uncharacterized diversity among the ribozyviruses. Currently, all ribozyviruses are classified as a single family, *Kolmioviridae*, containing 8 genera including *Deltavirus*.

Yet another recently discovered group of circular RNA agents are retrozymes, a class of retrotransposons that propagate via circular RNA intermediates of about 170 to 400 nucleotides. They do not encode any proteins but contain self-cleaving HHRs (19). These unique agents are not pathogenic and are not autonomous, unlike viroids. Rather, retrozyme replication appears to require the machinery encoded by autonomous retrotransposons. Like some satRNAs and some members of the family *Avsunviroidae*, these RNAs also adopt a branched conformation. Given the significant similarities between retrozymes and the members of the family Avsunviroidae, the relationship between these groups is of considerable interest, with a potential evolutionary scenario being that members of the family *Avsunviroidae* descend from retrozymes (20).

These several distinct, not necessarily evolutionarily related groups of mobile elements with small circular RNAs genomes fit the umbrella definition of 'viroid-like RNAs.' The molecular signatures of viroid-like RNAs are small, circular genomes, minimal or nonexistent coding capacity, replication via the rolling circle mechanism that is catalyzed by a hijacked DdRP or RdRP, and the frequent presence of ribozymes involved in replication intermediate processing. Given the small size of viroid-like RNA and advent of methods based high-throughput sequencing (21), it appears highly likely that additional types of such elements remain to be discovered.

There are currently no databases that would cover the panoply of the viroid-like agents. Previously developed databases (22,23) are out of date and, apparently, are not maintained and inaccessible. A complete database is a necessity for continued efficient study of these increasingly diverse agents. Thus, we developed ViroidDB to provide the community with a consistent, usable, and up-to-date resource for the analysis and classification of viroid-like RNA agents.

RESOURCE CONTENT

ViroidDB is a comprehensive database of viroids and viroid-like RNA agents, comprising almost 10 000 genomic sequences. ViroidDB draws all primary sequence data from the National Center for Biotechnology (NCBI) GenBank (24) and RefSeq (25) databases. A custom pipeline then produces multiple downstream analyses that are incorporated into the database (Figure 1).

Sequence data

Initially, all sequences marked as complete and associated with viroid-like taxa (retrozymes excepted) were downloaded from the NCBI Virus resource (26) and formatted consistently. Retrozymes, which are not taxonomically assigned, were identified by searching for the term 'retrozyme' in GenBank. All available metadata for these sequences were also downloaded.

Deduplication was first performed with respect not only to the exact sequences themselves but also to all rotations in both polarities. To this end, a canonical representation of the sequence is selected using the lexicographically minimal rotation (both polarities included). All subsequent sequences with the same canonical rotation are grouped together. From each group of rotationally identical sequences, a single reference sequence is chosen arbitrarily if no RefSeq-derived sequence is a member. The other rotationally identical sequences are discarded and their identifiers are noted on the record of the reference sequence. In addition, ambiguous sequences are removed in this phase of the pipeline.

The current version of ViroidDB contains 9691 sequences, of which the vast majority (9353) are from viroids. Among viroids, peach latent mosaic viroid is the most abundantly represented, with 4891 samples. The vast amount of these nearly identical sequences are the result of deep sequencing studies into the heterogeneity of quasispecies during infection (27). Ribozyviruses comprise the next most populous sequence type ($n = 243$), followed by retrozymes ($n = 73$), retroviroids ($n = 12$) and satRNAs ($n = 10$). A more detailed breakdown of the sequence content by genus is given in Table 1.

Secondary structures

The secondary structure for both polarities of each sequence was predicted using RNAfold (28). Additional metrics (e.g. percent of bases folded) are also computed and consolidated with visualizations generated by Forna (29). In the visualizations, each base is colored depending on its base-pairing interaction (such as stems, hairpins, interior loops) for easier examination (Figure 2).

Predicted ribozymes

Beyond the secondary structure, all sequences were scanned for signature motifs of auto-catalytic ribozymes because many viroid-like RNAs are known to employ self-cleaving ribozymes for replication (7). This scan was performed using the Infernal software (30), against selected ribozyme

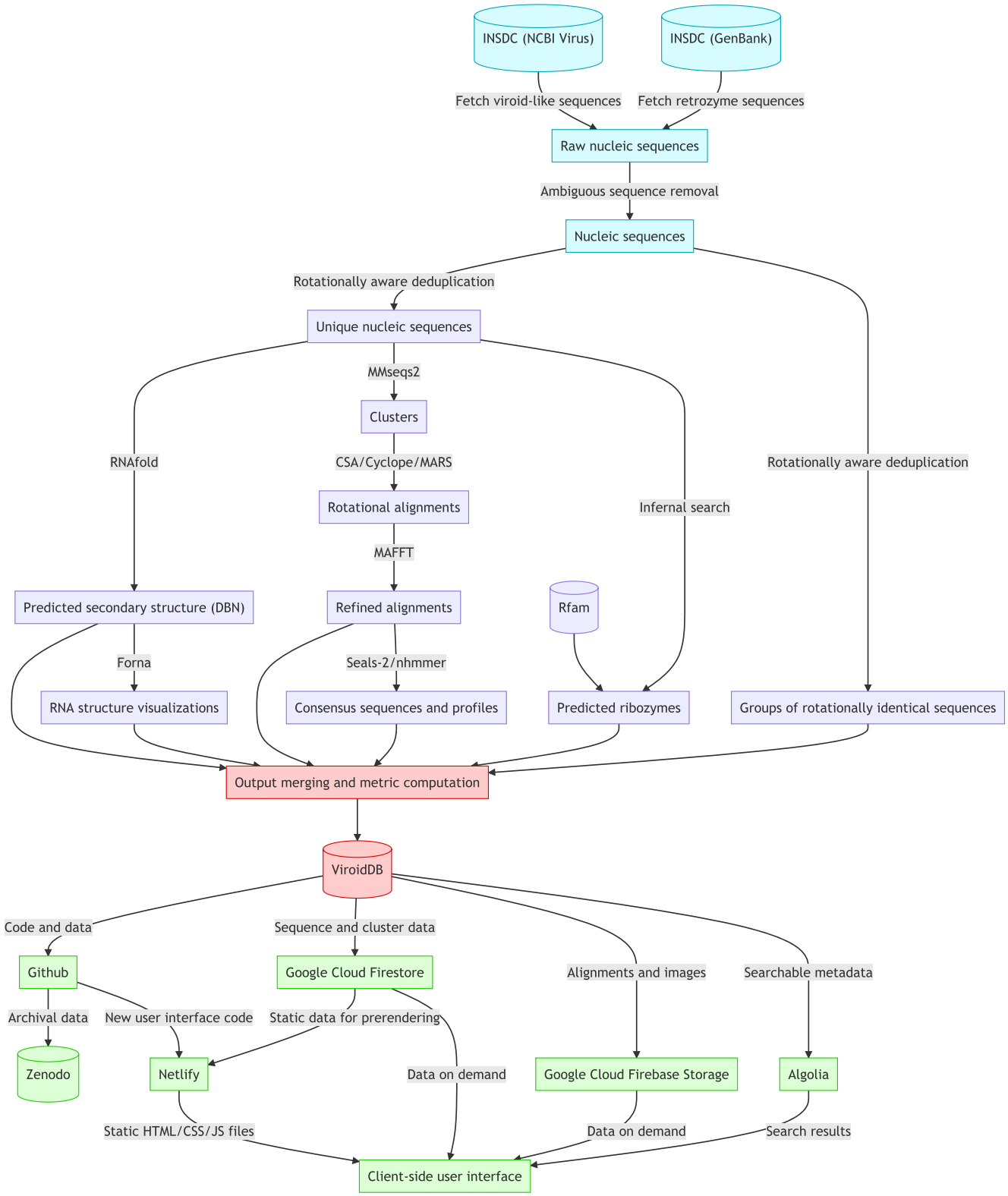

Figure 1. A flowchart of the complete data processing and release pipeline. Blue steps denote data fetching and pre-processing procedures while purple steps denote the analysis phase. Red steps correspond to post-processing organizational steps. The data release pipeline (green) is initiated upon every change to the underlying data or code ensures that the data are archived and that the website is up to date and efficient.

Table 1. The genera and number of sequences in the 2021–09-07 release of ViroidDB

Genus	Count
Pelamoviroid	5080
Pospiviroid	2363
Apscaviroid	939
Hostuviroid	488
Deltavirus	243
Cocadviroid	184
Elaviroid	102
Avsunviroid	60
Coleviroid	30

families from the Rfam database (31). All detected results with *E* values less than 0.1 were considered significant and reported. The database stores the location, type, *E* value, alignment and truncation status of each hit. Altogether, the database currently reports 11,713 putative ribozymes. The large number of ribozymes is a result of the database containing so many agents with symmetric rolling circle replication mechanisms which require one ribozyme per polarity combined with the *E* value threshold allowing alternative topologies for a given motif to be reported as well.

Sequence clusters

We devised a custom, standalone pipeline to cluster the sequences gathered from the reference database and generate multiple sequence alignments (MSAs) (Figure 1). Although these clusters are not intended to reflect taxonomic ranks, they greatly reduce the number of sequences in a consistent manner, streamlining advanced analyses, such as the identification of translocating motifs or conserved motif ordering. Initially, all deduplicated sequences are clustered using MMseqs2 (32) at multiple identity thresholds: 75%, 80%, 85%, 90%, 95% with a fixed minimum alignment coverage of 75% for both the query and the target sequences. Then, rotational alignment (in which the optimal starting position is identified) is attempted within every cluster using CSA (33). This step is necessary as similar sequences may have spuriously low alignment scores when the sequences are not oriented with consistent origins. We found that CSA was unable to produce meaningful output for all clusters for several technical reasons (namely, built-in limitation preventing handling inputs with more than 64 sequences). If CSA fails to produce an output, cyclic comparison is attempted using Cyclope (34). If Cyclope is also unsuccessful in producing an output, rotation is attempted using MARS (35).

The rotated clusters were then re-aligned using MAFFT (v7.475) using its automatic mode (36). The minority of the clusters, for which no cyclic comparison software successfully produced meaningful output, were forwarded directly to MAFFT, without rotational modifications. For each resulting MSA (one per cluster), a consensus sequence was generated using Seals-2 ((37), available at https://github.com/YuriWolf-ncbi/seals-2). The consensus sequence was then inserted into the alignment as the first sequence in the multi-FASTA file to act as a 'master sequence.' This final MSA was used to construct an nhmm profile using the HMMER suite v.3.3.2 (38).

The number of clusters at each identity threshold is given in Table 2.

DATA STORAGE AND INFRASTRUCTURE

To ensure the continuous functionality of ViroidDB, all code (both for processing and the user interface) and data are permanently archived externally in CERN's Zenodo database (https://zenodo.org). All software is available on GitHub under the open-source MIT license. The data is versioned by release date and releases are immutable. No special identifiers are presently assigned for sequences in ViroidDB; the existing accessions are used for simplicity. Because new data could change the cluster assignments, clusters are given a date- and threshold-dependent identifier.

The live version of the website was optimized for speed, scalability, and sustainability. At a high level, ViroidDB's web infrastructure consists of a browser-based application executed locally in JavaScript (albeit written in its strongly typed derivative, TypeScript) and served as static files via the Netlify content delivery network (CDN) (https://netlify.com) supported by a serverless database, file storage, and search services. The front-end of the browser application is built using the Vue (https://vuejs.org) and Nuxt (https://nuxtjs.org) frameworks, which are configured to automatically update the site upon every accepted change in the version-controlled repository hosted on GitHub. When a user loads the website, they download pre-generated HTML, CSS and JavaScript files stored entirely on the CDN's servers, which in turn fetch only the data needed to display the given page. Although the user interface relies heavily on dynamic JavaScript to retrieve the content being displayed, the compilation process is able to prefetch and statically render pages' default states (such as the truncated list of members of the family *Avsunviroidae*). This process completely eliminates the waiting period the user experiences after the page's scripts have loaded but before the data from the database has been loaded by the scripts, which in turn use it to render the document object model (DOM). It also reduces the number of database queries, which ensures that there is sufficient capacity for numerous users with minimal to no expenses resulting from operating the service.

Importantly, no servers must be maintained specifically for ViroidDB, thereby easing the development and maintenance burden. The downloaded client-side code simply queries and displays data stored in the database and file storage layer, which is built upon Google Cloud's Firebase platform (https://firebase.google.com/). Because the platform provides a free tier of service with a 50 000 query daily limit, the site is able to operate with no expected operating costs due to the database. Small metadata, such as records associating the accessions of sequences with their identified ribozymes, are stored in a serverless NoSQL database for fast retrieval while larger data, such as full-length alignments of clusters and images, are placed in an object store.

DATA QUERY AND ACCESS

The ViroidDB database supports the exploration and use of its data both via a graphical web interface and via downloadable raw data.

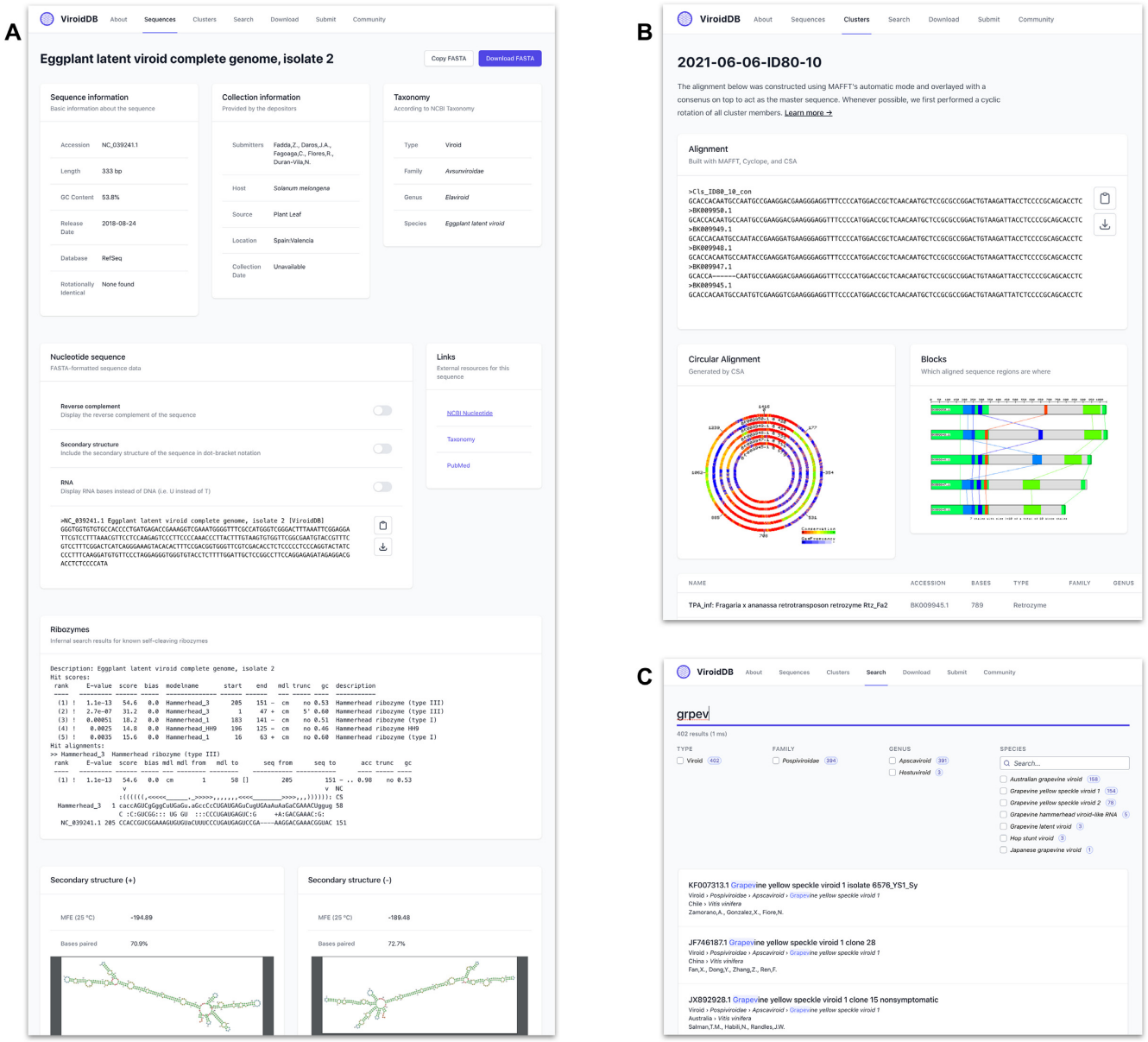

Figure 2. Examples of key pages on the ViroidDB website. (**A**) The sequence detail page. The first row of panels displays the sequence's basic information, sample collection information, and taxonomy. The next row of panels shows the sequence's nucleotide sequence in FASTA format and resource links. Users can toggle the nucleotide sequence to display the reverse complement, textual secondary structure, and/or convert between DNA and RNA bases. Users can also copy or download the nucleotide sequence in FASTA format through built-in buttons. The following panels displays alignments for any predicted self-cleaving ribozymes with secondary structures in both polarities displayed below. (**B**) The details for a cluster. The first panel shows multiple sequence alignments generated from the reference sequence after rotational alignment with the resultant nhmm profile. Visualizations of the alignments are immediately below. The final panel is a list of all sequences that are contained within the cluster. Users can jump directly to each sequence's individual page from this list. (**C**) The search page of ViroidDB demonstrates the site's fuzzy search, search-match highlighting, taxonomic refinement lists and species sub-search.

ViroidDB website

ViroidDB offers an interactive website that enables users to inspect, visualize, search, and download all the information stored in the database.

The site is organized around the sequence display page, with other pages of sequence collections linked to each sequence's page. These pages contain all the sequence-specific metadata and are organized with panels displaying basic information such as length and GC-content, taxonomy information, sample collection information, ribozymes, structures and the sequence itself. Sequences and structures can be easily copied to the user's clipboard or downloaded directly from the page. Links to external resources, such as associated PubMed articles and GenBank pages, are included as well. The sequences are listed on a sequence index page where the user can opt to only have sequences of a given type displayed for easier browsing.

Table 2. Cluster counts for varying identity thresholds (75% minimum reciprocal overlap required)

Threshold	Clusters
70%	226
75%	224
80%	252
85%	311
90%	458
95%	958

Clusters are treated analogously to sequences. A cluster index page lists each cluster and its reference sequence as well as the number of members. Clicking into a cluster page displays the list of the sequences comprising the cluster, the multiple sequence alignment, and visualizations of the rotational alignment results. Clicking a sequence within the cluster leads to the sequence display page.

We used the Algolia service (https://algolia.com) to provide typo-tolerant approximate searching of all sequences and associated metadata in the database. Searchable fields include the accession, sequence name, full taxonomy, authors, geographic location, host, and isolation site. This integration allows the user to instantly (<10 ms) perform full-text search with optional type (*e.g.* viroid), family, genus, and species filters. If a match to the input query is found, the exact part of the record for which the match is found is highlighted in the results. As the user types a query into the search area, the number of results in each taxon is displayed dynamically.

Because this site is designed to serve as a resource for the community, it also hosts a community section of the website. This section enables members of the community to post upcoming events and notices of interest to a virtual bulletin board. Additionally, email forwarding addresses using the viroids.org domain name are available upon demand at no cost. Finally, the site's community page provides access to an internet chat room to facilitate collaboration.

Data portal

In addition to facilitating the exploration of individual sequences and clusters, ViroidDB is also meant to enable computationally oriented researchers the ability to rapidly access high-quality data on viroid-like RNAs. To do so, the website also includes a data portal enabling users to download the cleaned sequence data and downstream analysis results in bulk form. The data download portal includes sequence data in FASTA format, RNA structure data in dot-bracket notation format, tabular and semi-structured output from Infernal, PDF documents containing structure visualizations, bitmap visualizations and alignments for each cluster, and a JavaScript Object Notation (JSON) file containing metadata. These files can be readily used for subsequent bioinformatic research.

FUTURE DIRECTIONS AND LIMITATIONS

Presently, the raw data ingestion and processed data deployment stages are performed manually. Although not particularly cumbersome (taking less than ten minutes each), relying on human intervention to create new database releases limits the frequency of releases. Both of these stages can be automated: the manual effort is limited in scope to data transfer rather than analysis. As the pace of viroid-like agent discovery increases, the ability to perform fully automatic updates will become critical. While the database is currently updated on an *ad-hoc* basis, it will be optimal to create regularly scheduled releases in the future.

Additionally, while all ViroidDB application and pipeline software is available under a permissive open-source license, not all of the software upon which the site depends is similarly licensed. Ideally, upgrades to the site will reduce the amount of non-open dependencies, thereby ensuring that the interactive aspect of the site will remain usable in the future even if the data infrastructure is no longer available.

CONCLUSION

To our knowledge, ViroidDB is the only currently available database of viroid-like RNAs. In addition to a comprehensive collection of sequences, it provides users with a wealth of information on these agents as well as analytical tools, making it a platform for bioinformatic studies. Recent metatranscriptome mining efforts have led to a dramatic expansion of the known diversity of RNA viruses. The search for viroid-like RNAs has not yet caught up, but can be expected to yield many new groups of these agents, too, as illustrated by the recent discovery of numerous ribozyme-containing RNAs. ViroidDB, which is expected to be regularly updated, should become an indispensable resource for researchers studying this remarkable class of mobile elements.

DATA AVAILABILITY

ViroidDB is freely accessible at viroids.org. The pipeline and user interface source is available under the MIT license at https://github.com/Benjamin-Lee/viroiddb and is archived at Zenodo (https://doi.org/10.5281/zenodo.5202945).

ACKNOWLEDGEMENTS

The authors would like to thank Soumil Singh for help with the site's documentation.

FUNDING

B.D.L. was supported by a fellowship from the National Institutes of Health Oxford-Cambridge Scholars Program; U.N. was partially supported by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University as well as the European Research Council [ERC-AdG 787514]; E.V.K. is supported by the Intramural Research Program of the National Institutes of Health of the USA (National Library of Medicine). Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Diener,T.O. (1971) Potato spindle tuber "virus". *Virology*, **45**, 411–428.

2. Daròs,J., Elena,S.F. and Flores,R. (2006) Viroids: an Ariadne's thread into the RNA labyrinth. *EMBO Rep.*, **7**, 593–598.
3. Branch,A.D. and Robertson,H.D. (1984) A replication cycle for viroids and other small infectious RNA's. *Science*, **223**, 450–455.
4. Nohales,M.-Á., Flores,R. and Daròs,J.-A. (2012) Viroid RNA redirects host DNA ligase 1 to act as an RNA ligase. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13805–13810.
5. Nohales,M.-Á., Molina-Serrano,D., Flores,R. and Daròs,J.-A. (2012) Involvement of the chloroplastic isoform of tRNA ligase in the replication of viroids belonging to the family Avsunviroidae. *J. Virol.*, **86**, 8269–8276.
6. Di Serio,F., Li,S.-F., Pallás,V., Owens,R.A., Randles,J.W., Sano,T., Verhoeven,J.Th.J., Vidalakis,G. and Flores,R. (2017) Viroid Taxonomy. In: *Viroids and Satellites*. Elsevier, pp. 135–146.
7. Flores,R., Minoia,S., López-Carrasco,A., Delgado,S., Martínez de Alba,Á.-E. and Kalantidis,K. (2017) Viroid Replication. In: *Viroids and Satellites*. Elsevier, pp. 71–81.
8. Wang,Y. (2021) Current view and perspectives in viroid replication. *Curr. Opin. Virol.*, **47**, 32–37.
9. Daròs,J.A. and Flores,R. (1995) Identification of a retroviroid-like element from plants. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 6856–6860.
10. Vera,A., Daròs,J.A., Flores,R. and Hernández,C. (2000) The DNA of a plant retroviroid-like element is fused to different sites in the genome of a plant pararetrovirus and shows multiple forms with sequence deletions. *J. Virol.*, **74**, 10390–10400.
11. Hegedus,K., Dallmann,G. and Balázs,E. (2004) The DNA form of a retroviroid-like element is involved in recombination events with itself and with the plant genome. *Virology*, **325**, 277–286.
12. Navarro,B., Rubino,L. and Di Serio,F. (2017) Small circular satellite RNAs. In: *Viroids and Satellites*. Elsevier, pp. 659–669.
13. Hepojoki,J., Hetzel,U., Paraskevopoulou,S., Drosten,C., Harrach,B., Zerbini,F.M., Koonin,E.V., Krupovic,M., Dolja,V.V. and Kuhn,J.H. (2020) Create one new realm (Ribozyviria) including one new family (Kolmioviridae) including genus Deltavirus and seven new genera for a total of 15 species. ResearchGate doi: https://doi.org/10.13140/RG.2.2.31235.43041, January 2021, preprint: not peer reviewed.
14. Paraskevopoulou,S., Pirzer,F., Goldmann,N., Schmid,J., Corman,V.M., Gottula,L.T., Schroeder,S., Rasche,A., Muth,D., Drexler,J.F. *et al.* (2020) Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent Proechimys semispinosus. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 17977–17983.
15. Hetzel,U., Szirovicza,L., Smura,T., Prähauser,B., Vapalahti,O., Kipar,A. and Hepojoki,J. (2019) Identification of a novel deltavirus in Boa constrictors. *mBio*, **10**, e00014–19.
16. Chang,W.-S., Pettersson,J.H.-O., Le Lay,C., Shi,M., Lo,N., Wille,M., Eden,J.-S. and Holmes,E.C. (2019) Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol.*, **5**, vez021.
17. Wille,M., Netter,H., Littlejohn,M., Yuen,L., Shi,M., Eden,J.-S., Klaassen,M., Holmes,E. and Hurt,A. (2018) A divergent hepatitis D-like agent in birds. *Viruses*, **10**, 720.
18. Bergner,L.M., Orton,R.J., Broos,A., Tello,C., Becker,D.J., Carrera,J.E., Patel,A.H., Biek,R. and Streicker,D.G. (2021) Diversification of mammalian deltaviruses by host shifting. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2019907118.
19. Cervera,A., Urbina,D. and de la Peña,M. (2016) Retrozymes are a unique family of non-autonomous retrotransposons with hammerhead ribozymes that propagate in plants through circular RNAs. *Genome Biol.*, **17**, 135.
20. de la Peña,M. and Cervera,A. (2017) Circular RNAs with hammerhead ribozymes encoded in eukaryotic genomes: The enemy at home. *RNA Biol.*, **14**, 985–991.
21. Weinberg,C.E., Olzog,V.J., Eckert,I. and Weinberg,Z. (2021) Identification of over 200-fold more hairpin ribozymes than previously known in diverse circular RNAs. *Nucleic Acids Res.*, **49**, 6375–6388.
22. Lafontaine,D.A., Mercure,S., Poisson,V. and Perreault,J.P. (1998) The viroid and viroid-like RNA database. *Nucleic Acids Res.*, **26**, 190–191.
23. Rocheleau,L. and Pelchat,M. (2006) The Subviral RNA Database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research. *BMC Microbiol.*, **6**, 24.
24. Sayers,E.W., Cavanaugh,M., Clark,K., Pruitt,K.D., Schoch,C.L., Sherry,S.T. and Karsch-Mizrachi,I. (2021) GenBank. *Nucleic Acids Res.*, **49**, D92–D96.
25. Li,W., O'Neill,K.R., Haft,D.H., DiCuccio,M., Chetvernin,V., Badretdin,A., Coulouris,G., Chitsaz,F., Derbyshire,M.K., Durkin,A.S. *et al.* (2021) RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res.*, **49**, D1020–D1028.
26. Hatcher,E.L., Zhdanov,S.A., Bao,Y., Blinkova,O., Nawrocki,E.P., Ostapchuck,Y., Schäffer,A.A. and Brister,J.R. (2017) Virus variation resource – improved response to emergent viral outbreaks. *Nucleic Acids Res.*, **45**, D482–D490.
27. Glouzon,J.-P.S., Bolduc,F., Wang,S., Najmanovich,R.J. and Perreault,J.-P. (2014) Deep-sequencing of the peach latent mosaic viroid reveals new aspects of population heterogeneity. *PLoS One*, **9**, e87297.
28. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
29. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.
30. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
31. Kalvari,I., Nawrocki,E.P., Ontiveros-Palacios,N., Argasinska,J., Lamkiewicz,K., Marz,M., Griffiths-Jones,S., Toffano-Nioche,C., Gautheret,D., Weinberg,Z. *et al.* (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.*, **49**, D192–D200.
32. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.
33. Fernandes,F., Pereira,L. and Freitas,A.T. (2009) CSA: An efficient algorithm to improve circular DNA multiple alignment. *BMC Bioinformatics*, **10**, 230.
34. Mosig,A., Hofacker,I. and Stadler,P.F. (2006) Comparative analysis of cyclic sequences: viroids and other small circular RNAs. In: *German Conference on Bioinformatics*, Vol. **83**, pp. 93–102.
35. Ayad,L.A.K. and Pissis,S.P. (2017) MARS: improving multiple circular sequence alignment using refined sequences. *BMC Genomics*, **18**, 86.
36. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
37. Walker,D.R. and Koonin,E.V. (1997) SEALS: a system for easy analysis of lots of sequences. In: *Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, pp. 333–339.
38. Wheeler,T.J. and Eddy,S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, **29**, 2487–2489.