Evolution of MicroRNAs Located Within *Hox* Gene Clusters

Andrea Tanzer [a,b], Chris T. Amemiya [c], Chang-Bae Kim [d], and Peter F. Stadler [a,b]

[a] *Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany*
{andrea,studla}@bioinf.uni-leipzig.de

[b] *Department of Theoretical Chemistry and Structural Biology, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*

[c] *Virginia Mason Research Center, Benaroya Research Institute, Molecular Genetics Dept. 1201 Ninth Avenue, Seattle, WA 98101 USA*
camemiya@benaroyaresearch.org

[d] *National Genome Information Center, Korea Research Institute of Bioscience and Biotechnology, 52 Eoun-dong, Yusong-gu, 305-333 Taejon, Korea*
changbae@kribb.re.kr

Abstract

MicroRNAs (miRNAs) form an abundant class of non-coding RNA genes that have an important function in post-transcriptional gene regulation and in particular modulate the expression of developmentally important transcription factors including *Hox* genes. Two families of microRNAs are genomically located in intergenic regions in the *Hox* clusters and vertebrates. Here we describe their evolution in detail.

Key words: micro RNA, mir-10, mir-196, iab-4, Hox Genes, Vertebrate Evolution

1 Introduction

MicroRNAs (miRNAs) form an abundant class of non-coding RNA genes. They are processed in the nucleus from a primary transcript to a hairpin-shaped precursor of about 80nt, which is exported to the cytoplasm where the single-stranded mature microRNAs of about 22nt in length is excised. These are incorporated in one or more RNP complexes that are instrumental

in the regulation of translation and degradation of mRNAs. We refer to a series of recent reviews [38, 44, 6] for a detailed discussion of their function and mechanisms as well as their history of discovery. The overall importance of microRNAs for development is highlighted by two facts: (1) many microRNAs have temporal and/or tissue-specific expression patterns, see e.g. [5, 30]; (2) most of the verified targets of microRNAs in both animals and plants are transcription factors, see e.g. Table 2 in [6]. An over-representation of putative mircroRNA targets in genes associated with transcriptional regulation has also been reported in computational studies [39], see also [52].

MicroRNAs genes which occur associated with genes encoding transcription factors are thus of particular interest. Indeed, there are at least three groups of microRNAs residing within the *Hox* gene clusters. *Hox* genes code for homeodomain containing transcription factors that are essential for embryonic patterning [41]. In many species they are organized in tightly linked clusters although in some cases the clusters have been broken up. The homology of the vertebrate *Hox* genes with the genes in the Drosophila homeotic gene clusters was demonstrated already a decade ago [2, 55]. The common ancestor of all recent gnathostomes (sharks, bony fish, and tetrapods) had four clusters homologous to the mammalian ones [29, 50]. The two agnathan lineages, lampreys and hagfish, also exhibit multiple *Hox* clusters which, however, arose through duplication events independent of those leading to the mammalian clusters [31, 17, 18, 59]. In contrast, protostomes and invertebrate deuterostomes (echinodermata, hemichordata, urochordata, and cephalochordata) have a single cluster [40, 48, 13, 21].

Mir-10 is located in the Antennapedia cluster of *Drosophila melanogaster* and has been reported in two mammalian *Hox* clusters [36, 37]. *Mir-196* has been found in a variety of vertebrates and is known to direct the cleavage of *HoxB8* mRNA in mouse embryos and also regulates the expression of *HoxC8*, *HoxD8*, and *HoxA7* [67]. The microRNA *iab-4* [11] is located in the bi-thorax cluster of *Drosophila melanogaster* [5, 37] and is predicted to target *Ubx*; it may well be analogue of the vertebrate *mir-196* but there is not recognizable sequence similarity.

The molecular evolution of microRNAs, maybe a bit surprisingly, has not been a main focus of research so far, with the exception of the *let-7*/*mir-125* family [46, 45, 43], which is present in metazoa with the exception of the most basal groups, and the *mir17*/*mir92* family which is also evolutionarily old and exhibits a complex history of tandem and cluster duplications in vertebrates [61]. In this short contribution we consider in detail the phylogenetic distribution and the evolutionary history of the three *Hox*-associated microRNAs *mir-10*, *mir-196*, and *iab-4*.

2 Materials and Methods

MicroRNA sequences were obtained from the `Rfam` microRNA registry, version 3.1. (April 2004) [23]. Genomic sequences of *Hox* clusters were retrieved from the NCBI database. In this study we use the previously described *Hox* clusters from *Homo sapiens* (Hs) [63], *Pan troglodytes* (Pt), *Mus musculus* (Mm) [66], *Rattus norvegicus* (Rn) [22], *Polypterus senegalus* (Ps) [10], *Takifugu rubripes* (Tr) [4], *Tetraodon nigroviridis* (Tn) [51], *Danio rerio* (Dr) [3], *Oreochromis niloticus* (On) [54], *Morone saxatilis* (Ms) [57], *Spheroides nephalus* (Sn) [4], *Heterodontus francisci* (Hf) [33], *Petromyzon marinus* (Pm) [31], *Drosophila melanogaster* (Dm) [65, 15], *Anopheles gambiae* (Ag) [49, 14], *Tribolium castaneum* (Tc) [8] , *Caenorhabditis elegans* (Ce) [9, 62], *Caenorhabditis briggsae* (Cb) [24]. *Gallus gallus* (Gg) sequences were taken from `pre-ensemble` in April 2004. After the our data analysis was complete a new release 4.0 of the microRNA registry was published in July 2004 containing predicted chicken microRNAs provided by the International Chicken Genome Sequencing Consortium. These sequences coincide with the results of our `blast` searches. *Xenopus tropicalis* (Xt) and *Ciona intestinalis* (Ci) data were taken from the JGI website [1] , *Strongylocentrotus purpuratus* date were obtained from the Baylor College of Medicine [2] , and *Ciona savignyi* sequences were downloaded from the Broad Institute at MIT [3] .

Furthermore, we used preliminary sequence data from the the NBCI databases GSS, WGS, and HTGS. Sequence for the following species were available in May 2004: *Apis mellifera* (Am), *Drosophila pseudoobscura* (Dp), *Bombyx mori* (Bm) [42], *Amia calva* (Ac), *Gasterosteus aculeatus* (Ga), *Sus scrofa* (Sc), *Bos taurus* (Bt), *Felis catus* (Fc), *Canis familiaris* (Cf), *Papio hamadryas* (Ph), *Carollia perspicillata* (Cp), *Otolemur garnettii* (Og).

The microRNA sequences in the *HoxB* and *HoxD* clusters of the hornshark *Hetrodontus francisci*, in the four *Hox* clusters of the coelacanth *Latimeria menadoensis* [12], and in the single *Hox* cluster of the amphioxus *Branchiostoma floridae* were obtained from unpublished complete cluster sequences.

We `blasted` (NCBI `blast` 2.2.8) the entire collection of microRNA precursor sequences from the `microRNA registry`, version 3.1., April 2004 [23] against all available *Hox* cluster sequences. We found only homologs of *mir-10*, *mir-196*, and *iab-4*. In addition, `blast` hits with very small *E*-values for *mir-333* were obtained in rodent sequences. These are related to rodent-specific repetitive elements rather than *bona fide* microRNAs (see below). We then `blast`ed the *Hox* cluster microRNAs against the genome databases.

[1] `www.jgi.doe.gov/`
[2] `ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Spurpuratus/`
[3] `http://www.broad.mit.edu/annotation/ciona/`

Fig. 1. Alignments of known and putative microRNA sequences and their secondary structures. Each matching pairs of parentheses denotes a base pair.
Top: Sequences of *mir-10* precursors from Latimeria and the sea urchin *Strongylocentrotus purpuratus* compared to two Human paralogs and a homolog from *Drosophila melanogaster*. Below: *mir-196* sequences from Latimeria and the lamprey *Petromyzon marinus* compared with the know rodent sequences. The mature miRNA is clearly identifiable as highly conserved block. The sea urchin and lamprey sequences have a longer stem-loop structure; the additional base pairs are indicated by brackets.

Putative microRNAs were then aligned using `clustalw` [64] and their consensus secondary structure was computed using `RNAalifold` [27]. This structure was compared to the secondary structure of the individual sequences (computed using `RNAfold` [28, 26]) to check whether the individual fold conforms with the consensus structure, Fig. 1. The boundaries of the precursor hairpin are determined by homology of both sequence and secondary structure with the known microRNAs from the `microRNA Registry`. Phylogenetic trees were computed using the neighbor-joining algorithm [53] implemented in the `phylip` package [16] with 1000 bootstrap replicates. All computations are preformed with the ∼80nt precursor molecules unless explicitly stated otherwise.

Possible homologies between distantly related sequences are assessed by computing the z-score of the sequence similarity score of a pairwise alignment in comparison to the distribution of sequence similarities of pairwise alignments of shuffled sequences as described in [61].

Accession numbers of genomic sequences and the sequences of both known and predicted microRNA precursors are listed in the electronic supplement [4]. These microRNA sequences have been submitted to the `rfam` microRNA registry.

[4] URL: `http:www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/04-013/`

3 Results

Fig. 2 summarizes the results of our survey for *mir-10*, *mir-196*, and *iab-4* homologs in metazoan sequences.

The known *mir-10* sequences are located between *Hox5* and *Hox4* in vertebrates and, correspondingly, between *Dfd* and *Scr* in arthropods. All *mir-10* homologs detected in our survey share this location. Of all genomes considered here, *mir-10* was absent only from the two nematodes *C. elegans* and *C. briggsae*, and from the two tunicate species *C. intestinalis* [13] and *C. savignyi* [5]. In both cases the *Hox* clusters have disintegrated into multiple pieces.

Vertebrate homologs of *mir-10* were found in *HoxB*, *HoxC*, and *HoxD* clusters of gnathostomes, while they are absent in all investigated *HoxA* clusters including shark, latimeria, bichir, various teleosts and tetrapods. The *mir-10* copy in the *HoxC* is present only in teleosts, Xenopus and Latimeria. At present there are no data available for the *HoxC* of a shark. A survey of the chicken genome and of all available mammalian genomes did not result in plausible candidates.

A *mir-10* homolog was identified in the sea urchin *Strongylocentrotus purpuratus*. Its precursor differs from its vertebrate homologs by an extended hairpin loop, Fig. 1. The genomic position of this putative microRNA is unknown since only individual contigs but not genome assembly is available. Sea urchins lack a *Hox4* gene [40], hence the detection of a *mir-10* was somewhat surprising since *mir-10* is located at a rather well-conserved distance of only about 1.5kb upstream of *Hox4* in vertebrates. Recently [25] showed, however, that *HoxB3a* in zebrafish is produced from two alternative primary transcripts, one of which starts already a short distance downstream of *Hox5* and contains *mir-10* in an intron. It is plausible to assume that the *mir-10* precursor is produced from the excised intron, so that *mir-10* expression would be linked to the expression of *Hox-3* rather than *Hox-4*.

The *Hox* cluster of both Ciona species is distributed over different scaffolds of their respective genome assemblies in the same way: only *Hox12/13-Hox11/12*, *Hox6/7-Hox5* and *Hox4-Hox3-Hox2*, resp., are located tightly linked on the same scaffolds, while the *Hox1* and *Hox10* genes appear on individual scaffolds [58]. A number of conserved sequence motifs are located between *Hox4* and the next gene upstream (a galactose 6-O-sulfotransferase). However, none of them forms a conserved hairpin structure. We therefore conclude that there is no analog of *mir10* in Ciona.

The *mir-196* sequences are located between *Hox10* and *Hox9*, or upstream

[5] URL: `http://www.broad.mit.edu/annotation/ciona/`.

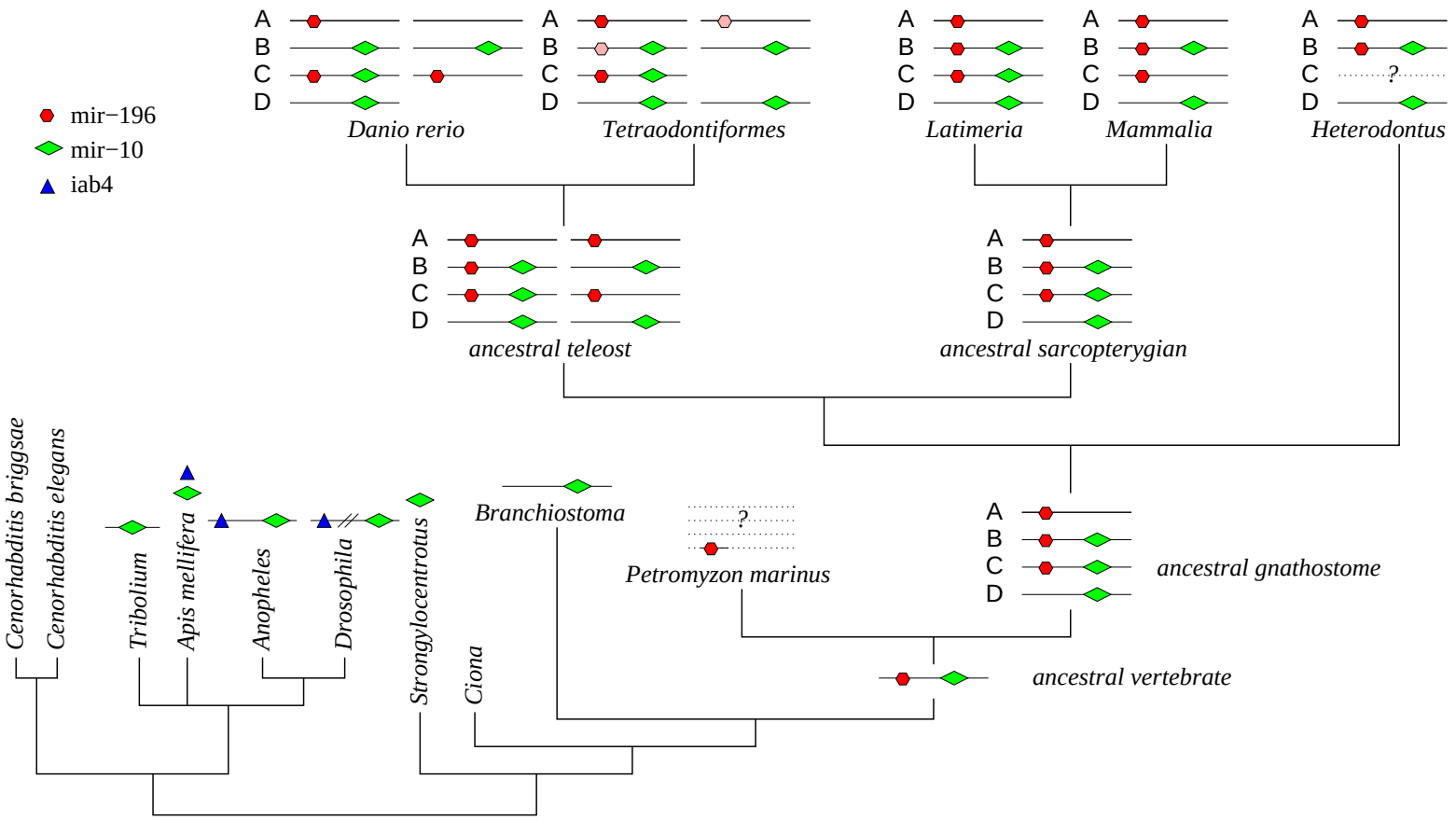


Fig. 2. MicroRNA within *Hox* clusters. Symbols on a line indicate that the microRNA is located in a cluster or at least physically linked to a *Hox* gene. Other homologs of the the *Hox* cluster related microRNAs are indicated by isolated symbols.

of *Hox9*, respectively. No invertebrate homologs of the known *mir-196* were found. Many of the *mir-196* sequences are listed already in the supplemental material of [67]. We were able to find additional members of this family in teleosts, and in particular in shark and latimeria. In addition, a *mir-196* is located downstream of a *Hox10* gene (designated *HoxW10a* in *Irvine:02*) of the lamprey *Petromyzon marinus*.

All gnathostome *mir-196* sequences are located in the *HoxA*, *HoxB*, and *HoxC* clusters, no candidates were detectable in any of the available *HoxD* cluster sequences. The distribution of *mir-196* sequences in teleost fishes is of particular interest. The only sequence in a *HoxB* clusters is located in the *HoxBa* cluster of the fugu (*Takifugu rubripes*). The *HoxA*-paralog of *mir-196* appears both in the *HoxAa* and the *HoxAb* cluster of pufferfishes, while the *HoxC*-paralog was retained after the duplication of the *HoxC* cluster in the zebrafish (*Danio rerio*) only.

As noted in [67] there is a different, unrelated microRNA, *iab-4*, in the corresponding region between *AbdB* and *AbdA* in the insect *Hox* clusters. We find that this sequence is conserved in larger group of insect species but probably not even throughout the arthropod clade.

Some of the microRNAs described here have been identified in previous studies as so-called "phylogenetic footprints", i.e., as conserved non-coding sequences, being identified as microRNAs. For example, CNS 6 in [25] is *mir-10-B*, while "footprint clique #169" in [51] is *mir-196-A*, and footprint A2 (10-9b) in [31, 18] corresponds to a lamprey homolog of *mir-196*.

Surprisingly, two of these microRNAs, namely human *mir-10-B* and *mir-196-A* from human, mouse, pig, and chicken, appear in ESTs. The *mir-196-A* sequence is located in the 5'UTR of *HoxA9* transcripts. Extensive alternative splicing has been reported for this gene in both human and mouse [20, 35, 47]. Since microRNAs have to undergo a maturation process already in the nucleus, in particular excision of the pre-microRNA by *Drosha*, they are most probably inactive when located on a mature mRNA. This suggests that the expression of at least of some microRNAs is linked to and regulated by alternative splicing of their host genes.

Both *mir-196* and the *mir-10* precursor sequences are very well conserved so that plausible alignments were obtained using `clustalw`. The reconstructed gene phylogenies, Fig. 3 are consistent with the established species phylogeny. They clearly reflect the duplication of the *Hox* clusters at the root of the vertebrates and the later duplication of the *Hox* clusters in the teleosts.

The gene trees in Fig. 3 suggest an elevated rate of evolution of *mir-10* in the *HoxBb* and *HoxDb* clusters of the teleosts. We used Tajima's relative rate test [60] to test this hypothesis and find that, indeed, the *mir-10-Bb* sequences of

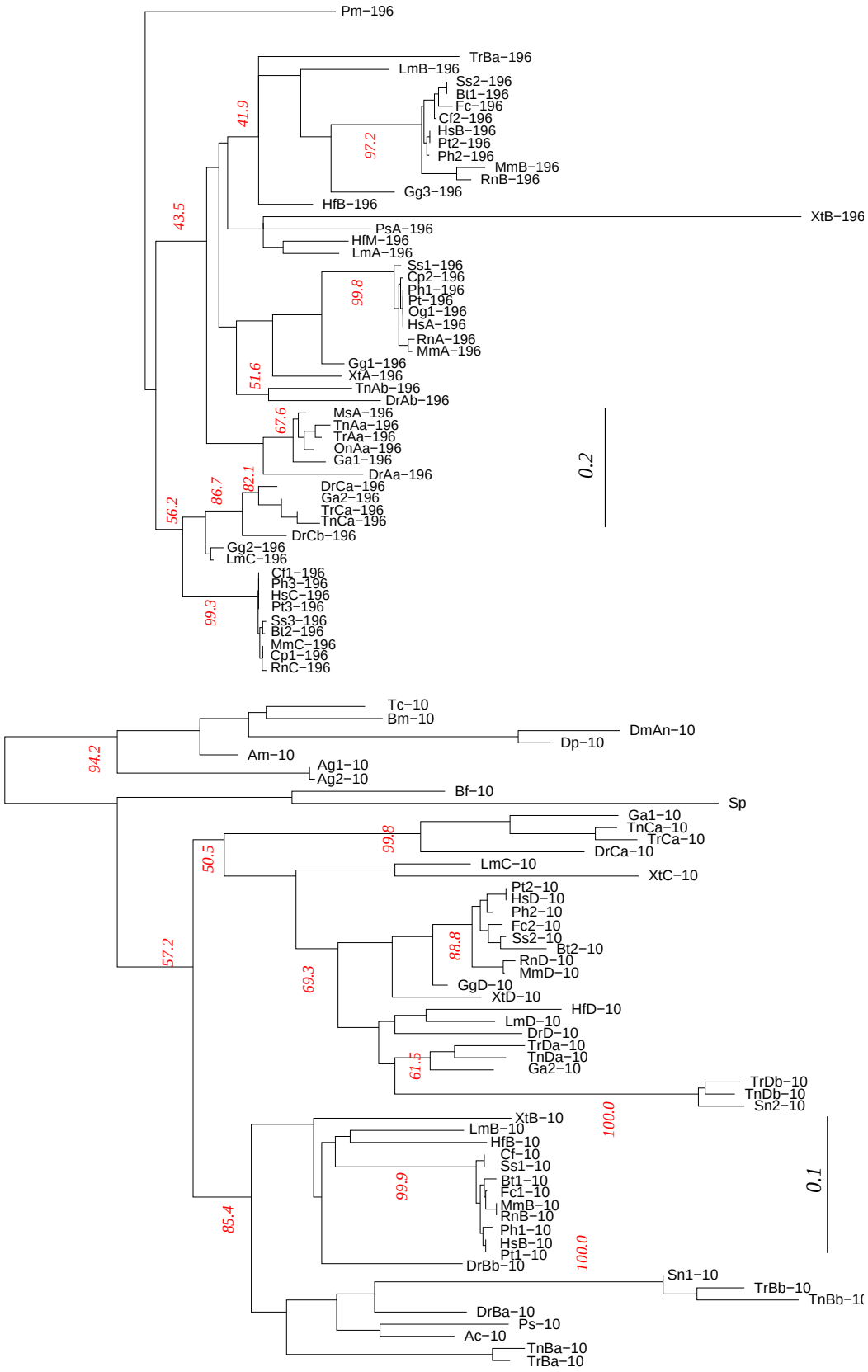


Fig. 3. Neighbor-joining trees of microRNA sequences. Bootstrap values of selected branches are marked in red.

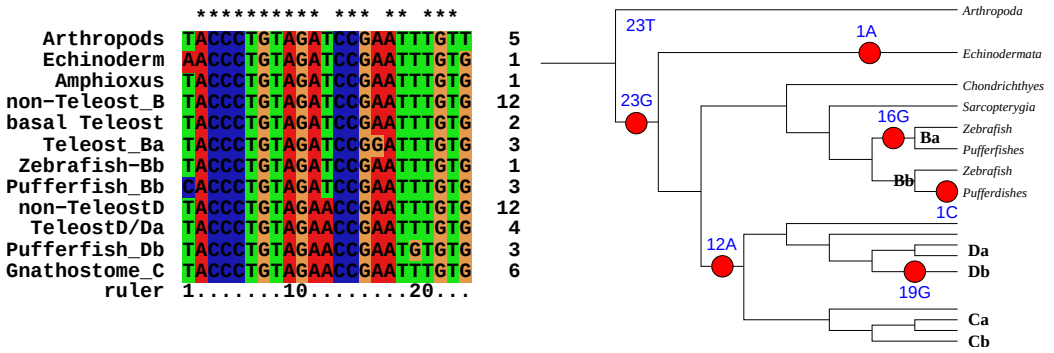


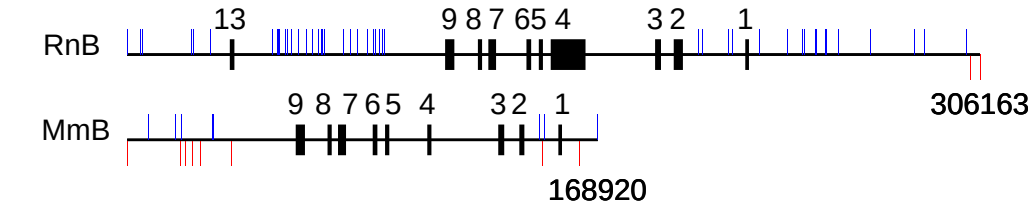
Fig. 4. Mutations in the mature *mir-10* sequence. Left: sequence alignment; the number of sequences in each group is indicated, additional mutations in only a single group member are ignored. Right: most parsimonious assignment of the mutations to the generally accepted phylogeny of the *Hox* clusters.

the pufferfishes evolve significantly faster (χ^2 values in the range from 4-6). The same is true for the pufferfish *mir-10-Db*. Rate comparisons along other branches did not yield significant rate differences. In contrast, there are no significant rate deviations in the *mir-196* family (with the exception of the highly derived Xenopus sequence).

The mature microRNAs are extremely well conserved. Interesting, most of the few mutations are characteristic for individual clades. For *mir-196* the variant in mammalian *HoxA* clusters differs by a single point mutation (12:G→A) from all other sequences. In the (evolutionarily older) *mir-10* family the situation is more interesting. Pufferfishes (but apparently not all percomorph fishes, judging from the single stickleback sequence) share mutations in the *mir-10-Db* (19:T→G) and *mir-10-Bb* sequences (1:T→C). This fits the observation of increased evolution rates in these microRNAs. The teleost *mir-10-Ba* sequences are set apart by (16:A→G). The gnathostome *HoxC* and *HoxD* share (12:T→A), while chordates and arthropod sequences differ in position 23 (G↔T).

The *iab-4* microRNA might be an analog of the vertebrate *mir-196* in arthropods [67]. We have therefore search for weak sequence similarities between the two classes of sequences using the z-score obtained by comparing the sequence identity with the distribution of sequence identities between shuffled sequences [61]. For all comparisons of an *mir-iab-4* with a *mir-196* sequence we find values of $z < 1.0$, far from the significance threshold.

In rodents we find a large number of `blast` hits of *mir-333* [34] with $E < 10^{-3}$ throughout the genome, including in and around the *Hox* clusters. The complete *mir-333* sequence, however, does not map to a *Hox* cluster in the rat *Rattus norvegicus*, from which is was originally obtained [34]. the Fig. 5 shows the *HoxB* locus as an example. The distribution of the `blast` hits in




```
MATURE          ---------------GTG-GTGTGCTAGTTACTTTT-------------------------------------------------------------
FOLD            ..((..(((((....(..-(((((((....((((((((((((.....))))))))))))....)))))..))..)..))))))..))..........
rno-mir-333     CCCCGGTGGAACCACGTG-GTGTGCTAGTTACTTTTGGGCTGGAGAGACGGCTCAGGGGTTAAGAGCACAGACTGCTCTTCCAGAGGTCCTGAGTT
RnB_59227       GTCAAGGAACAACATTAG-AAAAGACATTTACTCCAGGGCTGGAGAGATGGCTCAGGGTTTAAGAGCACTGTCTGCCCTTCCAGAGGTCCTGAGTT
RnA_79763       GTGTTTGCATAGCATGCATGAGACCTAG-GTTACCCAGGCTGGAGATATGGCTCAGTGGTTAAGAGCATTGACTGCTCTTCCAGAGGTCCTGAGTT
RnB_238554      -CATGAAACAGGCATGGATAAGAAGTAGAGGTTTTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCATTGACTGCTCTTCCAGAGGTCCTGAGTT
RnA_264360      ACTCTGGAGGAAGGCAAATTAGAACTAT-TTTCCTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACCGACTGCTCTTCCAGAGGTCCTGAGTT
RnB_252298      -TGGTGCATAGAAATGCATGCAGGGGGTTGGAGGGGAGGCTGGAGAGATGGCTCAGCGGTTAAGAGCACTGACTGCTCTTCCAGAGGTCCTGAGTT
                                          ********* * ******* * *********  * **** ********************
```

Fig. 5. Top: `blast` hits with $E < 10^{-3}$ of *rno-mir-333* in the genomic location of the *HoxB* clusters in mouse (Mm) and rat (Rn). The location of the hits on the positive and negative strands is indicated by spikes above or below the line, respectively. *Hox* genes are denotes by bars. Note that blast hits are depleted within the *Hox* clusters with the exception of the large intergenic region between *HoxB13* and *HoxB9*. Below: `clustalw` alignment of *rno-mir-333* and the five *B2-mm2* sequences from *Hox* clusters most similar to *rno-mir-333*. There is no resemblance of the the mature microRNA with one of the other sequences, while the 3'part of *rno-mir-333* is almost identical with the repetitive sequence element.

the area of *Hox* clusters indicates a drastically reduced density with the *Hox* clusters (with a few specific exception). This pattern matches the observation in [19] that repetitive DNA elements are strongly excluded from gnathostome *Hox* clusters. Sequence comparison shows that *mir-333* is almost identical with an abundant rodent-specific short retro-transposon (SINE), *B2-mm2* [32, 56], from its position 35 to the 3' end of the microRNA precursor, Fig. 5. This sequence interval does not include the mature microRNA, however. It is plausible that *mir-333* has originated from this rodent-specific element SINE, that itself is ancestrally derived from tRNAs [32]. Additional support for this hypothesis stems from the fact that the tRNA derived parts of B2 elements fold into a specific secondary structure that features a long hairpin loop [7, 32].

4 Discussion

The *mir-10* sequence is evolutionarily ancient and was probably present already in the common ancestor of protostomes and deuterostomes. Its absence in nematodes and tunicates is possibly related to the disintegration of the *Hox* clusters in these species. Both *C. briggsae* and *C. elegans* have lost their *Hox2-Hox4* paralogs, see [1], so that we do not expect to find a *mir-10* sequence in these species.

In contrast, *mir-196* homologs were detectable only in vertebrates. The *iab-4* microRNA, which might have an analogous regulatory function in arthropods [67], shows no detectable sequence homology. The fact that a *mir-196* was found in the agnathan *Petromyzon marinus* but not in more basal deuterostomia (amphioxus, tunicates, or sea urchins) suggests that the origin of *mir-196* is linked to the advent of the vertebrates.

The evolution of the *Hox*-cluster microRNAs closely follows the history of their "host' *Hox*-clusters. Subsequent to the genome duplication at the root of the vertebrate clade we observe loss of both *mir-10* and *mir-196* in one of the four paralog clusters. Given that the functional component of a microRNA is the mature 22-mer, which has remained almost identical throughout vertebrate evolution, it is surprising that this redundancy was not reduced more drastically.

The retention of most of the ancient microRNA paralogs suggest that the expression of the paralogs is regulated, probably linked to the *Hox9* and *Hox3* genes, resp., so that different paralogs act at different times and in different tissues. It is tempting to speculate, based on their extreme sequence similarity, that these paralog mircoRNAs cannot discriminate between different targets. The existence of paralog microRNA precursors would thus provide an additional degree of freedom for fine-tuning the spatio-temporal expression patterns of the mature miRNA which is further enhanced by means of both differential transcriptional regulation (e.g. *mir-10*/*HoxB3*) and alternative splicing (e.g. *mir-196*/*HoxA9*). An experimental test of this conjecture would require techniques for monitoring the precursor microRNA.

Acknowledgments

This work is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung, Project No P-15893 (AT,PFS), the German DFG Bioinformatics Initiative, BIZ-6/1-2 (AT,PFS), the Korean KRIBB Research Initiative Program (CBK), the National Institute of Health, HG-02526 (CTA), and the National Science Foundation, IBN-0321461 (CTA).

References

[1] A. A. Aboobaker. Hox gene loss during dynamic evolution of the nematode cluster. *Curr. Biol.*, 13:37–40, 2003.

[2] M. Akam. *Hox* and *HOM*: homologous gene clusters in insects and vertebrates. *Cell*, 57:347–349, 1989.

[3] A. Amores, A. Force, Y. L. Yan, L. Joly, C. Amemiya, A. Fritz, R. K. Ho,

J. Langeland, V. Prince, Y. L. Wang, M. Westerfield, M. Ekker, and J. H. Postlethwait. Zebrafish *Hox* clusters and vertebrate genome evolution. *Science*, 282:1711–1714, 1998.

[4] A. Amores, T. Suzuki, Y.-L. Yan, J. Pomeroy, A. Singer, C. Amemiya, and J. Postlethwait. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. *Genome Res.*, 14:1–10, 2004.

[5] A. A. Aravin, M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, G. T., J. Meyer, and T. Tuschl. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell*, 5:337–350, 2003.

[6] D. P. Bartel. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 116:281–297, 2004.

[7] T. S. Bladon, C. H. Fregeau, and M. W. McBurney. Synthesis and processing of small B2 transcripts in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, 10:4058–4067, 1990.

[8] S. J. Brown, J. P. Fellers, E. A. Shippy, Teresa D. Richardson, M. Maxwell, J. J. Stuart, and R. E. Denell. Sequence of the *Tribolium castaneum* homeotic complex: The region corresponding to the *Drosophila melanogaster* antennapedia complex. *Genetics*, 160:1067–1074, 2002.

[9] T. R. Burglin and G. Ruvkun. The *Caenorhabditis elegans* homeobox gene cluster. *Curr. Opin. Genet. Dev.*, 3:615–620, 1993.

[10] C.-H. Chiu, K. Dewar, G. P. Wagner, K. Takahashi, F. Ruddle, C. Ledje, P. Bartsch, J.-L. Scemama, E. Stellwag, C. Fried, S. J. Prohaska, P. F. Stadler, and C. T. Amemiya. Bichir *HoxA* cluster sequence reveals surprising trends in rayfinned fish genomic evolution. *Genome Res.*, 14:11–17, 2004.

[11] S. Cumberledge, A. Zaratzian, and S. Sakonju. Characterization of two RNAs transcribed from the cis-regulatory region of the *abd-A* domain within the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA*, 87:3259–3263, 1990.

[12] J. Danke, T. Miyake, T. Powers, J. Schein, H. Shin, I. Bosdet, M. Erdmann, R. Caldwell, and C. T. Amemiya. Genome resource for the indonesian coelacanth, *Latimeria menadoensis*. *J. Exp. Zool. A: Comp. Exp. Biol.*, 301:228–234, 2004.

[13] P. Dehal, Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. M. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H.-G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chinbow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B.-I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy,

N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. K. M. Levine, N. Satoh, and D. S. Rokhsar. The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. *Science*, 298:2157–2167, 2002.

[14] M. P. Devenport, C. Blass, and P. Eggleston. Characterization of the *Hox* gene cluster in the malaria vector mosquito, *Anopheles gambiae*. *Evol. Dev.*, pages 326–339, 2000.

[15] R. Drysdale. The *Drosophila melanogaster* genome sequencing and annotation projects: a status report. *Brief Funct. Genomic Proteomic*, 2:128–134, 2003.

[16] J. Felsenstein. Phylip – phylogeny inference package (version 3.2). *Cladistics*, 5:164–166, 1989.

[17] A. Force, A. Amores, and J. H. Postlethwait. Hox cluster organization in the jawless vertebrate *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)*, 294:30–46, 2002.

[18] C. Fried, S. J. Prohaska, and P. F. Stadler. Independent hox-cluster duplications in lampreys. *J. Exp. Zool., Mol. Dev. Evol*, 299B:18–25, 2003.

[19] C. Fried, S. J. Prohaska, and P. F. Stadler. Exclusion of repetitive dna elements from gnathostome *Hox* clusters. *J. Exp. Zool., Mol. Dev. Evol*, 302B:165–173, 2004.

[20] S. Fujimoto, K. Araki, O. Chisaka, M. Araki, K. Takagi, and K.-i. Yamamura. Analysis of the murine *Hoxa-9* cDNA: an alternatively spliced transcript encodes a truncated protein lacking the homeodomain. *Gene*, 16:77–85, 1998.

[21] J. Garcia-Fernández and P. W. Holland. Archetypal organization of the amphioxus hox gene cluster. *Nature*, 370:563–566, 1994.

[22] R. A. Gibbs and *et al.*: Rat Genome Sequencing Project Consortium. Genome sequence of the brown norway rat yields insights into mammalian evolution. *Nature*, 428:493–521, 2004.

[23] S. Griffiths-Jones. The microRNA Registry. *Nucl. Acids Res.*, 32:D109–D111, 2004. Database issue.

[24] B. P. Gupta and P. W. Sternberg. The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*. *Genome Biol.*, 4:#238, 2003.

[25] T. Hadrys, V. Prince, M. Hunter, R. Baker, and S. Rinkwitz. Comparative genomic analysis of vertebrate Hox3 and Hox4 genes. *J. Exp. Zool. B: Mol. Dev. Evol.*, 302:147–164, 2004.

[26] I. L. Hofacker. The Vienna RNA secondary structure server. *Nucl. Acids Res.*, 31:3429–3431, 2003.

[27] I. L. Hofacker, M. Fekete, and P. F. Stadler. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, 319:1059–1066, 2002.

[28] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures.

Mh. Chem., 125:167–188, 1994.

[29] P. W. Holland and J. Garcia-Fernández. Hox genes and chordate evolution. *Dev. Biol.*, 173:382–395, 1996.

[30] H. B. Houbaviy, M. F. Murray, and P. A. Sharp. Embryonic stem cell-specific microRNAs. *Dev. Cell*, 5:351–358, 2003.

[31] S. Q. Irvine, J. L. Carr, W. J. Bailey, K. Kawasaki, N. Shimizu, C. T. Amemiya, and F. H. Ruddle. Genomic analysis of Hox clusters in the sea lamprey, *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)*, 294:47–62, 2002.

[32] D. H. Kass, J. Kim, A. Rao, and P. L. Deininger. Evolution of B2 repeats: the muroid explosion. *Genetica*, 99:1–13, 1997.

[33] C. B. Kim, C. Amemiya, W. Bailey, K. Kawasaki, J. Mezey, W. Miller, S. Minosima, N. Shimizu, G. P. Wagner, and F. Ruddle. Hox cluster genomics in the horn shark, *heterodontus francisci*. *Proc. Natl. Acad. Sci. USA*, 97:1655–1660, 2000.

[34] J. Kim, A. Krichevsky, Y. Grad, G. D. Hayes, K. S. Kosik, G. M. Church, and G. Ruvkun. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl. Acad. Sci. USA*, 101:360–365, 2004.

[35] M. H. Kim, H. H. Chang, C. Shin, M. Cho, D. Park, and H. W. Park. Genomic structure and sequence analysis of human *HoxA-9*. *DNA Cell. Biol.*, 17:407–414, 1998.

[36] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858, 2001.

[37] M. Lagos-Quintana, R. Rauhut, J. Meyer, A. Borkhardt, and T. Tuschl. New microRNAs from mouse and human. *RNA*, 9:175–179, 2003.

[38] E. C. Lai. microRNAs: Runts of the genome assert themselves. *Curr. Biol.*, 13:R925–936, 2003.

[39] B. P. Lewis, I.-h. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge. Prediction of mammalian microRNA targets. *Cell*, 115:787–798, 2003.

[40] P. Martinez, J. R. Rast, C. Arena-Mena, and E. H. Davidson. Organization of an echinoderm *Hox* gene cluster. *Proc. Natl. Acad. Sci. USA*, 96:1469–1471, 1999.

[41] W. McGinnis and R. Krumlauf. Homeobox genes and axial patterning. *Cell*, 68:283–302, 1992.

[42] K. Mita, M. Kasahara, S. Sasaki, Y. Nagayasu, T. Yamada, H. Kanamori, N. Namiki, M. Kitagawa, H. Yamashita, Y. Yasukochi, K. Kadono-Okuda, K. Yamamoto, M. Ajimura, G. Ravikumar, M. Shimomura, Y. Nagamura, T. Shin-I, H. Abe, T. Shimada, S. Morishita, and T. Sasaki. The genome sequence of silkworm, *Bombyx mori*. *DNA Res.*, 29:27–35, 2004.

[43] E. Moss and L. Tang. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.*, 15:432–442, 2003.

[44] P. Nelson, M. Kiriakidou, A. Sharma, E. Maniataki, and Z. Mourelatos. The microRNA world: small is mighty. *Trends Biochem. Sci.*, 28:534–540, 2003.

[45] A. E. Pasquinelli, A. McCoy, E. Jiménez, S. Emili, G. Ruvkun, M. Q. Martindale, , and Baguñà. Expression of the 22 nucleotide let-7 heterochronic RNA throughout the metazoa: a role in life history evolution? *Evol. Dev.*, 5:372–378, 2003.

[46] A. E. Pasquinelli, B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kurodak, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Müller, A. Spring, Jürg Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408:86–89, 2000.

[47] C. V. Patel, S. Bandyopadhyay, and P. E. DiCorleto. Endothelial cells express a novel, tumor necrosis factor-α-regulated variant of *HoxA9*. *J. Biol. Chem.*, 274:1415–1422, 1999.

[48] J. Pendleton, B. K. Nagai, M. T. Murtha, and F. H. Ruddle. Expansion of the Hox gene family and the evolution of chordates. *Proc. Natl. Acad. Sci. USA*, 90:6300–6304, 1993.

[49] T. P. Powers, J. Hogan, Z. Ke, K. Dymbrowski, X. Wang, F. H. Collins, and T. C. Kaufman. Characterization of the hox cluster from the mosquito *Anopheles gambiae* (Diptera: Culicidae). *Evol. Dev.*, 2:311–325, 2000.

[50] S. J. Prohaska, C. Fried, C. T. Amemiya, F. H. Ruddle, G. P. Wagner, and P. F. Stadler. The shark HoxN cluster is homologous to the human HoxD cluster. *J. Mol. Evol.*, page 58, 2004. 212-217.

[51] S. J. Prohaska and P. F. Stadler. The duplication of the hox gene clusters in teleost fishes. *Th. Biosci.*, 123:89–110, 2004.

[52] N. Rajewski and N. D. Socci. Computational identification of microRNA targets. *Devel. Biol.*, 267:529–535, 2004.

[53] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol. Evol.*, 4:406–425, 1987.

[54] S. Santini, J. L. Boore, and A. Meyer. Evolutionary conservation of regulatory elements in vertebrate *Hox* gene clusters. *Genome Res.*, 13:1111–1122, 2003.

[55] F. R. Schubert, K. Nieselt-Struwe, and P. Gruss. The antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci. USA*, 90:143–147, 1993.

[56] I. M. Serdobova and D. A. Kramerov. Short retrotransposons of the B2 superfamily: Evolution and applications for the study of rodent phylogeny. *J. Mol. Evol.*, 46:202–214, 1998.

[57] E. A. Snell, J. L. Scemama, and E. J. Stellwag. Genomic organization of the *Hoxa4-Hoxa10* region from *Morone saxatilis*: implications for *Hox* gene evolution among vertebrates. *J. Exp. Zool. (Mol. Dev. Evol.)*, 285:41–49, 1999.

[58] A. Spagnuolo, F. Ristoratore, A. Di Gregorio, F. Aniello, M. Branno, and R. Di Lauro. Unusual number and genomic organization of *Hox* genes in the tunicate *Ciona intestinalis*. *Gene*, 309:71–79, 2003.

[59] P. F. Stadler, C. Fried, S. J. Prohaska, W. J. Bailey, B. Y. Misof, F. H. Ruddle, and G. P. Wagner. Evidence for independent *Hox* gene duplications in the hagfish lineage: A PCR-based gene inventory of *Eptatretus stoutii*. *Mol. Phylog. Evol.*, 2004. in press, DOI doi:10.1016/j.ympev.2004.03.015.

[60] F. Tajima. Simple methods for testing molecular clock hypothesis. *Genetics*, 135:599–607, 1993.

[61] A. Tanzer and P. F. Stadler. Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, 339:327–335, 2004.

[62] The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, 282:2012–2018, 1998.

[63] The Human Genome International Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

[64] J. D. Thompson, D. G. Higgs, and T. J. Gibson. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucl. Acids Res.*, 22:4673–4680, 1994.

[65] G. von Allmen, I. Hogga, A. Spierer, F. Karch, W. Bender, H. Gyurkovics, and E. Lewis. Splits in fruitfly *Hox* gene complexes. *Nature*, 380:116, 1996.

[66] R. H. e. M. G. S. C. Waterston. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420:520–562, 2002.

[67] S. Yekta, I.-h. Shih, and D. P. Bartel. MircoRNA-directed cleavage of *HoxB8* mRNA. *Science*, 304:594–596, 2004.