



Phylogenetic Analysis of the Endoribonuclease Dicer Family

Zeqian Gao¹, Miao Wang¹, David Blair², Yadong Zheng^{1*}, Yongxi Dou^{1*}

1 State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, CAAS, Lanzhou, Gansu, China, **2** School of Marine and Tropical Biology, James Cook University, Townsville, Qld, Australia

Abstract

Dicers are proteins of the ribonuclease III family with the ability to process dsRNA, involved in regulation of gene expression at the post-transcriptional level. Dicers are conserved from basal metazoans to higher metazoans and contain a number of functional domains that interact with dsRNA. The completed genome sequences of over 34 invertebrate species allowed us to systematically investigate Dicer genes over a diverse range of phyla. The majority of invertebrate Dicers clearly fell into the Dicer1 or Dicer2 subfamilies. Most nematodes possessed only one Dicer gene, a member of the Dicer1 subfamily, whereas two Dicer genes (Dicer1 and Dicer2) were present in all platyhelminths surveyed. Analysis of the key domains showed that a 5' pocket was conserved across members of the Dicer1 subfamily, with the exception of the nematode *Bursaphelenchus xylophilus*. Interestingly, *Nematostella vectensis* DicerB grouped into Dicer2 subfamily harbored a 5' pocket, which is commonly present in Dicer1. Similarly, the 3' pocket was also found to be conserved in all Dicer proteins with the exceptions of *Schmidtea mediterranea* Dicer2 and *Trichoplax adherens* Dicer A. The loss of catalytic residues in the RNase III domain was noted in platyhelminths and cnidarians, and the 'ball' and 'socket' junction between two RNase III domains in platyhelminth Dicers was different from the canonical junction, suggesting the possibility of different conformations. The present data suggest that Dicers might have duplicated and diversified independently, and have evolved for various functions in invertebrates.

Citation: Gao Z, Wang M, Blair D, Zheng Y, Dou Y (2014) Phylogenetic Analysis of the Endoribonuclease Dicer Family. PLoS ONE 9(4): e95350. doi:10.1371/journal.pone.0095350

Editor: John J. Rossi, Beckman Research Institute of the City of Hope, United States of America

Received: February 15, 2014; **Accepted:** March 25, 2014; **Published:** April 18, 2014

Copyright: © 2014 Gao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was financially sponsored by Gansu Natural Science Foundation (0710RJZA081), State Public-interest Institution Basal Research Fund and by Special Fund for Agro-scientific Research in the Public Interest (201103008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: douyongxi@caas.cn (YD); zhengyadong@caas.cn (YZ)

Introduction

Small regulatory RNA pathways are highly conserved mechanisms present in most eukaryotic organisms and play an important role in post-transcriptional gene regulation. The gene-regulatory function of microRNAs (miRNAs) and short interfering RNAs (siRNAs) is mainly through translational repression or degradation of cytoplasmic mRNAs by an RNA-induced silencing complex (RISC). miRNA and siRNA pathways share a common RNase III processing enzyme, Dicer, and together with other proteins it constitutes RISC for gene transcriptional repression [1]. Dicer is responsible for recognizing a hairpin (in pre-miRNA) or long double-strand RNA (dsRNA), and processing them into miRNA-miRNA* duplexes or siRNA duplexes [2]. These small RNA duplexes are converted to a single-stranded form and bound to Argonaute (AGO), a key component of RISC, through a process coordinated by Dicer and other RNA-binding proteins [3]. Then small RNAs target specific mRNA sequences, leading to cleavage or translational repression of these [4].

Dicer proteins are present in many eukaryotic organisms including plants, fungi, and metazoans [5,6]. Vertebrates and nematodes have only one Dicer gene (Dicer1), whereas insects and flatworms possess two, (Dicer1 and Dicer2). Dicers normally contain a number of functional domains: an N-terminal DEAD box, an RNA helicase domain, a Dicer dimer domain, a Piwi-

Argonaute-Zwille (PAZ) domain, two RNase III domains and a dsRNA binding domain [7,8]. The crystal structure of Dicer from *Giardia intestinalis* revealed that the PAZ domain was responsible for binding of the 3' terminus of dsRNA [9]. After the 3' end bound to the PAZ domain, pre-miRNAs or dsRNAs are cleaved by the two RNase III domains which form a single dsRNA processing center through intramolecular dimerization [10]. In Dicer1, binding of the PAZ domain to the 3' terminus of pre-miRNA is crucial for orienting the RNase III domains for cleavage, however, recent publications have revealed that 5' terminus recognition of pre-miRNAs is also important for mature miRNAs synthesis [11,12].

Previous studies have focused on Dicers of plants and model organisms, little is known about Dicers of invertebrates. The recent availability of genome sequences of over 34 invertebrate species from 10 phyla, including 1 choanoflagellate, 2 cnidarians, 1 placozoan, 2 annelids, 1 mollusc, 7 platyhelminths, 7 nematodes, 10 arthropods, 1 echinoderm and 3 chordates, have allowed us to perform an extensive phylogenetic analysis of Dicers.

Materials and Methods

Acquisition of sequence

For some well-annotated genomes, Dicer sequences were directly retrieved from the databases. In addition, BLASTP and TBLASTN

Table 1. Distribution of endoribonuclease Dicer genes in invertebrates.

Species	Species Abbreviation	Gene Symbol	Accession Number	Length ^a	Data origin ^b
Choanoflagellata					
<i>Monosiga brevicollis</i>	Mbrev	/	/	/	/
Cnidaria					
<i>Nematostella vectensis</i>	NvecA	Sca_186	ABZ10549 ^c	1099	NCBI JGI
	NvecB	167046039	ABZ10551 ^c	1469	
<i>Hydra magnipapillata</i>	HmagA	HYDRA_212274	Hma2.212274	835	Metazome
	HmagB	HYDRA_205202	Hma2.205202	769	
	HmagC	HYDRA_222700	Hma2.222700	565	
Placozoa					
<i>Trichoplax adhaerens</i>	TadhA	TRIADDRAFT_18215	XP_002108754	1018	NCBI JGI
	TadhB	TRIADDRAFT_51673	XP_002107798	1418	
	TadhC	TRIADDRAFT_51985	XP_002107959	685	
	TadhD	TRIADDRAFT_52058	XP_002108012	915	
	TadhE	DCL-E	ABZ10547 ^c	707	
Annelida					
<i>Capitella teleta</i>	Ctel	CAPTEDRAFT_223153	ELU12939	1651	NCBI
<i>Helobdella robusta</i>	Hrob	estExt_Genewise1.C_710069	103772	1316	JGI
Mollusca					
<i>Lottia gigantea</i>	Lgig	gw1.12.81.1	61365	1800	JGI
Platyhelminthes					
<i>Schmidtea mediterranea</i>	Smed1	Dicer-1	ASA.00018.01	1328	SmedGD
	Smed2	Dicer-2	mk4.000125.07.01	1589	
<i>Schistosoma mansoni</i>	Sman1	Dicer-1	Smp_169750.1	2319	Sanger
	Sman2	Dicer-2	Smp_033600	928	
<i>Schistosoma japonicum</i>	Sjap1	Dicer-1	Sjp_0069770	2480	Sanger
	Sjap2	Dicer-2	Sjp_0043700	923	
<i>Echinococcus granulosus</i>	Egra1	Dicer-1	EgrG_000085200	1904	Sanger
	Egra2	Dicer-2	EgrG_000181800	1924	
<i>Echinococcus multilocularis</i>	Emul1	Dicer-1	EmuJ_000085200	1906	Sanger
	Emul2	RNC3.1	EmuJ_000180900	1812	
	Emul2	RNC3.2	EmuJ_000181800	1929	
<i>Hymenolepis microstoma</i>	Hmic1	Dicer-1	HmN_000252400	1897	Sanger
	Hmic2	Endoribonuclease Dicer	HmN_000200100	1183	
<i>Taenia solium</i>	Tsol1	Dicer-1	TsM_000872800	1890	Sanger
	Tsol2	Endoribonuclease Dicer	TsM_000756400	829	
Nematoda					
<i>Caenorhabditis elegans</i>	Cele	Dicer-1	NP_498761	1910	NCBI
<i>Bursaphelenchus xylophilus</i>	Bxyl	Dicer-1	BUX_s00116.153	1040	Sanger
<i>Pristionchus pacificus</i>	Ppac	Dicer-1	WBGene00096444	2769	WormBase
<i>Strongyloides ratti</i>	Srat	Dicer-1	g5271	1824	WormBase
<i>Brugia malayi</i>	Bmal	Dicer-1/Bm5026	WBGene00225287	1930	WormBase
<i>Trichinella spiralis</i>	TspiA	Tsp_01223	XP_003377020	2029	NCBI
	TspiB	Tsp_00125	XP_003375890	1903	NCBI
<i>Loa loa filariasis</i>	Lloa	LOAG_02227	XP_003137813	1928	NCBI
Arthropoda					
<i>Daphnia pulex</i>	DpulA	DAPPUDRAFT_308316	EFX72380	1979	NCBI JGI
	DpulB	DAPPUDRAFT_329028	EFX69538	1459	
	DpulC	DAPPUDRAFT_309030	EFX86072	1604	
<i>Pediculus humanus corporis</i>	Phum	Dcier-1	XP_002429494	2179	NCBI

Table 1. Cont.

Species	Species Abbreviation	Gene Symbol	Accession Number	Length ^a	Data origin ^b
<i>Tribolium castaneum</i>	Tcas1	Dicer-1	XP_968993	1865	NCBI
	Tcas2	Dicer-2	NP_001107840	1623	
<i>Nasonia vitripennis</i>	Nvit1	Dicer-1	XP_001605287	1917	NCBI
	Nvit2	Dicer-2	XP_001602524	1450	
<i>Acyrtosiphon pisum</i>	ApisA	LOC100159500	XP_001943370	1626	NCBI
	ApisB	LOC100166428	XP_001945890	1691	
<i>Drosophila melanogaster</i>	Dmel1	Dicer-1	NP_524453	2249	NCBI FlyBase
	Dmel2	Dicer-2	NP_523778	1772	
<i>Anopheles gambiae</i>	Agam1	AgaP_AGAP002836	XP_003436256	2336	NCBI
	Agam2	AgaP_AGAP012289	XP_320248	1672	
<i>Aedes aegypti</i>	Aaeg1	AaeL_AAEL006794	XP_001652212	1658	NCBI
	Aaeg2	AaeL_AAEL001612	XP_001659747	2193	
<i>Culex pipiens quinquefasciatus</i>	Cpip1	CpipJ_CPIJ003169	XP_001844757	2270	NCBI
	Cpip2	CpipJ_CPIJ010534	XP_001855187	1165	
Echinodermata					
<i>Strongylocentrotus purpuratus</i>	Spur	LOC586001	XP_790894	1850	Metazome
Chordata					
<i>Branchiostoma floridae</i>	Bflo	BRAFLDRAFT_202604	XP_002610617	1868	NCBI
<i>Ciona intestinalis</i>	Ciona	Endoribonuclease Dicer	ENSCINP00000017117	1872	Ensemble
<i>Saccoglossus kowalevskii</i>	Skow	Sakowv30031161m.g	Sakowv30031161m	1905	Metazome

^aAmino acid length;

^bJGI: Joint Genome Institute; NCBI: National Center for Biotechnology Information; SmedGD: *Schmidtea mediterranea* Genome Database; Sanger: Wellcome Trust Sanger Institute; FlyBase: *Drosophila* database; SilkDB: silkworm database; Ensemble: Ensemble Genome Browser.

doi:10.1371/journal.pone.0095350.t001

were performed to search against their databases using *Drosophila melanogaster* Dicers (NP_524453 and NP_523778), *Caenorhabditis elegans* Dicer (NP_498761) or *Schistosoma mansoni* Dicers (Smp_169750.1 and Smp_033600) as query sequences. An E-value of 1×10^{-10} was used as a cutoff in BLAST searches and the hits were filtered to keep only those with at least 25% identity to the query sequence. Protein functional domains were identified using Pfam database and SMART database [13,14]. The species names, abbreviations and accession numbers are provided in Table 1.

Sequence alignment and phylogenetic analysis

The data sets contained a total of 58 sequences from 34 species (in a size from 565aa to 2769aa, Text S1). The amino acid sequences of Dicer were aligned by MUSCLE [15] with default parameters and manually optimized by Jalview 2.8 [16]. The alignments were subsequently processed using Gblocks v0.91b [17] for phylogenetic reconstruction, allowing gaps in 1/2 of the sequences. ProtTest 3.2 was applied to find an appropriate model of amino acid substitution for tree building analysis [18]. A maximum likelihood tree was constructed using PhyML 3.0 program [19]. Clade support was calculated using SH-like approximate likelihood ratio test, Bayes likelihood test and bootstrap proportions (500 replicates).

Results

Identification and distribution of Dicer genes across invertebrates

The final data sets contained 58 Dicer gene sequences from two cnidarians, one placozoa, two annelids, one mollusc, seven

platyhelminths, seven nematodes, eleven arthropods, one echinoderm and three chordates (Table 1). No Dicer homologues were identified in the choanoflagellate *Monosiga brevicollis*. Our results of genomic database searches revealed that one placozoan, two annelids, one mollusc, one echinoderm and three of the chordates investigated possessed only one Dicer1 gene. Each of nematodes had only one Dicer1 gene, except *Trichinella spiralis*, which expressed both Dicer1 and Dicer2 genes. Platyhelminths and arthropods possessed two Dicer genes in their genomes, with the exceptions of *Daphnia pulex* (three genes), *Pediculus humanus corporis* (one gene) and *Echinococcus multilocularis* (three genes).

Phylogenetic analysis of Dicers

As shown in the Maximum likelihood tree (Fig. 1), Dicers of invertebrates were grouped into two lineages: Dicer1 subfamily and Dicer2 subfamily. Almost all of the arthropods and platyhelminths surveyed possessed one member of each of these subfamilies, and annelids, molluscs, nematodes, echinoderms and chordates investigated had only one Dicer gene that belongs to Dicer1 subfamily. The placozoan *Trichoplax adhaerens* had the most copies of Dicer genes in our investigated species; however, all of the five Dicer genes were classed into the Dicer2 subfamily. The two cnidarians *N. vectensis* and *Hydra magnipapillata* each had only one Dicer2 gene, but possessed other Dicer genes that fell outside the two subfamilies.

Organization of functional domains of Dicer family

We identified the functional domains using the Pfam database and confirmed each inferred domain using the SMART database. As shown in Fig. 1, Dicers had significant variability in domain

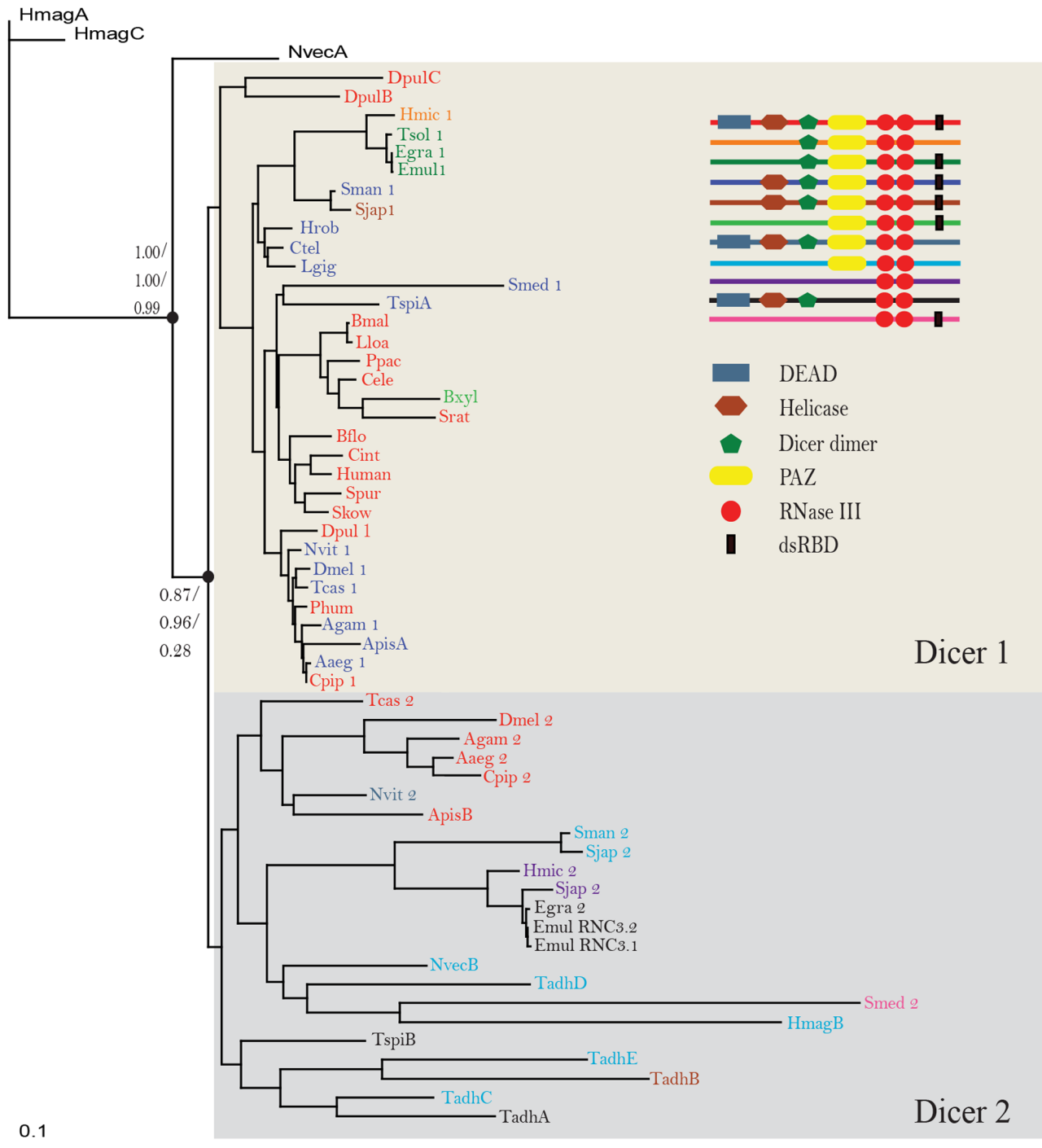


Figure 1. A maximum likelihood tree of invertebrate Dicers. The tree was constructed using maximum likelihood method. Two number sets, 1.00/1.00/0.99 and 0.87/0.96/0.28, at the nodes were SH-like approximate likelihood ratio, Bayes likelihood and bootstrap values, respectively. doi:10.1371/journal.pone.0095350.g001

organization. For instance, Dicers initially characterized in humans are multidomain proteins, consisting of an N-terminal DEAD box, an RNA helicase domain, a Dicer dimer domain, a PAZ domain, two RNase III domains and a dsRNA binding domain [10]. However, *Taenia solium* Dicer2 processed only one RNase III domain. We also observed the loss of the DEAD domain, which contains two RecA-like domains as a catalytic core

and can regulate various processes involving RNA [20], in Dicer1 of mollusks, annelids, platyhelminths and most arthropods.

A PAZ domain is an RNA-binding module found in PPD proteins (PAZ and Piwi domain proteins) and Dicer orthologs, and anchors the 2-nucleotide 3' overhang of dsRNA with its highly conserved binding pocket [10]. After searching annotated domains using Pfam and SMART databases, we did not find the PAZ domain in Dicer2 of the platyhelminths *Schmidtea mediterranea*,

Previous study suggested that the ‘ball’ consisted of the hydrophobic side chain of amino acids, and the interaction between ‘ball’ and ‘socket’ was precluded by charged side chain or the absence of any side chain in the position of the ‘ball’ [9]. Interestingly, we found that the ‘ball’ within the RNase III a domain. In platyhelminths Dicer1 proteins contained a hydrophilic side chain of threonine, but the residues of the hydrophobic ‘socket’ were conserved across different species. This substitution may result in reduced affinity between the ‘ball’ and the hydrophobic ‘socket’ and altered conformation.

References

- Lamontagne B, Larose S, Boulanger J, Elela SA (2001) The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Curr Issues Mol Biol* 3: 71–78.
- He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5: 522–531.
- Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293: 1146–1150.
- Wang Y, Juraneck S, Li H, Sheng G, Wardle GS, et al. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461: 754–761.
- Cerutti H, Casas-Mollano JA (2006) On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* 50: 81–99.
- Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23: 578–587.
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409: 363–366.
- Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, et al. (2006) The evolution and diversification of Dicers in plants. *FEBS Lett* 580: 2442–2450.
- Blaszczyk J, Tropea JE, Bubunencko M, Routzahn KM, Waugh DS, et al. (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* 9: 1225–1236.
- Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell* 118: 57–68.
- MacRae IJ, Zhou K, Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 14: 934–940.
- Park JE, Heo I, Tian Y, Simanshu DK, Chang H, et al. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475: 201–205.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–301.
- Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40: D302–305.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552.
- Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321.
- Kurimoto K, Muto Y, Obayashi N, Terada T, Shirouzu M, et al. (2005) Crystal structure of the N-terminal RecA-like domain of a DEAD-box RNA helicase, the *Dugesia japonica* vasa-like gene B protein. *J Struct Biol* 150: 58–68.
- Ji X (2006) Structural basis for non-catalytic and catalytic activities of ribonuclease III. *Acta Crystallogr D Biol Crystallogr* 62: 933–940.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, et al. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455: 1193–1197.
- de Jong D, Eitel M, Jakob W, Osigus HJ, Hadrys H, et al. (2009) Multiple dicer genes in the early-diverging metazoa. *Molecular Biology and Evolution* 26: 1333–1340.
- Krishna S, Nair A, Cheedipudi S, Poduval D, Dhawan J, et al. (2013) Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in *Hydra magnipapillata*. *Nucleic Acids Research* 41: 599–616.
- Moran Y, Praher D, Fredman D, Technau U (2013) The Evolution of MicroRNA Pathway Protein Components in Cnidaria. *Molecular Biology and Evolution* 30: 2541–2552.
- Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125: 887–901.
- Ji XH (2008) The mechanism of RNase III action: How Dicer dices. *Rna Interference* 320: 99–116.

Supporting Information

Text S1 Amino acid sequences of Dicers.
(TXT)

Author Contributions

Conceived and designed the experiments: YD ZG YZ. Analyzed the data: ZG MW DB. Contributed reagents/materials/analysis tools: ZG. Wrote the paper: ZG YZ DB.