

mirtronDB 2.0: enhanced database with novel mirtron discoveries

Fabiana Rodrigues de Goes^{1, }, Matheus Fujimura Soares^{2, }, Vitor Gregorio^{3, },
Bruno Thiago de Lima Nichio^{3, }, Alisson Gaspar Chiquitto^{3,4, }, Flavia Lombardi Lopes^{2, },
Mark Basham^{1, }, Douglas Silva Domingues^{5,*,}, Alexandre Rossi Paschoal^{1,3,*,}

¹Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, OX11 0QS, United Kingdom

²School of Veterinary Medicine, São Paulo State University (UNESP), Araçatuba, 16050-680, Brazil

³Department of Computer Science, Federal University of Technology-Parana, Cornelio Procopio, 86300-000, Brazil

⁴Federal Institute of Education, Science and Technology of Mato Grosso do Sul, Navirai, 79947-334, Brazil

⁵Department of Genetics, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba, São Paulo, 13418-900, Brazil

*Corresponding authors. Douglas Silva Domingues, Department of Genetics, “Luiz de Queiroz” College of Agriculture, 11 Avenue Pádua Dias, University of São Paulo, Piracicaba, São Paulo, 13418-900, Brazil. E-mail: dougsd@usp.br; Alexandre Rossi Paschoal, Rosalind Franklin Institute, Building R113 - Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, OX11 0QS, UK and Department of Computer Science, Federal University of Technology-Parana, 1640 Avenue Alberto Carazzai, Cornelio Procopio, PR, 86300-000, Brazil. E-mail: alexandre.paschoal@rfl.ac.uk or paschoal@utfpr.edu.br. Associate Editor: Jonathan Wren

Abstract

Motivation: MirtronDB provides a comprehensive and up-to-date resource for advancing mirtron research within RNA biology. Therefore, maintaining a specialized and continuously updated resource for mirtrons is essential to support ongoing discoveries and to serve as a key reference for researchers investigating the roles of mirtrons.

Results: Here, we present mirtronDB 2.0, an enhanced version that expands both content and functionality. This version integrates mirtron data published between 2017 and 2025, increasing the number of documented mirtrons across various species. In addition, it incorporates newly predicted mirtrons identified through a robust pipeline that combines advanced bioinformatics and machine learning approaches, with specific coverage of six mammalian species. We have introduced new website features, including an interactive dashboard to enhance usability and facilitate intuitive data exploration. These rigorous updates consolidate mirtronDB as a key resource for mirtron to the RNA biology community.

Availability and implementation: mirtronDB can be found under <http://mirtrondb.cp.utfpr.edu.br/>. The complete content of Database 2.0 and the source code for the analyses are also freely available in the FigShare repository: https://figshare.com/articles/dataset/MirtronDB_version2/29344775.

1 Introduction

Mirtrons are a subclass of pre-miRNAs encoded by a non-canonical pathway independent of the Drosha enzyme. Despite the existence of microRNA (miRNA) databases (de Amorim *et al.* 2022), e.g. miRBase (Kozomara *et al.* 2019) and MirGeneDB (Clarke *et al.* 2025), no resource was specifically dedicated to mirtrons. The available data were dispersed across publications, lacking standardization and centralization. To address this, the first version of mirtronDB (da Fonseca *et al.* 2019) introduced the only peer-reviewed and publicly accessible database dedicated to mirtrons, providing curated, species-specific data with standardized annotations. This highlights the importance of a continuously updated resource, prompting the development of this significantly expanded version.

In this work, we present mirtronDB 2.0, an updated resource featuring new content and enhanced functionality. Our systematic approach involved a comprehensive retrieval and integration of mirtron data from the recent literature, a rigorous curation process, and an *in silico* pipeline, based on similarity search and machine learning, to improve cross-species coverage and standardized annotation. We also introduce new website features, including an interactive dashboard, to improve usability and data exploration for the RNA biology community. To maintain consistency and standardization across the database while integrating newly collected mirtron data, we conducted a structured curation effort to review and remap the genomic coordinates of previously available and newly identified mirtrons to the latest Ensembl (Dyer *et al.* 2025) genome assemblies. To complement the literature-curated entries and explore

Received: 4 November 2025. Revised: 6 February 2026. Accepted: 27 February 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

the potential of discovering new mirtrons, we implemented an *in silico* prediction strategy across six mammalian species, identifying 165 potential novel homolog mirtrons. Finally, all data is open access from the download page on the website.

2 Materials and methods

The mirtronDB 2.0 pipeline integrates multiple data sources and computational strategies to ensure a comprehensive and high-quality dataset (Fig. 1). The process started with the review and update of mirtronDB 1.0 entries, along with the collection of additional mirtrons reported in the literature in recent years (Fig. 1A). Next, an *in silico* prediction pipeline was applied to identify similar mirtrons using bioinformatic and machine learning tools (Fig. 1B). Finally, we compared the entries from mirtronDB updated from literature only (Fig. 1A) with established miRNA databases to identify overlapping and unique entries (Fig. 1C). This structured approach ensures both the reliability and novelty of the database.

2.1 Mirtron data retrieval

We revisited studies published between 2007 and 2017 to review and recover data that had occasionally been missed in the mirtronDB 1.0 analyses (Table S1, available as supplementary data at Bioinformatics online, column “Revision”). In the end, we collected data from one article (Schamberger *et al.* 2012).

Next, we collected available mirtron data from articles published between 2017 and April 2025 by searching the National Center for the Biotechnology Information (NCBI) PubMed database using the query “(mirtron) AND (mirtrons)” with only the English language (Fig. 1A and Fig. S1, available as supplementary data at Bioinformatics online). This search yielded 42 articles (Collins *et al.* 2017, Monteleone and Lutz 2017, Baghdadi *et al.* 2018, Campbell *et al.* 2018, Hansen 2018, Ju *et al.* 2018, Mohammed *et al.* 2018, Rorbach *et al.* 2018, Titov and Vorozheykin 2018, Urbanek-Trzeciak *et al.* 2018, Zia and Flynt 2018, Bell *et al.* 2019, Cheng *et al.* 2019, da Fonseca *et al.* 2019, Kroupova *et al.* 2019, Zheng *et al.* 2019, Amourda and Saunders 2020, Dexheimer *et al.* 2020, Kim *et al.* 2020, Rivera *et al.* 2020, Yao *et al.* 2020, Zhao *et al.* 2020a, 2020b, Divisato *et al.* 2021, Mirzaei *et al.* 2021, Mohanta and Chakrabarti 2021, Orlans *et al.* 2021, Schamberger *et al.* 2021, Tasdelen and Sen 2021, Wong and Rasko 2021, de Amorim *et al.* 2022, Salim *et al.* 2022, Zapletal *et al.* 2022, Lee *et al.* 2023a, 2023b, Amaro *et al.* 2024, Buccheri *et al.* 2024, Gál *et al.* 2024, Khanal *et al.* 2024, Yao *et al.* 2024, Gao *et al.* 2025, Salim *et al.* 2025) (Table S1, available as supplementary data at Bioinformatics online). Next, we manually reviewed the full texts, supplementary materials, and references of all these articles to identify studies reporting novel mirtrons, selecting nine articles (Collins *et al.* 2017, Monteleone and Lutz 2017, Ju *et al.* 2018, Bell *et al.* 2019, Yao *et al.* 2020, Zhao *et al.* 2020a, Mirzaei *et al.* 2021, Amaro *et al.* 2024, Gao *et al.* 2025). For each selected study, we extracted information such as species,

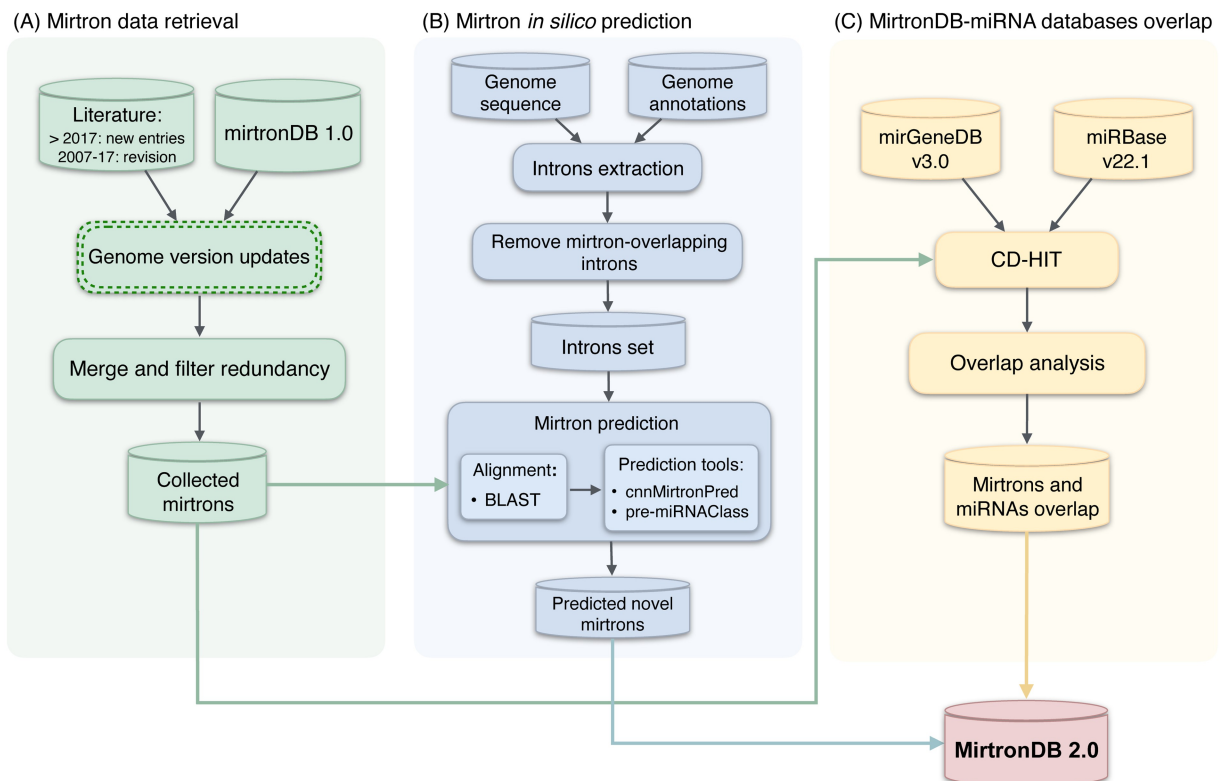


Figure 1 Overview of the mirtronDB 2.0 construction pipeline, comprising three main stages: (A) mirtron data retrieval, including review and update of mirtronDB 1.0 entries and collection of additional mirtrons reported in recent literature; (B) *in silico* prediction of mirtrons using similarity-search and machine learning tools to improve cross-species coverage and standardized annotation; and (C) comparison of retrieved mirtrons from (A) with reference miRNA databases to identify overlapping and unique entries.

genomic coordinates, and genome sequences, including data presented exclusively in supplementary files. Finally, we removed all entries already present in mirtronDB 1.0 by cross-checking species information, genomic coordinates, and sequence data through manual curation (Fig. S1, available as [supplementary data](#) at *Bioinformatics* online).

2.2 mirtronDB updates

The mirtrons included in mirtronDB 1.0 and the newly identified mirtrons were subjected to an update process (Fig. 1A) to their genomic coordinates across all 33 species in the mirtronDB 2.0 (Table S2, available as [supplementary data](#) at *Bioinformatics* online). First, we assessed whether the genome assembly of each species was updated according to the most recent version available in the Ensembl database. For cases where the assembly was outdated, we followed this workflow (Fig. S2 and Document S1, available as [supplementary data](#) at *Bioinformatics* online):

- 1) The genomic coordinates were updated according to the most recent genome assembly available in Ensembl.
- 2) If the mirtron sequence was 100% similar in the previous genome assembly and at least 90% similar in the updated assembly, we chose to update the sequence. The 90% similarity threshold was chosen to tolerate assembly-related variation while still preserving the integrity of mirtron annotation.
- 3) If these criteria were not satisfied, we opted to use the second most updated genome assembly.

In addition to the described workflow, the presence of paralogs was also identified (Document S1, available as [supplementary data](#) at *Bioinformatics* online). When the same sequence was detected at two or more distinct genomic loci, it was classified as a paralog, and its name was updated accordingly (e.g. mirtron-1, mirtron-2) to reflect its multiple genomic occurrences.

2.3 Mirtron *in silico* prediction

To achieve a comprehensive prediction of potential cross-species coverage of similar mirtrons, we developed a computational pipeline (Fig. 1B) that begins with the extraction of intronic sequences based on genome assemblies and transcript annotations from the Ensembl database. For each transcript, exons were identified, and introns were defined as the regions between adjacent exons. To eliminate redundancy caused by transcript isoforms, duplicate introns were removed by comparing their genomic coordinates, strand, and chromosome. Intronic sequences were then retrieved using the *getfasta* command from Bedtools (Quinlan and Hall 2010) with default parameters. Finally, introns overlapping known mirtrons were excluded using Bedtools *intersect*, with strandedness enforced (parameter *-s*) and all other parameters set to default.

Next, sequence similarity searches were performed with BLAST (version 2.16.0+, NCBI), using the full mirtronDB 2.0 (all species) as the subject database and the extracted introns as queries. Alignments were filtered applying thresholds of at least 98% identity and coverage, where coverage was defined as the ratio between alignment length and subject length. To ensure consistency with known mirtron characteristics, candidates with

intron lengths exceeding the maximum precursor length observed among known mirtrons (177 bp) were removed. These candidate introns filtered by length were then classified using two machine learning-based mirtron prediction tools: *cnnMirtronPred* (Zheng *et al.* (2019) available at <https://github.com/zhengxueming/cnnMirtronPred> (commit ID: f2f22f0), and *pre-miRNAClassification* (Aziz and Hasan 2019) available at <https://github.com/zahide-mon/Pre-miRNA-Classification> (commit ID: c02f743). Users could also download both tools from our FigShare (<https://doi.org/10.6084/m9.figshare.29344775>). The classification was performed using the pre-trained models available in their respective repositories, without any additional model training and using only the default execution parameters. The candidate introns were provided as input to the tools, which subsequently classified each sample as a mirtron or canonical miRNA. The selection of the two tools builds on previous work that analyzed the state of the art in mirtron classification (Gaspar Chiquitto *et al.* 2025). To account for distinct levels of prediction evidence, we adopted a confidence prediction scheme based on the agreement between the machine learning tools. Predictions supported by both tools (intersection) were classified as high confidence, whereas predictions identified by only one classifier were labeled as moderate confidence. This strategy allows a broader inclusion of candidate mirtrons while preserving interpretability of prediction reliability.

This procedure was carried out separately for each species. The analysis was conducted on six mammalian species: *B. taurus*, *C. familiaris*, *H. sapiens*, *M. musculus*, *M. mulatta*, and *P. troglodytes*.

2.4 Mirtrons and miRNAs similarity analysis

To assess the contribution of mirtronDB 2.0 in the context of existing miRNA annotations, we performed a comparative analysis between the entries from mirtronDB collected only from literature (Fig. 1A) and two widely used reference databases of annotated miRNAs: miRBase v22.1 (Kozomara *et al.* 2019) and MirGeneDB v3.0 (Clarke *et al.* 2025) (Fig. 1C). We used only mirtrons collected from literature because predicted mirtrons do not have mature information. This analysis served a dual purpose: (i) to quantify the extent of unique content in MirtronDB 2.0 by assessing how many mature mirtrons are not present in miRNA resources, thereby highlighting its novelty and (ii) to systematically annotate mirtrons that overlap with entries in these references, providing users with cross-references to their entries in other miRNA databases.

To identify overlapping sequences, pairwise sequence comparisons were performed using the CD-HIT-EST-2D tool (Huang *et al.* 2010), aligning mature mirtron sequences against miRBase and MirGeneDB entries. To ensure high-confidence matches, we applied a stringent threshold requiring a minimum of 9 nucleotides of overlap (parameter *-n* 9) and at least 98% sequence identity (parameter *-c* 0.98). Given the short length of mature miRNAs, the 98% identity ensures that only highly similar and potentially functionally related sequences are matched.

2.5 Target gene prediction

The human mirtron-mRNA interactions were predicted using the miRanda algorithm (Enright *et al.* 2003). Mature human

mirtron sequences were retrieved from mirtronDB 2.0, and the 3'UTR sequences, used as the target dataset, were obtained from Ensembl BioMart (Kinsella *et al.* 2011).

The analysis was conducted with a minimum alignment score threshold of 140, a maximum free energy cutoff of -10 kcal/mol, and the strict seed pairing requirement to ensure high-confidence predictions.

3 Results

3.1 mirtronDB 2.0: database content

mirtronDB 2.0 presents a significant update from its previous version by incorporating new findings from the literature. Compared to version 1, the database expanded from 18 to 33 species and increased the number of mirtron entries from 1475 to 1589 (Table 1). These advances consolidate mirtronDB as a central resource for investigating mirtrons across species.

3.2 Mirtron in silico prediction

The *in silico* prediction pipeline to improve cross-species coverage of similar mirtrons revealed substantial variability across species, both in the number of introns retained after alignment analysis and in the level of agreement between the classification tools (Table 2). Overall, the final number of predicted mirtrons

Table 1 Summary of content and functional improvements from the first (1.0) to the latest release (2.0) of mirtronDB.

	mirtronDB 1.0	mirtronDB 2.0
Articles	20	31
Species	18	33
Mirtrons entries	1475	1589
Mature mirtrons	2426	2470
Mirtron precursors	1407	1531
Predicted mirtrons	–	165
Cross-reference	miRBase	miRBase and MirGeneDB
Website	Non-interactive Statistics	Dashboard

Table 2 Results of each step in the *in silico* pipeline for the prediction of potential novel mirtrons.^a

Species	BLAST	cnnMirt	miRNAClass	High	Moderate
<i>H. sapiens</i>	242	6	13	4	2
<i>B. taurus</i>	4	–	–	–	–
<i>C. familiaris</i>	3	–	–	–	–
<i>M. mulatta</i>	78	18	22	18	2
<i>M. musculus</i>	74	2	3	2	1
<i>P. troglodytes</i>	531	130	150	123	13

^a BLAST shows introns filtered by sequence similarity; cnnMir and miRNAClass report predictions obtained independently by each classifier, cnnMirtronPred and pre-miRNAClassification, respectively; and the High and Moderate columns represent the confidence levels of the final predictions, indicating the number of mirtrons classified with high or moderate confidence.

was relatively low for most species, not only for high-confidence predictions (agreement between both tools), but also when considering moderate-confidence predictions. For example, in *H. sapiens*, 242 candidate introns passed the BLAST analysis; pre-miRNAClassification predicted up to 13 mirtrons, while cnnMirtronPred predicted 6. However, only 4 were identified with high confidence, highlighting the strictness of this agreement, and 2 with moderate confidence. In *M. mulatta*, the classifiers showed strong agreement, with 18 high-confidence mirtrons and 2 moderate-confidence predictions. A similar pattern was observed in *P. troglodytes*, which presented the highest number of introns selected from BLAST analysis (531 introns) and a high number of predicted mirtrons: 130 by cnnMirtronPred, 150 by pre-miRNAClassification, and 123 classified with high confidence. In contrast, *M. musculus* showed fewer predictions, with only 2 mirtrons confirmed by both tools from a smaller candidate set (74 introns).

For *H. sapiens* and *M. musculus*, the predicted introns showed sequence similarity exclusively with mirtrons annotated from their respective species. In contrast, for *M. mulatta* and *P. troglodytes*, sequence similarity across species was a key determinant factor. Specifically, in *M. mulatta*, among the 18 mirtrons predicted with high confidence, only two showed similarity with known *M. mulatta* mirtrons, while 14 were identified based on similarity to *H. sapiens* mirtrons, and 2 showed similarity to both *H. sapiens* and *P. troglodytes*. For *P. troglodytes*, considering the 123 mirtrons predicted with high confidence, 122 were identified through similarity with *H. sapiens* mirtrons, highlighting the strong conservation of mirtron sequences between these two species.

For *C. familiaris* and *B. taurus*, only 3 and 4 introns, respectively, were retrieved from the BLAST step. However, none of these sequences passed the length filter based on the maximum precursor size threshold, and therefore they were not included in the classification analysis using the prediction tools.

3.3 Mirtron availability in miRNA databases

Our findings reveal a relatively low degree of overlap between mirtronDB 2.0 and the reference databases. Specifically, only 950 (38.5%) of mature mirtrons were found only in miRBase, 3 (0.1%) only in MirGeneDB, and 11 (0.4%) mature mirtrons were found in both databases (Table 3). These results show that a substantial portion of the mature mirtrons in mirtronDB are not covered by widely referenced miRNA databases, highlighting the originality and importance of the data offered by mirtronDB.

4 Implementation, web portal, and statistics

4.1 Website and download

The mirtronDB web portal offers user-friendly access to comprehensive data, built on PostgreSQL 11.7 (Debian 11.7-0 +deb10u1), PHP 7.3.19-1, and JavaScript ES6. All information is organized by species and available in multiple file formats. Users can download data as detailed CSV files (including all relevant information), GFF annotation files, BED files, and FASTA

Table 3 Mature mirtrons availability in miRBase, and MirGeneDB, and data exclusively presented in mirtronDB.

Organism Group	miRBase v22.1		MirGeneDB v3.0		miRBase and MirGeneDB		mirtronDB	Exclusively available in mirtronDB	
	#	%	#	%	#	%	#	#	%
<i>H. sapiens/M. musculus</i>	867	39.9	2	0.1	0	0.0	2173	1304	60
Other chordates	24	41.4	1	1.7	0	0.0	58	33	56.9
Invertebrate	56	71.8	0	0.0	11	14.1	78	11	14.1
Plant	3	2.2	0	0.0	0	0.0	139	136	97.8
Protist	0	0.0	0	0.0	0	0.0	22	22	100
Total	950	38.5	3	0.1	11	0.4	2470	1506	61.0

sequence files. Downloads can be customized to include precursor sequences, mature sequences, or both. All the data for versions 1.0 and 2.0 are available in the Figshare repository (see the Data availability section).

4.2 Dashboard and network

We have also upgraded the mirtronDB web portal to include graphical visualizations of the database. mirtronDB now features three interactive pie charts displaying the number of mature and precursor mirtrons by species, as well as the number of articles included in our dataset, grouped by year of publication. Additionally, a phylogenetic tree is presented, organized by species groups (Chordates, Plants, and Invertebrates). Finally, two interactive box plots are available, showing the sequence length and GC content of the mature mirtrons. For the network, we ran the miRanda tool, and the output totaled >10 million interactions (10 114 676 results). To avoid false positives, we filter and consider only the top 50 best-score interactions for each mature mirtron (final total: 56 950 interactions), which are available on the mirtronDB website and in [Table S3](#), available as [supplementary data](#) at [Bioinformatics](#) online.

5 Discussion and conclusion

The mirtronDB 2.0 provides an expanded and curated resource dedicated to mirtrons, incorporating over 8 years of newly available data. Our systematic curation of the recent literature has included 44 new mature sequences, 124 new precursors, and 15 new species. This expansion provides a more comprehensive overview of the mirtron landscape and facilitates future comparative studies on their evolution and regulatory roles across a wider range of organisms.

A key finding from this update is that mirtronDB 2.0 significantly expands the mirtron landscape, highlighting the originality of our database. Our findings demonstrate that 61% of our curated content is not present in widely used miRNA repositories such as miRBase ([Kozomara et al. 2019](#)) or MirGeneDB ([Clarke et al. 2025](#)). This limited overlap highlights the importance of maintaining a specialized resource for mirtrons up to date, as general miRNA databases may not capture these non-canonical entries due to differences in their biogenesis pathways ([Rorbach et al. 2018](#)). Therefore, mirtronDB 2.0 provides a more complete and accurate view of the mirtron landscape.

To further develop the catalog of mirtrons beyond literature-curated entries, we implemented a computational prediction pipeline for six mammalian species. The combined use of BLAST-based similarity searches and length thresholds may be inherently more restrictive for non-model species, particularly when compared to humans, reflecting methodological bias rather than true biological absence. This approach, based on established mirtron criteria, identified 165 new potential mirtrons that still require experimental validation. The integration of these computationally predicted candidates together with experimentally validated mirtrons provides researchers with a useful set of targets for functional studies, accelerating the discovery of new mirtrons ([Gao et al. 2025](#)). This combined strategy ensures that mirtronDB remains a key resource for mirtron research by offering a comprehensive and up-to-date dataset.

Beyond content growth, we have also focused on enhancing the user experience and accessibility of mirtronDB 2.0. The new website, featuring an interactive dashboard and improved search functionality, allows users to explore the database with greater efficiency. These usability improvements ensure that researchers can quickly access the most relevant data for their studies, reinforcing mirtronDB's role as a user-friendly tool in the field.

mirtronDB 2.0 represents an updated version to help advance our understanding of mirtrons by providing the most comprehensive and standardized resource to date. By combining literature curation with novel *in silico* predictions, we have expanded the known mirtron landscape and highlighted a critical gap in existing miRNA repositories. The new version of the database, with its enriched content and usability, will serve as a reference for researchers studying mirtron biogenesis, function, and evolution. By providing a more comprehensive and standardized resource, mirtronDB 2.0 is expected to catalyze new discoveries, support comparative analyses, and ultimately contribute to a deeper understanding of non-canonical miRNA biology. Moreover, the growing amount of data on mirtrons within this database can positively impact the development of machine learning-based computational methods, enabling more accurate predictions and deeper biological insights.

Acknowledgements

The authors acknowledge the Advanced Research Computing team at the Rosalind Franklin Institute for providing the

computational infrastructure that supported the development of this work.

Supplementary material

Supplementary data are available at *Bioinformatics* online.

Conflicts of interest

None declared.

Funding

The authors acknowledge the support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [440412/2022–6 to A.R.P., #141494/2024–7 to F.L.L.]; Fundação Araucária NAPI Bioinformática [66.2021], and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) PDSE [88881.981333/2024–01] to B.T.L.N., V.G., A.G.C., D.S.D., and A.R.P.; and São Paulo Research Foundation (FAPESP) [2024/10086–4 to M.F.S.].

Data availability

The mirtronDB webserver is available at <http://mirtrondb.utfr.edu.br/>. The complete content of Database 1.0 is freely available in the FigShare repository: <https://figshare.com/s/63316eb72fcd8ef31cf1?file=14062844>. The complete content of Database 2.0 and the source code for the analyses are also freely available in the FigShare repository: https://figshare.com/articles/dataset/MirtronDB_version2/29344775 with DOI: <https://doi.org/10.6084/m9.figshare.29344775>. Requests for additional information or resources should be directed to the lead contact, Alexandre R. Paschoal (paschoal@utfr.edu.br).

Author contributions

Fabiana Rodrigues de Goes (Conceptualization [equal], Data Curation [equal], Formal Analysis [equal], Investigation [equal], Software [equal], Writing—Original Draft [equal], Matheus Fujimura Soares (Data Curation [equal], Validation [equal], Writing—Review & Editing [equal]), Vitor Gregorio (Data Curation [equal], Investigation [equal], Software [equal], Writing—Review & Editing [equal]), Bruno Thiago de Lima Nichio (Methodology [equal], Supervision [equal], Writing—Review & Editing [equal]), Alisson Gaspar Chiquitto (Formal analysis [equal], Writing—Review & Editing [equal]), Flavia Lombardi Lopes (Validation [equal], Writing—Review & Editing [equal]), Mark Basham (Funding Acquisition [equal], Resources [equal], Writing—Review & Editing [equal]), Douglas Silva Domingues (Conceptualization [equal], (Funding Acquisition [equal], Methodology [equal], Supervision [equal], Writing—Review & Editing [equal]), Alexandre Rossi Paschoal (Conceptualization [equal], Funding Acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Supervision [equal], Writing—Review & Editing [equal])

References

- Amaro F, González D, Gutiérrez JC. Micronas in tetrahymena thermophila: an epigenetic regulatory mechanism in the response to cadmium stress. *Microbiol Res* 2024;**280**:127565.
- Amourda C, Saunders TE. The mirtron mir-1010 functions in concert with its host gene skip to balance elevation of *nacrβ2*. *Sci Rep* 2020;**10**:1688.
- Aziz AZB, Hasan MAM. A mixed convolutional neural network for pre-miRNA classification. In: *2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh*. p.217–220. IEEE, 2019.
- Baghdadi MB, Firmino J, Soni K et al. Notch-induced mir-708 antagonizes satellite cell migration and maintains quiescence. *Cell Stem Cell* 2018;**23**:859–68.e5.
- Bell J, Larson M, Kutzler M et al. Mirwoods: enhanced precursor detection and stacked random forests for the sensitive detection of microns. *PLoS Comput Biol* 2019;**15**:e1007309.
- Buccheri V, Pasulka J, Malik R et al. Functional canonical rna in mice expressing a truncated dicer isoform and long dsrna. *EMBO Rep* 2024;**25**:2896–913.
- Campbell NV, Weitzenkamp DA, Campbell IL et al. “omics” data integration and functional analyses link enoyl-coa hydratase, short chain 1 to drug refractory dilated cardiomyopathy. *BMC Med Genomics* 2018;**11**:110. <https://doi.org/10.1186/s12920-018-0439-6>
- Cheng L, Li F, Jiang Y et al. Structural insights into a unique preference for 3’ terminal guanine of mirtron in *drosophila* tutase tailor. *Nucleic Acids Res* 2019;**47**:495–508.
- Clarke AW, Høye E, Hembrom AA et al. Mirgenedb 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microns families and updated covariance models. *Nucleic Acids Res* 2025;**53**:D116–D128.
- Collins DH, Mohorianu I, Beckers M et al. Microns associated with caste determination and differentiation in a primitively eusocial insect. *Sci Rep* 2017;**7**:45674.
- da Fonseca BHR, Domingues DS, Paschoal AR. Mirtrondb: a mirtron knowledge base. *Bioinformatics* 2019;**35**:3873–4.
- de Amorim TS, Pedro DLF, Paschoal AR. *MicroRNA Databases and Tools*. New York, NY, US: Springer, 2022, 131–66. https://doi.org/10.1007/978-1-0716-1170-8_7
- Dexheimer PJ, Wang J, Cochella L. Two microns are sufficient for embryonic patterning in *c. elegans*. *Curr Biol* 2020;**30**:5058–65.e5.
- Divisato G, Piscitelli S, Elia M et al. Microns and stem-like properties: the complex regulation underlying stemness maintenance and cancer development. *Biomolecules* 2021;**11**:1074.
- Dyer SC, Austine-Orimoloye O, Azov AG et al. Ensembl 2025. *Nucleic Acids Res* 2025;**53**:D948–D957.
- Enright A, John B, Gaul U et al. Microns targets in *drosophila*. *Genome Biol* 2003;**4**:P8.
- Gál L, Schamberger A, Wachtl G et al. The effect of alternative splicing sites on mirtron formation and arm selection of precursor microns. *Int J Mol Sci* 2024;**25**:7643.
- Gao Y, Yang L, Guo Y et al. Characterization. functional exploration, and evolutionary analysis of mirtronic microns reveal their origin in the invasive vector mosquito. *Aedes Albopictus* 2025:1–20.

- Gaspar Chiquitto A, Oliveira LS, Bugatti PH *et al.* Gennus: generative approaches for nucleotide sequences enhance mirtron classification. *NAR Genom Bioinform* 2025;**7**:lqaf072. <https://doi.org/10.1093/nargab/lqaf072>
- Hansen TB, Detecting agotrons in ago clipseq data. *miRNA Biogenesis: Methods and Protocols*. New York, NY: Springer, Vol. 1823, 2018, 221–32.
- Huang Y, Niu B, Gao Y *et al.* Cd-hit suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 2010; **26**:680–2.
- Ju Z, Jiang Q, Liu G *et al.* Solexa sequencing and custom microRNA chip reveal repertoire of micrornas in mammary gland of bovine suffering from natural infectious mastitis. *Anim Genet* 2018;**49**:3–18.
- Khanal S, de Cruz M, Strickland B *et al.* A tailed mirtron promotes longevity in drosophila. *Nucleic Acids Res* 2024;**52**:1080–9.
- Kim JH, Park JS, Lee CY *et al.* Dissecting seed pigmentation-associated genomic loci and genes by employing dual approaches of reference-based and k-mer-based gwas with 438 glycine accessions. *PLoS One* 2020;**15**:e0243085.
- Kinsella RJ, Kähäri A, Haider S *et al.* Ensembl biomarts: a hub for data retrieval across taxonomic space. *Database* 2011; **2011**:bar030.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. Mirbase: from microRNA sequences to function. *Nucleic Acids Res* 2019; **47**:D155–D162.
- Kroupova A, Ivascu A, Reimão-Pinto MM *et al.* Structural basis for acceptor rna substrate selectivity of the 3' terminal uridylyl transferase tailor. *Nucleic Acids Res* 2019;**47**:1030–42.
- Lee S, Jee D, Srivastava S *et al.* Promiscuous splicing-derived hairpins are dominant substrates of tailing-mediated defense of mirna biogenesis in mammals. *Cell Rep* 2023a;**42**:112111.
- Lee SH, Ng CX, Wong SR *et al.* Mirnas overexpression and their role in breast cancer: implications for cancer therapeutics. *Curr Drug Targets* 2023b;**24**:484–508.
- Mirzaei R, Babakhani S, Ajorloo P *et al.* The emerging role of exosomal mirnas as a diagnostic and therapeutic biomarker in mycobacterium tuberculosis infection. *Mol Med* 2021;**27**:34.
- Mohammed J, Flynt AS, Panzarino AM *et al.* Deep experimental profiling of microRNA diversity, deployment, and evolution across the drosophila genus. *Genome Res* 2018;**28**:52–65.
- Mohanta A, Chakrabarti K. Dbr1 functions in mrna processing, intron turnover and human diseases. *Biochimie* 2021; **180**:134–42.
- Monteleone NJ, Lutz CS. Mir-708-5p: a microRNA with emerging roles in cancer. *Oncotarget* 2017;**8**:71292–316.
- Orlans HO, McClements ME, Barnard AR *et al.* Mirtron-mediated rna knockdown/replacement therapy for the treatment of dominant retinitis pigmentosa. *Nat Commun* 2021;**12**:4934.
- Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2.
- Rivera B, Nadaf J, Fahiminiya S *et al.* Dgcr8 microprocessor defect characterizes familial multinodular goiter with schwannomatosis. *J Clin Invest* 2020;**130**:1479–90.
- Rorbach G, Unold O, Konopka BM. Distinguishing mirtrons from canonical mirnas with data exploration and machine learning methods. *Sci Rep* 2018;**8**:7560.
- Salim U, Kumar A, Kulshreshtha R *et al.* Biogenesis, characterization, and functions of mirtrons. *Wiley Interdiscip Rev RNA* 2022;**13**:e1680.
- Salim U, Menon MB, Dhamija S *et al.* Rna g-quadruplexes regulate mammalian mirtron biogenesis. *J Biol Chem* 2025; **301**:108276.
- Schamberger A, Sarkadi B, Orbán TI. Human mirtrons can express functional micrornas simultaneously from both arms in a flanking exon-independent manner. *RNA Biol* 2012; **9**:1177–85.
- Schamberger A, Várady G, Fóthi Á *et al.* Posttranscriptional regulation of the human abcg2 multidrug transporter protein by artificial mirtrons. *Genes (Basel)* 2021;**12**:1068.
- Tasdelen A, Sen B. A hybrid cnn-lstm model for pre-mirna classification. *Sci Rep* 2021;**11**:14125.
- Titov II, Vorozheykin PS. Comparing mirna structure of mirtrons and non-mirtrons. *BMC Genomics* 2018;**19**:114.
- Urbanek-Trzeciak MO, Jaworska E, Krzyzosiak WJ. Mirnamotif—a tool for the prediction of pre-mirna–protein interactions. *Int J Mol Sci* 2018;**19**:4075.
- Wong AC, Rasko JE. Splice and dice: intronic micrornas, splicing and cancer. *Biomedicines* 2021;**9**:1268.
- Yao J, Wu DC, Nottingham RM *et al.* Identification of protein-protected mrna fragments and structured excised intron rnas in human plasma by tgirt-seq peak calling. *Elife* 2020; **9**:e60743.
- Yao J, Xu H, Ferrick-Kiddie EA *et al.* Human cells contain myriad excised linear intron rnas with links to gene regulation and potential utility as biomarkers. *PLoS Genet* 2024;**20**:e1011416.
- Zapletal D, Taborska E, Pasulka J *et al.* Structural and functional basis of mammalian microRNA biogenesis by dicer. *Mol Cell* 2022;**82**:4064–79.e13.
- Zhao B, Fan C, Jin Y. Identification of novel mirnas in mouse embryonic stem cells and reprogrammed pluripotent cells. *Acta Biochim Biophys Sin (Shanghai)* 2020a;**52**:1040–3.
- Zhao M, Hou Y, Du Y *et al.* Drosha-independent mir-6778–5p strengthens gastric cancer stem cell stemness via regulation of cytosolic one-carbon folate metabolism. *Cancer Lett* 2020b; **478**:8–21.
- Zheng X, Xu S, Zhang Y *et al.* Nucleotide-level convolutional neural networks for pre-mirna classification. *Sci Rep* 2019;**9**:628.
- Zia MF, Flynt AS. Detection and verification of mammalian mirtrons by Northern blotting. *miRNA Biogenesis: Methods and Protocols*. New York, NY: Springer, 2018, 209–19.