JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 312B:164–170 (2009)

Simultaneous Expansions of MicroRNAs and Protein-Coding Genes by Gene/Genome Duplications in Early Vertebrates

XUN GU[1,2*], ZHIXI SU[2*], AND YONG HUANG[1]
[1]*Department of Genetics, Development and Cell Biology, Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa*
[2]*Institutes of Biomedical Sciences, School of Life Sciences, and Center for Evolutionary Biology, Fudan University, Shanghai, People's Republic of China*

ABSTRACT Does miRNAs underlie the origin of organismal complexity in vertebrates? The current controversy is focused on whether the inventory of vertebrate miRNAs can be explained by the classical two-round genome duplications. We estimate the age distribution of vertebrate miRNA duplication events, showing the evolutionary scenario that gene/genome duplications in the early stage of vertebrates may expand the protein-encoding genes and miRNAs simultaneously. We further speculate that genetically lying behind the evolution of vertebrate complexity may be the proteome doubling and alterations of the epigenetic (including miRNA) machinery. *J. Exp. Zool. (Mol. Dev. Evol.) 312B:164–170, 2009.*

How to cite this article: Gu X, Su Z, Huang Y. 2009. Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. J. Exp. Zool. (Mol. Dev. Evol.) 312B:164–170.

MicroRNA (miRNA) genes are small (~22 nucleotides) noncoding RNAs that have the ability to repress the expression of target genes posttranscriptionally (Bartel, 2004; Alvarez-Garcia and Miska, 2005; Chen and Rajewsky 2007). Since the discovery of miRNA genes, lin-4 (Lee et al., '93; Wightman et al., '93) and let-7 (Reinhart et al., 2000; Slack et al., 2000), a growing number of miRNAs have been found (Berezikov et al., 2006; Griffiths-Jones et al., 2006). There are two views about the evolutionary origin of new miRNAs: (i) the de novo inventory model for the ongoing, active process of miRNA innovation from random sequences (Allen et al., 2004; Lall et al., 2006; Rajagopalan et al., 2006; Chen and Rajewsky, 2007; Heimberg et al., 2008), and (ii) the duplication model (Tanzer and Stadler, 2004; Hertel et al., 2006; Maher et al., 2006; Sempere et al., 2006; Prochnik et al., 2007; Huang and Gu, 2007) for the expansion of existed miRNA gene families. In this article, we address this issue especially for vertebrate miRNA genes by estimating the age duplication of miRNA duplications to test the hypothesis of simultaneous expansion of miRNAs and proteins in the early stage of vertebrates.

ORIGINS OF VERTEBRATE miRNAs AND GENOME-WIDE DUPLICATIONS

Hertel et al. (2006) observed a wave of innovations (56 miRNAs) maps to the branch leading to the vertebrates, and an expansion of miRNAs owing to gene duplications (24 nonlocal and 11 local miRNA duplication events) in the early vertebrates. In spite of the relatively

Grant sponsors: Pioneer Computational Biology Fellowship; National Natural Science Foundation of China; Grant number: 30700140.
*Correspondence to: Xun Gu, Department of Genetics, Development and Cell Biology, Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011.
E-mail: xgu@iastate.edu (X.G.); zxsu@fudan.edu.cn (Z.S.)
Received 22 September 2008; Revised 3 November 2008; Accepted 20 November 2008
Published online 12 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21273

small dataset, Hertel et al. (2006) suggested that the two-round genome duplication(s) (Ohno '70) may play an important role for miRNA diversity. On the other hand, Heimberg et al. (2008) argued that genome duplications could not account for those identified 41 miRNA families that were found in lampreys, but not in either ascidians or amphioxus. They thus further claimed that lying behind the origin of vertebrate complexity is the dramatic expansion of miRNAs, rather than an increase in the protein-encoding inventory caused by genome duplication events.

Noticeably, both studies relied on another highly controversial issue, that is, whether the classical two-round (2R) genome duplication hypothesis (Ohno, '70; Holland et al., '94) that the first-round genome duplication occurred at the origin of vertebrates and the second-round at the origin of gnathostomes is solid. Numerous studies, especially after the first draft of human genome sequence (Lander et al., 2001; Venter et al., 2001), have reached very different conclusions and generated a hot debate (e.g., Hughes, '99; Wang and Gu, 2000; Gu et al., 2002; Larhammar et al., 2002; McLysaght et al., 2002; Friedman and Hughes, 2003). Together, one may reasonably argue a number of genomic duplication events including whole-genome duplications may occur in a broad time period of early vertebrates. However, there is indeed no strong evidence to support some detailed predictions for when, where and how many genome duplications had occurred. For many developmental gene families that have been thought to be the evidence for the 2R model, careful phylogenetic analyses yielded to a rather complicated pattern: The occurrence of these duplication events ranged from the time prior to amphioxus–vertebrate split to the time prior to the teleost–tetrapod split (Escriva et al., 2002; Panopoulou et al., 2003; Leveugle et al., 2004). Recently literatures, however, largely seems to favor the 2R genome duplication hypothesis, e.g., from the larger-scale synteny analysis (Sundström et al., 2008).

We (Gu et al., 2002) hold a more subtle view, that is, there existed an evolutionary time window around the origin of vertebrate or gnathostomes, during which multiple genome-wide duplications might have significant contributions to the follow-up evolution of the organismal complexity since the origin of vertebrates. In addition, a flux of small-scale duplications events should not be neglected.

MOLECULAR PHYLOGENY OF miRNA FAMILIES

In short, the main issue of miRNA inventory becomes whether most miRNA families identified in current vertebrate genomes were generated by the same genomic processes. We propose estimating the age distribution of duplicated miRNAs in vertebrates from the phylogenetic analysis of miRNA gene families. The result may reveal a concise, genome-wide pattern, and is useful to examine whether it has a similar pattern to protein-encoding duplicate genes. Here we address this issue. The sequences of animal miRNAs were retrieved from miRBase, Release 9.2 (Griffiths-Jones et al., 2006) (http://www.sanger.ac.uk/Software/Rfam/mirna/), including six model genomes: human, mouse, chicken, zebrafish, fruit fly and worm. We used a recently developed computational method called phylogeny-bootstrap-clustering (PBC) (Huang and Gu, 2007) to group homologous miRNAs into gene families. The PBC procedure combines sequence similarity, phylogenetic inference and bootstrap testing into the procedure of clustering, which has considerably improved the reliability of miRNA classifications. The outcome of PBC procedure results in the multiple-alignment of an miRNA family by the software Clustal X (Huang and Gu, 2007). With a few exceptions, the final classification of miRNA families is virtually the same dataset used by Hertel et al. (2006) and Heimberg et al. (2008).

We have noticed several technical issues that are nontrivial in the phylogenetic reconstruction of miRNAs, largely because of the short sequence length and the sophisticated nucleotide substitution pattern of RNA secondary structure. Extensive computer simulation studies have shown that, given such short sequence length, the performance of phylogenetic inference based on a relatively simple model may have a better statistical performance than the performance based on any sophisticated model (see Nei and Kumar, 2000 for a comprehensive review). Thus, following these recommendations, we adopted the neighbor-joining (NJ) method with simple Jukes–Cantor distance model to infer the miRNA tree from the precursor (hairpin) sequences (average 87 nucleotides in length), which was implemented in the software MEGA3.0 (http://megasoftware.net). Other methods, such as maximum parsimony (as implemented in PAUP4.0), and maximum likelihood (as implemented in PHYLIP) were used to examine whether the uncertainty of inferred

phylogeny can alter our main result. Bootstrapping was used to evaluate the statistical reliability of the inferred phylogeny.

MOLECULAR DATING OF miRNA DUPLICATION EVENTS

We adopted the split-time of primate–rodent (80 mya, million years ago), mammal–bird (310 mya), tetrapod–teleost (430 mya) and vertebrate–*Drosophila* splits (830 mya) as multiple calibrations; see Gu et al. (2002) for a detailed description. To estimate the duplication time under the inferred phylogeny, the method of nearest-neighbor clock involved two steps (Gu et al., 2002), which involves the following two steps:

First, one can determine the phylogenetic interval of a duplicate event (X) by the closest speciation, represented by (A, B) such that $A<X<B$. For example, ''(teleost–tetrapod, *Drosophila*–vertebrate)''means that the duplication event (X) occurred before the teleost–tetrapod split (A) but after the *Drosophila*–vertebrate split (B). In some cases, the available information can be only one sided, such as ''(teleost–tetrapod),'' meaning that the duplication event (X) occurred before the teleost–tetrapod split (A). Apparently, the phylogenetic interval is only based on the topology. Hence, as long as the phylogeny is largely correct, it is robust against the nonconstant evolutionary rate (violation of molecular clock), a major error resource in our study. Second, we estimated the age of the duplication event from the (fossil record) calibrations of the phylogenetic interval, taking an average if there were multiple fossil-record data points. The second step can be done under the option of linearlized tree in MEGA.

An example: We use the miR-10 gene family to illustrate our approach. From the inferred phylogeny (Fig. 1A), one can identify several phylogenetic intervals of duplication events. For instance, the duplication event of miR-10a/miR-10b is located after the *Drosophila*–vertebrate split but before the tetrapod–teleost split. Under our calibration system, it means that this duplication may occur sometime between 430 and 830 million years ago. Owing to miRNA data availability, in the case of the duplication between miR-100 and miR-99, we can only infer that it might occur after the *Drosophila*–vertebrate split but before the bird–mammal split, or 310–830 mya. Our analysis demonstrates that phylogenetic analysis of an miRNA family from multiple species can infer

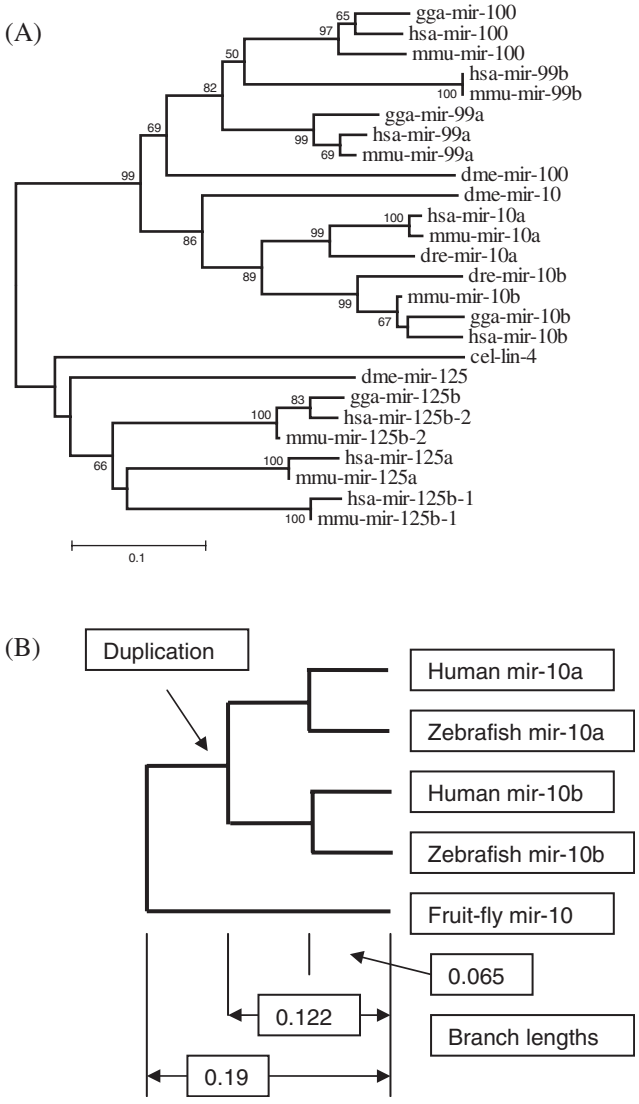

Fig. 1. (**A**) Phylogenetic tree of the miR-10 family inferred by the neighbor-joining method. Bootstrapping values less than 50% are not shown. hsa (human), mmu (mouse), gga (chicken), dre (zebrafish), dme (fruit fly) and cel (worm). (**B**) An example to show how to estimate the age of a gene duplication event.

the evolutionary interval for the timing of gene duplication. In practice, the evolutionary interval may differ among duplication events owing to the missing data in some species. In some cases one may solve it by adding more sequence data, but it may be difficult for miRNAs because they have experienced rapid gain and loss processes during the course of evolution. Consequently, it may not be uncommon that an miRNA gene has been lost in some species. In spite of that it unlikely causes any problem when a single miRNA family is focused, a genome-wide survey over multiple

miRNA families would raise some difficulties for drawing an overview about the miRNA evolution.

The second step for estimating the age of gene duplication provides a solution that is feasible in practical miRNA analysis. To avoid the effect of nonconstant rates among different miRNA members, we isolated the subtree of miR-10a/miR-10b from the whole phylogeny and then drew a linearlized tree (Fig. 1B). We measured the mean branch length (estimated by the Juke–Cantor model) of the *Drosophila*–vertebrate split is 0.19 ± 0.032, that of the tetrapod–teleost split is 0.065 ± 0.019 and that of the duplication event is 0.122 ± 0.016; the standard error for each branch length was estimated by the basic delta method when the inferred phylogeny was known (Nei and Kumar, 2000). Then, after assuming a local molecular clock between these two speciation events, the method of nearest-neighbor clock estimate the age of the miR-10a/10b duplication event (T) as follows:

$$T \approx 430 + (830 - 430) \times (0.122 - 0.065)/(0.19 - 0.065) \approx 612\,\text{mya}.$$

As the first term (430 mya) on the right hand of above formula represents the split time of ;tetrapod–teleost that is considered as a fixed number (though may be biased), the sampling variance is apparently from the second term. By the delta method, we approximately have $T \approx 430 + (182 \pm 57)\,\text{mya} = 612 \pm 57\,\text{mya}$. Instead, if we use a global molecular clock, we have $T \approx 830 \times 0.122/0.19 \approx 533 \pm 147\,\text{mya}$ (using the *Drosophila*–vertebrate split as a calibration), or $T \approx 430 \times 0.122/0.065 \approx 807 \pm 441\,\text{mya}$ (using the tetrapod–teleost split as a calibration).

This case study yields to the following conclusions. First, the method of nearest-neighbor clock can reduce the sampling error of the estimation considerably. Second, the method of nearest-neighbor clock is insensitive to whether the molecular timing of speciation is consistent to the fossil record (in this case, the tetrapod–teleost split versus the *Drosophila*–vertebrate split). Third, the method of nearest-neighbor clock is indeed sensitive to the criteria (molecular timing or fossil records) for the speciation used for the calibration. Hence, the interpretation of these age estimations of duplication events should be under the appropriate context (also see below).

SIMULTANEOUS EXPANSION OF DUPLICATED miRNAs AND PROTEIN-CODING GENES

Using the recently developed PBC method (Huang and Gu, 2007), we grouped conserved human miRNAs into 82 miRNA families, each of which contains at least two human miRNAs, as well as homologous miRNAs from other animal model organisms. We then reconstructed the phylogeny of each miRNA family from the precursor (hairpin) sequences (average 87 nucleotides in length). Based on the inferred miRNA family tree, the evolutionary age of each duplication event was estimated by the procedure as illustrated in Figure 1.

Together, we have dated 136 miRNA duplication events and obtained the *age distribution of vertebrate miRNA duplication events* (Fig. 2). Obviously, as the number of miRNAs has been increased gradually during the metazoan evolution, a rapid expansion of miRNAs appeared in the early stage of vertebrates. Interestingly, the age distribution of duplicated vertebrate miRNAs is similar to the well-known observation that vertebrate protein-encoding genes have been expanded in the early history of vertebrates (Gu et al., 2002). However, there is an important exception: Although we observed a recent increase in protein-encoding genes after the mammalian radiation, but no rapid miRNA expansion in the same evolutionary period. As tandem or segmental duplications were dominant in the mammalian

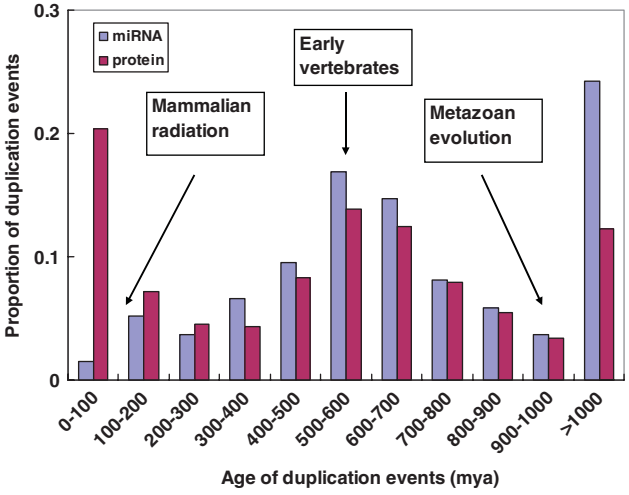

Fig. 2. Age distribution of miRNA duplication events with a bin of 100 million years, showing the hypothesis of simultaneous expansion of miRNA genes and proteins in the early stage of vertebrates.

TABLE 1. The mean duplication time of co-localized human miRNA families

	Co-localized miRNAs	Location	Mean times of duplication events
1	let-7a-2; miR-125b-1, miR-100	Chr11	Chr11–Chr5: 604 mya
	let-7e; miR-99b, miR-125	Chr5	Chr11–Chr21: 586 mya
	let-7c; miR-99a, miR-125b-2	Chr21	Chr5–Chr21: 567 mya
2	miR-93, miR-106b; miR-25	Chr7	Chr7–Chr13: 841 mya
	miR-17, miR-19a/19b-1; miR-92-1	Chr13	Chr7–ChrX: 591 mya
	miR106a; miR-19b-2; miR92-2	ChrX	Chr13–ChrX: 617 mya
3	miR-206; miR-133b	Chr6	Chr6–Chr18: 465 mya
	miR-1-2; miR-133a-1	Chr18	Chr6–Chr20: 465 mya
	miR-1-1; miR 133a-2	Chr20	Chr18–Chr20: 581 mya
4	miR-23b; miR-27b; miR-24-1	Chr9	Chr9–Chr19: 669 mya
	miR-23a; miR-27a; miR-24-2	Chr19	
5	miR-215; miR-194-1	Chr1	Chr1–Chr11: 532 mya
	miR-192; miR-194-2	Chr11	

radiation, it is possible that origin of new miRNAs may mainly through the de novo inventory in the case of no large-scale genome-wide duplication events.

Some human miRNA families are co-localized in the chromosome region (Table 1). A classical example is let-7 miRNA family, which was co-expanded with the HOX gene clusters (Tanzer and Stadler, 2004). We estimated the mean duplication time of each co-localized human miRNA families, supporting the hypothesis that these co-localized miRNAs may be the relics of (large-scale) genomic duplication events in the early stage of vertebrates.

Our analysis may have several caveats. The first one is the possibility of misclassification of miRNA gene families. We found some ambiguities in detecting ancient duplications (before the split between vertebrates and worm): In these cases the average sequence divergence between two ancient miRNA subfamilies is usually close to the dawn of random match-up. As we focused on the evolution since the origin of vertebrates rather at the origin of early animals, excluding these cases would not affect our main conclusion.

Secondly, it has been well known that phylogenetic inference from less than 100 nucleotide sites could be subject to a high level of statistical uncertainty. To reduce the random sampling errors, we used a simple tree-making method, with the risk of the increase in topological inference errors (Nei and Kumar, 2000). Nevertheless, our previous study (Gu et al., 2002) showed that the uncertainty in tree making owing to various tree-making methods or low statistical resolution might have some margin effects on the age distribution of duplication events. At any rate, we have conducted substantial case studies (miRNA families with three or more duplication events) to examine this issue, and found it is indeed the case. For instance, we used NJ, MP and ML methods to study the miR-10/miR-100 gene family. In spite of some minor tree topology differences and branch lengths, the general picture about the evolutionary pattern as shown in Fig. 1, remains virtually unchangeable (not shown).

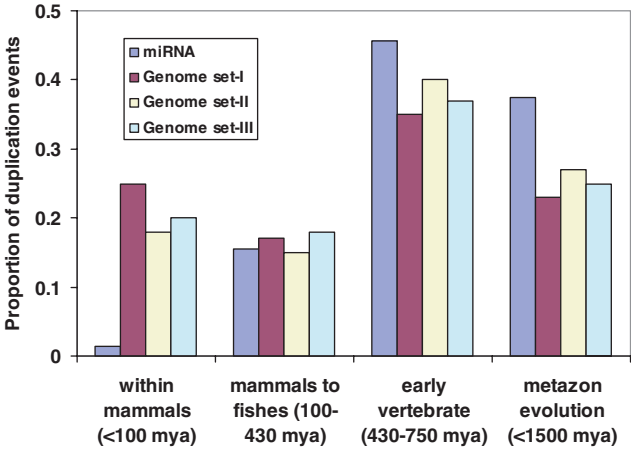

Fig. 3. Age distributions of protein-encoding genes by single calibration: genome set-I: the human–mouse (80 mya); II: the human–bird split (330 mya) and human–fish (430 mya).

Third, it should be noted that, in this study, the age distribution of miRNA duplications should be interpreted under the context of molecular timing, which may not completely compatible with the corresponding fossil records. Therefore, it is important to examine whether the simultaneous expansion hypothesis between miRNAs and proteins is sensitive to the specific calibration we used. We have carefully addressed this issue to validate the hypothesis. For instance, Figure 3 shows that the age distributions of protein duplications under

single calibration (the mammal–birds, the mammal–bird or the tetrapod–teleost split). Obviously, our main result is fairly robust.

OUTLOOK REMARKS: GENOME-WIDE DUPLICATIONS NOT ONLY DOUBLE PROTEIN-ENCODING GENES BUT ALSO MANY OTHERS

In this article, we hypothesize that the expansion of miRNA families that are currently conserved in mammals could be through many genomic duplication events in the early stage of vertebrates. Most likely, these same genomic processes also underlie the rapid expansion of protein-encoding genes. Our analysis is generally compatible to Hertel et al. (2006) and Heimberg et al. (2008). Our interpretation further indicates that genome-wide duplication(s) not only doubled the proteome, but also many other functionally important elements including miRNAs. In addition, our approach may provide a clear overview to summarize, at the genome level, the evolutionary pattern of miRNAs through gene/genome duplications.

One may put this issue further. Genome-wide duplications may have also initiated rapid and massive genomic alterations (Ohno, '70; Gu et al., 2005). Consequently, epigenetic status, such as methylation, imprinting or miRNA targeting, may differ between duplicated genes, a process that could occur much earlier than nucleotide mutations in the regulatory region (Rodin and Riggs, 2003; Rapp and Wendel, 2005; Rodin et al., 2005; Zheng, 2008). As a gene-wide duplication event is to simultaneously double the genetic elements related to the epigenetic machinery and protein-encoding genes, the co-evolution facilitated by these major genomic alterations in the early vertebrates has been an interesting issue but highly controversial. For instance, Heimberg et al. (2008) identified a number of novel vertebrate miRNA families that were found in lampreys, but not in ascidians or amphioxus. Though the authors did not favor the role of genome-wide duplications in the early-vertebrate miRNA expansion, the possibility that these miRNAs or developmental genes may have been lost in ascidians or amphioxus, which cannot be completely ruled out, similar to a number of developmentally important genes (Escriva et al., 2002; Panopoulou et al., 2003; Leveugle et al., 2004). Finally, we comment that all the aspects of organismal complexity in vertebrates may not have to appear instantly at the time of its origin. Some morphological diversity preserved in the fossil records may be driven by various ecological factors (Donoghue and Purnell, 2005). In short, we speculate that the rapid simultaneous increase of both coding and miRNA genes in the early stage of vertebrate should have some significant contributions to the evolution of vertebrate characters.

ACKNOWLEDGMENTS

We are grateful to Ying Zheng for the early data collection and analysis. This work was partly supported by the Pioneer Computational Biology Fellowship (to Y.H.) and the National Natural Science Foundation of China (Grant 30700140 to Z.S.).

REFERENCES

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet 36:1282–1290.

Alvarez-Garcia I, Miska EA. 2005. MicroRNA functions in animal development and human disease. Development 132:4653–4662.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297.

Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RH. 2006. Diversity of microRNAs in human and chimpanzee brain. Nat Genet 38:1375–1377.

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 8:93–101.

Donoghue PCJ, Purnell MA. 2005. Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol 20:312–319.

Escriva H, Manzon L, Youson J, Laudet V. 2002. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. Mol Biol Evol 19:1440–1450.

Friedman R, Hughes AL. 2003. The temporal distribution of gene duplication events in a set of highly conserved human gene families. Mol Biol Evol 20:154–161.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34:D140–D144.

Gu X, Wang Y, Gu J. 2002. Age-distribution of human gene families showing equal roles of large and small-scale duplications in vertebrate evolution. Nat Genet 31:205–209.

Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc Natl Acad Sci 102:707–712.

Heimberg AM, Sempere LF, Moy VN, Donoghue P, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci 105:2946–2950.

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF, Students of Bioinformatics Computer Labs 2004 and 2005. 2006. The expansion of the metazoan miRNA repertoire. BMC Genomics 7:25.

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. Development Suppl:125–133.

Huang Y, Gu X. 2007. A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs. BMC Genomics 8:66.

Hughes AL. 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48:565–576.

Lall S, Grün D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P, Kao HL, Gunsalus KC, Pachter L, Piano F, Rajewsky N. 2006. A genome-wide map of conserved MicroRNA targets in *C. elegans*. Curr Biol 16:460–471.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC et al. 2001 Initial sequencing and analysis of the human genome. Nature 409:860–921.

Larhammar D, Lundin LG, Hallbook F. 2002. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. Genome Res 12:1910–1920.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854.

Leveugle M, Prat K, Popovici C, Birnbaum D, Coulier F. 2004 Phylogenetic analysis of *Ciona intestinalis* gene superfamilies supports the hypothesis of successive gene expansions. J Mol Evol 58:168–181.

Maher C, Stein L, Ware D. 2006. Evolution of *Arabidopsis* microRNA familes through duplication events. Genome Res 16:510–519.

McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. Nat Genet 31:200–204.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press, Inc.

Ohno S. 1970. Evolution by gene duplication. London: George Allen and Unwin.

Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ et al. 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.

Prochnik SE, Rokhsar DS, Aboobaker A. 2007. Evidence for a microRNA expansion in the bilaterian ancestor. Dev Genes Evol 217:73–77.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–3425.

Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. New Phytol 168:81–91.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403:901–906.

Rodin SN, Riggs AD. 2003. Epigenetic silencing may aid evolution by gene duplication. J Mol Evol 56:718–729.

Rodin SN, Parkhomchuk DV, Rodin AS, Holmquist GP, Riggs AD. 2005. Repositioning-dependent fate of duplicate genes. DNA Cell Biol 24:529–542.

Sempere LF, Cole CN, McPeek MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zool B Mol Dev Evol 306:575–588.

Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G. 2000. The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell 5:659–669.

Sundström G, Larsson TA, Larhammar D. 2008. Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters. BMC Evol Biol 19, 8:254.

Tanzer A, Stadler PF. 2004. Molecular evolution of a microRNA cluster. J Mol Biol 339:327–335.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ et al. 2001. The sequence of the human genome. Science 291:1304–1351.

Wang Y, Gu X. 2000. Evolutionary patterns of gene families generated in the early stage of vertebrates. J Mol Evol 51:88–96.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75: 855–862.

Zheng D. 2008. Asymmetric histone modifications between the original and derived loci of human segmental duplications. Genome Biol 9:R105.