Molecular Evolution of a MicroRNA Cluster

Andrea Tanzer [a,b] and Peter F. Stadler [a,b]

[a] *Lehrstuhl für Bioinformatik am Institut für Informatik und Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany.*
{andrea,studla}@bioinf.uni-leipzig.de

[b] *Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
{at,studla}@tbi.univie.ac.at

Abstract

Many of the known microRNAs are encoded in polycistronic transcripts. Here we reconstruct the evolution of the **mir17** microRNA clusters which consist of *miR-17*, *miR-18*, *miR-19a*, *miR-19b*, *miR-20*, *miR-25*, *miR-92*, *miR-93*, *miR-106a*, and *miR-106b*. The history of this cluster is governed by an inital phase of local (tandem) duplications, a series of duplications of entire clusters and subsequent loss of individual microRNAs from the resulting paralogous clusters. The complex history of the **mir17** microRNA family appears to be closely linked to the early evolution of the vertebrate lineage.

Key words: microRNA, mir-17, polycistronic transcript, vertebrate evolution

1 Introduction

MicroRNAs (miRNAs) form a class of non-coding RNA genes whose products are small single-stranded RNAs with a length of about 22nt. These are involved in the regulation of translation and degradation of mRNAs. We refer to the recent review (1) for a discussion of their functions and mechanisms as well as their history of discovery. Almost 800 microRNAs from different animal and plant species have been reported so far [1] and a dedicated database, the `miRNA Registry` (2) has been created to collect information about them.

The expression of miRNAs in animals involves at least two processing steps (3). A long primary transcript, called the pri-miRNA, which may be polycistronic, is processed in the nucleus, yielding one or more hairpin precursor

[1] Rfam database Release 3.0: Jan. 2003,
`http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml`

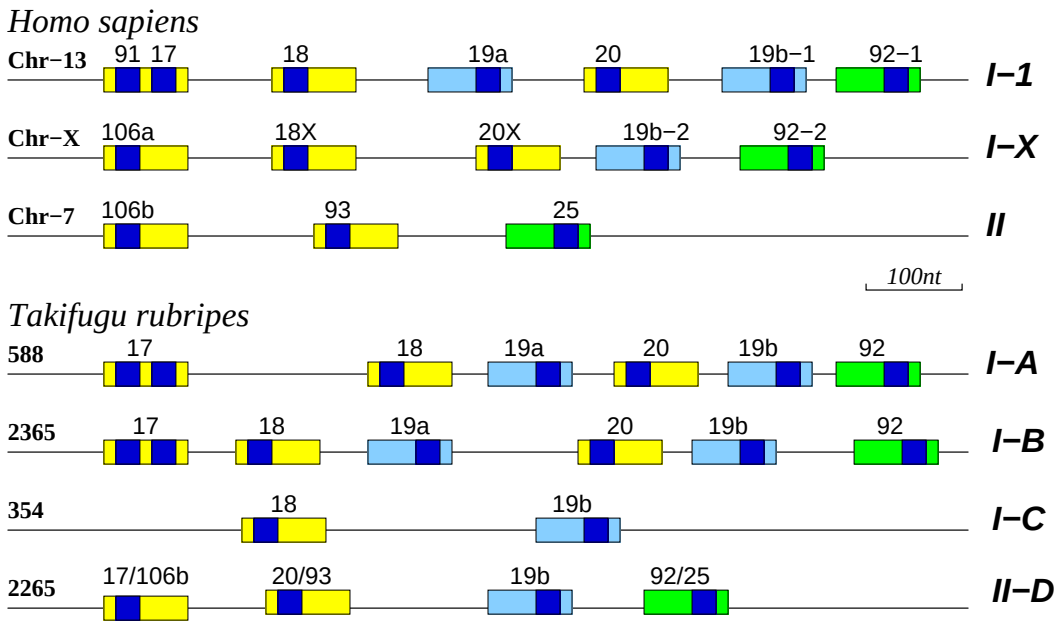


Fig. 1. The **mir17** microRNA clusters: Three paralog groups of microRNA precursors can be identified: *miR17/18/20/93/106/106b/93* (yellow), and *miR19/19b* (blue) and *miR92/25* (green). Mammals contain at least three paralog clusters while teleost fishes have at least four copies. The position of the mature miRNA is indicated by a dark box.

sequences (pre-miRNAs). These are exported to the cytoplasm by means of the *Exportin-5* pathway (4), where the mature miRNA is excised by the enzyme complex *Dicer*. The final destination of the mature miRNA is a ribonucleoprotein complex *miRNP* (5). Despite recent efforts on microRNA target prediction (6; 7; 8) little is know about the specific functionality of most miRNAs.

Many of the known microRNAs appear in clusters on a single polycistronic transcript (3; 9; 10; 11). In this contribution we reconstruct the evolutionary history of the **mir17** cluster and its paralogs in detail.

The human **mir17** cluster (3) contains 6 pre-miRNAs within about 1kb on chromosome 13, Fig. 1: *pre-mir-17* is the precursor of both *miR-17* on its 3'-arm (12) and of *miR-91* (*miR-17-5p*) at its 5'-arm (9). The other members of this cluster were already reported in (12) and/or in (9): *miR-18*, *miR-19*, *miR-19b*, *miR-20*, and *miR-92*. The **mir17** cluster is of particular interest because the human X-chromosome is known to contain paralogs of some of these micro RNAs: *miR-106* is a homolog of *miR-17* (9), *miR-19b-2* is a homolog of *miR-19b-1*, and *miR-92-2* is a homolog of *miR-92-1*. In fact, these sequences are located on a 1kb interval of the X-chromosome that also contains sequences homologous with *pre-mir-18* and *pre-mir-20*, Fig. 1.

Homologous microRNAs have been reported also in the mouse: *mmu-miR-18* and *mmu-miR-20* are located on chromosome 14 (13; 14), while *mmu-miR-19b*, *mmu-miR-92-2* and *mmu-miR-106a* (14) are located on the X-chromosome. A rat member of this family is *rno-miR-20* (15).

A miRNA similar to *miR-106a* is reported in (14) both as *miR-94* and *miR-106b*. Inspection of the adjacent region of the mouse chromosome 5 shows that *miR-106b* is indeed part of a microRNA cluster that also contains *miR-93* (14). For both sequences human homologs mapping to chromosome 7 are known as well (9). A third sequence, human *miR-25* (12), also belongs to this cluster. We find that *miR-25* is distantly related to *miR-92*. Parts of all three clusters are reported in (9, Fig.5).

Homologs of *mir-92* are also known in invertebrates. *Drosophila* contains two copies, *miR-92a* and *miR-92b*, that are located within some 5000nt on chromosome 3R (11; 16; 17).

2 Methods

The publicly available genome databases (see Appendix for URLs) were searched using `blastn` (18) against all pre-miRNAs of the **mir17** family mentioned in the previous section. Conversely, the entire MicroRNA Registry, version 3.0, was compared against the genomic sequences near the putative family members. Exact locations of homologs of known miRNAs were identified using `clustalw` (19) alignments and subsequent prediction of the secondary structure using `Vienna RNA Package` (20), in particular the programs `RNAfold`, `RNAalifold` (21), `RNALfold` (22), and `alidot` (23), in order to verify the hairpin structure of the precursor, see Fig. 2 for an example.

Phylogenetic trees were reconstructed both with Maximum Parsimony and Neighbor-joining using the `phylip` package (25) with standard parameters. The phylogeny of the entire clusters was computed using a concatenation of the alignments of the individual paralogous microRNAs according to their order in the cluster, and treating microRNAs that are not present in a particular cluster as missing data. This ensures that distances are measured based on nucleic acid substitution frequencies, not based on changes of cluster organization.

In order to identify distant sequence similarities between pre-miRNAs from different paralog groups we compute a similarity score based on the significance of the alignment score: The identity score $s(I, J)$ for the pairwise alignment of two pre-miRNAs I and J is computed using the implementation of the fast approximate Wilbur-Lipman algorithm (26) from the `clustalw` program. Then the mean score m and the variance v are estimated from a sample of 1000

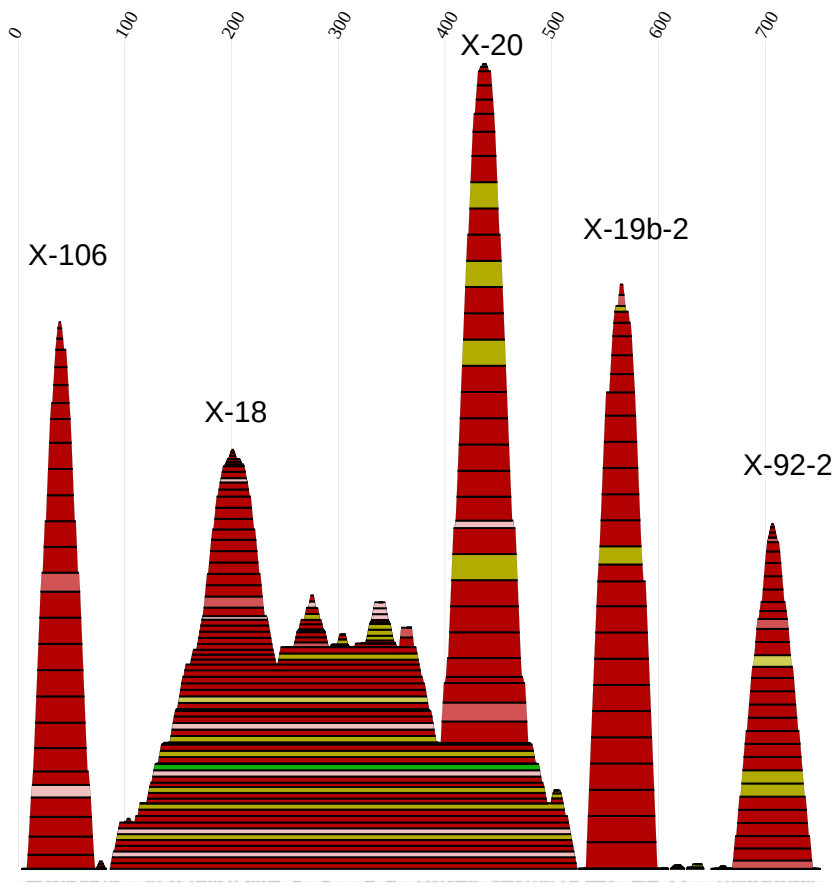


Fig. 2. Predicted consensus secondary structure of the **mir17** cluster paralogs I-X located at the mammalian X-chromosomes. The 5 pre-miRNAs appear as well-conserved hairpin structures. The computation was performed with `RNAfold` and `alidot` from a `clustalw` alignment as described in (21). The color-coded mountain plot (24) shows each basepair between sequence positions i and j as a slab from i to j with a thickness proportional to the probability that the base pair is formed. The saturation of the color indicates whether there are sequences in the alignment that cannot form a basepair so that basepairs with conflicts in one or two sequences appear pale. The color indicates the consistent and compensatory mutations: a conserved base pair is shown in red, two types of base pairs are shown in yellow, three in green. We see some structural ambiguities (pale base pairs) usually next to interior loops or bulges or at the base of the stem-loop structures, almost perfect sequence conservation for *pre-miR-106* and *pre-miR-18X*, while the structures of *pre-miR-20X*, *pre-miR-19b-2* and *pre-miR-92-2* receive additional support from a number of compensatory mutations.

alignments of sequences I_π and J_π that are obtained by randomly permuting the positions of I and J independently of each other. The z-score

$$z(I,J) = \frac{s(I,J) - m}{\sqrt{v}} \tag{1}$$

is then used as similarity measure of I and J for WPGMA clustering (27). This method produces robust similarity scores in regimes where reliable global alignments cannot be obtained.

The duplication history of the **mir17** family was reconstructed "by hand" based on the following assumptions: Edit operations are (a) duplications of individual microRNAs within a linked cluster, (b) the deletion of a microRNA, and (c) the duplication of an entire cluster. In other words, we explicitly exlude the possibility of recombination between paralog clusters within an organism and copying of individual microRNAs from one cluster to another. The available data do not contain any evidence that such processes might play a role.

3 Results

Fig. 1 summarized the results from `blast` searches: Mammals have 3 paralogs of the **mir17** cluster, teleost fishes contain 4 copies. Complete or nearly complete sequences have been obtained for *Homo sapiens* (Hs), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), *Takifugu rubripes* (Tr), *Tetraodon nigroviridis* (Tn), and *Danio rerio* (Dr). In addition homologs of partial clusters have been identified in the dog *Canis canis* (Cc) and frog *Xenopus tropicalis* (Xt) genomes. The 115 microRNA sequences and their locations are listed in the electronic supplemental material [2].

Fig. 3 provides an overview of the **mir17** sequences from different organisms. We can identify groups of pre-miRNA whose order in the **mir17** clusters are preserved: 17/106, 18, 19, 20/93, 19b, and 92/25. Only the 92/25 group has known invertebrate homologs, namely *mir-92a* and *mir-92b* in *Drosphila melanogaster* and *mir-235* in *Caenorhabditis elegans*. In (28, Fig.3) the *Caenorhabditis elegans* microRNA *cel-mir-254* is suggested as relative of the 19/19b group, the sequences are too distant to be certain about their homology, however.

Both the mosquito *Anopheles gambiae* and the nematode *Caenorhabditis briggsae* have distant relatives of *miR-92*, see electronic supplement for details. The mosquito genome also contains two paralogs that are separated by about 20kb.

We were not able to find unambiguous homologs of members of the **mir17** family in the *Ciona intestinalis* and *Strongylocentrotus purpuratus* genomes. In the light of the extreme conservation of many of the members of the **mir17**

[2] URL: `http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/04-002/`

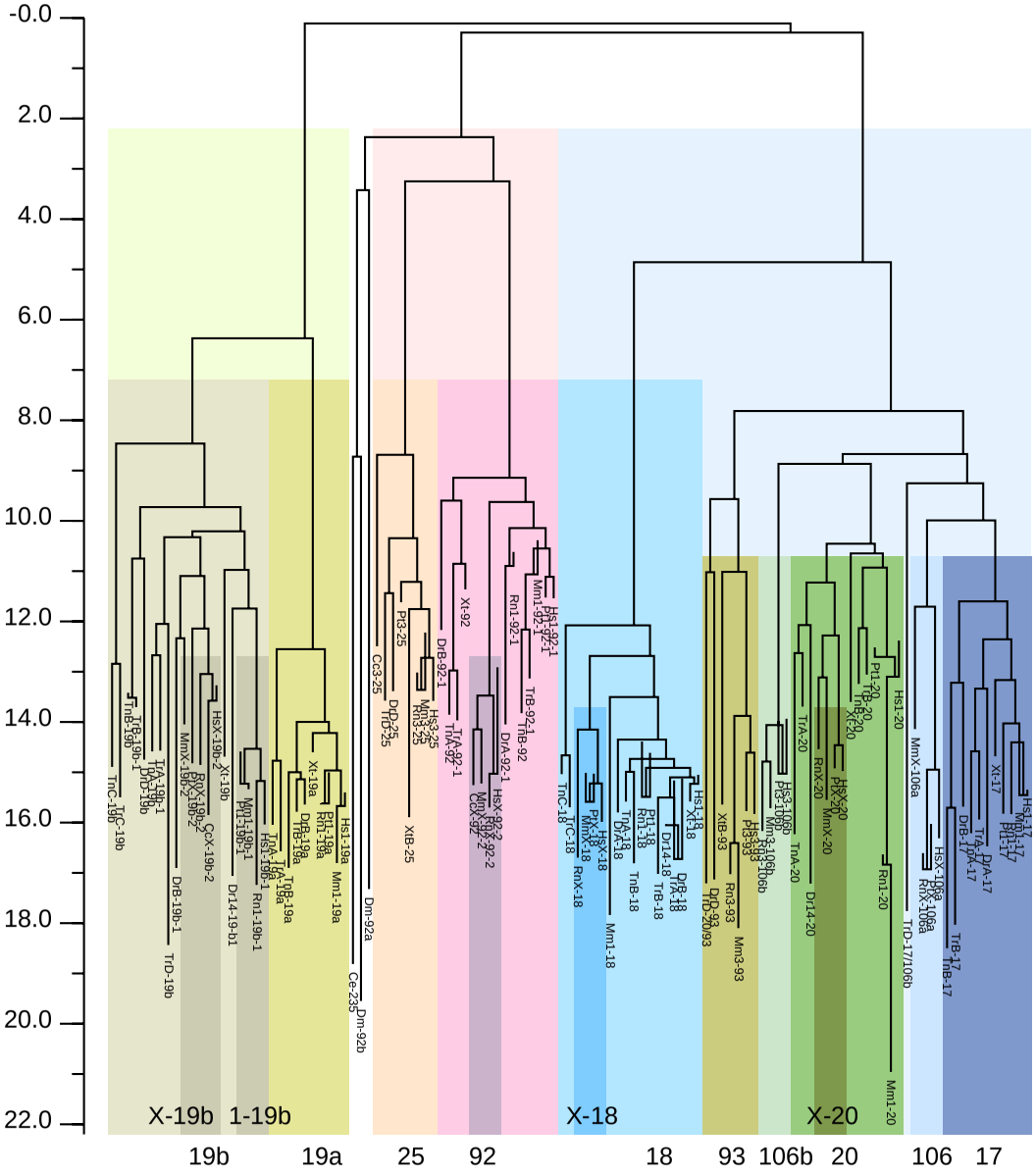
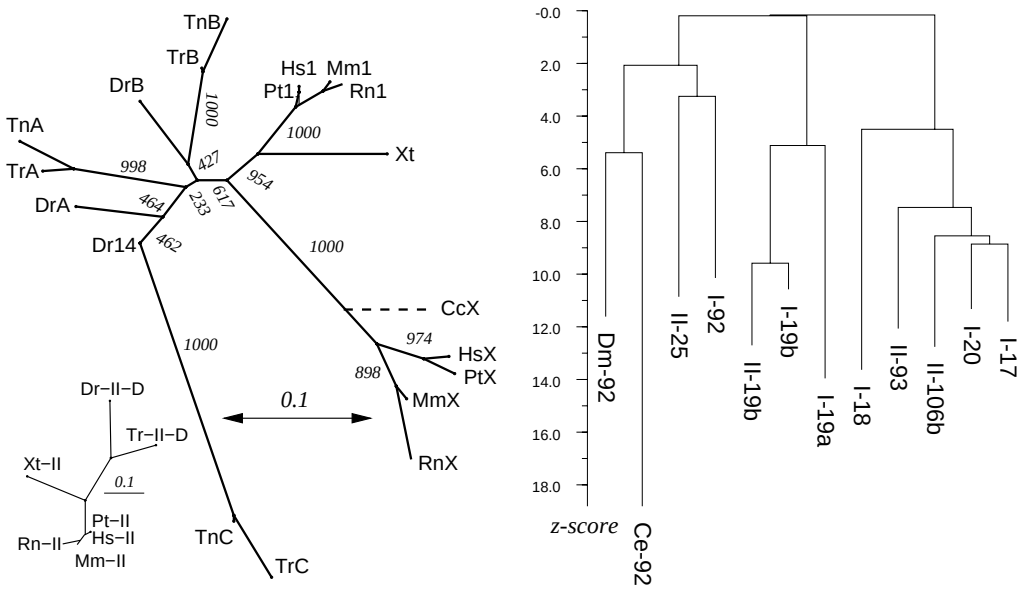


Fig. 3. WPGMA clustering of all available **mir17** sequences using the z-score defined in equ.(1) as similarity score. The main groups are highlighted. (Short negative branch length at the tips of the tree may appear due to the stochastic sampling of the distribution of the permutated sequences when the z-score similarity measures are computed for very similar sequences.)

family among vertebrates this suggests that their regulatory role(s) are predominantly vertebrate specific.

The **mir17** sequences from mammalian species belong to three clusters. One of them is homologous to the cluster containing the *hsa-mir-17* sequence, the second one, which contains *hsa-mir-106* is located at the X-chromosome in man, mouse, and rat. We label these two clusters as I-1 and I-X, respectively,



(a) (b)

Fig. 4. (a) Neighborjoining tree of the combined **mir17** type-I clusters and type-II clusters (inset). The fragment of the dog *CcX* cluster has been inserted based on a separate tree reconstruction leaving out the Dr14/pufferfish-C cluster since the absence of homologous sequences between the dog cluster fragment and some teleost clusters causes missing data in the distance matrix that cannot be dealt with properly by the `phylip` package. (b) WPGMA tree of the evolution of the paralog groups in the **mir17** family.

and collectively refer to them as type-I clusters. They are closely related and their component microRNAs can be easily aligned. The third mammalian cluster, II, is a distant relative containing *hsa-mir-93*. It cannot be aligned well with the type-I clusters; henceforth we will refer to this cluster as type-II. The frog genome also contains orthologs of both the I-1 and II cluster. A search from homologous microRNA sequences in the three teleost fish genomes revealed four microRNA clusters of which three, I-A, I-B, and I-C are homologous to the mammalian type-I clusters, and a single type-II cluster II-D that is clearly orthologous to the mammalian cluster II. To the extend that data are available, paralog clusters are located on different chromosomes.

A Neighborjoining tree, Fig, 4a summarizes the relationships of the type-I clusters. It follows that the various paralog type-I clusters share a common ancestor that pre-dates the divergence of the sarcopterygian and the actinopterygian lineages. The type-I clusters arose from this ancestor by means of duplications of the entire clusters. The type-II clusters simply reproduces the known species tree (see inset in Fig. 4a), indicating that they also originated very early in vertebrate evolution.

Clustering all 115 microRNA sequences of the **mir-17** family by means of the

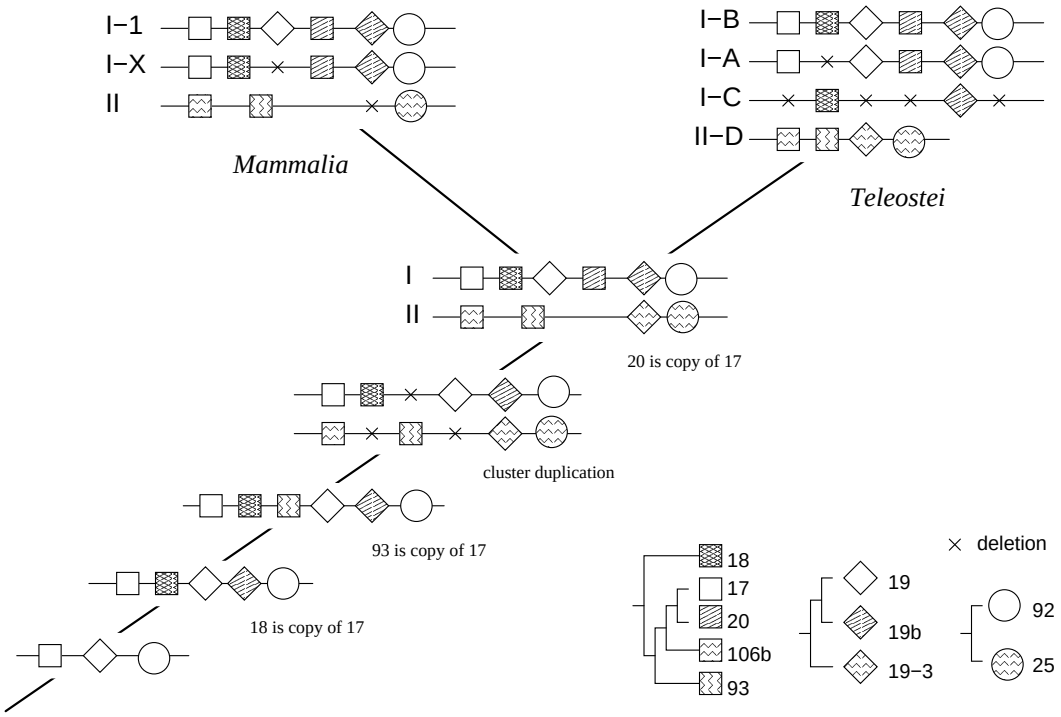


Fig. 5. A plausible scenario for the evolution of the **mir-17** family. For details see text.

z-score defined in equ.(1) shows that the microRNAs *miR-17*, *miR-106*, *miR-106b*, *mir-93*, *miR-20*, and *miR-18* are ancient paralogs. The same holds for *miR-19*, *miR-19b*, and *miR-II-19* as well as *miR-25* and *mir-92*, see Fig. 4b. It follows that the ancestral type-I and type-II clusters themselves arose through a complex sequence of tandem duplications, cluster duplications, as well as loss of microRNAs. A plausible scenario that is consistent with the history of the individual microRNAs in Figs. 3 and Fig. 4b is shown in Fig. 5.

4 Discussion

The microRNAs of the **mir17** clusters arose through a complex history of duplication and loss of individual members as well as duplications of entire clusters. It consists of three large groups of seemingly non-homologous microRNAs. Since we did not find homologs of *pre-mir-17* and *pre-mir-19* in invertebrates we suggest that these microRNAs are vertebrate innovations. In contrast, *pre-mir-92* is homologous to microRNAs from both *Drosophila melanogaster* and *Caenorhabditis elegans*.

We can only speculate how the ancestors of *pre-mir-17* and *pre-mir-19* arose. The stem-loop structures of microRNA precursors are among the most com-

mon local structures. In a recent simulation study (29) all locally stable RNA structures were computed using the `RNALfold` program (22) for the genome of *Caenorhabditis elegans*. More than 65000 subsequences were found in a single chromosome that are at least as similar to the *cel-mir-1* structure -than as] the other known microRNA precursors in terms of their tree-edit distance (30) to the *cel-mir-1* precursor structure. It seems plausible that a part of the transcript containing the ancestral *mir-92* may have accidentally formed a second hairpin structure that was processed and exported from the nucleus. If such a *de novo* precursor happens to be functional it could have been retained by stabilizing selection, otherwise it would have soon disappeared again as random drift destroys the *pre-mir*-like structural features. This might be a rather general mechanism for the origin of polycistronic microRNA transcripts. It could explain the large fraction of microRNAs that appear clustered (9; 10; 11).

The ancestral **mir17** cluster probably contained the ancestors of the *miR-17* group, the *miR-19* group, as well as the ancient *miR-92*. The first detectable duplication event was the split of *miR-18* from the *miR-17* group, preceeding the duplication of the entire cluster that created the ancestors of the 17/18/19/20/19b/92 family (type-I cluster) and the 106b/93/25 family (type-II cluster). WPGMA clustering, Fig. 4b, shows that the first duplication of the ancestral **mir17** cluster must have been preceeded by a complex series of tandem duplications: the cluster duplication cannot have happened before the split of *miR-19a* and *miR-19b* since the *miR-19* paralog in the type-II cluster clearly belongs to the *miR-19b* group. Similarly, *miR-106b* is orthologous to the ancestor of *miR-17* and *miR-20*, or possibly to *miR-20*. The origin of the *miR-93* group thus also pre-dated the cluster duplication. The large evolutionary distance between *miR-92* and *miR-25* is either the result of an enhanced rate of evolution in the *miR-92* group after the divergence of type-I and type-II clusters, or *miR-92* and *miR-25* might have been ancient tandem copies in the ancestral cluster. In the latter case a different paralog was lost in each of the two paralog clusters after the duplication of the ancestral cluster.

Since the genomic data of *Xenopus tropicalis* are still incomplete we cannot decide whether the two clusters, one type-I and one type-II cluster, represent the complete inventory or whether the frog genome also includes an ortholog on the mammalian X-chromosome cluster. The teleost sequences, unfortunately, are diverged so far that the duplication history of their three type-I clusters cannot be resolved unambiguously. It is very likely, however, that at least the teleost I-A and I-B clusters arose during the teleost-specific genome duplication (31). The homology of Dr14 with TnC/TrC is based on the cluster organization, the sequences are too short to provide an unambiguous signal. The tree in Fig. 4a suggests that the teleost I-A and I-C clusters are 1st order paralogs that subsequently lost some of their microRNAs, thereby resolving the redundancy that arose in the duplication event as predicted by the DDC model (32).

It is tempting to speculate that the expansion of the **mir17** family is linked to the origin of vertebrates and, more specifically, is associated with the two genome-wide duplications (33; 34) that also caused the quadruplicate *Hox*-clusters in vertebrates (35). A resolution of this question, however, will have to await data on microRNAs in lower vertebrates and amphioxus.

Acknowledgments

Useful comments by Claudia Fried and Sonja Prohaska are gratefully acknowledged. This work was supported in part by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Project No. P15893 (AT), and by the German *DFG* Bioinformatics Initiative BIZ-6/1-2 (PFS).

References

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

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

[3] Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.

[4] Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, **303**, 95–98.

[5] Schwarz, D. S. & Zamore, P. D. (2002). Why do miRNAs live in the miRNP. *Genes Dev.*, **16**, 1025–1031.

[6] Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C. & Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.*, **5**, R1.

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

[8] Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. (2003). Identification of *Drosophila* MicroRNA targets. *PLoS Biol.*, **1**, 397–409.

[9] Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**, 720–728.

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

[11] Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. (2003). Com-

putational identification of Drosophila microRNA genes. *Genome Biol.*, **4**, R42.

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

[13] Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Curr. Biol.*, **12**, 735–739.

[14] Houbaviy, H. B., Murray, M. F. & Sharp, P. A. (2003). Embryonic stem cell-specific microRNAs. *Dev. Cell*, **5**, 351–358.

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

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

[17] Sempere, L. F., Sokol, N. S., Dubrovsky, E. B., Berger, E. M. & Ambrosa, V. (2003). Temporal regulation of microrna expression in *Drosophila melanogaster* mediated by hormonal signals and Broad-Complex gene activity. *Devel. Biol.*, **259**, 9–18.

[18] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

[19] Thompson, J. D., Higgs, D. G. & Gibson, T. J. (1994). 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.

[20] Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucl. Acids Res.*, **31**, 3429–3431.

[21] Hofacker, I. L. & Stadler, P. F. (1999). Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comp. & Chem.*, **23**, 401–414.

[22] Hofacker, I. L., Priwitzer, B. & Stadler, P. F. (2004). Prediction of locally stable rna secondary structures for genome-wide surveys. *Bioinformatics*, **20**, 191–198.

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

[24] Hogeweg, P. & Hesper, B. (1984). Energy directed folding of RNA sequences. *Nucl. Acids Res.*, **12**, 67–74.

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

[26] Wilbur, W. J. & Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.

[27] Sokal, R. R. & Michner, C. D. (1958). A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.*, **38**, 1409–1438.

[28] Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B. & Bartel, D. P. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes Devel.*, **17**, 991–1008.

[29] Obernosterer, G., Tanzer, A., Hofacker, I. L., Stadler, P. F. & Höchsmann, M. (2003). Searching genomes for small temporal RNAs. In *The 18th Dubrovnik International Course & Conference on the Interfaces among Mathematics, Chemistry and Computer Science (MATH/CHEM/COMP/2003)*, Graovac, A., Pokrić, B. & Snrečki, V., eds., 65. Rudjer Bošković Institute, Zagreb, Croatia. Poster.

[30] Höchsmann, M., Töller, T., Giegerich, R. & Kurtz, S. (2003). Local similarity in RNA secondary structures. In *Proc of the Computational Systems Bioinformatics Conference, Stanford, CA, August 2003 (CSB 2003)*, 159–168.

[31] Taylor, J., Braasch, I., Frickey, T., Meyer, A. & Van De Peer, Y. (2003). Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.*, **13**, 382–390.

[32] Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.

[33] Holland, P. W. H., Garcia-Fernández, J., Williams, N. A. & Sidow, A. (1994). Gene duplication and the origins of vertebrate development. *Development*, **(Suppl.)**, 125–133.

[34] Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. & Lehrach, H. (2003). New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.*, **13**, 1056–1066.

[35] Bailey, W. J., Kim, J., Wagner, G. & Ruddle, F. H. (1997). Phylogenetic reconstruction of vertebrate Hox cluster duplications. *Mol. Biol. Evol.*, **14**, 843–853.

Web References

Rfam		http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml
Genomes		
Man	Hs	http://www.ensembl.org/Homo_sapiens/
Chimpanzee	Pt	http://pre.ensembl.org/Pan_troglodytes/
Mouse	Mm	http://pre.ensembl.org/Mus_musculus/
Rat	Rn	http://www.ensembl.org/Rattus_norvegicus/
Dog	Cc	http://www.ncbi.nlm.nih.gov/genome/seq/CfaBlast.html
Chicken	Gg	http://genome.wustl.edu/blast/client.pl
Frog	Xt	http://genome.jgi-psf.org/xenopus0/xenopus0.home.html
Tetraodon	Tn	http://www.genoscope.cns.fr/externe/tetraodon/
Fugu	Tr	http://www.ensembl.org/Fugu_rubripes/
Zebrafish	Dr	http://pre.ensembl.org/Danio_rerio/
Tunicate	Ci	http://genome.jgi-psf.org/ciona4/ciona4.home.html
Sea Urchin	Sp	http://www.hgsc.bcm.tmc.edu/projects/seaurchin/
Drosophia	Dm	http://www.ensembl.org/Drosophila_melanogaster/
Mosquito	Ag	http://www.ensembl.org/Anopheles_gambiae/
Nematode	Ce	http://www.ensembl.org/Caenorhabditis_elegans/
Nematode	Cb	http://www.ensembl.org/Caenorhabditis_briggsae/