Calcino et al. BMC Evolutionary Biology (2018) 18:160
https://doi.org/10.1186/s12862-018-1274-2

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Diverse RNA interference strategies in early-branching metazoans

Andrew D. Calcino[1,4], Selene L. Fernandez-Valverde[1,5], Ryan J. Taft[2,3] and Bernard M. Degnan[1*]

Abstract

Background: Micro RNAs (miRNAs) and piwi interacting RNAs (piRNAs), along with the more ancient eukaryotic endogenous small interfering RNAs (endo-siRNAs) constitute the principal components of the RNA interference (RNAi) repertoire of most animals. RNAi in non-bilaterians – sponges, ctenophores, placozoans and cnidarians - appears to be more diverse than that of bilaterians, and includes structurally variable miRNAs in sponges, an enormous number of piRNAs in cnidarians and the absence of miRNAs in ctenophores and placozoans.

Results: Here we identify thousands of endo-siRNAs and piRNAs from the sponge *Amphimedon queenslandica*, the ctenophore *Mnemiopsis leidyi* and the cnidarian *Nematostella vectensis* using a computational approach that clusters mapped small RNA sequences and annotates each cluster based on the read length and relative abundance of the constituent reads. This approach was validated on 11 small RNA libraries in *Drosophila melanogaster*, demonstrating the successful annotation of RNAi-associated loci with properties consistent with previous reports. In the non-bilaterians we uncover seven new miRNAs from *Amphimedon* and four from *Nematostella* as well as sub-populations of candidate cis-natural antisense transcript (cis-NAT) endo-siRNAs. We confirmed the absence of miRNAs in *Mnemiopsis* but detected an abundance of endo-siRNAs in this ctenophore. Analysis of putative piRNA structure suggests that conserved localised secondary structures in primary transcripts may be important for the production of mature piRNAs in *Amphimedon* and *Nematostella*, as is also the case for endo-siRNAs.

Conclusion: Together, these findings suggest that the last common ancestor of extant animals did not have the entrained RNAi system that typifies bilaterians. Instead it appears that bilaterians, cnidarians, ctenophores and sponges express unique repertoires and combinations of miRNAs, piRNAs and endo-siRNAs.

Keywords: RNAi, miRNA, Endo-siRNA, piRNA, Non-bilaterian, Demosponge, Cnidarian, Ctenophore

Background

RNA interference (RNAi) evolved prior to the divergence of extant eukaryotic lineages, possibly in response to threats from parasitic double-stranded RNA species such as retroviruses and transposons [1]. In contemporary animals, three independent RNAi systems comprise the bulk of the small RNA (sRNA) repertoire: micro RNAs (miRNAs); Piwi interacting RNAs (piRNAs); and endogenous small interfering RNAs (endo-siRNAs). Amongst non-bilaterian animals - sponges, cnidarians ctenophores and placozoans - miRNAs appear to be lost in placozoans and ctenophores with these lineages also lacking key miRNA biogenic enzymes [2–4]. The absence of miRNAs in the sister lineages to the animal kingdom - choanoflagellates and other unicellular holozoans - and fungi [2, 5, 6], suggests the miRNA system has either been lost or evolved independently multiple times [7]. Nonetheless, animal miRNAs play fundamental roles in cell type differentiation and maintenance, and their emergence and proliferation is linked to the evolution of complex multicellularity [8]. The prevalence of miRNAs in plants and algae [9] lends further support to the hypothesis that miRNAs may be important regulators of multicellular development. However, miRNAs do not appear to be essential for animal multicellularity given they are missing from the morphologically complex non-bilaterian metazoans, the ctenophores [3, 4].

There are some marked differences in the miRNA systems of sponges, cnidarians and bilaterians. In contrast to bilaterians, which express a complex repertoire of

* Correspondence: b.degnan@uq.edu.au
[1]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia
Full list of author information is available at the end of the article

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

miRNAs in somatic tissues [10–16], miRNA expression in cnidarians is consistently dwarfed by piRNAs [2, 17–21]. The miRNAs of the cnidarian *Nematostella vectensis*, while capable of bilaterian-like silencing by transcript destabilisation or translational inhibition [22], also regularly silence their targets through extensive base pairing followed by cleavage, as observed in plants [19]. The miRNA repertoire in sponges is substantially less than in cnidarians and bilaterians with only eight, eleven and nineteen currently reported from the demosponges *Amphimedon queenslandica*, *Stylissa carteri* and *Xestospongia testudinaria,* respectively [2, 23]. In *Amphimedon*, these differ from other metazoan miRNAs in having a peculiar plant-like pre-miRNA secondary structure, and have no discernible homology with any animal miRNAs, except those found in in other demosponges [2, 24–27].

Although questions about miRNA evolution in animals remain unresolved, the presence and roles of endo-siRNA and piRNA systems in non-bilaterian metazoans have received less attention. The endo-siRNA pathway is an ancient eukaryotic feature and was likely to have been present in the last common ancestor of extant metazoans [28, 29], although the complex repertoire of endo-siRNAs in most non-bilaterians has yet to be fully documented. In contrast, piRNAs appear to be a metazoan innovation, being present in sponges and cnidarians but not in placozoans [2]; piRNAs have not been studied in ctenophores. A functional PIWI-piRNA pathway is present in *Hydra, Nematostella* and the anemone *Anemonia viridis* [30–32].

Given the apparent diversity of RNAi systems amongst representatives of non-bilaterian phyletic lineages, we developed an in silico approach to detail the sRNA components in representatives of these lineages with small RNA libraries and assembled genomes (i.e., *Amphimedon*, *Nematostella* and the ctenophore *Mnemiopsis leidyi*). We first confirmed the efficacy and accuracy of this approach on 11 well-annotated developmental small RNA libraries from *Drosophila melanogaster.* When applied to the non-bilaterians, this approach identified novel miRNAs, piRNAs and endo-siRNAs and revealed that *Amphimedon*, *Mnemiopsis* and *Nematostella* have markedly different RNAi repertoires from each other and from bilaterians.

Results

The uniformity index as a tool for discriminating RNAi classes

To investigate the sRNA repertoires of *Amphimedon, Mnemiopsis* and *Nematostella*, we developed a method for the annotation of putative precursor transcripts of endo-siRNAs, piRNAs and miRNAs based on Illumina sequenced small RNA libraries (see Methods). This method leverages on the fact that the biogenesis of miRNAs reliably produces sRNAs of a predictable length and sequence [33].

Variation around the most abundant reads within a cluster of a miRNA loci is limited, leading to large numbers of sRNA reads exhibiting low sequence diversity. In contrast, without the guidance of binding partners involved in miRNA production, Dicer cleaves dsRNA with less discrimination, producing endo-siRNAs of a regular length, typically 21–22 nucleotides (nts), but with far greater sequence variability [34–42]. As a consequence, endo-siRNAs loci typically generate a higher diversity of sRNAs that are lower in relative abundance compared to miRNA loci. Likewise, piRNA biogenesis involves limited specificity over the 5′ and 3′ ends produced by the catalytic components of the pathway, resulting in a highly diverse population of piRNAs generally 26–30 nt in length arising from each loci [43–47]. Cluster diversity can be further increased by posttranscriptional modifications such as the trimming and tailing of sRNAs by TUTases and nucleases such as Nibbler [48–51].

The uniformity of sRNA reads comprising a given cluster can be measured by what we term the *uniformity index* - the ratio of the total abundance of sRNA reads comprising a cluster and the number of distinct sRNA reads from that same cluster. For example, a miRNA-like hairpin comprised of 16 counts but only three distinct reads results in a uniformity index of (16/3) or 5.3 while an endo-siRNA like hairpin comprised of 16 counts comprising 12 distinct reads results in a uniformity index of (16/12) or 1.3 (Additional file 1). Calculating this index for each sRNA cluster enables segregation of high uniformity (HU) clusters (such as miRNAs) from low uniformity (LU) endo-siRNA and piRNA clusters, as we demonstrate in *Drosophila*. Amongst the segregated HU clusters are repetitive sequences as well as miRNAs and other biologically significant sRNA clusters which can be secondarily annotated. Increasing library depth results in increasing UI values, in particular for HU clusters (Additional file 2a), however if cluster UI comparisons between libraries are required, dividing the UI by library depth can normalise the UI (Additional file 2b).

Developmental small RNA libraries from *Nematostella* [19] and *Amphimedon* and two replicate small RNA libraries from *Mnemiopsis* [52], were included in our analysis. In addition to the non-bilaterian datasets, we analysed eleven developmental small RNASeq libraries from *Drosophila* [53]. As the sRNA repertoire of *Drosophila* is well characterised, we first determined if the classification pipeline produced results consistent with prior published analyses [46, 53–55].

Discrimination and annotation of RNAi classes in *Drosophila*

Drosophila is one of the most well-annotated and widely studied model organisms in terms of its small RNA repertoire. Of the three RNAi classes, miRNAs are the best

annotated. In total there are 258 miRNAs currently deposited in miRBase (release 21) and 150 of these have been annotated with high confidence [56]. We were able to identify 139 previously reported miRNAs clusters (54%) including 121 high confidence miRNAs (81%; Additional files 3, 4). The UI of miRNA clusters averaged 122.5 compared to 1.8 for endo-siRNA clusters. No new miRNA candidates were identified in *Drosophila*.

The piRNA repository piRBase currently details over 28 million individual piRNA sequences in *Drosophila* [57]. From the 11 *Drosophila* datasets examined here, we identified 8929 putative piRNA clusters. Of these, 8915 (99.8%) overlap a previously reported piRNA sequence (Additional file 4).

In *Drosophila*, endo-siRNAs are less well annotated than either miRNAs or piRNAs. As no central endo-siRNA database has yet been established, we produced a reference database of endo-siRNA loci from those reported in six previous publications (Additional file 4) [35–38, 58, 59]. This reference database comprises 1210 clusters spanning ~ 5.7 million base pairs (bp) or 3.3% of the *Drosophila* genome. Our analysis identified 3517 endo-siRNA clusters covering approximately 1.4% of the *Drosophila* genome (Additional file 4). An intersection of our reference dataset based on previous publications with the newly identified endo-siRNA cluster loci identified 13.3% congruence (467 loci) between the two. This represents a significant enrichment compared to what would be expected if the reference dataset and the newly identified endo-siRNA clusters had uncorrelated genomic distributions ($p < 0.00001$; see Additional file 5: Supplementary Methods). 86.7% of the endo-siRNA clusters identified by our pipeline were not found in the reference. This may be due to the incompleteness of the limited reference endo-siRNA dataset.

Evidence of a ping-pong biogenesis signature (a bias for a uridine at position one and an adenosine at position 10) [43, 44, 46] was found in the putative piRNAs from both the *Drosophila* adult female and adult male body libraries as well one of the 2–4 day old pupal libraries (Additional file 6). Such a signature was not found in any of the putative endo-siRNAs in which, as expected, only a position one-uridine bias was observed (Additional file 7) [34, 40].

To confirm an association between transposons and the putative endo-siRNA and piRNA clusters, the genomic positions of all clusters were intersected with those of annotated coding sequences, including exons, introns, 5′ untranslated regions (5' UTRs) and 3′ untranslated regions (3' UTRs), and known and unknown transposons (based on sequence similarity to Repbase entries). Clusters that did not overlap with these genomic elements were deemed to be 'intergenic'. As anticipated, multi-mapping endo-siRNAs and piRNAs derive primarily from transposons (Fig. 1) [60, 61]. In addition, we found that unique endo-siRNA clusters frequently map to exons, 5′ and 3' UTRs in coding genes (Fig. 1), with unique endo-siRNA clusters underrepresented in introns suggesting that endo-siRNA production occurs after intron splicing.

The program Randfold [62] was used to test the likelihood that the secondary structures predicted to form from the precursor transcripts of endo-siRNA and piRNA clusters could occur by chance. Briefly, Randfold compares the minimum free energy of the predicted secondary structure of a native sequence to the minimum free energies of randomised versions of itself. For each library, Randfold scores were generated for endo-siRNA and piRNA clusters and these were compared to all other clusters (i.e. all clusters other than those under investigation) from the libraries in question (Fig. 2). Both unique and multi-mapping endo-siRNA clusters in *Drosophila* show evidence of secondary structure while putative piRNA transcripts do not (Fig. 2). This is consistent with most models of endo-siRNA and piRNA biogenesis in bilaterians in which some endo-siRNAs are cleaved from secondarily structured primary transcripts while piRNAs are not [60].

Given that the putative piRNA and endo-siRNA clusters identified here have proven to be consistent with previously reported properties, we deemed our method to be satisfactory for naive identification and annotation.

Discrimination and annotation of RNAi classes in non-bilaterians

Using the same approach undertaken in *Drosophila*, we surveyed the miRNA, piRNA and endo-siRNA repertoire of *Amphimedon*, *Nematostella* and *Mnemiopsis*. The numbers of clusters corresponding to each RNAi class in each species are summarised in Table 1.

Our analysis identified all eight previously reported miRNAs from *Amphimedon*, 62 of the previously reported 141 miRNAs from *Nematostella* and confirmed the absence of miRNAs in *Mnemiopsis*. In addition, we identified seven new miRNA candidates from *Amphimedon* including a second copy of aqu-miR-2016 located just over 1 kilobase (kb) from the originally annotated copy, and four new miRNAs in *Nematostella*, all of which are copies of previously reported miRNAs (Additional files 8, 9, 10). None of the newly identified sponge miRNAs share sequence similarity with the miRNAs recently identified in two other sponges *Stylissa carteri* and *Xestospongia testudinaria* [23].

Using miRDeep2, the current standard tool for animal miRNA detection [63], we were able to detect all four of the new miRNAs from *Nematostella* from at least one developmental library (Additional file 11). None of the newly identified *Amphimedon* miRNAs were detected by miRDeep2 most likely because these sponge miRNAs

Fig. 1 Genomic context of endo-siRNA and piRNA cluster expression for unique and multi-mapping clusters. Each colour-coded segment represents the percentage of endo-siRNA or piRNA clusters mapping to the specified genomic elements. Percentages slightly exceed 100% due to some regions of the genome encoding multiple types of element. The genome column shows the percentage of the genome covered by the specified genomic elements. For *Drosophila*, d1) 12–24 h embryo, d2) first instar larvae 1, d3) first instar larvae 2, d4) third instar larvae 1, d5) third instar larvae 2, d6) 0–1 day pupae, d7) 2–4 day pupae 1, d8) 2–4 day pupae 2, d9) male adult body, d10) female adult body, d11) female adult head; *Amphimedon* a1) pre-competent larvae, a2) competent larvae, a3) juvenile, a4) adult; *Mnemiopsis,* m1) *Mnemiopsis* 1, m2) *Mnemiopsis* 2; *Nematostella,* n1) unfertilized eggs, n2) blastula, n3) gastrula, n4) early planula larvae, n5) late planula larvae, n6) metamorphosing, n7) primary polyp, n8) male adult, n9) female adult

are structurally different from canonical bilaterian miRNAs ([2] and this study).

Despite not being detected by miRDeep2, three of the new miRNA hairpins (aqu-miR-temp-1,4,6) structurally resemble canonical metazoan pre-miRNAs while the remaining three (aqu-mir-temp-2,3,5) are more similar to the eight previously described long-form miRNAs in *Amphimedon* (Additional file 8) [2]. All of these candidates possess either low numbers of reads mapping to their passenger strands or variable passenger strand 5′ ends [64]. However as these characteristics are present in some high confidence miRNAs, such as human hsa-miR-126 [64], we annotate these six loci as candidate novel miRNAs. The remaining HU endo-siRNA-like clusters consist of a mixture of snoRNA, tRNA and rDNA loci, and clusters with highly multi-mapping dominant reads, endogenous hairpin RNAs (hp-RNA; Additional file 12) [65] and secondary structures not consistent with any known sRNA class.

Unlike in *Drosophila* where evidence of a ping-pong biogenesis signature was only found in two of the 11 libraries, a bias for a 5′ uridine and an adenosine at position 10 was detected in all *Amphimedon* and *Nematostella* libraries and one of the two *Mnemiopsis* libraries (Additional file 6). As expected, endo-siRNA clusters only exhibit a bias for a 5′ uridine (Additional file 7) [34, 40].

As in *Drosophila*, unique *Amphimedon* endo-siRNA clusters frequently map to coding genes (Fig. 1). In contrast, distributions of unique endo-siRNAs do not show a bias towards coding genes in *Mnemiopsis* or *Nematostella.* Unique endo-siRNAs in these species map to coding genes with a frequency more similar to that which would be expected if they were randomly distributed throughout the genome (Fig. 1). In all species, multi-mapping endo-siRNA and piRNA clusters tend to map to transposons. This is at odds with one study that concluded endo-siRNAs were not associated with transposons in most phyla [66].

Randfold analysis of unique endo-siRNA clusters in *Amphimedon* and *Nematostella* show that they are no more likely to form secondary structures than the putative transcripts of all other unique clusters (i.e. all identified sRNA clusters not including endo-siRNAs). In contrast, unique endo-siRNA clusters in *Mnemiopsis,* as in *Drosophila,* show evidence of secondary structuring, as do multi-mapping endo-siRNA clusters in all four species (Fig. 2).

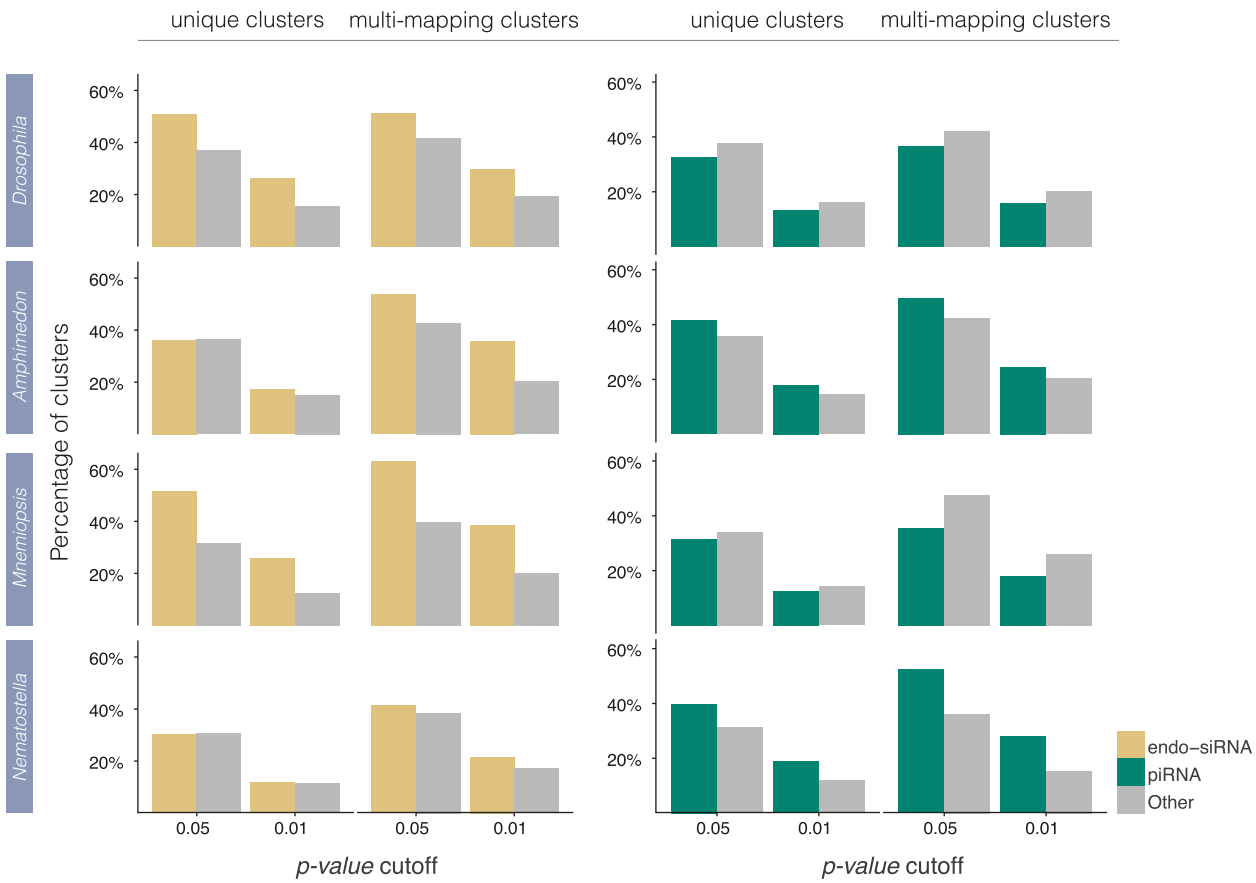

Fig. 2 Randfold results for endo-siRNA and piRNA clusters. Each bar represents the percentage of clusters with Randfold *p-values* equal to or less than the values stated on the X-axis. The more stringent the *p-value* cutoff, the more confidence there is that the secondary structure of the native sequence is more stable than a randomised version of itself. For each graph, the Randfold scores of either endo-siRNAs or piRNAs are compared to the Randfold scores of all clusters not annotated as endo-siRNAs or piRNAs. For each species, all available datasets were pooled

Unexpectedly, putative piRNA transcripts of *Amphimedon* and *Nematostella* show evidence of secondary structure for both unique and multi-mapping clusters while those in *Mnemiopsis* are more similar to the unstructured piRNAs known from bilaterians (Fig. 2).

Variation in overall RNAi complements in basal metazoans

The relative contributions of sRNAs markedly differ amongst *Amphimedon, Mnemiopsis, Nematostella* and *Drosophila* (Fig. 3). In *Amphimedon* and in all but one *Drosophila* developmental stage, miRNAs comprise the bulk of mapped sRNAs while endo-siRNAs and piRNAs are dominant in *Mnemiopsis* and *Nematostella* respectively. Except for the *Nematostella* libraries, a substantial proportion of each library remains unassigned to one of the three RNAi classes (Fig. 3). This is likely due to the stringent requirements set here for annotating sRNA clusters (see Methods) and the presence of non-RNAi related sRNAs produced by each animal.

Developmental dynamics of endo-siRNA and piRNA expression

Co-expression of endo-siRNA and piRNA clusters across developmental time was investigated in *Amphimedon,*

Table 1 Number of annotated miRNA, piRNA and endo-siRNA clusters in *Drosophila*, *Amphimedon*, *Nematostella* and *Mnemiopsis*

	miRNA clusters	unique endo-siRNA clusters	multi-mapping endo-siRNA clusters	unique piRNA clusters	multi-mapping piRNA clusters
Drosophila	139	1877	3517	897	8929
Amphimedon	15	1390	7218	9547	25,840
Nematostella	32	232	4524	7912	48,465
Mnemiopsis	0	4408	23,579	1235	2231

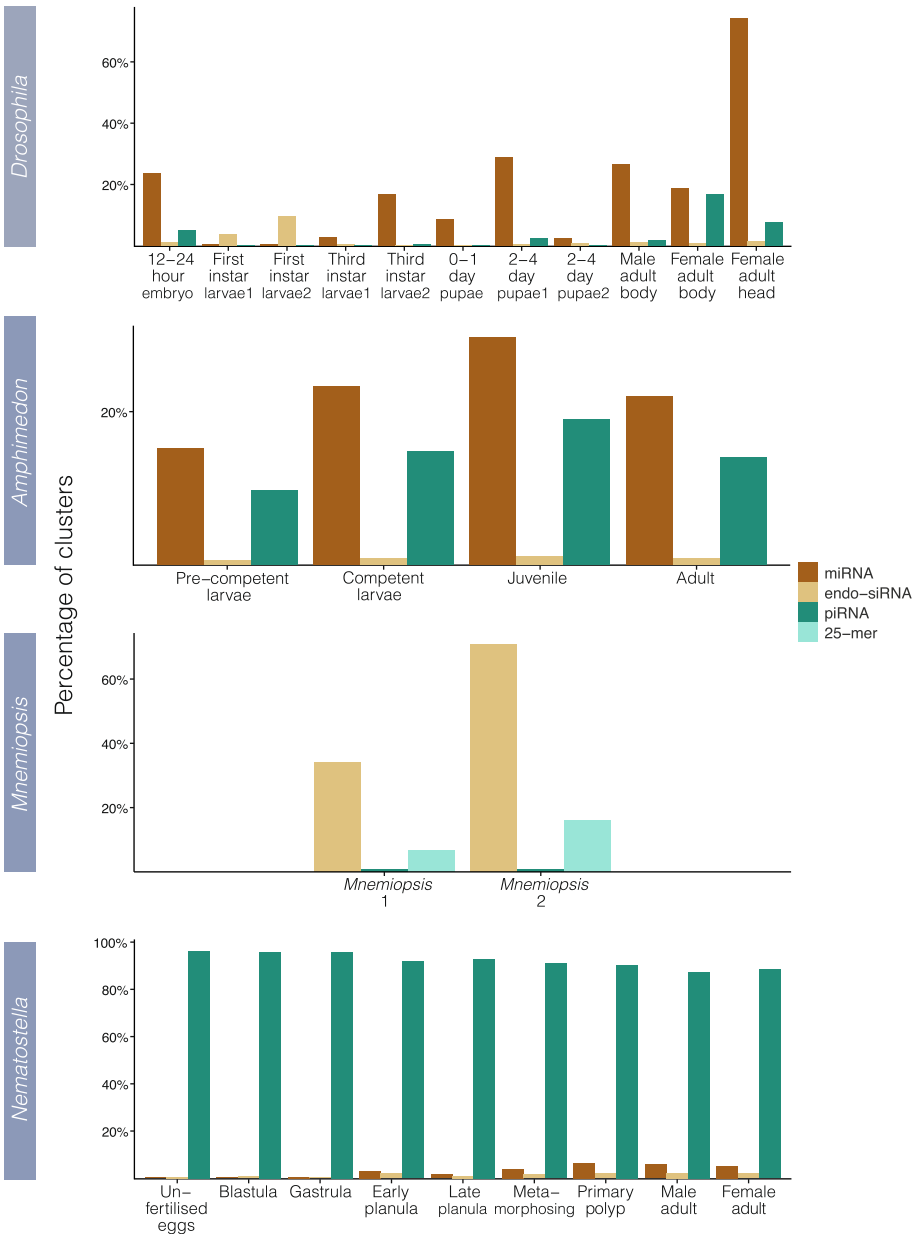

Fig. 3 Library contributions from each RNAi component as a percentage of total library depth. Total contributions of miRNAs, endo-siRNAs, piRNAs and *Mnemiopsis* 25-mer clusters to total library depth. For each, only a single copy of each multi-mapping read was considered

Nematostella and *Drosophila* (Fig. 4; Additional file 13); *Mnemiopsis* was excluded due the absence of developmental data. This analysis highlights differences in the expression dynamics of endo-siRNAs and piRNAs; while many endo-siRNAs are co-expressed in the *Nematostella* male adult and female adult libraries, the populations of piRNAs in these two samples appear to be more distinct. Likewise for *Amphimedon*, endo-siRNA co-expression is highest for the two larval libraries whereas piRNAs appear to be more consistently expressed in all four developmental stages.

Mnemiopsis 25-mer cluster annotation

In addition to putative endo-siRNA and piRNA clusters, a substantial proportion of *Mnemiopsis* reads were found to be approximately 25 nt in length (Fig. 5a). As the *Mnemiopsis* clusters producing ~25 nt reads (hereafter referred to as 25-mer clusters) may constitute a new class or type of sRNA, these clusters were further investigated.

Small RNAs from *Mnemiopsis* 25-mer clusters have a bias for a uridine at their 5′ end, as is common for endo-siRNAs but no ping-pong biogenesis signature was identified (Fig. 5b). These 25-mer clusters did not appear

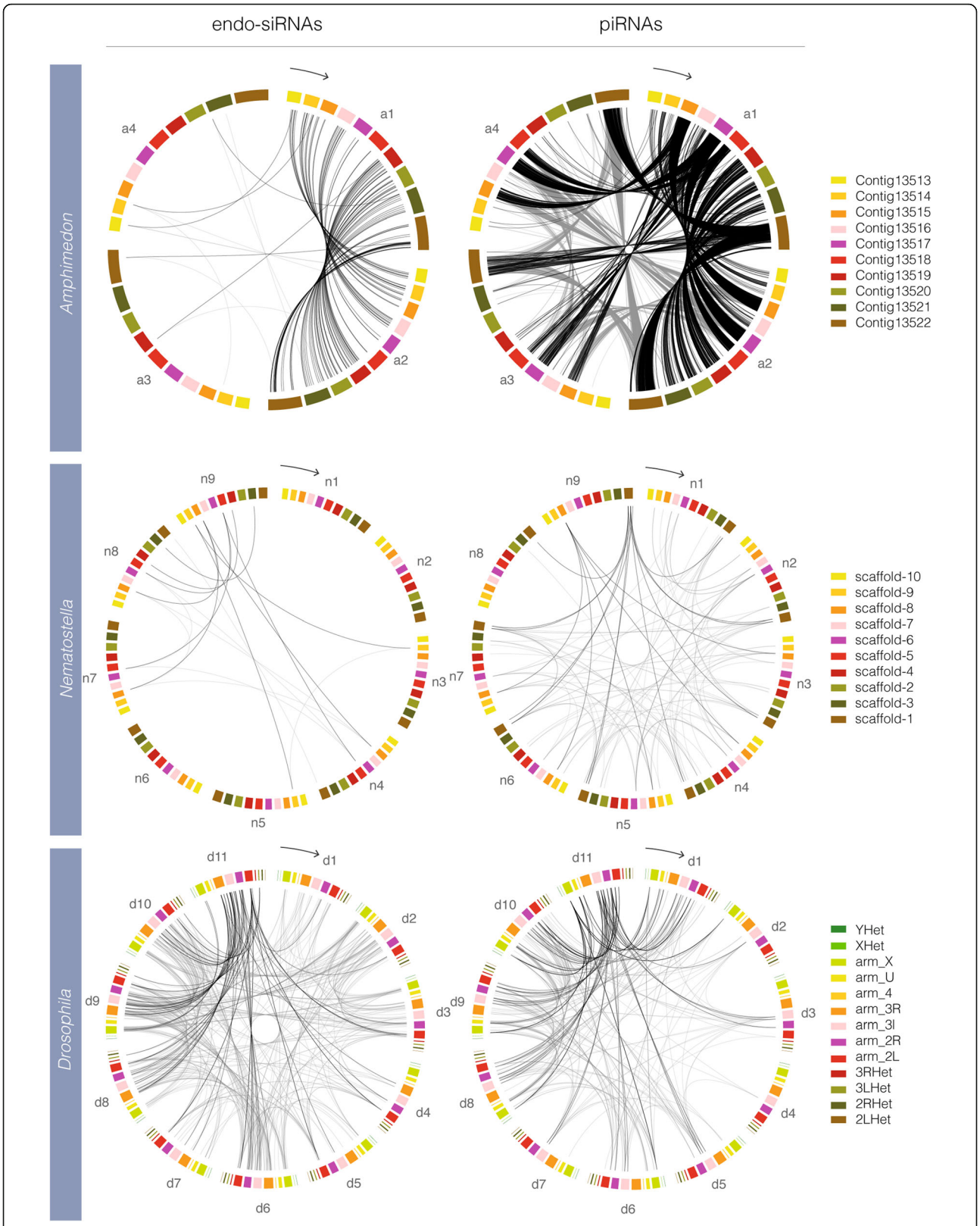

Fig. 4 (See legend on next page.)

(See figure on previous page.)
Fig. 4 Co-expression of uniquely-mapping endo-siRNA and piRNA clusters. Each plot is divided in to groups of coloured scaffolds/chromosomes, each of which represents a developmental stage; four stages in *Amphimedon*, nine stages in *Nematostella* and 11 stages in *Drosophila*. For each plot, the earliest developmental stage is marked with an arrow indicating the chronological order of developmental stages. Links between scaffolds/chromosomes indicate co-expression from a particular endo-siRNA or piRNA cluster in the two linked developmental stages. For *Drosophila*, all chromosomes are represented while for *Amphimedon* and *Nematostella*, the ten largest genomic scaffolds were used. Beginning with the developmental stage indicated by the arrow, the stages for *Amphimedon*, *Nematostella* and *Drosophila* are as per Fig. 1. For each species, the links shared with a single developmental stage are coloured black for emphasis while the rest are coloured grey. For *Amphimedon* the emphasised stage is the pre-competent larvae (a1), for *Nematostella* the female adult (n9) and for *Drosophila*, the female adult head (d11)

to differentially map to any genic feature, including transposons (Fig. 5c) and there was no evidence of secondary structure amongst the putative 25-mer precursor transcripts (Fig. 5d). As no evidence of structure or function was identified for 25-mer clusters, further work is required to determine whether they are biologically significant.

Discussion

Although RNA interference systems are important post-transcriptional regulators in metazoans, a detailed understanding of the repertoire, role and developmental dynamics of these systems is lacking for most animal taxa, resulting in an incomplete picture of their evolution and function. Here we developed a method for the clustering and annotation of mapped sRNA libraries, and applied it to annotating small RNA components in the demosponge *Amphimedon*, the ctenophore *Mnemiopsis* and the cnidarian *Nematostella*. We used this approach to identify miRNAs, piRNAs and endo-siRNAs in these non-bilaterian metazoans, and thereby address the early evolution of metazoan RNAi systems. As the application of this method to the bilaterian *Drosophila* recapitulated the results of previous studies [35–38, 43, 44, 58, 59], it appears that this approach can be applied to other species.

As expected based on previous work [59], our method found that *Drosophila* miRNAs account for the highest number of mapped reads, and piRNAs and endo-siRNAs are dynamically expressed and frequently map to transposons. Endo-siRNAs display a bias for a 5′ uridine and the ping-pong biogenesis signature can be detected in annotated piRNAs. 99.8% of the *Drosophila* piRNA clusters identified using this method map to previously reported piRNAs and that at least 13% of endo-siRNAs also correspond to previously reported endo-siRNA generating loci. In agreement with the established models of endo-siRNA and piRNA biogenesis, secondary structure appears to be important for *Drosophila* endo-siRNA clusters but not for piRNA clusters.

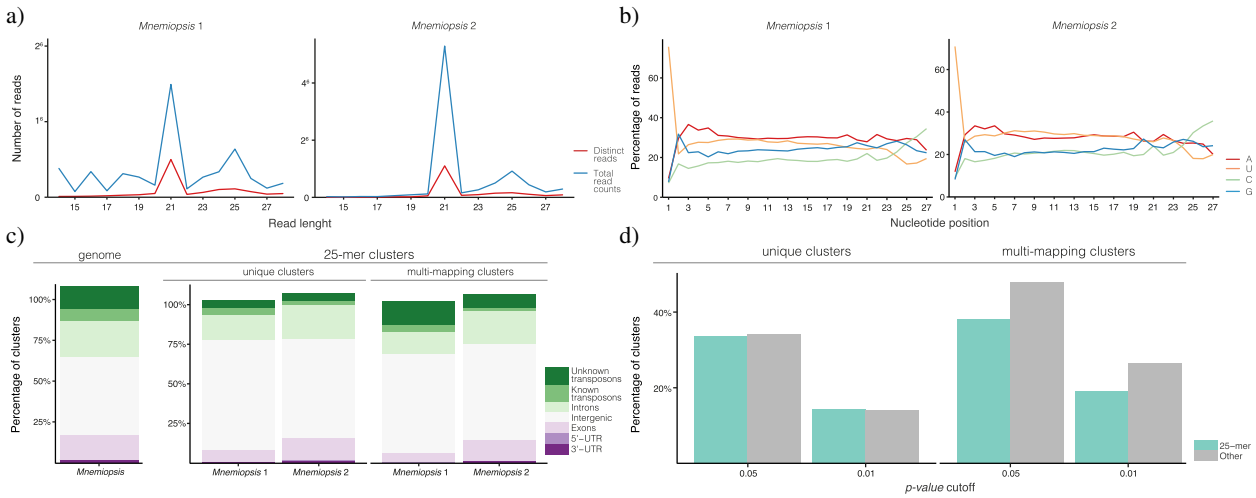

Fig. 5 Characterisation of *Mnemiopsis* 25-mer clusters. The *Mnemiopsis* 25-mer clusters were annotated using the same methods employed for characterisation of the three known RNAi classes. **a** Read length distribution of all mapped sRNAs from the Woods Hole, MA, USA library (*Mnemiopsis* 1) and the Miami, FL, USA library (*Mnemiopsis* 2). Distinct reads (red) and total read counts (blue) of all mapped sRNA size classes reveals peaks of mapped sRNAs at 21 and 25 nt in both libraries. **b** Nucleotide biases along the length of all sRNAs mapping to 25-mer clusters. sRNAs were anchored at their 5′ nucleotide and biases are displayed as a percentage of the contribution of each nucleotide at each position. Of note is the tendency for a uracil at position 1. **c** Genomic context of 25-mer cluster expression (as per Fig. 1) demonstrates the lack of enrichment of 25-mer clusters from coding genes or transposons. **d** Randfold results (as per Fig. 2) demonstrate a lack of evidence for secondary structure in 25-mer clusters

Using this strategy for the clustering and annotation of mapped sRNA libraries, we detected all previously reported miRNAs in the *Amphimedon* datasets and 44% of known miRNAs from *Nematostella*. We also confirmed the absence of miRNAs in *Mnemiopsis,* and showed that endo-siRNAs and piRNAs are the most abundant RNAi classes in *Mnemiopsis* and *Nematostella* respectively. In *Amphimedon*, as in *Drosophila*, unique endo-siRNAs derive primarily from the exons and UTRs of coding genes, consistent with these being derived from mature spliced mRNAs in both species.

The ping-pong piRNA biogenesis signature is a feature of secondary piRNA biogenesis produced during the silencing of active transposons [46]. The detection of such a signal in all but one of the *Amphimedon*, *Mnemiopsis* and *Nematostella* sRNA libraries suggests heightened transposon activity during the development of these species in comparison to *Drosophila*. In libraries in which a ping-pong biogenesis signature was not identified, secondary piRNA biogenesis may be occurring at levels below detection by this method. Those piRNA clusters that do not exhibit a bias for a position 10 adenosine may represent sites of primary piRNA biogenesis.

Primary transcript secondary structure does not appear to be a requirement for piRNA biogenesis [67], although a role for the RNA helicase MOV10L1/Armitage in unwinding localised secondary structures of piRNA precursors in mice and *Drosophila* has been described [68, 69]. Orthologues of this helicase can be found in *Amphimedon* (NCBI: XP_019853676.1), *Nematostella* (NCBI: XP_001626596.1, XP_001637169.1) and *Mnemiopsis* (NHGRI: ML005359a). Our analysis did not find any evidence of conserved piRNA cluster secondary structure in *Drosophila* or *Mnemiopsis*, however *Amphimedon* and *Nematostella* piRNA clusters do appear to be structured.

In mammals, RNA secondary structural elements known as G quadruplexes appear to act as landmarks for piRNA biogenesis [68]. This association creates a characteristic genomic signature in piRNA producing loci consisting of a guanosine enrichment downstream from position 25. We observed a similar enrichment in the piRNAs of *Amphimedon*, *Mnemiopsis* and *Nematostella* but not in those of *Drosophila* suggesting that sites of localised secondary structure within primary piRNA transcripts may be an ancestral feature of metazoan piRNA biogenesis (Additional file 6).

Unique and multi-mapping endo-siRNA clusters in *Drosophila* and *Mnemiopsis* appear to have a propensity to form secondary structures while only multi-mapping endo-siRNA clusters appear to in *Amphimedon* and *Nematostella*. As endo-siRNA directed RNA interference is most efficient for targets with full-length complementarity [70], most uniquely mapping endo-siRNAs are expected to silence transcripts arising from the antisense strand from which their host gene was transcribed [71]. Consistent with this, Randfold analysis of the predicted secondary structures formed by *Amphimedon* unique endo-siRNA clusters showed that they are more likely to occur by chance than are the secondary structures formed by multi-mapping endo-siRNA clusters.

Given that (i) *Amphimedon* does not encode an RNA dependent RNA polymerase (RdRP), (ii) secondary structure is probably less important for the biogenesis of most unique endo-siRNAs and (iii) the most efficient targets of unique endo-siRNAs are likely found antisense to themselves, it follows that most unique endo-siRNAs are likely to be the products of cis-Natural Antisense Transcripts (cis-NATs) [38, 72] rather than hairpin RNAs. Of the 40,122 coding gene models for *Amphimedon* [73], 8133 are predicted to be cis-NATs. While this only represents 20.3% of the total coding genes, nearly 50% of all unique endo-siRNA clusters that align to coding genes, align to these putative cis-NAT genes.

Unique endo-siRNA clusters in *Drosophila* also align to coding genes, although both unique and multi-mapping endo-siRNA clusters show evidence of forming secondary structures. Despite this, the 16% of genes that form cis-NAT pairs in this species account for 22% of all mature coding gene-mapping unique endo-siRNA clusters, suggesting that cis-NATs are the source of some uniquely mapping endo-siRNAs in *Drosophila*. Differences in the rate of cis-NAT endo-siRNA production observed between *Drosophila* cell types [59] may account for the lower overall rate detected in comparison to *Amphimedon*. The more compact *Amphimedon* genome may also be responsible for a higher rate of overlapping antisense transcripts [73, 74].

Conclusions

The RNAi repertoires of non-bilaterian metazoans - sponges, ctenophores and cnidarians – differ both from each other and from the canonical RNAi repertoire of bilaterians. Although largely comprised of the same three major systems that constitute the bilaterian RNAi repertoire, the degree to which miRNAs, piRNAs and endo-siRNAs are expressed varies substantially between the sponge *Amphimedon*, the ctenophore *Mnemiopsis* and the cnidarian *Nematostella*. The unexpected differences in the RNAi repertoire of bilaterians, cnidarians, ctenophores and sponges uncovered here, suggests that while the last common ancestor of extant animals employed miRNA, piRNA and endo-siRNA systems, these were not integrated into an ancestral gene regulatory system. This is in contrast to bilaterians, which appear to use a common RNAi system [10–16], although some RNAi innovations have also been identified in select bilaterian linages [75, 76]. Following the emergence of these major metazoan RNAi pathways,

lineage-specific evolutionary trajectories appear to have resulted in divergent RNAi strategies evolving in each basal metazoan lineage.

Methods

Biological sampling, small RNA library preparation and sequencing

Additional details on methods can be found in Supplementary Methods. Briefly, *Amphimedon* material was collected from Heron Island, Australia and RNA extracted using Tri Reagent (Sigma Aldrich). The adult small RNA library was prepared with the Illumina TruSeq Small RNA Sequencing Kit as per the manufacturer's instructions. Pre-competent larval, competent larval and juvenile small RNA libraries were prepared with the Epicentre ScriptMiner Small RNA-Seq Library Preparation Kit as per the manufacturer's instructions. Libraries were indexed and the four libraries were pooled with eight others unrelated samples giving a total of 12 samples. These were then split and sequenced over four lanes on an Illumina HiSeq 2000.

Mapping of sRNA libraries to genomes

Mnemiopsis (SRS355925, SRS355926) [3], *Nematostella* (SRR039731, SRR039754, SRR039764, SRR039762, SRR039760, SRR039758, SRR039756, SRR039726, SRR039727) [19] and *Drosophila* (SRR013604, SRR018039, SRR016854, SRR013601, SRR013603, GSM360260, SRR013600, SRR013602, GSM360256, GSM360257, SRR014367) [53] sRNA datasets were acquired either from NCBI's Sequenced Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) or from NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo).

All fastq files were checked for quality with FastQC. 3′ adaptor sequences were removed with fastx_clipper from the FASTX-Toolkit (v0.0.13). Collapsed reads were mapped to their respective genomes with bowtie (v0.12) [77] allowing for up to 51 mappings per read but no mismatches between the read and the genome. Those reads that were mapped to the genome 51 times were then removed from the library, leaving only reads that mapped between 1 and 50 times. A second file was produced from those reads that only mapped to a single genomic location.

Small RNA cluster generation and minimum free energy

All sRNAs that map to annotated rRNAs were removed from the libraries and the remaining reads were clustered using bedCluster.pl [78]. A 150 bp window was defined for cluster generation, reflecting the approximate length of the long pre-miRNAs typical of *Amphimedon* [2] and in recognition that miRNA, piRNA and endo-siRNA biogenesis results in products located in overlapping or close genome proximity to one another, all of which derive from an original primary transcript (or two in the case of natural antisense endo-siRNAs) [38]. Only clusters composed of at least three distinct reads (non-perfectly overlapping) and at least 51 bp in length were considered. Clusters corresponding to previously reported miRNAs were annotated as such. tRNAs were predicted with tRNA-scan-SE [79] and snoRNAs with snoSeeker [80], and clusters mapping to these locations were annotated. The minimum free energy of each cluster was defined using RNALfold from the Vienna RNA package (v2.05) [81]. For those clusters comprised of reads from both strands, both strands were submitted to RNALfold with the strand that produced the lowest minimum free energy (MFE) retained. If both strands produced equal MFEs, a strand was selected arbitrarily. To assess the likelihood that the structures predicted by RNALfold could have arisen by chance, each was submitted to Randfold [62] with 100 randomisations. Randfold measures the MFE of these randomisations and compares the results to the MFE of the native sequence. The result is a *p-value* assigned to each cluster that describes the likelihood that the native sequence of that cluster will fold to form a secondary structure that is more stable than a randomised version of itself. This can be interpreted as the likelihood that the secondary structure predicted for a cluster has not occurred by chance and thus is likely functionally important.

Endo-siRNA, piRNA and 25-mer cluster annotation

Endo-siRNA, piRNA and 25-mer clusters were annotated based on the read length composition of their constituent sRNAs. For endo-siRNAs, clusters with peaks of expression at 20, 21 or 22 nt were first selected, reflecting the typical length of Dicer cleavage products. If the sum of the reads constituting the peak read length plus or minus one nucleotide was greater than the total number of reads of all other size classes, these were annotated as endo-siRNA clusters. For piRNA annotation, sRNA peaks of 26, 27 or 28 nt were required for the non-bilaterians while for *Drosophila*, 24, 25 or 26 nt were selected, reflecting the shorter length of piRNAs in this species [43]. For *Mnemiopsis* 25-mer clusters, 24, 25 or 26 nt peak clusters were also selected.

Genomic context

The four genomes were annotated according to their coverage by transposons or coding genes before being intersected with sRNA clusters. RepeatModeler and RepeatMasker [82] were used to identify transposons in all four genomes both with (known) and without (unknown) homology to those in RepBase [83]. Exons, introns, 5′ and 3' UTRs were obtained from publicly available sources. Exons, introns, 5′ and 3' UTRs that overlapped with predicted transposons were removed. All elements were mapped to the genome with GenomeCoverageBed from the BEDTools package (v2.5.0) [84].

The genomic context of endo-siRNA and piRNA clusters were assessed using overlapSelect from UCSC [85] to determine which elements clusters aligned to. At least 51% of the length of a cluster was required to overlap with a particular feature, otherwise it was deemed to be intergenic. To determine if any HU clusters derive from tRNAs or snoRNAs, HU clusters were intersected with tRNAs as predicted with tRNAScan-SE [79] and snoRNAs as predicted by snoSeeker [80].

Cis-NAT prediction of gene models

Gene models for the four species were overlapped with themselves using overlapSelect [85]. Gene models from opposing strands were aligned to one another and those overlapping another by at least one nucleotide were considered to be cis-NAT genes.

Circos plots

Circos plots [86] were constructed that describe the co-expression of clusters in different developmental contexts. Links were formed between corresponding genomic loci from two developmental stages if those loci co-expressed either an endo-siRNA or piRNA cluster in both temporal contexts.

Additional files

Additional file 1: Demonstration of High Uniformity and Low Uniformity sRNA clusters. Two hypothetical hairpin RNAs demonstrating the difference between a high uniformity and a low uniformity clustering. In (a), a total of 16 reads composed of just three distinct reads map to a hairpin RNA giving a uniformity index of 5.3. In (b), 16 reads also map to a hairpin RNA but these are composed of 12 distinct reads resulting in a uniformity index of just 1.3. The high uniformity cluster (a) is composed of an equal number of reads to the low uniformity cluster (b) however these reads are less evenly distributed along the length of the source hairpin RNA. (PDF 304 kb)

Additional file 2: Effect of library depth on uniformity index. Random sampling of reads from the *Amphimedon* juvenile library (1, 5, 10, 25, 50, 75, 100%) show a trend towards increasing UI for high uniformity miRNA clusters as library depth increases (a). Dividing the UI by the library depth acts to normalise these values (b). Library depth normalised UIs can be more accurately compared between libraries. (PDF 138 kb)

Additional file 3: Uniformity of *Drosophila* endo-siRNA and miRNA clusters. Endo-siRNA clusters (yellow) display a consistently lower uniformity of small RNA expression (ratio of total read counts:distinct reads) in comparison to miRNA clusters (red) for both unique clusters (above) and multi-mapping clusters (below). (PDF 3341 kb)

Additional file 4: Locations of annotated RNAi loci from *Drosophila*, *Amphimedon*, *Nematostella* and *Mnemiopsis*. Genomic loci of annotated miRNA, piRNA, endo-siRNA and 25-mer clusters in all four species. (GZ 1305 kb)

Additional file 5: Supplementary Methods. Detailed methods (DOCX 66 kb)

Additional file 6: Nucleotide biases of piRNA clusters. Nucleotide biases along the length of all sRNAs mapping to predicted piRNA clusters. sRNAs were anchored at their 5′ nucleotide and biases are displayed as a percentage the contribution of each nucleotide at each position. Of note is the tendency for a uracil at position 1 and an adenosine at position 10 in most libraries that together comprise the ping-pong piRNA biogenesis signature. Arrows indicate guanosine enrichments downstream of position 25. (PDF 323 kb)

Additional file 7: Nucleotide biases of endo-siRNA clusters. Nucleotide biases along the length of all sRNAs mapping to predicted endo-siRNA clusters. sRNAs were anchored at their 5′ nucleotide and biases are displayed as a percentage of the contribution of each nucleotide at each position. Of note is the tendency for a uracil at position 1 which is present in all libraries except the *Drosophila* 1st instar larval libraries. (PDF 328 kb)

Additional file 8: New *Amphimedon* miRNA candidates. Wiggle plots and predicted secondary structures of mapped reads across the length of previously described miRNA miR-2016a, the newly identified miR-2016b and six novel miRNA candidates (aqu-mir-temp-1-6). For each cluster, the library with the most mapped reads to each loci was used to construct the graph. (PDF 483 kb)

Additional file 9: New *Nematostella* miRNA candidates. Wiggle plots and predicted secondary structures of four newly identified miRNAs in the sea anemone. All four miRNAs are new copies of previously identified miRNAs. (PDF 205 kb)

Additional file 10: New miRNA data. Sequence and genomic location data for the newly identified *Amphimedon* and *Nematostella* miRNAs. (XLSX 48 kb)

Additional file 11: miRDeep2 identification of new *Nematostella* miRNA candidates. Results of miRDeep2 annotation of the newly identified miRNA candidates from *Nematostella*. (a) nve-miR-temp-1, (b) nve-miR-temp-2, (c) nve-miR-temp-3, (d) nve-miR-temp-4. (PDF 2157 kb)

Additional file 12: *Amphimedon* endogenous hairpin RNAs. Wiggle plots and predicted secondary structure of three long highly complementary endo-siRNAs from *Amphimedon* with unevenly distributed mapped sRNA populations. (PDF 348 kb)

Additional file 13: Co-expression of multi-mapping endo-siRNA and piRNA clusters across development. Each plot is divided into groups of coloured scaffolds/chromosomes, each of which represents a developmental stage. For each plot, the earliest developmental stage is marked with an arrow indicating the chronological order of the following developmental stages. Links between scaffolds/chromosomes indicate co-expression from a particular endo-siRNA or piRNA cluster in the two linked developmental stages. For *Drosophila*, all chromosomes are represented while for *Amphimedon* and *Nematostella*, the ten largest genomic scaffolds were used. Beginning with the developmental stage indicated by the arrow, the stages for *Amphimedon*, *Nematostella* and *Drosophila* are as per Fig. 1. For each species, the links shared with a single developmental stage are coloured black for emphasis while the rest are coloured grey. For *Amphimedon* the emphasised stage is the pre-competent larvae, for *Nematostella* the female adult and for *Drosophila*, the female adult head. (PDF 23669 kb)

Abbreviations

bp: Base pairs; cis-NAT: Cis-natural antisense transcript; endo-siRNA: Endogenous small interfering RNA; hp-RNA: hairpin RNA; HU: High uniformity; kb: Kilo base; LU: Low uniformity; miRNA: Micro RNA; nts: Nucleotides; piRNA: Piwi-interacting RNA; RNAi: RNA interference; sRNA: Small RNA

Acknowledgements

We thank Kelin Ru and and Anupma Choudhary for their assistance and advice with the production of the *Amphimedon* RNASeq libraries and to Paulina Tapia for her assistance with document design. We also thank Carmel McDougall for feedback on the manuscript.

Funding

This research was supported by an Australian Research Council grant to BMD.

Availability of data and materials

The *Amphimedon queenslandica* RNA-seq datasets are deposited in the SRA database under the accession SRP158555.

Authors' contributions

ADC contributed to the design of the project, analysis and interpretation of the data, collected the required material from the field, conducted the laboratory procedures and drafted the manuscript. SLF-V, RJT and BMD contributed to the design of the project, analysis and interpretation of the data

and drafted the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate
No specific ethics approval was required for this project.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Authors' Note A paper was published after acceptance of this manuscript providing evidence for animal-like microRNAs and the miRNA biogenesis machinery in the unicellular ichthyosporeans (Bråte J, Neumann RS, Fromm B, Haraldsen AAB, Tarver, JE., Suga H, et al. (2018). Unicellular Origin of the Animal MicroRNA Machinery. Current Biology, https://doi.org/10.1016/j.cub.2018.08.018). Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia. [2]Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia. [3]Illumina Inc, San Diego, California 92122, USA. [4]Present address: Department of Integrative Zoology, University of Vienna, Althanstraße 1, 4A-1090 Vienna, Austria. [5]Present address: CONACYT, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio). CINVESTAV, Irapuato, Guanajuato, Mexico.

Received: 19 March 2018 Accepted: 8 October 2018
Published online: 01 November 2018

References
1. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Phil Trans R Soc B. 2009; 364:99–115.
2. Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008;455:1193–7.
3. Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD. MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics. 2012;13:714.
4. Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, et al. The ctenophore genome and the evolutionary origins of neural systems. Nature. 2014;510:109–14.
5. Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, et al. Diverse pathways generate microRNA-like RNAs and dicer-independent small interfering RNAs in fungi. Mol Cell. 2010;38:803–14.
6. Chen R, Jiang N, Jiang Q, Sun X, Wang Y, Zhang H, et al. Exploring microRNA-like small RNAs in the filamentous fungus *Fusarium oxysporum*. PLoS One. 2014;9:e104956.
7. Moran Y, Agron M, Praher D, Technau U. The evolutionary origin of plant and animal microRNAs. Nat Ecol Evol. 2017;1:27.
8. Kosik KS. MicroRNAs and cellular phenotypy. Cell. 2010;143:21–6.
9. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature. 2010;465:617–21.
10. Stoeckius M, Maaskola J, Colombo T, Rahn H-P, Friedländer MR, Li N, et al. Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. Nat. Methods. 2009;6:745–51.
11. Wei Y, Chen S, Yang P, Ma Z, Kang L. Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust. Genome Biol. 2009;10:R6.
12. Ohnishi Y, Totoki Y, Toyoda A, Watanabe T, Yamamoto Y, Tokunaga K, et al. Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. Nucleic Acids Res. 2010;38:5141–51.
13. Wang J, Czech B, Crunk A, Wallace A, Mitreva M, Hannon GJ, et al. Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. Genome Res. 2011;21:1462–77.
14. Wei C, Salichos L, Wittgrove CM, Rokas A, Patton JG. Transcriptome-wide analysis of small RNA expression in early zebrafish development. RNA. 2012;18:915–29.
15. Yao Y, Ma L, Jia Q, Deng W, Liu Z, Zhang Y, et al. Systematic characterization of small RNAome during zebrafish early developmental stages. BMC Genomics. 2014;15:117.
16. Zhao X, Yu H, Kong L, Liu S, Li Q. High throughput sequencing of small RNAs transcriptomes in two *Crassostrea* oysters identifies microRNAs involved in osmotic stress response. Sci Rep. 2016;6:22687.
17. Krishna S, Nair A, Cheedipudi S, Poduval D, Dhawan J, Palakodeti D, et al. Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in *Hydra magnipapillata*. Nucleic Acids Res. 2013;41:599–616.
18. Liew YJ, Aranda M, Carr A, Baumgarten S, Zoccola D, Tambutté S, et al. Identification of microRNAs in the coral *Stylophora pistillata*. PLoS One. 2014;9:e91101.
19. Moran Y, Fredman D, Praher D, Li XZ, Wee LM, Rentzsch F, et al. Cnidarian microRNAs frequently regulate targets by cleavage. Genome Res. 2014;24:651–63.
20. Gajigan AP, Conaco C. A microRNA regulates the response of corals to thermal stress. Mol Ecol. 2017;26:3472–83.
21. Baumgarten S, Cziesielski MJ, Thomas L, Michell CT, Esherick LY, Pringle JR, et al. Evidence for miRNA-mediated modulation of the host transcriptome in cnidarian-dinoflagellate symbiosis. Mol Ecol. 2018;27:403–18.
22. Mauri M, Kirchner M, Aharoni R, Ciolli Mattioli C, van den Bruck D, Gutkovitch N, et al. Conservation of miRNA-mediated silencing mechanisms across 600 million years of animal evolution. Nucleic Acids Res. 2016;45:938–50.
23. Liew YJ, Ryu T, Aranda M, Ravasi T. miRNA Repertoires of Demosponges *Stylissa carteri* and *Xestospongia testudinaria*. PLoS ONE. 2016;11:e0149080.
24. Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, et al. The deep evolution of metazoan microRNAs. Evol Dev. 2009; 11:50–68.
25. Sperling EA, Robinson JM, Pisani D, Peterson KJ. Where's the glass? Biomarkers, molecular clocks, and microRNAs suggest a 200-Myr missing Precambrian fossil record of siliceous sponge spicules. Geobiology. 2010;8:24–36.
26. Tarver JE, Donoghue PCJ, Peterson KJ. Do miRNAs have a deep evolutionary history? BioEssays. 2012;34:857–66.
27. Robinson JM, Sperling EA, Bergum B, Adamski M, Nichols SA, Adamska M, et al. The identification of microRNAs in calcisponges: independent evolution of microRNAs in basal metazoans. J Exp Zool B Mol Dev Evol. 2013;320:84–93.
28. Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet. 2006;50:81–99.
29. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr Biol. 2003;13:807–18.
30. Juliano CE, Reich A, Liu N, Goetzfried J, Zhong M, Uman S, et al. PIWI proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. Proc Natl Acad Sci U S A. 2014;111:337–42.
31. Praher D, Zimmermann B, Genikhovich G, Columbus-Shenkar Y, Modepalli V, Aharoni R, et al. Characterization of the piRNA pathway during development of the sea anemone *Nematostella vectensis*. RNA Biol. 2017;14:1727–41.
32. Urbarova I, Patel H, Forêt S, Karlsen BO, Jørgensen TE, Hall-Spencer JM, et al. Elucidating the small regulatory RNA repertoire of the sea anemone *Anemonia viridis* based on whole genome and small RNA sequencing. Genome Biol. Evol. 2018;10:410–26.
33. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15:509–24.
34. Chung W-J, Okamura K, Martin R, Lai EC. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. Curr Biol. 2008;18:795–802.
35. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. An endogenous small interfering RNA pathway in *Drosophila*. Nature. 2008;453:798–802.
36. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. Science. 2008;320:1077–81.
37. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature. 2008;453:793–7.

38. Okamura K, Balla S, Martin R, Liu N, Lai EC. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. Nat Struct Mol Biol. 2008;15:581–90.
39. Okamura K, Chung W-J, Ruby JG, Guo H, Bartel DP, Lai EC. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. Nature. 2008;453:803–6.
40. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008;453:534–8.
41. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. 2008;453:539–43.
42. Farazi TA, Juranek SA, Tuschl T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development. 2008; 135:1201–14.
43. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell. 2007;128:1089–103.
44. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. Science. 2007;315:1587–90.
45. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell. 2007;129:69–82.
46. Czech B, Hannon GJ. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. Trends Biochem Sci. 2016;41:324–37.
47. Huang X, Fejes-Toth K, Aravin AA. piRNA biogenesis in *Drosophila melanogaster*. Trends Genet. 2017;33:882–94.
48. Han BW, Hung J-H, Weng Z, Zamore PD, Ameres SL. The 3′-to-5′ exoribonuclease nibbler shapes the 3′ ends of microRNAs bound to *Drosophila* Argonaute1. Curr Biol. 2011;21:1878–87.
49. Liu N, Abe M, Sabin LR, Hendriks G-J, Naqvi AS, Yu Z, et al. The exoribonuclease nibbler controls 3′ end processing of microRNAs in *Drosophila*. Curr Biol. 2011;21:1888–93.
50. Hayashi R, Schnabl J, Handler D, Mohn F, Ameres SL, Brennecke J. Genetic and mechanistic diversity of piRNA 3′-end formation. Nature. 2016;539:588–92.
51. Modepalli V, Moran Y. Evolution of miRNA tailing by 3′ terminal uridylyl transferases in Metazoa. Genome Biol Evol. 2017;9:1547–60.
52. Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science. 2013;342:1242592.
53. Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung J-H, et al. Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. Genome Res. 2011;21:203–15.
54. Yates LA, Norbury CJ, Gilbert RJC. The long and short of microRNA. Cell. 2013;153:516–9.
55. Claycomb JM. Ancient endo-siRNA pathways reveal new tricks. Curr Biol. 2014;24:R703–15.
56. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152–7.
57. Zhang P, Si X, Skogerbø G, Wang J, Cui D, Li Y, et al. piRBase: a web resource assisting piRNA functional study. Database. 2014;2014:bau110.
58. Okamura K, Robine N, Liu Y, Liu Q, Lai EC. R2D2 organizes small regulatory RNA pathways in *Drosophila*. Mol Cell Biol. 2011;31:884–96.
59. Wen J, Mohammed J, Bortolamiol-Becet D, Tsai H, Robine N, Westholm JO, et al. Diversity of miRNAs, siRNAs, and piRNAs across 25 *Drosophila* cell lines. Genome Res. 2014;24:1236–50.
60. Gebert D, Rosenkranz D. RNA-based regulation of transposon expression. Wiley Interdiscip Rev RNA. 2015;6:687–708.
61. Zhang D, Tu S, Stubna M, Wu W-S, Huang W-C, Weng Z, et al. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science. 2018;359:587–91.
62. Bonnet E, Wuyts J, Rouzé P, Van de Peer Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics. 2004;20:2911–7.
63. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40:37–52.
64. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42:D68–73.
65. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. Genes Dev. 2008;22:2773–85.
66. Waldron FM, Stone GN, Obbard DJ. Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. PLoS Genet. 2018;14:e1007533.
67. Le Thomas A, Toth KF, Aravin AA. To be or not to be a piRNA: genomic origin and processing of piRNAs. Genome Biol. 2014;15:204.
68. Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. 2015;29:617–29.
69. Fu Q, Pandey RR, Leu NA, Pillai RS, Wang PJ. Mutations in the MOV10L1 ATP hydrolysis motif cause piRNA biogenesis failure and male sterility in mice. Biol Reprod. 2016;95:103.
70. Haley B, Zamore PD. Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol. 2004;11:599–606.
71. Zinad HS, Natasya I, Werner A. Natural antisense transcripts at the interface between host genome and mobile genetic elements. Front Microbiol. 2017;8:1068.
72. Zhang Y, Liu XS, Liu Q-R, Wei L. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res. 2006;34:3465–75.
73. Fernandez-Valverde SL, Calcino AD, Degnan BM. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge *Amphimedon queenslandica*. BMC Genomics. 2015;16:720.
74. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, et al. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature. 2010;466:720–6.
75. Lewis SH, Quarles KA, Yang Y, Tanguy M, Frézal L, Smith SA, et al. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat. Ecol. Evol. 2018;2:174–81.
76. Mondal M, Klimov P, Flynt AS. Rewired RNAi-mediated genome surveillance in house dust mites. PLoS Genet. 2018;14:e1007183.
77. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
78. Nahkuri S, Taft RJ, Mattick JS. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle. 2009;8:3420–4.
79. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:0955–64.
80. Yang J-H, Zhang X-C, Huang Z-P, Zhou H, Huang M-B, Zhang S, et al. snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. Nucleic Acids Res. 2006;34: 5112–23.
81. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.
82. Smit A, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013. Available from: http://www.repeatmasker.org
83. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7.
84. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
85. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013;14:144–61.
86. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.