A miRNA catalogue and ncRNA annotation of the short-living fish *Nothobranchius furzeri*

Mario Baumgart [1,*] Emanuel Barth [1,2,*], Aurora Savino [4], Marco Groth [1], Philipp Koch[1], Andreas Petzold [5], Ivan Arisi [6], Matthias Platzer [1], Manja Marz [1,2,†] and Alessandro Cellerino [1,3 †]

[1]Leibniz Institute for Age Research - Fritz Lipmann Institute (FLI), Beutenbergstraße 11, 07745 Jena, Germany
[2]Bioinformatics/High Throughput Analysis, Friedrich Schiller University Jena, Leutragraben 1, 07743 Jena, Germany
[3]Laboratory of Biology, Scuola Normale Superiore, 56126, Pisa, Italy
[4]Babraham Institute, Cambridge, England
[5]Dresden University of Technology, Dresden, Germany
[6]European Brain Research Institute (EBRI), Rome, Italy

Associate Editor: XXXXXXX

ABSTRACT

Background: The short-lived fish *Nothobranchius furzeri* is the shortest-lived vertebrate that can be cultured in captivity and was recently established as a model organism for aging research. Small non-coding RNAs, especially miRNAs, are implicated in age-dependent control of gene expression.

Results: Here, we present a comprehensive catalogue of miRNAs and several other non-coding RNA classes (ncRNAs) for *Nothobranchius furzeri*. Analyzing multiple small RNA-Seq libraries, we show most of these identified miRNAs are expressed in at least one of seven *Nothobranchius* species. Additionally, duplication and clustering of *N. furzeri* miRNAs was analyzed and compared to the four fish species *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus* and *Takifugu rubripes*. A peculiar characteristic of *N. furzeri* as compared to other teleosts was a duplication of the miR-29 cluster.

Conclusion: The completeness of the catalogue we provide is comparable to that of zebrafish. This catalogue represents a basis to investigate the role of miRNAs in aging and development in this species.

Availability: All supplementary material can be found online at http://www.rna.uni-jena.de/en/supplements/nothobranchius-furzeri-mirnome/.

Keywords: miRNome, fish miRNA evolution, Nothobranchius furzeri, ncRNA

Contact: manja@uni-jena.de, alessandro.cellerino@sns.it

1 BACKGROUND

The annual teleost *Nothobranchius furzeri* is a recent experimental animal model in biomedical research. In the wild, this fish inhabits

---

*shared first authorship

†shared last authorship; to whom correspondence should be addressed

ephemeral pools in semi-arid *bushveld* of Southern Mozambique. It has adapted to the seasonal drying of its natural environment by producing desiccation-resistant eggs, which can remain dormant in the dry mud for one and maybe more years by entering into diapause. Due to very short duration of the rainy season, the natural lifespan of these animals is limited to a few months. They represent the vertebrate species with the shortest captive lifespan of only 4 - 12 months and also the fastest maturation. In addition, they express a series of conserved aging markers and are amenable to genetic manipulations, making them an attractive model system for aging research (for a review see [8]).

Here, we provide a comprehensive microRNA (miRNA) catalogue for *N. furzeri*. MiRNAs are abundant non-coding RNAs between 18 and 24 nucleotides in length that are produced in a complex biosynthetic process starting from longer transcripts and are established as key players in the post-transcriptional regulation of gene expression. MiRNA genes can be hosted within an intron of a protein-coding gene (and their transcriptional regulation follows that of the hosting gene) or can arise from primary transcripts that are regulated independently of any protein-coding RNA. Several miRNAs are grouped in genomic clusters containing mostly two to six individual miRNAs with an intra-miRNA distance of $<10\,\text{kb}$, which are co-transcribed, however unusually large clusters were also found in some species, like the miR-430 cluster in zebrafish consisting of 57 miRNAs [30, 46, 51]. The advantage of this accumulation is unclear, it could be possible that multiple loci are required to increase the copy-number and therefore the expression level of specific miRNAs in particular conditions, like miR-430 in the maternal-zygotic transition in zebrafish (*Danio rerio*) [12]. MiRNA genes are grouped into families based on sequence homology and can be defined as a collection of miRNAs that are derived from a common ancestor [13]. On the contrary, miRNA clusters may contain miRNAs belonging to different miRNA families, but are located in relative close proximity to each other.

Both the evolutionary conservation of some miRNA families and the innovations leading to appearance of novel miRNAs are well-described. An expansion of miRNA inventory due to genome duplications in early vertebrates and in ancestral teleosts has been described [16].

MiRNAs bind target mRNAs, due to sequence complementarity in the seed region (nucleotides 2 - 7), mostly in the 3' untranslated region, thereby silencing expression of the gene product via translational repression and/or transcript degradation. Up to now, several thousands of miRNAs have been predicted and identified in animals, plants and viruses and one single species can express more than one thousand miRNAs [14]. They frequently represent the central nodes of regulatory networks and may act as "rheostat" to provide stability and fine-tuning to gene expression networks [34, 39].

Before a sequence of the *N. furzeri* genome assembly became available [36], we could show by use of the *Danio rerio* reference from miRBase, that aging in *N. furzeri* brain displays evolutionary conserved miRNA regulation converging in a regulatory network centered on the antagonistic actions of oncogenic MYC and tumor-suppressor TP53 [2] and the expression of miR-15a and the miR-17/92 cluster is mainly localized in neurogenetic regions of the adult brain [7]. Two draft genome sequences for *N. furzeri* were recently produced [36, 50], in this paper, we now provide a comprehensive annotation of the *N. furzeri* miRNome based on a combination of Illumina based small RNA-Seq, different *in silico* prediction methods on the genome assembly and a final manual curation. Using the newly created miRNA reference, we analysed a large dataset of 162 small RNA-Seq libraries and report tissue-specific miRNA expression of conserved and non-conserved miRNAs in *N. furzeri*. We further used the *N. furzeri* reference to analyse the miRNA expression in other Nothobranchius species and one closely-related non-annual killifish species, which were previously used to analyse positive selection [36] to identify when in the evolutionary history of *N. furzeri* non-conserved miRNAs arose.

RESULTS AND DISCUSSION

Small RNA-Seq libraries

For this study 150 small RNA-Seq libraries of *N. furzeri* from different ages and tissues were sequenced, making a total of around 2.2 billion reads. In more detail, we had 75 libraries for both of the strains *N. furzeri MZM-0410* (in the following called MZM) and *N. furzeri GRZ*. We investigated the three tissues brain, liver and skin at five different ages in *N. furzeri GRZ* (5, 7, 10, 12, 14 weeks) and *N. furzeri MZM* (5, 12, 20, 27, 39 weeks) with five biological replicates each. The only exception are the *N. furzeri MZM* brain libraries where we have four biological replicates for each age, but an additional time point of 32 weeks with five replicates. Additionally, seven embryonic small RNA-Seq libraries of *N. furzeri* were sequenced (two of the strain *MZM-0403* and five of *MZM-0410*). After preprocessing all libraries a total of around 1.9 billion high-quality reads was used for further analysis (see Methods section and Supplement Tab. 1).

For each of the six killifish species *A. striatum*, *N. kadleci*, *N. rachovii*, *N. pienaari*, *N. kunthae* and *N. korthausae* we generated two biological replicates of small RNA-Seq libraries from the brain of mature animals. The average size per library was 24.5 million reads with a minimum of 16.8 million and a maximum of 36.1 million reads (for more details about the small RNA-Seq libraries see Supplement Tab. 1).

Annotation of ncRNAs

We could identify more than 750 non-coding RNA (ncRNA) genes in the *N. furzeri* genome based on small RNA-Seq reads (see Tab. 1, Supplement Material 1 and Supplement Tab. 5). In line with other eukaryotes, we identified multiple copies of rRNA operons, tRNAs, several major spliceosomal RNAs, signal recognition particle (SRP) RNAs and one copy of a minor spliceosomal RNA set. Further housekeeping RNA genes, such as RNase P, RNase MRP, and the 7SK RNA are found as expected once in the entire genome. We annotated the widely distributed TPP riboswitch, capable of binding thiamine pyrophosphate and thereby regulating genes that are in charge of the thiamine balance. We could also identify more RNA elements directly involved in the regulation of gene expression (3 copies of *IRE1* – controlling iron responsive proteins, CAESAR – controlling tissue growth factor CTGF, DPB – controlling DNA polymerase β), localization of mRNAs (Vimentin3), DNA replication (four copies of the *Y RNA* gene, and Telomerase RNA TERC), or of unknown function (12 vault RNAs). Additionally, ncRNAs responsible for editing certain mRNAs have also been found (two copies of Antizyme FSE, one U1A polyadenylation inhibition element (PIE), 26 Potassium channel RNA editing signals (KRES), and six copies of *GABA3*). Two promising candidate long non-coding RNAs (lncRNAs), SPYR-IT1 and MEG8, were also included in the annotation, even though we were not able to identify all of their exons. Two vague candidates for XIST and MALAT can be viewed in the supplemental material. The MALAT-derived masc and men RNA gene was clearly detected in 42 copies throughout the genome of *N. furzeri*.

Table 1. Number of annotated ncRNAs.

ncRNA class	No.	ncRNA class	No.
18S rRNA	3	7SK	1
5.8S rRNA	11	Vault	12
28S rRNA	6	TPP	1
5S rRNA	28	IRE1	3
tRNA	570	CAESAR	1
U1	6	DPB	1
U2	8	Vimentin3	1
U4	8	Antizyme FSE	2
U5	8	U1A PIE	1
U6	6	KRES	26
U11	1	GABA3	6
U12	1	Y RNA	4
U4atac	1	TERC	1
U6atac	1	mascRNA-menRNA	42
RNase P	1	SPYR-IT1*	1
RNase MRP	1	MEG8*	1
SRP	3		

Besides housekeeping RNAs RNA elements controlling metabolism and protein editing were identified. * lncRNA candidate.
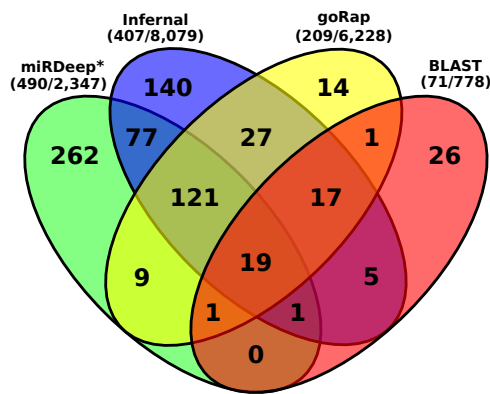
Fig. 1. Venn diagram of predicted miRNA genes from four tools `miRDeep*`, `Infernal`, `goRap` and `BLAST`. Only 2 of the 33 candidates predicted by `CID-miRNA` overlapped with any of the other miRNA candidates. Nevertheless all 33 candidates were selected as miRNAs after manual inspections. The total number of miRNA predictions after and before applying any filtering step are shown in brackets for each tool.

Mapping and miRNA prediction results

For the identification of putative miRNA genes we used five methods, each following a different prediction approach (`BLAST`, `CID-miRNA`, `Infernal`, `GoRAP`, `miRDeep*`) and `Blockbuster` as verification (see Fig. 2 for an example).

The five tools identified 71, 33, 407, 209, 490 miRNA candidates, respectively (Fig. 1 shows the variety and the overlap of the different tools). All predictions were merged and redundant loci were removed (for details see methods section). Of the remaining 788 candidate miRNAs 617 (78.3 %) showed expressions and were verified by `blockbuster` and manually investigated for miRNA-like profiles. By this 34 (4.3 %) candidates were removed, all predicted by `miRDeep*`. Candidates showing no expression in any of the sequenced small RNA-Seq libraries were still kept as putative miRNAs. In total we predict a final amount of 754 miRNAs in *N. furzeri* by union of these methods (see Supplementary Material 2).

Most of the small RNA-Seq reads (up to 88.81%) mapped onto the identified 754 miRNAs. Interestingly, the number of miRNA related reads varies broadly between the tissue samples (see Tab. 2). Possibly, this difference correlates with different regenerative capacities of these tissues. This suggests the existence of a class of ncRNAs, yet to be identified and annotated in *N. furzeri*, in line with a relatively high amount of reads mapping to unannotated locations in the liver samples. About half of the miRNA annotations are overlapping genes coding for proteins and are therefore intragenic. A minor fraction of reads (see Tab. 2) maps to other ncRNAs and proteins. Whereas 333 of the predicted miRNAs can be assigend to one of the known miRBase families, based on sequence identity, 421 miRNAs did not match any known family and can therefore be considered as novel or non-conserved miRNAs (for details see Table 3).

Effects of tissue and age on global miRNA expression

We used principal component analysis (PCA) to visualize the effects of tissue type and age on global miRNA expression (see Fig. 3).

Table 2. Total number of genes known in *N. furzeri* and number of ncRNAs covered by small RNA-Seq reads and percental distribution of reads for brain, skin and liver of *N. furzeri*. See Supplement Tab. 1 for more details and Tab. 4 for details about the read libraries.

	Total	Brain		Skin		Liver	
miRNA	754	470	88.81 %	438	75.01 %	368	58.10 %
tRNA	570	293	0.35 %	361	1.33 %	361	4.27 %
rRNA	48	37	1.73 %	43	3.61 %	43	6.97 %
ncRNA	250	136	0.09 %	217	0.15 %	204	0.41 %
protein*	22,627	0	6.60 %	0	4.59 %	0	6.35 %
none			2.42 %		15.31 %		23.90 %
Sum	24,125	936	100 %	1059	100 %	976	100 %

Asteriks indicates annotation of protein-coding genes from NCBI. ncRNA – All ncRNAs except miRNA, tRNA and rRNA; None – Reads mapping to non-annotated genomic areas.

A strong component of tissue-specific expression was detected and samples corresponding to different tissues clustered tightly and widely apart in the plane defined by the first two principal component axes (collectively accounting for 77 % of variance). Remarkably, the third principal component axis (3 % of variance explained) identifies an age-dependent component of miRNA expression, that is common to all three tissues with the youngest samples (5 weeks) clearly separated from the rest. A detailed analysis of age- and tissue-dependent miRNA expression including embryonic development will be part of a separate publication.
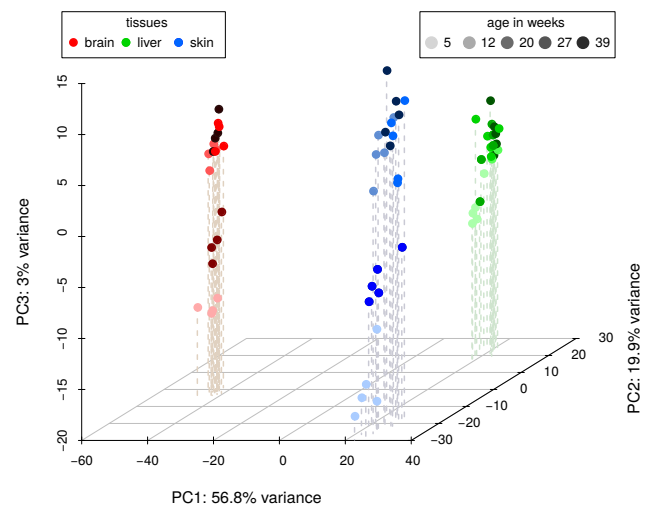

Fig. 3. A three-dimensional PCA plot of the *N. furzeri MZM* small RNA-Seq libraries of all three tissues (brain – red, liver – green, blue – skin) and all investigated ages (from light to dark: 5, 12, 20, 27, 39 weeks). Whereas the samples cluster well according to their tissue belongings, a distinct separation regarding the ages can only be observed for the youngest samples in each tissue.

MiRNA expression in closely related killifish

To compare and validate miRNA composition in *N. furzeri* we created for each of the six related killifish species two small RNA-Seq libraries (see Tab. 4). These libraries were mapped simultaneously on all available miRBase (release 21) sequences and our annotated miRNAs of *N. furzeri*, to observe *N. furzeri* miRNA-candidates expressed in other killifish and conserved miRNAs possibly missing in *N. furzeri*, but not in the closely related species. In total, 546 (93.7 %) of the 583 expressed and 17 (9.9 %) of the 171 non-expressed miRNA candidates in *N. furzeri* showed expression in at least one of the related killifish (Fig. 5 shows a miRNA not expressed in *N. furzeri* but in several of the other killifish). Of these expressed miRNAs 299 belong to the 421 non-conserved *N. furzeri* miRNA genes. To investigate whether miRNA sequences reflect known phylogenetic relationships, we concatenated the sequences of all expressed miRNAs and constructed a phylogenetic tree. This tree perfectly reflected the evolution of the *Nothobranchius* linage [10]. It is also interesting that the number of *N. furzeri* miRNAs expressed in other killifish species (indicated above the branch in Figure 4) is inversely correlated to the evolutionary distance, i.e. this number is higher for killifish the closer they are related to *N. furzeri*.

A. striatum, *N. korthausae*, *N. pienaari*, *N. rachovii*, *N. kunthae* and *N. kadleci* showed expression for 352, 428, 488, 473, 496 and 534 miRNAs, respectively. Most of these expressed miRNAs (> 89 %) are among the 333 conserved miRNAs of *N. furzeri* (see

Supplement Tab. 3). The composition of expressed miRNAs from the six killifish varies only marginally. The *Nothobranchius* species (except *N. furzeri*) had in total 395 expressed miRNAs in common (of which 148 are non-conserved), and *A. striatum* expressed 324 of them (of which 116 are non-conserved). These 324 miRNAs represent a core of miRNAs from *Nothobranchiidae* whose origin predates the emergence of annualism in this clade.

MiRNA clusters and gene duplications

MiRNAs are known to often occur in clusters [16]. We define a miRNA cluster to consist of at least two miRNAs with a maximum distance of 10 kb. Examining the localization and distances of the miRNA genes in the five fish species with assembled genomes, we identified 83, 96, 58, 68, and 59 different clusters in *N. furzeri*, *D. rerio*, *O. latipes*, *G. aculeatus* and *T. rubripes*, respectively (see Tab. 3, Fig. 6-A).

In all investigated fish species but *T. rupripes* the largest cluster is the *mir-430* cluster (containing 7 to 55 miRNAs; see Fig. 6-C). This cluster is extremely divergent and evolving relatively quickly in each lineage. Not only the number of miR-430 copies within each cluster varies greatly, but also the number and organization of the members of this miRNA family. Whereas miR-430a and miR-430c can be found in all five fish species, miR-430b and miR-430d seem to occur only in *D. rerio* and *O. latipes*, respectively.
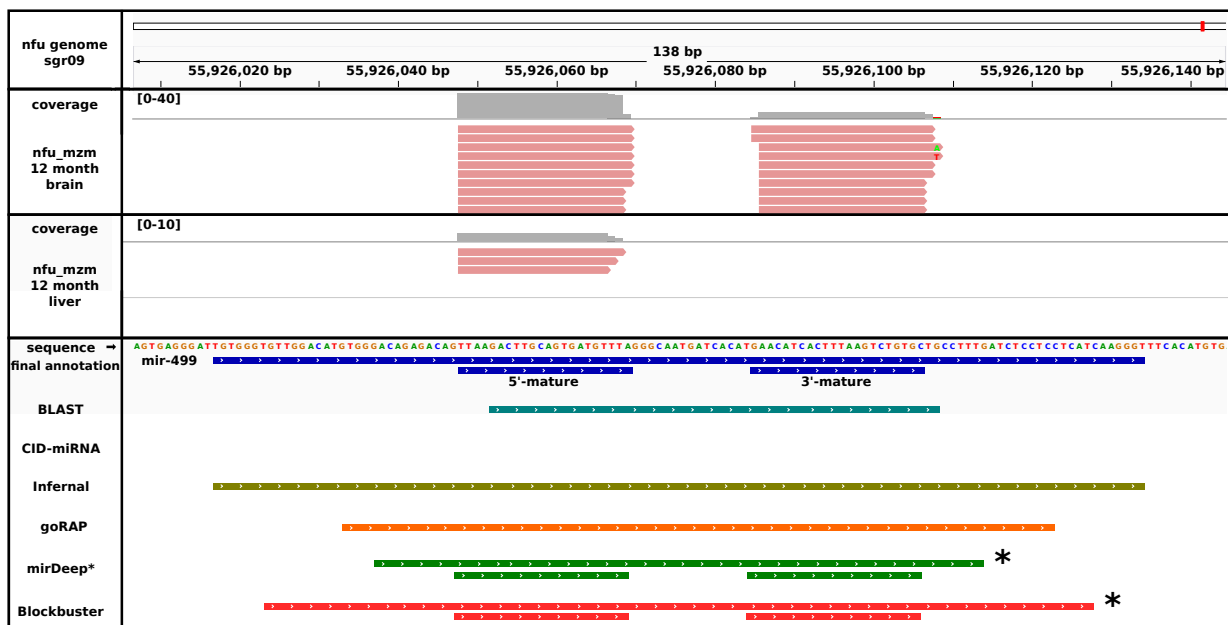

Fig. 2. Annotation, expression profiles and prediction comparison for miR-499. We annotated the pre-miR-499 on sgr09, position 55,926,017–55,926,134 and the two mature miRNAs at 55,926,048–55,926,069 and 55,926,085–55,926,106. The six methods used for miRNA detection are displayed, CID-miRNA was not able to detect this miRNA. Tools working independent of the small RNA-Seq data (BLAST (cyan), Infernal (olive green) and goRAP (orange)) vary in their annotation length. The latter two programs are based on covariance models identifying mostly the complete pre-miRNA. The remaining two programs miRDeep* and Blockbuster are based on small RNA-Seq data (*) and therefore accurately annotate the mature miRNAs. MiR-499 is expressed weakly within *N. furzeri MZM* 12 month liver library and therefore could not be detected by miRDeep* and blockbuster. In the *N. furzeri MZM* 12 month brain library miR-499 was expressed high enough to be detected by both programs.
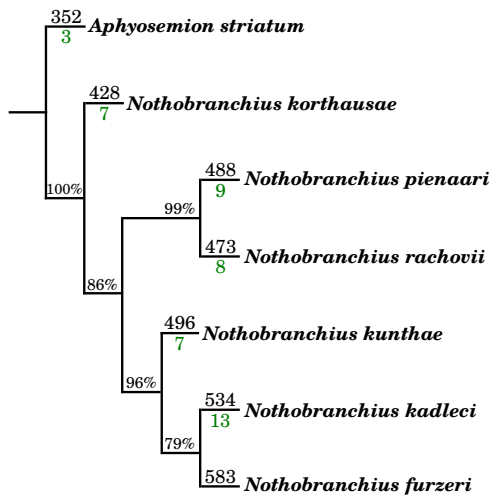
Fig. 4. Killifish phylogeny based on the expressed miRNAs calculated via hierarchical clustering using the `R` package `pvclust` [41]. Bootstrap values are given as percentages at the corresponding branches. The amount of identified expressed miRNAs is given next to the species names. The numbers in green indicate the number of these expressed miRNAs, which were annotated but not expressed in any of the sequenced *N. furzeri* samples.
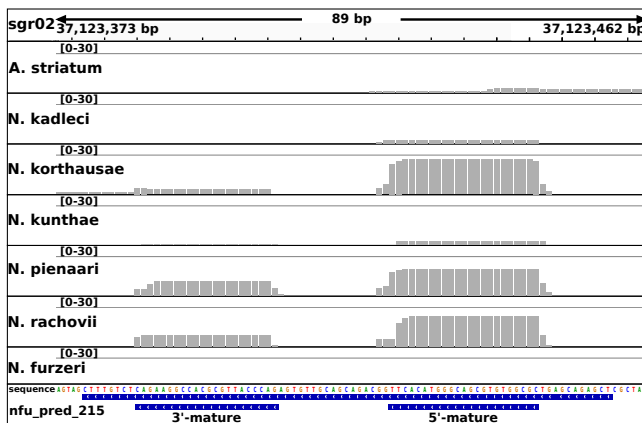

Fig. 5. Expression profiles of the predicted miR-215. Grey bars indicate the amount of aligned reads and therefore coverage at the specific postions. Whereas no expression can be observed for this miRNA in *N. furzeri*, clear activation can be seen in *N. korthausae*, *N. pienaari* and *N. rachovii*. *A. striatum*, *N. kadleci* and *N. kunthae* show a weak expression for at least the 5'-mature variant of this miRNA.

Additionally, no structural similarities or shared repetition patterns can be observed for this miRNA cluster, which is an additional indication of the low purifying selection on this specific gene cluster. However, a clear duplication pattern can be observed for the miR-430 cluster in *D. rerio* (the order miR-430c/b/a is repeated with only few excpetions) and *N. furzeri* (the order miR-430c/a/a/c/a/a/a is repeated). For *O. latipes* and *G. aculeatus* the order of miR-430 variants appears to be more random and *T. rubripes* has too few copies to show any repeated pattern.

Fig. 6-B depicting the miR-17-92 cluster shows an example of the

Table 3. Amount of annotated miRNAs, identified miRNA clusters and the number of miRNAs in clusters, as well as known conserved and non-conserved miRNA families in *N. furzeri* (Nfu), *D. rerio* (Dre), *O. latipes* (Ola), *G. aculeatus* (Gac) and *T. rubripes* (Fru).

Species	#miRNAs	#miRNA clusters	#miRNAs in clusters	#mirBase families	#unknown families
nfu	754	83	213	94 (333)	383 (421)
dre	765	96	305	99 (307)	302 (458)
ola	366	58	151	104 (269)	55 (97)
gac	504	68	299	102 (413)	63 (91)
fru	337	59	143	99 (278)	51 (59)

In brackets the amount of miRNAs associated to the identified miRNA families are given. For detailed lists of miRNA family assignments see Supplement Tab. 4.

other extreme. In all five investigated fish species, two perfectly conserved clusters can be found. These represent a duplication of an ancestral cluster present in all vertebrates and the order of the different members is perfectly conserved. It is known that the miR-17-92 cluster is transcribed polycystronically and acts in oncogenic and tumor suppressor pathways [15, 43]. Furthermore, up to two smaller and lesser conserved clusters, containing at least two miRNAs of the miR-17 or miR-92 family were identified per fish species, similar to what is known for mammals. Having correctly identified this highly conserved cluster in *N. furzeri* is again good evidence for the high quality of its newly assembled genome and completeness of our miRNA catalogue.

Another example for an evolutionary conserved miRNA cluster is the miR-29 cluster depicted in Fig. 6-D. Mir-29 family members are up-regulated during aging in a variety of different tissues including muscle, skin, brain and aorta [2, 11, 33, 40, 42, 49] and appear to be key regulators of age dependent gene expression [4, 37]. This cluster consists of miR-29a (which is identical to the mammalian miR-29c) and its variant miR-29b and is duplicated at leasts once. In some fish species an additional variant miR-29c is known, which is identical to the miR-29a in mammals with one nucleotide being different outside the seed region [29]. As from RFAM (version 12.1) and miRBase (release 21) miR-29 genes are mainly identified in vertebrates as well as one *Hemichordata* and one *Arthropoda*, so we can only speculate that the original cluster duplication event arose in the early metaza lineage. In *O. latipes* and *T. rubripes* both miR-29 clusters are still present, whereas *D. rerio* seems to has lost one copy of the miR-29a gene. For *G. aculeatus* we were only able to identify one miR-29 cluster. However, since its genome assembly is incomplete we assume that the second cluster may not be lost but is missing in the current version of its miRNA annotation. Interestingly, in *N. furzeri* we identified an additional miR-29a/b pair and a fourth single copy of miR-29b. Assuming a complete genome assembly, different scenarios could explain this finding: (1) both original miR-29 clusters were individually duplicated once more and the fourth miR-29a gene was later lost, (2) one of the two clusters was duplicated as a whole, whereas in the other only miR-29b was copied, or (3) both original clusters were duplicated during the same event and again one of the miR-29a genes was later lost.

About the same amount of different miRBase miRNA families could be identified for all five fish species, despite their big differences in the number of identified miRNA genes. All miRNA
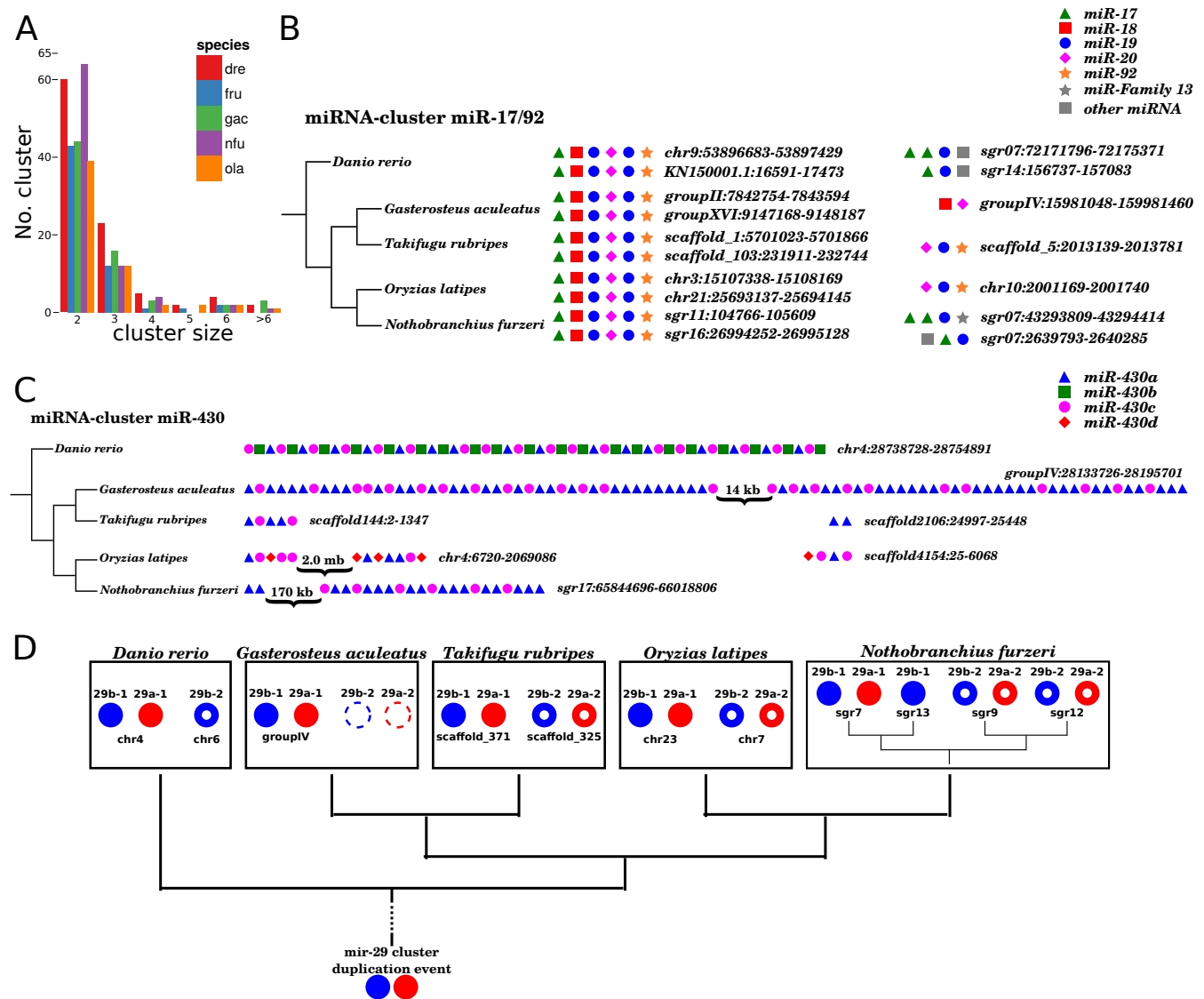
Fig. 6. MiRNA cluster comparison between fish. (A) Amount of clusters and their respective sizes with a maximum distance of 10,000 bp between two miRNAs. (nfu - *Nothobranchius furzeri*, dre - *Danio rerio*, ola - *Oryzias latipes*, gac - *Gasterosteus aculeatus*, tru - *Takifugu rubripes*) (B) Structure comparison of the miR-17/92 cluster. Two highly conserved clusters could be identified for each species, as well as some smaller less conserved clusters, containing at least two miRNAs of the miR-17/92 cluster. (C) Structure comparison of the miR-430 cluster. No structural similarity between the different species can be observed. However, *D. rerio*, *D. aculeatus* and *N. furzeri* show some distinct but individual repeating pattern. Even though the gene variants miR-430b and miR-430d seem to be unique to *D. rerio* and *O. latipes*, they can be clearly distinguished, based on sequence alignments. (D) After the ancestral duplication event the mir-29 cluster is distinguished in the mir-29a/b-1 (filled red and blue dots) and the mir-29a/b-2 cluster (red and blue circles). Whereas for *D. rerio* the mir-29a-2 gene seems to be lost, we assume that for *G. aculeatus* the whole second mir-29 cluster (dashed circles) is only missing, because of the low quality genome sequencing and assembly. In *N. furzeri* we observe an additional copy for each of the two clusters, except that the mir-29a/b-1 pair was only partially duplicated or the second mir-29a-1 gene was lost again.

genes not matching any known mirRBase family were clustered based on their sequence identity in order to estimate the amount of miRNA 'families' not covered by the miRBase database (see Tab. 3 and Supplement Tab. 4).

CONCLUSION

This study involves a multitude of small RNA-Seq libraries from several tissues, ages, strains and embryos of *N. furzeri* and closely related species. The aim was the characterization of the *N. furzeri* miRNome and a detailed annotation in the recently published genome [36]. The inclusion of other killifish species allowed us to analyze the occurrence of novel miRNAs in the group of annual fish.

Due to the fact, that we identified roughly the same number of miRNAs in *N. furzeri* as known in *D. rerio* and both fish species share almost equal amounts of miRBase families and unknown miRNA families, we assume that our miRNA catalogue is as complete as that of the model organism *D. rerio*.

METHODS

RNA extraction

Animal maintenance was performed as described [44, 45]. To avoid effects of circadian rhythms and feeding, animals were always sacrificed at 10 a.m. in fasted state. For tissue preparation, fish were euthanized with MS-222 and cooled on crushed ice. The protocols of animal maintenance and experiments were approved by the local authority in the State of Thuringia (Veterinaer- und Lebensmittelueberwachungsamt). Total RNA was extracted as described [2]. RNA quality and amount was determined using the Agilent Bioanalyzer 2100 and the RNA 6000 Nano Kit (Agilent Technologies).

small RNA library preparation and sequencing

Library preparation and sequencing was done using *Illuminas* NGS platform [3]. One μg of total RNA was used for library preparation using *Illuminas* TruSeq small RNA sample preparation kit following the manufacturers instruction. The libraries were quantified on the Agilent DNA 1000 chip and subjected to sequencing-by-synthesis on an *Illumina* HiSeq2500 in high-output, 50 bp single-read mode. Sequencing chemistry v3 was used. Read data were extracted in FastQ format using the Illumina supported tool `bcl2fastq` (v1.8.3 and v1.8.4). The only exceptions were three of the *N. furzeri* embryo samples, which were sequenced on an *Illumina* HiSeq2000 in 50 bp single-read mode and where read data was extracted in FastQ format using the tool `CASAVA` (v1.8.2). Sequencing resulted in around $4-50$ million reads per sample with pooling eight samples per lane.

In total 169 small RNA-Seq libraries from seven different killifish species were created. 157 of them were obtained from *N. furzeri* strains *GRZ* and *MZM-0410* at several ages from the three tissues brain, liver and skin. The remaining RNA-Seq libraries obtained from *Aphyosemion striatum*, *N. kadleci*, *N. rachovii*, *N. pienaari*, *N. kunthae* and *N. korthausae* were used to identify expression patterns at predicted miRNA locations in *N. furzeri* and miRBase pre-mature miRNA sequences. For details see Tab. 4, Supplement Tab. 1 and Supplement Tab. 2.

Validation of mature miRNA expression by quantitative real-time PCR

For validation of mature miRNAs expression quantitative real-time PCR was performed an published in [2]. The used primers were Qiagen miScript primer: tni-miR-15a (cat# MSC0001523), tni-miR-101a (MSC0001279), tni-miR-101b (MSC0001283), dre-miR-145 (MSC0001529), has-miR-29c-1 (which is 100 % identical to dre-miR-29a, MS00003269), hsa-let-7a-5p (MS00031220), hsa-miR-124a-1 (MS00006622), has-miR-1-2 (MS00008358), ola-miR-21 (MS00049546), ola-miR-183-5p (MS00049392), and from cluster dre-miR-17a/18a/19a (MSC0002184–MSC0002186), and dre-miR-20a (MSC0002183).

Table 4. SmallRNA-Seq samples from *Nothobranchius* strains generated in this study. * – unknown; # – number of replicates; + – two weeks post-fertilization plus diapause

Species	Tissue	Age (weeks)	No. library	#
N. furzeri MZM	whole embryos	diapause II$^+$	7	7
N. furzeri MZM	brain, liver, skin	5,12,20, 27,32,39	75	4-5
N. furzeri GRZ	brain, liver, skin	5,7,10, 12,14	75	5
A. striatum	brain	*	2	2
N. kadleci	brain	*	2	2
N. rachovii	brain	*	2	2
N. pienaari	brain	*	2	2
N. kunthae	brain	*	2	2
N. korthausae	brain	*	2	2

Small RNA-Seq library processing and mapping

In-house scripts were used to cut the RA3 adapter of the TruSeq small RNA preparation kit (5'-TGGAATTCTCGGGTGCCAAGG) from the reads. Additionally, `PRINSEQ` (v0.20.3) [38] was used to trim the reads from both sides, in order that the read bases had an minimum quality of 20 and reads were at least 15 bases long. Mapping onto the *N. furzeri* genome was performed with `segemehl` (v0.2.0) [18] using the `-H 1` option, allowing single reads to be mapped to multiple best fitting locations. Visualization of mapped reads was done using `IGV` (v2.0.34) [47]. Since `Bowtie` (v1.0.0) [28] is the built-in method in `miRDeep*` for mapping it was also used for the genomes of *N. furzeri*, *D. rerio*, *O. latipes* and *T. rubripes*.

Genomes and annotations

The recently published high-quality draft genome assembly and annotation of *N. furzeri* and the small RNA-Seq libraries described above were used for mapping as well as for miRNA and other ncRNA predictions and annotations [36]. Additionally, these RNA-Seq libraries were also mapped on the following fish genomes: *Danio rerio* (GRCz10) [19], *Oryzias latipes* (HdrR) [22] and *Takifugu rubripes* (FUGU5) [20]. For annotation comparison, the latest complete genomic information of those three fish and of *Gasterosteus aculeatus* (BROAD S1) [20] were downloaded from the ensembl database [9].

ncRNA and miRNA annotation

Non-coding RNAs in general were annotated with `GoRAP 2.0`, which is based on the RFAM database currently holding 2,450 ncRNA families (v12.0) [31].

For an initial prediction of candidate miRNAs a combination of five tools was used, each of them following a different annotation strategy: `miRDeep*` (v32) [1], `Infernal` (v1.1) [32], `BLAST` (v2.2.30) [6], `GoRAP` (v2.0, unpublished) and `CID-miRNA` (version from April 2015) [48]. A detailed description of the individual searches can be found below.

All results were merged and putative miRNAs overlapping with

genes of the recently published *N. furzeri* annotation were removed. Expression profiles of the remaining non-redundant candidate miRNA genes were analyzed automatically using `blockbuster` (v1) [26] and in-house scripts in order to mark candidates that did not exhibit a typical miRNA expression profil (according to [27, 21]). All candidates were additionally manually examined and filtered, leading to the final set of miRNA predicitions.

miRDeep* search Mappings of 39 MZM brain, 15 GRZ brain, 25 GRZ liver, 28 MZM liver, 3 MZM skin and 7 MZM embryo small RNA-Seq libraries were used on four different fish genomes (*N. furzeri*, *D. rerio*, *O. latipes*, *T. rubripes*) as input for `miRDeep*` (for a detailed list of used libraries see Supplement Tab. 1). Predictions from all 117 mappings were pooled together, in order to obtain a comprehensive representation of the `miRDeep*` results. To each predicted miRNA hairpin sequence, we assigned the average of the `miRDeep*` score computed across the multiple samples were the sequence was found.

The merged non-redundant list of identified miRNA sequences was re-mapped with `BLAT` [23] on the *N. furzeri* genome, and only gap-free alignments were accepted. These loci underwert further filtering steps: (i) A hairpin sequence was considered reliable if it showed a `BLAT` hit (one mismatch allowed) in miRBase (release 20) [24] or a miRDeep* score $>= 7$ and (ii) overlapping hairpin loci (i.e., within 100 nt) were discarded and the sequence with the highest score was kept. Predictions where no hits in miRBase could be obtained were further analyzed based on their secondary structure. Therefore, corresponding sequences were extended by 50 nt on either side and were compared with Rfam using `Infernal`. All predicted loci that had a significant hit to a known miRNA secondary structure or no hit at all were kept, while loci hitting other ncRNAs were discarded.

Infernal search For the `Infernal` search on the *N. furzeri* genome 155 hand curated pre-miRNA covariance models were used as input [5, 17] and only significant hits with a p-value of $p < 0.005$ were kept.

BLAST search In order to identify candidates from the most conserved miRNA families, `blastn` was used with all mature and pre-mature miRNA sequences available on miRBase (release 21) [25]. Only non-redundant hits were kept if they spanned the complete sequences of their corresponding input miRNAs to at least 90 % with no gaps allowed. To further reduce false positive hits a stringent cutoff of $p < 10^{-7}$ was chosen.

CID-miRNA search Being based on a stochastic context free grammar model to identify possible pre-miRNAs, `CID-miRNA` follows a similar approach as `Infernals` covariance models. The *N. furzeri* genome was given as input with the following thresholds: putative miRNAs have a length between 60 bp and 120 bp and the grammar and structural cutoff were set to the recommended values of -0.609999 and 23, respectively.

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Supplementary material can be found online at http://www.rna.uni-jena.de/en/supplements/nothobranchius-furzeri-mirnome/.
Data were deposited in GEO with the accession number GSE92854. The presented miRNome annotation is accessible and downloadable via the NFINgb Nothobranchius furzeri Genome Browser (http://nfingb.leibniz-fli.de/).

Competing interests

The authors declare no competing interests.

Authors contribution

MB, MP and AC conceived and designed the study; MB performed the experiments; MG performed the RNA-Seq; IA, AP, EB, AS and MM analyzed and interpreted the data; AC, MB, EB and MM wrote the manuscript with contributions from all other authors.

REFERENCES

[1] J. An, J. Lai, M. L. Lehman, and C. C. Nelson. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic acids research*, 41(2):727–737, 2013.

[2] M. Baumgart, M. Groth, S. Priebe, J. Appelt, R. Guthke, M. Platzer, and A. Cellerino. Age-dependent regulation of tumor-related microRNAs in the brain of the annual fish Nothobranchius furzeri. *Mechanisms of ageing and development*, 133(5):226–233, 2012.

[3] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218):53–59, 2008.

[4] R. A. Boon, K. Iekushi, S. Lechner, T. Seeger, A. Fischer, S. Heydt, D. Kaluza, K. Tréguer, G. Carmona, A. Bonauer, et al. Microrna-34a regulates cardiac ageing and function. *Nature*, 495(7439):107–110, 2013.

[5] I. Braasch, A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki, J. Pasquier, A. Amores, T. Desvignes, P. Batzel, J. Catchen, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature genetics*, 48(4):427–437, 2016.

[6] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. BLAST+: architecture and applications. *BMC bioinformatics*, 10(1):421, 2009.

[7] A. Cellerino. Regulation of microRNA expression in the neuronal stem cell niches during aging of the short-lived annual fish Nothobranchius furzeri. *Regulatory RNAs in the Nervous System*, page 112, 2015.

[8] A. Cellerino, D. R. Valenzano, and M. Reichard. From the bush to the bench: the annual Nothobranchius fishes as a new model system in biology. *Biological Reviews*, 91(2):511–533, 2016.

[9] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, et al. Ensembl 2015. *Nucleic acids research*, 43(D1):D662–D669, 2015.

[10] A. Dorn, Z. Musilová, M. Platzer, K. Reichwald, and A. Cellerino. The strange case of east african annual fishes: aridification correlates with diversification for a savannah aquatic group? *BMC evolutionary biology*, 14(1):1, 2014.

[11] A. M. Fenn, K. M. Smith, A. E. Lovett-Racke, M. Guerau-de Arellano, C. C. Whitacre, and J. P. Godbout. Increased micro-rna 29b in the aged brain correlates with the reduction of insulin-like growth factor-1 and fractalkine ligand. *Neurobiology of aging*, 34(12):2748–2758, 2013.

[12] A. J. Giraldez, Y. Mishima, J. Rihel, R. J. Grocock, S. Van Dongen, K. Inoue, A. J. Enright, and A. F. Schier. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *science*, 312(5770):75–79, 2006.

[13] S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy. Rfam: an RNA family database. *Nucleic acids research*, 31(1):439–441, 2003.

[14] S. Griffiths-Jones, R. J. Grocock, S. Van Dongen, A. Bateman, and A. J. Enright. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, 34(suppl 1):D140–D144, 2006.

[15] L. He, J. M. Thomson, M. T. Hemann, E. Hernando-Monge, D. Mu, S. Goodson, S. Powers, C. Cordon-Cardo, S. W. Lowe, G. J. Hannon, et al. A microrna polycistron as a potential human oncogene. *Nature*, 435(7043):828–833, 2005.

[16] J. Hertel, M. Lindemeyer, K. Missal, C. Fried, A. Tanzer, C. Flamm, I. L. Hofacker, and P. F. Stadler. The expansion of the metazoan microRNA repertoire. *BMC Genomics*, 7:25, 2006.

[17] J. Hertel and P. F. Stadler. The expansion of animal microRNA families revisited. *Life*, 5(1):905–920, 2015.

[18] S. Hoffmann, C. Otto, S. Kurtz, C. M. Sharma, P. Khaitovich, J. Vogel, P. F. Stadler, and J. Hackermüller. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol*, 5(9):e1000502, 2009.

[19] K. Howe, M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446):498–503, 2013.

[20] F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392):55–61, 2012.

[21] C.-H. Jung, M. A. Hansen, I. V. Makunin, D. J. Korbie, and J. S. Mattick. Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. *BMC genomics*, 11(1):77, 2010.

[22] M. Kasahara, K. Naruse, S. Sasaki, Y. Nakatani, W. Qu, B. Ahsan, T. Yamada, Y. Nagayasu, K. Doi, Y. Kasai, et al. The medaka draft genome and insights into vertebrate genome evolution. *Nature*, 447(7145):714–719, 2007.

[23] W. J. Kent. BLAT – the BLAST-like alignment tool. *Genome research*, 12(4):656–664, 2002.

[24] A. Kozomara and S. Griffiths-Jones. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research*, page gkq1027, 2010.

[25] A. Kozomara and S. Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, page gkt1181, 2013.

[26] D. Langenberger, C. Bermudez-Santana, J. Hertel, S. Hoffmann, P. Khaitovich, and P. F. Stadler. Evidence for human microrna-offset rnas in small rna sequencing data. *Bioinformatics*, 25(18):2298–2301, 2009.

[27] D. Langenberger, C. Bermudez-Santana, P. F. Stadler, and S. Hoffmann. Identification and classification of small RNAs in transcriptome sequence data. In *Pacific Symposium on Biocomputing*, volume 15, pages 80–87. World Scientific, 2010.

[28] B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biol*, 10(3):R25, 2009.

[29] A. Liston, A. S. Papadopoulou, D. Danso-Abeam, and J. Dooley. Microrna-29 in the adaptive immune system: setting the threshold. *Cellular and Molecular Life Sciences*, 69(21):3533–3541, 2012.

[30] A. Marco, M. Ninova, M. Ronshaugen, and S. Griffiths-Jones. Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic acids research*, 41(16):7745–7752, 2013.

[31] E. P. Nawrocki, S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate, et al. Rfam 12.0: updates to the RNA families database. *Nucleic acids research*, page gku1063, 2014.

[32] E. P. Nawrocki and S. R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22):2933–2935, 2013.

[33] K. Nolan, M. R. Mitchem, E. M. Jimenez-Mateos, D. C. Henshall, C. G. Concannon, and J. H. Prehn. Increased expression of microrna-29a in als mice: functional analysis of its inhibition. *Journal of Molecular Neuroscience*, 53(2):231–241, 2014.

[34] M. Osella, C. Bosia, D. Corá, and M. Caselle. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput Biol*, 7(3):e1001101, 2011.

[35] M. Reichard, A. Cellerino, and D. R. Valenzano. Turquoise killifish. *Current Biology*, 25(17):R741–R742, 2015.

[36] K. Reichwald, A. Petzold, P. Koch, B. R. Downie, N. Hartmann, S. Pietsch, M. Baumgart, D. Chalopin, M. Felder, M. Bens, et al. Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell*, 163(6):1527–1538, 2015.

[37] R. Ripa, L. Dolfi, M. Terrigno, L. Pandolfini, V. Arcucci, M. Groth, E. T. Tozzini, M. Baumgart, and A. Cellerino. Microrna mir-29 controls a compensatory response to limit neuronal iron accumulation during adult life and aging. *bioRxiv*, page 046516, 2016.

[38] R. Schmieder and R. Edwards. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–864, 2011.

[39] V. Siciliano, I. Garzilli, C. Fracassi, S. Criscuolo, S. Ventre, and D. di Bernardo. MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nature communications*, 4, 2013.

[40] M. Somel, S. Guo, N. Fu, Z. Yan, H. Y. Hu, Y. Xu, Y. Yuan, Z. Ning, Y. Hu, C. Menzel, et al. Microrna, mrna, and protein expression link development and aging in human and macaque brain. *Genome research*, 20(9):1207–1218, 2010.

[41] R. Suzuki and H. Shimodaira. Pvclust: an r package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542, 2006.

[42] M. Takahashi, A. Eda, T. Fukushima, and H. Hohjoh. Reduction of type iv collagen by upregulated mir-29 in normal elderly mouse and klotho-deficient, senescence-model mouse. *PLoS one*, 7(11):e48974, 2012.

[43] A. Tanzer and P. F. Stadler. Molecular evolution of a microrna cluster. *Journal of molecular biology*, 339(2):327–335, 2004.

[44] E. Terzibasi, C. Lefrançois, P. Domenici, N. Hartmann, M. Graf, and A. Cellerino. Effects of dietary restriction on mortality and age-related phenotypes in the short-lived fish Nothobranchius furzeri. *Aging cell*, 8(2):88–99, 2009.

[45] E. Terzibasi, D. R. Valenzano, M. Benedetti, P. Roncaglia, A. Cattaneo, L. Domenici, and A. Cellerino. Large differences in aging phenotype between strains of the short-lived annual fish Nothobranchius furzeri. *PLoS One*, 3(12):e3866, 2008.

[46] E. J. Thatcher, I. Bond, I. Paydar, and J. G. Patton. Genomic organization of zebrafish microRNAs. *BMC genomics*, 9(1):1, 2008.

[47] H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, page bbs017, 2012.

[48] S. Tyagi, C. Vaz, V. Gupta, R. Bhatia, S. Maheshwari, A. Srinivasan, and A. Bhattacharya. CID-miRNA: a web server for prediction of novel miRNA precursors in human genome. *Biochemical and biophysical research communications*, 372(4):831–834, 2008.

[49] A. P. Ugalde, A. J. Ramsay, J. de la Rosa, I. Varela, G. Mariño, J. Cadiñanos, J. Lu, J. M. Freije, and C. López-Otín. Aging and chronic dna damage response activate a regulatory pathway involving mir-29 and p53. *The EMBO Journal*, 30(11):2219–2232, 2011.

[50] D. R. Valenzano, B. A. Benayoun, P. P. Singh, E. Zhang, P. D. Etter, C.-K. Hu, M. Clément-Ziza, D. Willemsen, R. Cui, I. Harel, et al. The african turquoise killifish genome provides insights into evolution and genetic architecture of lifespan. *Cell*, 163(6):1539–1554, 2015.

[51] Y. Wang, J. Luo, H. Zhang, and J. Lu. microRNAs in the same clusters evolve to coordinately regulate functionally related genes. *Molecular Biology and Evolution*, page msw089, 2016.