Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*

Yi-Chien Lu, Magda Smielewska, Dasaradhi Palakodeti, et al.

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2009/06/25/rna.1702009.DC1.html
P<P	Published online June 24, 2009 in advance of the print journal.
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

REPORT

Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*

YI-CHIEN LU,[1] MAGDA SMIELEWSKA,[1] DASARADHI PALAKODETI,[1] MICHAEL T. LOVCI,[2] STEFAN AIGNER,[3] GENE W. YEO,[2] and BRENTON R. GRAVELEY[1]

[1]Department of Genetics and Developmental Biology, University of Connecticut Stem Cell Institute, University of Connecticut Health Center, Farmington, Connecticut 06030-3301, USA
[2]Department of Cellular and Molecular Medicine, Stem Cell Program, UCSD Moores Cancer Center, University of California, San Diego, La Jolla, California 92037, USA
[3]Laboratory of Genetics, Salk Institute, La Jolla, California 92037, USA

ABSTRACT

MicroRNAs (miRNAs) play important roles in directing the differentiation of cells down a variety of cell lineage pathways. The planarian *Schmidtea mediterranea* can regenerate all lost body tissue after amputation due to a population of pluripotent somatic stem cells called neoblasts, and is therefore an excellent model organism to study the roles of miRNAs in stem cell function. Here, we use a combination of deep sequencing and bioinformatics to discover 66 new miRNAs in *S. mediterranea*. We also identify 21 miRNAs that are specifically expressed in either sexual or asexual animals. Finally, we identified five miRNAs whose expression is sensitive to γ-irradiation, suggesting they are expressed in neoblasts or early neoblast progeny. Together, these results increase the known repertoire of *S. mediterranea* miRNAs and identify numerous regulated miRNAs that may play important roles in regeneration, homeostasis, neoblast function, and reproduction.

Keywords: planarians; microRNAs; stem cells; deep sequencing; bioinformatics

INTRODUCTION

Planarians are an excellent and increasingly popular model system for studying stem cell biology, as they have the ability to completely replace tissue lost after injury (Newmark and Sanchez Alvarado 2002). This remarkable regenerative capacity is made possible by the presence of a population of pluripotent somatic stem cells called neoblasts that can give rise to all cell types in the animal, including germ cells (Baguñà et al. 1989; Wang et al. 2007), and comprise ∼30% of the entire organism. In addition to being required for regeneration, neoblasts are also essential for homeostasis where they replace cells lost during normal tissue turnover.

The planarian species that is predominantly studied and particularly amenable to molecular biology analysis is *Schmidtea mediterranea* (Sanchez Alvarado 2006). Two strains of *S. mediterranea* have been identified that differ significantly in their mode of reproduction (Newmark and Sanchez Alvarado 2002). One strain undergoes sexual reproduction as cross-fertilizing hermaphrodites and develops both testes and ovaries. In contrast, the second strain never develops reproductive organs but rather reproduces by asexual fission. The only known genetic difference between these two strains is a chromosomal translocation (Baguñà et al. 1999), although the precise loci involved are unknown.

MicroRNAs (miRNAs) are small (∼22 nucleotides [nt]) noncoding RNAs that regulate gene expression by base pairing to mRNAs where they direct either mRNA cleavage or translation repression (Eulalio et al. 2008). miRNAs are generated from longer primary transcripts called pri-miRNAs through a series of processing events (Kim 2005). In the first step of miRNA synthesis, the stem–loop structure containing the miRNA is cleaved from the pri-miRNA by the RNaseIII enzyme Drosha and its cofactor DGCR8 to generate a pre-miRNA. The pre-miRNA is then exported from the nucleus to the cytoplasm by exportin5 where it is

further processed by a second RNaseIII enzyme Dicer and its cofactor TRBP. The mature miRNA is then loaded onto an Argonaute protein (Ago1, Ago2, Ago3, or Ago4 in humans) where it then interacts with and regulates the mRNA target.

We previously described the cloning and characterization of 63 miRNAs from *S. mediterranea* (Palakodeti et al. 2006). As these miRNAs were identified using traditional cloning and sequencing methods, these efforts were not exhaustive and it is likely that additional miRNAs remained to be identified in *S. mediterranea*. Here we use deep sequencing to further explore the diversity of *S. mediterranea* miRNAs and their expression patterns. Specifically, we expand the miRNA repertoire of *S. mediterranea* and identify 66 new miRNA encoding loci. We also identify 21 miRNAs that are expressed specifically in the sexual or asexual strains and five miRNAs that decrease upon irradiation. The regulated miRNAs are particularly intriguing as they may play important roles in the development of reproductive organs, the control of reproductive fission, and regulating the function of neoblasts or their progeny during regeneration and homeostasis.

RESULTS AND DISCUSSION

Deep sequencing of planarian miRNAs

To identify additional *S. mediterranea* miRNAs we prepared four small RNA libraries from RNA isolated from nonirradiated sexual and asexual animals and from sexual and asexual animals four days after γ-irradiation. FACS analysis confirmed that the γ-irradiation treatment significantly depleted the animals of both dividing (X1) and nondividing (X2) neoblasts (Fig. 1A). Furthermore, RT-PCR of the four RNA samples demonstrated that the neoblast marker *smedwi-1* (Reddien et al. 2005) was present at significantly lower levels in the irradiated samples than in the nonirradiated samples, while *smed-actin* was unaffected by irradiation (Fig. 1B). As piRNAs are significantly more abundant than miRNAs in *S. mediterranea* (Palakodeti et al. 2008), we first gel purified RNAs between 18 nt and 28 nt in length to enrich the libraries for miRNAs. Each library was sequenced in one lane of an Illumina Genome Analyzer II for 36 cycles. We obtained 11,059,352 and 11,826,579 individual reads from the nonirradiated sexual and asexual libraries, respectively, and an additional 9,343,092 and 10,703,240 individual reads from the irradiated sexual and asexual libraries, respectively. A total of 19,635,184 of these reads were between 18 and 24 nt in length.

Detailed analysis of the 18–24 nt reads revealed that they map to all 63 previously identified miRNA precursors. However, our data suggests that three of the previously reported miRNAs are not actual miRNAs. First, although we obtained 9949 and 21,604 reads that perfectly aligned to
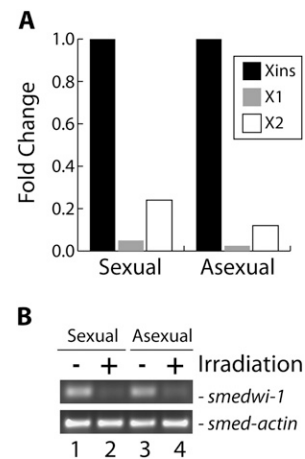

FIGURE 1. Sample quality control. (*A*) FACS analysis to monitor the neoblast population in untreated or γ-irradiated animals. The fold change of the X1, X2, and Xins populations in the irradiated sample compared with the untreated sample is graphed. The fold change of the Xins population was arbitrarily set to a value of 1.0 and the values for X1 and X2 adjusted accordingly. (*B*) Analysis of *smedwi-1* mRNA level. The level of *smedwi-1* mRNA, a neoblast marker (Reddien et al. 2005), was analyzed using semiquantitative RT-PCR. As a control, a parallel experiment was performed for *smed-actin* mRNA.

the *mir-749* and *mir-751* pre-miRNAs, respectively, the boundaries of the reads on the predicted pre-miRNA structure are inconsistent with that predicted for Drosha and Dicer processing events (data not shown). Second, *mir-753* appears to correspond to a repetitive element as it is present >520 times in the draft assembly of the *S. mediterranea* genome. Thus, our current data suggest that the previously reported miRNAs *mir-749*, *mir-751*, and *mir-753* (Palakodeti et al. 2006) are not authentic miRNA encoding loci in *S. mediterranea*.

The sequencing data has also allowed us to more accurately annotate the known miRNAs (Fig. 2A). For example, in our initial study we obtained two sequences of different lengths for *mir-755* leading to the annotation of the *mir-755* mature strand as either 18 or 20 nt in length (Palakodeti et al. 2006). We have now obtained 27,268 reads that map perfectly to the *mir-755* pre-miRNA and analysis of the most abundant reads that map to the 5′ or 3′ stems (Fig. 2A) suggests that the mature strand is actually 22 nt. Thus, the deep sequencing has dramatically improved the annotation of the known *S. mediterranea* miRNAs.

We have also utilized an algorithm called MIResque (GW Yeo and S Aigner, unpubl.) to identify new miRNA encoding loci from the deep-sequencing data (Table 1). This algorithm analyzes sequence reads that have been mapped to the genome, searches for clusters of reads, performs RNA folding on the sequence encompassed by the reads, and then assigns a score to the potential miRNA hairpin. The assigned score is derived from a regularized least-squares classification algorithm designed to score unknown miRNA hairpins given a positive set of known miRNA
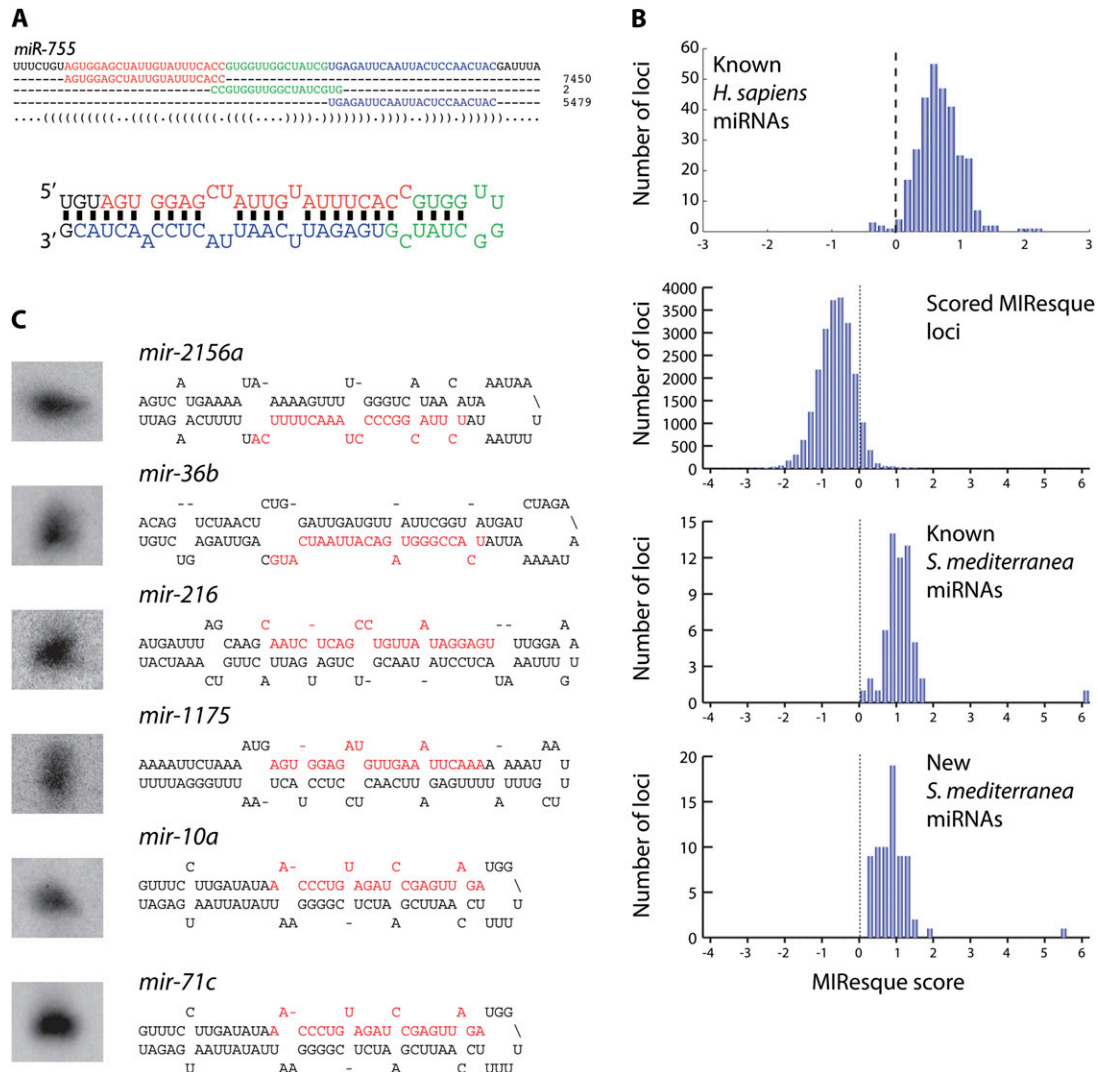
FIGURE 2. Deep sequencing of planarian miRNAs. (*A*) Alignment of sequence reads to *pre-miR-755*. The most frequently obtained mature, loop, and miRNA* strand sequences are highlighted in red, blue, and green, respectively, and the number of total reads obtained for each sequence are shown on the *right*. The location of these sequences on the secondary structure of the pre-miRNA is depicted below the alignment. (*B*) Histograms of the MIResque scores for known human miRNAs, all *S. mediterranea* loci containing clusters of deep sequence reads, known *S. mediterranea* miRNAs, and new *S. mediterranea* miRNAs identified in this study. The dotted line indicates the threshold at which a genomic region is considered to be a candidate miRNA locus. (*D*) Northern blot validation of MIResque predicted miRNAs. The structure for each pre-miRNA is shown on the *right*.

hairpins generated by deep sequencing. The classification strategy encompasses highly characteristic structural and sequence features of miRNAs and their precursors, such as duplex quality, average length of the mature miRNAs, and heterogeneity of the ends of the reads. MIResque utilized a data set of deep-sequencing reads aligned to a subset of known human miRNA hairpins as a training set. After training, MIResque analyzed the remainder of the known human miRNA hairpins and scored 98% (297/304) of these above the threshold assigned for a new miRNA (Fig. 2B). Interestingly, some of the poorly scored human miRNAs are actually not miRNAs, but rather ribosomal RNA (GW Yeo and S Aigner, unpubl.).

To identify as many new planarian miRNAs as possible, we pooled the sequence data obtained from all four samples and mapped them to the draft assembly of the *S. mediterranea* genome while allowing for up to two mismatches. Of the 60 previously identified *S. mediterranea* miRNA loci with sequence reads mapping to both the 5′ and 3′ stems, 57 (95%) scored above the threshold (Fig. 2B). Interestingly, three miRNAs—*let-7a*, *mir-1c*, and *mir-754a*—were not scored by MIResque. First, *mir-754a* did not receive a score because we only obtained reads that mapped to the mature strand. Second, *mir-1c* was not scored because numerous reads were obtained whose 5′ ends mapped to the beginning of the loop and the 3′ ends mapped in the

TABLE 1. New miRNA encoding loci identified from deep-sequencing data

miRNA	5p Sequence	5p Reads	3p Sequence	3p Reads	Orthologs	Genomic location	MIResque score
mir-2206	AGGGCAGGCAUAGCUAAGUG	826	ACCACUGGUGCGAUCCUGCC	287	N	Contig9421:12897-12977	5.55
mir-9a	UCUUUGGUUUUCUAGCUGUAUGA	4801	ACACACUAGAUUACCAAUUUC	485	Y	Contig2917:42342-42422	1.85
mir-36b	GAUUGAUGUUAUUCGGUAUGAUC	21	UCACCGGUAGACAUUAAUCAU	37,678	Y	Contig21980:6311-6403 Contig20276:577-669	1.48
mir-753b-1	AGGCUUGAAUUGUGAUCUCA	119	UGAGAUCAUUGUUCAAGCUCUUA	501	N	Contig7065:2463-2533	1.4
bantam-c-2	AGGCUUUCAUUAUUGAUUUUAA	6	UGAGAUCAUUAUGAAAGCUUUU	3996	Y	Contig483:150256-150320	1.39
mir-10a	AACCCUGUAGAUCCGAGUUAGA	23,674	AAUUCGAAUCUCGGGAAUU	12641	Y	Contig3771:6006-5925	1.33
mir-216	CAAUCUCAGCCUGUUAAUAGGAGU	281	UCCUAUAAACGCUCGAUGAUUAC	5590	Y	Contig23847:7701-7593	1.31
mir-87d	AUGUCUUAUAUUUUGCUAUUAUC	104	GUGAGCAAAGUUUAAGACAUUCA	2416	Y	Contig19991:5073-5173 Contig14046:2088-2174	1.31
mir-2159	UGCCUUUUCUUUGAGUAGCU	208	GAGCUGCCAAAUAAAGGGUGAU	310	N	Contig9447:14581-14679	1.3
mir-76	UCAUUCUUAAACAAUGAAUA	25	UUCGUUGUUUAUGAAUGAGAGA	158	Y	Contig20679:1219-1316 Contig18565:5370-5464 Contig26227:10793-10696	1.28
mir-2164	UAUAUUUGAUUCACGUAGAGA	69	UUGUGCGUGAUAUCAAGUAUAAU	128	N	Contig13809:14676-14764 Contig5497:46150-46062	1.28
mir-2151	AGAUCAUGUUAAGUGGAAUUUGAU	137	GAUUGCACUCAAUAUGGUCAGA	2497	N	Contig2742:61496-61583	1.24
mir-61b	AGAUCGUGUCGUUCUAGAUUA	634	UGACUAGAAUGUUCACCUUCUUU	2219	Y	Contig3451:4510-4605	1.2
mir-2168	CAUAUUCGUUAAGGUUGUAAC	42	CACAACCUUUGACGAAUGAG	26	N	Contig508:30920-31012	1.19
mir-1175	AGUGGAGAUGUUGAAAUUCAAA	44,455	UGAGAUUCAACUCCUCCUACU	1023	Y	Contig3917:32836-32925	1.18
mir-2172	GAAAUAAUGACUAUGCAUGGGACA	531	UACAUGUAUAGUAUUAUUUGCC	33	N	Contig2233:53295-53396	1.17
mir-36c	GAUGAUUGUCCAACUGGUUU	986	UCACCGGUAGACAUUCAUUCU	34,631	Y	Contig3934:37301-37396	1.13
mir-1993	GAUGAAUCAACAGUAUAAAACG	207	UAUUAUGCUGUAUUCAUGAG	5454	Y	Contig1563:42995-42882	1.05
mir-2163	AACUUGAUAGAUAUAUGCAUC	809	GCAUAGUCUAUUAAGUUGGU	483	N	Contig21390:5650-5760	1.04
mir-2173	ACAACUUGUAUAUGCAUACGUA	1	UACAUGUACAUAUAAGUUGAAU	303	N	Contig5049:3190-3280 Contig11322:1643-1543	1.03
mir-87c	UCAAUUGAUAAUGGGCCCCA	104	AUGAGCAAAUUAUCAAUUGACU	551	Y	Contig2469:10419-10510	1.01
mir-1992	UGUCUUGAACAACUGCUUAU	89	UCAGCAGUUGUUCCAUUGACA	157	Y	Contig11270:11976-11879	1
mir-753b-2	AGGCUUGAAUUGUGAUCUCA	109	UGAGAUCAUUAUUCAAGCUCUUC	1548	N	Contig7504:7760-7822	0.97
mir-2156a	AAAGUUUUGGGUCAUAACAUA	253	UCUUACGGCCCUAAAACUUUUCA	18,215	N	Contig1400:30116-30022	0.96
mir-9b	UCUUUGGUUGUUUAGCUAUAUGA	1174	AUAUACUGCAUAAUCAAAGUUG	134	Y	Contig1910:62757-62855	0.96
mir-2149	CGCCAAUGACAAACUGUUGAU	212	UCAGCAGUUGUCCAUUGACGUU	440	N	Contig997:5619-5728	0.96
mir-2148	AGAAAUUCUGUGCGUUGAU	1093	UACUGCCCACAAAAUUUCUUU	59	N	Contig12516:7003-6916	0.95
mir-2177	AAGAAAUUCAUUGGUUAAUUUC	56	AAAUGCCCAGUGAAUUUCUCCGGU	15	N	Contig12516:6882-6810	0.95
mir-2161	UAUCUUUUUAUUUACACACAAA	3	UGUGCGUAGUUAAUAGAGGUAUC	363	N	Contig33:107536-107642	0.93
mir-2169	AACUUUGAAAUUCGAUCUGAU	588	UAAGCUCGAUAUUUAAAGUGGU	121	N	Contig573:149093-149189	0.93
mir-133a	AGCUGUUGACGAAGUAUCAAAA	400	UUGGUCCCCGUCAACCAGCUGU	53	Y	Contig7719:22251-22358	0.93
mir-2162	UGUGCCUUAUUUUGUAUAUUUGCA	51	UAUUAUGCAAAUAUUCACAAU	66	N	Contig946:66516-66403	0.91
mir-96a	UUUGGCACUUAAGGAAUUUGUCAC	45,136	GACAAUACUAAAGAGUCAAGCG	136	Y	Contig10175:11154-11052	0.89
mir-124e	CCAUUCUCAGUUGGAGUCUU	6	UAAGGCACGCUGUGAAUGCCA	212	Y	Contig2459:610-520	0.88

(continued)

TABLE 1. *Continued*

miRNA	5p Sequence	5p Reads	3p Sequence	3p Reads	Orthologs	Genomic location	MIResque score
mir-2157	UGAGUAUUGCAUCAAGAACCGA	61,280	AGUUUUGAUGAAUUACUUAA	16	N	Contig23847:7429-7337	0.86
mir-754c-1	GUUGCUUGCGGUAUUACUAU	309	UUUAUCUUCAAGCAAUAUCGG	13	N	Contig329:118407-118328	0.86
mir-124d	ACAUUUACAAGCGAGCCUUAAU	13	UAAGGCACCGCGGUAAGUGGGU	226	Y	Contig791:74276-74186	0.83
mir-31b-3	AGGCAAGAUGCUGGCAUAGCUGA	1859	AGCUAUUCAGGAUCUUUCCAU	14	Y	Contig27737:5583-5673	0.82
mir-31b-2	AGGCAAGAUGCUGGCAUAGCUGA	1826	AGCUAUUCAGGAUCUUUCCAU	24	Y	Contig27737:4905-4998	0.8
mir-2158	UGAGUAUUUUUUUCUUAUACCGA	6	CGGUAACGAAGAAUACUUGGA	587	N	Contig1616:21960-21880	0.8
mir-2160-2	UGGCCGCUUAUGUGUAGAACCGG	1659	GGUUUUUCAUGUAAGAGUCC	1	N	Contig31371:1012-1114	0.76
mir-2165	CAUUCUACAGGGAUCGAAUCAGU	268	UAAUUCGAUUCUUUAGAGUAUU	687	N	Contig15254:515-634	0.74
mir-2154-1	UCAGCUGUACAAUUGGCAUUGUGU	6246	CAAGUCUUUGUAUAGCUUAU	222	N	Contig21503:11132-11233	0.74
mir-124c-2	GCAUUAACCCUGUUGUCUUUAGAU	615	UAAGGCACCGCGGUGAAUGCCA	1810	Y	Contig8879:8241-8334	0.73
mir-2154-2	UCAGCUGUACAAUUGGCAUUGUGU	6246	CAAGUCUUUGUAUAGCUUAU	222	N	Contig21857:11141-11242	0.69
mir-7d	UGGAAGACUAGUAAUUUUGUUGU	3955	AACAUUAUUUCUGUUUUCCAA	5	Y	Contig17271:6048-6167	0.67
mir-754b-1	GUUGCUUGCGGGUAUUACUA	5817	AUUUUAUCAUCAAGCAAUAU	6	N	Contig23853:3949-4032	0.67
mir-754c-2	GUUGCUUGCGGGUAUUACUAU	350	UAGUUUUAUCUUCAUGCAAUAUCA	4	N	Contig5823:26722-26803	0.65
mir-2203	ACGAAGGCUUAAUGAUUAUCAU	3	AUCAAGAUCGUUAGUAAGUU	60	N	Contig43306:3411-3483	0.64
mir-754b-3	GUUGCUUGCGGGUAUUACUA	5821	AUUUUAUCAUCAAGCAAUAU	1	N	Contig29676:7219-7124	0.62
mir-753g	CAGGGCCUUGGAUUGUGAUCUC	1011	AGACAUUAUUCAAGCUCUCU	331	N	Contig12996:14591-14529	0.56
mir-2201	AGUUACUGAAGAAGCUGAUGCUG	8	UAUCUGAAGAGCUGGAGCAUC	687	N	Contig3745:19515-19340	0.53
mir-8b	CUUCUUACUUGAUGGUAAGAGA	3	UAAUACUGUCAGGUAAGAAUACU	3461	Y	Contig2498:385-476	0.5
mir-2202	UGCGCAAAAGAAUAUUAAUCGA	101	GAUUAAUUUCCUUUUACGCUUA	3	N	Contig28205:2670-2579 Contig4044:8963-9054	0.49
mir-2171	GUCAUCUAUUUUAGUCGUGGCU	57	GUAAAAUUAAGAUAGAUGCAGU	3	N	Contig1252:68864-68755	0.46
mir-315	UUUUGAUUGUUGCUCUGAGAGUU	679	CUGUCUAACAACCAUCAAAUU	2	Y	Contig20390:2443-2554	0.45
mir-96b	UUUGGCACUUAAGGAAUCGUCAC	33,734	ACAUUUCUAAGUGUCUAUUAU	1	Y	Contig4508:35049-34970	0.45
mir-2205	UGGAUAGUGAGUAAAGCGGUGA	31	UCUGUAUUUCUCAUCUUCCCA	4	N	Contig5365:29004-28919	0.44
mir-2160-1	UGGCCGCUUAUGUGUAGAACCGG	1693	UUUUCAUGUAAGAGCCAAA	3	N	Contig12997:19612-19717	0.41
mir-2147d	CGGAAAACUCGCACCGUGAUAU	7	UUUCACUGCGACUGUUCCCAU	2	N	Contig3339:35247-35345	0.41
mir-2156b	AAGGUUUUUGAGCCGUAAAUCA	9	CCUUACGGCCCCUAAACCUUAUU	1158	N	Contig877:48519-48609	0.36
mir-2f	AUCUGAAGCUGGAUGUGUUAU	88	UAUCACAGCCAUGCUACAGAGUUA	10	Y	Contig272:56588-56514	0.33
mir-2e	AAACCUAGUUGGUUGCCGAUAUG	8	UAUCACAGCCUUGCUCUAGGUUUA	6	Y	Contig272:56477-56381	0.33
mir-2204	UGAAAUUAGGACGACGAAAUUACAC	44	AAACAUCCGGACUAAGUUUAU	82	N	Contig5212:12435-12343	0.33
mir-2200	UGAAGACGAUUAUUAUUGAUUAUU	82	AAAAAAUCCGAAUUCAGCUGU	2	N	Contig1580:76060-76147	0.25
mir-2179	UGCAAAAAUAUUUAUAUUUAACUCA	2	AGACUAUAGAGAUUUUUGCUAU	10,123	N	Contig16538:4132-4042	0.2

middle of the mature strand, which is inconsistent with Dicer processing. Finally, *let-7a* was not scored because the pre-miRNA sequence is not present in the draft assembly of the *S. mediterranea* genome even though several overlapping sequence reads containing the *let-7a* pre-miRNA generated during sequencing of the genome are present in the NCBI trace archives. Nonetheless, as we obtained 352, 54,437, and 654,138 reads for *mir-754a*, *mir-1c*, and *let-7a*, respectively, we believe that these are authentic miRNAs.

MIResque identified 200 hairpins that scored above the miRNA threshold and had a total of at least 50 reads that mapped to the pre-miRNA with up to two mismatches. As MIResque was trained on human miRNAs, not planarian miRNAs, we manually analyzed the predicted miRNA hairpins to assess whether they were actual miRNAs. This analysis eliminated 77 hairpins that clearly did not encode miRNAs, but received high scores due to the number of mapped reads and secondary structure. The discarded hairpins typically corresponded to highly expressed piRNA clusters that fortuitously resembled miRNA hairpins and for which 18–25 nt sequences were obtained due to partial degradation of piRNAs (data not shown). As 57 of the remaining 123 MIResque hairpins corresponded to previously known miRNAs, MIResque identified 66 new pre-miRNA hairpins corresponding to 72 loci in the *S. mediterranea* genome (Table 1). We have validated the expression of seven of the new MIResque predicted miRNAs by Northern blots (Figs. 2C, 3C). These results suggest that most, if not all of the miRNA encoding loci predicted by MIResque are in fact authentic. When combined with the 60 previously identified miRNAs, we have now identified a total of 126 distinct pre-miRNA hairpins corresponding to 138 loci in the *S. mediterranea* genome (Supplemental Table 1). Thirty-one of the new miRNA loci encode miRNAs that belong to miRNA families present in other species (e.g., *mir-124d*, *mir-36b*, etc.), while 41 of these loci encode miRNAs that have only been identified in *S. mediterranea*. Thus, *S. mediterranea* contains at least 48 miRNAs that are specifically found in planarians and 78 miRNAs that are members of evolutionarily conserved miRNA families.

We previously described 6 miRNA clusters that contain between two to four miRNAs (Palakodeti et al. 2006). Analysis of the genomic coordinates of the newly identified miRNAs revealed six additional miRNAs clusters—*mir-2148/mir-2177*, *mir-mir-31b-2/mir-31b-3*, *mir-747/mir-2158*, *mir-36a/mir-36c*, *mir-216/mir-2157*, and *mir-mir-2e/mir-2f*—where the miRNAs are located between 35 and 617 nt apart (Table 1). Thus, *S. mediterranea* contains at least 12 miRNA clusters encoding a total of 28 miRNAs.

Strain-specific miRNA expression

We next examined whether the expression of any miRNAs differed between the sexual and asexual strains. To do this, we compared the relative abundance of each miRNA between the sexual and asexual libraries. While most miRNAs were expressed at similar levels in each strain, we identified 21 miRNAs whose expression was significantly restricted to either the sexual or asexual strains (Fig. 3). Eleven of these miRNAs are expressed at least two- to threefold higher in sexual animals than in asexual animals (Fig. 3A), while the other 10 miRNAs are expressed at least threefold higher in asexual animals than in the sexual animals (Fig. 3B).

To validate these findings, we performed Northern blots using probes complementary to *mir-754b-1* and *mir-2160-1* which our sequence data indicated were significantly enriched in sexual and asexual animals, respectively. Specifically, *mir-754b-1* was sequenced 3228 times in nonirradiated sexual animals, but only 23 times in nonirradiated asexual animals (Supplemental Table 2). In contrast, *mir-2160-1* was sequenced 740 times in nonirradiated asexual animals, but only 36 times in nonirradiated sexual animals (Supplemental Table 2). Consistent with our sequence data, the Northern blot experiments show that *mir-754b-1* and *mir-2160-1* are indeed expressed in a strain-specific manner, whereas *mir-71c* is expressed equally in all samples (Fig. 3C). Interestingly, the Northern blots also showed
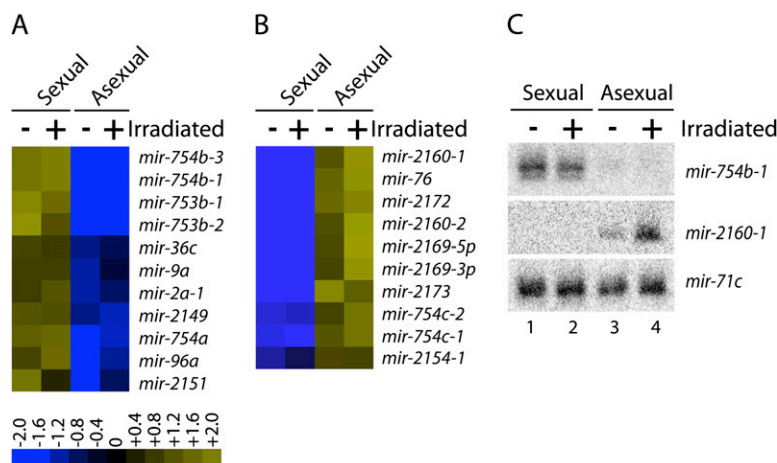

FIGURE 3. Strain-specific miRNA Expression. (*A*) MicroRNAs that are expressed at least twofold higher in sexual animals than in asexual animals were identified, and the expression ratios were log$_2$ transformed and depicted in heat map representation. (*B*) MicroRNAs that are expressed at least twofold higher in asexual animals than in sexual animals were identified, and the expression ratios were log$_2$ transformed and are depicted in heat map representation. (*C*) Northern blots for *mir-754b-1* and *mir-2160-1* were performed with total RNA isolated from intact (−) and irradiated (+) sexual and asexual animals. The microRNA *mir-71c* was used as a probe to control for equal RNA loading.

that *mir-2160-1* is expressed higher in irradiated animals than in nonirradiated animals (Fig. 3C, lanes 3,4), again consistent with our sequencing data. Thus, these data demonstrate that a subset of miRNAs are expressed in a strain-specific manner.

Interestingly, the majority of the strain-specific miRNAs are planarian-specific: only five of the 21 strain-specific miRNAs are members of miRNA families that exist in other organisms. Moreover, only one miRNA, *mir-76*, belonging to a known miRNA family is expressed specifically in the asexual strain. Thus, asexual animals express nine planarian-specific miRNAs at much higher levels than sexual animals. This raises the intriguing possibility that these miRNAs participate in either repressing development of the reproductive tissues or controlling reproductive fission, which only occurs in asexual animals. Conversely, it is likely that the miRNAs that are expressed specifically in the sexual animals enhance the development of testes and ovaries or repress fission.

Irradiation sensitive miRNAs

To identify miRNAs that may be expressed in neoblasts or neoblast progeny and potentially play a role in stem cell function we examined whether the expression of any miRNAs was reduced upon irradiation. This analysis identified 5 miRNAs—*let-7a*, *mir-71b*, *mir-756*, *mir-13*, and *mir-752*—that were reduced at least twofold upon irradiation (Fig. 4). For example, *let-7a* was sequenced 266,326 times from nonirradiated asexual animals, but only 92,535 times from irradiated asexual animals (Supplemental Table 2). Similarly, *mir-13* was sequenced 2554 and 1257 times from nonirradiated and irradiated asexual animals, respectively. Consistent with the sequence results, Northern blot analysis confirmed that the levels of *let-7a* and *mir-13* decrease upon irradiation, while the levels of a control miRNA, *bantam-a*, were similar between all samples (Fig. 4B). Intriguingly, at least one other member of the *let-7* family appears to be expressed in irradiation sensitive cells. Specifically, our sequence data indicated that *let-7b* levels decrease 1.5-fold upon irradiation (Supplemental Table 2) and Northern blot experiments confirm this observation (data not shown). These results demonstrate that the expression of these miRNAs is sensitive to irradiation in both sexual and asexual animals. Together, this suggests that these miRNAs are expressed in either neoblasts or their progeny and are strong candidates for miRNAs that are involved in some aspect of stem cell function. It is important to note that the sequencing and Northern blot experiments were all conducted 4 d after irradiation. Thus, it is quite likely that the levels of other miRNAs will decrease at later time points.

Interestingly, *mir-13*, *mir-71b*, and *mir-752* are contained in a cluster that also contains *mir-2d*. Within this cluster, *mir-2d* is located between *mir-752* and *mir-71b* (Fig. 4C).
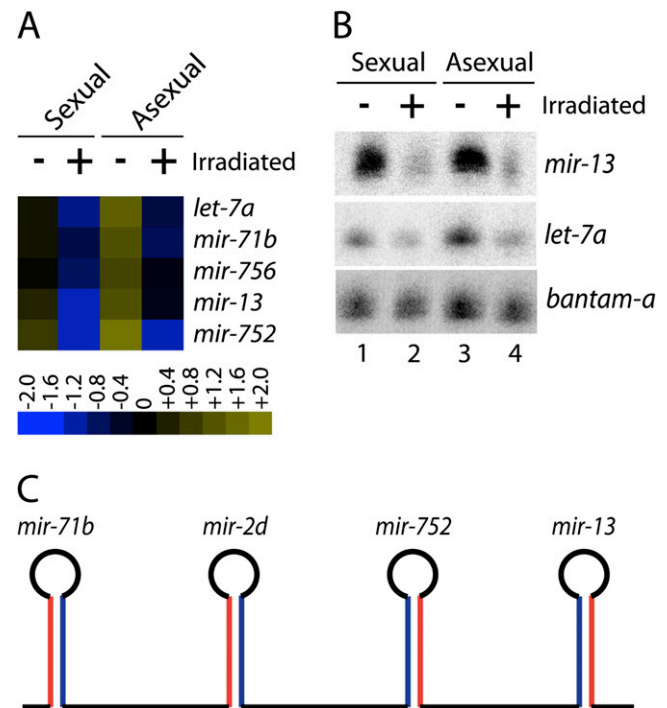

FIGURE 4. Irradiation-sensitive miRNA Expression. (*A*) MicroRNAs whose levels are reduced at least twofold after irradiation were identified and the expression ratios were log$_2$ transformed and depicted in heat map representation. (*B*) Northern blots for *mir-13* and *let-7a* were performed with total RNA isolated from intact (−) and irradiated (+) sexual and asexual animals. The microRNA *bantam-a* was used as a probe to control for equal RNA loading. (*C*) Schematic depiction of the organization of the *mir-71b/mir-2d/mir-752/mir-13* miRNA cluster (red, mature strand; blue * strand).

Surprisingly, while the expression of *mir-13*, *mir-71b*, and *mir-752* decreases at least twofold upon irradiation, the levels of *mir-2d* does not change (1.08-fold decrease). This suggests that the expression of the miRNAs in this cluster may be individually regulated at a post-transcriptional step.

The observation that *let-7a*, and to a lesser extent *let-7b*, is expressed in irradiation sensitive cells (most likely neoblasts) is intriguing, as *let-7* has been linked to cellular proliferation and differentiation in many organisms. In *Caenorhabditis elegans*, *let-7* is required for the terminal differentiation of seam cells—in *let-7* mutants, the seam cells fail to exit the cell cycle and instead continue to proliferate (Reinhart et al. 2000). Similarly, vertebrate *let-7* has been shown to be involved in the differentiation of embryonic and cancer stem cells. For example, *let-7* expression is low in both cancer and embryonic stem cells and becomes upregulated upon differentiation (Johnson et al. 2005; Lee et al. 2005; Yu et al. 2007). Finally, expression of the *let-7* family has been found to change significantly during regeneration in the newt (Tsonis et al. 2007). Specifically, *let-7* expression decreases during the initiation of lens and inner ear hair cell regeneration that involves dedifferentiation. Thus, in most organisms, *let-7 is* expressed in differentiated cells, not in stem cells. While

our data links *let-7a* expression to neoblast function, it is not yet clear how *let-7a* functions in planarians. One possibility is that rather than being involved in inducing differentiation, in planarians *let-7a* could be expressed in neoblasts where it functions to prevent differentiation. In this case, neoblast differentiation would require *let-7a* down-regulation. It is worth noting that we have not yet identified a miRNA that perfectly matches the canonical *let-7a* miRNA, which is 100% conserved in all other metazoans (Pasquinelli et al. 2000); *smed-let-7a* differs by 5 nt from the cannonical *let-7a* (Palakodeti et al. 2006). On the other hand, *let-7a* could function similarly in planarians and other organisms where it acts to induce differentiation. In this regard, we imagine that *let-7a* would either be transiently expressed in the neoblasts or the early neoblast progeny (Eisenhoffer et al. 2008) to initiate differentiation, but that its expression would decrease soon after differentiation began. This scenario would explain the absence of *let-7a* expression in both irradiated (Fig. 4) and *smedwi-2(RNAi)* (data not shown) animals. Nonetheless, our results strongly suggest that *let-7a* plays an important role in some aspect of neoblast function.

MATERIALS AND METHODS

Planarian cultures

The hermaphroditic sexual (Zayas et al. 2005) and asexual (Sanchez Alvarado et al. 2002) strains of *S. mediterranea* used for these experiments were provided by P. Newmark (University of Illinois) and A. Sánchez-Alvarado (University of Utah). The animals were maintained essentially as described (Cebria and Newmark 2005) at 22°C in ddH$_2$O supplemented with 1.6 mM NaCl, 1.0 mM CaCl$_2$, 1.0 mM MgSO$_4$, 0.1 mM MgCl$_2$, 0.1 mM KCl, and 1.2 mM NaHCO$_3$, and fed homogenized organic beef liver. All animals were starved 1 wk prior to experiments.

Small RNA cloning and sequencing

Libraries for the Illumina small RNA sequencing were constructed using the Illumina Small RNA Sample Prep Kit, essentially as described by the manufacturer. Briefly, small RNAs of 18–28 nt in length were gel-purified from 10 μg of total RNA isolated from sexual and asexual animals that were either untreated or irradiated. Eluted RNAs were successively ligated to 5′ and 3′ adaptors with RNA ligase and gel purified after each ligation step. RNAs were reverse transcribed and the resultant cDNA amplified by PCR for 16 cycles. PCR products were purified on a nondenaturing acrylamide gel and sequenced for 36 cycles on the Illumina Genome Analyzer II.

Small RNA genomic analysis

Sequences from the Illumina Genome Analyzer image files were extracted using Firecrest and Bustard. Sequences were trimmed for adaptor sequences and 18–25 base sequences were mapped to the supercontigs of the draft *S. mediterranea* genome assembly

(version 3.1) with Bowtie (Langmead et al. 2009). For the MIResque analysis, the mapping was performed to identify reads that map to the genome with up to two mismatches. To compare the expression levels of all miRNAs, reads were only considered that mapped perfectly to the pre-miRNA sequences. MIResque is implemented as a combination of Perl scripts and Matlab code for assigning scores to candidate miRNA hairpins (GW Yeo and S Aigner, unpubl.).

γ-Irridiation and FACS analysis

Animals were γ-irradiated with a cesium source essentially as previously described (Palakodeti et al. 2008) except that the animals were exposed to one 30-Gy dose and the animals analyzed 4 d later. FACS analysis of the neoblast populations were carried out as previously described (Palakodeti et al. 2008) based on earlier protocols (Reddien et al. 2005; Hayashi et al. 2006). Planarians were cut into small pieces on ice in cold calcium, magnesium-free medium (CMF) (15 mM HEPES, 400 mg/L NaH$_2$PO$_4$, 800 mg/L NaCl, 1200 mg/L KCl, 800 mg/L NaHCO$_3$, 240 mg/L glucose [pH 7.3]) and washed twice with CMF. The fragments were soaked in CMF containing 0.25% (w/v) Trypsin (GIBCO) and 1% BSA (Fisher Scientific) at 20°C for ∼60 min. The fragments were then completely dissociated into single cells by gentle pipetting. The cell mixture was sequentially filtered through 40 μm and 20 μm nylon filters (Millipore). Cells were collected by centrifugation, resuspended in fresh CMF supplemented with 1% BSA, 0.5 μg/mL Calcein AM (Sigma), and 10 μg/mL Hoescht 3342 (Sigma) and incubated at 20°C for 2 h. Finally, 2 μg/mL propidium iodide (Sigma) was added and the samples were kept on ice for analysis with a LSR II flow cytometer (Becton-Dickinson).

Northern blotting

Twenty miocrgrams of total RNA were resolved on a denaturing polyacrylamide gel, transferred to nitrocellulose, and hybridized as described previously (Palakodeti et al. 2006). Three 10-min post-hybridization washes were performed at room temperature in 2× SSC and 0.1% SDS. The membranes were then visualized with a Cyclone PhosphorImager (Perkin-Elmer).

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

REFERENCES

Baguñà J, Saló E, Auladell C. 1989. Regeneration and pattern formation in planarians III. Evidence that neoplasts are totipotent stem cells and the source of blastema cells. *Development* **107:** 77–86.

Baguñà J, Carranza S, Pala M, Ribera C, Giribet G, Arnedo MA, Ribas M, Riutort M. 1999. From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario Benazzi. *Ital J Zool (Modena)* **66:** 207–214.

Cebria F, Newmark P. 2005. Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development* **132:** 3691–3703.

Eisenhoffer G, Kang H, Alvarado A. 2008. Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* **3:** 327–339.

Eulalio A, Huntzinger E, Izaurralde E. 2008. Getting to the root of miRNA-mediated gene silencing. *Cell* **132:** 9–14.

Hayashi T, Asami M, Higuchi S, Shibata N, Agata K. 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev Growth Differ* **48:** 371–380.

Johnson S, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert K, Brown D, Slack F. 2005. RAS is regulated by the let-7 microRNA family. *Cell* **120:** 635–647.

Kim VN. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6:** 376–385.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lee Y, Kim H, Chung S, Kim K, Dutta A. 2005. Depletion of human micro-RNA miR-125b reveals that it is critical for the proliferation of differentiated cells but not for the down-regulation of putative targets during differentiation. *J Biol Chem* **280:** 16635–16641.

Newmark P, Sanchez Alvarado A. 2002. Not your father's planarian: A classic model enters the era of functional genomics. *Nat Rev Genet* **3:** 210–219.

Palakodeti D, Smielewska M, Graveley B. 2006. MicroRNAs from the planarian *Schmidtea mediterranea*: A model system for stem cell biology. *RNA* **12:** 1640–1649.

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

Pasquinelli A, Reinhart B, Slack F, Martindale M, Kuroda M, Maller B, Hayward D, Ball E, Degnan B, Muller P, et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408:** 86–89.

Reddien P, Oviedo N, Jennings J, Jenkin J, Sanchez Alvarado A. 2005. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* **310:** 1327–1330.

Reinhart B, Slack F, Basson M, Pasquinelli A, Bettinger J, Rougvie A, Horvitz H, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403:** 901–906.

Sanchez Alvarado A. 2006. Planarian regeneration: Its end is its beginning. *Cell* **124:** 241–245.

Sanchez Alvarado A, Newmark P, Robb S, Juste R. 2002. The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells, and regeneration. *Development* **129:** 5659–5665.

Tsonis P, Call M, Grogg M, Sartor M, Taylor R, Forge A, Fyffe R, Goldenberg R, Cowper-Sallari R, Tomlinson C. 2007. MicroRNAs and regeneration: *Let-7* members as potential regulators of dedifferentiation in lens and inner ear hair cell regeneration of the adult newt. *Biochem Biophys Res Commun* **362:** 940–945.

Wang Y, Zayas R, Guo T, Newmark P. 2007. Nanos function is essential for development and regeneration of planarian germ cells. *Proc Natl Acad Sci* **104:** 5901–5906.

Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, et al. 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131:** 1109–1123.

Zayas R, Hernandez A, Habermann B, Wang Y, Stary J, Newmark P. 2005. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: Analysis of ESTs from the hermaphroditic strain. *Proc Natl Acad Sci* **102:** 18491–18496.