Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants

Krishanu Mukherjee,*[1] Henry Campos,[2] and Bryan Kolaczkowski[1,3]

[1]Department of Microbiology & Cell Science, University of Florida

[2]Interdisciplinary Center for Biotechnology Research, University of Florida

[3]Genetics Institute, University of Florida

***Corresponding author:** E-mail: Krishanu@ufl.edu.

Associate editor: Matthew Hahn

Abstract

RNA interference (RNAi) is a eukaryotic molecular system that serves two primary functions: 1) gene regulation and 2) protection against selfish elements such as viruses and transposable DNA. Although the biochemistry of RNAi has been detailed in model organisms, very little is known about the broad-scale patterns and forces that have shaped RNAi evolution. Here, we provide a comprehensive evolutionary analysis of the Dicer protein family, which carries out the initial RNA recognition and processing steps in the RNAi pathway. We show that Dicer genes duplicated and diversified independently in early animal and plant evolution, coincident with the origins of multicellularity. We identify a strong signature of long-term protein-coding adaptation that has continually reshaped the RNA-binding pocket of the plant Dicer responsible for antiviral immunity, suggesting an evolutionary arms race with viral factors. We also identify key changes in Dicer domain architecture and sequence leading to specialization in either gene-regulatory or protective functions in animal and plant paralogs. As a whole, these results reveal a dynamic picture in which the evolution of Dicer function has driven elaboration of parallel RNAi functional pathways in animals and plants.

Key words: RNAi, RNA interference, dicer, positive selection, antiviral dicer, PAZ domain.

Introduction

RNA interference (RNAi) is a widespread molecular process that plays two primary cellular roles: 1) regulating the activities of endogenous genes and 2) protecting the cell from selfish genetic material such as viruses and transposable elements. RNAi's gene-regulatory function is largely facilitated by degradation of cytoplasmic mRNAs by the RNA-induced silencing complex (RISC), a collection of proteins and template RNA that targets and degrades complementary RNA sequences (Zamore et al. 2000; Preall and Sontheimer 2005; Liu et al. 2006; Wang, Noland, et al. 2009). RNAi is additionally thought to affect gene regulation by altering chromatin structure (Hall et al. 2003; Volpe et al. 2003; Fukagawa et al. 2004). The protective role of RNAi appears to occur primarily through direct identification, processing and degradation of unwanted RNA molecules by the RISC (Galiana-Arnoux et al. 2006).

The RNAi process occurs through an interaction between RNA molecules and multiple proteins that collectively identify, process, and regulate RNA. Dicer is the primary RNA recognition and processing protein. Dicer anchors a double-stranded RNA (dsRNA) molecule and cuts it to produce short dsRNAs (Ketting et al. 2001). These short RNAs are converted to single-stranded form and bound to an Argonaute (AGO) protein through a process coordinated by Dicer and other RNA-binding proteins (Hammond et al. 2001). Once bound to AGO, the template RNA binds complementary RNA sequences, at which point they are cleaved or otherwise regulated by AGO and other components of the RISC (Wang, Juranek et al. 2009).

Because of its primary role in dsRNA recognition and processing, Dicer has received considerable research attention (Merritt et al. 2010; Rossbach 2010; Pecot et al. 2011; Davalos and Esteller 2012; Grimm 2012). The typical Dicer found in animals and plants is a large protein consisting of a series of functional domains that interact with RNA in different ways (fig. 1). The PAZ domain anchors the 3′-end of the dsRNA, after which it is cleaved by twin RNase3 domains (Bernstein et al. 2001). This "catalytic core" is sufficient for Dicer function in some organisms (Macrae et al. 2006). Some Dicer proteins additionally anchor the 5′ end of the dsRNA molecule using an N-terminal extension of the PAZ domain. This 5′ anchoring seems particularly important for micro-RNA (miRNA) recognition and processing, supporting Dicer's main gene-regulatory role (Park et al. 2011).

Dicer's DEAD/Helicase domain is thought to facilitate the movement of the protein along long dsRNA molecules (Welker et al. 2011), suggesting it may be important for supporting Dicer's protective role by allowing it to find the end of viral and transposable-element RNAs, which can then be assayed by the PAZ domain. Further support for this model

Article

Fig. 1. Dicer protein consists of multiple RNA-interacting functional domains. We plot the domain architecture of human Dicer along the protein sequence; other Dicer proteins from animals and plants have similar domain architectures (figs. 2 and 4). Functional domains were identified via sequence searches of the PFam database (Punta et al. 2011), the SMART database (Letunic et al. 2012) and the NCBI conserved-domain database (Marchler-Bauer et al. 2011). The 5′ RNA-binding pocket extension of the canonical PAZ domain found in some animal Dicers is also shown (Park et al. 2011).

comes from examination of *Drosophila melanogaster* Dicer1, which seems to have specialized in miRNA processing by losing its functional DEAD/Helicase domain (Welker et al. 2011). Other Dicer functional domains appear to coordinate the "hand-off" of processed RNAs to AGO, either through direct Dicer–RNA interaction or through interactions with other partner proteins (Maniataki and Mourelatos 2005; Koscianska et al. 2011).

Although the biochemical functions of Dicer have been detailed in model organisms, the evolution of the Dicer "superfamily" remains poorly characterized. Dicer is absent from bacteria and archaea but is found throughout eukaryotes, suggesting an early eukaryote origin (Cerutti and Casas-Mollano 2006; Shabalina and Koonin 2008). Current evidence suggests that the Dicer family diversified independently in animals, plants, and fungi (Cerutti and Casas-Mollano 2006) and was lost from many parasitic protozoa (Ullu et al. 2004; Baum et al. 2009) as well as model fungi lacking RNAi (Drinnenberg et al. 2009). However, the support in favor of this model is relatively weak, and alternative hypotheses have not been thoroughly evaluated.

Vertebrates and nematodes have only one Dicer gene, whereas insects have two (Hammond 2005), suggesting an insect-specific duplication followed by functional divergence into miRNA-based gene regulation and antiviral immunity (de Jong et al. 2009). This hypothesis is supported by evidence for strong positive selection affecting fly Dicer2—which performs an antiviral function (Obbard et al. 2006; Heger and Ponting 2007; Kolaczkowski et al. 2011)—and a parallel loss of DEAD/Helicase function in Dicer1, which appears to focus this protein's function on miRNA processing (Welker et al. 2011). All of this is consistent with a model of gene duplication followed by functional divergence in insects or arthropods. However, phylogenetic analysis—the real test of macro-evolutionary hypotheses (Huelsenbeck and Rannala 1997)—has so far failed to strongly support the insect-specific duplication hypothesis (de Jong et al. 2009).

Most model plant genomes encode four Dicer genes (DCLs 1–4), which—similar to the case in animals—appear to have diverged to function in miRNA-based gene regulation vs. antiviral immunity (Blevins et al. 2006; Bouche et al. 2006). However, there may be some functional overlap among plant Dicer paralogs, particularly in the case of antiviral Dicers, where one Dicer may compensate for loss of a paralog's function (Gasciolli et al. 2005). How plant Dicers functionally diverged is completely unknown, so it is impossible to evaluate whether there is any similarity with what we observe in animals.

Here, we examine the broad patterns of Dicer evolution using a combination of phylogenetic, structural-modeling and sequence-analysis approaches. We show that: 1) Dicer independently diversified in animal and plant lineages, coincident with the origins of multicellularity and requirements for complex gene regulation; 2) animal Dicer did not duplicate in insects but much earlier in metazoan evolution, with antiviral Dicer2 being subsequently lost from lineages developing alternative antiviral strategies; 3) the main plant antiviral Dicer (DCL-4) has been a repeated target of intense positive selection for changes in RNA recognition and/or binding, suggesting a long-term evolutionary arms race between this protein and viral molecules; and 4) although the biochemical capacity to recognize miRNAs appears ancestral, efficient miRNA recognition like that employed by humans arose later and possibly independently in animals and plants. These results provide a thorough picture of the forces and patterns shaping Dicer evolution and suggest that many common assumptions about the evolution of RNAi may warrant more careful investigation.

Results

Evolution of Eukaryote Dicers

The availability of complete genome sequences from a variety of eukaryotes provides an opportunity to systematically examine the origin and evolution of important biomolecular systems by studying the evolution of the genes encoding their component parts. To deepen our understanding of how the RNAi system evolved, we examined the evolution of the Dicer gene family, which carries out the initial RNA recognition and processing steps that initiate RNAi (Hamilton and Baulcombe 1999; Bernstein et al. 2001).

We used homology-based gene identification based on confirmed functional Dicers to find novel Dicer genes from a large number of fully sequenced genomes (supplementary Excel spreadsheet, Supplementary Material online). As expected, we found full-length Dicer genes throughout animals, plants, fungi, and many protozoan lineages, although Dicers were absent from yeast and some parasitic protozoa, consistent with previous results (Cerutti and Casas-Mollano 2006). The functional domain building-blocks of Dicer proteins are encoded in bacteria and archaea, but we found no full-length or multi-domain Dicer proteins in these lineages, suggesting that Dicer—and indeed the entire canonical RNAi machinery—is a eukaryote innovation that arose very early in this lineage's evolution.

We inferred the evolutionary history of eukaryote Dicers using maximum-likelihood and Bayesian techniques, finding

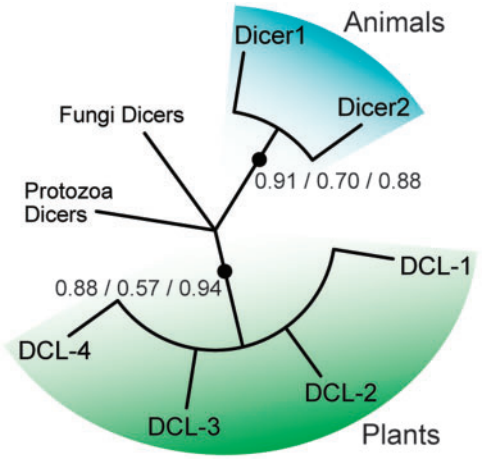

Fig. 2. Phylogenetic analysis supports independent expansions of an ancient eukaryote Dicer protein in animals and plants. We plot the support for monophyletic expansions of Dicer paralogs in animals and plants. Support is given as SH-like aLRT scores/maximum-likelihood bootstrap proportions/Bayesian posterior probabilities. See supplementary figures S1–S5, Supplementary Material online, for full trees and additional analyses.

strong support for independent expansions in animals, plants, and fungi (fig. 2 and supplementary fig. S1, Supplementary Material online). These independent expansions were consistently supported by multiple statistical-support measures and alignment strategies, suggesting our results are robust to alignment ambiguity and do not depend on any particular approach for inferring clade support. Support for key nodes increased when we used an elision strategy that combines information from multiple alignments (Wheeler et al. 1995), suggesting that the most reliable alignment positions consistently support a phylogeny in which Dicer genes independently expanded in the major multicellular eukaryote lineages.

Support for independent Dicer expansions increased when we removed the fast-evolving taxa most likely to contribute to phylogenetic error, suggesting our result is robust to taxon sampling and not strongly affected by long-branch attraction (supplementary fig. S2, Supplementary Material online). Our phylogeny was also robust to evolutionary model uncertainty. Alternative evolutionary models identified using model-selection procedures all produced the same maximum-likelihood phylogeny (supplementary table S1, Supplementary Material online), and Bayesian analyses that integrate over multiple evolutionary models recovered the same independent expansions of animal, plant, and fungi Dicers with high posterior probability (supplementary fig. S1, Supplementary Material online).

We also reconstructed the Dicer gene family tree using advanced heterogeneous evolutionary models that have been shown to prevent errors and improve phylogenetic accuracy under realistic conditions. First, we used a Bayesian technique that incorporates heterogeneity in amino acid frequencies across the protein sequence (Lartillot et al. 2009). This approach recovered the same independent Dicer expansions in animals, plants, and fungi, although statistical support was somewhat reduced, as expected with a more complex model (supplementary fig. S3, Supplementary Material online). Animal and plant expansions were also present in the maximum-likelihood phylogeny reconstructed using a model that incorporates across-site branch-length heterogeneity (Kolaczkowski and Thornton 2008), although fungi Dicers were not monophyletic in this tree (supplementary fig. S4, Supplementary Material online).

We were concerned that the high variability of the overall Dicer domain architecture across eukaryotes (supplementary Excel spreadsheet, Supplementary Material online) could lead to phylogenetic artifacts, so we additionally reconstructed the Dicer phylogeny using only the aligned "catalytic core," consisting of the PAZ and twin RNase3 domains conserved across all Dicers. This phylogeny also supported independent Dicer expansions in animals, plants, and fungi with high statistical confidence (supplementary fig. S5, Supplementary Material online).

Together these results strongly support an evolutionary history in which the Dicer gene family originated early in eukaryote evolution and independently expanded in animals, plants and fungi, coincident with the evolution of multicellularity—and requirement for more complex gene regulation—in these major eukaryote lineages. This result is robust to all the major factors that have been shown to affect phylogenetic accuracy, including the reconstruction strategy (maximum-likelihood vs. Bayesian), the method for assessing statistical confidence, alignment ambiguity, taxon sampling and evolutionary model ambiguity and inadequacy.

Our overall Dicer phylogeny provides some clues about the patterns of expansions in animals and plants (supplementary figs. S1–S5, Supplementary Material online). Specifically, we observed a main clade of animal Dicer1 homologs spanning arthropods, nematodes, and vertebrates as well as a smaller clade of Dicer2 homologs from arthropods, which fell toward the unresolved base of the animal tree. In plants, the main DCL types (DCLs 1–4) appear to have originated early, as monocots and dicots are represented in all four DCL clades, and DCLs 1 and 3 contain clear moss and *Selaginella* orthologs.

However, many of the fine-scale branching patterns remained unresolved in the full Dicer tree, limiting our ability to draw strong conclusions about the patterns of Dicer evolution within animals and plants. We hypothesized that aligning highly divergent Dicer homologs from across eukaryotes reduced the alignable data set size, which could contribute to lack of resolution in the full tree. Indeed, after processing alignments to remove ambiguously aligned regions, the combined sequence alignment was much shorter than individual alignments of only animal or plant sequences, and clade support was noticeably lower (supplementary table S2, Supplementary Material online). Because our full Dicer tree strongly supported independent expansions in animals and plants, we reconstructed separate animal and plant phylogenies to make better use of the available sequence data and generate a clearer picture of precisely when and how these independent expansions occurred.

Evolution of Animal Dicers

We reconstructed the evolutionary history of animal Dicers using maximum-likelihood and Bayesian methods (fig. 3 and supplementary fig. S6, Supplementary Material online). We found strong support for an ancient duplication, giving rise to Dicer1 and Dicer2 genes very early in metazoan evolution. This is in stark contrast to the widely held belief that the Dicer1/2 duplication occurred later in insects (de Jong et al. 2009).

As with our evolutionary analysis of eukaryote Dicers as a whole, we evaluated the robustness of our animal Dicer tree to all the major factors known to affect phylogenetic accuracy. We found that our results were robust to alignment ambiguity and use of different methods for assessing statistical confidence (supplementary fig. S6, Supplementary Material online). Support for an ancient Dicer duplication in animals remained after removing fast-evolving taxa (supplementary fig. S7, Supplementary Material online) and was robust to model uncertainty (supplementary table S1 and fig. S6, Supplementary Material online). Use of advanced heterogeneous evolutionary models less susceptible to phylogenetic errors also recovered this ancient Dicer1/2 duplication event (supplementary figs. S8 and S9, Supplementary Material online), as did reconstructing the animal Dicer phylogeny using only the aligned "catalytic core" of the protein (supplementary fig. S10, Supplementary Material online).

Although it is impossible to completely rule out phylogenetic error in any analysis, we were unable to detect a strong affect on our results from any of the major factors known to cause such errors, suggesting our results are indeed reliable. Furthermore, we found that non-insect animals—the prawn *Litopenaeus vannamei* and the planarians *Clonorchis sinensis* and *Schmidtea medeterranea*—had two Dicer genes, one of which grouped with insect Dicer1, and the other of which strongly grouped with insect Dicer2, indicating that the Dicer1/Dicer2 duplication occurred prior to insect divergence (fig. 3). Statistical topology tests also strongly rejected an arthropod- or insect-specific gene duplication as a possible explanation for our data, using full-length Dicer alignments ($P < 3.13 \times 10^{-3}$), alignments of only the DEAD/Helicase domains ($P < 2.65 \times 10^{-2}$) or alignments of the conserved Dicer "catalytic core" ($P < 1.01 \times 10^{-3}$, supplementary table S3, Supplementary Material online).

Synthesizing results from multiple phylogenetic analyses of various alignments, we identified Dicer2 orthologs from the basal metazoa, *Trichoplax* and *Nematostella*, suggesting that the Dicer1/2 duplication may have occurred very early in animal evolution. However, the relatively sparse taxon sampling and long-branch lengths at the base of the animal phylogeny left much of the basal branching pattern unresolved.

To attempt to better resolve this early branching pattern, we aligned additional animal Dicer RNase3 sequence data from a previous study (de Jong et al. 2009) to homologous regions from our sequence database and reconstructed the animal Dicer phylogeny from this expanded dataset. Although the same early Dicer1/2 duplication was recovered, the additional data did not add much taxonomic diversity, and the basal branching pattern remained weakly resolved (supplementary fig. S11, Supplementary Material online).

In addition to retrieving data from additional taxa, phylogenetic resolution can often be improved by including additional sequence data; unfortunately, this approach is not applicable to gene-tree reconstruction, as the length of the gene's sequence cannot be extended. However, it has long been observed that physically and functionally interacting proteins often have highly correlated patterns of gene duplication (Fryxell 1996; Pazos et al. 1997; Goh et al. 2000; Wu et al. 2011). Based on this observation, if the phylogenies of genes that interact in the same biochemical pathway all exhibit the same duplication pattern, this not only suggests that the entire pathway duplicated at the same time but also reinforces our confidence that the individual gene trees are robust, as phylogenetic artifacts would have to strongly affect all the gene trees in the same way to recover a consistent—but erroneous—phylogeny.

In the case of RNAi, Dicer's function is tightly linked to AGO proteins, with which Dicer directly interacts (van Rij et al. 2006; Marques et al. 2010). Furthermore, studies in model organisms have shown that Dicer1 interacts with AGO1, whereas Dicer2 primarily interacts with AGO2 (Liu et al. 2003; Okamura et al. 2004). When we reconstructed the AGO gene family tree by maximum likelihood, we found strong support for an AGO1/AGO2 gene duplication occurring at the same time in early animal evolution that we observed in our Dicer phylogeny (supplementary fig. S12, Supplementary Material online). We also observed a striking correlation between the taxonomic distributions of Dicer-AGO interacting partners, with Dicer2/AGO2 found in arthropods, prawn, planaria, and *Trichoplax*. Our AGO tree also recovered a strongly supported *Trichoplax* AGO2 along with a strongly supported *Nematostella* AGO1, further suggesting that this duplication event may have occurred very early in animal evolution. This result reinforces our confidence in an ancient Dicer1/2 duplication and suggests that—in addition to Dicer—many of the core proteins involved in RNAi may have duplicated around the same time early in animal evolution.

Under this early duplication scenario, both Dicer2 and AGO2 appear to have been lost from deuterostomes and nematodes. These genes may have also been lost from mollusks/annelids, although more extensive taxon sampling from these lineages will be required to definitively address this issue. We also observed losses of Dicer1 and AGO1 from *Trichoplax*, suggesting that this simple organism has lost the RNAi machinery required for miRNA-based gene regulation. These findings are consistent with recent studies suggesting *Trichoplax* lacks miRNA genes (Grimson et al. 2008; Hertel et al. 2009).

It appears that the Dicer1/2 and AGO1/2 duplication events both occurred after sponges diverged from the main animal branch but before the cnidarian split. However, the precise timings of the Dicer1/2 and AGO1/2 duplications are difficult to confidently ascertain, given the long branch lengths and lack of resolution at the base of the animal phylogeny in both the Dicer and AGO trees.

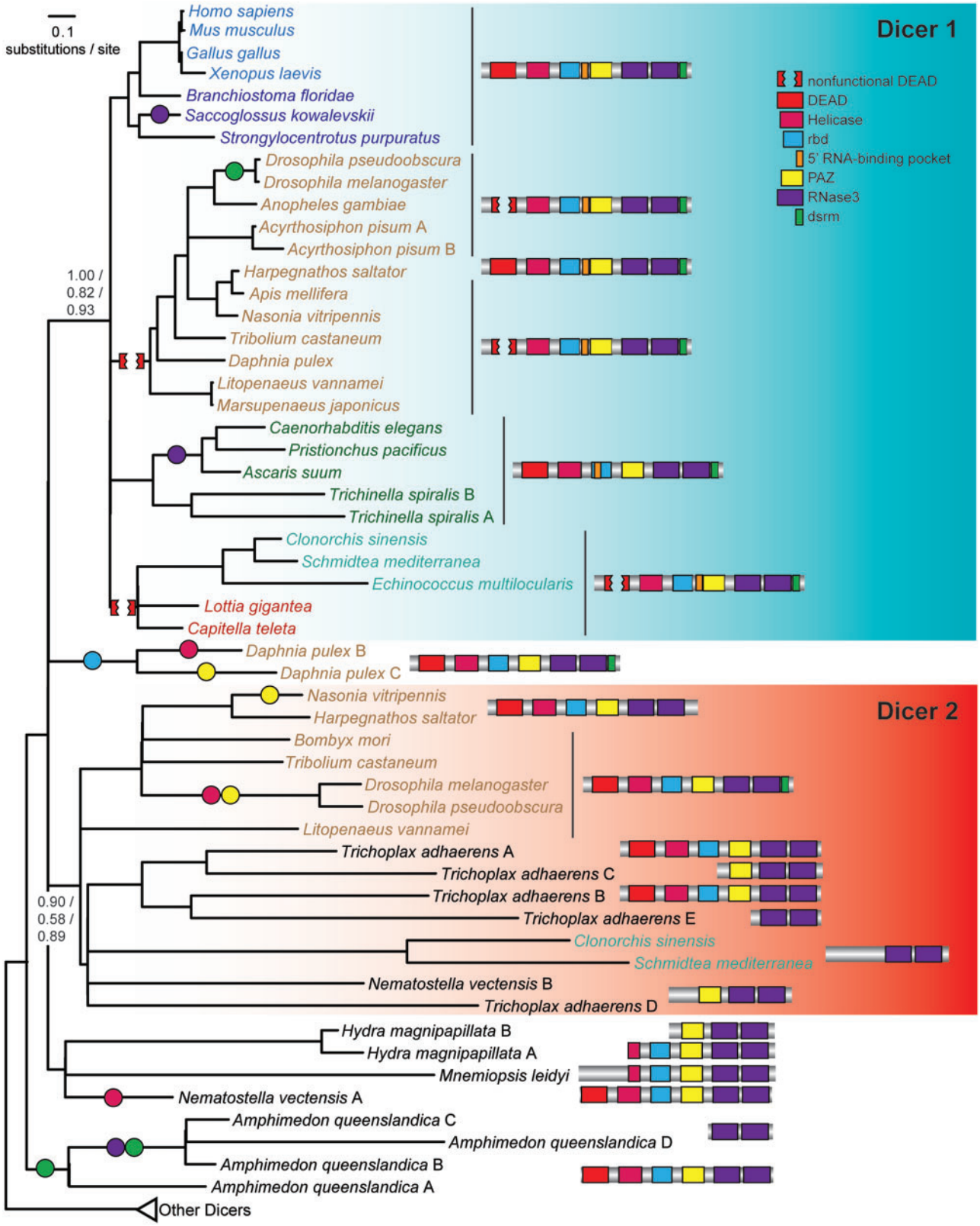

Fig. 3. Dicer duplicated early in animal evolution. We inferred the metazoan Dicer family phylogeny using maximum likelihood and Bayesian methods. Support is shown for key nodes as SH-like aLRT scores/bootstrap proportions/Bayesian posterior probabilities (supplementary figs. S6–S11, Supplementary Material online, for additional support calculations and analyses). We also show the inferred domain architecture for each sequence. We used a branch-sites model to identify protein-coding adaptation from aligned codon sequences (Zhang et al. 2005). Colored circles indicate branches showing significant support for adaptation in each functional domain ($P < 0.05$ after correcting for multiple tests). Species names are colored by taxonomic group.

The domain architecture of large, multi-domain proteins can change over evolutionary timescales, leading to major changes in the protein's functional repertoire. To examine the evolution of Dicer's domain architecture in animals, we identified known functional domains in each sequence using the PFam database (Punta et al. 2011) and confirmed each inferred domain architecture using the SMART (Letunic et al. 2012) and CDD (Marchler-Bauer et al. 2011) databases (supplementary Excel spreadsheet, Supplementary Material online). As figure 3 shows, we observed a complete loss of the DEAD functional domain in arthropod, mollusk, annelid, and planarian Dicer1. Additionally, we observed a corruption of the Walker-B "DECH" motif in the majority of these sequences, which is likely to disrupt the ATP-binding properties of this domain and lead to a complete loss of DEAD/Helicase function (Sengoku et al. 2006). Nematode Dicer1 retained a recognizable DEAD/Helicase domain that appears to be functional. Since the loss of a functional DEAD/Helicase domain in fly Dicer1 has been linked to specialization in miRNA-based gene regulation (Welker et al. 2011), our results suggest that this functional specialization may have occurred multiple times in various protostome lineages, including early in arthropod evolution.

Previous studies have found that antiviral Dicer2 is under intense positive selection in *Drosophila melanogaster* and across the *Drosophila* phylogeny (Obbard et al. 2006; Heger and Ponting 2007; Kolaczkowski et al. 2011). We confirmed that Dicer2 DEAD/Helicase and PAZ domains have experienced positive selection in flies using branch-sites analyses to identify adaptive protein-coding changes (Zhang et al. 2005). However, other than a spattering of adaptive substitutions in various lineages, we found little evidence for recurrent adaptation across the animal Dicer phylogeny (fig. 3). We note that adaptively driven loss of domain function is unlikely to be detectable using branch-sites models based on the nonsynonymous/synonymous substitution rate ratio (dN/dS), as any signal will be rapidly lost due to the fast evolution of nonfunctional protein sequences. For this reason, it is not possible to determine whether the loss of DEAD/Helicase function observed in various protostomes was adaptive.

Evolution of Plant Dicers

We used homology-based gene prediction to identify Dicer genes from fully sequenced plant genomes and reconstructed their phylogeny using maximum-likelihood and Bayesian methods. We found that the Dicer gene underwent a rapid four-way duplication early in plant evolution, giving rise to four distinct Dicer groups (DCL-1 to DCL-4) after the divergence of single-celled green algae but before or around the divergence of moss from higher plants (fig. 4). This result mirrors what we observed for animal Dicer—rapid diversification of the Dicer gene family coincident with the origins of multicellularity—suggesting that diversification of RNAi pathways may have played important and parallel roles in the development of multicellular animals and plants.

As with our analysis of animal Dicers, our plant Dicer phylogeny was robust to different analysis methods, alignment strategies, support measures, and evolutionary models (supplementary table S1 and fig. S13, Supplementary Material online). The same tree topology was recovered using heterogeneous evolutionary models (supplementary figs. S14 and S15, Supplementary Material online) and alignment of the DCL "catalytic core" (supplementary fig. S16, Supplementary Material online), suggesting that our plant phylogeny is reliable.

Although DCL-1 and DCL-3 have clear moss orthologs in all our analyses, suggesting this gene duplication occurred before the divergence of moss from higher plants, the precise timing of the DCL-2/DCL-4 duplication was unresolved in some trees (fig. 4). To examine the DCL-2/4 duplication in more detail, we aligned additional Dicer homologs from a variety of plant genomes and reconstructed the plant DCL phylogeny using this expanded dataset.

The resulting tree identified a moss DCL-4 ortholog and a *Selaginella* DCL-2 ortholog, both with fairly strong statistical support (supplementary fig. S17, Supplementary Material online), consistent with what we found in our phylogeny built from the DCL "catalytic core" (supplementary fig. S16, Supplementary Material online). Although we remain cautious when interpreting these findings—as the timing of the DCL-2/4 duplication was unresolved in some trees—these results suggest that the DCL-2/4 duplication may have also occurred before moss diverged from higher plants, in which case all four DCLs would have originated very early in plant evolution, coincident with the origin of multicellularity.

We observed some changes in the Dicer domain architecture across the plant phylogeny (fig. 4)—similar to what we observed in animals—including lineage-specific losses of functional DEAD/Helicase domains in monocot and dicot DCL-1, which is known to function primarily in miRNA-based gene regulation (Welker et al. 2011). The major difference between plant and animal Dicer domain architecture appears to be the presence of twin C-terminal dsrm domains in many plant Dicers; this architecture was not observed in the animal proteins. Twin dsrm domains were found throughout DCL-4 proteins and in DCL-1s.

The functional consequences of these changes in C-terminal dsrm architecture are unknown. Dsrm domains typically function in coordinating the hand-off of the RNA template from Dicer to an AGO protein, either through protein–RNA or protein–protein interactions (Parker et al. 2008). This suggests that differences in C-terminal domain architecture may play a role in defining specific RNAi pathways by determining which downstream partners a specific Dicer interacts with (Marques et al. 2010).

In contrast to the sparse protein-coding adaptation we observed for animal Dicers, we found strong evidence for recurrent adaptation targeting the PAZ domain of plant DCL-4 throughout monocots and dicots (fig. 4). There was little evidence for adaptation in other functional domains or other DCL groups, suggesting that adaptation has specifically

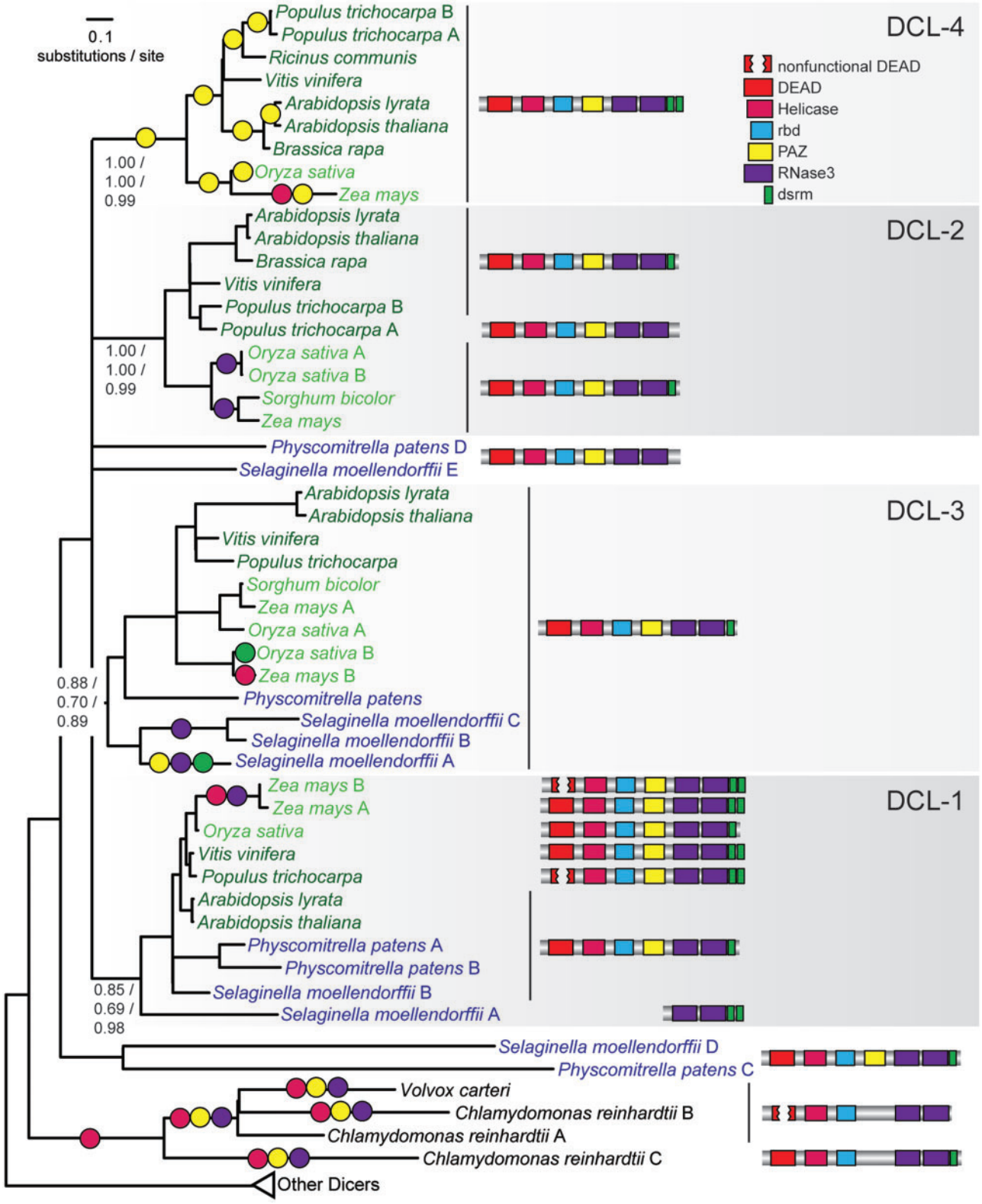

FIG. 4. Dicer duplicated early in plant evolution. We inferred the plant Dicer family phylogeny using maximum likelihood and Bayesian methods. Support is shown as SH-like aLRT scores / bootstrap proportions / Bayesian posterior probabilities (supplementary figs. S13–S17, Supplementary Material online). We additionally display the inferred domain architecture for each sequence. Colored circles on each branch indicate significant support ($P < 0.05$ after correcting for multiple tests) for adaptive protein-coding changes in each functional domain, inferred using branch-sites analysis (Zhang et al. 2005). Species names are colored by taxonomic group.

driven changes in antiviral DCL-4 PAZ function throughout plant evolution.

The Dicer PAZ domain anchors the end of the RNA molecule to facilitate cutting by twin RNase3 domains. As is the case with many similar RNA-binding domains, Dicer PAZ contains a large, positively charged pocket that binds the RNA-end structure primarily through electrostatic interactions (Ma et al. 2004). As expected, we found that this pocket was highly conserved across Dicers, including both plant and animal proteins (fig. 5). However, the plant DCL-4 PAZ domain showed extreme variation in electrostatic distribution across the RNA-binding pocket, even among very closely related plants.

Particularly in monocots, there appears to be a complete reversal of the electrostatic charge of the DCL-4 PAZ RNA binding pocket to a primarily negative charge. This radical change in pocket charge is completely novel among Dicer PAZ domains and is uncharacteristic of PAZ domains in general (Lingel et al. 2004; Wang, Juranek, et al. 2009). Although the functional consequences of these changes in binding-pocket electrostatics are unknown, such extreme changes are highly likely to affect RNA binding, suggesting that DCL-4 may exhibit radically different RNA-binding properties compared with other Dicer proteins.

Given the importance of plant DCL-4 for antiviral immunity (Bouche et al. 2006; Deleris et al. 2006; Diaz-Pendon et al. 2007) and the probable role of the PAZ domain in recognizing specific RNA motifs (Lingel et al. 2004; Ma et al. 2004; Wang, Juranek, et al. 2009), our results suggest that the DCL-4 PAZ domain is likely involved in a long-term evolutionary arms race with viral factors, leading to rapid variation in RNA-binding properties. When we identified specific adaptive substitutions in various DCL-4 homologs, we found that many of them clustered around the RNA-binding pocket, with a particularly high concentration found in the RNA-binding loop between helices 1 and 2 (fig. 6). One position in the middle of this loop is highly variable across DCL-4s and has undergone adaptation in multiple plant lineages; this position is extremely conserved across other Dicer proteins (supplementary fig. S18, Supplementary Material online) and forms a critical RNA-contact in *Giardia* Dicer (Simon et al. 2011). The radical difference in both conservation and amino acid physiochemical properties at this critical RNA-contact residue between DCL-4 and other Dicers further suggests that DCL-4 RNA-binding properties are likely to be distinct from those of other Dicers, although the precise nature of these functional differences is unknown.

Evolution of miRNA Recognition

Recent analyses suggest that miRNA genes may have originated independently in animals and plants, suggesting that miRNA-based gene regulation may have arisen at least twice in eukaryote evolution (Murphy et al. 2008; Lee et al. 2010; Axtell et al. 2011). Dicer recognizes the 3′-end of miRNAs via specific contact residues within the PAZ domain (Zhang et al. 2002, 2004). Efficient processing of miRNAs in humans requires additional contacts with the 5′-end of the miRNA molecule, facilitated by conserved residues within the C-terminal part of the PAZ domain as well as an N-terminal PAZ extension. Recent analyses suggest that human miRNA processing relies primarily on 5′ miRNA recognition (Park et al. 2011).

We examined the patterns of sequence evolution and intron/exon structure around key miRNA-contact residues in animal and plant Dicers to better understand how miRNA processing evolved. We found that a number of 3′-contacting residues and surrounding PAZ-domain sequence were highly conserved across not only all animal Dicer1s but also all plant DCL-1 proteins, suggesting that the biochemical basis for 3′ miRNA recognition was present in the ancestral Dicer before the animal/plant split (fig. 7). Maximum-likelihood ancestral reconstruction confirmed this conclusion, as all four 3′-contacting residues were inferred in animal, plant, and animal/plant/fungi ancestral sequences with strong support (supplementary table S4, Supplementary Material online). These contact residues and surrounding sequence were not conserved in animal Dicer2s or plant DCLs 2–4, suggesting that the mode of RNA recognition and/or binding differs between miRNA-processing and other Dicers. Interestingly, some of the Dicer homologs falling in the unresolved base of the animal tree had this conserved Dicer1-like signature, suggesting that they may have retained the ability to bind the 3′-end of miRNAs (fig. 7).

Sequence alignment of the 5′ miRNA-binding pocket revealed that while 5′-binding residues within the PAZ domain (R996 and R1003) are conserved across eukaryote Dicers (but not other PAZ-containing proteins, supplementary fig. S19, Supplementary Material online), key RNA-contact residues within the N-terminal extension (R778, R780, and R811) are conserved only among bilaterian Dicer1 proteins and absent from other animal and plant Dicers (fig. 7). Gene modeling identified a bilateria-specific exon encoding the entirety of the 5′-binding N-terminal extension sequence that is not present in other animal or plant Dicers, suggesting that this 5′-binding pocket is a bilaterian innovation.

Although all plant Dicer homologs lack the bilaterian N-terminal extension of the PAZ domain required for efficient 5′ miRNA binding, we identified a similar but non-homologous N-terminal extension of the PAZ domain conserved throughout plant DCL-1 but absent from other DCLs (fig. 7). This DCL-1 insertion does not align to the animal 5′ miRNA-binding pocket and has different intron/exon boundaries, supporting the conclusion that it is not homologous to the animal pocket but represents a unique plant DCL-1 sequence.

Both the animal 5′-pocket and the plant DCL-1 insertion contain a large number of positively charged residues (Arg and Lys), which in animals have been shown to be important for miRNA-binding (Park et al. 2011). Although functional analyses will be required to confirm the role of the DCL-1 insertion, evolutionary analysis suggests that the plant DCL-1 insertion may be a novel motif supporting efficient DCL-1-based gene regulation by facilitating 5′ miRNA-binding.

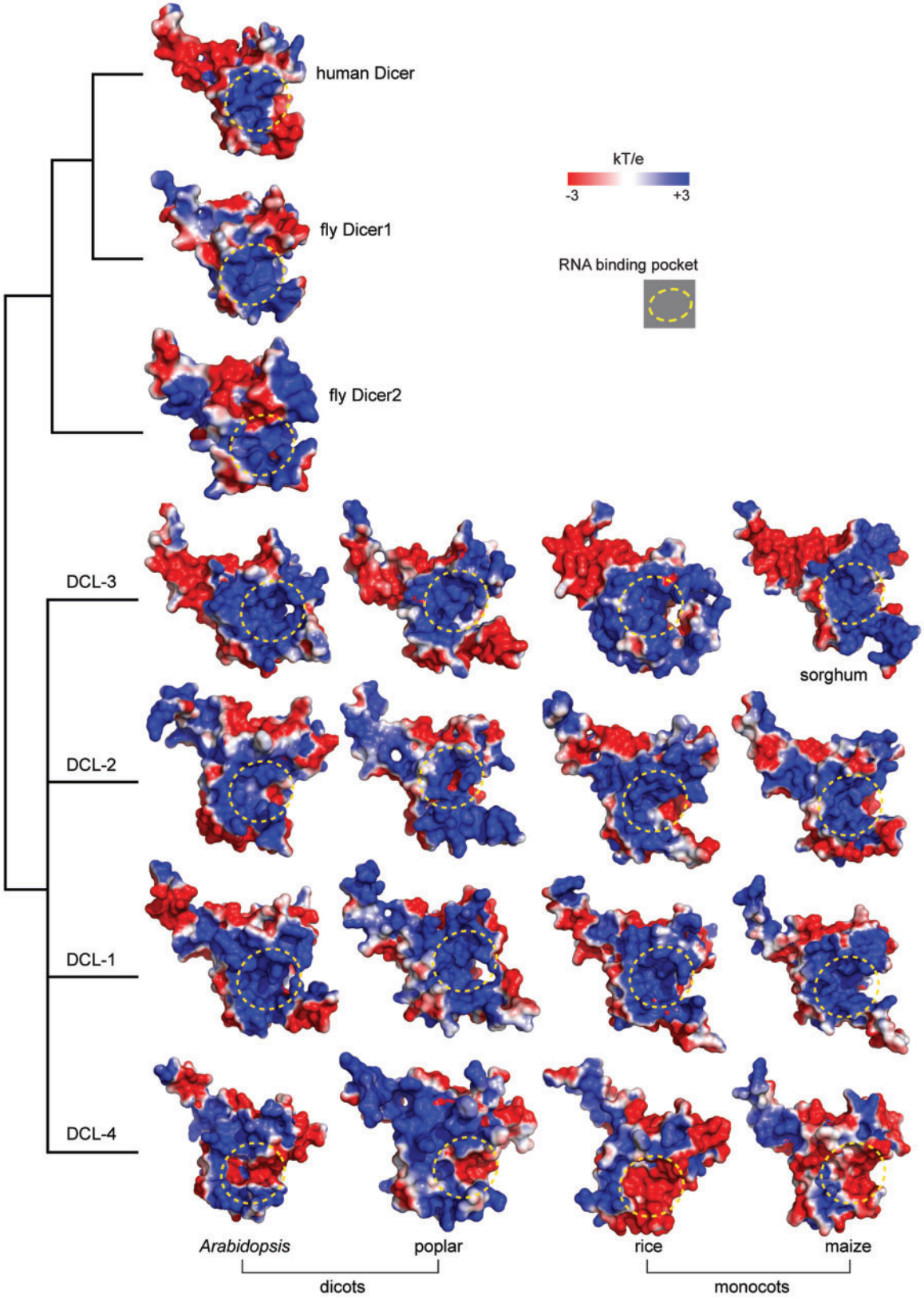

Fig. 5. Adaptive substitutions altered the electrostatic distribution across the RNA-binding pocket of plant DCL-4 PAZ. We inferred the 3D structure and electrostatic distribution of plant and animal Dicer PAZ domains (see Materials and Methods). We plot the electrostatic distribution (kT/e) across the RNA-binding pocket (yellow-dotted outline) of each protein.

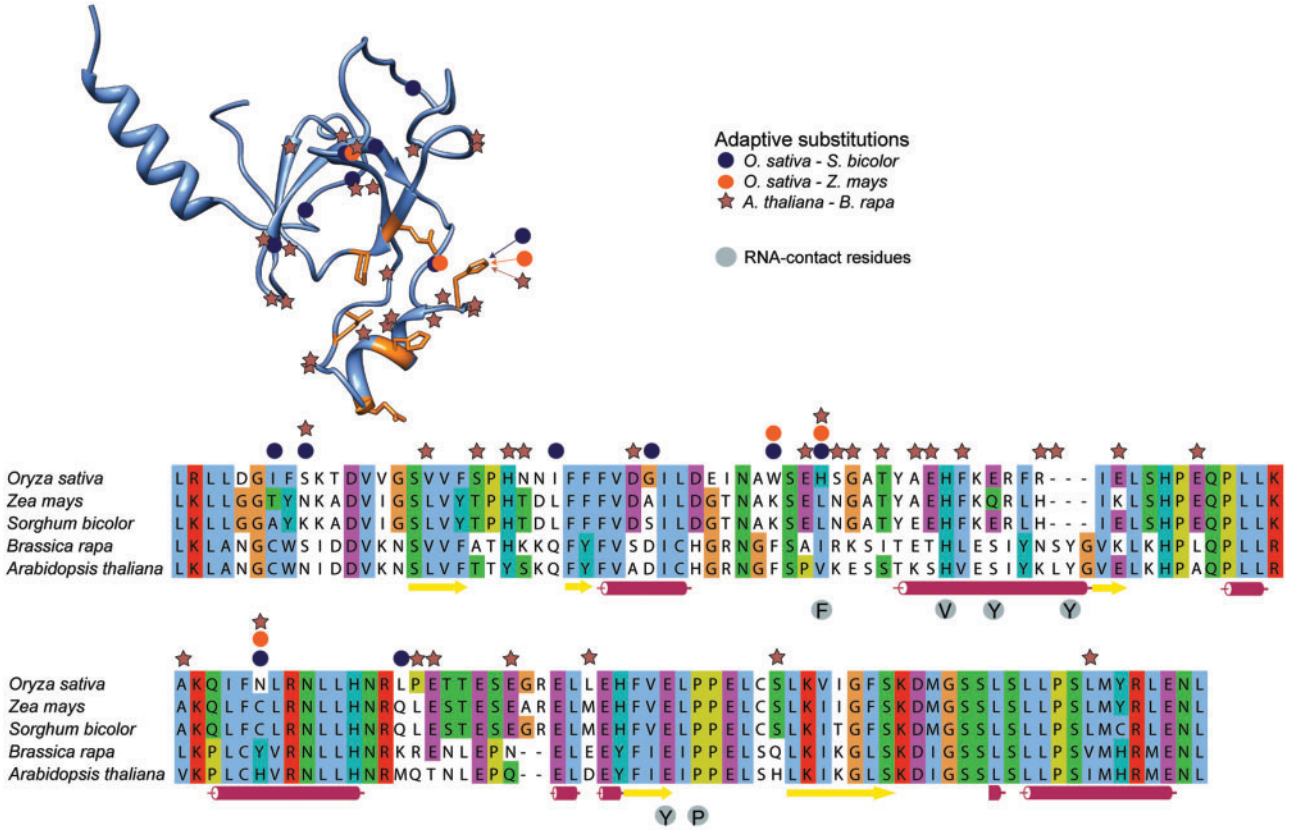

FIG. 6. Adaptive substitutions in plant DCL-4 PAZ domain are likely to affect RNA binding. We inferred adaptive protein-coding substitutions in monocot and dicot DCL-4 PAZ domains using branch-sites models (see Materials and Methods). We plot adaptive substitutions (posterior probability > 0.95) along the protein structure (top) and multiple sequence alignment (bottom). RNA-contacting residues identified in *Giardia* Dicer PAZ (Simon et al. 2011) and AGO PAZ (Wang, Juranek, et al. 2009) are indicated below the sequence alignment.

Discussion

Our results suggest a model of Dicer evolution in which an ancestral eukaryote Dicer duplicated independently and early in animal and plant lineages, coincident with the origins of multicellularity. These early Dicer family expansions may have coincided with duplications of other proteins in the RNAi pathway (such as AGO, supplementary fig. S12, Supplementary Material online) as well as with the independent origins of miRNA genes in animals and plants (Murphy et al. 2008; Lee et al. 2010; Axtell et al. 2011), suggesting that the requirement for complex gene regulation to support multicellular organisms may have independently recruited RNAi pathways and driven their elaboration in animals and plants.

Our results suggest that the major functional repertoire of the Dicer family was likely present in early multicellular animals and plants, with subsequent functional refinement and gene loss events "fine-tuning" RNAi function in various lineages. Specifically, bilaterian Dicer1 appears to have evolved a novel 5′ miRNA binding mechanism, which increased miRNA-processing efficiency, a trait that may have arisen independently in plant DCL-1. We also observed parallel losses of the DEAD/Helicase domain of Dicer1 and DCL-1 in various protostome and plant lineages, which likely focused these proteins' functions on miRNA processing.

Although the importance of RNAi for antiviral immunity in vertebrates is unclear (Cullen 2006), Dicer2 and DCL-4 are known to play critical antiviral roles in invertebrate animals and plants, respectively (Hamilton and Baulcombe 1999; Zamore et al. 2000; Aliyari and Ding 2009). Our model of animal Dicer evolution suggests that Dicer2's antiviral role probably arose very early in animal evolution, with Dicer2 being subsequently lost from deuterostomes and nematodes. This raises the possibility that alternative antiviral strategies may have compensated for loss of RNAi-based antiviral immunity in these lineages.

Although we detected no evidence for long-branch attraction or other phylogenetic artifacts from any of the major factors known to cause such errors, there is always the possibility that the inferred phylogeny is not the correct tree. In the case of the animal Dicer phylogeny, the presence of multiple long Dicer2 branches falling at the base of the tree is consistent with a long-branch attraction hypothesis in which fast-evolving Dicer2 duplicates—perhaps arising in multiple lineages—artifactually group at the base of the animal tree. More extensive taxon sampling of basal animal lineages will help resolve this issue.

Perhaps, the most striking result from our study is the extremely strong signature of recurrent adaptation throughout the evolution of the plant DCL-4 PAZ domain and the radical variability in its RNA-binding pocket compared with

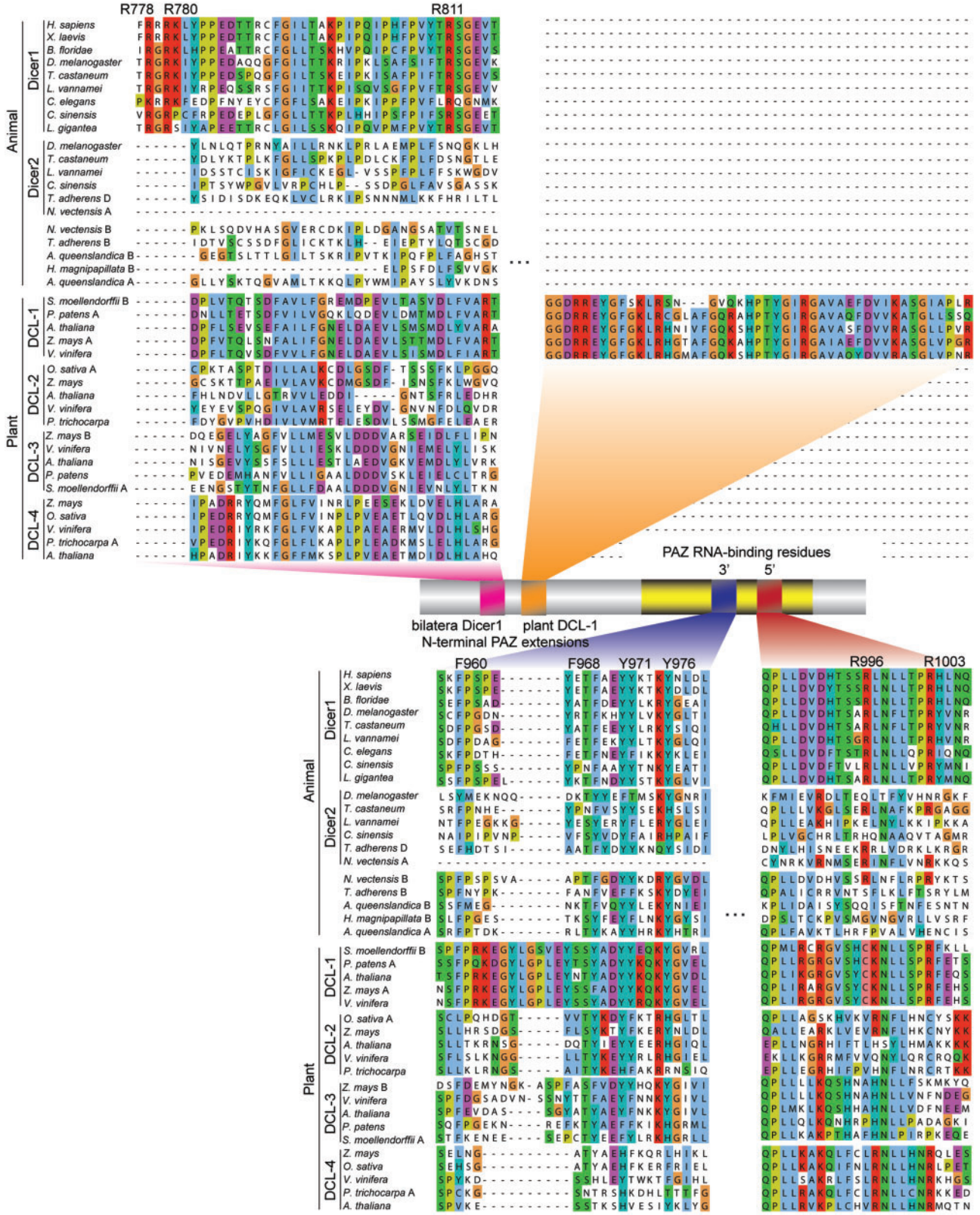

FIG. 7. Examination of Dicer sequence alignment reveals bilaterian- and DCL-1-specific N-terminal PAZ extensions likely to contribute to 5′ RNA-binding. We display parts of aligned Dicer sequences from representative taxa. Protein domain architecture is show in the middle, with the PAZ domain indicated in yellow. Specific residues previously shown to contribute to 3′ and 5′ RNA-binding are indicated at the top of each alignment (Lingel et al. 2004; Wang, Juranek, et al. 2009; Park et al. 2011; Simon et al. 2011). Alignment of conserved N-terminal PAZ extensions from bilateria and plant DCL-1 are indicated in pink and orange, respectively.

those of other Dicers. The PAZ domain anchors the end of the RNA molecule and may play an important role in recognizing unwanted RNAs (Simon et al. 2011). Most of the adaptive protein-coding substitutions we identified in DCL-4 PAZ are located around the RNA-binding pocket, and some affect key RNA-contact residues, suggesting that changes in viral RNA structure may have driven adaptation in DCL-4 PAZ. Alternatively, several viral suppressors of RNAi (VSRs, viral proteins functioning to disrupt the host's RNAi-based immune response) are known to antagonize the PAZ domain and may have driven its rapid evolution in plants (Baumberger et al. 2007; Singh et al. 2009; Hamera et al. 2012). Uncovering the precise interactions driving this clear evolutionary arms race will help determine the general biochemical forces governing host-pathogen interactions.

RNAi biochemistry operates on the elegantly simple principle of using an RNA template in coordination with enzymatic proteins to target and regulate complementary RNA molecules. Organisms have exploited this simple principle to support multiple functional roles, such as gene regulation, antiviral immunity and preserving genome integrity. Our analyses have revealed the RNAi system to be highly dynamic across evolutionary timescales, with parallel gains and losses of entire functional pathways, long-term evolutionary arms races, and changes in protein domain architecture and sequence driving functional divergence. Given this dynamic backdrop, we expect RNAi to function differently in different organisms; future comparative-functional studies are expected to further illuminate the details of how evolution has shaped RNAi function.

Materials and Methods

Gene Identification

We used confirmed functional Dicer genes as queries to identify new Dicer homologs from fully sequenced genomes using TBLASTN, with a liberal e-value cutoff of 1.0 to identify all potential homologs (Altschul et al. 1997). Sequences recovered from one round of TBLASTN were recursively used as queries until no further sequences were detected. We additionally performed TBLASTN searches using individual functional domains as query sequences, to identify potential Dicer homologs with different domain architectures.

We used homology-based gene prediction based on hidden Markov models (HMMs) to identify the complete open reading frame of each potential Dicer homolog. We excised ±2,000 nucleotides of genomic sequence surrounding each TBLASTN hit and calculated log-odds scores for coding exons using FGENESH+ with a log-odds cutoff of 50.0 to identify strongly supported exons (Salamov and Solovyev 2000). Potential exon mis-predictions were manually refined using dot-plot analysis and 3-frame translations of the genomic sequence. The closest known homolog recovered from PSI-BLAST search of Genbank was used as a guide for homology-based gene prediction. Coding exons were re-predicted, if necessary, following multiple sequence alignment of the resulting protein sequences (see below), and non-homologous sequences were removed.

Protein Domain Identification

Protein functional domains were identified by sequence search of the PFam database (Punta et al. 2011), using an e-value cutoff of $1.0e^{-5}$ to identify strongly supported domains. PFam hits were verified via sequence searches of the SMART (Letunic et al. 2012) and NCBI conserved domain databases (Marchler-Bauer et al. 2011). We also manually verified loss of functional DEAD domains by examining the highly conserved "DECH" motif in each sequence. We identified Dicer 5′ RNA-binding pockets by aligning Dicer PAZ domains and relevant N-terminal sequences and manually examining these alignments for the presence or absence of experimentally determined RNA-binding residues (Park et al. 2011).

Sequence Alignment

Sequences were aligned using Muscle v3.8.31 (Edgar 2004) (default parameters), Mafft v6.850b (Katoh et al. 2002) (1,000 iterations and –genafpair option) and Kalign v2.03 (Lassmann and Sonnhammer 2005) (default parameters). All alignments were manually refined to remove obvious alignment errors and preserve alignment of identified functional domains. Resulting alignments were processed using Gblocks v0.91b (Castresana 2000) to remove ambiguously aligned regions prior to phylogenetic analysis, allowing gaps in 1/2 of the sequences. Additionally, we produced an elision alignment by concatenating all three individual Gblocks-processed alignments (Wheeler et al. 1995). In addition to full-length and elision alignments, we produced individual-domain alignments consisting of only the DEAD/Helicase domains and only the Dicer "catalytic core" PAZ+RNase3 domains. These individual-domain alignments were not processed using Gblocks.

Phylogenetic Analysis

Maximum likelihood trees were inferred using PhyML v3.0 (Guindon and Gascuel 2003; Guindon et al. 2010), with the best-fit evolutionary model being selected using ProtTest v3.0 (Darriba et al. 2011). Clade support was calculated using SH-like approximate likelihood ratio tests (Anisimova et al. 2011) and bootstrap proportions (500 replicates).

Bayesian phylogenies were reconstructed using MrBayes v3.1.2 (Ronquist et al. 2012), assuming the default priors. We integrated over all available substitution models and assumed a discrete-gamma model of among-site rate variation. We evaluated chain convergence by performing two independent runs and terminating the analysis when the average standard deviation in clade probabilities between the two runs fell below 0.01.

We also reconstructed phylogenetic trees using two approaches that have been shown to improve phylogenetic accuracy. First, we used a Bayesian approach implemented in PhyloBayes v2.3 that incorporates variation in amino acid frequencies across the sequence alignment (Lartillot et al. 2009). Second, we used a maximum-likelihood method that incorporates across-site heterogeneity in branch lengths (Kolaczkowski and Thornton 2008). For this approach, the

best-fit number of branch-length categories was identified using the Akaike information criterion (AIC).

Protein Adaptation

We evaluated protein-coding adaptation across the phylogeny using a branch-sites test implemented in PAML v4.5 (Zhang et al. 2005; Yang 2007). We used a four-parameter branch-sites model: a proportion of sites in the alignment (p_0) are assumed to be under purifying selection, evolving at nonsynonymous/synonymous rate $0 < w_0 < 1$. A second class of p_1 sites are neutral ($w_1 = 1$), and a third class of sites ($p_2 = [1 - p_0 - p_1]p_0/[p_0 + p_1]$) evolve under positive selection ($w_2 > 1$). The null hypothesis constrains $w_2 = 1$.

For each branch on the phylogeny, we calculated the likelihood ratio of the positive-selection model ($w_2 > 1$) versus the constrained neutral model ($w_2 = 1$) and estimated P values using the chi-square distribution with one degree of freedom (Zhang et al. 2005). We corrected for multiple testing using a Bonferroni correction. We additionally confirmed any positive results by re-testing after removing fast-evolving taxa, which might cause false positives. We report only those results that were positively identified ($P < 0.05$) in both analyses.

Individual codons were classified as adaptive or non-adaptive using Bayes empirical Bayes (BEB) posterior probabilities calculated by PAML. For each codon, we compare the probability that the codon evolved with $w_2 > 1$ in a given lineage to the sum of the probabilities that the codon evolved with $0 < w_0 < 1$, $w_1 = 1$ or $w_2 > 1$, each weighted by the estimated proportion of sites evolving under that category. Parameter uncertainty is incorporated by integrating over diffuse prior distributions (Zhang et al. 2005). Sites with BEB posterior probability >0.95 were considered adaptive protein-coding changes.

Structural Modeling

We identified homologous X-ray structures via sequence search of the Protein Data Bank (Bernstein et al. 1978). Sequences were aligned to the structural template manually following initial multiple-sequence alignment by Muscle. We built five structural models of each sequence using Modeller v9.10 (Sali and Blundell 1993) and report results obtained using the best model, selected based on DOPE score (Shen and Sali 2006).

Structural models were processed using PROPKA and PDB2PQR to determine residue side-chain pKas, optimize the structure for favorable hydrogen bonding and calculate charge and radius parameters from electrostatic force fields (Dolinsky et al. 2004; Rostkowski et al. 2011). Electrostatic surface potentials were estimated from processed structures using APBS (Baker et al. 2001) and projected onto the molecular surface for visualization.

Supplementary Material

Supplementary Excel spreadsheet, tables S1–S4, and figures S1–S19 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors are grateful to Ryan Joseph for genomic sequence data. This work was supported by the University of Florida.

References

Aliyari R, Ding SW. 2009. RNA-based viral immunity initiated by the Dicer family of host immune receptors. *Immunol Rev.* 227:176–188.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 60:685–699.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.* 12:221.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A.* 98:10037–10041.

Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, Cowman AF, Crabb BS, de Koning-Ward TF. 2009. Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res.* 37:3788–3798.

Baumberger N, Tsai CH, Lie M, Havecker E, Baulcombe DC. 2007. The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. *Curr Biol.* 17:1609–1614.

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.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1978. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys.* 185:584–591.

Blevins T, Rajeswaran R, Shivaprasad PV, et al. (11 co-authors). 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* 34:6233–6246.

Bouche N, Lauressergues D, Gasciolli V, Vaucheret H. 2006. An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* 25:3347–3356.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17: 540–552.

Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet.* 50: 81–99.

Cullen BR. 2006. Is RNA interference involved in intrinsic antiviral immunity in mammals? *Nat Immunol.* 7:563–567.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.

Davalos V, Esteller M. 2012. Rolling the dice to discover the role of DICER in tumorigenesis. *Cancer Cell.* 21:717–719.

de Jong D, Eitel M, Jakob W, Osigus HJ, Hadrys H, Desalle R, Schierwater B. 2009. Multiple Dicer genes in the early-diverging metazoa. *Mol Biol Evol.* 26:1333–1340.

Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O. 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* 313:68–71.

Diaz-Pendon JA, Li F, Li WX, Ding SW. 2007. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* 19:2053–2063.

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665–W667.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in budding yeast. *Science* 326:544–550.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Fryxell KJ. 1996. The coevolution of gene family trees. *Trends Genet.* 12: 364–369.

Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M. 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol.* 6: 784–791.

Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler JL. 2006. Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nat Immunol.* 7:590–597.

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H. 2005. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol.* 15:1494–1500.

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE. 2000. Co-evolution of proteins with their interaction partners. *J Mol Biol.* 299:283–293.

Grimm D. 2012. All for one, one for all: new combinatorial RNAi therapies combat hepatitis C virus evolution. *Mol Ther.* 20:1661–1663.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455: 1193–1197.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52: 696–704.

Hall IM, Noma K, Grewal SI. 2003. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc Natl Acad Sci U S A.* 100:193–198.

Hamera S, Song X, Su L, Chen X, Fang R. 2012. Cucumber mosaic virus suppressor 2b binds to AGO4-related small RNAs and impairs AGO4 activities. *Plant J.* 69:104–115.

Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286:950–952.

Hammond SM. 2005. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett.* 579:5822–5829.

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.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res.* 17:1837–1849.

Hertel J, de Jong D, Marz M, Rose D, Tafer H, Tanzer A, Schierwater B, Stadler PF. 2009. Non-coding RNA annotation of the genome of *Trichoplax adhaerens*. *Nucleic Acids Res.* 37:1602–1615.

Huelsenbeck JP, Rannala B. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276:227–232.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15: 2654–2659.

Kolaczkowski B, Hupalo DN, Kern AD. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol Biol Evol.* 28:1033–1042.

Kolaczkowski B, Thornton JW. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy. *Mol Biol Evol.* 25: 1054–1066.

Koscianska E, Starega-Roslan J, Krzyzosiak WJ. 2011. The role of Dicer protein partners in the processing of microRNA precursors. *PLoS One* 6:e28548.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.

Lassmann T, Sonnhammer EL. 2005. Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6:298.

Lee HC, Li L, Gu W, et al. (11 co-authors). 2010. Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol Cell.* 38:803–814.

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40:D302–D305.

Lingel A, Simon B, Izaurralde E, Sattler M. 2004. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol.* 11:576–577.

Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X. 2003. R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 301:1921–1925.

Liu X, Jiang F, Kalidas S, Smith D, Liu Q. 2006. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA* 12:1514–1520.

Ma JB, Ye K, Patel DJ. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429: 318–322.

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. 2006. Structural basis for double-stranded RNA processing by Dicer. *Science* 311:195–198.

Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* 19:2979–2990.

Marchler-Bauer A, Lu S, Anderson JB, et al. (27 co-authors). 2011. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39:D225–D229.

Marques JT, Kim K, Wu PH, Alleyne TM, Jafari N, Carthew RW. 2010. Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat Struct Mol Biol.* 17:24–30.

Merritt WM, Bar-Eli M, Sood AK. 2010. The dicey role of Dicer: implications for RNAi therapy. *Cancer Res.* 70:2571–2574.

Murphy D, Dancis B, Brown JR. 2008. The evolution of core proteins involved in microRNA biogenesis. *BMC Evol Biol.* 8:92.

Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol.* 16:580–585.

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18:1655–1666.

Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. 2011. Dicer recognizes the 5′ end of RNA for efficient and accurate processing. *Nature* 475:201–205.

Parker GS, Maity TS, Bass BL. 2008. dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J Mol Biol.* 384:967–979.

Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. 1997. Correlated mutations contain information about protein-protein interaction. *J Mol Biol.* 271:511–523.

Pecot CV, Calin GA, Coleman RL, Lopez-Berestein G, Sood AK. 2011. RNA interference in the clinic: challenges and future directions. *Nat Rev Cancer.* 11:59–67.

Preall JB, Sontheimer EJ. 2005. RNAi: RISC gets loaded. *Cell* 123:543–545.

Punta M, Coggill PC, Eberhardt RY, et al. (16 co-authors). 2011. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.

Rossbach M. 2010. Small non-coding RNAs as novel therapeutics. *Curr Mol Med.* 10:361–368.

Rostkowski M, Olsson MH, Sondergaard CR, Jensen JH. 2011. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct Biol.* 11:6.

Salamov AA, Solovyev VV. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* 10:516–522.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 234:779–815.

Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S. 2006. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* 125:287–300.

Shabalina SA, Koonin EV. 2008. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol.* 23:578–587.

Shen MY, Sali A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15:2507–2524.

Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, Sehr P, Pillai RS, Carlomagno T. 2011. Recognition of 2′-*O*-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. *Structure* 19:172–180.

Singh G, Popli S, Hari Y, Malhotra P, Mukherjee S, Bhatnagar RK. 2009. Suppression of RNA silencing by Flock house virus B2 protein is mediated through its interaction with the PAZ domain of Dicer. *FASEB J.* 23:1845–1857.

Ullu E, Tschudi C, Chakraborty T. 2004. RNA interference in protozoan parasites. *Cell Microbiol.* 6:509–519.

van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, Andino R. 2006. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* 20:2985–2995.

Volpe T, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, Allshire RC. 2003. RNA interference is required for normal centromere function in fission yeast. *Chromosome Res.* 11:137–146.

Wang HW, Noland C, Siridechadilok B, Taylor DW, Ma E, Felderer K, Doudna JA, Nogales E. 2009. Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol.* 16:1148–1153.

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ. 2009. Nucleation, propagation, and cleavage of target RNAs in Ago silencing complexes. *Nature* 461:754–761.

Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, Liu Q, Bass BL. 2011. Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol Cell.* 41:589–599.

Wheeler WC, Gatesy J, DeSalle R. 1995. Elision: a method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. *Mol Phylogenet Evol.* 4:1–9.

Wu DD, Irwin DM, Zhang YP. 2011. Correlated evolution among six gene families in *Drosophila* revealed by parallel change of gene numbers. *Genome Biol Evol.* 3:396–400.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101:25–33.

Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W. 2002. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* 21:5875–5885.

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.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.