Published online 17 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D223–D230***
doi: 10.1093/nar/gkv1265

piRNA cluster database: a web resource for piRNA producing loci

David Rosenkranz*

Institute of Anthropology, Johannes Gutenberg University, Mainz 55099, Germany

Received August 15, 2015; Revised October 30, 2015; Accepted November 03, 2015

ABSTRACT

Piwi proteins and their guiding small RNAs, termed Piwi-interacting (pi-) RNAs, are essential for silencing of transposons in the germline of animals. A substantial fraction of piRNAs originates from genomic loci termed piRNA clusters and sequences encoded in these piRNA clusters determine putative targets for the Piwi/piRNA system. In the past decade, studies of piRNA transcriptomes in different species revealed additional roles for piRNAs beyond transposon silencing, reflecting the astonishing plasticity of the Piwi/piRNA system along different phylogenetic branches. Moreover, piRNA transcriptomes can change drastically during development and vary across different tissues.

Since piRNA clusters crucially shape piRNA profiles, analysis of these loci is imperative for a thorough understanding of functional and evolutionary aspects of the piRNA pathway. But despite the ever-growing amount of available piRNA sequence data, we know little about the factors that determine differential regulation of piRNA clusters, nor the evolutionary events that cause their gain or loss.

In order to facilitate addressing these subjects, we established a user-friendly piRNA cluster database (http://www.smallrnagroup-mainz.de/piRNAclusterDB.html) that provides comprehensive data on piRNA clusters in multiple species, tissues and developmental stages based on small RNA sequence data deposited at NCBI's Sequence Read Archive (SRA).

INTRODUCTION

Based on sequence complementarity, Piwi-interacting (pi-) RNAs function as guide molecules, thus ensuring target specificity of Piwi-subclass Argonaute proteins (1). Transposon silencing in the animal's germline is considered as the main function of Piwi proteins and piRNAs. Further, a growing number of studies report on regulatory functions beyond transposon silencing in diverse species including the regulation of protein coding genes (2–7). From the evolutionary point of view, piRNA transcriptomes are extremely adaptive, in particular because of the fact that piRNA clusters can incorporate foreign DNA that provides the substrate for new antisense piRNAs (8). In case of propagating transposons, piRNA clusters can this way provide an adaptive immunity against selfish DNA. However, regular mRNA can likewise be subject to reverse transcription and genomic reintegration into existing piRNA clusters, giving rise to processed pseudogenes that yield piRNAs with gene-regulatory function (9). Finally, gene-related piRNAs can also originate from loci that encode functional protein coding genes and piRNA clusters on opposite strands (5).

So far, different online databases have been established to provide information on piRNAs in different species (10–12). However, piRNA cluster database is the first database that focuses on genomic piRNA clusters as a biological source of piRNAs rather than piRNA sequences itself. Moreover, the algorithm that was used for piRNA cluster prediction was shown to produce better results compared to methods applied by other online databases and it was recently used to annotate piRNA clusters in many different species (5,13–19).

proTRAC annotation of piRNA clusters using more than 100 Sequence Read Archive (SRA) data sets from 12 species, together comprising over 250 million non-identical small RNA sequences represents the heart of this database. Our main goal is to provide reliable, reproducible and comparable information on piRNA producing loci not only in different species but also in, for example, different developmental stages or different tissues of the same species in order to facilitate research that focuses on both functional and evolutionary aspects. Knowing the genomic loci that can produce piRNAs under certain conditions is the basis to study aspects of their transcriptional regulation including analysis of putative promoter sequences, transcription factor binding sites or DNA/histone methylation/modifications. In addition, it represents the starting point for comparative studies that address evolutionary changes of piRNA cluster sequences that promote adaptations of piRNA transcriptomes.

*To whom correspondence should be addressed. Tel: +49 6131 3924004; Fax: +49 6131 3923799; Email: rosenkranz@uni-mainz.de

DATA COLLECTION AND ANALYSIS

Data collection and initial curation/processing

We screened NCBI's SRA database for small RNA sequence data obtained from animal germline tissues. In addition, we considered SRA data sets from studies that reported the presence of piRNAs or piRNA-like sequences in other than germline tissues. We collected a total of 112 SRA data sets from 12 different species, together comprising more than 3.8 billion reads representing ∼250 million non-identical sequences. Metadata associated with each data set (such as gender, tissue, developmental stage or information on associated Piwi proteins) is available in the SRA data set overview section of this database. Adapter sequences were clipped from the raw data and only sequence reads ranging from 18 to 34 nt in length were considered for further analysis. We removed identical sequences but retained information on read counts using the collapse tool from the NGS toolbox (20). Further, we removed low-complexity reads using the duster tool from the NGS toolbox applying the default settings (20). Figure 1 displays the relevant processing steps and the general structure of the database.

sRNA mapping and piRNA cluster prediction

We mapped the curated sequence sets to the respective genome with sRNAmapper (19) requiring a perfect match from position 1 to 18 and allowing one internal mismatch from position 19 onwards considering a commonly declining base calling quality toward the 3′ end of a sequence read. In addition, we allowed up to two non-templated 3′ nucleotides in order to successfully map sequences that were subject to post-transcriptional 3′ editing which is a common phenomenon for small RNAs including piRNAs (19). For each sequence, only the best-scoring alignments were taken into account (option: -alignments best).

After mapping, we used the program reallocate (http://www.smallrnagroup-mainz.de/software.html) to apportion read counts of multiple mapping sequences according to estimated local transcription rates based on uniquely mapping sequences. In contrast to a normal fractioning of read counts, we assume this method to yield a picture that is closer to the biological reality, especially considering sequence data sets with a high amount of transposon-related sequences. This method can also improve proTRAC performance as it generally favors to assign sequence read counts to piRNA clusters, which typically exhibit many uniquely mapping sequences, rather than piRNA targets (active transposons) that are characterized by few or no uniquely mapping sequences. When applying this sorting algorithm, running proTRAC resulted in a higher number of predicted piRNA clusters as well as more non-identical sequences and more sequence read counts assigned to piRNA clusters for most sRNA data sets (Supplementary Figure S1).

Since sequence data sets may contain large amounts of miRNAs, particularly when obtained from somatic tissues, we screened successfully mapped sequences for perfect matches to mature miRNAs and miRNA precursor molecules annotated at miRBase (21). miRNA-related sequences were removed from the map files and piRNA clusters were predicted using proTRAC version 2.0.5. The length distribution and positional nucleotide composition for mapped and clustered sequence reads was analyzed using the basic-analyses tool from NGS toolbox.

Building generic data sets

Obviously, SRA data sets can vary in quality depending on tissue material quality, RNA extraction efficiency and further downstream procedures. In order to compensate for putative variance between the different SRA data sets, we pooled multiple SRA data sets from one species to generate generic data sets for testis, ovary, oocytes and embryo where applicable. The generic data sets aim to provide a picture on piRNA expression under specific conditions (such as expression in adult testis) that is less affected by bias or contamination of single SRA data sets. When merging different data sets, read count values for each sequence in a given SRA data set were re-scaled in such a way that the total read count for each SRA data set finally summed up to one million. This approach compensates for the different number of read counts in each SRA data set and gives each data set the same weight without losing any sequence information. We then analyzed the generic data sets in the same way as the original SRA data sets.

Repeat annotation

We used RepeatMasker (http://www.repeatmasker.org/) annotation to obtain a detailed repeat annotation and determine the amount of repetitive sequences in predicted piRNA clusters and the amount of repeat-related small RNAs in the SRA data set. In addition, for the purpose of clarity, we grouped repeat families into five major categories: LTR (long terminal repeat, e.g. ERV1), LINE (long interspersed nuclear element, e.g. L1), SINE (short interspersed nuclear elements, e.g. Alu), DNA (DNA transposons, e.g. hAT) and others (e.g. tRNA, rRNA or unknown). In case of small RNA annotation, we also considered whether a sequence produced a hit in sense or antisense orientation. Read counts were apportioned accordingly if one sequence matched to loci with different repeat annotation. In order to asses a relative enrichment or depletion in repeat sequences compared to the whole genome situation, we annotated the genomes of the relevant species accordingly.

Ping-pong signatures

As a result of their biogenesis during the so-called ping-pong amplification loop, mapped piRNAs typically exhibit a bias for 10 bp 5′ overlaps which is referred to as the ping-pong signature. In contrast, other classes of small non-coding RNAs such as miRNAs or siRNAs produce 5′ overlaps of ∼18–22 bp owing to Dicer processing. We analyzed the 5′ overlaps of mapped sequences for each SRA data set and calculated a significance score (ping-pong score) according to the method described by Zhang *et al.* (22). The distribution of 5′ overlaps allows to draw inferences on the ratio of piRNAs and siRNAs in the SRA data set and can reveal ongoing ping-pong amplification.

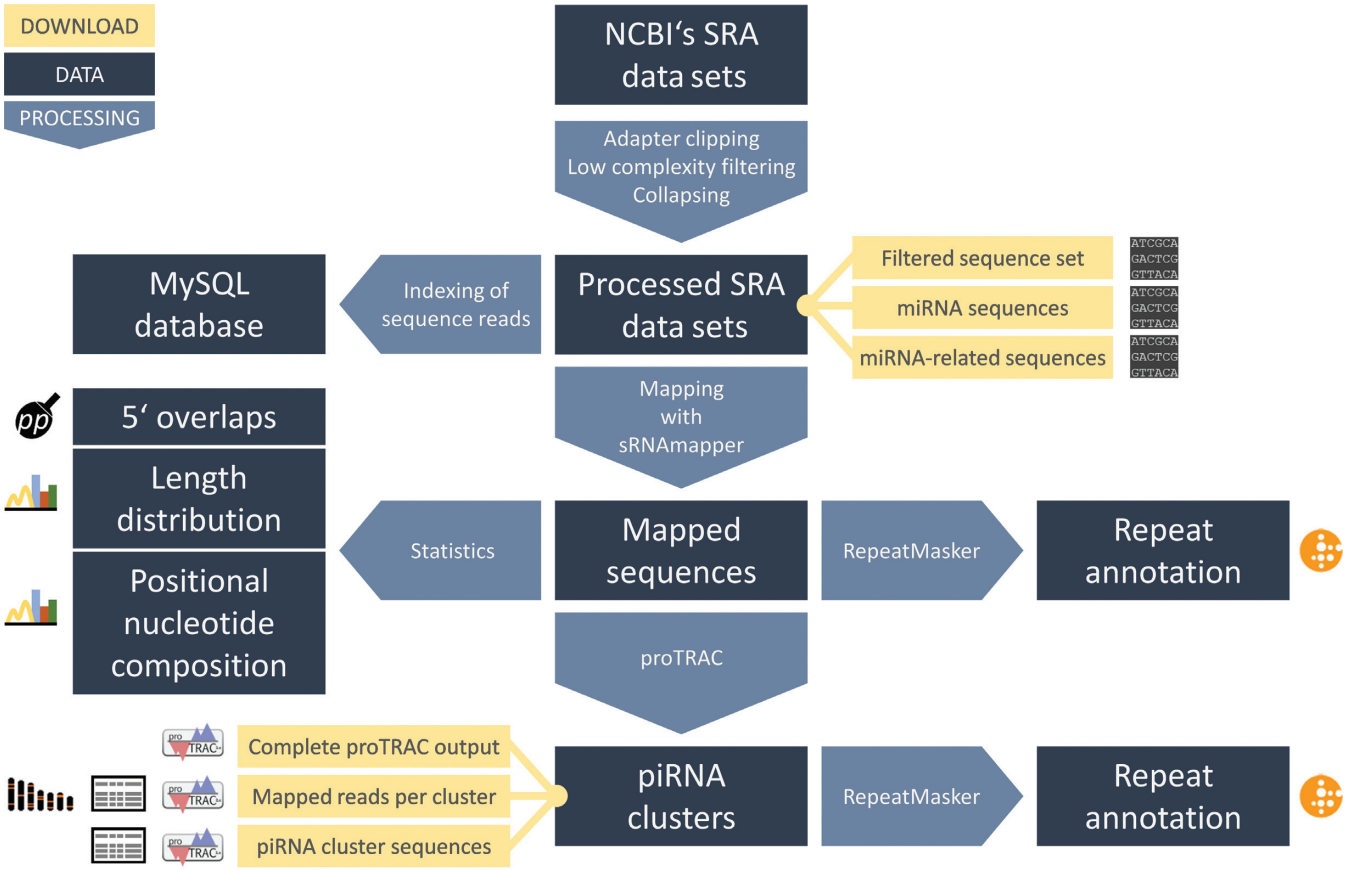

Figure 1. Organization of piRNA cluster database. Icons indicate the links to access the according data.

Transcriptional activity of piRNA producing loci in time and space

Since piRNA clusters are often dynamically expressed in different tissues or at different developmental stages, we combined proTRAC annotations using different SRA data sets as well as the generic data sets for each species in order to obtain a complete set of genomic loci that can produce piRNAs within one species. Adjacent piRNA clusters with a distance less than 10 kb were merged. All piRNA producing loci for a species including detailed annotation can be accessed conveniently via the pi-browser (see below). In addition, we compared sequence read coverage of these loci for different SRA data sets and prepared heatmaps that visualize the relative normalized small RNA coverage for each piRNA producing locus across different SRA data sets.

HOW TO ACCESS THE DATA

Choosing a SRA data set

An easy way to access information is to browse by species and select the desired SRA data set. If the user does not want to choose piRNA clusters by species but by means of other criteria such as tissue, gender, developmental stage, etc., he can select an appropriate SRA data set from a searchable and sortable table that comprises the available SRA data sets with associated metadata. The table also provides links to the 'Browse piRNA clusters' section for each SRA data set. Once the user has selected a specific data set, he can access the provided data using several datalinks (Figure 2A).

Datalinks, downloads and interface

Basic information such as read counts, number of mapped reads, number of miRNA reads, number of piRNA clusters and number of clustered reads for each SRA data set is accessible via the statistics link (Figure 2B). Here, we also provide data on length distribution and positional nucleotide composition for both mapped and clustered reads in a graphical and tabular form.

The physical location of piRNA clusters for each SRA data set along chromosomes is displayed on a schematic karyogram (Figure 2C). The location of piRNA clusters is indicated with colored bars with the color encoding the relative sequence read coverage of a given piRNA cluster. Moving the mouse over a piRNA cluster shows the read coverage along the cluster sequence and additional data associated with the piRNA cluster. A FASTA file comprising mapped sequence reads for each piRNA cluster can be downloaded by right-click on the desired location.

We further provide a graphical presentation of the RepeatMasker annotation for mapped sequences reads, piRNA clusters and the species genome comprising exact values for sense/antisense reads or covered bp for each transposon family, respectively (Figure 2D).

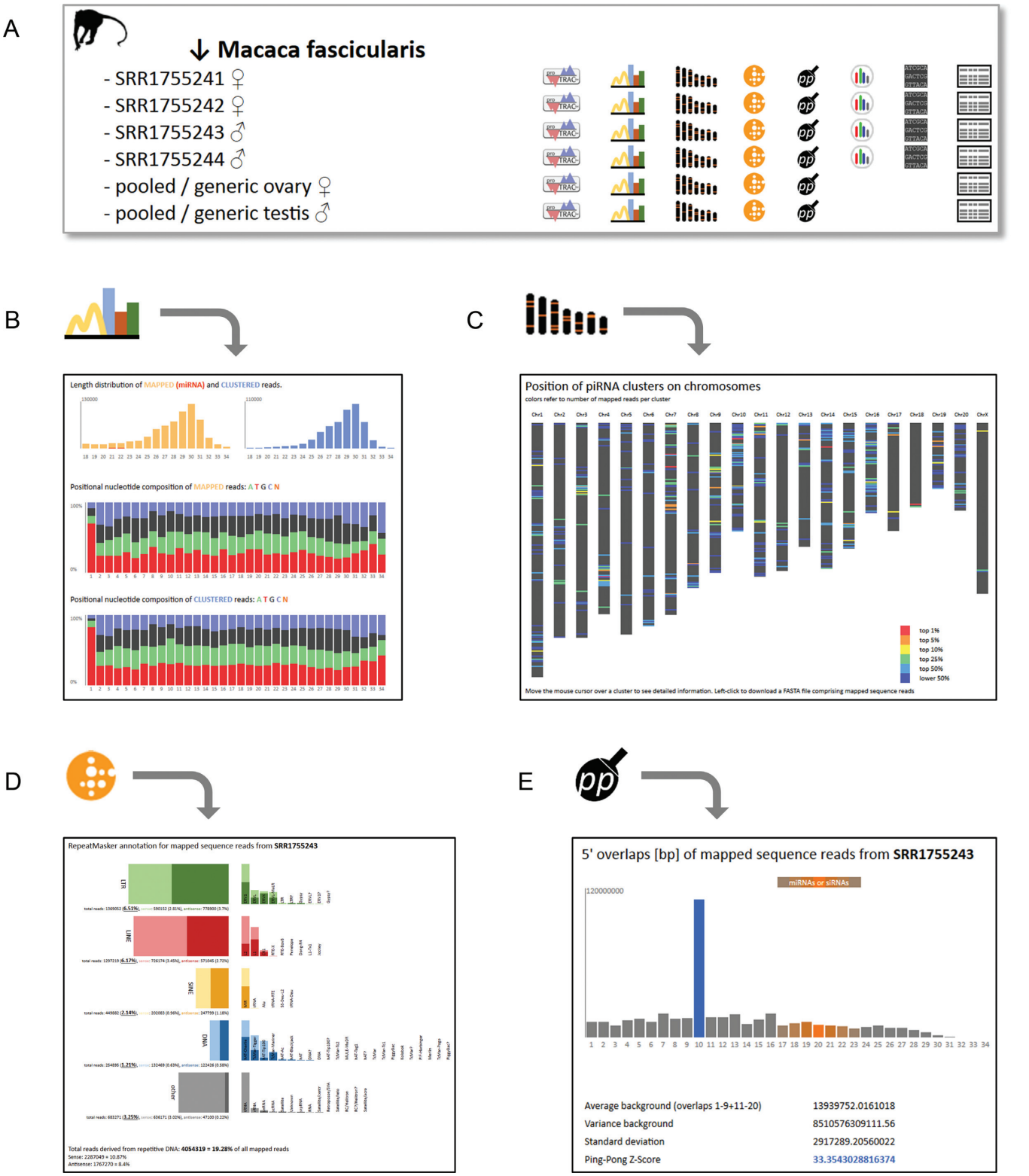

Figure 2. Graphical user interfaces of the piRNA cluster database. (**A**) Database frontend lists different small RNA data sets that are available for the desired species and shows links to access the connected data. (**B**) Length distribution and positional nucleotide composition of mapped and clustered sequence reads. (**C**) A karyogram shows the physical distribution of piRNA clusters and provides download links for mapped sequence reads. (**D**) RepeatMasker annotation for mapped and clustered sequence reads and the genome of the species in question. (**E**) Distribution of 5′ overlaps for mapped sequence reads and related statistics.

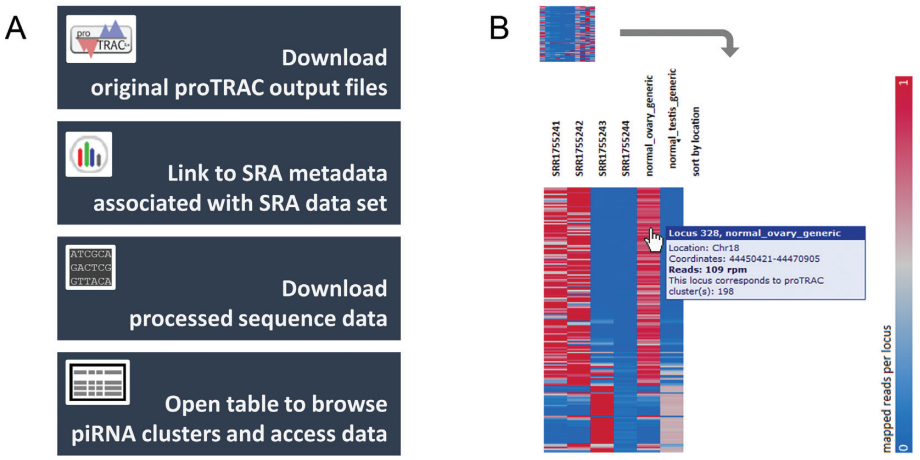

Figure 3. More information and direct download links. (**A**) The complete proTRAC output is available for download. The user can also browse, search and sort the proTRAC results by piRNA clusters using the piRNA cluster tables. Links to SRA annotation and processed sequence data sets (e.g. files that served as input for proTRAC) are provided. (**B**) Heatmaps illustrate read coverage of genomic loci across different SRA data sets. The heatmaps can be sorted according to the number of normalized mapped reads for each SRA data set separately.

For each SRA data set, the user can access a graphical representation of 5′ overlaps for mapped sequence reads. Values for a 10-bp overlap (as they result from piRNA ping-pong amplification loop) are highlighted in blue, whereas overlaps of 18–22 nt that may point to the presence of putative Dicer products are indicated with orange coloring. The scores for 5′ overlaps and values related to the calculation of ping-pong Z-scores according to Zhang *et al.* (22) are provided in a tabular format (Figure 2E).

The processed sequence data that were used for genomic mapping as well as the complete original proTRAC output files are available using the direct download links. In addition, single piRNA clusters can be browsed via sortable and searchable tables that comprise all clusters for a given proTRAC run including associated data and further download links for (i) genomic piRNA cluster sequence, (ii) sequences of mapped reads and (iii) proTRAC cluster image files. Metadata for any SRA data set can be found by following the direct link to the run browser at NCBI's SRA (Figure 3A).

pi-browser – Browse all piRNA producing loci in a species

For each species, the user can browse all genomic loci that were annotated as a piRNA cluster in any of the included SRA data sets. The pi-browser contains a sortable table that comprises a complete list of piRNA producing loci for the species in question. In addition, a karyogram visualizes the physical location of these loci on the chromosomes (Figure 4). To access the data connected to each locus, a desired locus can be selected from the table or directly from the karyogram. Once the user has chosen a locus, the sequence read coverage for each SRA data set along the locus in question is shown in the sequence read coverage field. SRA data sets can be selected or de-selected in order to display only SRA data sets of interest. Exact values for plus/minus-strand sequence read coverage are shown when moving the mouse cursor over the desired location. Gene annotation based on Ensembl Gene Sets, RepeatMasker annotation and annotation of transcription factor binding sites (proTRAC version 2.1) for each locus are provided at the bottom of the read coverage field. Moving the mouse cursor over a specific element in the RepeatMasker or Gene Set annotation table will highlight the respective element in the RepeatMasker or Gene Set track, respectively. For piRNA producing loci in mouse we additionally added the complete set of known mouse piRNA precursor transcripts identified by Li *et al.* (23).

Heatmaps display piRNA expression across SRA data sets

Users can compare the transcriptional activity of genomic loci across all SRA data sets for a given species using the provided heatmaps, where rows represent a genomic locus and columns represent a specific SRA data set (Figure 3B). By default, the sorting corresponds to genomic coordinates but the heatmaps can also be sorted according to the number of mapped reads for a desired SRA data set in descending or ascending order. For each locus, the SRA data set that produces the highest coverage is indicated by a red box and colors for the remaining SRA data sets derive from comparing rpm (reads per million) values relative to the maximum value where 0% of the maximum rpm value is indicated by a blue box. This means that the coloring is comparable within one line (for one locus) but not within one column. Consequently, when sorting the loci according to rpm values for a desired SRA data set the coloring will not necessarily be continuous from red to blue (or converse), although this tendency can usually be observed. For purposes of detailed analysis, we provide a tab-delimited table comprising the underlying data with rpm values per locus and SRA data set for download.

DISCUSSION AND PERSPECTIVES

In 2006, four groups independently characterized a novel class of Piwi binding small non-coding RNAs in the male mouse germline which were named piRNAs (24–27). piRNAs where found to originate from discrete genomic loci that are now commonly referred to as piRNA clusters. In

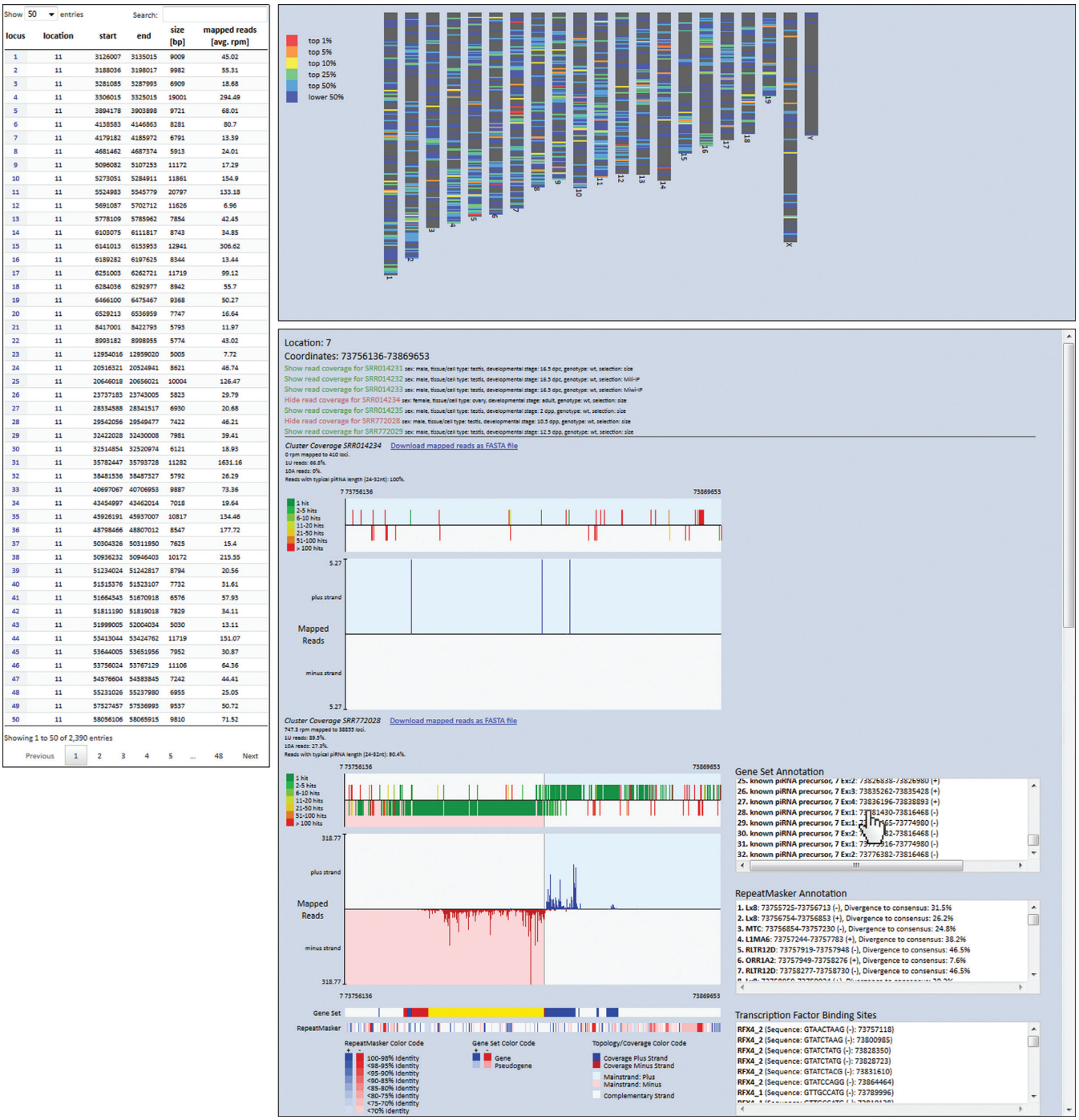

Figure 4. The pi-browser interface. A desired locus can be selected from the table or from the karyogram. piRNA read coverage is displayed seperately for each SRA data set available for the species in question. Further annotation for the selected locus (genes, transposons, transcription factor binding sites) can be found below the sequence read coverage plots.

the subsequent years, numerous studies shed light upon many aspects of piRNA biogenesis and function in species ranging from ciliates to human (1,28,29). Based on findings from the mouse model, the function of the mammalian Piwi/piRNA system was initially assumed to be limited to transposon control in the male germline. However, recent studies on mammalian piRNAs report on their presence in the female germline, in early embryos as well as in somatic tissues (19,30). Moreover, piRNA function is clearly not restricted to transposon silencing and evidence for a conserved gene regulatory role in mammals exist (4–7). Today we know that the Piwi/piRNA system is highly adaptive and manifold realizations have been established in different species during evolution. Understanding aspects of piRNA function inside and outside the germline, in evolution and

development, is a Herculean task that is far from being completed.

Owing to readily accessible high throughput sequencing techniques, the number of annotated piRNAs as well as the number of known piRNA source loci in different species and tissues is growing faster than our understanding of piRNA biology does. With their initial description of piRNAs in 2006, Girard *et al.* identified 123 piRNA clusters in mouse based on testis-expressed piRNAs (26). Currently, piRNA cluster database comprises 879 piRNA clusters based on pooled SRA sequence sets from mouse testis. In addition, 336 piRNA clusters were annotated using small RNA sequence data from mouse ovaries. A similar progress can be observed for other species such as human, cattle or the crab-eating macaque, where piRNAs were recently identified in the female germline (19). One of the interesting facts considering piRNAs from different sexes is that female piRNAs originate from largely different clusters compared to their male counterparts. The exact mechanisms that are responsible for their oocyte-specific expression are obscure and yet unknown factors may be involved since a critical transcription factor that binds pachytene piRNA clusters during mouse spermatogenesis is expressed in testis but not in ovaries (31,32). Moreover, evolutionary aspects like piRNA sequence conservation during mammalian evolution have thus far been studied focusing on loci that are transcriptionally active in testis (18,33), while nothing is known about the evolution of loci that produce piRNAs in oocytes. Similarly, it appears reasonable to assume that piRNAs target different protein coding genes in testis compared to ovary (or any other) tissue. However, comparative studies in this field are yet outstanding.

We argue that addressing these and related issues requires reliable and detailed knowledge regarding which genomic loci produce piRNAs under which conditions. So far, this information is commonly hidden in supplemental files and comparisons across studies are hampered by the fact that fundamentally different criteria were applied for the annotation of piRNA clusters. Thus, piRNA cluster database was launched to provide easy access to comprehensive data related to genomic piRNA clusters while at the same time facilitating effortless comparisons of clusters predicted using sequence data obtained from different experimental settings.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The author would like to thank René F. Ketting from the Institute of Molecular Biology (IMB) in Mainz, Germany, for friendly advice regarding zebrafish piRNA biology. Hans Zischler and Daniel Gebert are gratefully acknowledged for their assistance in construction and testing of this database and careful revising of this manuscript. A warm Thank-You goes to Christian Heckelmann and David Gemmer for help in database programming. Finally, thanks go to the members of the research initiative GeneRED, which is funded as a Forschungsschwerpunkt ('research focus') by Rhineland-Palatinate's Ministry of Education, Science, Further Education and Cultural Affairs for valuable scholarly exchange.

FUNDING

Intramural research funding program MAIFOR from the University Medical Centre Mainz. Funding for open access charge: Institute of Anthropology, Johannes Gutenberg University Mainz, Germany.
Conflict of interest statement. None declared.

REFERENCES

1. Iwasaki,Y.W., Siomi,M.C. and Siomi,H. (2015) PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu. Rev. Biochem.*, **84**, 405–433.
2. Watanabe,T., Tomizawa,S., Mitsuya,K., Totoki,Y., Yamamoto,Y., Kuramochi-Miyagawa,S., Iida,N., Hoki,Y., Murphy,P.J., Toyoda,A. *et al.* (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science.*, **332**, 848–852.
3. Rajasethupathy,P., Antonov,I., Sheridan,R., Frey,S., Sander,C., Tuschl,T. and Kandel,E.R. (2012) A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell*, **149**, 693–707.
4. 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.
5. 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.
6. 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.
7. 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.
8. Kawaoka,S., Mitsutake,H., Kiuchi,T., Kobayashi,M., Yoshikawa,M., Suzuki,Y., Sugano,S., Shimada,T., Kobayashi,J., Tomari,Y. *et al.* (2012) A role for transcription from a piRNA cluster in de novo piRNA production. *RNA*, **18**, 265–273.
9. 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.
10. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
11. Sarkar,A., Maji,R.K., Saha,S. and Ghosh,Z. (2014) piRNAQuest: searching the piRNAome for silencers. *BMC Genomics*, **15**, 555.
12. Zhang,P., Si,X., Skogerbø,G., Wang,J., Cui,D., Li,Y., Sun,X., Liu,L., Sun,B., Chen,R. *et al.* (2014) piRBase: a web resource assisting piRNA functional study. *Database*, **2014**, bau110.
13. Rosenkranz,D. and Zischler,H. (2012) proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, **13**, 5.
14. Gerstl,M.P., Hackl,M., Graf,A.B., Borth,N. and Grillari,J. (2013) Prediction of transcribed PIWI-interacting RNAs from CHO RNAseq data. *J. Biotechnol.*, **166**, 51–57.
15. Rengaraj,D., Lee,S.I., Park,T.S., Lee,H.J., Kim,Y.M., Sohn,Y.A., Jung,M., Noh,S.J., Jung,H. and Han,J.Y. (2014) Small non-coding RNA profiling and the role of piRNA pathway genes in the protection of chicken primordial germ cells. *BMC Genomics*, **15**, 757.
16. Di,R., He,J., Song,S., Tian,D., Liu,Q., Liang,X., Ma,Q., Sun,M., Wang,J., Zhao,W. *et al.* (2014) Characterization and comparative profiling of ovarian microRNAs during ovine anestrus and the breeding season. *BMC Genomics*, **15**, 899.
17. Itou,D., Shiromoto,Y., Yukiho,S.Y., Ishii,C., Nishimura,T., Ogonuki,N., Ogura,A., Hasuwa,H., Fujihara,Y.,

Kuramochi-Miyagawa,S. *et al.* (2015) Induction of DNA methylation by artificial piRNA production in male germ cells. *Curr. Biol.*, **25**, 901–906.
18. Rosenkranz,D., Rudloff,S., Bastuck,K., Ketting,R.F. and Zischler,H. (2015) Tupaia small RNAs provide insights into function and evolution of RNAi-based transposon defense in mammals. *RNA.*, **21**, 911–922.
19. Roovers,E.F., Rosenkranz,D., Mahdipour,M., Han,C.T., He,N., Chuva de Sousa Lopes,S.M., van der Westerlaken,L.A.J., Zischler,H., Butter,F., Roelen,B.A.J. *et al.* (2015) Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep.*, **10**, 2069–2082.
20. Rosenkranz,D., Han,C.T., Roovers,E.F., Zischler,H. and Ketting,R.F. (2015) Piwi proteins and piRNAs in mammalian oocytes and early embryos: from sample to sequence. *Genom Data*, **5**, 309–313.
21. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. Maartje,J., Luteijn,M.J. and Ketting,R.F. (2013) PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.*, **14**, 523–534.
29. Gebert,D. and Rosenkranz,D. (2015) RNA-based regulation of transposon expression. *Wiley Interdiscip Rev RNA*, **6**, 687–708.
30. Yan,Z., Hu,H.Y., Jiang,X., Maierhofer,V., Neb,E., He,L., Hu,Y., Hu,H., Li,N., Chen,W. *et al.* (2011) Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.*, **39**, 6596–6607.
31. Mettus,R.V., Litvin,J., Wali,A., Toscani,A., Latham,K., Hatton,K. and Reddy,E.P. (1994) Murine A-myb: evidence for differential splicing and tissue-specific expression. *Oncogene*, **9**, 3077–3086.
32. 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.
33. 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.