Supplementary Discussion

***Trichoplax* phylogeny.** *Trichoplax* has neither miRNAs nor piRNAs, and the canonical machinery required for their synthesis is incomplete. Our interpretation of these observations is that during the evolution of the *Trichoplax* lineage, miRNAs and piRNAs were lost — an interpretation that rests on the assumption that *Amphimedon*, which possesses both, diverged prior to *Trichoplax*. Such a phylogeny, with *Amphimedon* basal to *Trichoplax*, is robustly supported by a recent analysis of the *Trichoplax* genome together with other completed genome sequences[17]. An earlier hypothesis based on analysis of mitochondrial genes places the *Trichoplax* divergence as more basal[36]. This alternative phylogeny was explicitly tested using available complete genome sequences and rejected as improbable ($P = 0.07$, ref. 17). In the unlikely event that *Trichoplax* is basal to *Amphimedon*, we would be unable to differentiate the possibility that miRNAs and piRNAs were lost in *Trichoplax* from the possibility that the origin of miRNAs and piRNAs occured after the divergence of *Trichoplax* (essentially the same situation as that which exists for *Monosiga*). Irrespective of the phylogenetic position of *Trichoplax*, however, our major conclusions remain: (i) miRNAs were present before the divergence of the Bilateria and existed early in the evolution of the Metazoa; (ii) miRNA evolution has been highly plastic, with precursor sizes and miRNA sequences differing greatly between *Nematostella*, *Amphimedon* and bilaterians; (iii) two classes of piRNAs are also ancient, predating the divergence of the Porifera; and, (iv) an ancestral function of piRNAs has been to repress transposon activity.

High abundance of *Nematostella* and *Amphimedon* piRNAs. The piRNAs were abundant in both *Nematostella* and *Amphimedon*, even though these species do not have true gonads. In *Amphimedon*, germ cells apparently derive from choanocytes and archeocytes of the adult, which populate much of the animal, but only a minority of which appear to transdifferentiate into germ cells[37]. In *Nematostella*, germ cells also differentiate from somatic cells in the adult body, and cells exhibiting markers of germ-cell specification are present from early juvenile stages[38]. Perhaps the high fraction of piRNAs in these animals corresponds to the high number of cells that have the potential to generate germ cells. This scenario resembles that in *Planaria*, whose piRNAs are expressed in neoblasts, somatic cells distributed throughout the animal that have the potential to differentiate into any cell type, including germ cells[39]. Perhaps the likely broad somatic distribution of piRNAs in *Planaria*, *Nematostella* and *Amphimedon* represents the ancestral condition.

Supplementary References

36 S. L. Dellaporta *et al.* Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum *Proc Natl Acad Sci U S A* **103**, 8751-8756 (2006).

37 S. P. Ereskovsky Leys, A.E. Embryogenesis and larval differentiation in sponges *Can. J. Zool.* **84**, 262-287 (2006).

38 C. G. Extavour, K. Pang, D. Q. Matus & M. Q. Martindale vasa and nanos expression patterns in a sea anemone and the evolution of bilaterian germ cell specification mechanisms *Evol Dev* **7**, 201-215 (2005).

39 D. Palakodeti, M. Smielewska, Y. C. Lu, G. W. Yeo & B. R. Graveley The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians *RNA* **14**, 1174-1186 (2008).

40 Y. Tomari, T. Du & P. D. Zamore Sorting of Drosophila small silencing RNAs *Cell* **130**, 299-308 (2007).

41 N. C. Lau, L. P. Lim, E. G. Weinstein & D. P. Bartel An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans *Science* **294**, 858-862 (2001).

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

Supplementary Table 1. Read counts for *Nematostella* miRNAs in periodate-treated (periodate) and mock-treated (mock) libraries. To calculate the ratio of reads from the two libraries (mock / periodate), the read counts were first normalized to the total number of genome-matching reads present in each library. The ratio for miRNAs not cloned in the periodate-treated library but observed in the mock-treated library is not defined (N.D.); miRNAs observed in neither library are omitted.

The *Amphimedon* miRNA reads were ~300-fold reduced in the treated compared to the untreated sample (Supplementary Table 2), which showed that like most other metazoan miRNAs, *Amphimedon* miRNAs are not methylated at their 2' terminal oxygen. For *Nematostella*, the pattern was more complex; reads for certain miRNAs (including miR-100) were not reduced in the treated sample, which suggested that they were 2' modified, whereas reads for others were reduced >20 fold. This pattern resembled that observed in *Drosophila*, wherein miRNAs associated with Argonaute2 are methylated, whereas those associated with Argonaute1 are not[23,40].

miRNA	Periodate	Mock	Mock/Periodate
miR-100	297	161	1.2
miR-2022	33	13	0.9
miR-2023	13,510	3,754	0.6
miR-2024a	14	302	48.4
miR-2024b	0	11	N.D.
miR-2024c	39	450	25.9
miR-2024d	30	230	17.2
miR-2024e	46	472	23
miR-2024f	45	232	11.6
miR-2024g	44	154	7.9
miR-2025	2,414	1,025	1
miR-2026	421	164	0.9
miR-2027	5	25	11.2
miR-2028	0	25	N.D.
miR-2029	210	68	0.7
miR-2030	164	99	1.4
miR-2031	14	12	1.9
miR-2032a	0	2	N.D.
miR-2032b	0	1	N.D.
miR-2033	4	1	0.6
miR-2034	0	7	N.D.
miR-2035	0	13	N.D.
miR-2036	10	3	0.7
miR-2037	40	5	0.3
miR-2038	0	7	N.D.
miR-2039	0	1	N.D.
miR-2040a	0	3	N.D.
miR-2041	0	1	N.D.
miR-2042	0	1	N.D.
miR-2046	3	2	1.5
miR-2048	3	1	0.7
miR-2049	0	6	N.D.
miR-2051	1	0	0

Supplementary Table 2. Read counts for *Amphimedon* miRNAs in periodate-treated (periodate) and mock-treated (mock) libraries. To calculate the ratio of reads from the two libraries (mock / periodate), the read counts were first normalized to the total number of genome-matching reads present in each library. The ratio for miRNAs not cloned in the periodate-treated library but observed in the mock-treated library is not defined (N.D.).

miRNA	Periodate	Mock	Mock/Periodate
miR-2014	0	3,932	N.D.
miR-2015-3p	12	538	32
miR-2015-5p	6	549	66
miR-2016	26	20,056	554
miR-2017	0	120	N.D.
miR-2018	2	243	87
miR-2019	4	1,637	294
miR-2020	1	2,640	1,896
miR-2021	4	1,950	350

Supplementary Table 3. *Nematostella* genomic piRNA loci. Scaffold and coordinate values reference the *Nematostella* genome[13]. Read counts for small RNAs (27–30 nt) either possessing a 5'-U, or not (5'-V), were normalized by the number of genome matches for each small RNA. Strand bias is the percentage of match-normalized 5'-U reads on the strand that contains the majority of such reads. Periodate ratio is calculated as the number of match-normalized 5'-U reads from the periodate-treated library divided by the number from the mock-treated library.

Locus	Scaffold	Coordinates	5'-U reads	5'-V reads	Strand bias	Periodate ratio
1	11	1660000-1680000	7314.5	244.8	98.10%	1.24
2	15	1390000-1410000	2841.0	189.7	99.88%	1.18
3	23	460000-470000	1024.9	83.6	98.60%	1.05
4	24	600000-610000	2706.2	437.7	97.07%	1.30
5	24	1130000-1150000	6847.2	232.7	99.97%	1.22
6	25	610000-620000	1222.9	103.7	100.00%	1.12
7	27	1330000-1350000	6816.5	3010.6	99.88%	1.17
8	29	830000-840000	1708.5	243.4	99.99%	1.21
9	29	850000-860000	1356.0	60.1	99.80%	1.20
10	36	1030000-1040000	16163.2	88.6	99.23%	1.29
11	36	1050000-1070000	56062.0	320.7	100.00%	1.42
12	36	1080000-1090000	28982.9	151.5	100.00%	1.23
13	36	1130000-1140000	43039.7	11231.0	100.00%	1.21
14	55	920000-930000	1277.9	117.1	99.91%	0.96
15	56	60000-90000	5073.0	281.1	99.32%	1.08
16	56	130000-140000	1064.2	73.9	97.87%	1.70
17	56	170000-190000	3936.8	241.7	99.96%	1.22
18	57	1010000-1020000	1346.0	244.2	99.80%	1.25
19	60	780000-810000	6421.0	241.9	99.40%	1.22
20	68	70000-90000	2723.8	168.8	99.75%	1.21
21	68	260000-270000	1452.9	61.6	99.80%	1.10
22	80	180000-190000	3035.4	502.5	98.96%	1.48
23	82	130000-140000	1770.6	129.0	99.83%	1.34
24	83	190000-200000	1201.2	102.3	98.66%	1.59
25	87	340000-350000	2321.1	318.8	98.47%	1.03
26	89	680000-690000	10940.0	664.8	100.00%	1.36
27	90	570000-580000	1505.5	156.8	92.35%	1.38
28	105	690000-710000	2665.4	369.1	88.53%	1.41
29	106	160000-170000	1996.6	165.4	99.91%	1.22
30	106	580000-610000	17449.8	1976.2	99.96%	1.21
31	113	210000-240000	3836.2	210.4	99.43%	1.22
32	115	30000-40000	1516.6	84.3	99.92%	1.25
33	120	140000-160000	5879.7	273.9	99.97%	1.07
34	123	210000-230000	7610.1	389.3	99.99%	1.20
35	123	600000-620000	5477.5	337.2	96.97%	1.10
36	128	10000-20000	1204.8	86.6	99.04%	1.14
37	128	30000-40000	1984.5	66.8	99.96%	0.96
38	128	70000-90000	3213.3	163.2	99.99%	1.27
39	131	160000-200000	20780.9	740.0	99.95%	1.34
40	134	80000-100000	4436.6	265.8	99.98%	1.03
41	134	340000-360000	3289.8	928.3	99.12%	1.15
42	143	140000-160000	3878.4	414.2	99.93%	1.18
43	150	470000-490000	6718.2	653.1	99.90%	1.35
44	150	510000-520000	1365.5	250.3	96.53%	1.47
45	174	150000-160000	1940.6	115.7	99.96%	1.32
46	176	460000-470000	1257.2	55.2	100.00%	1.42
47	196	340000-360000	3763.2	834.0	99.93%	1.25
48	201	220000-230000	1409.5	16.7	99.90%	0.82
49	205	390000-400000	4187.9	223.5	99.48%	1.59
50	211	0-10000	1315.2	77.4	99.85%	1.38

51	218	310000-330000	3906.3	70.8	99.49%	1.52
52	220	340000-350000	1099.4	46.2	99.97%	1.07
53	220	360000-370000	1704.0	39.4	99.84%	1.11
54	226	90000-100000	1196.6	135.6	100.00%	1.28
55	227	170000-180000	1275.9	191.3	99.57%	1.29
56	228	340000-350000	1123.1	94.3	100.00%	0.99
57	252	290000-300000	1416.5	147.7	97.48%	1.05
58	272	60000-70000	1804.9	130.0	96.20%	1.46
59	278	150000-160000	3816.0	141.9	99.80%	1.26
60	279	210000-220000	1202.0	58.7	99.95%	1.23
61	279	230000-240000	5089.6	627.0	99.90%	1.13
62	281	160000-170000	1014.9	39.5	99.72%	1.07
63	282	290000-300000	1701.7	249.5	97.18%	1.35
64	287	80000-90000	1755.6	115.5	99.60%	1.40
65	291	70000-90000	7965.8	559.3	98.81%	0.98
66	293	30000-40000	1242.9	83.7	98.29%	0.97
67	293	50000-60000	1385.6	40.4	99.94%	0.83
68	311	20000-50000	9889.5	1017.5	99.96%	1.35
69	311	160000-170000	1278.5	88.0	81.51%	1.62
70	317	90000-120000	77720.8	1801.4	99.94%	0.99
71	328	60000-130000	12312.2	599.3	98.53%	1.31
72	377	110000-120000	1171.6	39.1	99.90%	1.41
73	383	10000-20000	1590.9	97.1	99.45%	1.18
74	392	60000-80000	4555.6	1282.5	99.93%	1.15
75	435	30000-40000	2175.2	141.5	99.95%	0.84
76	441	100000-110000	1416.9	69.7	85.90%	1.22
77	481	60000-70000	1605.2	108.7	99.87%	1.16
78	534	80000-90000	1318.3	228.0	99.47%	1.17
79	586	0-10000	1249.5	146.9	99.72%	1.51
80	755	10000-20000	2086.4	3175.8	97.09%	0.96
81	883	30000-40000	1029.6	66.0	99.96%	1.23
82	889	0-10000	1805.7	144.6	99.93%	1.21
83	943	30000-50000	6116.6	6237.6	99.05%	1.10
84	1084	20000-38783	59204.1	723.4	100.00%	1.37
85	1483	20000-28446	4955.9	296.8	99.94%	1.04
86	1799	10000-20000	1839.1	89.6	99.74%	1.19
87	1971	0-17665	6626.1	306.7	99.81%	1.21
88	2053	0-10000	2550.4	100.9	99.95%	1.24
89	2400	0-10000	1057.4	33.0	99.92%	1.56

Supplementary Table 4. *Amphimedon* genomic piRNA loci. Scaffold and coordinate values reference a preliminary *Amphimedon* assembly[16]. Read counts for small RNAs (24-30 nt) either possessing a 5'-U, or not (5'-V), were normalized by the number of genome matches for each small RNA. Strand bias is the percentage of match-normalized 5'-U reads on the strand that contains the majority of such reads. Periodate ratio is calculated as the number of match-normalized 5'-U reads from the periodate-treated library divided by the number from the mock-treated library.

Locus	Scaffold	Coordinates	5'-U reads	5'-V reads	Strand bias	Periodate ratio
1	497	0-5000	118.1	46.8	100.00%	2.22
2	2101	0-5000	339.9	82.4	100.00%	2.13
3	4121	0-5000	108.3	2.5	99.25%	1.25
4	6857	0-5000	113.5	49.2	96.01%	2.39
5	8973	0-5000	122.7	23.5	81.11%	2.74
6	9433	0-5000	218.8	63.1	76.19%	3.64
7	10098	0-5000	144.0	36.0	100.00%	2.25
8	10330	0-5000	171.8	67.8	80.49%	2.09
9	10809	0-5000	172.9	42.9	96.53%	2.55
10	11666	5000-10000	117.8	27.3	86.12%	2.49
11	11673	0-5000	129.2	9.5	99.47%	2.56
12	11994	0-5000	128.5	54.5	99.61%	2.31
13	12254	5000-10000	139.7	28.0	78.06%	2.59
14	12435	5000-10000	139.8	13.0	99.43%	2.56
15	12643	15000-20000	221.6	11.5	100.00%	2.43
16	12809	25000-30000	131.3	22.4	95.18%	2.27
17	13010	0-5000	153.7	179.1	64.38%	2.41
18	13072	30000-35000	157.1	45.7	91.99%	2.08
19	13111	90000-95000	130.1	35.0	99.98%	2.05
20	13125	85000-90000	149.8	31.0	88.65%	2.50
21	13132	35000-40000	156.6	13.3	93.55%	2.44
22	13165	45000-50000	182.3	20.2	98.90%	2.14
23	13165	85000-90000	129.8	5.7	99.98%	2.29
24	13176	105000-110000	212.7	21.3	99.89%	2.25
25	13219	35000-40000	112.7	17.1	50.27%	2.42
26	13224	60000-65000	167.0	43.0	100.00%	2.98
27	13229	95000-100000	248.6	28.8	99.60%	2.47
28	13244	65000-70000	133.0	4.9	100.00%	2.23
29	13248	130000-135000	169.9	11.6	98.90%	2.23
30	13261	45000-50000	107.0	4.8	99.92%	2.28
31	13273	40000-45000	104.3	22.8	99.76%	2.40
32	13273	55000-60000	106.7	3.8	100.00%	3.05
33	13282	20000-25000	116.0	83.1	100.00%	1.56
34	13352	90000-95000	1061.0	53.0	100.00%	2.05
35	13385	120000-125000	193.0	9.0	100.00%	3.28
36	13385	185000-190000	166.0	7.8	98.46%	2.56
37	13389	75000-80000	124.0	25.0	88.71%	2.61
38	13391	95000-100000	287.0	1249.0	76.66%	2.64
39	13405	225000-230000	115.3	15.7	93.73%	1.93
40	13406	60000-65000	117.6	97.7	87.99%	2.69
41	13406	205000-210000	148.5	66.5	93.60%	2.82
42	13418	5000-10000	142.7	40.2	97.19%	2.23
43	13418	25000-30000	111.0	13.2	99.00%	3.04
44	13419	160000-165000	142.0	27.0	83.10%	2.18
45	13459	160000-180000	2137.4	571.4	99.98%	2.11
46	13460	165000-170000	149.0	86.0	100.00%	1.78
47	13472	55000-60000	160.0	21.3	99.79%	1.89
49	13473	190000-195000	258.5	36.0	99.03%	3.06

50	13474	310000-315000	190.2	33.0	100.00%	2.43
51	13475	220000-225000	107.4	76.7	95.41%	3.17
52	13475	310000-315000	179.6	85.2	84.65%	1.78
53	13476	20000-25000	100.2	2.1	99.63%	1.47
54	13487	345000-350000	100.7	11.7	76.46%	2.55
55	13491	105000-110000	230.0	41.0	97.83%	2.35
56	13492	25000-30000	163.5	16.3	97.15%	2.33
57	13501	310000-315000	294.1	72.8	92.18%	2.43
58	13502	410000-415000	116.0	37.6	96.51%	2.16
59	13503	435000-440000	100.7	13.6	99.76%	2.12
60	13503	460000-465000	117.1	25.2	91.46%	2.44
61	13503	525000-530000	231.8	1.2	99.86%	1.42
62	13516	245000-250000	175.1	52.8	86.53%	2.24
63	13516	335000-340000	157.1	15.5	80.11%	2.06
64	13517	470000-475000	123.5	35.3	99.54%	2.04
65	13517	810000-815000	144.5	15.5	97.23%	1.59
66	13517	890000-895000	133.0	61.0	98.50%	2.32
67	13521	140000-145000	461.0	89.0	100.00%	1.83
68	13521	505000-510000	120.1	13.3	97.22%	2.61
69	13521	680000-685000	195.0	35.0	98.97%	2.19
70	13521	1075000-1080000	226.5	6.6	90.67%	2.46
71	13521	1345000-1350000	103.2	14.2	99.94%	2.43
72	13521	1505000-1510000	151.9	7.6	97.59%	2.82
73	13521	1875000-1880000	281.0	11.0	100.00%	2.91

Supplementary Table 5. Annotated *Nematostella* coding regions with many reads resembling piRNAs. Listed are *Nematostella* predicted coding sequences (identifiers reference *Nematostella* genome project[13]) with the highest number of 27–30mer 5'-U match-normalized read counts from the periodate-treated library (5'-U reads). Coding sequences with no uniquely matching reads were omitted. All 27–30mer counts were separated into sense and antisense match-normalized read counts. Proteins or domains with similarity to the *Nematostella* annotations are indicated, together with whether the *Nematostella* sequence was considered likely to encode a transposase.

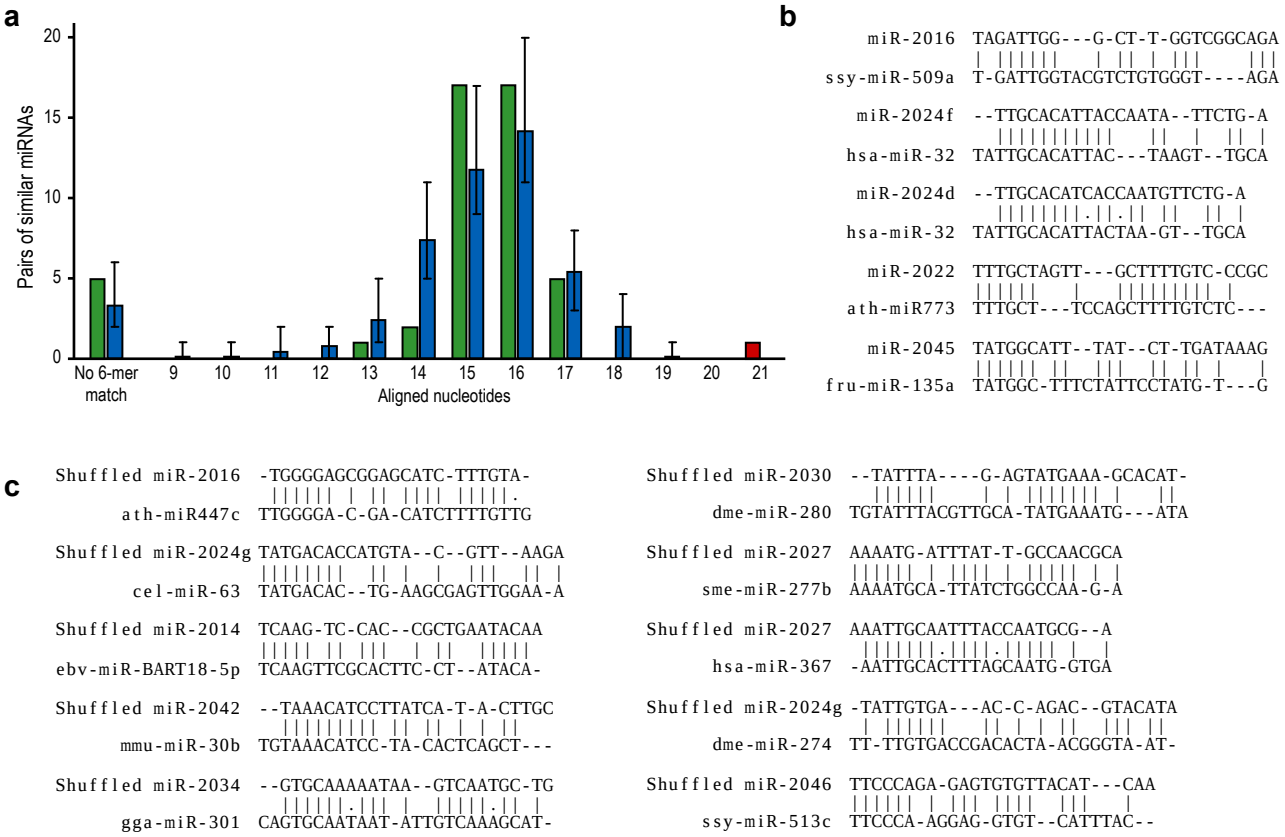
Identifier	5'-U reads	Antisense reads	Sense reads	Similarity	Transposase
>215197	1712.7	1865.5	6.3	Gag-pol	Yes
>245476	1487.0	1949.0	93.5	Novel	No
>210822	1396.3	1496.8	1.9	Novel	No
>200314	1168.5	1770.0	259.0	Endonuclease/RTase	Yes
>242950	1119.0	1372.0	66.5	Novel	No
>210517	1115.8	1224.3	3.5	Potassium channel	No
>207394	1050.0	1470.7	473.0	Endonuclease/RTase	Yes
>244204	1017.0	1128.5	4.0	Novel	No
>248843	1002.4	1034.9	37.4	Zinc finger	No
>243277	874.2	1070.7	30.3	Gag-pol transposase	Yes
>247570	846.6	605.6	811.6	Novel	No
>246865	840.3	1174.4	361.9	Fuktinin	No
>246936	768.6	1103.3	244.2	Novel	No
>244474	754.0	949.0	1.0	Novel	No
>220709	746.0	910.5	37.5	Protease	No
>208216	697.0	883.0	0.0	Piggybac Transposase	Yes
>220102	662.0	87.5	1307.5	Ubiquitin ligase	No
>50157	629.0	718.0	27.0	ATPase	No
>210818	626.8	696.4	1.8	Novel	No
>200367	622.5	918.0	348.0	ReqQ helicase	No
>205414	599.5	670.5	42.5	Protein kinase	No
>241523	587.0	512.0	856.0	Ty3- RTase	Yes
>114840	580.0	623.2	7.0	Transposase	Yes
>52268	562.0	627.0	5.0	Endonuclease/RTase	Yes
>221645	541.6	623.1	0.5	Proteoglycan	No
>211219	498.0	523.0	0.0	Ubiquitin ligase	No
>248766	493.9	709.3	17.7	Gag-pol	Yes
>154922	491.6	552.2	5.3	Endonuclease/RTase	Yes
>248752	481.7	553.5	19.3	Novel	No
>245073	468.7	478.7	18.0	Fbox	No
>217332	461.1	500.0	3.0	Zinc finger	No
>211739	454.2	455.2	0.0	Ubiquitin ligase	No
>219373	444.8	437.8	2594.8	NOD protein	No
>222811	422.2	494.2	11.7	Gag-pol	Yes
>140434	421.0	523.0	0.0	Gag-pol	Yes
>221673	415.7	468.0	1.6	Mucin	No
>219929	414.1	466.8	15.7	Novel	No
>218285	409.0	495.0	10.0	Exonuclease	No
>199093	408.3	514.1	3.3	Endonuclease/RTase	Yes
>244691	403.0	452.5	1.5	Novel	No

Supplementary Table 6. Annotated *Amphimedon* coding regions with many reads resembling piRNAs. Listed are *Amphimedon* predicted coding sequences[16] with the highest number of 24–30mer 5'-U match-normalized read counts from the periodate-treated library (5'-U reads); otherwise as in Supplementary Table 5.

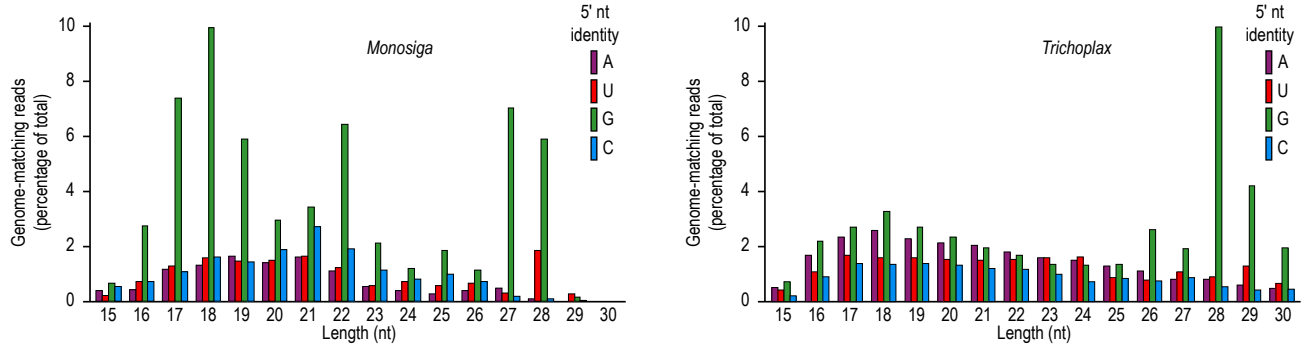
Identifier	5'-U reads	Antisense reads	Sense reads	Similarity	Transposase
>g23214.t1	500.5	656.5	13.0	Gag-pol	Yes
>g5547.t1	482.0	544.0	93.0	Splicing factor	No
>g15322.t1	409.0	151.0	270.0	Hobo transposase	Yes
>g31604.t1	373.3	418.7	0.6	Novel	No
>g29521.t1	350.0	441.0	0.0	Novel	No
>g15134.t1	301.0	367.0	53.0	Transposase	Yes
>g21226.t1	251.0	262.5	0.0	Acetyl-CoA transferase	No
>g6662.t1	248.0	281.0	85.0	K-channel	No
>g12684.t1	246.0	289.5	7.5	ReqQ helicase	No
>g6913.t1	235.0	262.5	0.0	TNF receptor	No
>g22087.t1	223.0	244.0	0.0	ReqQ helicase	No
>g5109.t1	213.0	239.0	21.0	Zinc finger	No
>g7287.t1	189.5	202.5	0.0	Novel	No
>g9667.t1	186.0	199.5	42.5	Transposase	Yes
>g10820.t1	184.0	202.8	0.0	Kinase	No
>g4684.t1	179.0	241.8	10.0	Without children	No
>g12681.t1	178.0	212.0	9.5	Novel	No
>g22507.t1	175.0	189.5	8.5	Transposase	Yes
>g15924.t1	171.0	180.0	51.0	Ubiquitin ligase	No
>g26911.t1	162.5	208.5	0.0	Novel	No
>g5558.t1	162.0	178.5	0.0	NAD dependent Epimerase	No
>g23203.t1	155.5	169.0	37.0	Transposase	Yes
>g23181.t1	155.0	206.0	23.5	Zinc finger	No
>g2110.t1	145.9	164.4	0.6	Novel	No
>g13978.t1	144.0	156.8	3.1	Novel	No
>g13979.t1	140.5	155.0	0.6	Novel	No
>g1147.t1	135.0	187.3	0.0	Novel	No
>g25864.t1	132.0	160.0	17.0	Novel	No
>g14890.t1	130.0	127.0	68.0	Integrase	Yes
>g10998.t1	129.5	176.5	56.5	Transposase	Yes

Supplementary Table 7. Libraries sequenced on the Illumina platform.

Sample	Size selection (nt)	Sequencing reactions	Genome-matching reads
Nematostella	15-24	1	1,140,549
Nematostella	25-30	1	1,136,886
Nematostella	15-30	1	664,381
Trichoplax	15-30	1	420,634
Amphimedon-adult	15-30	3	1,014,098
Amphimedon-embryonic	15-30	3	1,457,341
Nematostella [periodate-treated]	15-30	2	770,629
Nematostella [mock-treated]	15-30	1	343,395
Amphimedon-adult [periodate-treated]	15-30	2	490,047
Amphimedon-adult [mock-treated]	15-30	1	682,537

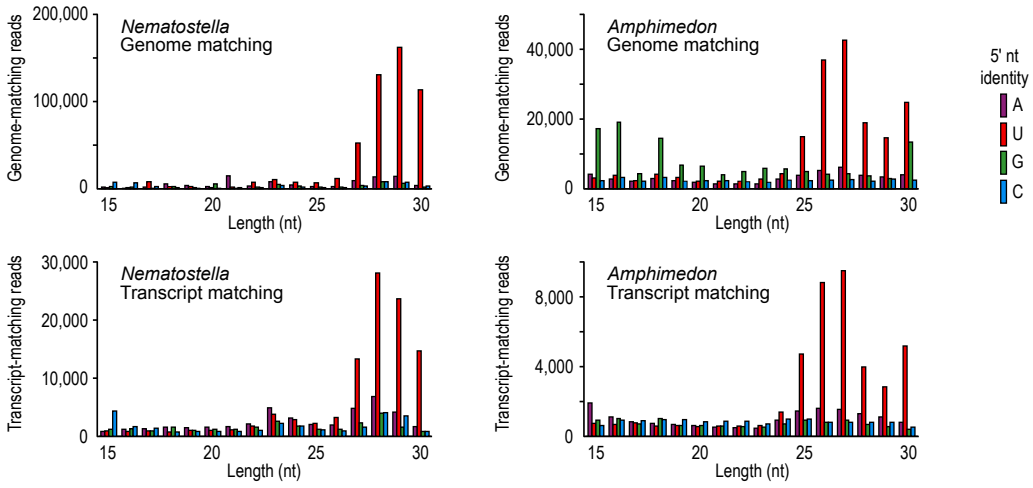


Supplementary Figure 1: Similarity between newly identified miRNAs and previously annotated miRNAs. a, Number of aligned nucleotides between miRNA pairs comprised of *Nematostella* and *Amphimedon* miRNAs each paired with its best match to a previously annotated miRNA. To look for miRNAs related to those of *Nematostella* and *Amphimedon*, we first compared each *Nematostella* and each *Amphimedon* miRNA to all miRNAs annotated in miRBase, searching for pairs of miRNAs that shared a 6-mer within their first 8 nucleotides. For all such pairs, we then aligned the two miRNAs (Needleman-Wunsch alignment with gap penalty of 3 and gap extension penalty of 0.5), and chose the highest scoring pairing for each of the *Nematostella* and *Amphimedon* miRNAs (plotted in green). To simulate the distribution of high-scoring pairs attributable to random similarities between these short oligonucleotide sequences, we repeated the analysis with ten cohorts of dinucleotide-shuffled *Nematostella* and *Amphimedon* miRNAs (plotted in blue is the average value across ten cohorts, with error bars indicating the range of values obtained for different cohorts). Overall, there was no significant difference between the distributions derived from real and shuffled miRNAs (Wilcoxon-rank sum test, $P = 0.54$). The outlying value, (in red) is miR-100 (Fig. 2e). **b,** The top-scoring pairs of basal animal and miRBase miRNAs. Two of these pairs illustrate similarity between miR-2024d/f and miR-32, a conserved bilaterian miRNA, whereas the others illustrate similarity to miRNAs that are not conserved among Bilateria. Although the similarity to miR-32 is less likely than the others to be due to chance because miR-32 is broadly conserved among Bilateria, homology between miR-2024d/f and miR-32 is impossible to assert with confidence because the number of identical nucleotides (17) observed for these pairs was frequently observed for the shuffled control miRNAs (panel a). Moreover, the 5' terminal nucleotides critical for miRNA function were offset by 2 nucleotides. **c,** The ten top-scoring pairs involving unique shuffled control sequences and miRbase miRNAs. Two of these pairs include members of conserved bilaterian miRNA families (miR-30b and miR-367), illustrating further the difficulty of asserting that the similarity observed between miR-2024d/f and miR-32 was more than expected by chance.

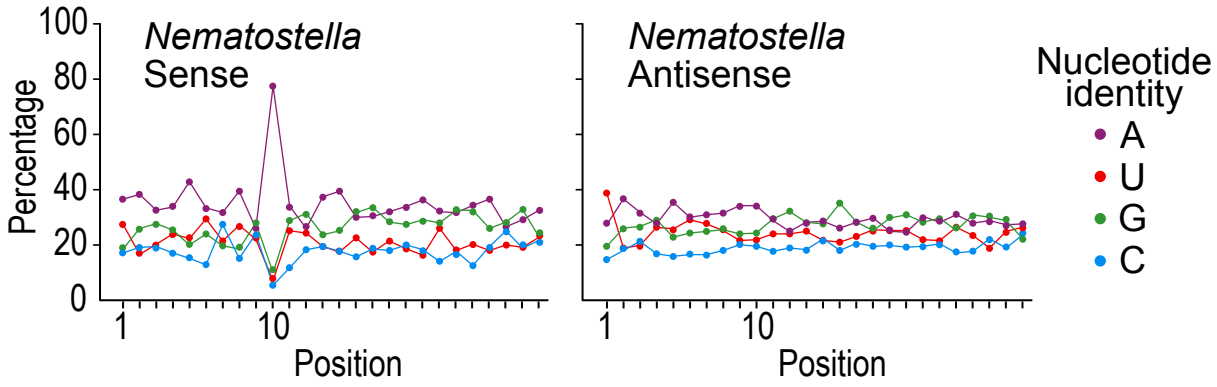


Supplementary Figure 2: *Monosiga brevicollis* and *Trichoplax adhaerens* sequencing results. Shown is the length distribution of genome-matching sequencing reads, plotted by 5'-nucleotide identity as in Fig. 2a. A total of 16,064 and 420,634 reads were obtained from the *Monosiga* and *Trichoplax* libraries, respectively; matches to ribosomal DNA were omitted in the figure. Because we found no evidence for miRNAs in *Trichoplax* or *Monosiga*, we explored whether the size of the datasets obtained for these species would have been sufficient to identify miRNAs in *Nematostella* and *Amphimedon*, by simulating the fraction of *Nematostella* and *Amphimedon* miRNAs we would have identified if the datasets for these analyses were reduced to the size of the *Trichoplax* and *Monosiga* datasets. (Dozens of miRNA genes were previously identified in Bilateria and plants using small-RNA datasets with only ~300 reads[41,42], but the fraction of reads attributed to miRNAs in both *Nematostella* and *Amphimedon* was far lower, in large part because piRNAs contribute such a large fraction of reads in these species, and thus the cases of *Nematostella* and *Amphimedon* provide more stringent tests.) Our simulations (100) required that the previously identified miRNA and miRNA* sequences were found at least twice and at least once, respectively (the same requirements used in our the original analysis), in randomly selected sets of sequences of the size of the *Monosiga* and *Trichoplax* datasets. For *Amphimedon* we would have found 3.7 (s.d., 0.9; range 2–6) miRNAs (out of a total of 8) if the dataset was reduced to 16,064 reads, and 7.3 (s.d., 0.7; range 6–8) if the dataset was reduced to 420,634 sequences. For *Nematostella*, the analogous values were 4.3 (s.d., 1.7; range 1–9) and 19.2 (s.d., 1.8; range 15–24) out of a total of 40. Therefore, even if the fraction of small RNAs deriving from miRNA were as low as observed in *Nematostella* or *Amphimedon*, a possibility made all the more unlikely by the absence of piRNAs in *Monosiga* and *Trichoplax*, the *Trichoplax* and *Monosiga* datasets would have been sufficient to find miRNAs.

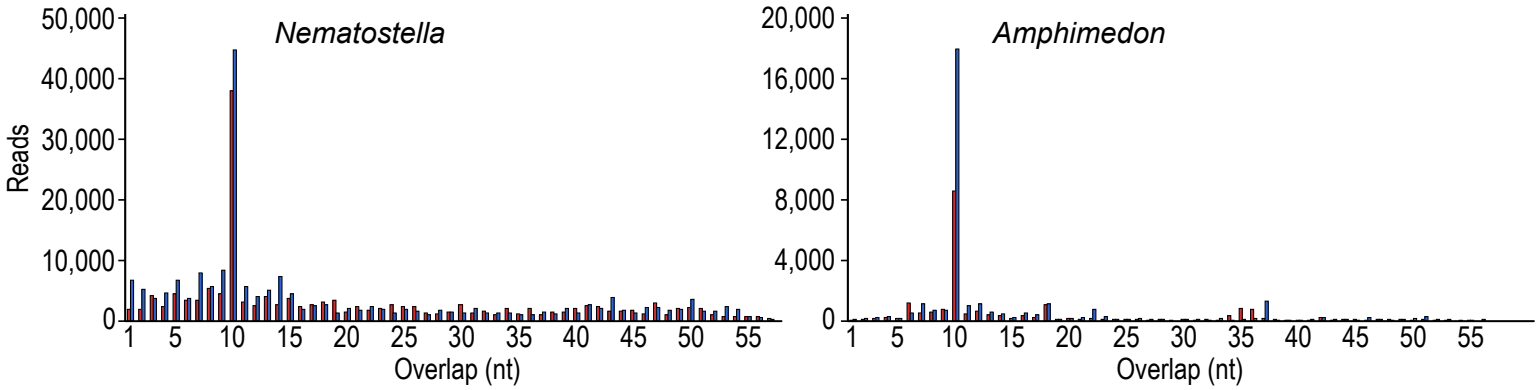
Supplementary Figure 3: Example of a predicted *Trichoplax adhaerens* hairpin that was rejected as a miRNA hairpin. The sequence of the hairpin is depicted above its bracket-notation secondary structure, as predicted by RNAfold. The sequenced small RNAs mapping to the hairpin are aligned below, with the number of reads shown on the right and the designated miRNA and inferred miRNA* species colored red and blue, respectively. The candidate hairpin exhibited multiple features inconsistent with encoding a miRNA: (i) the predicted star sequence was not cloned, rendering it impossible to confidently predict this locus as a miRNA gene; (ii) the miRNA:miRNA* duplex would contain three large bulged segments, features inconsistent with known miRNAs:miRNA* duplexes; (iii) numerous species derived from regions of the hairpin distinct from those of the candidate miRNA and miRNA*; (iv) numerous variant miRNA species were sequenced, many of a length atypical of miRNAs. These characteristics, especially point (ii), indicate that this locus, the most miRNA-like we found in *Trichoplax*, does not encode a miRNA.



Supplementary Figure 4: Length distribution of *Nematostella* and *Amphimedon* sequenced small RNAs resistant to periodate treatment. Reads matching the respective genomes (top panels) or those matching annotated coding regions (bottom panels) are plotted by 5'-nucleotide.



Supplementary Figure 5: Nucleotide composition of periodate-resistant 22–26-nucleotide RNAs matching the indicated strand of *Nematostella* annotated coding regions. The analysis was performed as for Figure 4C, except that only 22–26-nucleotide small RNAs were considered. In contrast to the results for the longer reads (Fig. 4C), the antisense matching reads of this shorter length did not exhibit a pronounced bias for a 5' U; the sense matching reads, however, did exhibit an enrichment for an A at position 10. These data suggest that the 22–26-nucleotide sense small RNAs can be produced as or derived from secondary piRNAs via a ping-pong mechanism [4], whereas the 22–26-nucleotide antisense reads are either rarely produced as piRNAs via a ping-pong mechanism, or they are degradation intermediates in a pathway that trims off the 5' terminal residues. Consistent with the production of sense and antisense piRNAs of different sizes, the sense piRNAs of *Drosophila* (Ago3-bound) tend to be shorter than antisense piRNAs (Piwi/Aub-bound). We did not observe this phenomenom in our *Amphimedon* data.



Supplementary Figure 6: Pairing between small RNAs that are sense and antisense matches to annotated coding regions. Sense and antisense piRNA partners that initiate and arise from the ping-pong amplification cascade pair to each other at their 5' ends, with an overlap of 10 nucleotides[4]. To examine whether overlapping sense and antisense piRNAs tended to exhibit this characteristic pairing, we calculated the number of nucleotides of overlap between all sense and antisense periodate-resistant small RNAs that matched annotated coding regions. For *Nematostella*, sense-matching reads of 22–30-nucleotides (Supplemental Fig. 5) and antisense-matching reads of 27–30-nucleotides (Fig. 4C) were considered. For *Amphimedon*, sense- and antisense-matching reads of 24–30-nucleotides were considered. For each instance of overlap, we tallied the number of times the sense (red bars) and antisense (blue bars) small RNA was sequenced, normalized by the number of times that sequence matched annotated coding sequences. Plotted are the sums of the tallies at each overlap register. For overlaps of 10 nucleotides, 98% (*Nematostella*) and 88% (*Amphimedon*) of the sense reads had an A at position 10, and 96% (*Nematostella*) and 93% (*Amphimedon*) of the antisense reads had a 5'-U.