Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Biology and Chemistry
Computational Biology and Chemistry 30 (2006) 249–254
www.elsevier.com/locate/compbiolchem

MicroRNA sequence motifs reveal asymmetry between the stem arms

J. Gorodkin [a,*,1], J.H. Havgaard [a,1], M. Ensterö [b], M. Sawera [a], P. Jensen [a], M. Öhman [b], M. Fredholm [a]

[a] *Division of Genetics and Bioinformatics, IBHV and Center for Bioinformatics, The Royal Veterinary and Agricultural University, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark*
[b] *Department of Molecular Biology and Functional Genomics, University of Stockholm, SE-106 91 Stockholm, Sweden*

Received 17 April 2006; accepted 24 April 2006

Abstract

The processing of micro RNAs (miRNAs) from their stemloop precursor have revealed asymmetry in the processing of the mature and its star sequence. Furthermore, the miRNA processing system between organism differ. To assess this at the sequence level we have investigated mature miRNAs in their genomic contexts. We have compared profiles of mature miRNAs within their genomic context of the 5′ and 3′ stemloop precursor arms and we find asymmetry between mature sequences of the 5′ and 3′ stemloop precursor arms. The main observation is that vertebrate organisms have a characteristic motif on the 5′ arm which is in contrast to the 3′ arm motif which mainly show the conserved U at the position of the mature start. Also the vertebrate 5′ arm motif show a semi-conserved G 13 nucleotides upstream from the first position. We compared the 5′ and 3′ arm profiles using the average log likelihood ratio (ALLR) score, as defined by Wang and Stormo (2003) [Wang T., Stormo, G.D., 2003. Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics 2369–2380.] and computing a *p*-value we find that the two profiles differs significantly in their 3′ end where the 5′ arm motif (in contrast to the 3′ arm motif) has a semi-conserved GU rich region. Similar findings are also obtained for other organisms, such as fly, worm and plants. The observed similarities and differences between closely and distantly related organisms are discussed and related to current knowledge of miRNA processing.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: miRNA; Mature miRNA organization; 5′ and 3′ hairpin arms; Sequence logos; Sequence profiles

1. Introduction

Mature miRNA forms a ∼ 22 nucleotide RNA duplex together with its star sequence, miRNA*, and is processed out in an asymmetric fashion from its stemloop precursor structure (reviewed by Bartel, 2004).

The asymmetry results from the two (metazoan) processing steps conducted by the nuclear Drosha and the cytoplasmic Dicer. Both these RNase III endonucleases act on different precursor signals. Drosha is thought to interact with the hairpin apex loop and cuts the hairpin near the terminal base, thus defining one end of the mature miRNA (Zeng and Cullen, 2004; Lee et al., 2003). The Drosha processing of a hairpin structure is further coordinated by Pasha which have two RNA binding motifs, a homolog to the mammalian DGCR8 (DiGeorge syndrome chromosomal region 8) (Denli et al., 2004; Gregory et al., 2004). Dicer is acting via its PAZ-domain which is known to interact with the 2 nucleotide 3′ overhang (Ma et al., 2004). Dicer then cuts away the loop subsequently defining the miRNA::miRNA* duplex.

Dicer has been shown to associate with a variety of different proteins including another highly conserved group—the Argonaute family which also share the PAZ-domain (Hammond et al., 2001). The RISC (RNA induced silencing complex) is a multi-protein complex and the understanding of the biogenesis scheme from miRNA::miRNA* duplex to final single stranded mature miRNA is not fully delineated. The Ago2 has been shown to be the actual slicer within the RISC (Meister et al., 2004). While siRNA specifically degrade their target through Ago2 miRNAs is believed to mainly interact with the first ∼ 7 nucleotides hence diversifying the target repertoire. Another important RISC component for fine tuning strand selection is R2D2 (*Drosophila melanogaster*) which can sensor the different thermodynamic inequalities for accurate strand incorporation (Tomari et al., 2004).

It has been observed that miRNAs are less stable in the 5′ end than in their 3′ end (5′ end of the star sequence)

* Corresponding author. Tel: +45 3528 3578; fax: +45 3528 3042.
E-mail address: gorodkin@bioinf.kvl.dk (J. Gorodkin).
[1] These authors contributed equally to this work.

1476-9271/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2006.04.006

Table 1
The table gives an overview of the miRNAs used as well as the distribution of the left and right matures

Org	Genome	DB	Hair	U-hair	5′arm	3′arm	Mat	$5'_{red}$arm	$3'_{red}$arm
hsa	NCBI35	332	332	332	203	191	394	122	119
mml	MMUL0.1	71	63	62	44	23	67	32	20
mmu	NCBIM34	270	276	267	159	148	307	101	101
rno	RGSC3.4	234	228	228	138	116	254	92	88
gga	WASHUC1	144	144	144	88	64	152	42	40
dre	WTSIZv5	372	335	310	149	198	347	49	45
fru	FUGU4	131	132	130	68	65	133	36	39
tni	TETRAODON7	131	142	131	72	70	142	35	39
dme	BDGP4.0	78	78	78	34	51	85	29	29
dps	DPSE2.0	73	73	73	28	46	74	25	28
cbr	cb25.agp8	79	82	79	26	56	82	22	37
cel	WS140	114	114	113	41	75	116	30	55
ath	Refseq[a]	117	117	117	62	57	119	28	21
osa	TIGR3.0	178	124	123	62	62	124	20	21

Org, the organism; genome, the release (see text for details); DB, the number of hairpins in the miRNA database; hair, the number of hairpins with genome coordinates; U-hair, the number unique hairpins with genome coordinates; 5′ arm (3′ arm, respectively), the number mature miRNAs on the 5′ arm of (3′ arm, respectively) of the hairpin; mat, the number of mature sequences with genome coordinates; $5'_{red}$arm ($3'_{red}$arm, respectively), the number of mature miRNAs with genome coordinates on the 5′ arm (3′ arm, respectively) left after similarity reduction (see text for details).

[a] Consist of GenBank accessions: NC_003070.5, NC_003071.3, NC_003074.4, NC_003075.3, NC_003076.4.

(Schwarz et al., 2003; Khvorova et al., 2003; Krol et al., 2004) and that the molecular processing machinery can sensor this (Tomari et al., 2004). Also, recent findings for intronic miRNAs in zebrafish suggest a non-canonical asymmetry in the process of strand selection acting concurrently with thermodynamical properties (Lin et al., 2005). Here, we further investigate this asymmetry and show that the organization in the genomic sequence context is asymmetric with respect to the mature sequence in the 5′ and 3′ arms of the stemloop precursor. This organization is similar for related organisms, but different for distantly related organisms.

2. Materials and methods

2.1. Data

Organisms represented in mirBASE version 8.0 (Griffiths-Jones et al., 2005) was extracted in their genomic contexts and only miRNA hairpins with genome gff coordinates was used. All coordinates were checked by comparing the extracted sequence and the sequence in the registry. Hairpins for which genomic coordinates were not given were ignored. One hairpin from *cel* was removed as it had no mature sequence annotated. For each organism the sequence data was divided into two sets one containing the mature sequences on the 5′ arm in the stemloop precursor and one where the mature sequences are on the 3′ arm in the stemloop precursor. For stemloops containing mature sequences on both the 5′ and 3′ arms, the mature sequences were used in their respective contexts. The number of such cases is in general low.

Furthermore, we made similarity reduced sets by grouping the sequences into families by the nucleotides 2–8 of the mature sequences, using only one sequence from each family (Lewis et al., 2005). Only organisms with at least 20 sequences left for both the 5′ and 3′ arms were included in the data set. These are: *Arabidopsis Thaliana* (*ath* (Arabidopsis Genome Initiative, 2000)), *Caenorhabditis briggsae* (*cbr* (C. elegans Sequencing Consortium, 2006)), *Caenorhabditis elegans* (*cel* (C. elegans Sequencing Consortium, 1998)), *Drosophila melanogaster* (*dme* (Celniker et al., 2002)), *Drosophila pseudoobscura* (*dps* (Richards et al., 2005)), *Danio rerio* (*dre* (The Zebrafish Sequencing Group, 2006)), *Fugu rubripes* (*fru* (Aparicio et al., 2002)), *Gallus Gallus* (*gga* (Int. Chicken Genome Sequencing Consortium, 2004)), *Homo Sapiens* (*hsa* (Int. Human Genome Sequencing Consortium, 2004)), *Macaca mulatta* (*mml* (HGSC at Baylor College of Medicine, 2006)), *Mus Musculus* (*mmu* (Mouse Genome Sequencing Consortium, 2002)), *Oryza sativa* (*osa* (Yuan et al., 2003)), *Rattus norvegicus* (*rno* (Rat Genome Sequencing Project Consortium, 2004)) and *Tetraodon nigroviridis* (*tni* (Jaillon et al., 2004)). The miRNA sequences ("hairpins" as in mirBASE) were then matched with their genomic context, and whole segments typically of 3000 nucleotides were extracted. The details of the data are listed in Table 1.

2.2. Sequence profiles

To construct sequence profiles the miRNAs along with their surrounding genomic context, were aligned by the start of their mature sequence. For each of the considered organisms this was done for the 5′ and 3′ arm mature sequences, respectively. Next, sequence logos (Schneider and Stephens, 1990) were generated by computing the relative entropy as by Gorodkin et al. (1997), with nucleotide frequencies computed for each position of the aligned sequence. Briefly, the information content for each position in the alignment is defined as $I = \sum_l q_l \log_2 q_l/p_l$, where l belong to the set of nucleotides. The fraction q_l is the observed nucleotide distributions, whereas the fraction p_l is the expected (background) nucleotide frequencies drawn from the miRNA hairpin excluding the mature sequence. For each position in the logo the correspond to the information content I, and the height of the letter l is the portion $q_l I$. When $q_l < p_l$ letter l is displayed upside down.

2.3. Comparing distributions

To compare the significance of the difference between 5′ and 3′ arm motifs the corresponding weight matrices (profiles) from the sequence logos were stored for computing the average log-likelihood ratio (ALLR) score as defined by Wang and Stormo (2003). This measure can be used to distinguish two corresponding columns from each weight matrix. It is the joint probability of observing the data generated by one distribution given the likelihood ratio of the other distribution. The ALLR score is a log-likelihood test of how one data set fits into another and vice versa. It is the average of the two log-likelihood ratios. When the data sets are unrelated the ALLR is expected to be negative. For details, see Wang and Stormo (2003).

Here, we compute the ALLR score for the two profiles (5′ arm and 3′ arm, respectively) when aligning them up- and downstream from the beginning of the mature sequence. For this fixed alignment we compute the ALLR score across several different regions. When comparing the two profiles, an ALLR score is computed over the corresponding regions of the two profiles. One of the nice features of the ALLR score is that it takes into account that the profiles compared can be made from different numbers of sequences.

Empirical *p*-values for significance of the obtained ALLR scores are computed in a given region by keeping the columns of a window from the 5′ arm (3′ arm, respectively) fixed while shuffling the columns (100 times) in the 3′ arm window (5′ arm, respectively) and for each shuffling, computing the ALLR score. The rank of the true ALLR score gives the empirical *p*-value.

3. Results

The data sets for the organisms considered here are shown in Table 1, where we observe the following: for the non-reduced data sets of the organisms *has, mml, mmu, rno, gga* that there seems to be a slight over representation of 5′ arm mature miRNAs. In contrast for fly and worm the over representation seems to be for the 3′ arm miRNAs. For plants, the number on both arms appears to be the same. The latter have also been noticed in by others (Bartel and Bartel, 2003).

However, here we focus on the similarity reduced sets of mature miRNAs in the 5′ and 3′ arms and unless mentioned otherwise we refer to this set. Results similar to those presented for reduced sets are obtained on the full non-reduced 5′ and 3′ arm data sets (not shown). For each organism we constructed profiles of the 5′ and 3′ arms with the precursor in the genomic context as described in Section 2. The profiles can be represented by sequence logos (Schneider and Stephens, 1990) as shown in Fig. 1 for the human case. The corresponding profiles for the organisms listed in Table 1 are shown in the supplementary material Figure S1. Note that the first position of the mature miRNA is indicated as position zero in the logos.

By inspection, we observe that all organism profiles show different characteristics between their 5′ and 3′ arm motifs. For the vertebrate organisms (*hsa, mml, mmu, rno, gga, dre, fru, tni*) we observe that they have a characteristic motif on the 5′ arm. In contrast, the 3′ arm motif essentially only displays the well

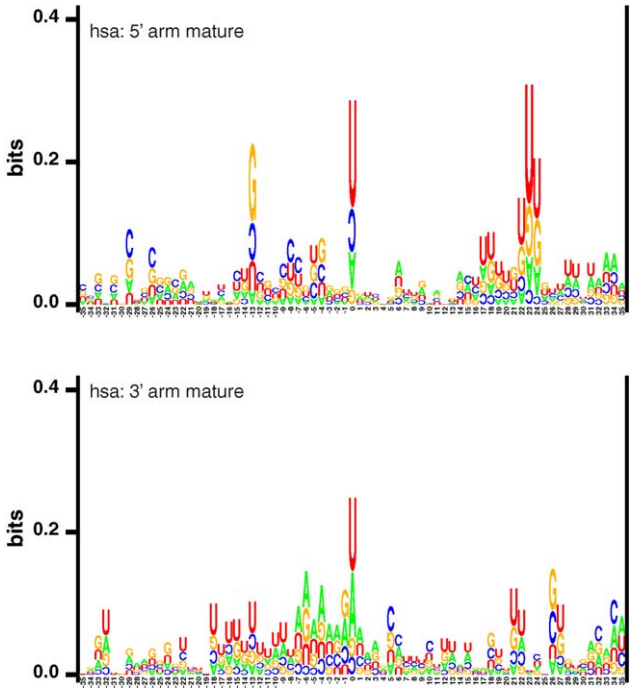

Fig. 1. The sequence logos of the 5′ (top) and 3′ (bottom) arms of the human miRNAs in their genomic context. Position zero corresponds to mature sequence start. The 5′ arm logo was generated from 122 sequences and the 3′ arm logo from 119 sequences. Letter sizes are shown according to their frequencies. (Upside-down is less than the expected frequency.).

known conserved U at the start of the mature sequence. For the invertebrates organisms flies and worms (*dme, dps, cel, cbr*) the 3′ arm motif is more characteristic showing a highly conserved U at the mature start. For plants (*ath* and *osa*) both the 5′ and the 3′ arm motifs show characteristic, but different motifs, the 5′ arm motif having a strongly conserved U at the mature and the 3′ arm a conserved C at the mature end. However, as there are relatively few plant sequences in the reduced sets, more sequences will are likely to provide more information.

For the characteristic vertebrate 5′ arm motif it contains the well known U conservation at the beginning of the mature sequence (position zero in logos) and a GU rich region in the 3′ end (of the 5′ arm) around positions 18–25. Interestingly, the 5′ arm motif also contains an upstream semi-conserved G at position −13. For the invertebrate organisms (*dme, dps, cel, cbr*), the 3′ arm motif seems characteristic. Fly seems to have more conserved positions in the neighborhood of the mature start in particular a semi-conserved U at position −9. (See Figure S1 in the supplementary material for details.)

To compute which parts of the 5′ and 3′ motifs are similar and different, we utilized a sliding window across the two profiles and computed the ALLR score and an empirical *p*-value at corresponding positions (Section 2). Window sizes from 6 to 14 were utilized all providing the same information with different resolutions. In Fig. 2 we show the scan on human for window size 7. It is in particular notable that around positions 15–20 the ALLR score drops significantly while the *p*-value at the same time is getting close to one. This indicates that the 3′ end of the two types of mature sequences differs (low ALLR score) and

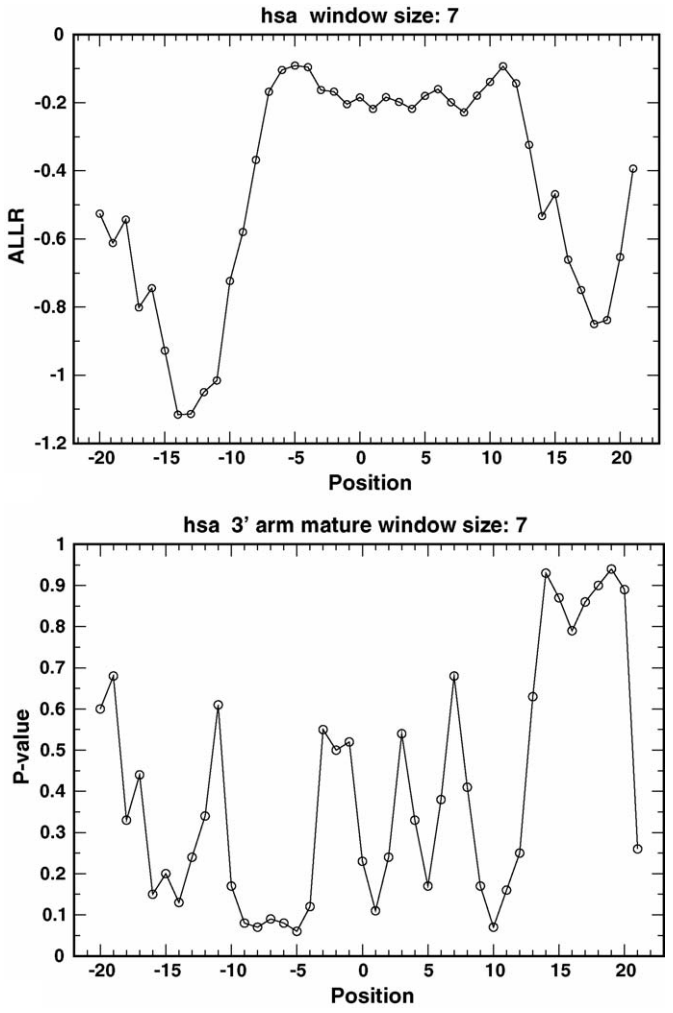

Fig. 2. Profiles of ALLR scores (top) and p-values (bottom) for human using a window size of 7 nucleotides across the two profiles, that are assumed aligned to the corresponding regions, see Wang and Stormo (2003) for details. For each position, the three neighboring nucleotides on both sides were used to compute the ALLR score. The p-values for each of the sliding windows are computed empirically by shuffling the columns (100 times) in each of the windows of the 5′ arm motif while fixing 3′ arm motif, see Section 2 for details. An almost identical plot is obtained by shuffling the 3′ arm motif while fixing the 5′ arm motif (not shown). The profiles in the top row are computed by keeping the columns of the 5′ arm motif fixed while shuffling the columns of the 3′ arm motif (corresponding positions).

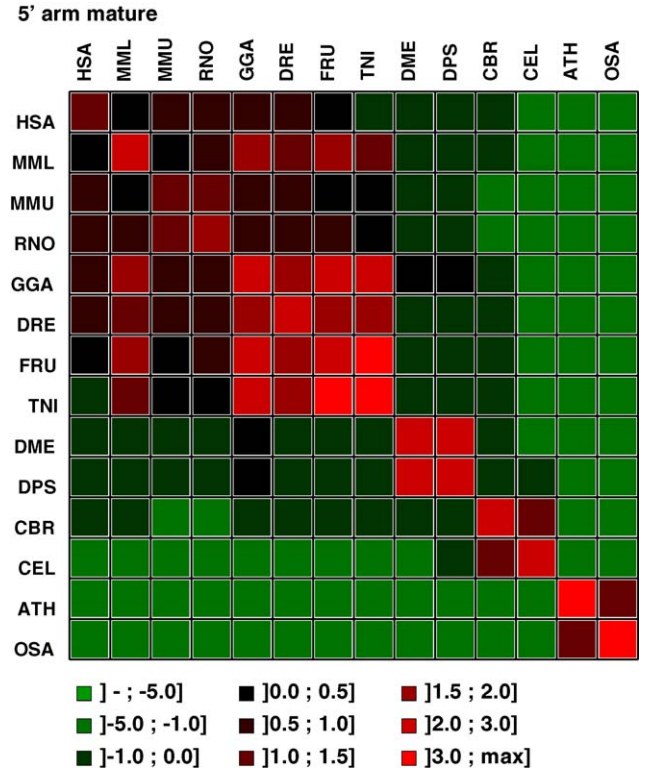

Fig. 3. Pairwise comparisons for the organisms listed in Table 1. The ALLR score was computed of the region spanning from the mature start position and 25 nucleotides downstream.

also differs significantly as the p-value is high. Similar type of observations are obtained for the other organisms listed in Table 1, however the curves do in some instances vary differently upstream from the 3′ end of the compared mature regions (data not shown). In few cases the p-value signal on positions 15–20 is not so strong, and the p-value drops to 0.55 in a case (*cel*). Also the peak might be shifted towards position zero.

In contrast to the difference observed between the 3′ ends of the 5′ and 3′ arm mature sequences, we for the 5′ ends observe that the ALLR score, although negative it is only slightly negative and the p-value is close to zero. This indicates that the ALLR score in this region is not significantly different from what would be expected when comparing with shuffled sequences. Hence no conclusion can drawn about similarity between the 5′ ends of the mature sequences in the 5′ and 3′ stem arms.

We also compared the 5′ arm (3′ arm, respectively) among the organisms computing the ALLR score. We compared the region covering the entire mature sequence starting at position 0 ending 22–29 nucleotides downstream. For each comparison we find that the related organisms have a relatively high score among themselves and score lower with more distant organisms. An example using a region spanning 24 nucleotides downstream is shown in Fig. 3. Note that the ALLR score of a profile against itself is exactly the information content of the profile within the considered region. The only less inconsistent pattern is the *mml* comparison. This is likely to be due to the relatively few sequences compared to the many sequences for the close related organism *hsa, mmu, rno*. This is also likely to be reflected in the higher score of *mml* against itself than *hsa, mmu, rno* against themselves.

4. Discussion

We did find that there is an asymmetry (difference) between miRNA sequence motifs when the mature sequence is located in the 5′ and 3′ arms of the stemloop precursor. A key question is, whether the 5′ and 3′ arm of the mature miRNA sequences are processed in the same way. Given the asymmetry observed here, this could be possible if, for example, the 3′ arm of the mature miRNA* sequence contains the same features as the mature 5′ arm sequence. However, recent studies have provided biochem-

ical verification showing that a mature miRNA is less stable at its 5′ end than its 3′ end (5′ end of star sequence) (Schwarz et al., 2003; Khvorova et al., 2003; Krol et al., 2004). This shows that the star sequences of the 3′ arm cannot have the same properties as the mature sequences on the 5′ arm and vice versa. These observations suggest that the miRNA processing machinery not only acts in an asymmetric fashion with respect to the mature and its star sequence, as showed by Tomari et al. (2004), but is also asymmetric with respect to processing 5′ and 3′ arm sequences. For vertebrate organisms, the upstream conserved G of the 5′ arm motif is a candidate for playing a role in the asymmetry.

We also observed that profiles among the organisms somewhat differs and that more closely related organisms have a higher score among themselves than more distantly related organisms. The vertebrate profiles seems to share very similar profiles of the 5′ arm motif, but interestingly fish appears to differ on the 3′ arm motif with and A-dominated signal at position 13 in the logos. However, in particular the plant appears to a specific feature, namely semi-conserved C in 3′ end of both the 5′ and 3′ arm mature sequences.

In agreement with this, the plant species are well known to have major differences in the biogenesis. They lack the processing of Drosha which instead is mediated through Dicer-like endonucleases— and more specifically DCL1 (Dicer-like protein 1) (Papp et al., 2003). DCL1 acts, in contrast to metazoan homologs, in the nucleus as the first processing steps of the pre-miRNAs which are further categorized differently from the metazoan intermediates. It is both more variable in size and have a high turn-over rate most likely from a coupled processing in the nucleus from the DCL endonuclease, resulting in a temporary precursor intermediate (Reinhart et al., 2002). Moreover, plant miRNA::target interaction is also more precise and shows a near-perfect complementarity (Rhoades et al., 2002). No obvious conservation of any miRNA gene and lack of Drosha homologs between the animal and plant kingdoms even propose an independent origin of this mechanism, as suggested by Bartel (2004).

Even though there is no experimental results concerning the different organization of 5′ and 3′ arm mature sequences between fly/worm and amniotic deployment, related distinctions have been observed. Note, that the RISC complex have only been studied in detail for fly (Tomari and Zamore, 2005) and similar studies might reveal variation in processing, for example, between human and fly. A related difference in the RISC complex with respect to RNAi have been observed, where Argonaute 2 is the only slicer in human that provides a fully functional RISC complex (Liu et al., 2004). Mammals do not have an endogenous siRNA expression in contrast to the lower eukaryotic species (reviewed by Bartel, 2004). The mammalian miRNA biogenesis has evolved to a state of fine tuning the processing steps solely with miRNA expression at hand. The other clustered groups have different silencing pathways (DNA methylation, siRNAs) relying in general on the same set of processing machinery hence, signals for biogenesis properties had to evolve coordinately in contrast to the mammalian way where miRNA associated proteins and miRNA signals have evolved synchronously. Another aspect is also target substrates. Although many miRNA families seem to be evolutionary conserved, which also is a trait distinguishing them from siRNA, there is a rising number of mammal specific miRNAs, for example the mir-196 involved in regulating expression from the HOX-gene clusters (Yekta et al., 2004).

Our observations also indicate differences between worm and fly and it has been suggested that they could have different RNAi pathways (Zamore, 2002). In fact, recent findings (Lin et al., 2005) show that the location of the mature sequence in intronic miRNAs in zebrafish are crucial for proper processing. Here, it is suggested that Dicer promotes asymmetry in strand selection possibly due to sequence bias within the apex loop. Hence, our observations are not in conflict with the current knowledge of miRNA processing, but contribute further to the possibility of variations in the miRNA processing machinery.

Acknowledgements

Thanks to Gary D. Stormo and Ting Wang for many long discussions and suggestions to comparing the two types of motifs. Thanks to Anders Krogh on comments on an early version of the manuscript. This work was supported by Danish Research Councils (SJVF/STVF) and the Danish Center for Scientific Computation. MÖ and ME were supported by Wallenberg Consortium North.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.compbiolchem.2006.04.006.

References

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M., Roach, J., Oh, T., Ho, I., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S., Clark, M., Edwards, Y., Doggett, N., Zharkikh, A., Tavtigian, S., Pruss, D., Barnstead, M., Evans, C., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., Brenner, S., 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301–1310.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Bartel, B., Bartel, D.P., 2003. MicroRNAs: at the root of plant development? Plant Physiol. 132 (2), 709–717.

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

Celniker, S., Wheeler, D., Kronmiller, B., Carlson, J., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S., Frise, E., Hodgson, A., George, R., Hoskins, R., Laverty, T., Muzny, D., Nelson, C., Pacleb, J., Park, S., Pfeiffer, B., Richards, S., Sodergren, E., Svirskas, R., Tabor, P., Wan, K., Stapleton, M., Sutton, G., Venter, C., Weinstock, G., Scherer, S., Myers, E., Gibbs, R., Rubin, G., 2002. Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. Genome Biol. 3 (RESEARCH0079).

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282, 2012–2018.

C. elegans Sequencing Consortium, 2006. http://www.sanger.ac.uk/pub/wormbase/cbriggsae/cb25.agp8.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., Hannon, G.J., 2004. Processing of primary microRNAs by the microprocessor complex. Nature 432, 231–235.

Gorodkin, J., Heyer, L.J., Brunak, S., Stormo, G.D., 1997. Displaying the information contents of structural RNA alignments: the structure logos. CABIOS 13, 583–586, http://www.cbs.dtu.dk/gorodkin/appl/slogo.html.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., Shiekhattar, R., 2004. The microprocessor complex mediates the genesis of micrornas. Nature 432, 235–240.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., Bateman, A., 2005. Rfam: annotating non-coding rnas in complete genomes. Nucl. Acids Res. 1, D121–D124.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., Hannon, G.J., 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 293, 1146–1150.

HGSC at Baylor College of Medicine, 2006. http://www.ensembl.org/pub/current_macaca_mulatta/data/fasta/dna.

Int. Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Int. Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.

Jaillon, O., Aury, J., Brunet, F., Petit, J., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K., McEwan, P., Bosak, S., Kellis, M., Volff, J., Guigo, R., Zody, M., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E., Weissenbach, J., Roest, C.H., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431, 946–957.

Khvorova, A., Reynolds, A., Jayasena, S.D., 2003. Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209–216.

Krol, J., Sobczak, K., Wilczynska, U., Drath, M., Jasinska, A., Kaczynska, D., Krzyzosiak, W.J., 2004. Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J. Biol. Chem. 279, 42230–42239.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature 425 (6956), 415–419.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. Cell 120, 15–20.

Lin, S.-L., Chang, D., Ying, S.-Y., 2005. Asymmetry of intronic pre-miRNA structures in functional RISC assembly. Gene. 356, 32–38.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., Hannon, G.J., 2004. Argonaute2 is the catalytic engine of mammalian RNAi. Science 305, 1437–1441.

Ma, J.-B., Ye, K., Patel, D.J., 2004. Structural basis for overhangspecific small interfering RNA recognition by the PAZ domain. Nature 429, 318–322.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T., 2004. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. Mol. Cell 10, 544–550.

Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., Matzke, A.J., 2003. Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol. 132, 1382–1390.

Rat Genome Sequencing Project Consortium, 2004. Genome sequence of the brown norway rat yields insights into mammalian evolution. Nature 428, 493–521.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P., 2002. MicroRNAs in plants. Genes Dev. 16 (13), 1616–1626.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., Bartel, D.P., 2002. Prediction of plant microRNA targets. Cell 110 (4), 513–520.

Richards, S., Liu, Y., Bettencourt, B., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M., Chen, R., Meisel, R., Couronne, O., Hua, S., Smith, M., Zhang, P., Liu, J., Bussemaker, H., van, B.M., Howells, S., Scherer, S., Sodergren, E., Matthews, B., Crosby, M., Schroeder, A., Ortiz-Barrientos, D., Rives, C., Metzker, M., Muzny, D., Scott, G., Steffen, D., Wheeler, D., Worley, K., Havlak, P., Durbin, K., Egan, A., Gill, R., Hume, J., Morgan, M., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K., Dubchak, I., Noor, M., Anderson, W., White, K., Clark, A., Schaeffer, S., Gelbart, W., Weinstock, G., Gibbs, R., 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. Genome Res. 15, 1–18.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display concensus sequences. Nucl. Acids Res. 18, 6097–6100.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D., 2003. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208.

The Zebrafish Sequencing Group, 2006. http://www.ensembl.org/pub/current_danio_rerio.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., Zamore, P.D., 2004. A protein sensor for siRNA asymmetry. Science 306, 1377–1380.

Tomari, Y., Zamore, P.D., 2005. Perspective: machines for RNAi. Genes Dev. 19, 517–529.

Wang, T., Stormo, G.D., 2003. Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics 2369–2380.

Yekta, S., Shih, I.H., Bartel, D.P., 2004. Microrna-directed cleavage of hoxb8 mrna. Science 304, 594–596.

Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J., Buell, C., 2003. The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. Nucleic Acids Res. 31, 229–233.

Zamore, P.D., 2002. Ancient pathways programmed by small RNAs. Science 296, 1265–1269.

Zeng, Y., Cullen, B.R., 2004. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res. 32 (16).