Dev Genes Evol (2007) 217:73–77
DOI 10.1007/s00427-006-0116-1

SEQUENCE CORNER

Evidence for a microRNA expansion in the bilaterian ancestor

Simon E. Prochnik · Daniel S. Rokhsar · A. Aziz Aboobaker

Received: 24 July 2006 / Accepted: 2 October 2006 / Published online: 14 November 2006

Abstract Understanding how animal complexity has arisen and identifying the key genetic components of this process is a central goal of evolutionary developmental biology. The discovery of microRNAs (miRNAs) as key regulators of development has identified a new set of candidates for this role. microRNAs are small noncoding RNAs that regulate tissue-specific or temporal gene expression through base pairing with target mRNAs. The full extent of the evolutionary distribution of miRNAs is being revealed as more genomes are scrutinized. To explore the evolutionary origins of metazoan miRNAs, we searched the genomes of diverse animals occupying key phylogenetic positions for homologs of experimentally verified human, fly, and worm miRNAs. We identify 30 miRNAs conserved across bilaterians, almost double the previous estimate. We hypothesize that this larger than previously realized core set of miRNAs was already present in the ancestor of all Bilateria and likely had key roles in allowing the evolution of diverse specialist cell types, tissues, and complex morphology. In agreement with this hypothesis, we found only three, conserved miRNA families in the genome of the sea anemone *Nematostella vectensis* and no convincing family members in the genome of the demosponge *Reniera* sp. The dramatic expansion of the miRNA repertoire in bilaterians relative to sponges and cnidarians suggests that increased miRNA-mediated gene regulation accompanied the emergence of triploblastic organ-containing body plans.

Keywords MicroRNA · Bilateria · Post-transcriptional regulation

Communicated by N. Satoh

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00427-006-0116-1 and is accessible for authorized users.

S. E. Prochnik · D. S. Rokhsar
DOE Joint Genome Institute,
2800 Mitchell Avenue,
Walnut Creek, CA 94598, USA

D. S. Rokhsar · A. A. Aboobaker
Center for Integrative Genomics,
Department of Molecular and Cell Biology,
University of California,
Berkeley, CA 94720, USA

A. A. Aboobaker (✉)
Institute of Genetics,
University of Nottingham, Queens Medical Centre,
Nottingham NG7 2UH, UK
e-mail: Aziz.Aboobaker@nottingham.ac.uk

Introduction

MicroRNAs (miRNAs) play key roles in animal development, particularly in the establishment, temporal control, and maintenance of cell-, tissue-, and organ-specific identity (Wienholds and Plasterk 2005). The growing appreciation for the role of miRNAs and other noncoding RNAs in regulating gene expression suggests they might be key players in the evolution of developmental processes, particularly in facilitating the establishment of more complex transcriptional control mechanisms. To assess their importance in the evolution of development across animals, we would like an accurate picture of how many miRNAs are conserved across different groups. In particular, we wanted to know the core set of miRNAs shared among the Bilateria as this would provide an insight into how many shared features of bilaterian development

already involved miRNA regulation before the evolution of the extant animal phyla. So far, invertebrate and vertebrate model organisms have been experimentally searched for miRNAs by sequencing from small RNA libraries and confirming the presence of short 21–23mers by Northern blot analyses (Wienholds and Plasterk 2005). It has then been straightforward to search for these experimentally verified miRNAs in other genomes (Hertel et al. 2006; Sempere et al. 2006).

The developmental "toolkit" of protein-coding genes is largely the same across animals (Technau et al. 2005), with secondary losses in some animals occurring frequently (Kortschak et al. 2003; Aboobaker and Blaxter 2003). This means that increasing complexity and the number of developmental genes are not necessarily correlated among animal phyla. We therefore need alternate explanations for the large increase in morphological complexity shared by some protostome and deuterostome phyla in comparison to nonbilaterians, such as cnidarians and sponges. miRNA-mediated regulation represents one possible novel source for increased complexity. For example, the conservation of the founding miRNA, *let-7*, across bilaterians, but not in nonbilaterian taxa, provides a single data point linking miRNAs to the transition from early animals to the more complex triploblastic bilaterians (Pasquinelli et al. 2003). If the evolutionary distribution of *let-7* is typical of other conserved bilaterian miRNAs, this group of regulatory genes would be implicated in the evolution of regulatory complexity and, as a result, morphological complexity. To test this hypothesis, we characterized the phylogenetic distribution and conservation of miRNAs across diverse metazoan genomes. We sample a wider distribution of animal taxa than previously available and find a greater number of miRNAs conserved across the Bilateria than previously thought (Hertel et al. 2006). Our analysis is an essential first step in understanding the role of miRNAs in the diversification of bilaterians and assessing their role in the evolution of complexity.

Materials and methods

Search algorithm

Five hundred forty-nine mature miRNA sequences for cloned *Homo sapiens*, *Caenorhabditis elegans*, and *Drosophila melanogaster* miRNAs (355, 114, and 80 sequences, respectively) were downloaded from miRBase (Griffiths-Jones et al. 2006). Our search algorithm consisted of running WU-blastn (9) with S=15, E=1,000, W=7, M=1, N=−1, Q=3, and R=2 with each mature miRNA sequence against genomes of *Lottia gigantea* (J. Chapman, assembly of NCBI Trace Archive data, personal communication), *Strongylocentrotus purpuratus* (Baylor College of Medicine Human Genome Sequencing Center; Spur 0.5, 15 April 2005 http://www.hgsc.bcm.tmc.edu/projects/seaurchin/), *Tribolium castaneum* (Baylor College of Medicine Human Genome Sequencing Center v2, Tcas 20051011-genome, http://www.hgsc.bcm.tmc.edu/projects/tribolium/), *Ciona intestinalis* (JGI v2.0 unmasked, Oct 2002), *Nematostella vectensis* (JGI draft assembly, Sept 30, 2005), *Reniera* sp. (J. Chapman, assembly of NCBI Trace Archive data, pers. comm.), *Fugu rubripes* (JGI v4.0 unmasked, Oct 2004), with *D. melanogaster* (FlyBase assembly Release 4.0, http://flybase.net/), and *C. elegans* (WormBase assemblyWS100, May 10 2003 http://wormbase.org) providing positive controls. Hits were discarded if the seed site of the query (defined in this study as bases 1–6 or 2–7 of the mature miRNA sequence) was not a perfect match or four or more nucleotides (nt) shorter than the query. Remaining hits were extended at least 10 and 50 nt and up to 20 and 80 nt at each end within the genome sequence and folds predicted with mfold 3.1 (Zuker 2003) with T=25, filtered for dG<−25 kcal/mol (see below) and displayed on a website (http://www.evodevoteam.org/mirna/basal_mirna). Hits were manually inspected for structural statistics that lie within the distribution of statistics predicted by mfold for cloned miRNAs (data not shown) and previously published requirements for miRNAs (Ambros et al. 2003) including: dG<−25 kcal/mol, fewer than 7 nt from the mature miRNA lying in even-sided internal loops and fewer than 6 nt lying in uneven-sided internal loops. Genomic coordinates for hits were recorded allowing clustering of miRNA genes to be trivially observed. miRNAs occurring in more than one search genome (and therefore very unlikely to be false positives, $p<0.01$) were counted as hits and grouped in families with identical seed sites. Most miRNAs occurring in only one search genome were within either the vertebrate or insect lineages and are not presented, as they are probably restricted to these groups. Other hits to single genomes were all with human/vertebrate-specific miRNAs. Where there structural statistics are acceptable, they are included in the supplementary Table S1 but are presented as conserved miRNAs as they are likely to be false positives (see below).

Randomized control search

To assess the false positive rate of the algorithm, over 40 conserved miRNA sequences were each randomized ten times and the search algorithm run as before. In the resulting set of folding predictions, we found no cases of miRNAs with acceptable structural statistics conserved in more than one subject organism. The false positive rate was hence calculated to be <1%.

Results and discussion

To explore the evolution of the miRNA repertoire in animals, we devised an algorithm for finding homologs of known miRNAs. Our search method exploits conservation of the 5′ seed site and 3′ sequence in mature ~21 nucleotide miRNAs (Brennecke et al. 2005) as well as conservation of the hairpin structure formed by pre-miRNAs. We searched for homology to all experimentally verified fly, nematode, and human mature microRNAs (80, 114, and 355 miRNAs, respectively) from the Sanger Institute miRBase (Griffiths-Jones et al. 2006). We searched the genomes of the teleost fish *F. rubripes*, beetle *T. castaneum*, limpet *L. gigantea*, sea squirt *C. intestinalis*, sea urchin *S. purpuratus*, sea anemone *N. vectensis*, and demosponge *Reniera* sp. We estimate that the shotgun sequences for these genomes are well over 90% complete, based on their content of conserved eukaryotic proteins and ESTs from each animal (data not shown). To improve confidence in our predictions, and to focus on ancient families with correspondingly broad phylogenetic distribution, we required miRNAs to be detected in at least two search species. Matches were extended and folded using the mfold program (Zuker 2003). We compared the resulting fold statistics including free energy (dG, kcal/mol), and the number balanced and unbalanced nucleotides in the miRNA to those of known miRNAs and only accepted candidates that fell within this distribution. Candidates at the upper limit of the distribution were discounted as being false positives unless they also showed very broad phylogenetic distribution and/or were clustered with other conserved miRNAs (see Table S1).

Using these criteria, we found that at least 30 miRNA families are conserved across the Bilateria and were therefore present in Urbilateria, the last common ancestor of all living bilaterians (Fig. 1 and http://www.evodevoteam.org/miRNA/basal_miRNA for full data set). Our search also uncovered many examples of lineage specific expansions in miRNA families and increased the number of miRNA gene clusters that are likely to be ancient (Table S1 and Fig 1). In some cases, clustering of miRNAs in the search species provides additional evidence that miRNA candidates are likely to be real, as well as insights into the evolutionary origins of other lineage-restricted miRNAs. As miRNAs are processed from large primary transcripts, miRNAs that are clustered are cotranscribed (Aboobaker et al. 2005). It is feasible that post-transcriptional regulation leads to unequal activity of clustered miRNAs (Thomson et al. 2006). Nonetheless, clustered miRNAs may be involved in regulating the same pathways and possibly overlapping sets of target mRNAs. This is especially likely to be the case when miRNA clusters are populated by the members of the same or a related miRNA family (see Table S1). These miRNAs may have evolved by duplication of a single ancestral miRNA rather then by de novo sequence evolution. miRNA duplication may be a simple evolutionary mechanism for increasing miRNA dosage in cells. Alternatively, the subtle changes in sequence of the mature miRNA that are observed within miRNA families may have

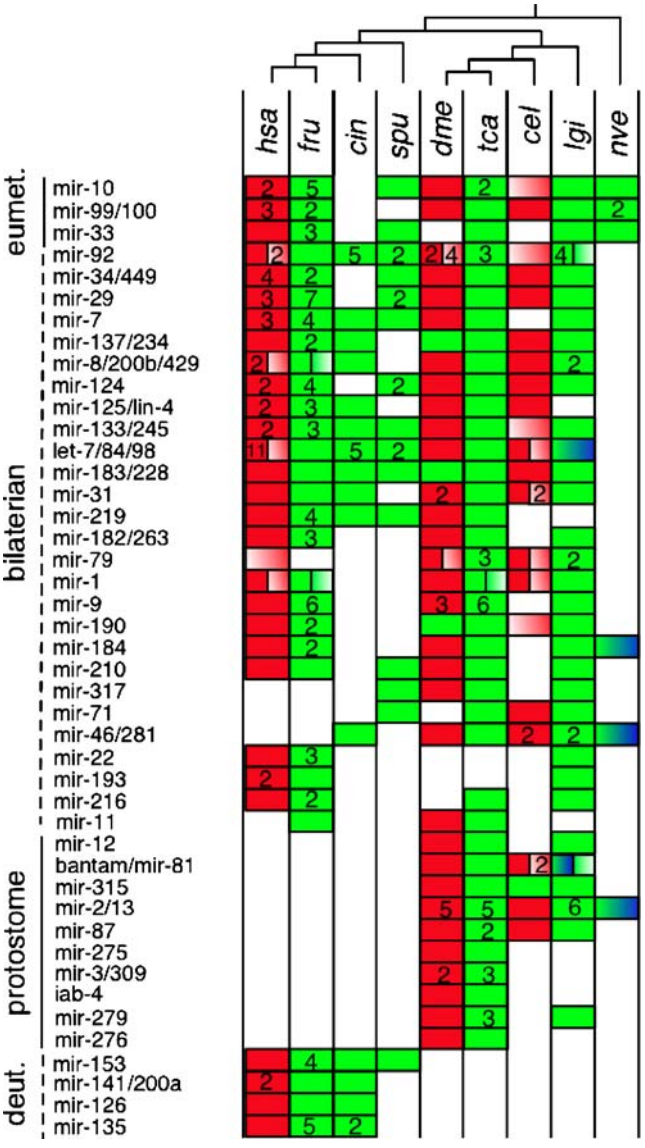

Fig. 1 miRNA distribution across the Bilateria. The presence of an experimentally verified (i.e., cloned) miRNA is represented by a *solid red box*, and a miRNA predicted by this study is represented by a *solid green box*. *Numbers in boxes* indicate the number of paralogs present in a species. miRNA family members with divergent sequence are indicated by *partial white shading* for cloned (*red* and *white*) and predicted (*green* and *white*) miRNAs. In cases where conserved and divergent family members are present, *two smaller boxes* are used. *Blue* shading identifies miRNAs with structural statistics that are close to the upper limit of the distribution of experimentally verified miRNAs. We do not show miRNAs that were found only in vertebrates or insects. The phylogenetic relationships between the species are shown at the top of the table. *hsaH. sapiens*, *fruF. rubripes*, *cinC. intestinalis*, *spuS. purpuratus*, *dmeD. melanogaster*, *tcaT. castaneum*, *celC. elegans*, *lgiL. gigantea*, *nveN. vectensis*, *renReniera* sp.; *eumet.* eumetazoan, *deut.* deuterostome

allowed fine-tuning to particular target mRNAs or particular target sites. An increase in the number of family members in a cluster, along with differences in the mature sequence, may increase the number of target sites that can be regulated.

We find that many more miRNAs than previously realized (and by deduction their target sites, though not necessarily their target mRNAs) are broadly conserved across Bilateria (Hertel et al. 2006; Sempere et al. 2006). The real number of conserved bilaterian miRNAs is therefore considerably higher then previous estimates of just 18 conserved families (Hertel et al. 2006; Sempere et al. 2006). We recovered all 18 of the previously described families, which were found either experimentally or by sequence homology (Hertel et al. 2006). The inclusion of new animal genomes that had not previously been searched for miRNAs, particularly a divergent chordate (*Ciona*), a nonchordate deuterostome (*S. purpuratus*), and a member of the lophotrochozoa (*L. gigantea*), was critical for observing the extent of miRNA conservation (Fig 1). Also, our approach of initially searching with just the mature miRNA sequence appears to be a more sensitive method than using the pre-miRNA hairpin sequence. This may be because our method allows for more sequence evolution of the opposite side of the hairpin as well as length changes in pre-miRNA hairpin (Lai et al. 2003). Formal testing of this is problematic because it requires exacting knowledge of the algorithms used by other studies (Hertel et al. 2006; Sempere et al. 2006).

Our results support an important role for miRNAs in the establishment of the bilaterian lineage. Considering the expression patterns of the conserved miRNAs in vertebrates (Wienholds et al. 2005) and flies (Aboobaker et al. 2005), and their computationally predicted targets (Enright et al. 2003, Griffiths-Jones et al. 2006), it is likely that most developmental pathways involved some level of miRNA-mediated regulation in the ancestor of all bilaterians.

We also looked for conserved miRNAs in the genomes of the nonbilaterian cnidarian and demosponge. Surprisingly, we could confidently predict only four miRNAs representing just three miRNA families (mir-99/100, mir-10, and mir-33) in the cnidarian *N. vectensis*. In addition to these found weak candidates for each of the mir-2/13 and mir-281 pan-bilaterian miRNA families (Fig 1), these candidates have free energy scores close to the upper limit for cloned miRNAs and so we cannot be confident that they are real miRNAs. In agreement with our findings, another recent study (Sempere et al. 2006) also detected the presence of mir-100 and mir-10 in cnidarians by Northern blot analysis, although they were unable to find mir-10 by informatic means. This may be due to their search method being less sensitive than the one we use in this study. They do not report the expression of either mir-2/13 or mir-281 in a cnidarian, suggesting that these weak candidates are not real miRNAs. We predicted no miRNAs with confidence in the demosponge *Reniera* sp. We conclude that the bilaterian/cnidarian ancestor probably had orthologs of mir-99/100, mir-10, and mir-33. We find homology between mir-99/100 and mir-10 miRNA families, suggesting that they arose by duplication from a single ancestral miRNA even earlier. The mir-100 family of miRNAs was found clustered with the let-7 and mir-125 miRNAs. Our data suggest that mir-100 is the most ancient of these and that both let-7 and mir-125 arose in the bilaterian ancestor. Target predictions for mir-10 (Enright et al. 2003) and its expression pattern in vertebrates (Woltering and Durston 2006) and flies (Aboobaker et al. 2005) suggest that it is likely to regulate Hox gene mRNAs. The presence of mir-10 in the cnidarian lineage suggests that the involvement of miRNA-mediated regulation in homeobox gene function was a feature of the eumetazoan ancestor and not just of bilaterians. miRNA-mediated regulation may therefore have played an important role in the evolution of early axis specification mechanisms. Further study of the role of mir-10 in the Cnidaria will help to assess this.

From our analysis, it appears that the vast majority of conserved bilaterian miRNAs emerged after the divergence of bilaterians from the bilaterian/cnidarian ancestor. We cannot rule out the possibility that the bilaterian/cnidarian ancestor had a complex miRNA repertoire that was secondarily simplified in modern cnidarians, but consider this unlikely. Another possibility is that the miRNAs that are conserved across Bilateria have diverged beyond our levels of detection in cnidarians and/or sponges. If this were true, however, they would also likely be binding divergent target sites that would all have to have coevolved. This seems less likely than the appearance of these miRNAs after the bilaterian/cnidarian split. Our results suggest that an increase in miRNA-mediated gene regulation could have played an important role in the evolution of increased regulatory complexity, cell type number, as well as morphological complexity, observed in bilaterian animals. The emerging picture of miRNA function in model animals, with some acting as developmental switches and with many more possibly acting as tissue-specific monitors of aberrant transcription (Brennecke et al. 2005; Bartel and Chen 2004), supports this model.

Acknowledgements SEP is funded by the US DOE Joint Genome Institute; DSR by the Center for Integrative Genomics; and AAA by a Wellcome Trust International Research Fellowship. We are grateful to the Baylor College of Medicine Genome Sequencing Center for the prepublication use of the *Tribolium castaneum* and *Strongylocentrotus purpuratus* genome data, and the DoE Joint Genome Institute for the prepublication use of the *Nematostella vectensis* and *Reniera* sp. genome data. We are grateful to Scott Nichols for comments on the manuscript. Simon E. Prochnik and A. Aziz Aboobaker contributed equally to this work.

References

Aboobaker AA, Blaxter ML (2003) Hox gene loss during dynamic evolution of the nematode cluster. Curr Biol 13:37–40

Aboobaker AA, Tomancak P, Patel NH, Rubin GM, Lai E (2005) Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc Natl Acad Sci USA 102:18017–18022

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T (2003) A uniform system for microRNA annotation. RNA (3):277–279

Bartel D, Chen C (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 5:396–400

Brennecke J, Stark A, Russell R, Cohen S (2005) Principles of microRNA-target recognition. PLoS Biol 3:e85

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS (2003) MicroRNA targets in Drosophila. Genome Biol 5(1):R1

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

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF (2006) Students of bioinformatics computer labs 2004 and 2005. The expansion of the metazoan microRNA repertoire. BMC Genomics 7:25

Kortschak RD, Samuel G, Saint R, Miller DJ (2003) EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Curr Biol 13:2190–2195

Lai EC, Tomancak P, Williams RW, Rubin GM (2003) Computational identification of Drosophila microRNA genes. Genome Biol 4:R42

Pasquinelli AE, McCoy A, Jimenez E, Salo E, Ruvkun G, Martindale MQ, Baguna J (2003) Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? Evol Dev 5:372–378

Sempere LF, Cole CN, McPeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol. DOI 10.1002/jez.b

Technau U, Rudd S, Maxwell P, Gordon PM, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, Ball EE, Miller DJ (2005) Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends Genet 21:633–639

Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev 20:2202–2207

Wienholds E, Plasterk RH (2005) MicroRNA function in animal development. FEBS Lett 579 (26):5911–5922

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, Plasterk RH (2005) MicroRNA expression in zebrafish embryonic development. Science 309:310–311

Woltering JM, Durston AJ (2006) The zebrafish hoxDb has been reduced to a single miRNA. Nat Genet 38:601–602

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415