

piRNAclusterDB 2.0 - update and expansion of the piRNA cluster database

David Rosenkranz^{1,2}, Hans Zischler¹ and Daniel Gebert^{1,3,*}

¹ Institute of Organismic and Molecular Evolution, Anthropology, Johannes Gutenberg University, Mainz 55099, Germany

² Senckenberg Centre for Human Genetics, Facharztzentrum Frankfurt-Nordend gGmbH, Frankfurt am Main 60314, Germany

³ Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom

* To whom correspondence should be addressed. Email: dg572@cam.ac.uk

ABSTRACT

PIWI-interacting RNAs (piRNAs) and their partnering PIWI proteins defend the animal germline against transposable elements and play a crucial role in fertility. Numerous studies in the past have uncovered many additional functions of the piRNA pathway, including gene regulation, anti-viral defense, and somatic transposon repression. Further, comparative analyses across phylogenetic groups showed that the PIWI/piRNA system evolves rapidly and exhibits great evolutionary plasticity. However, the presence of so-called piRNA clusters as the major source of piRNAs is common to nearly all metazoan species. These genomic piRNA-producing loci are highly divergent across taxa and critically influence piRNA populations in different evolutionary lineages. We launched the initial version of the piRNA cluster database to facilitate research on regulation and evolution of piRNA-producing loci across tissues and species. In recent years the amount of small RNA sequencing data that was generated and the abundance of species that were studied has grown rapidly. To keep up with this recent progress, we have released a major update for the piRNA cluster database (<https://www.smallrnagroup.uni-mainz.de/piRNAclusterDB>), expanding it from 12 to a total of 51 species with hundreds of new datasets, and revised its overall structure to enable easy navigation through this large amount of data.

INTRODUCTION

PIWI proteins and piRNAs represent a mainly metazoan system for the regulation of a range of target sequences (1-4), including transposable elements (5,6), protein-coding genes (7-11) and long non-coding RNAs (12). These targets can be regulated transcriptionally (13), as well as post-transcriptionally (6,14) and are recognized through sequence complementarity by piRNAs, which guide their associated PIWI proteins to their destination. While in vertebrates the PIWI/piRNA pathway is mostly restricted to the germline, in invertebrate groups such as arthropods and mollusks, piRNAs are in addition ubiquitously found in somatic tissues (15,16). Moreover, PIWI proteins and piRNAs were identified in somatic stem cells of sponges and cnidarians (17,18).

In general, piRNA-producing loci, called piRNA clusters, are considered to lie at the very center of the PIWI pathway (6). These loci are transcribed from one or from both DNA strands into large precursor RNAs, which in turn are processed into 23-31 nucleotide (nt) mature piRNAs as they are loaded onto PIWI proteins. Further, the ping pong amplification cycle, which acts during post-transcriptional silencing, additionally contributes to piRNA biogenesis (6,14). In a typical metazoan genome, up to a few hundred piRNA clusters can be identified, ranging in their size between a few thousand base pairs (kb) to more than 100 kb. Though these regions make up overall only small portions of a genome with 0.1 to 5%, they produce the vast majority of piRNAs. In *Drosophila*, over 90% of all sequenced germline piRNAs can be derived from these genomic loci (19). Similarly, in mammals, up to 95% of pachytene piRNAs are produced from clusters, while still more than 55% of pre-pachytene piRNAs can be attributed to these distinct loci (5). Generally, piRNA clusters are more or less dispersed in the genome, though they do not occupy similar regions in different phylogenetic groups. In flies, piRNA precursors stem from mostly pericentric heterochromatic loci (6), whereas in mammals, (pachytene) piRNA-producing loci are euchromatic A-MYB promoter-dependent RNA polymerase II transcription units (20).

In all species studied so far, piRNA clusters evolve rapidly, appearing and disappearing rather quickly on evolutionary time scales and evolving neutrally on the sequence level (21-23). Therefore, each species has a unique set of piRNA clusters with varying numbers of homologous clusters shared between lineages. Noteworthy, it has also been shown that piRNA clusters are highly divergent within species, such as observed in human (24). Additionally, in those clades in which somatic piRNAs are common, piRNA clusters show distinct expression levels in different tissues (19,16), similar to the differential activity at various developmental stages that is observed more broadly (5,13,25,26).

The first release of the piRNA cluster database (27), which was launched to as a central resource for piRNA cluster research, comprised more than 100 Sequence Read Archive (SRA) datasets from 12 species. Noteworthy, while extensive analysis showed that many piRNA databases are contaminated with non-coding RNA (ncRNA) fragments, especially when concerning somatic tissues, the piRNA cluster database stands out with a remarkably low amount of such contaminations (28), due to our stringent criteria for piRNA definition and piRNA cluster identification (29). Further, while different databases for piRNA sequences are maintained (30-32), it is to date still the only database that is dedicated specifically to piRNA clusters. In recent years, the amount of small RNA sequencing data has drastically increased, while our understanding of the piRNA pathway and the expression of piRNAs has grown continuously. In order to incorporate this progress into the piRNA cluster database, we have now released a major update, which includes more than 350 SRA datasets from 51 species, comprising more than 15,000 piRNA clusters in total. This set of species contains mollusks, arthropods, fishes, amphibians, reptiles, birds, and mammals. We further significantly improved our small RNA transcriptome analysis, which now includes an extensive set of non-coding RNAs, and we provide bibliographic information, as well as easy access to reference data used in our analysis.

DATA COLLECTION AND ANALYSIS

Dataset search and initial processing

We systematically searched the NCBI sequence read archive (SRA) (33) for candidate small RNA sequencing datasets from a range of different tissue samples from metazoan species for which a reference genome is available. The raw sequence reads were first subjected to processing with unitas version 1.7.5 (34), including adapter trimming and low-complexity read filtering (Figure 1). The clean reads were then mapped with bowtie (35) to the corresponding reference genome, which was obtained from the NCBI Genome resource (36).

The mapped reads were annotated using unitas with reference datasets for coding sequences and non-coding RNAs, such as microRNAs, tRNAs, rRNAs and long non-coding RNAs. To determine the presence of piRNA-like reads, we examined the length distribution, rates of uridine at position 1 (1U) and adenine at position 10 (10A), as well as the rate of 10 nt 5' read overlaps, known as ping-pong signature, of those reads that did not match to annotated ncRNAs in the previous step. The first threshold that was applied is a minimum share of piRNA-sized (23-31 nt) reads of 20%, while the largest peak was required to be within the piRNA size range. Additionally, since piRNAs typically show a high degree of sequence diversity with a considerable amount of unique sequence reads compared to other small RNA (sRNA) classes, a threshold for a minimum of 10% of unique reads within all piRNA-sized reads was employed. Finally, a 1U or 10A rate of at least 35% and a ping-pong signature with a significant ($p < 0.05$) z-score of more than 1.65 (37) were required to pass our filter.

Genome mapping and piRNA cluster prediction

In preparation for the identification of piRNA clusters, we removed all previously annotated reads, as well as all reads that do not fall into the piRNA size range of 23-31 nt from the initial map file. A resulting map file with uniquely-mapping filtered reads without seed mismatches, produced by bowtie (35), and the corresponding genome was then used, along with an available repeatmasker annotation and a GFF gene set that were obtained from the NCBI Genome resource, to predict piRNA clusters with proTRAC version 2.4.4 (29). We applied a minimum cluster size of 5 kb and a minimum rate of 1U or 10A of 50%. Further, we used a sliding window size of 5 kb with an increment of 1kb and defined a p-value of 0.01 for minimum number of normalized read counts per kb. To ensure similar cluster prediction power in all species, no assumptions on strand directionality were made (option: -clstrand 0). Finally, resulting piRNA cluster loci with a distance smaller than 10kb were merged. However, if no piRNA clusters could be identified the dataset was discarded.

Overall, 358 SRA datasets from 51 species passed all filters and could be used for successful piRNA cluster prediction, totaling in 15,857 piRNA-producing loci with a median of 250 per species. The total number of unique clustered piRNA sequences amounts to nearly 88 million with a median of more than 1.1 million reads per species. All custom Perl scripts used in the data processing are available at GitHub (<https://github.com/d-gebert/piRNAclusterDB>).

HOW TO ACCESS THE DATA

Database structure and dataset access

We completely redesigned the overall structure of the piRNA cluster database in order to accommodate to the vast number of datasets and species that we added in this major update. The primary entry point of the interface is the species selector, which is represented as an interactive phylogenetic tree (Figure

2A) on the one hand and as a tabular list on the other hand with additional information on taxonomy, number of piRNA clusters and total amount of clustered piRNA sequences (Figure 2B). The table also provides links to the corresponding genome assembly data that were used in our analysis, including genome, gene set (GFF) and repeatmasker file. A graphical representation of the locations of piRNA clusters on chromosomes is linked to the number of piRNA clusters in each species. Further, files on piRNA cluster coordinates (GTF), sequence (FASTA), and differential piRNA expression in reads per million (RPM) for each dataset, as well as pooled clustered piRNA reads are available for download. In addition to that, we provide a comprehensive list of all publications that are associated with small RNA (sRNA) datasets of the database, including PubMed IDs and direct links (Figure 2C).

Cluster browser

Once a species of interest is selected, the user can browse piRNA-producing loci along all datasets from the chosen species in the cluster browser section (Figure 3). The available loci are provided in a list with selectable piRNA cluster IDs (Figure 3A), which contains additional information on location, size and reads per million. The cluster view incorporates tracks for gene and repeat annotation alongside of piRNA read coverage in rpm for plus and minus strands, which all can be individually inspected to receive further information, for example on gene or transposon name and repeat class, as well as exact rpm per position (Figure 3B). Finally, each SRA dataset can be individually selected and deselected to produce a customized view of piRNA cluster expression across different datasets, tissues, or developmental stages.

SRA dataset section

The third section of the database presents detailed analyses of processed, filtered, mapped, and annotated sRNA reads for each SRA dataset from a selected species, which were generated by unitas (34). SRA datasets are selectable from a list that includes information on the tissue of origin, number of reads and PubMed ID of the associated publication (Figure 4A). For each of the annotated RNA types, such as miRNA, rRNA, tRNA-derived sRNA, lncRNA, mRNA and more, a table offers read counts, as well as links for the download of reads in FASTA format and info files with length distribution and positional nucleotide composition (Figure 4B). The fraction that most likely represents or contains to a large part piRNAs can be accessed under the type "unknown", as these sequences could not be annotated as any other known sRNA. Moreover, graphical output on read composition, length distribution, positional nucleotide composition and ping-pong signature are provided for total sRNA reads and piRNA reads, which gives an accessible insight into the sRNA make-up of each SRA dataset and the contribution of piRNAs to the total pool of reads.

DISCUSSION

Unsurprisingly, the first species in which piRNAs and piRNA clusters were identified and characterized were mice (*Mus musculus*) (1-4) and flies (*Drosophila melanogaster*) (6,14). Subsequent studies focused on other model organisms (38) and eventually extended towards non-model organisms, especially in recent years, including hitherto less well-studied taxa (17,18,25,26,15,16). Besides, many

sRNA studies, e.g. in which miRNAs are the main focus, produced readily available piRNA fractions that were not yet examined.

This has created both the necessity and opportunity to considerably expand and restructure the piRNA cluster database to open up further potential for piRNA research. It is now clear that the PIWI pathway and piRNA-producing loci in particular evolve rapidly, leading to greatly different sets of clusters across taxa (21-23). The vast number of species included in this substantial update of the piRNA cluster database will enable evolutionary studies on an unprecedented scale, including previously less studied phylogenetic groups. In the past, studies in non-model organisms regularly yielded unexpected and novel insights into different aspects of the piRNA pathway (18,25,15,16).

Despite the considerable progress that has been achieved in recent years, many aspects especially concerning the evolution of piRNA clusters are still not fully understood. It has yet to be determined how exactly piRNA clusters initially emerge and what drives their genesis, their maintenance across evolutionary times and their demise. Furthermore, intriguing differences between phylogenetic groups, such as flies and mammals, regarding transposon enrichment or genomic location are likewise still not elucidated. For instance, it is conceivable that the very nature and origin of the transposon-rich, heterochromatic and pericentromeric piRNA clusters of *Drosophila* (6) are inherently different from those of mammalian clusters, which are more dispersed in the genome and less enriched for transposon sequences (5). Similarly, clusters of recently identified somatic piRNAs in non-vertebrates (15,16) remain to be thoroughly studied, to detect putative functions that might be different from those in gonads. Finally, differences in piRNA biology of male and female germline have not yet been widely studied but have so far yielded important insights (39). We believe that these outlined areas of research will greatly benefit from the resources that this update of the piRNA cluster database provides.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Felipe Karam Teixeira and the Human Frontier Science Program (HFSP) for the support (CDA-00032/2018) during the completion of this database update. We gratefully acknowledge the GenEvo RTG funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2526/1 – Projectnr. 407023052 enabling the conception of this research project.

FUNDING

Walter Benjamin Fellowship (GE 3407/1-1) of the German Research Foundation DFG (to D.G.).

Funding for open access charge: DFG Research Training Group (RTG) “GenEvo - Gene Regulation in Evolution: From Molecular to Extended Phenotypes” (GRK 2526/1, Projectnr. 407023052).

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Aravin,A., Gaidatzis,D., Pfeffer,S., Lagos-Quintana,M., Landgraf,P., Iovino,N., Morris,P., Brownstein,M.J., Kuramochi-Miyagawa,S., Nakano,T. et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442, 203-207.
2. Girard,A., Sachidanandam,R., Hannon,G.J. and Carmell,M.A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442, 199-202.
3. Grivna,S.T., Beyret,E., Wang,Z. and Lin,H. (2006) A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.*, 20, 1709-1714.
4. Watanabe,T.,Takeda,A.,Tsukiyama,T.,Mise,K.,Okuno,T., Sasaki,H., Minami,N. and Imai,H. (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.*, 20, 1732-1743.
5. Aravin,A.A., Sachidanandam,R., Girard,A., Fejes-Toth,K. and Hannon,G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316, 744-747.
6. Brennecke,J., Aravin,A.A., Stark,A., Dus,M., Kellis,M., Sachidanandam,R. and Hannon,G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128, 1089-103.
7. Gou,L.T., Dai,P., Yang,J.H., Xue,Y., Hu,Y.P., Zhou,Y., Kang,J.Y., Wang,X., Li,H., Hua,M.M. et al. (2014) Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res.*, 24, 680-700.
8. Zhang,P., Kang,J.Y., Gou,L.T., Wang,J., Xue,Y., Skogerboe,G., Dai,P., Huang,D.W., Chen,R., Fu,X.D. et al. (2015) MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.*, 25, 193-207.
9. Goh,W.S.S., Falcatori,I., Tam,O.H., Burgess,R., Meikar,O., Kotaja,N., Hammell,M. and Hannon,G.J. (2015) piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.*, 29, 1032-1044.
10. Pantano,L., Jodar,M., Bak,M., Ballescà,J.L., Tommerup,N., Oliva,R. and Vavouri,T. (2015) The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA*, 21, 1085-1095.
11. Gebert,D., Ketting,R.F., Zischler,H. and Rosenkranz,D. (2015) piRNAs from pig testis provide evidence for a conserved role of the Piwi pathway in post-transcriptional gene regulation in mammals. *PLOS One*, 10, e0124860.

12. Watanabe,T., Cheng,E.C., Zhong,M. and Lin,H. (2015) Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res.*, 25, 368-380.
13. Aravin,A.A., Sachidanandam,R., Bourc'his,D., Schaefer,C., Pezic,D., Fejes-Toth,K., Bestor,T. and Hannon,G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell*, 31, 785-799.
14. Gunawardane,L.S., Saito,K., Nishida,K.M., Miyoshi,K., Kawamura,Y., Nagami,T., Siomi,H. and Siomi,M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, 315, 1587-1590.
15. Lewis,S.H., Quarles,K.A., Yang,Y., Tanguy,M., Frézal,L., Smith,S.A., Sharma,P.P., Cordaux,R., Gilbert,C., Giraud,I. et al. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.*, 2, 174-181.
16. Jehn,J., Gebert,D., Pipilescu,F., Stern,S., Kiefer,J.S.T., Hewel,C. and Rosenkranz,D. (2018) PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Commun. Biol.*, 1, 137.
17. Funayama,N., Nakatsukasa,M., Mohri,K., Masuda,Y. and Agata,K. (2010) Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. *Evol. Dev.*, 12, 275-287.
18. Juliano,C.E., Reich,A., Liu,N., Götzfried,J., Zhong,M., Uman,S., Reenan,R.A., Wessel,G.M., Steele,R.E. and Lin,H. (2014) PIWI proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. *Proc. Natl. Acad. Sci. U. S. A.*, 111, 337-342.
19. Malone,C.D., Brennecke,J., Dus,M., Stark,A., McCombie,W.R., Sachidanandam,R. and Hannon,G.J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell*, 137, 522-535.
20. Li,X.Z., Roy,C.K., Dong,X., Bolcun-Filas,E., Wang,J., Han,B.W., Xu,J., Moore,M.J., Schimenti,J.C., Weng,Z. et al. (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell*, 50, 67-81.
21. Assis,R. and Kondrashov,A.S. (2009) Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 106, 7079-7082.

22. Chirn,G., Rahman,R., Sytnikova,Y.A., Matts,J.A., Zeng,M., Gerlach,D., Yu,M., Berger,B., Naramura,M., Kile,B.T. et al. (2015) Conserved piRNA expression from a distinct set of piRNA cluster loci in eutherian mammals. *PLOS Genet.*, 11, e1005652.
23. Gebert,D., Zischler,H. and Rosenkranz,D. (2019) Primate piRNA cluster evolution suggests limited relevance of pseudogenes in piRNA-mediated gene regulation. *Genome Biol. Evol.*, 11, 1088-1104.
24. Özata,D.M., Yu,T., Mou,H., Gainetdinov,I., Colpan,C., Cecchini,K., Kaymaz,Y., Wu,P.H., Fan,K., Kucukural,A. et al. (2020) Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol.* 4, 156-168.
25. Praher,D., Zimmermann,B., Genikhovich,G., Columbus-Shenkar,Y., Modepalli,V., Aharoni,R., Moran,Y. and Technau,U. (2017) Characterization of the piRNA pathway during development of the sea anemone *Nematostella vectensis*. *RNA Biol.*, 14, 1727-1741.
26. Llonga,N., Ylla,G., Bau,J., Belles,X. and Piulachs,M.D. (2018) Diversity of piRNA expression patterns during the ontogeny of the German cockroach. *J. Exp. Zool. B. Mol. Dev. Evol.*, 330, 288-295.
27. Rosenkranz,D. (2016) piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids Res.*, 44, D223-D230.
28. Tosar,J.P., Rovira,C. and Cayota,A. (2018) Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol.*, 1, 2.
29. Rosenkranz,D. and Zischler,H. (2012) proTRAC-a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, 13, 5.
30. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, 36, D173-177.
31. Sarkar,A., Maji,R.K., Saha,S. and Ghosh,Z. (2014) piRNAQuest: searching the piRNAome for silencers. *BMC Genomics*, 15, 555.
32. Wang,J., Zhang,P., Lu,Y., Li,Y., Zheng,Y., Kan,Y., Chen,R. and He,S. (2019) piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.*, 47, D175-D180.
33. Kodama,Y., Shumway,M., Leinonen,R., International Nucleotide Sequence Database Collaboration. (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40, D54-D56.

34. Gebert,D., Hewel,C. and Rosenkranz,D. (2017) unitas: the universal tool for annotation of small RNAs. BMC Genomics, 18, 644.
35. Langmead,B., Trapnell,C., Pop,M., and Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol., 10, R25.
36. Sayers,E.W., Beck,J., Bolton,E.E., Bourexis,D., Brister,J.R., Canese,K., Comeau,D.C., Funk,K., Kim,S., Klimke,W. et al. (2021) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 49, D10-D17.
37. Zhang,Z., Xu,J., Koppetsch,B.S., Wang,J., Tipping,C., Ma,S., Weng,Z., Theurkauf,W.E. and Zamore,P.D. (2011) Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol. Cell, 44, 572–584.
38. Lau,N.C., Seto,A.G., Kim,J., Kuramochi-Miyagawa,S., Nakano,T., Bartel,D.P. and Kingston,R.E. (2006) Characterization of the piRNA complex from rat testes. Science, 313, 363-367.
39. Chen,P., Kotov,A.A., Godneeva,B.K., Bazylev,S.S., Olenina,L.V. and Aravin,A.A. (2021) piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in *D. melanogaster* male germline. Genes Dev. doi: 10.1101/gad.345041.120. Epub ahead of print.

FIGURE LEGENDS

Figure 1. Processing and analysis pipeline for SRA dataset analysis and piRNA cluster prediction. The highlighted final results are accessible through the piRNA cluster database.

Figure 2. Top section and primary entry point of the piRNA cluster database. (A) Interactive phylogenetic tree, comprising the 51 species that are included in the database. (B) Complete tabular species list with quantitative information on piRNAs and piRNA clusters and associated download links, as well as links to the assemblies used in the analysis. (C) List of publications underlying the SRA datasets with PubMed IDs.

Figure 3. Cluster browser section of the piRNA cluster database. (A) List of all piRNA cluster loci of the selected species with coordinates, size, and maximum rpm. (B) Browser with gene and repeat annotation tracks and piRNA clusters read coverage. Track elements are selectable for additional information.

Figure 4. SRA dataset section of the piRNA cluster database. (A) List of SRA datasets of the selected species with information on tissue, data size, read count and PubMed ID. (B) Annotation of SRA datasets with download links for each RNA class and graphical output of analysis on read composition, positional nucleotide frequencies, length distribution and ping-pong signature (sense/antisense 5' overlap lengths).

Figure 1

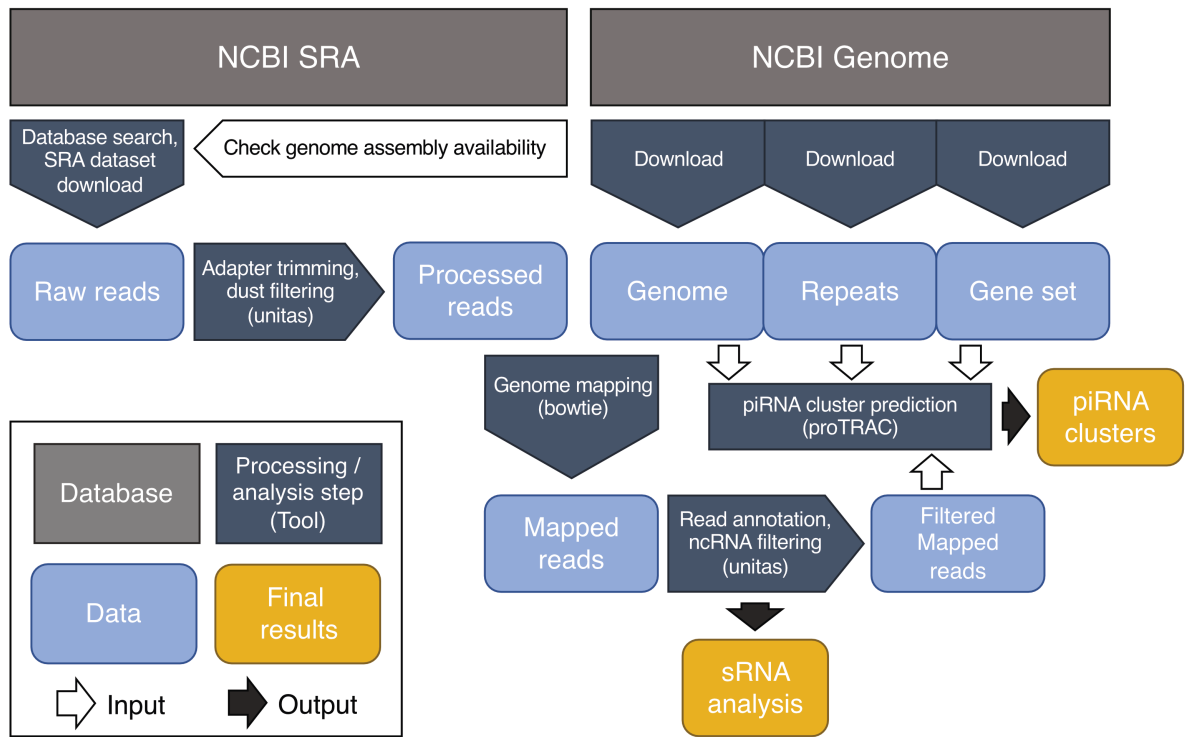
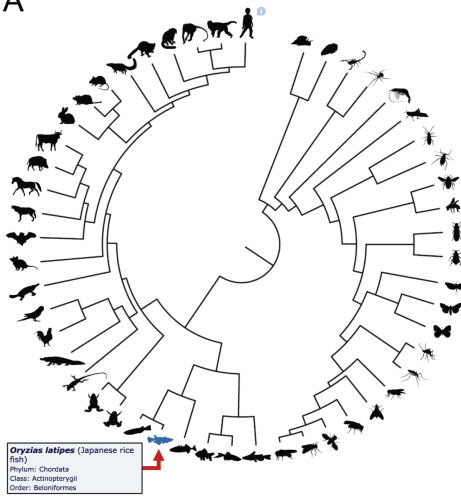


Figure 2

A



B

SPESIES ^D

Show entries

Search:

Species (click to select)	Common name	Phylum	Class	Order	Assembly	piRNA clusters	Clustered piRNAs
<i>Monodelphis domestica</i>	Gray short-tailed opossum	Chordata	Mammalia	Didelphimorphia	MonDom3	135 .fas .gff .rpm	157410
<i>Mus musculus</i>	House mouse	Chordata	Mammalia	Rodentia	GRCm38.p6	939 .fas .gff .rpm	4522284
<i>Musca domestica</i>	Housefly	Arthropoda	Insecta	Diptera	Musca_domestica-2.0.2	107 .fas .gff .rpm	627281
<i>Nicrophorus vespilloides</i>	Herbert's burying beetle	Arthropoda	Insecta	Coleoptera	Nicve_v1.0	174 .fas .gff .rpm	479204
<i>Oncopeltus fasciatus</i>	Large milkweed bug	Arthropoda	Insecta	Hemiptera	Ofas_2.0	119 .fas .gff .rpm	1286320
<i>Ornithorynchus anatinus</i>	Platypus	Chordata	Mammalia	Monotremata	Ornithorynchus_anatinus_S.D.1	101 .fas .gff .rpm	65741
<i>Oryctolagus cuniculus</i>	European rabbit	Chordata	Mammalia	Lagomorpha	OryCun2.0	451 .fas .gff .rpm	1290642
Oryzias latipes	Japanese rice fish	Chordata	Actinopterygii	Belontiiformes	ASM223467.v1	168 .fas .gff .rpm	1070683
<i>Oryzias melastigma</i>	Blue-eyed rice fish	Chordata	Actinopterygii	Belontiiformes	ASM292280.v2	258 .fas .gff .rpm	1833580
<i>Pararge aegeria</i>	Speckled wood	Arthropoda	Insecta	Lepidoptera	ilParAeg1.1	149 .fas .gff .rpm	379954

Showing 31 to 40 of 51 entries

Previous 1 2 3 **4** 5 6 Next

C

PUBLICATIONS ^D

Show entries

Search:

Authors	Title (click to select SRA datasets)	Journal	Bibliographic	Species	PubMed ID
Gay S, Bugeon J, Bouchareb A, Henry L, Delahaye C, Legasi F, Montfort J, Le Cam A, Siegel A, Bobe J, Thiernes V	MIR-202 controls female fecundity by regulating medaka oogenesis.	PLoS Genet	2018 14:e1007593	Oryzias latipes	30199527
Lai KP, Li JW, Wang SY, Chiu JM, Tse A, Lau K, Luk S, Au DW, Tse WK, Wong CK, Chan TF, Kong RY, Wu RS	Tissue-specific transcriptome assemblies of the marine medaka Oryzias melastigma and comparative analysis with the freshwater medaka Oryzias latipes.	BMC Genomics	2015 16:135	Oryzias melastigma	25765076

Showing 1 to 2 of 2 entries (filtered from 83 total entries)

Previous Next

Figure 3

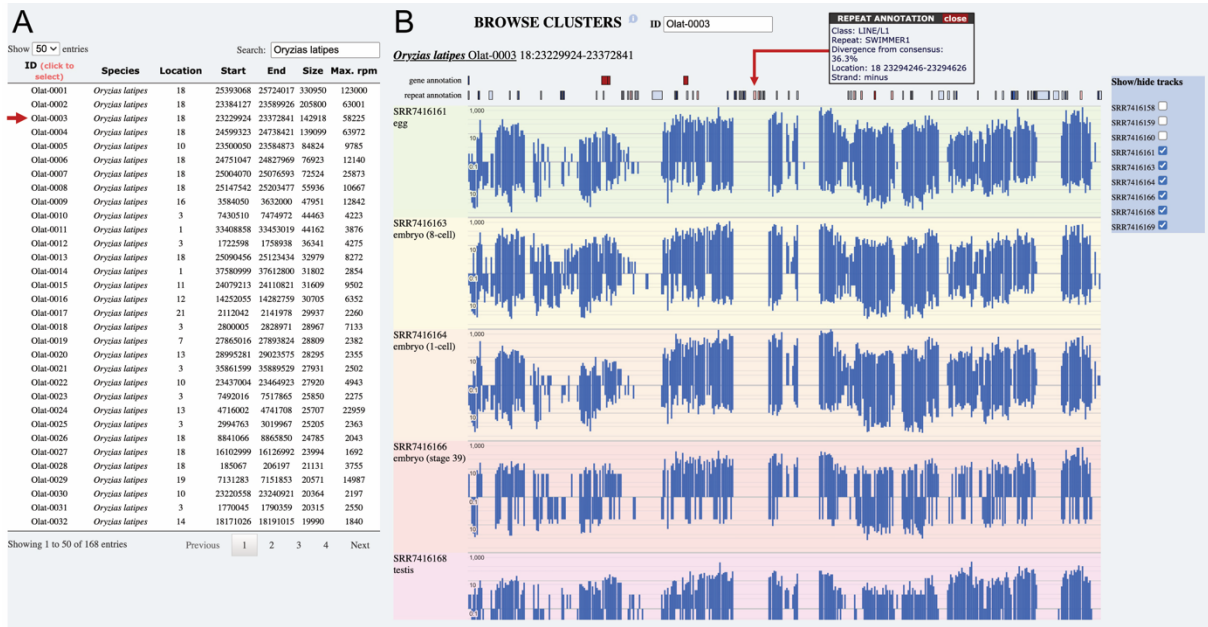


Figure 4

A

Show 20 entries Search: *Oryzias latipes*

Species	Tissue	ID	Size [Mb]	Reads	PubMed ID
<i>Oryzias latipes</i>	ovary	SRR7416169	504	8298060	30199527
<i>Oryzias latipes</i>	testis	SRR7416168	367	5718125	30199527
<i>Oryzias latipes</i>	egg	SRR7416161	723	11205365	30199527
<i>Oryzias latipes</i>	embryo (1-cell)	SRR7416164	496	7558158	30199527
<i>Oryzias latipes</i>	embryo (8-cell)	SRR7416163	481	6351845	30199527
<i>Oryzias latipes</i>	embryo (stage 15-16)	SRR7416158	419	4723399	30199527
<i>Oryzias latipes</i>	embryo (stage 31-32)	SRR7416160	371	6578967	30199527
<i>Oryzias latipes</i>	embryo (stage 31-36)	SRR7416159	456	8462769	30199527
<i>Oryzias latipes</i>	embryo (stage 39)	SRR7416166	354	6388584	30199527

Showing 1 to 9 of 9 entries (filtered from 358 total entries) Previous 1 Next

B

SRA DATASET SECTION

