Analytical Biochemistry 602 (2020) 113781

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Identification and characterization of microRNAs (miRNAs) and their transposable element origins in the saltwater crocodile, *Crocodylus porosus*

Arnab Ghosh[a], Roy N. Platt II[a,1], Michael W. Vandewege[a,b], Rabia Tabassum[c], Chuan-Yu Hsu[d], Sally R. Isberg[e,f], Daniel G. Peterson[d], John W. Finger Jr.[g,h], Troy J. Kieran[g], Travis C. Glenn[g], Jaime Gongora[e], David A. Ray[a,*]

[a] *Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA*
[b] *Department of Biology, Eastern New Mexico University, Portales, NM, USA*
[c] *Charles Darwin University, Australia*
[d] *Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, MS, USA*
[e] *Sydney School of Veterinary Science, Faculty of Science, University of Sydney, Sydney, NSW, Australia*
[f] *The Centre for Crocodile Research, Darwin, NT, Australia*
[g] *Department of Environmental Health Science, University of Georgia, Athens, GA, USA*
[h] *Department of Biological Sciences, Auburn University, Auburn, AL, USA*

ARTICLE INFO

Keywords:
miRNA
Crocodylus porosus
miRDeep2
Crocodilian
Phylogeny
Small RNA

ABSTRACT

MicroRNAs (miRNAs) are 18–24 nucleotide regulatory RNAs. They are involved in the regulation of genetic and biological pathways through post transcriptional gene silencing and/or translational repression. Data suggests a slow evolutionary rate for the saltwater crocodile (*Crocodylus porosus*) over the past several million years when compared to birds, the closest extant relatives of crocodilians. Understanding gene regulation in the saltwater crocodile in the context of relatively slow genomic change thus holds potential for the investigation of genomics, evolution, and adaptation. Utilizing eleven tissue types and sixteen small RNA libraries, we report 644 miRNAs in the saltwater crocodile with > 78% of miRNAs being novel to crocodilians. We also identified potential targets for the miRNAs and analyzed the relationship of the miRNA repertoire to transposable elements (TEs). Results suggest an increased association of DNA transposons with miRNAs when compared to retrotransposons. This work reports the first comprehensive analysis of miRNAs in *Crocodylus porosus* and addresses the potential impacts of miRNAs in regulating the genome in the saltwater crocodile. In addition, the data suggests a supporting role of TEs as a source for miRNAs, adding to the increasing evidence that TEs play a significant role in the evolution of gene regulation.

1. Introduction

MicroRNAs (miRNA) are 18–24 nucleotide (nt) long eukaryotic non-coding sequences that act as post-transcriptional regulators of gene and gene networks through either mRNA degradation or translational repression. These small RNAs are generated from their source genes mostly through bi-directional transcription by RNA polymerase-II. This generates a primary miRNA (pri-miRNA) which is then processed by the RNAse-III-like enzyme, Drosha, in the nucleus. The pri-miRNA is then processed into a ~70-nt long hairpin-shaped precursor miRNA (pre-miRNA), which is transported out of the nucleus by the Exportin-5 protein. Once inside the cytoplasm, the pre-miRNA undergoes a round of Dicer-mediated cleavage to generate the duplex miRNA. The more thermodynamically stable mature miRNA strand is bound by an Argonaute (AGO) protein and forms the functional ribonucleoprotein complex known as the RISC (RNA-induced silencing complex). The RISC is then guided by the mature miRNA to a complimentary or partially complimentary mRNA target site to carry out gene silencing either through mRNA cleavage or translational repression [1–3]. The less stable complement to the mature miRNA will usually undergo degradation and be lost [4–6].

miRNAs in animals are known to be most effective in targeting gene transcripts when there is full complementarity between the 3′ untranslated region (UTR) of the target mRNA with a 7-8 nt stretch in the 5' end of miRNA (i.e., the *seed sequence*) [7–10]. Nearly 60% of the protein coding genes are subject to miRNA regulation [11]. In animals,

* Corresponding author. Department of Biological Sciences, Texas Tech University, Box 43131, Lubbock, TX, 79409, USA.
[1] Current address: Disease Intervention and Prevention Program, Texas Biomedical Research Institute, San Antonio, TX.
E-mail address: david.4.ray@gmail.com (D.A. Ray).

https://doi.org/10.1016/j.ab.2020.113781
Received 12 February 2020; Received in revised form 11 May 2020; Accepted 19 May 2020
Available online 30 May 2020
0003-2697/

miRNAs are known to be involved in regulation of several critical biological processes-including the time and pattern of development, apoptosis, control of the cell cycle, disease pathogenesis and host response, and stress response [12–15].

Birds are the closest extant relatives to crocodilians in the clade Archosauria, which also encompasses non-avian dinosaurs and pterosaurs [16,17]. Crocodilian genomes have been evolving very slowly over the past several million years when compared to birds [18]. Understanding the evolution, regulation and adaptive capabilities of crocodilian genomes and their genetic diversity can therefore provide information on how slowly evolving genomes manage to stay viable in the face of competition from other taxa and changing environmental conditions. Reptilian genomic studies are also necessary for better understanding of the patterns of genomic evolution across amniotes (mammals, birds and non-avian reptiles).

miRNAs provide us with the opportunity to investigate the potential causes behind the slow evolution of crocodilians. At present, only two studies, both done in American alligators (*Alligator mississippiensis*), have examined small RNA dynamics in crocodilians. Rice et al. [19] targeted a related crocodilian, *Alligator mississippiensis*, but only a single tissue. In an earlier study, some conserved miRNAs, also in *Alligator mississippiensis,* were annotated by mapping small RNA reads to miRNAs from other species [20]. For this study, we used the most recent annotated genome assembly for the saltwater crocodile, *Crocodylus porosus* [21], to identify and characterize 644 miRNAs using Ilumina's® deep sequencing HiSeq 2500 platform. The miRNAome includes known and novel miRNAs. Tissue specific patterns of expression were also observed, providing the first comprehensive analysis of crocodilian miRNAs.

In addition to examining the general catalog of these small RNAs, we were interested in the genomic context of miRNAs. One feature that reflects the low rate of genomic change in crocodilians is the lack of recent transposable element (TE) accumulation in crocodilian genomes [18,22]. Transposable elements are genetic features that can copy and, in some cases, physically move from one location to another in a genome. Such activity can result in the modification of gene expression through multiple processes including disruption of coding and non-coding regions, promotion of non-homologous recombination, exaptation, and epigenetic regulation. All of these processes could impact the evolution of diversity, both taxonomic and genomic [23–31]. TEs are broadly classified into two classes, Class I retrotransposons and Class II DNA transposons [32]. An increasing body of research has suggested that TEs are major players in the evolution of gene regulation [33–37] and miRNAs have been known to be either derived from or involved in the silencing of TEs [38–45]. Because better understanding of regulatory networks and TEs can help us assess how gene expression phenotypes evolve and the TE landscape is maintained in the genome, we chose to investigate the connection between these two genomic features.

2. Materials and methods

2.1. Tissue collection

Ten hatchling saltwater crocodiles (of which five were phenotypically normal and five were runts) were euthanized for immediate tissue collection (University of Sydney Animal Ethics protocol N00/5-2012/3/5729) using methods previously described [46,47]. All tissues (Table 1) were flash frozen in liquid nitrogen. Tissues included were belly skin, brain, cloaca, heart, jaw skin, liver, small intestine, stomach, spleen, testis, and tongue. Normal vs. runt status of subjects was determined based on size differences. The five runts were, on average, 359 days old with a total length of 386.2 ± SD 23.6 mm compared to the normal, healthy crocodiles, which were 373 days old and with an average length of 740.2 ± 73.5 mm. These animals were denoted R1-R5 and N1-N5 respectively.

Table 1
Small RNA library details for this study.

Tissue	Individual	Library ID	RQN
Testis	N2	EL4410	8.2
Testis	N1	EL4411	7.1
Testis	R1	EL4412	7.2
Testis	R5	EL4413	7.7
Brain	R5	EL5564	9.1
Brain	N2	EL5565	4.6
Cloacal Gland	N3	EL5566	10
Liver	N1	EL5567	9.7
Jaw Skin	N3	EL5568	9.8
Heart	N2	EL5569	9.3
Heart	R3	EL5570	5
Belly skin	R5	EL5571	9
Small Intestine	R4	EL5572	8.3
Spleen	R2	EL5573	10
Stomach	R3	EL5574	8.5
Tongue	R5	EL5575	8.4

N = 'Normal' non-runt Individuals (N1-N3).
R = Runted hatchlings (R1-R5).
RQN = RNA Quality Number.

2.2. RNA processing and small RNA library preparation

Total RNA was isolated using Trizol (Life Technologies Inc.) and the Direct-Zol RNA Miniprep Kit (Zymo Research Inc.). Total RNA was quantified using the Qubit RNA assay and Qubit 2.0 fluorometer (Life Technologies Inc.). RNA integrity was confirmed using the Total RNA Standard Sensitivity assay reagents and Fragment Analyzer system (Agilent Inc.). Only samples having an RNA Integrity Number (RIN) greater than 7 or samples with clear 18s and 28s ribosomal RNA peaks as seen with the Bioanalyzer were used in further studies, the only exception being one sample each from the brain and the heart where a RIN of < 7 was considered acceptable. Note that some of the brain and heart samples had RINs > 7 and these were compared to RIN < 7 samples later in the procedure.

One microgram of total RNA was used as input material in building each NEBNext Small RNA Library Prep for Illumina (NEB Inc.). To prepare samples for sequencing, small non-coding RNA species underwent 3′ RNA ligation and reverse transcriptase primer hybridization followed by 5' RNA ligation and first strand cDNA generation using Superscript III reverse transcriptase (Life Technologies Inc.). The resulting cDNA which contained a portion of the adapter sequences for Illumina sequencing, underwent PCR enrichment (94 °C for 30 s, 12 cycles [94 °C for 15sec, 62 °C for 30 s, 70 °C for 15 s], 70 °C for 5 min) using a universal primer and a multiplex specific primer. The final product was small RNA libraries with sample-specific 6-nt indexes that allow for multiplex sequencing on the Illumina platform. Following PCR, samples were purified using a 1.8x SPRI bead purification (Beckman Coulter Inc) and the final sequencing construct was size selected on the Pippin Blue system (Sage Science Inc.) using a 3% w/v gel cassette and internal marker for collecting the range of fragments from 135 to 170 base pairs (bp).

Library validation was performed using the NGS Hi Sensitivity assay reagents and the Fragment Analyzer system followed by quantitation using the Qubit HS DNA assay and a qPCR kit for Illumina (Kapa Biosystems Inc). Libraries were diluted using the Qubit or qPCR information and loaded onto a single HiSeq2500 (Illumina Inc.) flow cell. Single-end 75 bp reads were generated using the pooled libraries. Pooled sequence data was demultiplexed into individual files with bcl2fastq (v2) representing the sixteen unique libraries. Using a series of programs available in the FASTX-toolkit (v 0.0.14; http://hannonlab.cshl.edu/fastx_toolkit/), sequences were quality filtered. In brief, the Fastx_clipper program was used to remove adapter sequence (5′-AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C-3′) from each read as well as culling reads with ambiguous nucleotides ("N"). Reads which

lacked adapter sequence or were less than 10-nt after clipping were not included in downstream analyses. These reads were then piped into the Fastq_quality_filter program and any reads with Phred quality values less than 20 across more than half of the read were removed. Small RNA reads have been deposited in the Gene Expression Omnibus archive and can be accessed through GEO accession number GSE150461.

2.3. miRNA prediction and annotation

The miRNAome in *C. porosus* was identified and characterized in three phases: Identification of conserved miRNAs, identification of novel crocodilian miRNAs, and characterization of expression profiles. After initial library preparation, sequencing and filtering, miRDeep2 (version 2_0_0_8) [48] was used to identify and annotate both known and novel miRNA. Because of the relatively limited current knowledge of crocodilian miRNA, with no information available for the saltwater crocodile, two birds-the chicken and zebra finch (the only two birds with data in miRBase v21), were chosen as taxa for comparison. The mature and hairpin precursor sequences for the above avian miRNAs, as downloaded from miRBase (v21; www.mirbase.org), were used as a training set after using a custom UNIX script (https://github.com/davidaray/cPorosus_miRNAs) to select chicken and zebra finch mature miRNAs and corresponding hairpin precursor sequences. This script was also used to process all remaining steps in our pipeline unless otherwise indicated. All mature and hairpin precursor sequences were concatenated for both species in two separate files and denoted AVIAN_MATURE and AVIAN_HAIRPIN, respectively, to be used as training sets in the miRDeep2 analysis. The mapper.pl script (part of the miRDeep2 suite) was then used to map the deep sequencing reads to a reference genome. Default options for mapper.pl were used except reads less than 18 bases were removed (-l 18), identical reads were collapsed (-m), and sequences with non-canonical bases were removed (-j). miRDeep2 predicted miRNAs were scored based on the ability of the flanking genomic sequence to form a hairpin structure typical of miRNAs. Throughout the rest of this work, any miRNA identified in the saltwater crocodile but previously known and catalogued in miRbase were considered as “known”, while those miRNAs that were not found in miRbase were designated as “novel” in crocodilians.

The miRDeep2 core algorithm assigns a distinct “score” to each mapped small RNA read based on its conformity to several miRNA biogenesis hallmarks. Based on a previous study [48], any miRNA with a miRDeep2 score of < 2 was removed from further analyses. After prediction by the miRdeep2 algorithm, several steps were taken to remove low confidence predictions from the dataset and subsequent analyses. First, all predicted known and novel miRNA hairpin precursor sequences were filtered to remove matches to tRNA and/or rRNA datasets. Any matches to either of these two datasets irrespective of e-value or bitscore, were removed from the predicted entries. The eukaryotic genomic tRNA database (http://gtrnadb.ucsc.edu) and the rRNA dataset (http://www.arb-silva.de/) were used. BLASTn in conjunction with our UNIX script was used to select out unique predicted miRNA entries that were produced from the miRDeep2.pl run but did not match to the rRNA and/or tRNA datasets.

There were some cases where multiple hairpin precursor sequences were identified from overlapping regions, often offset by only a few bases. We removed any overlapping miRNA predictions. The procedure was accomplished in two steps using Bedtools (v 2.26.0) [49]. First, the ‘sort’ and ‘merge’ functions from Bedtools were used to identify overlaps between the predicted miRNA precursor sequences. Only one instance was found where a novel and a known predicted miRNA candidate precursor overlapped and that pair was removed from both lists. For all remaining precursor candidates, the script removeOverlappingMirnas.pl was used to retain the highest scoring partner from each overlapping pair and remove the partners with lower scores. The score here refers to a variable representing the difference between start and end coordinate of the miRNA precursor. Essentially, any shorter precursor(s) completely overlapping and “contained within” another precursor (hence lower score) was discarded. The longer precursor (essentially the highest scoring partner) was retained.

The miRDeep2 core algorithm assigns a distinct “score” to each mapped small RNA read based on its conformity to several miRNA biogenesis hallmarks. Based on a previous study [48], a miRDeep2 score of 2 was chosen as a cutoff value to remove any candidate miRNA that had a miRDeep2 score of < 2.

The miRDeep2 algorithm utilizes the RNAfold (v 2.0) program to predict secondary structures of putative miRNA precursors, where the core algorithm assigns a “yes” flag to the “randfold” value for potential true miRNA candidate precursors. At this step, another custom bash script was used to select those miRNA candidates that were assigned a “yes” on the significant randfold value category by the miRDeep2.pl script. These selected miRNA candidates were then piped into the output file of the miRDeep2score filter from the previous step and used to generate a preliminary list of high-quality filtered miRNA candidates.

We removed any candidate mature miRNAs that mapped fewer than 10 times on the hairpin precursor sequence. The output file was ultimately piped into genHQseq.pl script to generate a final list of high-quality mature miRNA and their corresponding hairpin precursor files in FASTA format.

Final manual filtering steps were included as follows. Any candidate miRNA with ten or more single nucleotide repeats were removed irrespective of prediction by the miRDeep2 program. Ten such candidate miRNAs were removed and none of these candidate miRNAs were from the previously ‘known’ category. The remaining candidates were scanned for duplicates (generated due to biogenesis from multiple loci in the genome or being retained as a result of existing in both chicken and zebra finch) and only one unique candidate sequence, based on its highest normalized expression was retained. All retained predicted miRNA candidates had at least a 70% estimated probability of being a true miRNA and a final total of 644 unique high-quality candidate miRNAs along with their hairpin precursors were identified and annotated in *C. porosus*. This included 132 known and 532 novel miRNA candidates. A general overview representing key steps of the miRNA prediction and annotation procedure is provided in Fig. 1.

2.4. miRNA expression

Quality filtered small RNA reads from the individual libraries were mapped to the curated precursor and mature miRNA sequences. Once mapped, the quantifier.pl script from the miRDeep2 package was used to calculate both raw counts and normalized counts for each of the distinct mature miRNA candidates. Expression scores were weighted based on the number of small RNA read mapping positions. We applied a serial nomenclature to the predicted novel miRNAs in *C. porosus*. For known miRNAs, the miRNA family names from the avian datasets as well as other species were retained with the addition of the prefix “cPor”. Two separate heatmaps were generated expressing normalized read counts of the novel and known miRNA candidates using R v3.3 and output from quantifier.pl script. Separate normalized expression profile for several distinct miRNA candidates were generated due to their unique expression profile in a certain tissue type (library) or a group of tissues (libraries). All novel miRNAs were deposited in miRBase.

2.5. Target identification

miRNA target identification was performed by using two independent target prediction tools – miRanda (version 3.3a) and RNAhybrid (version 2.1.2). For miRanda, the option “–strict” was enforced so that the miRNA seed region had to show complete complementarity to a 3′UTR region. Naturally, this option prevented detection of target sites containing gaps or non-canonical base pairing. A free energy of -20 kcal/mol and an alignment threshold score of 170 were chosen as cutoffs in option parameters for miRanda. These stricter

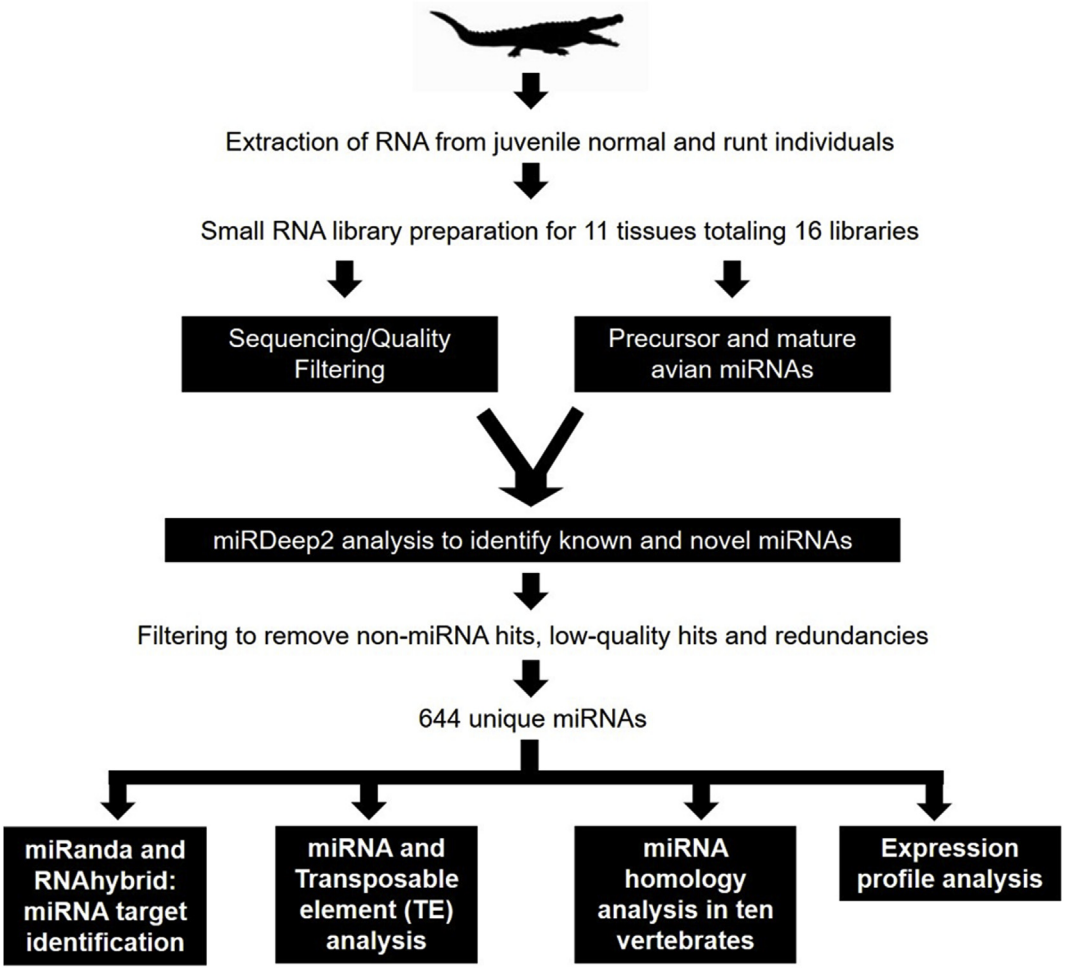

Fig. 1. Workflow representing key steps undertaken for the identification and characterization of miRNAs in the saltwater crocodile.

values were used to reduce the possibility of false positives when compared to previous analyses [50–55]. A FASTA file containing all the 3′ UTR regions of *C. porosus* genes were provided as the subject for miRNA queries against the newly annotated salt water crocodile genome [21]. Bedtools' "getfasta" command was used to identify the 3′UTR regions of genes and pipe it into the miRanda command line script.

For RNAhybrid, RNAcalibrate (a part of the RNAhybrid package) was run on the mature miRNA sequence FASTA file and the same 3′UTR sequence file as above. The free energy cut off was kept identical to miRanda (− 20 kcal/mol) and the p value was set at ≤ 0.01. In an effort to reduce false positives for RNAcalibrate and RNAhybrid, the seed region folding option of the mature miRNA (–f 2, 8) was again enforced.

2.6. miRNAs and transposable elements (TE) intersections

The TE profile in *C. porosus* was determined using Repeatmasker (version 4.0.6) [56]. The '-species Crocodylus' option was invoked. rmblastn version 2.2.27+ and RM database and RepBase version 20150807 was used. The coordinates of the mature miRNA sequences, corresponding star and hairpin precursor sequences were all obtained from the miRDeep2 run. The Bedtools "intersect" package was used to identify overlaps with coordinates generated by RepeatMasker.

2.7. Taxonomic distribution of miRNAs

We used MapMi (version 1.5.9-b31) to determine phylogenetic distributions of the identified miRNAs relative to other vertebrates. Mature miRNAs were mapped to the American alligator (*A. mississippiensis;* GCA_000281125.4), the Indian gharial (*G. gangeticus*; JRWT00000000), human (*Homo sapiens,* GCA_000001405.25), platypus (*Ornithorhynchus anatinus*; GCF_000002275.2), African clawed frog (*Xenopus tropicalis*; GCA_000004195.1), chicken (*Gallus gallus*; GCA_000002315.3), Anna's Hummingbird (*Calypte anna*; GCA_000699085.1), zebra finch (*Taeniopygia guttata*; ABQF01000000), green anole (*Anolis carolinensis*; GCA_000090745.1), painted turtle (*Chrysemys picta;* AHGY00000000.2) and the saltwater crocodile (*C. porosus*), by allowing 1 mismatch between the miRNAs and each genome. The genome assemblies were accessed from NCBI Genbank and www.ensembl.org (Ensembl Release 90). MapMi results were processed to retain only unique miRNAs in each species and numbers of shared and lineage-specific loci were plotted on a tree depicting established relationships [57]. Because the core algorithm of MapMi differs from that of miRDeep2, the total number of miRNAs mapped was reduced compared to the 644 reported with miRDeep2. For example, the MapMi algorithm assigns a score based on several biogenesis criteria to a predicted miRNA in a species. Based on matches with miRBase, the default cutoff for the MapMi score is 35 and removal of all miRNAs below this score is recommended by the MapMi program. These factors resulted in 29 fewer miRNAs being mapped through MapMi.

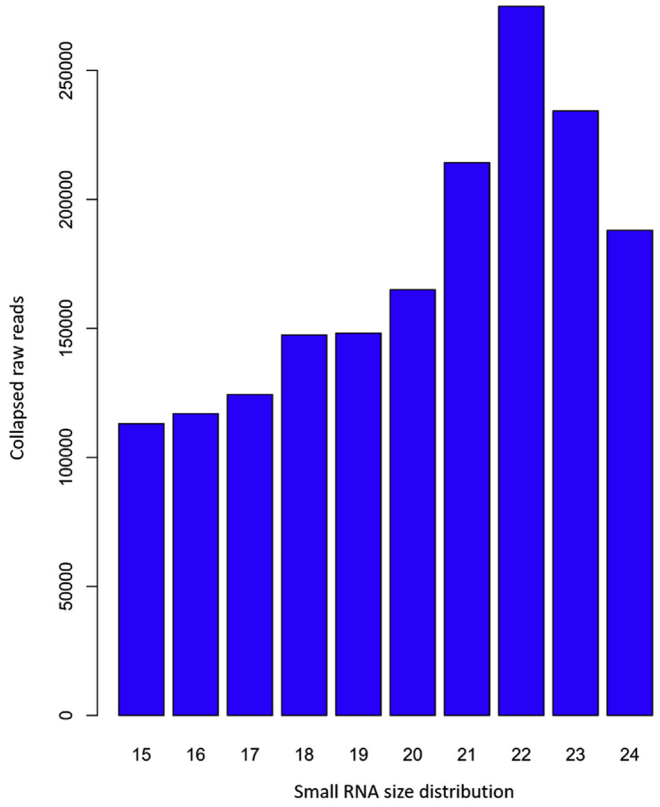

Fig. 2. Size class distribution of small RNA reads (collapsed) from all libraries.

3. Results

3.1. Prediction and annotation of known and novel miRNAs in C. porosus

644 unique miRNAs, along with their corresponding star and hairpin precursor sequences were identified. A majority of the small RNA reads fell in the size class of 21-nt and 22-nt. In animals, miRNAs predominate at these two size classes [58–60]. The total number of collapsed raw reads for each size class of small RNA reads from all 16 libraries combined is shown in Fig. 2. Although all small RNA size classes do not qualify for miRNA, our dataset of 644 miRNAs had size classes ranging between 18 through 22-nt. A 5′- T (or U) is a hallmark of eukaryotic miRNA biogenesis [61] and this bias was observed in the *C. porosus* data (Supplemental Fig. 1, Supplementary Material online).

As expected, the mature miRNAs, being the functional and stable component, were associated with higher read counts compared to the star sequences [62,63]. Out of the 644 miRNAs identified, 139 were known miRNAs from birds or other species as catalogued in miRbase (ver 21). Many highly conserved miRNAs like the let-7 family, miR-10, miR-1, miR-30, miR-7, miR-16 were identified in *C. porosus* suggesting that the additional, novel miRNAs are likely valid. A representation of some known and novel miRNAs in crocodilians and their hairpin precursors are shown as identified by miRDeep2 in this study (Fig. 3). A list of all mature miRNAs, both known and novel, in crocodilians identified in this study is provided in Supplemental file S1, Supplemental Material online. Several miRNAs with varying numbers of loci of origin were found. One deeply conserved miRNA (let7a/j -5p), had four distinct loci, while several others had 3 loci (cPor-miR-N143, cPor-miR-133-5p, cPor-miR-135a/b-5p, cPor-miR-199-5p to name a few).

3.2. Expression profiles of miRNAs identified in Crocodylus porosus

miRNAs are often expressed in a tissue specific manner. This can be attributed to signaling, compartmentalization, or pattern formation during development. It can also occur in response to localized stress or defense mechanisms and other factors [64–68]. The expression profile of the known and novel miRNAs in *C. porosus* varied, with the widely conserved miRNAs (and some novel crocodilian miRNAs) displaying very high expression in all tissue types and some novel crocodilian miRNAs peaking only in certain tissue types. There were also several miRNA candidates with very low expression in all tissue types. Despite this low expression, these candidate miRNAs passed all criteria of miRNA biogenesis, possessed all hallmarks of a true miRNA (as processed through the miRDeep2 pipeline) and were hence retained in the final dataset. Heatmaps were generated representing normalized expression profile for the top 20 known and novel crocodilian miRNAs as found in the different small RNA libraries in the saltwater crocodile (Fig. 4a and b).

3.3. Target prediction and annotation

We identified miRNA targets by using the most recent annotated *C. porosus* genome as reference [21] and enforcing strict seed complementarity at the 3' end in miRanda [52]. Previous studies indicate that this improves the binding score and provides more efficient target prediction [69–71]. To be conservative, we used miRanda with an alignment match score threshold of 170 and free energy threshold of −20 kcal/mol. One or multiple targets were identified for 583 of the 644 unique miRNAs. We attribute the reduction in number of identified miRNAs with targets from the total number of identified miRNAs to the steep alignment threshold score cutoff of 170, which is high compared to earlier studies that used a threshold score of 155–160 [52,53]. When the cutoff score was lowered to 160 (still high considering previous publications) while keeping the free energy value the same, miRanda identified targets for 640 out of the 644 miRNAs. We also performed miRNA target identification using RNAhybrid [72–76].

3.4. miRNAs and transposable elements

A majority of the TE-associated miRNA precursors are associated with Class II TEs, aka DNA transposons. A total of 283 miRNA-TE overlaps were found, of which nearly 78.3% were DNA transposons (subclass I + subclass II) overlaps, while 21.6% of the overlaps arose from Class I TEs, i.e. retrotransposons. The remaining 0.1% were derived from unidentified TEs.

Many of the TE-associated miRNA precursors appear to arise from these DNA transposons as they had the same orientation as the TE. Others were complimentary to the TE, possibly indicating the TE to be a target of miRNA-mediated regulation. However, this hypothesis is not testable using data for this study. Not all TE-derived miRNA precursors were fully overlapped by their respective TEs. Therefore, additional analyses were performed to identify four separate categories of miRNA-precursor and TE overlap situations. We expected the most prevalent category of these four types would be when the entire miRNA precursor (including the mature and star sequences) would overlap the TE. This was indeed the case, as ~92.5% of the miRNA-TE hits were found to be complete overlaps of the precursor and TE (Fig. 5). A list of all miRNA precursors that completely overlap with a TE is provided in Supplementary file S2, Supplemental Materials Online.

3.5. Orthologs of C. porosus miRNAs

To investigate the taxonomic distributions of *C. porosus* miRNAs, we selected ten species representing all four classes of vertebrates. miRNAs were mapped to the genomes of each species with MapMi [77]. Given established relationships [57], Fig. 6 shows the relationship of miRNA presence/absence within each of the selected vertebrates.

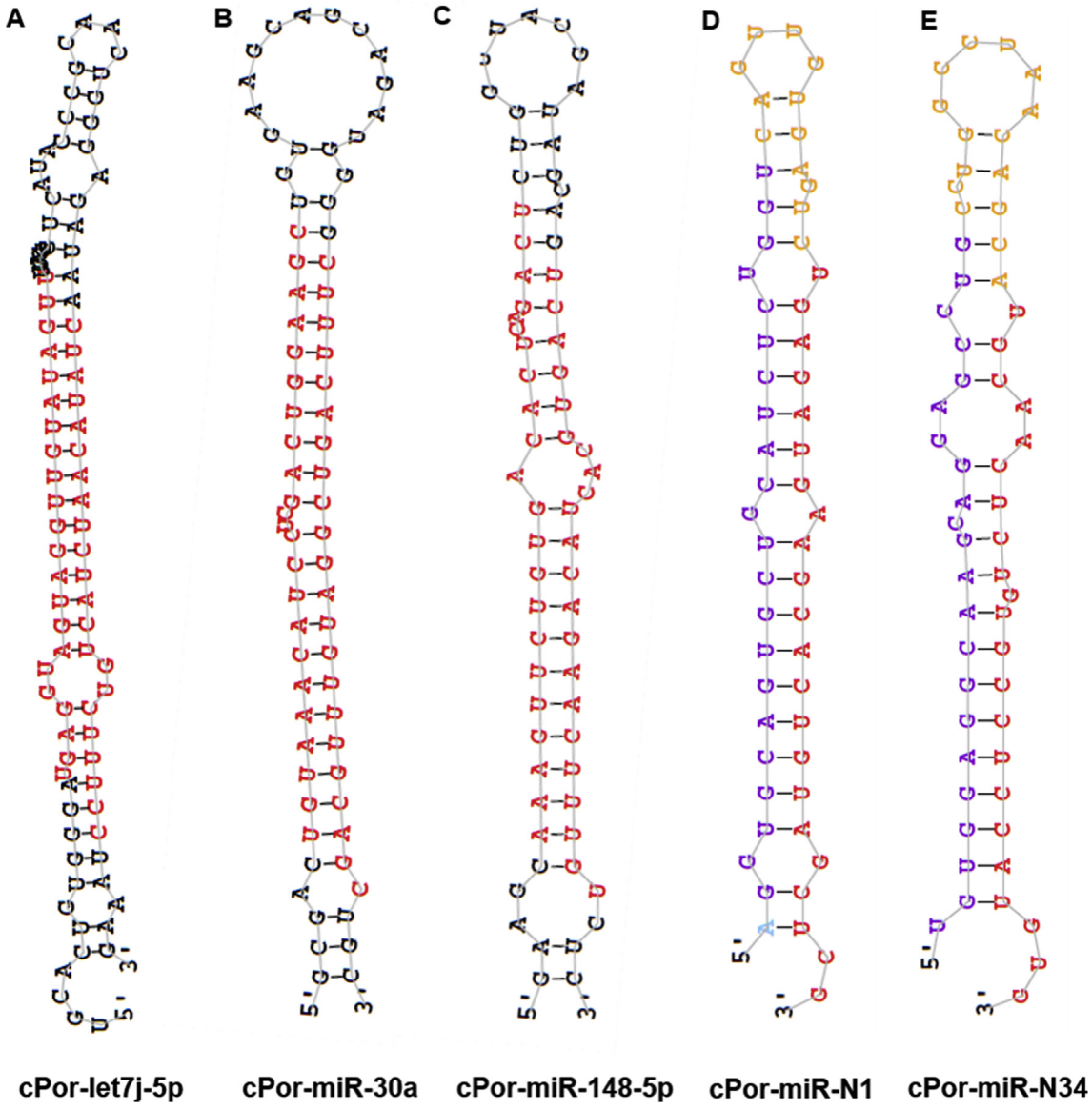

Fig. 3. Representation of five selected hairpin precursors of three known (A-C) and two novel crocodilian miRNAs (D-E) identified in the saltwater crocodile. For the known miRNAs (A-C), both the mature miRNA and star sequences are represented in red. For the novel crocodilian miRNAs (D-E), the mature miRNA is in red and the star sequence is in purple, while the loop is golden.

4. Discussion

This work represents the first *de novo* analysis of miRNAs in the saltwater crocodile. Small RNAs were isolated, sequenced and analyzed from eleven different tissues in as many as four different individuals per tissue. A total of 16 different small RNA libraries were prepared. More than 187 million reads were generated with each library represented by an average of ~11.7 million reads. Stringent quality control measures were taken to reduce low quality reads, and only those, where more than 50% of the reads scored higher than Phred 20 were accepted for downstream analysis. After filtering, the number of high-quality reads per library varied from just under a quarter million to more than 2.25 million.

For all 644 distinct miRNAs identified in this study, miRDeep2 identified mature and star sequences as well as miRNA precursors and precursor coordinates. A score calculated based on the miRdeep2 algorithm was assigned to each miRNA candidate. The higher the score, usually the better the predicted miRNA candidate. Examples of a mature miRNA and star sequence represented on their hairpin precursor as generated by miRDeep2 are shown in Supplemental Fig. 2, Supplementary Material online. The terms “gga” and “tgu” refer to miRNAs being already identified and present in chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*), respectively. A list of all mature miRNA consensus sequences, their corresponding consensus star sequences and precursor sequences generated by the miRDeep2 program is provided in Supplementary file S3, Supplementary Material online.

Several tissue specific miRNA expression profiles could be identified and corroborated from previous studies. One such example is miR-21, which is known to target several apoptotic genes in human spleen [12]. Given that the spleen is the site for expression of various apoptotic genes like Bcl-2, Bak-1, and Caspase-3, it was expected that the crocodilian spleen would also show high expression of miR-21 in *C. porosus* and this was observed as well (Fig. 4a). In murine tissues (*Mus musculus*), miR-21 is detected at much higher levels in liver, lung, kidney as compared to the brain [78]. Though kidney and lung tissues were absent in our small RNA library data, miR-21 expression data in *C. porosus* brain and liver tissues exhibited a similar pattern. miR-133 has been associated with mammalian cardiac muscle and regulation of cardiac

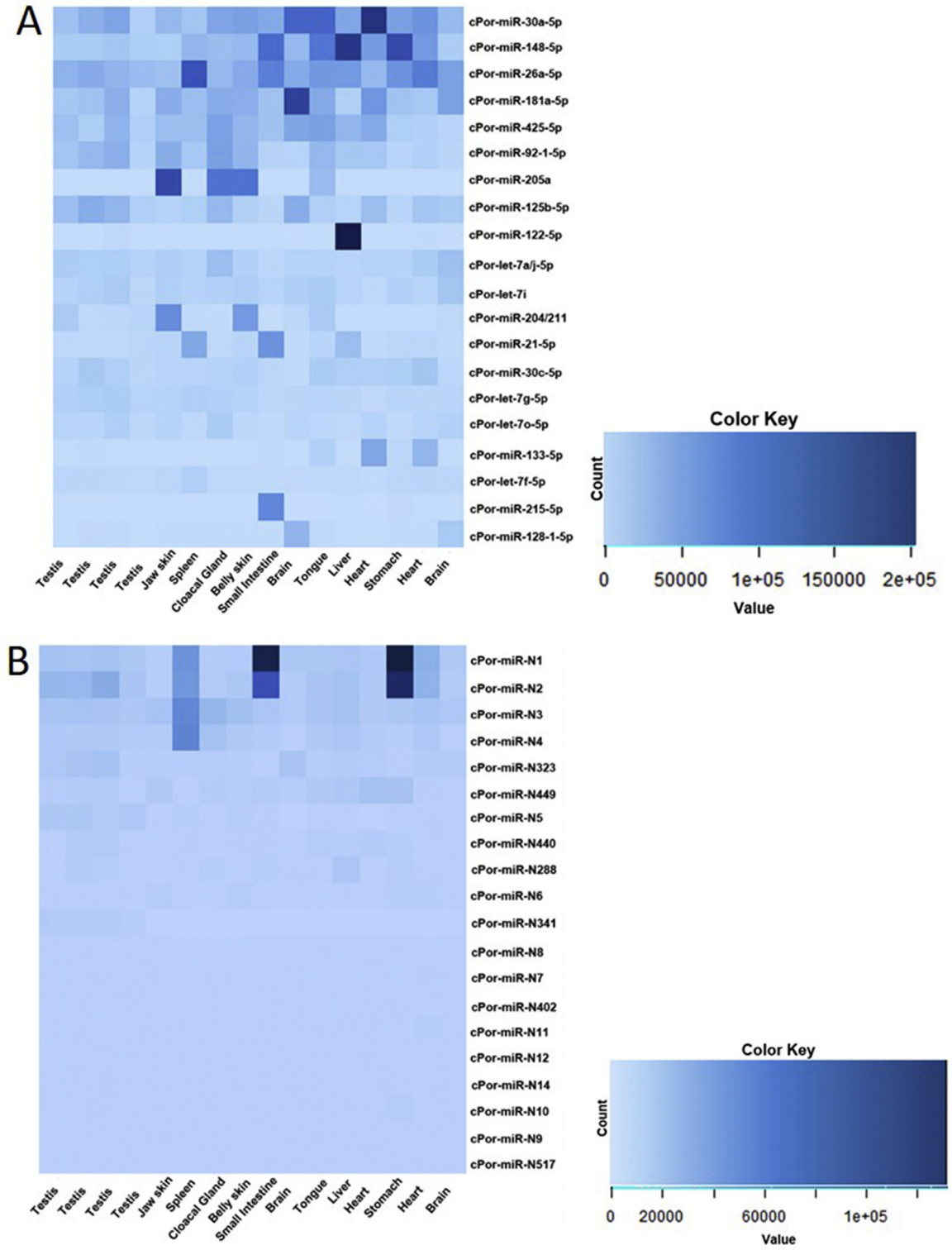

Fig. 4. Heatmap showing expression profile of the top 20 highest expressing miRNAs in *C. porosus.* 4a showing the known miRNAs while 4b shows the novel miRNAs in crocodilians.

hypertrophy as well as functioning as an inhibitor of cardiomyocyte proliferation [79]. Both miR-1 and miR-133 are bicistronic in the mammalian genome and can be thus transcribed together. They are encoded by loci that are duplicated bicistronic and hence they possess identical miRNA mature sequence. Both of these miRNAs are expressed in the heart [80]. We found that both miR-1 and miR-133 achieved peak expression in heart tissue in *C. porosus*.

Experimental validation of these numerous miRNAs was beyond the scope of this work. However, certain miRNAs like miR-1, miR-20, miR-21, miR-27, miR-30, miR-122, miR-126, miR-132, miR-199, miR-214, miR-338 and miR-363 identified in this study were found in the adult chicken [81] and chicken embryo [82,83] with small RNA blot validation. As expected, the patterns of tissue specific expression of these miRNAs are similar in both chicken and the saltwater crocodile. While some of these miRNAs are highly conserved, some are not and that certainly points to an interesting aspect of miRNA expression and

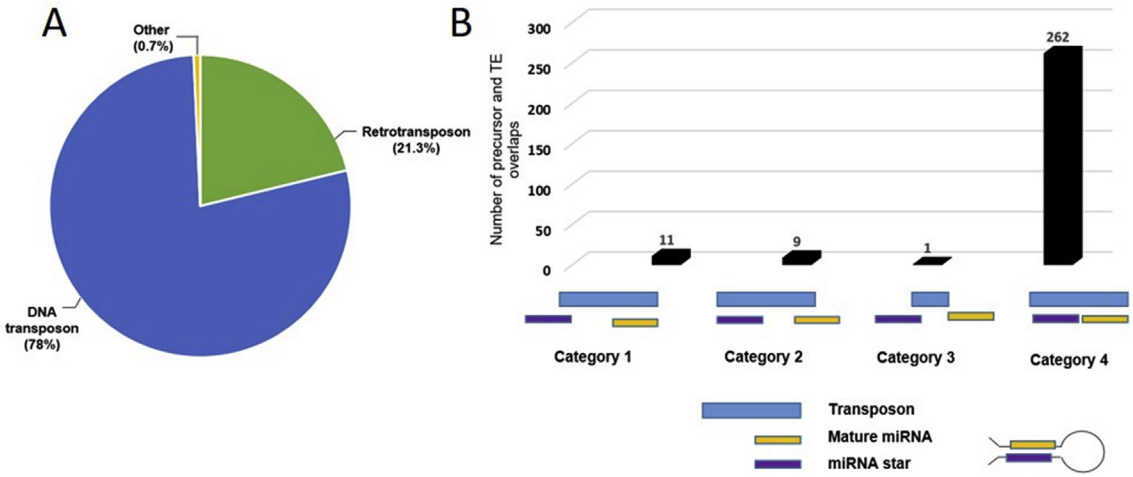

Fig. 5. miRNA-TE overlap rates in *C. porosus*. A) Rates of miRNA origination relative to TE class. B). Histogram depicting the numbers of *C. porosus* miRNAs that fall into the four overlap categories.

evolution in these two relatively closely related species. Our expression profiles clearly demonstrate highly tissue specific as well as conserved expression patterns (Fig. 7) suggesting that our analysis accurately identifies expression profiles for miRNAs.

With this assurance, we investigated the expression patterns of selected novel crocodilian miRNAs that exhibited tissue-specific patterns. cPor-miR-N1 peaked in expression in the gastro-intestinal regions (small intestine and stomach tissues) of the crocodile while two of the other novel candidates - cPor-miR-N16 and cPor-miR-N17 peaked in the spleen tissue. Some others like cPor-miR-N170 had highest expression in the tongue. Interestingly, cPor-miR-N341 - one of the novel miRNA candidates, had expression in all testes tissues, but it's expression was essentially zero to very low in all other tissues. Specific functions of these miRNAs are unknown but these patterns suggest regulatory processes related to these tissues. A comprehensive list of the normalized expression profiles of all identified miRNAs in this study as determined by the quantifier.pl script of the miRDeep2 program has been provided in Supplemental file S4, Supplemental Material online. Lastly, although not the focus of this work-it was observed that the total normalized expression of miRNAs in the heart tissue type was higher in the normal hatchlings as compared to the runts. The opposite was true for the miRNA expression pattern found in the brain between the two types of hatchlings. In future work, it will be interesting to follow up on a comparative miRNA expression profiles of normal and runt hatchlings.

Although miRNA targeting in animals primarily takes place in the 3′ UTR region of targeted genes, reports have shown the 5′ UTR of genes and open reading frames of some highly repeat rich elements to be miRNA targets as well [84–86]. Binding and targeting are most effective if there is complete complementarity between the target with the seed region, the 2-8th position on the 5' end of the mature miRNA [1,87–92]. It has been noted previously that target prediction tools which require Watson-Crick base pairing with the query miRNA seed region, perform better in predicting true targets as compared to those programs that do not have this requirement [93,94].

As mentioned above, we used miRanda and RNAhybrid to predict the miRNA targets. RNAhybrid mainly works based on free energy of folding of miRNA around the target and a calibration step improves overall target estimation for the miRNAs. For RNAhybrid, we used the same value for free energy cutoff as that of miRanda but found that the miRanda results were more comprehensive. Thus, the miRNA-target data shown here only corresponds to those identified by miRanda. The four miRNA candidates that failed to be predicted by miRanda with a

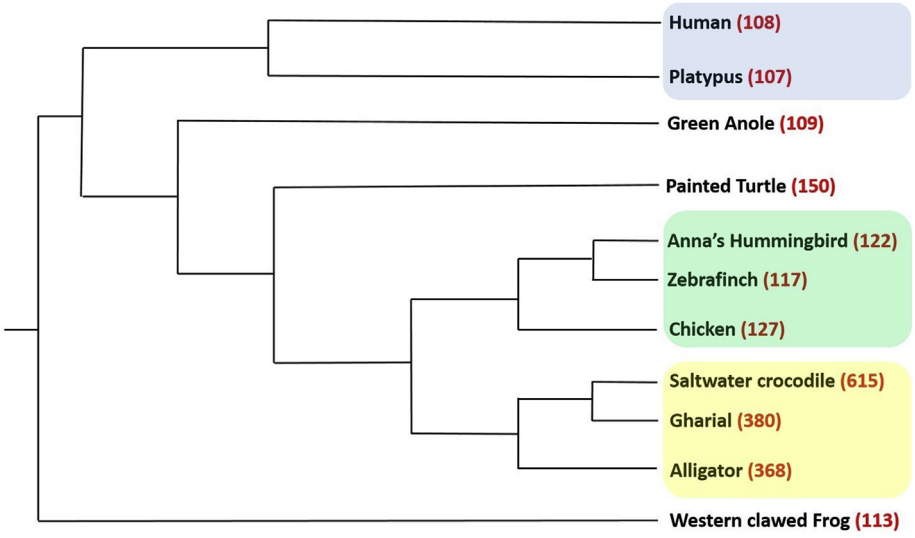

Fig. 6. Phylogeny of vertebrates examined for orthologous miRNAs. Numbers in parentheses indicate the number of *C. porosus* miRNAs identified in this study that are also identifiable in each species. The tree is based on and redrawn from Chiari et al. [57].

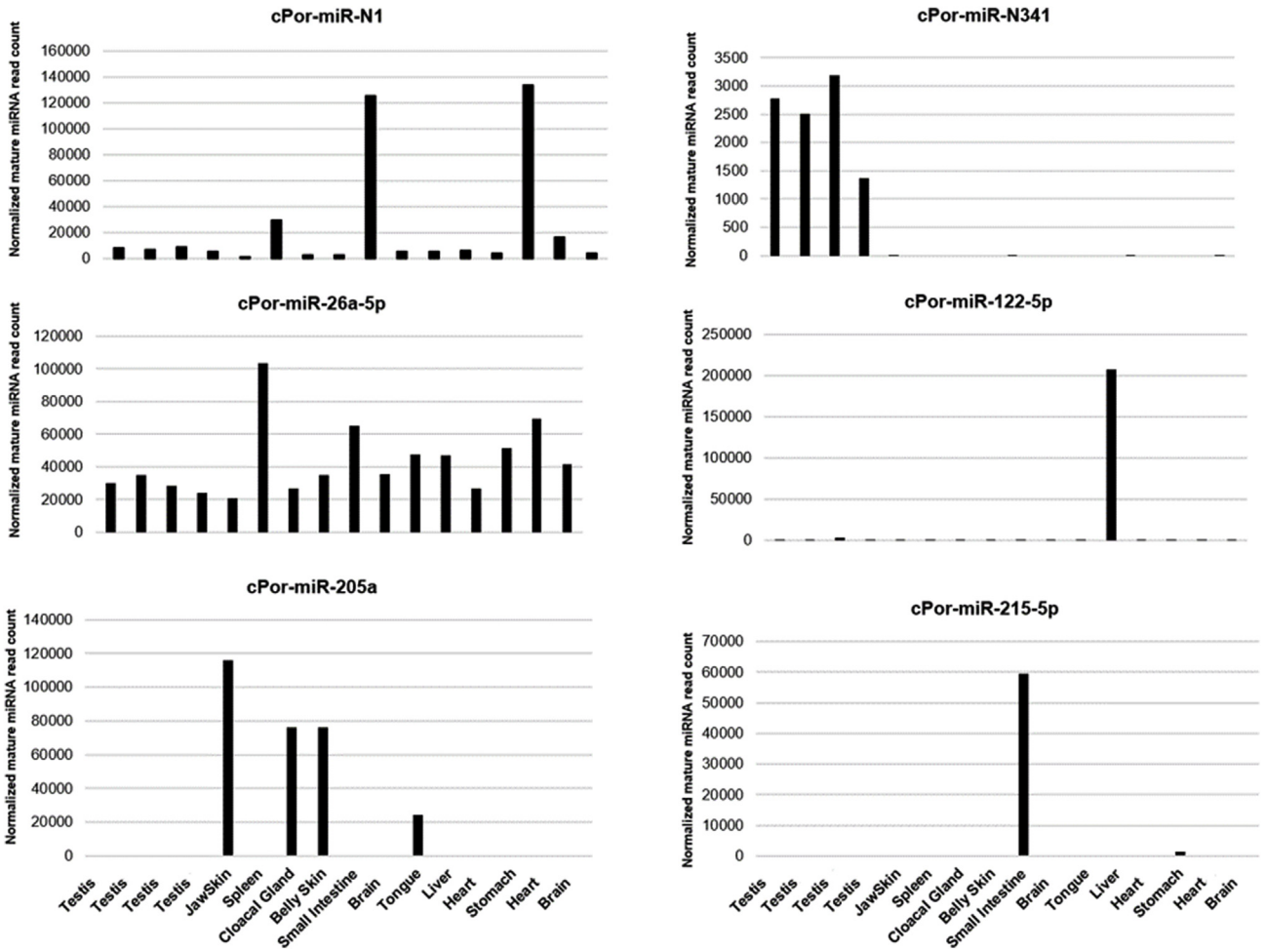

Fig. 7. Tissue-specific expression profiles for selected known and novel *C. porosus* miRNAs.

lowered complementary match cutoff score of 160 included miR-32-5p, miR-126-5p, miR-N512 and miR-208a. All of these miRNA candidates had zero normalized counts in multiple libraries and extremely low total normalized count overall. However, having satisfied all criteria of being a true miRNA by the miRDeep2 pipeline and possessing all hallmarks of a true miRNA biogenesis, these candidates were retained. It is possible, that with a larger target dataset, these miRNAs might find a suitable target protein.

Because of the significant role miRNAs play in posttranscriptional gene regulation and the increasing focus on TEs as a major player in the same, we investigated the relationship between the two. Our results indicate that nearly half (44%) of the miRNAs we identified overlapped in some way with TE loci and 94% of those arose completely from within the TE. DNA transposons are the mobile element fraction most highly associated with miRNAs. A mixed distribution of both complementary and sense orientated population of miRNA (with respect to the TEs) suggested that miRNAs could target the TEs from which they originated. This suggests a potential mechanism for suppressing the expression of the associated TEs. This is interesting in that TEs have not accumulated at significant rates in crocodilians over the past several million years [18].

Our results are in good agreement with previous studies that demonstrate the role of TEs in the evolution of functional miRNAs [44,95]. Specifically, 62% of miRNAs in the vespertilionid bat *Eptesicus fucus* have been known to be associated with TEs while that percentage changes to 23.9% and 16.2% in dog and horse respectively [96]. In the same vein, miRNA association has been found to be more abundant with retrotransposons in other mammals, including humans. In addition, miRNAs have been also known to be primarily associated with DNA transposons in the zebrafish and the western clawed frog [97]. This further relates TEs to the evolution of phenotypic and phylogenetic diversity. Several hypotheses exist suggesting that TEs play a significant role in mechanisms of diversification [98–100]. There is also some evidence for the temporal correlation between bursts of TE accumulation (and therefore miRNA deposition) with speciation [101–106]. Thus, we believe that the TE-derived miRNAs described here are good candidates for further investigation with regards to identifying lineage-specific functional differences in gene expression across Crocodylia.

Finally, when analyzing the origin of *C. porosus* miRNAs, we observed that as taxonomic distance increases, the number of shared miRNAs decreases. The vast majority of miRNAs recovered from *C. porosus* were identifiable in other crocodilians. As would be expected given, the more recent divergence of gharials from true crocodiles (~55 mya vs. ~85 mya for the alligator divergence) [107], more miRNAs were shared by *G. gangeticus* and *C. porosus*, when compared to the two available alligator genomes (*A. mississippiensis* and *A. sinensis*) (Fig. 8).

5. Conclusion

We identified both known and novel miRNAs in the saltwater crocodile, many of which are orthologous to several other vertebrates. A weakness of this work was our inability to validate miRNAs by identifying associations with RISC proteins or demonstrating definitive impacts on mRNAs and/or phenotype. Future work including target validation through 5' Rapid amplification of cDNA ends (RACE) as well as approaches like degradome assays [108,109] might provide target

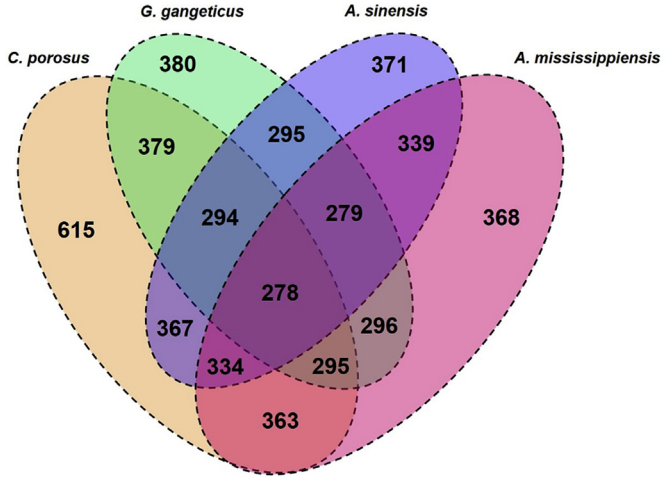

Fig. 8. Homologous miRNAs shared among sequenced crocodilian genomes. The numbers indicate number of orthologous crocodilian miRNAs that are shared between respective sister clades of order Crocodilia.

validation of some of these miRNAs identified here. However, our conservative scoring strategies and the fact that many of the miRNAs we identified were orthologous with other known and validated miRNAs suggests that our pipelines were effective.

The data here provide an intriguing start to analyses of gene regulation in this ecologically important clade that is, nonetheless, rarely studied from a molecular perspective. Additional analyses, especially those that can be compared with the only other extant member of Archosauria, birds, could provide insight into the ancient and extinct relatives of both groups, the ever-interesting dinosaurs as well as into the evolution of gene regulation in non-mammalian vertebrates. For example, similar and/or identical miRNA target sites can reveal potentially conserved target sites among vertebrates. More direct applications arise from an increased understanding of gene regulation as it may relate to the development of commercially important features, especially hide, which is highly valued.

Author contributions

Arnab Ghosh: Conceptualization, Methodology, Visualization, Writing-Original Draft, Writing-Review & Editing, Data Curation, Investigation, Formal Analysis, Validation. **Roy N. Platt II**: Methodology, Writing-Reviewing & Editing. **Michael W. Vandewege**: Methodology, Writing-Reviewing & Editing. **Rabia Tabassum**: Data Curation, Writing-Reviewing & Editing, Resources . **Chuan-Yu Hsu**: Methodology, Investigation. **Sally R. Isberg**: Conceptualization, Supervision, Funding Acquisition, Writing-Review & Editing, Resources, Funding Acquisition, Project Administration. **Daniel G. Peterson**: Conceptualization, Funding Acquisition, Supervision Writing-Review & Editing. **John W. Finger**: Resources, Writing-Review & Editing **Troy J. Kieran**: Methodology, Investigation. **Travis C. Glenn**: Conceptualization, Methodology, Visualization, Writing-Original Draft, Writing-Review & Editing, Data Curation, Formal Analysis, Funding Acquisition, Project Administration. **Jaime Gongora**: Conceptualization, Writing-Original Draft, Writing-Review & Editing, Resources, Funding Acquisition, Project Administration. **David A. Ray**: Conceptualization, Methodology, Visualization, Writing-Original Draft, Writing-Review & Editing, Data Curation, Investigation, Validation, Funding Acquisition, Project Administration.

Declaration of competing interest

None.

Acknowledgements

This work was supported by the National Science Foundation (MCB-0841821, DEB-1020865, and DEB-1838283 to D.A.R.) and Rural Industries Research and Development Corporation (PRJ-005355 and PRJ-002461 to S.I. and J.G.). A University of Sydney Bridging Support grant to J.G. also supported the project. The High Performance Computing Center (HPCC) at Texas Tech University provided substantial logistical support. All miRNAs and their precursor sequences have been deposited at miRBase (www.miRBase.org).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ab.2020.113781.

References

[1] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2) (2004) 281–297.
[2] R.W. Carthew, E.J. Sontheimer, Origins and Mechanisms of miRNAs and siRNAs, Cell 136 (4) (2009) 642–655, https://doi.org/10.1016/j.cell.2009.01.035.
[3] K. Chen, N. Rajewsky, The evolution of gene regulation by transcription factors and microRNAs, Nat. Rev. Genet. 8 (2) (2007) 93–103, https://doi.org/10.1038/nrg1990.
[4] D.M. Tyler, K. Okamura, W.J. Chung, J.W. Hagen, E. Berezikov, G.J. Hannon, E.C. Lai, Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci, Genes Dev. 22 (1) (2008) 26–36, https://doi.org/10.1101/gad.1615208.
[5] A. Stark, N. Bushati, C.H. Jan, P. Kheradpour, E. Hodges, J. Brennecke, D.P. Bartel, S.M. Cohen, M. Kellis, A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands, Genes Dev. 22 (1) (2008) 8–13, https://doi.org/10.1101/gad.1613108.
[6] W. Bender, MicroRNAs in the Drosophila bithorax complex, Genes Dev. 22 (1) (2008) 14–19.
[7] A. Grimson, K.K. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, D.P. Bartel, MicroRNA targeting specificity in mammals: determinants beyond seed pairing, Mol. Cell 27 (1) (2007) 91–105, https://doi.org/10.1016/j.molcel.2007.06.017.
[8] I. Güller, A.P. Russell, MicroRNAs in skeletal muscle: their role and regulation in development, disease and function, J. Physiol. 588 (21) (2010) 4075–4087.
[9] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (1) (2005) 15–20.
[10] T.W. Nilsen, Mechanisms of microRNA-mediated gene regulation in animal cells, Trends Genet. 23 (5) (2007) 243–249, https://doi.org/10.1016/j.tig.2007.02.011.
[11] R.C. Friedman, K.K.-H. Farh, C.B. Burge, D.P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs, Genome Res. 19 (1) (2009) 92–105.
[12] J.A. Chan, A.M. Krichevsky, K.S. Kosik, MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells, Canc. Res. 65 (14) (2005) 6029–6033, https://doi.org/10.1158/0008-5472.CAN-05-0137.
[13] K.N. Ivey, D. Srivastava, MicroRNAs as regulators of differentiation and cell fate decisions, Cell stem cell 7 (1) (2010) 36–41.
[14] A.K. Leung, P.A. Sharp, MicroRNA functions in stress responses, Mol. Cell 40 (2) (2010) 205–215, https://doi.org/10.1016/j.molcel.2010.09.027.
[15] M. Ronshaugen, F. Biemar, J. Piel, M. Levine, E.C. Lai, The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings, Genes Dev. 19 (24) (2005) 2947–2952, https://doi.org/10.1101/gad.1372505.
[16] A. Janke, U. Arnason, The complete mitochondrial genome of Alligator mississippiensis and the separation between recent archosauria (birds and crocodiles), Mol. Biol. Evol. 14 (12) (1997) 1266–1272.
[17] A. Sennikov, The first ctenosauriscid (Reptilia: Archosauromorpha) from the lower Triassic of eastern Europe, Paleontol. J. 46 (5) (2012) 499–511.
[18] R.E. Green, E.L. Braun, J. Armstrong, D. Earl, N. Nguyen, G. Hickey, M.W. Vandewege, J.A. St John, S. Capella-Gutierrez, T.A. Castoe, C. Kern, M.K. Fujita, J.C. Opazo, J. Jurka, K.K. Kojima, J. Caballero, R.M. Hubley, A.F. Smit, R.N. Platt, C.A. Lavoie, M.P. Ramakodi, J.W. Finger Jr., A. Suh, S.R. Isberg, L. Miles, A.Y. Chong, W. Jaratlerdsiri, J. Gongora, C. Moran, A. Iriarte, J. McCormack, S.C. Burgess, S.V. Edwards, E. Lyons, C. Williams, M. Breen, J.T. Howard, C.R. Gresham, D.G. Peterson, J. Schmitz, D.D. Pollock, D. Haussler, E.W. Triplett, G. Zhang, N. Irie, E.D. Jarvis, C.A. Brochu, C.J. Schmidt, F.M. McCarthy, B.C. Faircloth, F.G. Hoffmann, T.C. Glenn, T. Gabaldon, B. Paten, D.A. Ray, Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs, Science 346 (6215) (2014) 1254449, https://doi.org/10.1126/science.1254449.
[19] E.S. Rice, S. Kohno, J.S. John, S. Pham, J. Howard, L.F. Lareau, B.L. O'Connell, G. Hickey, J. Armstrong, A. Deran, I. Fiddes, R.N. Platt 2nd, C. Gresham, F. McCarthy, C. Kern, D. Haan, T. Phan, C. Schmidt, J.R. Sanford, D.A. Ray, B. Paten, L.J. Guillette Jr., R.E. Green, Improved genome assembly of American alligator genome reveals conserved architecture of estrogen signaling, Genome Res. 27 (5) (2017) 686–696, https://doi.org/10.1101/gr.213595.116.

[20] T.R. Lyson, E.A. Sperling, A.M. Heimberg, J.A. Gauthier, B.L. King, K.J. Peterson, MicroRNAs support a turtle + lizard clade, Biol. Lett. 8 (1) (2012) 104–107, https://doi.org/10.1098/rsbl.2011.0477.
[21] A. Ghosh, M.G. Johnson, A.B. Osmanski, S. Louha, N.J. Bayona-Vasquez, T.C. Glenn, J. Gongora, R.E. Green, S. Isberg, R.D. Stevens, D.A. Ray, A high-quality reference genome assembly of the saltwater crocodile, Crocodylus porosus, reveals patterns of selection in Crocodylidae, Genome Biol. Evol. 12 (1) (2020) 3635–3646, https://doi.org/10.1093/gbe/evz269.
[22] C.G. Sotero-Caio, R.N. Platt 2nd, A. Suh, D.A. Ray, Evolution and diversity of transposable elements in vertebrate genomes, Genome Biol. Evol. 9 (1) (2017) 161–177, https://doi.org/10.1093/gbe/evw264.
[23] D.V. Babushok, E.M. Ostertag, H.H. Kazazian Jr., Current topics in genome evolution: molecular mechanisms of new gene formation, Cell. Mol. Life Sci. 64 (5) (2007) 542–554, https://doi.org/10.1007/s00018-006-6453-4.
[24] C. Feschotte, Transposable elements and the evolution of regulatory networks, Nat. Rev. Genet. 9 (5) (2008) 397–405, https://doi.org/10.1038/nrg2337.
[25] J. Jurka, Origin and Evolution of Alu Repetitive Elements. The Impact of Short Interspersed Elements (SINEs) on the Host Genome, (1995), pp. 25–41.
[26] H.H. Kazazian Jr., Mobile elements: drivers of genome evolution, Science 303 (5664) (2004) 1626–1632, https://doi.org/10.1126/science.1089670.
[27] G. Kunarso, N.-Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.-S. Chan, H.-H. Ng, G. Bourque, Transposable elements have rewired the core regulatory network of human embryonic stem cells, Nat. Genet. 42 (7) (2010) 631.
[28] V.J. Lynch, R.D. Leclerc, G. May, G.P. Wagner, Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals, Nat. Genet. 43 (11) (2011) 1154–1159, https://doi.org/10.1038/ng.917.
[29] P. Nigumann, K. Redik, K. Matlik, M. Speek, Many human genes are transcribed from the antisense promoter of L1 retrotransposon, Genomics 79 (5) (2002) 628–634, https://doi.org/10.1006/geno.2002.6758.
[30] A.E. Peaston, A.V. Evsikov, J.H. Graber, W.N. de Vries, A.E. Holbrook, D. Solter, B.B. Knowles, Retrotransposons regulate host genes in mouse oocytes and pre-implantation embryos, Dev. Cell 7 (4) (2004) 597–606, https://doi.org/10.1016/j.devcel.2004.09.004.
[31] M. Speek, Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes, Mol. Cell Biol. 21 (6) (2001) 1973–1985, https://doi.org/10.1128/MCB.21.6.1973-1985.2001.
[32] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, Nat. Rev. Genet. 8 (12) (2007) 973–982, https://doi.org/10.1038/nrg2165.
[33] E.B. Chuong, N.C. Elde, C. Feschotte, Regulatory activities of transposable elements: from conflicts to benefits, Nat. Rev. Genet. 18 (2) (2017) 71–86, https://doi.org/10.1038/nrg.2016.139.
[34] J. Judd, C. Feschotte, Transposons take remote control, Elife 7 (2018), https://doi.org/10.7554/eLife.40921.
[35] P. Krishnan, L. Meile, C. Plissonneau, X. Ma, F.E. Hartmann, D. Croll, B.A. McDonald, A. Sanchez-Vallet, Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat, BMC Biol. 16 (1) (2018) 78, https://doi.org/10.1186/s12915-018-0543-2.
[36] M. Trizzino, A. Kapusta, C.D. Brown, Transposable elements generate regulatory novelty in a tissue-specific fashion, BMC Genom. 19 (1) (2018) 468, https://doi.org/10.1186/s12864-018-4850-3.
[37] M. Trizzino, Y. Park, M. Holsbach-Beltrame, K. Aracena, K. Mika, M. Caliskan, G.H. Perry, V.J. Lynch, C.D. Brown, Transposable elements are the primary source of novelty in primate gene regulation, Genome Res. 27 (10) (2017) 1623–1633, https://doi.org/10.1101/gr.218149.116.
[38] K. Ahn, J.-A. Gim, H.-S. Ha, K. Han, H.-S. Kim, The novel MER transposon-derived miRNAs in human genome, Gene 512 (2) (2013) 422–428.
[39] G.M. Borchert, N.W. Holton, J.D. Williams, W.L. Hernan, I.P. Bishop, J.A. Dembosky, J.E. Elste, N.S. Gregoire, J.A. Kim, W.W. Koehler, J.C. Lengerich, A.A. Medema, M.A. Nguyen, G.D. Ower, M.A. Rarick, B.N. Strong, N.J. Tardi, N.M. Tasker, D.J. Wozniak, C. Gatto, E.D. Larson, Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins, Mobile Genet. Elem. 1 (1) (2011) 8–17, https://doi.org/10.4161/mge.1.1.15766.
[40] E.J. Devor, A.S. Peek, W. Lanier, P.B. Samollow, Marsupial-specific microRNAs evolved from marsupial-specific transposable elements, Gene 448 (2) (2009) 187–191, https://doi.org/10.1016/j.gene.2009.06.019.
[41] B. Mugat, A. Akkouche, V. Serrano, C. Armenise, B. Li, C. Brun, T.A. Fulga, D. Van Vactor, A. Pelisson, S. Chambeyron, MicroRNA-dependent transcriptional silencing of transposable elements in Drosophila Follicle cells, PLoS Genet. 11 (5) (2015) e1005194, , https://doi.org/10.1371/journal.pgen.1005194.
[42] J. Piriyapongsa, I.K. Jordan, A family of human microRNA genes from miniature inverted-repeat transposable elements, PloS One 2 (2) (2007) e203, , https://doi.org/10.1371/journal.pone.0000203.
[43] J. Piriyapongsa, L. Marino-Ramirez, I.K. Jordan, Origin and evolution of human microRNAs from transposable elements, Genetics 176 (2) (2007) 1323–1337, https://doi.org/10.1534/genetics.107.072553.
[44] J.T. Roberts, S.E. Cardin, G.M. Borchert, Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences, Mobile Genet. Elem. 4 (3) (2014) e29255.
[45] N.R. Smalheiser, V.I. Torvik, Mammalian microRNAs derived from genomic repeats, Trends Genet. 21 (6) (2005) 322–326, https://doi.org/10.1016/j.tig.2005.04.008.
[46] J.W. Finger Jr., P.C. Thomson, A.L. Adams, S. Benedict, C. Moran, S.R. Isberg, Reference levels for corticosterone and immune function in farmed saltwater crocodiles (Crocodylus porosus) hatchlings using current Code of Practice guidelines, Gen. Comp. Endocrinol. 212 (2015) 63–72, https://doi.org/10.1016/j.ygcen.2015.01.023.
[47] C. Shilton, G.P. Brown, L. Chambers, S. Benedict, S. Davis, S. Aumann, S.R. Isberg, Pathology of runting in farmed saltwater crocodiles (Crocodylus porosus) in Australia, Vet. Pathol. 51 (5) (2014) 1022–1034, https://doi.org/10.1177/0300985813516642.
[48] M.R. Friedlander, S.D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, Nucleic Acids Res. 40 (1) (2012) 37–52, https://doi.org/10.1093/nar/gkr688.
[49] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.
[50] W. Chen, W. Zhao, A. Yang, A. Xu, H. Wang, M. Cong, T. Liu, P. Wang, H. You, Integrated analysis of microRNA and gene expression profiles reveals a functional regulatory module associated with liver fibrosis, Gene 636 (2017) 87–95, https://doi.org/10.1016/j.gene.2017.09.027.
[51] C. Cowled, C.R. Stewart, V.A. Likic, M.R. Friedlander, M. Tachedjian, K.A. Jenkins, M.L. Tizard, P. Cottee, G.A. Marsh, P. Zhou, M.L. Baker, A.G. Bean, L.F. Wang, Characterisation of novel microRNAs in the Black flying fox (Pteropus alecto) by deep sequencing, BMC Genom. 15 (2014) 682, https://doi.org/10.1186/1471-2164-15-682.
[52] A.J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, D.S. Marks, MicroRNA targets in Drosophila, Genome Biol. 5 (1) (2003) R1, https://doi.org/10.1186/gb-2003-5-1-r1.
[53] E.D. Hutchins, W.L. Eckalbar, J.M. Wolter, M. Mangone, K. Kusumi, Differential expression of conserved and novel microRNAs during tail regeneration in the lizard Anolis carolinensis, BMC Genom. 17 (2016) 339, https://doi.org/10.1186/s12864-016-2640-3.
[54] X.G. Wang, J.F. Yu, Y. Zhang, D.Q. Gong, Z.L. Gu, Identification and characterization of microRNA from chicken adipose tissue and skeletal muscle, Poultry Sci. 91 (1) (2012) 139–149, https://doi.org/10.3382/ps.2011-01656.
[55] L. Zhang, P.S. Sullivan, J.C. Goodman, P.H. Gunaratne, D. Marchetti, MicroRNA-1258 suppresses breast cancer brain metastasis by targeting heparanase, Canc. Res. 71 (3) (2011) 645–654, https://doi.org/10.1158/0008-5472.CAN-10-1910.
[56] A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0. 1996–2004, Institute for Systems Biology, 2004.
[57] Y. Chiari, V. Cahais, N. Galtier, F. Delsuc, Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria), BMC Biol. 10 (1) (2012) 65.
[58] T. Du, P.D. Zamore, microPrimer: the biogenesis and function of microRNA, Development 132 (21) (2005) 4645–4652, https://doi.org/10.1242/dev.02070.
[59] Y.K. Kim, V.N. Kim, Processing of intronic microRNAs, EMBO J. 26 (3) (2007) 775–783, https://doi.org/10.1038/sj.emboj.7601512.
[60] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, V.N. Kim, The nuclear RNase III Drosha initiates microRNA processing, Nature 425 (6956) (2003) 415–419, https://doi.org/10.1038/nature01957.
[61] F. Frank, N. Sonenberg, B. Nagar, Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2, Nature 465 (7299) (2010) 818–822, https://doi.org/10.1038/nature09039.
[62] W. Filipowicz, S.N. Bhattacharyya, N. Sonenberg, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9 (2) (2008) 102–114, https://doi.org/10.1038/nrg2290.
[63] L. Guo, Z. Lu, The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? PloS One 5 (6) (2010) e11387.
[64] L.P. Lim, N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, J.M. Johnson, Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs, Nature 433 (7027) (2005) 769.
[65] J.H. Mansfield, B.D. Harfe, R. Nissen, J. Obenauer, J. Srineel, A. Chaudhuri, R. Farzan-Kashani, M. Zuker, A.E. Pasquinelli, G. Ruvkun, P.A. Sharp, C.J. Tabin, M.T. McManus, MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression, Nat. Genet. 36 (10) (2004) 1079–1083, https://doi.org/10.1038/ng1421.
[66] E.A. Miska, E. Alvarez-Saavedra, M. Townsend, A. Yoshii, N. Sestan, P. Rakic, M. Constantine-Paton, H.R. Horvitz, Microarray analysis of microRNA expression in the developing mammalian brain, Genome Biol. 5 (9) (2004) R68, https://doi.org/10.1186/gb-2004-5-9-r68.
[67] G. Wheeler, S. Ntounia-Fousara, B. Granda, T. Rathjen, T. Dalmay, Identification of new central nervous system specific mouse microRNAs, FEBS Lett. 580 (9) (2006) 2195–2200, https://doi.org/10.1016/j.febslet.2006.03.019.
[68] E. Wienholds, W.P. Kloosterman, E. Miska, E. Alvarez-Saavedra, E. Berezikov, E. de Bruijn, H.R. Horvitz, S. Kauppinen, R.H. Plasterk, MicroRNA expression in zebrafish embryonic development, Science 309 (5732) (2005) 310–311, https://doi.org/10.1126/science.1114519.
[69] B.J. Reinhart, F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, G. Ruvkun, The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans, Nature 403 (6772) (2000) 901–906, https://doi.org/10.1038/35002607.
[70] F.J. Slack, M. Basson, Z. Liu, V. Ambros, H.R. Horvitz, G. Ruvkun, The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor, Mol. Cell 5 (4) (2000) 659–669.
[71] M.C. Vella, E.Y. Choi, S.Y. Lin, K. Reinert, F.J. Slack, The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR, Genes Dev. 18 (2) (2004) 132–137, https://doi.org/10.1101/gad.1165404.

[72] J. Kruger, M. Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible, Nucleic Acids Res. 34 (2006) W451–W454, https://doi.org/10.1093/nar/gkl243 Web Server issue.
[73] M. Rehmsmeier, P. Steffen, M. Hochsmann, R. Giegerich, Fast and effective prediction of microRNA/target duplexes, RNA 10 (10) (2004) 1507–1517, https://doi.org/10.1261/rna.5248604.
[74] D. Betel, A. Koppal, P. Agius, C. Sander, C. Leslie, Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, Genome Biol. 11 (8) (2010) R90, https://doi.org/10.1186/gb-2010-11-8-r90.
[75] W.L. Eckalbar, E.D. Hutchins, G.J. Markov, A.N. Allen, J.J. Corneveaux, K. Lindblad-Toh, F. Di Palma, J. Alfoldi, M.J. Huentelman, K. Kusumi, Genome reannotation of the lizard Anolis carolinensis based on 14 adult and embryonic deep transcriptomes, BMC Genom. 14 (2013) 49, https://doi.org/10.1186/1471-2164-14-49.
[76] P. Landgraf, M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A.O. Kamphorst, M. Landthaler, C. Lin, N.D. Socci, L. Hermida, V. Fulci, S. Chiaretti, R. Foa, J. Schliwka, U. Fuchs, A. Novosel, R.U. Muller, B. Schermer, U. Bissels, J. Inman, Q. Phan, M. Chien, D.B. Weir, R. Choksi, G. De Vita, D. Frezzetti, H.I. Trompeter, V. Hornung, G. Teng, G. Hartmann, M. Palkovits, R. Di Lauro, P. Wernet, G. Macino, C.E. Rogler, J.W. Nagle, J. Ju, F.N. Papavasiliou, T. Benzing, P. Lichter, W. Tam, M.J. Brownstein, A. Bosio, A. Borkhardt, J.J. Russo, C. Sander, M. Zavolan, T. Tuschl, A mammalian microRNA expression atlas based on small RNA library sequencing, Cell 129 (7) (2007) 1401–1414, https://doi.org/10.1016/j.cell.2007.04.040.
[77] J.A. Guerra-Assuncao, A.J. Enright, MapMi: automated mapping of microRNA loci, BMC Bioinf. 11 (2010) 133, https://doi.org/10.1186/1471-2105-11-133.
[78] J.M. Thomson, J. Parker, C.M. Perou, S.M. Hammond, A custom microarray platform for analysis of microRNA gene expression, Nat. Methods 1 (1) (2004) 47–53, https://doi.org/10.1038/nmeth704.
[79] M. Abdellatif, The role of microRNA-133 in cardiac hypertrophy uncovered, Circ. Res. 106 (1) (2010) 16–18, https://doi.org/10.1161/CIRCRESAHA.109.212183.
[80] B. Meder, H.A. Katus, W. Rottbauer, Right into the heart of microRNA-133a, Genes Dev. 22 (23) (2008) 3227–3231.
[81] H. Xu, X. Wang, Z. Du, N. Li, Identification of microRNAs from different tissues of chicken embryo and adult chicken, FEBS Lett. 580 (15) (2006) 3610–3616, https://doi.org/10.1016/j.febslet.2006.05.044.
[82] D.K. Darnell, S. Kaur, S. Stanislaw, J.H. Konieczka, T.A. Yatskievych, P.B. Antin, MicroRNA expression during chick embryo development, Dev. Dynam. 235 (11) (2006) 3156–3165, https://doi.org/10.1002/dvdy.20956.
[83] J.A. Hicks, P. Tembhurne, H.C. Liu, MicroRNA expression in chicken embryos, Poultry Sci. 87 (11) (2008) 2335–2343, https://doi.org/10.3382/ps.2008-00114.
[84] W.P. Kloosterman, E. Wienholds, R.F. Ketting, R.H. Plasterk, Substrate requirements for let-7 function in the developing zebrafish embryo, Nucleic Acids Res. 32 (21) (2004) 6284–6291, https://doi.org/10.1093/nar/gkh968.
[85] J.R. Lytle, T.A. Yario, J.A. Steitz, Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR, Proc. Natl. Acad. Sci. Unit. States Am. 104 (23) (2007) 9667–9672.
[86] M. Schnall-Levin, O.S. Rissland, W.K. Johnston, N. Perrimon, D.P. Bartel, B. Berger, Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs, Genome Res. 21 (9) (2011) 1395–1403.
[87] V. Ambros, The functions of animal microRNAs, Nature 431 (7006) (2004) 350–355, https://doi.org/10.1038/nature02871.
[88] J. Brennecke, A. Stark, R.B. Russell, S.M. Cohen, Principles of microRNA–target recognition, PLoS Biol. 3 (3) (2005) e85.
[89] A. Krek, D. Grun, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, N. Rajewsky, Combinatorial microRNA target predictions, Nat. Genet. 37 (5) (2005) 495–500, https://doi.org/10.1038/ng1536.
[90] E.C. Lai, Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation, Nat. Genet. 30 (4) (2002) 363.
[91] B.P. Lewis, I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, C.B. Burge, Prediction of mammalian microRNA targets, Cell 115 (7) (2003) 787–798, https://doi.org/10.1016/s0092-8674(03)01018-3.
[92] H. Siomi, M.C. Siomi, Posttranscriptional regulation of microRNA biogenesis in animals, Mol. Cell 38 (3) (2010) 323–332, https://doi.org/10.1016/j.molcel.2010.03.013.
[93] D. Baek, J. Villén, C. Shin, F.D. Camargo, S.P. Gygi, D.P. Bartel, The impact of microRNAs on protein output, Nature 455 (7209) (2008) 64.
[94] M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, N. Rajewsky, Widespread changes in protein synthesis induced by microRNAs, Nature 455 (7209) (2008) 58–63, https://doi.org/10.1038/nature07228.
[95] J.A. Gim, H.S. Ha, K. Ahn, D.S. Kim, H.S. Kim, Genome-wide identification and classification of MicroRNAs derived from repetitive elements, Genom. Inf. 12 (4) (2014) 261–267, https://doi.org/10.5808/GI.2014.12.4.261.
[96] R.N. Platt 2nd, M.W. Vandewege, C. Kern, C.J. Schmidt, F.G. Hoffmann, D.A. Ray, Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats, Mol. Biol. Evol. 31 (6) (2014) 1536–1545, https://doi.org/10.1093/molbev/msu112.
[97] S. Qin, P. Jin, X. Zhou, L. Chen, F. Ma, The role of transposable elements in the origin and evolution of MicroRNAs in human, PloS One 10 (6) (2015) e0131365, , https://doi.org/10.1371/journal.pone.0131365.
[98] K.R. Oliver, W.K. Greene, Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates, Mobile DNA 2 (1) (2011) 8, https://doi.org/10.1186/1759-8753-2-8.
[99] K.R. Oliver, W.K. Greene, Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis, Ecol. Evol. 2 (11) (2012) 2912–2933, https://doi.org/10.1002/ece3.400.
[100] D.W. Zeh, J.A. Zeh, Y. Ishida, Transposable elements and an epigenetic basis for punctuated equilibria, Bioessays 31 (7) (2009) 715–726, https://doi.org/10.1002/bies.200900026.
[101] A. Belyayev, Bursts of transposable elements as an evolutionary driving force, J. Evol. Biol. 27 (12) (2014) 2573–2584.
[102] A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J.-N. Volff, Transposable elements as drivers of genomic and biological diversity in vertebrates, Chromosome Res. 16 (1) (2008) 203–215, https://doi.org/10.1007/s10577-007-1202-6.
[103] G. Bonchev, C. Parisod, Transposable elements and microevolutionary changes in natural populations, Mol. Ecol. Resour. 13 (5) (2013) 765–775, https://doi.org/10.1111/1755-0998.12133.
[104] S.J. Klein, R.J. O'Neill, Transposable elements: genome innovation, chromosome diversity, and centromere conflict, Chromosome Res. 26 (1–2) (2018) 5–23, https://doi.org/10.1007/s10577-017-9569-5.
[105] K.R. Oliver, J.A. McComb, W.K. Greene, Transposable elements: powerful contributors to angiosperm evolution and diversity, Genome Biol. Evol. 5 (10) (2013) 1886–1901, https://doi.org/10.1093/gbe/evt141.
[106] L. Schrader, J. Schmitz, The impact of transposable elements in adaptive evolution, Mol. Ecol. 28 (6) (2019) 1537–1549, https://doi.org/10.1111/mec.14794.
[107] J.R. Oaks, A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles, Evolution: Int. J. Org. Evol. 65 (11) (2011) 3285–3297.
[108] C. Addo-Quaye, T.W. Eshoo, D.P. Bartel, M.J. Axtell, Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome, Curr. Biol. 18 (10) (2008) 758–762, https://doi.org/10.1016/j.cub.2008.04.042.
[109] M.A. Frohman, M.K. Dush, G.R. Martin, Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer, Proc. Natl. Acad. Sci. U. S. A. 85 (23) (1988) 8998–9002, https://doi.org/10.1073/pnas.85.23.8998.