BioOne COMPLETE

Computational Identification and Characterization of Putative miRNAs in Nasonia Species

Authors: Sathyamurthy, Ganesh, and Swamy, N. Ramachandra

Source: International Journal of Insect Science, 2(1)

Published By: SAGE Publishing

URL: https://doi.org/10.4137/IJIS.S4197

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

International Journal of Insect Science

Libertas Academica
FREEDOM TO RESEARCH

OPEN ACCESS
Full open access to this and thousands of other papers at http://www.la-press.com.

ORIGINAL RESEARCH

Computational Identification and Characterization of Putative miRNAs in Nasonia Species

Ganesh Sathyamurthy and N. Ramachandra Swamy

Department of Biochemistry, Bangalore University, Bangalore-560 001, India. Email: sganesh@bub.ernet.in

Abstract: MicroRNAs are important at post transcriptional regulation in eukaryotes. Nasonia genus is becoming increasingly popular model in present days due to genetic advantages it possesses over Drosophila. Nasonia species are found distributed throughout the world, expect for *N. longicornis*, and *N. giraulti*. In this study, we use the sequential method of blasting all known invertebrate miRNA genes against the *Nasonia vitripennis, Nasonia longicornis*, and *Nasonia giraulti* genomes. We identify 40, 31 and 29 putative pre-miRNAs and mature sequences in *N. vitripennis, N. giraulti* and *N. longicornis* respectively. A cross species comparison of putative miRNA sequences and their statistical characteristics reveals that there are no huge differences between the species, except for few miRNAs which are reported. We also find that the minimal folding energy index for three Nasonia species pre-miRNA's average is around -0.85 ± 0.11. Further, we report that U is predominant at the 5` end of mature sequence, which being a typical characteristic of plant miRNAs. Using MiRanda, we predict nearly 471 potential sites in the N. vitripennis genome. Thus concluding our study to be the beginning of understanding the Nasonia's non coding RNAs and may play an important role in effective pest management in near future.

Keywords: evolution, divergence, invertebrates, computational biology, homologs, potential miRNA targets

International Journal of Insect Science 2010:2 7–19

This article is available from http://www.la-press.com.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

Introduction

microRNAs are a class of conserved small noncoding regulatory RNAs. miRNAs are involved in regulation of gene expression at the post transcriptional level. In animals they inhibit translation by partially binding to their target mRNAs followed by their cleavage, while in plants they have perfect or near perfect base pairing on their target sequence. Apart from regulation of gene expression, they are also involved in control of organ development, stem cell differentiation and developmental timing.[1] Different studies reveal role of miRNAs in diseases such as cancer and other infections.[2,3] Recent studies on miRNAs have begun to explore exact effect of the microRNA clusters or individual microRNA on cellular processes.[4]

Preliminary surveys of miRNAs across the animal kingdom demonstrated a very compelling feature of miRNA evolution when compared with the evolution of the protein coding repertoire.[5] Many studies on evolutionary surveys have been reported on comparing the conserved miRNAs between the taxonomies of interest.[6–9] If a taxon increased its mutation rate or lost miRNAs, then the phylogenetic history of that particular miRNA would be inaccurately constructed, leading to potentially spurious claims about the import of miRNAs with respect to organism's evolution. Yet, there are many studies exploring the miRNA repertoire independent of phylogenetic conservation.[5,10,11] Further, miRNAs have enormous potential in disease diagnosis and find application in gene therapy.[12]

Nearly, 9539 hairpin sequence entries have been registered by March, 2009 in miRBase. Sequence analysis have shown that some mature miRNAs are phylogenetically conserved, particularly in the first 8 residues at the 5` end in species of the same kingdom.[13–15] It has also been reported that quite a few mature miRNA sequences are conserved between animals and plants. Examples include, mir-854, identified in *C. elegans*, mouse, human and plants.[16]

Currently there are three methods for identifying miRNAs, which are the classic cloning method, deep sequencing method and computational approaches. The classic and deep sequencing methods are generally used for validation of miRNAs, but are not efficient as compared to the computational approach for detecting miRNAs. The computational approach could be further classified into 3 types: *ab initio* prediction based on the sequence and structural features, comparative genomic strategy based on conservation and the integrated approach. Most of the known miRNAs presently are detected by computational approaches in diverse organisms from plants to higher animals.[17–21]

Many computational prediction algorithms of miRNAs have been developed to aid in experimental studies of miRNA discovery. Commonly, all algorithms have different approach with the aim of reducing the false positive and/ or to increase specificity of prediction.[22] Evolutionary conservation is considered an important feature of the hairpin sequence and analysis thereof is often used to identify and focus comparison on the conserved noncoding sequences space in different genomes.[23–25] Phylogenetic shadowing has been used for combined selection and filtering of miRNA candidates.[26] Other filtering criterias include intragenomic matching of candidate miRNAs and their potential targets,[27] expression profile, thresholds such as minimal folding energy (MFE), minimal folding energy index (MFEI), occurrence of intergenic regions or existing in close proximity of known miRNA clusters.[28–30]

Nasonia is approximately 120MY diverged from Honeybees, in a second major branch of Hymenoptera. The Order Hymenoptera is diverse and are natural enemies of a broad range of vector arthropods of medical, veterinary and agricultural significance. Nasonia belongs to family Pteromalidae, whose adults lay their eggs in or various life stages of other arthropods subsequently regulating the insect population. There are three closely related species in the genus Nasonia, they are *Nasonia vitripennis*, *Nasonia longicornis* and *Nasonia giraulti*. These three species differ in their host preferences—*N.* vitripennis parasitize a wide range of flies (including blowflies, fleshfiles and houseflies), while the other two species appears to be specialists for their host. They are haplodiploidy, which allows geneticists to exploit many of the advantages of haploid genetics. As a result, Nasonia, are emerging as models for studies of complex genetic traits. Nasonia is well positioned phylogenetically to assist in identifying orthologs of important genes in insects and a genetically traceable system for functional analysis.[31] On a higher level these genomes may help us to understand features such as regulatory domain evolution, frequency and type of noncoding DNA, and metabolic capabilities. Thus in our

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

studies here, we identify putative miRNA homologs in *N. vitripennis*, *N. longicornis* and *N. giraulti*, and compare the sequences among the species. Further we also describe the statistical sequence characteristics of putative miRNA gene sequences and predict their possible targets in *N. vitripennis* species using MiRanda.[32,33]

Materials and Methods

microRNA sequence dataset

The pre-miRNAs (miRNA genes) for all known invertebrates, 1742 sequences were obtained from miRBase Sequence Database, release 13.0, March 2009.[34] These sequences include all miRNA reported species from protozoans to echinodermata. These miRNAs in database were either obtained by direct cloning or/ and confirmed by a variety of experimental approaches, including northern blotting, polymerase chain reaction or microarray. Besides these, many of them were obtained by computational identification of their homologs in closely related species.

Detection of miRNA homolog in all *Nasonia* species genome

Here we performed BLAST[35] (expect value 0.01, mismatch -2) search using all previously published invertebrate pre-miRNAs and queried against the *N. vitripennis* (6.2X), *N. longicornis* (1X), *and N. giraulti* (1X) genome assemblies.[31] All the hits were downloaded in FASTA format[36] and used for further analysis. The hits were iterated to remove the duplicates of same miRNA gene based on optimal values of identity, alignment length and gaps between the query and hit sequences. A flowchart is shown summarizing the method for identifying putative miRNAs in *Nasonia* (Fig. 1). Since the complete assembly of the genome is not available or completed, it may be too early to identify the clustered miRNAs in Nasonia.

Sequence characteristic analysis

The sequence characteristic analysis was performed by identifying base frequencies of A, G, C, U, A + U, G + C of the hit sequences. Independently all invertebrate miRNA gene sequences statistical analysis was performed and compared with the iterated pre-miRNA sequences. MFE of the secondary structure was obtained using m-FOLD[37,38] for all the putative pre-miRNA genes. The adjusted minimal folding energy (AMFE) and the MEFI, was calculated as previously described by Zhang.[39]

Results and Discussion

Pre- and mature miRNAs of nasonia

Pre-miRNA sequences of all invertebrates studied till date were downloaded from the miRBase.[34] These pre-miRNA sequences were used as query for BLAST tool against the *N. vitripennisi* (6.2x), *N. longicornis* (1X), *and N. giraulti* (1X), genome assemblies. All the hits were stored in multi FASTA format and used for further analysis. The iteration of the raw BLAST data is summarized in Figure 1. For those sequences with less than 85% identity in all three species were validated using miPred. The resulting sequences were used as putative pre-miRNA sequences (Supplementary Table 1, 2, 3). In *N. vitripennis* after iteriation 40 putative miRNAs were obtained, of which mir-iab-4 alone had 100% Identity at pre-miRNA level, 13 miRNAs had identity between 95%–99.9% Identity, followed by 14 miRNAs between 90%–94.9% and 12 between 80%–89.9% Identity (Fig. 2) to their query sequences at pre-miRNA level. Even in case of *N. giraulti* after iteriation of the BLAST hits, 31 putative miRNAs were obtained of which only mir-iab-4 had 100% identity, followed by 9 pre-miRNAs between 95%–99.9%, 10 pre-miRNAs

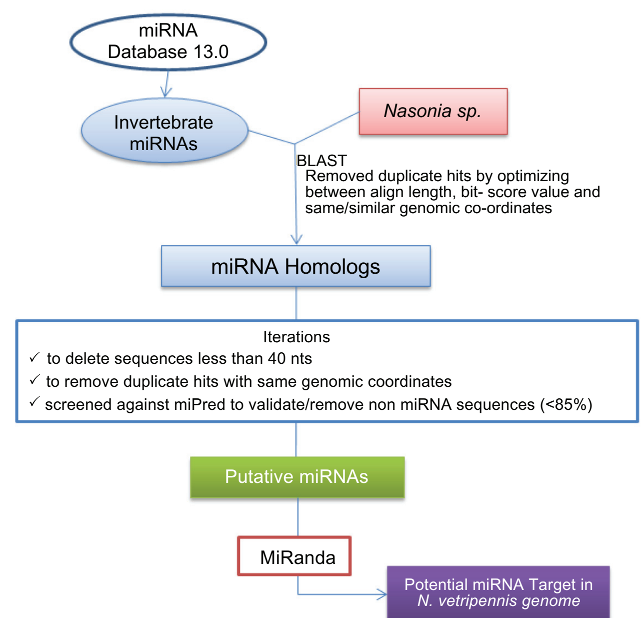

Figure 1. Flowchart summarizing the method for identifying miRNA homologs from Nasonia species.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

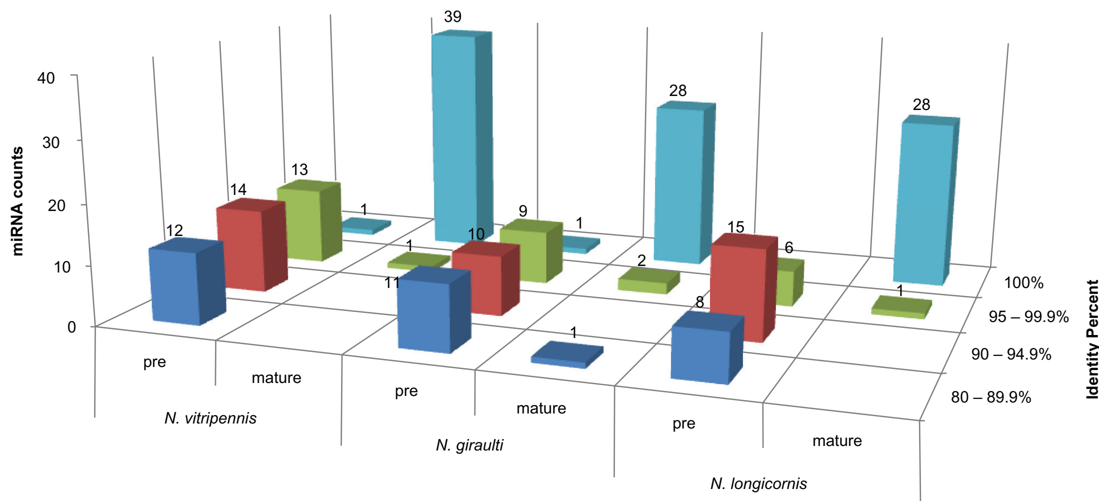

Figure 2a. Count of pre- and mature miRNAs in *N. vitripennis, N. giraulti*, and *N. longicornis* with Identity percent to query sequences.

between 90%–94.9% and 11 pre-miRNAs between 80%–89.9%. While in *N. longicornis* we obtained 29 putative pre-miRNAs and interestingly there were no 100% identity miRNAs, 6 pre-miRNAs between 95%–99.9%, 15 between 90%–94.9% and 8 between 80%–89.9% to their query sequences (Fig. 2a). It is quite interesting to note that most of the query sequences after complete screening were mostly from *Apis mellifera*—95.12% (*N. vitripennis*), 96.77% (*N. giraulti*), 100%,[40] (Fig. 2b) suggesting their phylogenetic closeness. They both belong to same Order Hymenoptera and sub order Apocrita. The organisms of this order have membranous wings, which is a unique characteristic. Further all these sequences have 100% conserved mature miRNA sequences.

At mature miRNA level very interestingly, of 40 miRNAs, 39 pre-miRNAs though they had varying identity between 80%–99.9% identity; 38 of them had 100% identity at mature sequence level, showing conservation of sequences at mature sequence level during evolution. In *N. giraulti* of 31 miRs obtained, 28 of them had no change in the mature sequences compared to their query sequences. While miR-281 has 95.24% of mature identity (Supplementary Table 2) and miR-100 had 81% mature identity with 4 mismatches (Supplementary Table 2, Figure 3a) at mature sequence level. *N. longicornis* was more interesting to analyze than the other two species as at pre-miRNA level there was no 100% identity sequence, while at mature sequence level 28 of 29 sequences had 100%

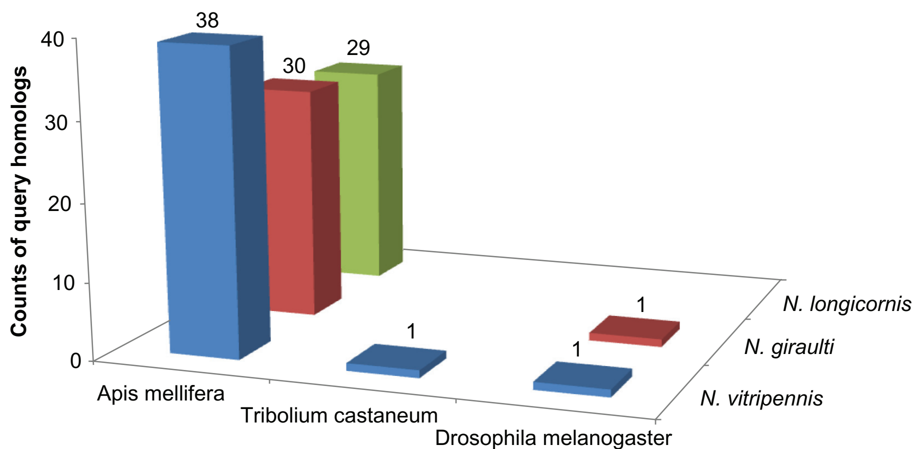

Figure 2b. Total miRNA query counts with maximum homolog to the Nasonia species.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

Table 1. Cross species comparison of pre- and mature miRNAs sequences of *N. vitripennis, N. longicornis, N. giraulti*.

N. vitripennis	*N. giraulti*	*N. longicornis*	Pre % Identity	Mature % Identity
Bantam			×	×
let-7	let-7	let-7	100	100
miR-1	miR-1		100	100
miR-10		miR-10	100	100
miR-100	miR-100		92.3	75
miR-12		miR-12	100	100
miR-124			×	×
miR-125		miR-125	100	100
miR-133	miR-133	miR-133	100	100
miR-137	miR-137	miR-137	100	100
miR-13a	miR-13a	miR-13a	98.8	100
miR-14	miR-14		100	100
miR-184	miR-184	miR-184	100	100
miR-210	miR-210		97.6	100
miR-219	miR-219	miR-219	100	100
miR-2-3	miR-2-3	miR-2-3	100	100
miR-252	miR-252	miR-252	98.9	100
miR-263b	miR-263b	miR-263b	100	100
miR-275	miR-275	miR-275	100	100
miR-276	miR-276	miR-276	98.8	100
miR-277	miR-277	miR-277	100	100
miR-279			×	×
miR-281	miR-28l	miR-281	100	100
miR-282		miR-282	100	100
miR-283	miR-283		100	100
miR-29b	miR-29b		100	100
miR-305	miR-305	miR-305	100	100
miR-315		miR-315	100	100
miR-317	miR-317		98.9	100
miR-31s	miR-31a	miR-31a	100	100
miR-7		miR-7	100	100
miR-8		miR-8	100	100
miR-927	miR-927	miR-927	95.7	100
miR-929	miR-929	miR-929	100	100
miR-92a	miR-92a	miR-92a	96.8	100
miR-932	miR-932	miR-932	100	100
miR-993	miR-993	miR-993	92.1	100
miR-9a	miR-9a	miR-9a	100	100
miR-iab-4	miR-iab-4		100	100
miR-iab-4as	miR-iab-4as		100	100
	miR-34	miR-34	100	100
	miR-375		×	×

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

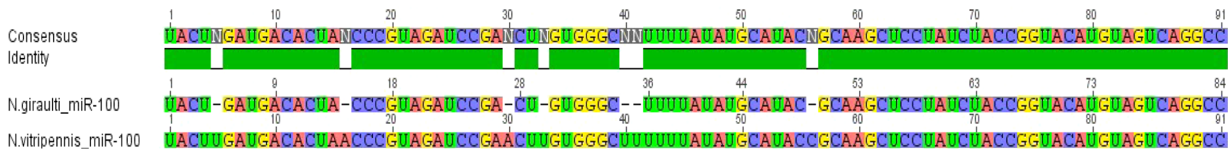

Figure 3a. Pairwise alignment of pre-mir-100 of *N. giraulti* and *N. vitripennis*. Sequence has 92.3% and 75% identity at pre- and mature miRNA respectively.

mature identity and again miR-281 had 95.24% identity to their query sequences, making this miRNA sequence genus specific. A cross species comparison of Identity percent of miRNA sequences at pre- and mature sequence level is determined (Table 1). The pre-miRNA sequence analysis gives a clear picture that there has been divergence during speciation of Nasonia.

Sequence Analysis

We looked upon the effect of sequences on their secondary structure and stability during their divergence into speciation. At pre-miRNA level except for few miRNAs such as mir-100, mir-13a, mir-210, mir-252, mir-276, mir-317, mir-92a, mir-927, and mir-993 others have 100% identity (Table 1). In *N. giraulti* and *N. vitripennis* putative mir-100 had only 92.3% identity at pre-miRNA level and 75% Identity at mature miRNA (Table 1), yet the hairpin structure between the species is almost conserved (Fig. 3a, 3b) the minimal variation between the structures is accounted for mismatches and gaps between the sequences. Also, mir-100 is probably lost during divergence of *N. longicornis* or could be due to differences in genome sequence coverage. mir-993 is another example for divergence among the Nasonia species which has affected both the secondary structure and MFE. Other extreme divergent example among the species is putative mir-252 at 46th position has uracil instead of cytosine, which has affected the secondary structure (Fig. 4a, 4b) and hence their free energy at very minimal difference, but the mature sequence remains unaltered. It is interesting to see mir-317 (Fig. 5a, 5b), though there is a single base mismatch (A-G) at 38th position, it has affected neither the secondary structure nor the MFE (minimal folding energy).[41] Few other examples wherein a single base has not affected the secondary structure or MFE are mir-13a, mir-276, mir-92a (Supplementary Figure 1, 2, 3). Whilst of these divergence among the species, the seed region sequences have been highly conversed with all the miRNAs, maintaining their functional integrity in the genome.

Previously it has been demonstrated that compared to other noncoding RNAs, pre-miRNAs have low MFE.[41] Therefore, MFE was considered as one of the important factors to identify miRNA genes.[42,43] We have also compared the values of the MFE, AMFE (adjusted minimal folding energy) and MFEI (minimal folding energy index). Even though the identity values of few above mentioned miRNAs have diverged during divergence among the species, amazingly, their AMFE and MFEI values are greatly conversed except in very few cases such as mir-276, mir-993 making them species specific (Table 2). We find that the MFEI values for *N. vitripennis*, *N. giraulti* and *N. longicornis* pre-miRNA's average being -0.86 ± 0.12, -0.85 ± 0.11 and -0.85 ± 0.11, suggesting that

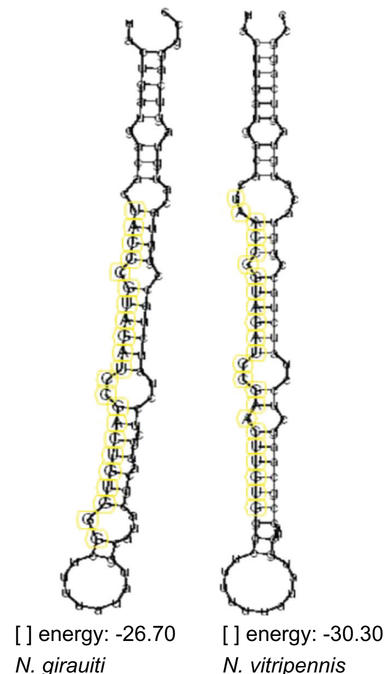

Figure 3b. Comparison of Secondary structure comparison of mir-100 between *N. giraulti* and *N. vitripennis.* Structure is drawn using m-FOLD. The structure is nearly conserved.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

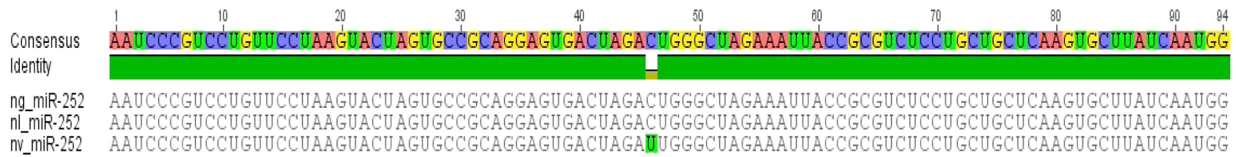

Figure 4a. Pairwise alignment of pre-mir-252 between Nasonia species. Nv = *N. vitripennis;* Ng = *N. giraulti;* Nl = *N. longicornis.* The sequence is almost conserved expect for 46th position in *N. vitripennis*.

though there are divergence mismatches or gaps in sequence, yet the MFEI have been highly conserved between the species. Altogether, these features could play a significant role in further understanding the evolutionary pattern of the pre-miRNAs among the species.

Statistical sequence characteristics

Sequence characteristics of pre- and mature miRNAs are reported in plants and very recently in animals.[9,19,34,44] We performed a detailed analysis of sequence characteristics of *N. vitripennis*, *N. longicornis* and *N. giraulti* putative pre-miRNA sequences and all known invertebrate miRNA sequences (Table 3). Our study shows that the length of all three Nasonia species pre-miRNAs varies between 41 to 100 nucleotides with an average of ~83 ± ~10-nts for all three species together, which is quite consistent with the other known invertebrate sequences, particularly within class insecta. The average base composition

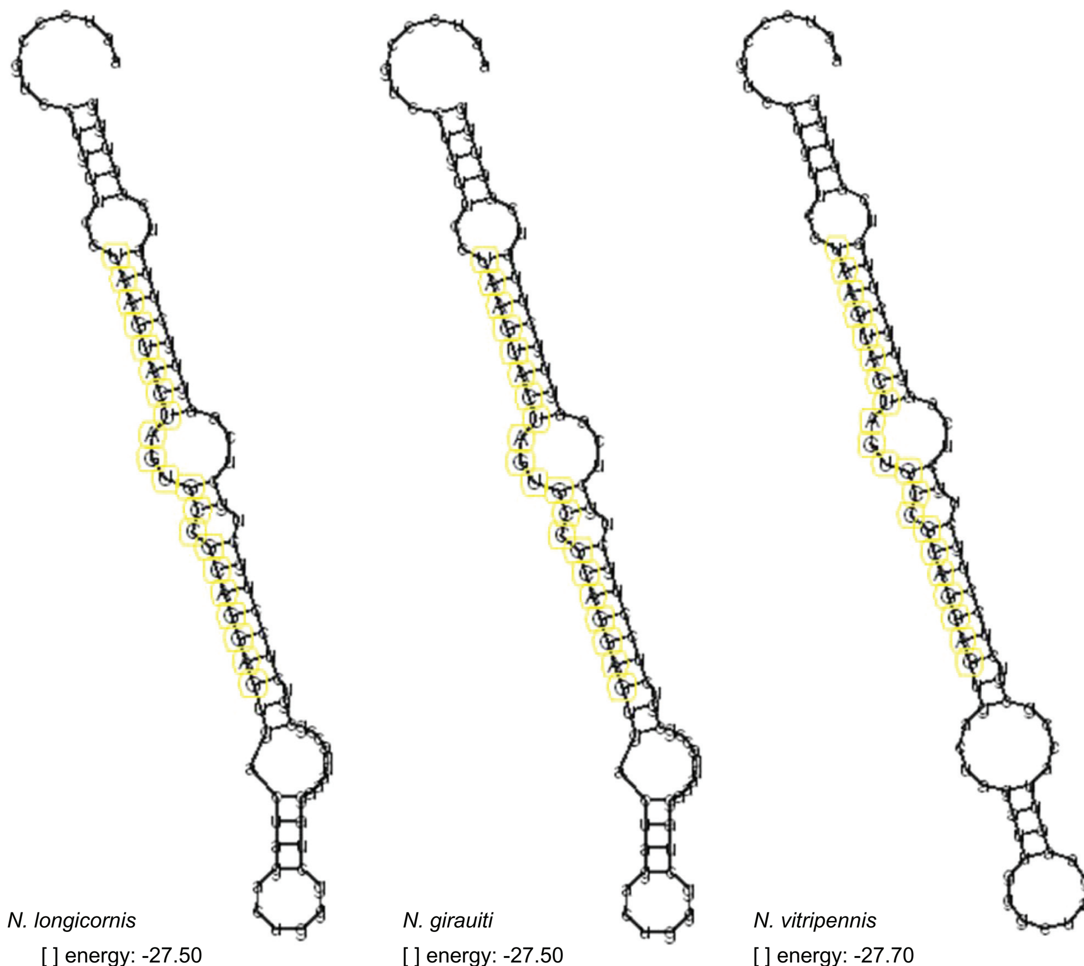

Figure 4b. Comparison of Secondary structure comparison of mir-252 between Nasonia species. Structure is drawn using m-FOLD. Uracil at 46th position has altered one of the bulge near the loop region in *N. vitripennis* and is accounted in mfe value.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

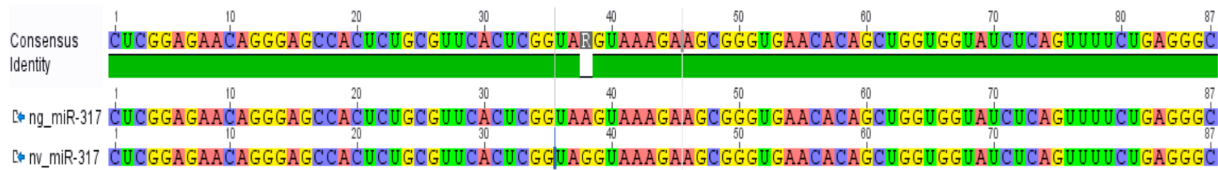

Figure 5a. Pairwise alignment of pre-mir-317 between two Nasonia species. Nv = *N. vitripennis;* Ng = *N. giraulti*.

of pre-miRNA sequences in all three Nasonia species and other invertebrates are almost identical (Table 3). The average frequency of G + C% in Nasonia species is higher than other invertebrates, which is quite unique observation for this genus. Due to the high frequency of G + C, the base frequency of A + U is generally found much less in Nasonia. It is well known that the A + U composition decreases the stability of the pre-miRNA secondary structure,[39] yet in Nasonia we find a totally different observation from what is been previously observed in other animals. Such a unique observation has not been reported so far. This high frequency of G + C, could be unique to genus, differing from other close relatives. Further, we find that G/C and U/A ratio for Nasonia (average of all 3 species) and invertebrates are 1.16, 0.81 and 1.11, 0.79 respectively. We further analyzed the frequency of nucleotides at each position in the mature miRNA of all Nasonia species.

Earlier studies have shown that U is the predominant nucleotide at 5′ end of the mature miRNA in plants. Based on this it has been proposed that the 5′ end may play an important role in biogenesis of mature miRNA through recognition of the targeted miRNA precursors by RISC.[39] Consistent to this, our studies have shown that uracil is present predominantly at 5′ end of the mature miRNAs in all three species of Nasonia (Fig. 6a). Therefore, our study suggests that there are a certain degree of similarity at sequence level between plant and Nasonia mature miRNAs at a distant view, in addition to other invertebrate miRNA sequences. We further find apart for 1st position in mature miRNA, U is predominant at positions 9, 17, 20, and least at 7, 14, 22 positions (Fig. 6b). In plants, cytosine is the dominant nucleotide at position 19; however that is not observed in animals.[39] Instead we find in the case of Uracil in Nasonia is dominant at the same position (average of 3 species 33.6%). These positional differences in miRNA sequences could affect their miRNA:mRNA target binding energy and making Nasonia unique from other related species. Also, we find that G + C frequency is higher than A + U at position 3. All these sequence characteristics features could play an important role in further understanding the evolution of miRNAs and their potential target sequences.

miRNA Target prediction

miRNA targets are usually located in the 3`-UTR region of mRNAs. These UTRs have already been recognized as an important regulatory region even

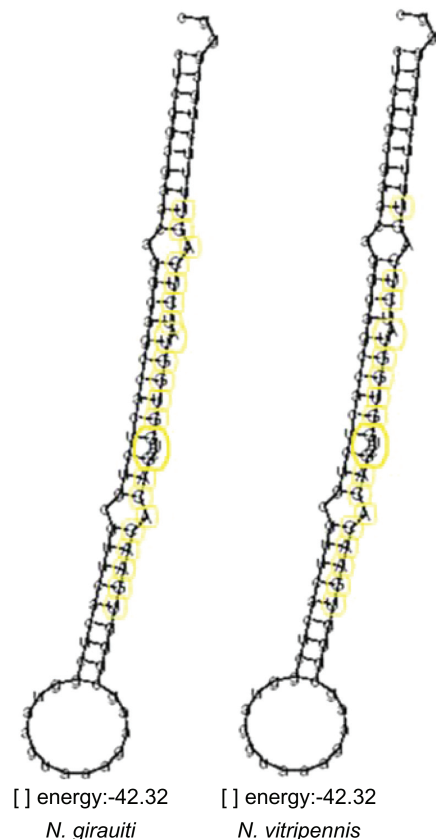

Figure 5b. Comparison of Secondary structure comparison of mir-317 between two Nasonia species. The difference in base nucleotide at 38th position, has neither affected the structure nor mfe value.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

Table 2. Comparative list of MFE, AMFE and MFEI of *N. vitripennis, N. longicornis, N. giraulti.*

miRNA	Nasonia vitripennis			Nasonia giraulti			Nasonia longicornis		
	MFE	AMFE	MFEI	MFE	AMFE	MFEI	MFE	AMFE	MFEI
let-7	–49.2	–48.71	–0.99	–49.2	–48.71	–0.99	–49.2	–48.71	–0.99
mir-1	–26.4	–38.26	–0.75	–26.4	–38.26	–0.75			
bantam	–30.1	–40.68	–1.20						
mir-10	–39	–37.86	–0.86				–39	–37.36	–0.86
mir-100	–30.3	–33.30	–0.72	–26.7	–29.34	–0.60			
mir-133	–34.9	–39.21	–0.92	–34.9	–39.21	–0.92	–34.9	–39.21	–0.92
mir-12	–23.3	–35.85	–0.88				–23.3	–35.85	–0.88
mir-124	–33.3	–43.25	–0.85						
mir-125	–41.3	–40.49	–0.71				–41.3	–40.49	–0.71
mir-137	–33.7	–40.60	–0.82	–33.7	–40.60	–0.82	–33.7	–40.60	–0.82
mir-13a	–34.8	–40.94	–0.92	–34.8	–40.94	–0.92	–34.8	–40.94	–0.94
mir-14	–32.9	–39.17	–0.75	–32.9	–39.17	–0.75			
mir-184	–34.2	–37.58	–0.76	–34.2	–37.58	–0.76	–34.2	–37.58	–0.76
mir-210	–40.52	–48.24	–0.86	–43.22	–50.85	–0.90			
mir-219	–34	–41.46	–0.84	–34	–41.46	–0.84	–34	–41.46	–0.84
mir-2-3	–29.4	–37.69	–0.84	–29.4	–37.69	–0.84	–29.4	–37.69	–0.84
mir-252	–27.7	–29.47	–0.59	–27.5	–29.26	–0.57	–27.5	–29.26	–0.57
mir-263b	–38.3	–44.53	–0.85	–38.3	–44.53	–0.85	–38.3	–44.53	–0.85
mir-275	–34.3	–44.55	–0.77	–34.3	–44.55	–0.77	–34.3	–44.55	–0.77
mir-276	–38.34	–45.11	–0.95	–38.34	–45.11	–0.95	–31.94	–37.58	–0.79
mir-277	–33.89	–39.87	–0.72	–33.89	–39.87	–0.72	–33.39	–39.87	–0.72
mir-231	–32.6	–44.05	–0.97	–32.6	–44.05	–0.97	–32.6	–44.05	–0.97
mir-279	–24.5	–37.69	–0.83						
mir-282	–30.6	–37.32	–0.80				–30.6	–37.32	–0.80
mir-283	–26.56	–33.20	–0.75	–26.56	–33.20	–0.75			
mir-29b	–36.3	–37.81	–0.81	–36.3	–37.31	–0.81			
mir-305	–33.7	–38.30	–0.80	–33.7	–38.30	–0.80	–33.7	–38.30	–0.80
mir-315	–28.7	–37.76	–0.89				–28.7	–37.76	–0.89
mir-317	–42.32	–48.64	–0.88	–42.32	–48.64	–0.90			
mir-31a	–27.72	–37.46	–0.75	–27.72	–37.46	–0.75	–27.72	–37.46	–0.75
mir-34				–30.5	–37.20	–0.73	–30.5	–37.20	–0.73
mir-7	–37.3	–43.37	–1.10				–37.3	–43.37	–1.10
mir-8	–31.44	–36.14	–0.94				–31.44	–36.14	–0.94
mir-375				–20	–27.78	–0.69			
mir-927	–23.9	–31.08	–0.99	–29.2	–32.09	–1.00	–28.9	–31.76	–0.99
mir-929	–44.8	–47.16	–1.02	–44.8	–47.16	–1.02	–44.8	–47.16	–1.02
mir-92a	–32.4	–33.75	–0.78	–32.4	–33.75	–0.81	–32.4	–33.75	–0.80
mir-932	–41.1	–41.94	–0.82	–41.1	–41.94	–0.82	–41.1	–41.94	–0.82
mir-993	–37.86	–42.54	–0.95	–30.4	–34.55	–0.75	–29.76	–33.44	–0.74
mir-9a	–40	–44.94	–1.00	–40	–44.94	–1.00	–40	–44.94	–1.00
mir-iab-4	–33.8	–39.76	–0.89	–33.8	–39.76	–0.89			
mir-iab-4as	–20.9	–34.83	–0.91	–20.9	–34.83	–0.91			

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

Table 3. Comparison of Statistical sequence characters between all Nasonia species and all invertebrates.

Organism	Length	A%	U%	G%	C%	A + U%	G + C%	
N. vitripennis	60.00	14.29	21.84	17.39	12.16	42.25	31.52	Minimal
	100.00	38.04	40.63	33.33	29.58	68.48	57.75	Maximal
	83.45	23.88	29.36	25.02	21.75	53.24	46.76	Mean
	9.76	4.51	4.14	3.55	3.72	5.84	5.84	SD
N. giraulti	60.00	14.29	21.84	17.78	14.44	42.25	32.22	Minimal
	95.00	35.00	34.48	33.75	29.58	67.78	57.75	Maximal
	83.03	23.95	28.40	25.37	22.28	52.35	47.65	Mean
	8.28	4.26	3.72	3.99	3.30	5.44	5.44	SD
N. longicomis	41.00	15.00	22.54	17.78	9.76	42.25	32.22	Minimal
	100.00	36.67	40.63	46.34	29.58	67.78	57.75	Maximal
	83.14	23.45	29.51	25.88	21.16	52.96	47.04	Mean
	11.66	4.24	4.15	5.31	4.26	5.87	5.87	SD
Invertebrates	49.00	11.88	15.69	9.68	7.69	31.40	19.32	Minimal
	215.00	44.29	46.15	40.00	37.21	80.68	68.60	Maximal
	90.00	25.00	31.52	22.33	20.00	57.73	42.27	Mean
	13.82	4.58	4.91	3.80	4.29	6.85	6.85	SD

before the discovery of miRNAs, due to the presence of numerous regulatory signals involved in the control of nuclear export, subcellular localization, transcript stability amongst other processes.[45] Additionally, this regions frequently contains multiple target sites for more than one miRNA to interact.[45] However, it is been shown that target sequences inserted in the coding or 5′ UTRs can also be functional.

It is well known that animal miRNA targets are difficult to predict, unlike plant targets since miRNA:

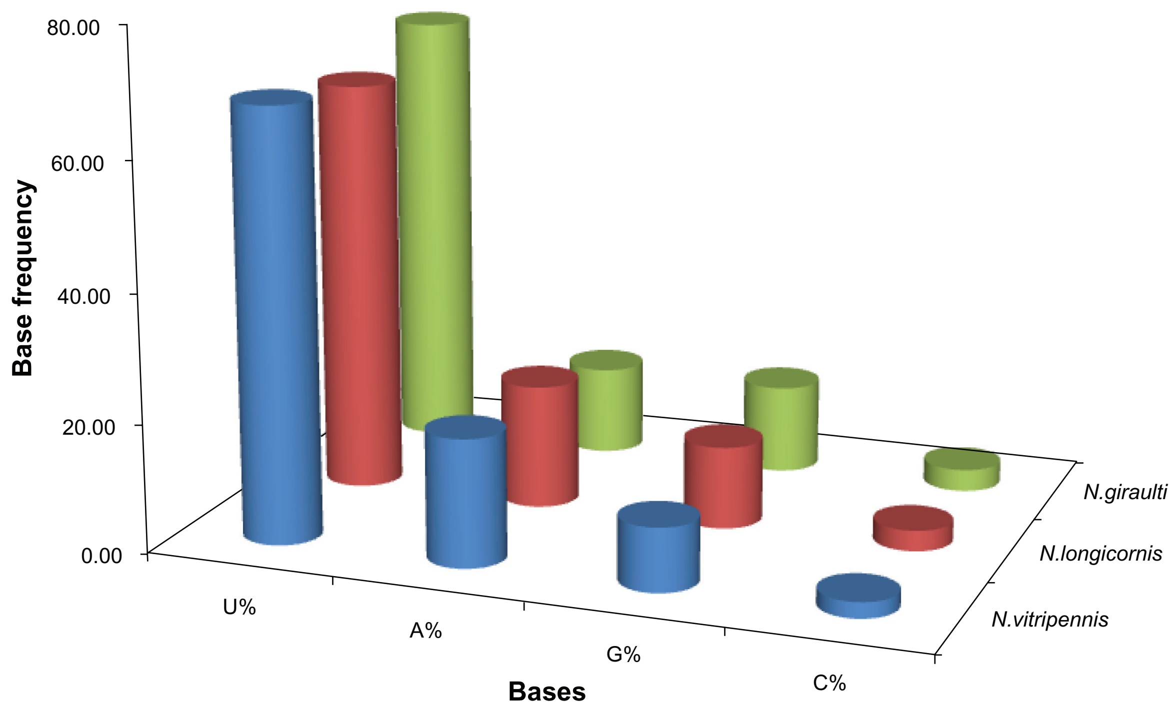

Figure 6a. Base frequency at 5`end of mature miRNA in all Nasonia species.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

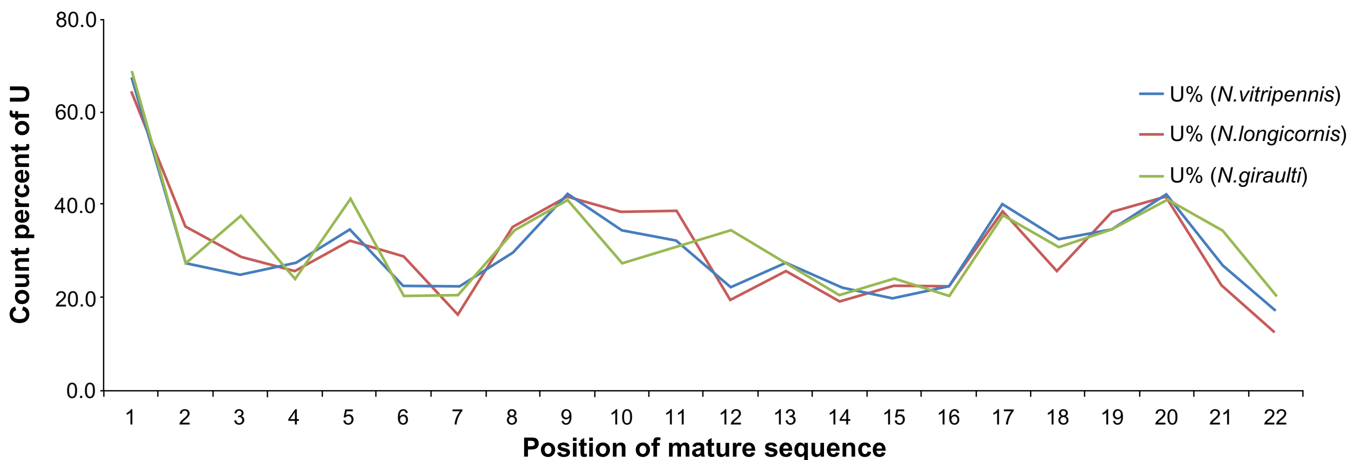

Figure 6b. Distribution of U frequency along the length of mature miRNA sequence in all Nasonia species.

mRNA duplexes often contain several mis-matches, gaps and G:U base pairs in many positions.[45] However, it is increasingly recognized that near-perfect complementarity between a few bases at the 5` end of miRNA and the 3` UTR targets is instrumental in metazoan target site recognition, these sites are referred as "seed sites/sequences".[23,46] In our study we use MiRanda algorithm,[32] which encompasses the thermodynamic stability of miRNA:mRNA duplex as one of the entity in detecting the potential binding site on the 3`UTRs. As we have seen although the pre-miRNAs sequences were greatly varying between the query sequences and the hit sequences, the mature sequences were almost conserved. We looked upon the conservation of the mature sequences with the well studied invertebrate model *D. melanogaster*, as predicted atleast 80% of the mature sequence in all three Nasonia species were 100% conserved (Fig. 7). This conservation gives an overwhelming opinion that they are functionally conserved and hence their target sites to a certain extent. Thus we extracted all the 3`UTRs from the *N. vitripennis* species, converted into FASTA format and further used in the MiRanda algorithm. We thus identify 471 poteinal target sites in 46 GIs (Gene IDs) (Supplementary Table 4). To reduce the false positive binding sites we used cut off value of −14 kcal/mol and score value of 80. As the *N. giraulti* and *N. longicornis* sequences were from trace archived and the sequences are not completed, it could be too early to predict the potential targets at this point of time as it may lead to many false positive hits. Further, experimental evidences are required to validate these targets in *in vivo* conditions which are beyond the scope of our study. Thus, we consider that

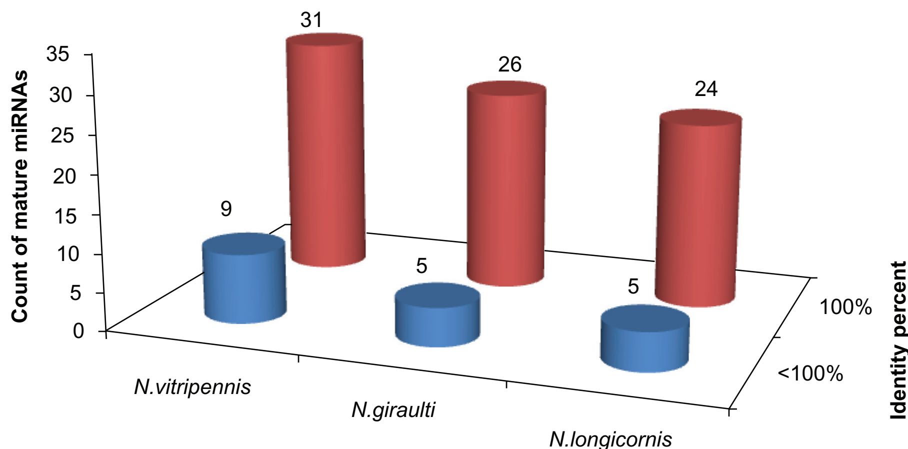

Figure 7. Identity percent of mature miRNA sequences with *D. melanogaster*.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

our study leads to the beginning of understanding the Nasonia's miRNA integrity.

Conclusion

In this study, we report 40, 31 and 29 putative miRNAs in *N. vitripennis*, *N. giraulti* and *N. longicornis* respectively, by modifying the existing computational methods using all reported invertebrate miRNA genes. It is important to note that *N. vitripennis* (6.2x) has higher genome sequence coverage than the other two species (1x), which will influence the number of miRNAs reported and could change with the genome coverage. Sequences with less than 85% have been validated using miPred. The statistical sequence characteristics of the putative sequences are found to be quite consistent with other invertebrate sequences. Further, we find this integrated method of detecting putative miRNA genes has been reasonable as we started with all invertebrate miRNA genes containing reported miRNA genes from majority of invertebrate phyla, after iterations, we find that most of the query sequences come from *Apis mellifera*, the closest organism in evolution. We did not find any close homolog miRNA genes from other query phyla used. The positional differences of bases in the sequences could make these miRNAs unique for Nasonia. Among the putative hits, although we find some of the pre-miRNA sequences are not 100% homolog, yet the 100% conservation of mature sequences reveals the functional domains are probably unchanged during the evolution of Nasonia genome. Further insights to this conservation the miRNA targets and find that these targets are again conserved during evolution when compared to the most well studied organism *D. melanogaster*. It may be too early to determine the miRNA clusters as the assembly of the Nasonia genomes are still under progress. As we have basically used comparative analysis for detecting homolog miRNAs, there could be more miRNAs yet to be identified in the genome, which could have aroused after speciation and remains unidentified in our analysis. We are very sure that miRNAs could be probably used in the potential pest management for controlling of insects in agricultural fields. This method could make the limited use of chemical active ingredients on the crops, thereby reducing the pollution and ecological imbalances caused in recent years. We hope that this paper could be a start in understanding of miRNAs in the Nasonia genomes. There could be many more interesting evolutionary features yet to be discovered when we start looking at the miRNA target sites.

Acknowledgement

We extend our thanks to BCM's (Bayer College of Medicine) Nasonia Genome Project group and all other groups associated with Nasonia genome sequencing project for sequencing the genome and making available for the public.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors report no conflicts of interest.

References

1. Kloosterman WP, Plasterk RH. The diverse functions of microRNAs in animal development and disease. *Dev Cell*. 2006;11(4):441–50.
2. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6(11):857–66.
3. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer*. 2006;6(4):259–69.
4. Lu LF, Liston A. MicroRNA in the immune system, microRNA as an immune system. *Immunology*. 2009;127(3):291–8.
5. Hertel J, et al. The expansion of the metazoan microRNA repertoire. *BMC Genomics*. 2006;7:25.
6. Lagos-Quintana M, et al. Identification of novel genes coding for small expressed RNAs. *Science*. 2001;294(5543):853–8.
7. Lau NC, et al. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*. 2001;294(5543):858–62.
8. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. *Science*. 2001;294(5543):862–4.
9. Zhang B, Stellwag EJ, Pan X. Large-scale genome analysis reveals unique features of microRNAs. *Gene*. 2009;443(1–2): 100–9.
10. Palakodeti D, Smielewska M, Graveley BR. MicroRNAs from the Planarian Schmidtea mediterranea: a model system for stem cell biology. *Rna*. 2006;12(9):1640–9.
11. Fu X, Adamski M, Thompson EM. Altered miRNA repertoire in the simplified chordate, Oikopleura dioica. *Mol Biol Evol*. 2008;25(6):1067–80.
12. Zhang B, Farwell MA. microRNAs: a new emerging class of players for disease diagnostics and gene therapy. *J Cell Mol Med*. 2008;12(1):3–21.
13. Griffiths-Jones S, et al. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34 (Database issue):D140–4.
14. Griffiths-Jones S. The microRNA Registry. *Nucleic Acids Res*. 2004;32(Database issue):D109–11.
15. Ambros V, et al. A uniform system for microRNA annotation. *Rna*. 2003; 9(3):277–9.
16. Arteaga-Vazquez M, Caballero-Perez J, Vielle-Calzada JP. A family of microRNAs present in plants and animals. *Plant Cell*. 2006;18(12): 3355–69.
17. Xie FL, et al. Computational identification of novel microRNAs and targets in Brassica napus. *FEBS Lett*. 2007;581(7):1464–74.
18. Sunkar R, Jagadeeswaran G. In silico identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol*. 2008; 8:37.

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use

19. Zhou M, et al. In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. *Genomics*. 2009;94(2):125–31.
20. Baev V, Daskalova E, Minkov I. Computational identification of novel microRNA homologs in the chimpanzee genome. *Comput Biol Chem*. 2009;33(1):62–70.
21. Zhang B, Pan X, Stellwag EJ. Identification of soybean microRNAs and their targets. *Planta*. 2008;229(1):161–82.
22. Lindow M, Gorodkin J. Principles and limitations of computational microRNA gene and target finding. *DNA Cell Biol*. 2007;26(5):339–51.
23. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281–97.
24. Lim LP, et al. The microRNAs of Caenorhabditis elegans. *Genes Dev*. 2003;17(8):991–1008.
25. Lai EC, et al. Computational identification of Drosophila microRNA genes. *Genome Biol*. 2003;4(7):R42.
26. Berezikov E, et al. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*. 2005;120(1):21–4.
27. Lindow M, et al. Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS Comput Biol*. 2007;3(11):e238.
28. Stark A, et al. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. *Genome Res*. 2007;17(12): p. 1865–79.
29. Pfeffer S, et al. Identification of microRNAs of the herpesvirus family. *Nat Methods*. 2005;2(4):269–76.
30. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna*. 2005;11(3):241–7.
31. Savard J, et al. Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res*. 2006;16(11):1334–8.
32. Enright AJ, et al. MicroRNA targets in Drosophila. *Genome Biol*. 2003;5:R1.
33. John B, et al. Human MicroRNA targets. *PLoS Biol*. 2004;2(11):e363.
34. Griffiths-Jones S, et al. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008;36(Database issue):D154–8.
35. Altschul SF, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215(3): 403–10.
36. Durand P, Canard L, Mornon JP. Visual BLAST and visual FASTA: graphic workbenches for interactive analysis of full BLAST and FASTA outputs under MICROSOFT WINDOWS 95/NT. *Comput Appl Biosci*. 1997;13(4): 407–13.
37. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406–15.
38. Mathews DH, et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*. 1999;288(5):911–40.
39. Zhang B, et al. Conservation and divergence of plant microRNA genes. *Plant J*. 2006;46(2):243–59.
40. Deloukas P, et al. The DNA sequence and comparative analysis of human chromosome 10. *Nature*. 2004;429(6990):375–81.
41. Zhang BH, et al. Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci*. 2006;63(2):246–54.
42. Jiang P, et al. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Res*. 2007;35(Web Server issue):W339–44.
43. Wang X, et al. MicroRNA identification based on sequence and structure alignment. *Bioinformatics*. 2005;21(18):3610–4.
44. Bonnet E, et al. Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. *Proc Natl Acad Sci U S A*. 2004;101(31):11511–6.
45. Stark A, et al. Identification of Drosophila MicroRNA targets. *PLoS Biol*. 2003;1(3):E60.
46. Doench JG, Sharp JA. Specificity of microRNA target selection in translational repression. *Genes Dev*. 2004;18:504–11.

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

http://www.la-press.com

Downloaded From: https://bioone.org/journals/International-Journal-of-Insect-Science on 25 Sep 2024
Terms of Use: https://bioone.org/terms-of-use