

Phylogenetic Analysis of Canonical/non-canonical Dicers and RNase III Containing Proteins in Fungal Kingdom

Jiayao Wu

University of Helsinki

Jaeyoung Choi

KIST Gangneung Institute of Natural Products

Fred O. Asiegbu

University of Helsinki

Yong-Hwan Lee (✉ yonglee@snu.ac.kr)

Seoul National University

Research Article

Keywords: RNase III domain, dsRNA binding domain, evolution, fungi

Posted Date: December 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1111034/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Dicers were member of RNase III containing proteins family with important RNAi function in eukaryotes. In this study, we tried to address the potential distribution of all RNase III containing proteins among the fungal kingdom, as well as their possible evolution paths including canonical Dicers, non-canonical Dicers and non-canonical Dicer-like proteins.

Result: RNase III containing proteins were collected from 83 species, and discussed the features and possible evolution pathways of RNase III containing protein family. In general, RNase III containing protein family could be characterized into three different groups as canonical Dicer, non-canonical Dicers and Dicer-like proteins based on their domain structures and functional annotation. Most eukaryotes obtained multiple RNase III protein of different types at the same time, including canonical Dicers and Dicer-like proteins. Phylogenetic analysis showed that the RNase III domains were different between canonical Dicers and Dicer-like proteins, for the first and second RNase III domains had different insertions in different regions with certain extent of conservation. Regardless of the types, RNA-binding domains in RNase III protein family were very similar to each other. Furthermore, short insertions were found in different positions from the first and second RNase III domains in canonical Dicers separately. RNA-binding domains from all types were quite similar to each other.

Conclusion: RNase III containing proteins in general widely exist in eukaryotes with minor divergent among different types and groups. Suggested by the RNase III and RDB domains, the canonical/non-canonical Dicers Dicer-like proteins might share the same ancestor and have evolved from separate mild-types through different directions.

Background

Dicer was a type of endoribonuclease, one of the critical components required for RNA interference (RNAi) that widely conserved in eukaryotes to regulate gene expression at transcriptional (TGS) or post-transcriptional level (PTGS) [1]. It formed RNA-induced silencing complex (RISC) and cleaved double-stranded small RNAs (dsRNA) that triggered the degradation of homologous mRNAs [2]. RNAi played essential roles in many cellular processes in fungal kingdom, including genomic stability and defense, heterochromatin formation, meiotic silencing and other gene regulation function [3–5]. Model fungi *Neurospora crassa* was one of the best studied fungal species of their RNAi proteins and diverse functional small RNAs [6]. MSUD RNAi proteins are paralogs but sharing the same proteins comparing to Quelling machinery [7, 8]. Another fungal species *Magnaporthe oryzae*, the causal agent of the rice blast, obtained similar RNAi pathways and functions [9, 10]. RNAi was also associated with heterochromatin and epigenetic gene silencing at specific chromosomal loci in fission yeast *Schizosaccharomyces pombe* [11]. It seemed that budding yeast completely lost RNAi, only with rare cases of Dicers and Argonautes other than classical RNAi function existed in a few budding yeasts [12]. To date, many RNAi pathways have been well discussed numbers of fungal species [13].

Dicers belonged to Ribonuclease III protein family, a big family of endoribonucleases played important dsRNA processing in RNAi as well as RNA decay and maturation [3, 14–16]. As for a canonical Dicer, two RNase III domains (RIIID), DEAD boxes at N-terminal, RNA helicase domain, double-stranded RNA-binding domains (dsRBD) and Dicer dimerisation domain were functional necessary [17]. Besides, an additionally Piwi-Argonaute-Zwille (PAZ) domain was also important domain for Dicers in plant and animal [15, 18], but much less common in fungi [19]. Dicers were mainly responsible for small RNA biogenesis from long double-stranded RNA (dsRNA) by diverse pathways especially through RNAi, as well as other RNA-processing and DNA damage response functions [2, 20, 21]. The ubiquitous members of Ribonuclease III family commonly obtained RNase III domains and double-stranded RNA-binding domains, and functioned in a broader range of RNA processing [22]. Drosha, class 3 Ribonuclease III family, encoded with double RIIID enzymes and essential for canonical miRNA generating pathway especially in high eukaryotes [23]. Class 1 and 2 Ribonuclease III members contained single RIIID, such as RNase III that regulated mRNA stability in *Escherichia coli* [24] and *Bacillus subtilis* Mini-III catalyzed maturation of rRNA without dsRNA binding domain [25]. RNase III containing proteins in yeast mostly involved in snoRNA and rDNA processing, RNT1 in *Saccharomyces cerevisiae* for instance [26, 27]. Pac1 in *Schizosaccharomyces pombe* additionally involved sexual development as well [28]. These RNase III-containing proteins made up a big protein family with related functions and appeared in species of different taxonomy ranges. Interestingly, the Class 3 Ribonuclease III (Drosha) is not found in the fungi Kingdom, while the Class 4 Ribonuclease IIIs (canonical Dicers) are much conserved in the eukaryotes [29].

Recently, features in Dicer sequences and distribution and evolution of Dicers have also been discussed by bioinformatics point of view [30, 31]. Recent reports showed such proteins in *Candida albicans* Dicer and *Kluyveromyces polysporus* possessed Dicer-like activity and capable of producing siRNAs [32, 33]. They were of a much simpler structure with only one RNase III domain adjacent to a dsRNA-binding domain (dsRBD) and an additional C-terminal dsRBD [34], and likely classified into Class 2 Ribonuclease III [35]. So far, these non-canonical Dicers only presented in budding yeasts missing with canonical Dicers and RNAi [36]. Non-canonical Dicer with identical structure with its small RNA producing function was characterized in human protozoan parasite [37], thusly non-canonical RNAi-related pathways were expected in *Plasmodium parasites* as well [38]. These RNase III-containing enzymes with dicer-like activity indicated their structure and functional relations with the canonical Dicers. Besides, non-canonical functions of canonical Dicers were also discovered [39, 40]. Till now, only canonical Dicer was highlighted in these *in silico* analysis, lacking of other RNase III-containing proteins and non-canonical Dicers in yeast as well [41]. These reminded us that non-canonical Dicers should be included when discuss the question of how dicers evolved.

In this study, we aimed at the distribution and comparative evolution analysis of all RNase III containing proteins focused in fungal kingdom, including canonical, non-canonical Dicers and RNase III domain-containing proteins, i.e. non-canonical Dicer-like proteins (NCDL). By considering both RNase III domain and dsRBD sequences, we tried to reveal the relationship of RNase III family members as well as a few rare protein structures which may help to understand the possible evolution paths of fungal RNase III

family that were thusly suggested. The lines of evidences brought up the need for more comprehensive analysis of RNase III domain-containing proteins and especially in fungi for their different structures in canonical and non-canonical Dicers.

Results

Domain structures of characterized Dicers and their distribution across the taxonomy

There were limited numbers of well characterized Dicers in fungal kingdom. Taking model fungi *Neurospora crassa* and *Magnaporthe oryzae* for instances, they both encoded two redundant Dicers with typical structures of two separate RNase III domains and at least one double-stranded RNA-binding domain [42, 43], aforementioned as “canonical” Dicers. Despite the presence of predicted Argonaute and/or RDRP-encoding genes (RNA-dependent RNA polymerase), a number of species like *Saccharomyces castellii* [34], *Candida albicans* [41], and *Kluyveromyces polysporus* [32] lacking of canonical Dicers kept single non-canonical Dicer for dsRNA processing instead. These non-canonical Dicers shared identical domain profile of one RNase III domain and two dsRBDs.

To address the existing of RNase III family including canonical and non-canonical Dicers and their relationship, a total number of 427 RNase III domain-containing protein sequences were retrieved from the proteome sequences from 83 species covering 12 bacteria, three Oomycetes, and 68 fungi. Canonical Dicers, non-canonical Dicers and NCDLs shared essential domains but varied in the numbers and configuration resulting in differentiated functions. All types of proteins were of different composition of three domains, i.e. RNase III domain, a RBD preceding RNase III domain normally found in canonical Dicers (Dicer dimerisation domain, IPR005034) and RBDs following RNase III domain (IPR014720). Based on this observation, all RNase III-containing proteins were divided into eight types of three subgroups (Figure 1A): canonical Dicers (TYPE1 and 2), non-canonical Dicers (TYPE3) and non-canonical Dicer-like proteins (TYPE4 to 8). To discuss them across the kingdoms, additional types exclusively found in five model species of animals and plants were also included (TYPE9-10, Supplementary Figure S1).

Canonical Dicers were subdivided into two types depending upon the presence of additional dsRBD at C-terminus (as TYPE1 and TYPE2), and non-canonical Dicers were assigned as TYPE3. These eight types of proteins distributed interestingly in a taxon-specific manner that Dicers and NCDLs always presented together in most species each phylum or subphylum had a relatively fixate combinations of different types of Dicers and NCDLs (Figure 1B).

Clearly, no canonical Dicers appeared in Bacteria, the simplest forms of TYPE4 and 8 proteins processed all dsRNA. Canonical Dicers started to present with NCDLs in low eukaryotes. TYPE1 much outnumbered of TYPE2 and presented in the majority of fungal species and Oomycota, while TYPE2 appeared only in Pezizomycotina (Figure 1B). Canonical Dicers (TYPE1 and 2) were completely absent in Saccharomycotina, while Ustilaginomycotina and Microsporidia with only canonical Dicer. The

interesting parallel existing of TYPE1 and TYPE 2 Dicers was only found in Ascomycetes, mostly Sordariomycetes (12/15), meanwhile partial species in other class Leotiomycetes (1/2), Dothideomycetes (2/4) and Eurotiomycetes (2/10) obtained both types of canonical Dicers. It was not clear how TYPE2 differed from TYPE1 in its function and why necessary it was to keep both, TYPE2 were not simply duplication of TYPE1. Although *Neurospora crassa* encoded both types of Dicers as functional redundancy [42], *Magnaporthe oryzae* Dicers showed clear diversification in function and structure as well [43, 44]. There were also three species obtained only TYPE2 Dicers: *Verticillium albo-atrum* VaMs.1021, *Botrytis cinerea* and *Mycosphaerella graminicola*. Budding yeast seemed completely lost the RNAi pathway, TYPE3 instead solely presented in four *Saccharomyces* with validated in RNA processing function, including previously characterized non-canonical Dicers ScaDcr 1 (TYPE3) in *Saccharomyces castellii* and CaDcr1 in *Candida albicans* SC5314. Two additional proteins as TYPE3 structure in *Pichia stipites* and *Kluyveromyces polysporus* were also included in our data set. Furthermore, non-canonical Dicer was also reported in *Entamoeba histolytica* [37], showing a broad existing of such protein.

Among all NCDLs, TYPE4 and TYPE8 were very commonly in most species neglecting the taxa. TYPE4 could found in all species except for one bacterium, one Oomycete and seven fungi species. TYPE8 as the simplest form of all, was absent in most Bacteria (11 out of 13) and 7 fungi species somehow. It was obvious that TYPE5, 6 and 7 were rare: Type5 presented only in two Oomycetes and seven fungi; TYPE6 found in seven fungi and TYPE7 were in barely three. All these types existed in the five animal and plant species, with a few proteins that do not fit into any types described.

Phylogenetic analysis on RNase III domain

The relationship of RNase III domains showed in phylogenetic tree constructed by all RIIIDs despite of protein types. There was a quite clear distinction: RIIIDs from canonical Dicers separated from the ones from NCDLs (Figure 2. Clade A and B vs Clade C and D); RIIIDs in the same position were quite similar for the first RIIIDs were more similar to each other than the second RIIIDs. The first RIIIDs from canonical Dicers (S1 and S3) were exclusively grouped in the vicinity of each other in Clade A, as well as S2 and S4 (the second RIIIDs in the canonical Dicers) in Clade B. None of S1-4 (canonical Dicer RIIIDs) was found in the other half of the tree, where S6 and S10 were prevalent. Meanwhile, the S5 from non-canonical dicers scattered mostly in Clade C with S6 rather than clustered independently, showing its closer relationship with non-canonical RIIIDs. Most S8 and S9 from TYPE6 NCDLs located with S1/3 in Clade A and S2/4 in Clade B separately (10 out of 13), while S10 and S11 from TYPE7 NCDLs were all located with S6. As for RIIIDs in NCDLs, S6 and S12 formed two major clades of NCDLs individually with slightly mixed. Most S6 RIIIDs formed an S6-dominant clade (Clade C) harboring with a few of S5, S7, S10 and a small group of S12. There were some S6 RIIIDs interleaved with the majority of S12 in Clade D. Shown in the phylogenetic tree, S1/3/8 and S2/4/9 were close to each other, respectively, which was in accordance with their position in the protein sequences, i.e. S1/3/8 were closer to the N-terminus and S2/4/9 were near the C-terminus. The topology implied that RIIIDs from Dicers might have evolved from RIIIDs from NCDLs ancestors and obtained the function of canonical and/or non-canonical Dicers.

Furthermore, two regions showed high variation between the conserved sites were identified in canonical Dicer RIIIDs without proof of affection of catalytic ability: one conserved region in the N-terminal of S1/3 RIIIDs and the one clear insertion in the middle of S2/4 RIIIDs (Figure 3, green boxes). There was a short pattern of high converseness presented in 28 S1/3 but not in the rests comparing to S2/4. This pattern was mostly in Ascomycetes, not in Basidiomycetes. The other region was a relatively long insertion presented in 45 S2/4 RIIIDs from 40 Ascomycetes and Basidiomycetes. This insertion were less conserve, and mostly in TYPE2 (14 out of 18) had insertion while 31 out of 99 S2 (TYPE1) showing a bias between protein types. Interestingly, none of the sequences obtained S13 pattern in first RIIIDs (S1/3) and second RIIID insertions (S2/4) at the same time (Supplementary Figure S2). Formerly reported that Dicer sequence variation was related with function that a long insertion in Human Drosha RIIIDa that worked as a “ruler” comparing bacterial RIIIDs [45]. Such insertion was not the same insertion we described in here, may duel to that fungal RIIIDs were similar to human Drosha RIIIDa than bacterial RIIIDs and Drosha did not exist in fungi or plant.

Phylogenetic analysis on RBD domain

There were two types of dsRNA binding domains among all RNase III domain-containing proteins: IPR005034 which was always in front of RIIIDs (assigned as A in Figure 4A) and the ones following RIIIDs were IPR014720 (Assigned as B in Figure 5A). Phylogenetic trees of IPR005034 and IPR014720 were respectively constructed in order to find their relationships among Dicers and NCDLs. Among RBDs preceding RIIIDs (IPR005034) in TYPE1-2 and TYPE5, most A2s (10 out of 18) gathered in one subclade and the rests scattered among A1s with four A2 RBDs formed a tiny subclade (Figure 4B). Only one A3 mixed with A2, while the rests gathered with A1s. In RBDs following the RIIIDs, B4s (RBDs in TYPE 4) were the most frequently found. All of B1s located at the outermost clade, while other RDBs (B2, B3, and B5) were interspersed among the B4 RBDs. As for B2s and B3s double RBDs in non-canonical Dicers, they were similar but not duplication, like the double RIIIDs in canonical Dicers. B5s from TYPE7 were mixed with in B4s and far away from B1 to B3 RBDs. Based on the topology of the C tree, there might be possibility either that B1 was the ancestral form of B1-3 or that B1 was very different from other RDBs considering they obtained by canonical Dicers (Figure 4C).

Structural comparison of non-canonical Dicers

In order to find structural converseness of four non-canonical Dicers, their protein sequences were subjected to SWISS-MODEL (<https://swissmodel.expasy.org/interactive>) for 3D-structures [46] and then visualized by PyMol [47]. Each RIIID and the adjacent dsRBD were selected for prediction, and the best model of highest GMQE with acceptable QMEAN (no less than -4.0) were picked. Results showed conserved structures among the predicted regions by comparing their structure pairwise (Figure 5A), most of the helices shared by two or more proteins when superimposed together (Figure 5B). In *Pichia stipitis* genome which lacked of Argonaute and RdRP, putative Dicer shared structural conserveness with the other three characterized non-canonical Dicers. For the great importance of *P. stipitis* in industrial aspects, more studies deciphering its RNAi mechanism may be required. In addition, the 3D structure of Rnt1 (a TYPE4 protein) in *S. cerevisiae* was also modeled by the same approach, showing a largest overlap with

Dcr1 in *K. polysporus* and *P. stipitis* (Figure 5C). It suggested that TYPE4 might be an intermediate or building block for a non-canonical Dicer.

A conserved motif at the C-terminus of canonical Dicers

In *Schizosaccharomyces pombe*, a C-terminal domain including a zinc-binding motif was responsible for nuclear retention of Dcr1, a canonical Dicer [48]. Such CHCC motifs were identified from the C-terminal region following the second RNase III domain of canonical Dicers and NCDLs in different proportion. Among the 116 canonical Dicers, CHCC motif presented in 58 out of 98 TYPE1 and 4 out of 18 TYPE2 canonical Dicers. Meanwhile, four TYPE6, one TYPE4 and one TYPE5 were found with CHCC motif as well. A number of pathogenic fungi to animals or plants had Dicer(s) with CHCC motif at the C-terminus, as previously reported. Here we found this motif existed in all fungal species with canonical Dicers except for Oomycetes and Bacteria, involved saprotrophic fungi like *Neosartorya fischeri*, *Neurospora crassa*, *Trichoderma virens*, *S. octosporus*, and *Podospora anserina* (Supplementary Table S1).

Both Dicers in *Neurospora crassa* obtained different features: NCU06766T0 with insertion in S4 RIIID and NCU08270T0 with S13 conserved pattern in S1 RIIID and CHCC motif. *Magnaporthe oryzae* Dicers had the very same situation of different features: DCL2 or MDL-2 that produce siRNA [43] had insertion in the second RIIID while DCL1 possessed S13 pattern and CHCC at the same time.

Interestingly, there was an incompatible relationship between CHCC motif and the S2/4 insertion: the sequences with CHCC motifs did not possessed insertions in S2/4 RIIIDs or vice versa. The 68 sequences with CHCC motif largely overlapped with 28 inserted S1/3 RIIIDs, for 26 RIIIDs obtained both features. Above all, there were seven canonical Dicers obtained none of the features mentioned here, and all of them were TYPE1 proteins (Supplementary Figure S2).

Evolution of RNase III domain-containing proteins

Considering a parsimonious way of evolution, we assumed that the earliest ancestors were of a simple structure with single RIIID, or TYPE 8 type proteins. Based on the phylogenetic tree of RIIIDs, S12 or TYPE8 might have diverged into the other types following the respective course of evolution. As acquisition of additional domain(s), some of RNase III domain-containing proteins gained their relative functions and eventually evolved into Dicers. The steps of acquisition and function might not happen at random. Furthermore, Dicers (canonical and non-canonical ones) and Dicer-like proteins should have their own ancestors separately. We speculated the direct ancestor of canonical Dicers should be TYPE4 like proteins.

From Bacteria kingdom, TYPE8-like ancestor firstly got one RNA-binding domain and became Ribonuclease, like E.coli Ribonuclease and yeast Rnt1 that not necessarily RNA silencing function but dsRNA processing. TYPE4 proteins remained in all Bacterial genomes instead of TYPE8, while most eukaryotes genomes processed both TYPE4 and TYPE8 proteins. For unclear reasons, they are very necessary for only very rare cases missed both types at the same time. For instance, *Puccinia graminis* possessed one TYPE6 as the only NCDL alongside six TYPE1 Dicers; meanwhile *Nosema ceranae* only

obtained one TYPE1 without NCDLs. TYPE4 outside Bacteria diverged a little bit in their architecture. The distance between RBD and RIIIDs in bacteria were very shortly or merely next to each other, and later increased and became more common in fungi. Such distances was not strict, but might reflect a certain tendency of how Dicers evolved.

By its look, TYPE3 proteins seemed like coming from TYPE4 by gaining an extra RBD (Figure 6A), and only existed in the budding yeast species. Taking together the existing of non-canonical Dicers and Argonautes in some Saccharomycotina genomes, TYPE3 proteins here possibly work as substitution of canonical Dicers thusly Saccharomycotina may obtain partial RNA silencing ability.

In eukaryotes, two signatures marked the differences between canonical Dicer and NCDLs: double RNase III domains and Dicer dimerisation domain. If we suppose the additional RIIID was added in two possible steps depending on the order of gaining domains: firstly evolving into double a RIIIDs structure either like TYPE6 (Figure 6B), or through TYPE5 by gaining Dicer dimerisation domain first (Figure 6C). The RIIID tree indicated most S8 and S9 RIIIDs were already very similar to S1/3 and S2/4 respectively. Interestingly, the S10 RIIIDs in TYPE5 were very similar to S12 RIIID while their Dicer dimerisation domains were similar to the ones in canonical Dicers. So far, TYPE5 became an important middle-type during canonical Dicers evolution. Till these stages, neither TYPE5 nor 6 proteins became functional and left in the genomes as very rare cases. Another supported was that S7 from TYPE5 were all closer to S6 in TYPE4 than RIIIDs in canonical Dicers indicating RIIID duplication. TYPE5 proteins were relatively rare, only 13 sequences somehow in a variety of Taxa including one Oomycete, one Zygomycete, two Pezizomycetes, three Pucciniomycetes and six Agaricomycetes. After that, NCDLs gained the second RIIIDs and then RBSs became canonical Dicers. There was no middle type of one RIIID, one Dicer dimerisation domain and one following RBD, really suggested that TYPE2 structure should came from TYPE1 in higher kingdom of plant and animal (Figure 6, red box). It is still not solid to determent that canonical Dicers and NCDLs evolved through which way with current proves. Never the less, by logical, TYPE4 to TYPE6 direction would be easier to explain the evolution paths of all types. There were other protein types found in plant and animal that could also be included in these middle types and paths.

In conclusion, we identified classified the RNase III-containing proteins in fungi by domain structures. The different types reflected their function as well as their occurrence among the taxa. There were regions of sequence variations between two RIIIDs of canonical Dicers correlated with CHCC motif. By RIIID and RBD relationships, possible evolution paths included all protein types were indicated. From the simplest form (TYPE8) there were two distinguish directions: canonical Dicers (TYPE1, and TYPE2), non-canonical Dicers (TYPE3) and Dicer-like proteins (TYPE4 to 8). They consistently accumulated the essential domains by different strategy in each direction. And different directions have different evolution outcomes with different functions, and build up the complex and diverse RNase III family including canonical Dicer.

Material And Method

Collection of characterized sequences and proteome

Canonical Dicer protein sequences were collected from funRNA database (<http://funrna.riceblast.snu.ac.kr/>). In order to determine the domain profile of the characterized non-canonical Dicers by InterPro scan [49], three characterized protein sequences were retrieved from NCBI protein database, of which are from three yeasts including *Saccharomyces castellii*, *Kluyveromyces polysporus* and *Candida albicans* [32, 34, 41].

Non-canonical Dicer and Non-canonical Dicer-like proteins

Domain profile of the non-canonical Dicers was defined by InterPro scan, showing one RNase III domain and two trailing dsRNA binding domains (Figure 1A). Protein sequences of canonical Dicers were retrieved from funRNA database accordingly [19], as well as other RNase III-containing proteins. In addition, RNase III domain-containing proteins that do not belong to canonical and non-canonical Dicers were referred to as NCDL proteins.

Data set

To identify non-canonical Dicers and NCDLs, all RNase III domain-containing proteins were retrieved from 83 proteomes covering 12 bacteria, 3 Oomycetes and 68 fungi, including 45 Ascomycetes and 20 Basidiomycetes. A total of five non-canonical Dicers and 427 non-canonical Dicers-like proteins were found, together with 116 canonical Dicers were retrieved from funRNA database (Supplementary Table S1).

Multiple sequence alignment, phylogenetic tree construction

All RNase III-containing sequences were assigned as TYPE1 to 8 according to their domain structures, and sequences of RNase III domains were assigned as S1-12 in order to distinguish each domain among the different types of proteins as well (Figure 2A). Besides, dsRBDs were denoted as A and B according: Dicer dimerisation domains as A; and double-stranded RNA-binding domain (IPR014720) as B. Multiple sequence alignment and the phylogenetic trees construction by neighbor-joining method was performed by MUSCLE [50]. Species phylogeny was constructed by using CVTree (v4.2.1; source code distribution) [51]. The whole proteome sequences were used as input and K-tuple length was set to be seven.

Discussion

RNase III containing proteins, canonical Dicers or non-canonical Dicer-like proteins were much more common than we expected. This phenomenon indirectly confirmed their diverse but important physical roles. In general, there is very few species (merely six among all) completely lost canonical Dicers in eukaryotes. Furthermore, none of the species in our data set completely lost RNase III-containing proteins (Dicers and NCDLs) indicating the existence of none canonical Dicer-like proteins should not be ignored.

So far, there isn't sufficient information to determine the whole picture of all RIIID containing proteins, the abundance of protein types in fungi makes it possible to discuss by two functional directions. The NCDLs seemed apparently have simpler structures and completely different functions comparing to canonical ones. Their RIIID sequences suggested different directions and paths from canonical Dicers. The TYPE4 and 8 proteins are commonly found and functional in some genomes, but it is not the focus of the story how RNase III containing family evolved. As for other NCDLs so far, they are quite rare and lack of any RNAi related functions, they were regarded as incomplete proteins or traces of the ancient types that left in the genomes during the evolution. It has been long believed that the yeast genomes have lost the canonical RNAi machinery. Among the four yeast species: *C. albicans*, *K. polysporus*, *S. castellii* and *P. stipites* encoded with non-canonical Dicers, three of them (except for *P. stipites*) obtained single copy of Argonaute in their genomes at the same time. The non-canonical Dicer-like proteins have divergent function and evolution paths as well. If non-canonical Dicers (TYPE3) work as substitution of canonical Dicer, then it is reasonable to speculate that at least these three yeasts have the potential of partial RNAi by a non-canonical pathway or other RNA related functions.

The structures of RNase III containing proteins are diverse and complicate. Taking canonical Dicers for instance, Dicer proteins with the same function still differentiate in their essential domains. Although canonical Dicers in fungi have fewer domains than those in higher eukaryotes; they still have a complex structure that should evolve in steps. There are several possible ways to achieve the canonical structure. Without transitional canonical Dicer, we studied the protein structure constructed by the important domains. In this study, we found that compare of full length Dicer protein sequences were very complicated and messy, and the relationship of all RNase III domains instate would be easier to reflect the true relationship of the whole proteins. Since the structure of double RNase III in canonical Dicers distinguished them from the non-canonical Dicers, the key problem of canonical Dicer evolution is to answer how such structure was formed. Although we have approved that the first and second RNase III domains are not simple replication of each other, a protein with double RNase III domains structure that works as transitional type is still undetected. Furthermore, whether such structure obtained RNA related functions to some extent long already will largely support the double RIIIDs evolution pathway. The evolution pathways suggested here is only one of the many possibilities, and the true nature of double RNase III domains still needs more proves in the future.

Conclusions

This study we summarized all RNase III containing proteins in fungal kingdom, and showed the complex RNase III family obtained mainly by fungi based on function and domain structures. Regardless of canonical Dicers and Dicer-like proteins, most Eukaryotes obtained multiple RNase III protein types at the same time. We identified RNase III domains of canonical Dicers with different insertions in the first and second RNase III domains, while the RNA-binding domains in RNase III protein family were very similar to each other. Dicers (canonical and non-canonical) and NCDL proteins might have evolved from separate ancestors through different directions. Our speculations of evolution paths include all RNase-III

containing protein family might help to elucidate how the different functions of RNase-III containing protein have evolved, and further to explain the RNAi and some other relevant function in fungi.

Abbreviations

RIIID: RNase III domains; dsRDB: RNA-binding domains; PAZ: Piwi-Argonaute-Zwille; dsRNA: double-stranded RNA; dsRBD: dsRNA-binding domain; RdRp: RNA-dependent RNA polymerase; NCDL: non-canonical Dicer-like proteins.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All sequences data generated or analyzed during this study are included in supplementary table.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Finland Distinguished Professor Programme (FiDiPro # 138116) from Academy of Finland, the National Research Foundation of Korea grant funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A2A1A10051434) and the Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01115401), Rural Development Administration, Republic of Korea.

Authors' contributions

WJ and CJ retrieved the non-canonical Dicer-like protein sequences, and CJ constructed the database. WJ applied the data analysis and interpretation, and wrote the manuscript. LYH and CJ contribute to the drafting of the manuscript. LYH and FOA conceived the study and contributed to the experimental design.

Acknowledgements

Not applicable.

References

1. Hammond SM, Caudy AA, Hannon GJ. Post-transcriptional gene silencing by double-stranded RNA. *Nat Rev Genet.* 2001;2(2):110-119.
2. Tijsterman M, Plasterk RH. Dicers at RISC; the mechanism of RNAi. *Cell.* 2004;117(1):1-3.
3. Court DL, Gan J, Liang YH, Shaw GX, Tropea JE, Costantino N, et al. RNase III: Genetics and function; structure and mechanism. *Annual review of genetics.* 2013;47:405-431.
4. Billmyre RB, Calo S, Feretzaki M, Wang X, Heitman J. RNAi function, diversity, and loss in the fungal kingdom. *Chromosome Res.* 2013;21(6-7):561-572.
5. Torres-Martinez S, Ruiz-Vazquez RM. The RNAi Universe in Fungi: A Varied Landscape of Small RNAs and Biological Functions. *Annu Rev Microbiol.* 2017;71:371-391.
6. Chang SS, Zhang Z, Liu Y. RNA interference pathways in fungi: mechanisms and functions. *Annual review of microbiology.* 2012; 66:305-323.
7. Shiu PKT, Metzenberg RL. Meiotic silencing by unpaired DNA: Properties, regulation and suppression. *Genetics.* 2002;161(4):1483-1495.
8. Cogoni C, Macino G. Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. *Proc Natl Acad Sci USA.* 1997;94(19):10233-10238.
9. Murata T, Kadotani N, Yamaguchi M, Tosa Y, Mayama S, Nakayashiki H. siRNA-dependent and -independent post-transcriptional cosuppression of the LTR-retrotransposon MAGGY in the phytopathogenic fungus *Magnaporthe oryzae*. *Nucleic Acids Res.* 2007;35(18):5987-5994.
10. Nunes CC, Gowda M, Sailsbery J, Xue M, Chen F, Brown DE, et al. Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. *BMC Genomics.* 2011;12:288.
11. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science.* 2004;303(5658):672-676.
12. Bernstein DA, Vyas VK, Fink GR. Genes come and go: the evolutionarily plastic path of budding yeast RNase III enzymes. *RNA Biol.* 2012;9(9):1123-1128.
13. Nicolas FE, Ruiz-Vazquez RM. Functional diversity of RNAi-associated sRNAs in fungi. *Int J Mol Sci.* 2013;14(8):15348-15360.
14. Aguado LC, tenOever BR. RNase III Nucleases and the Evolution of Antiviral Systems. *Bioessays.* 2018;40(2):1700173.
15. Lamontagne B, Larose S, Boulanger J, Elela SA. The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Current issues in molecular biology.* 2001;3(4):71-8.
16. Nicholson AW. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA.* 2014;5(1):31-48.

17. Song MS, Rossi JJ. Molecular mechanisms of Dicer: endonuclease and enzymatic activity. *The Biochemical journal*. 2017;474(10):1603-1618.
18. Gao Z, Wang M, Blair D, Zheng Y, Dou Y. Phylogenetic analysis of the endoribonuclease Dicer family. *PLoS One*. 2014;9(4):e95350.
19. Choi J, Kim KT, Jeon J, Wu J, Song H, Asiegbu FO, et al. funRNA: a fungi-centered genomics platform for genes encoding key components of RNAi. *BMC Genomics*. 2014;15(9): 1-10 Suppl 9:S14.
20. Francia S, Michelini F, Saxena A, Tang D, de Hoon M, Anelli V, et al. Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*. 2012;488(7410):231-235.
21. Kurzynska-Kokorniak A, Koralewska N, Pokornowska M, Urbanowicz A, Tworak A, Mickiewicz A, et al. The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. *Nucleic Acids Res*. 2015;43(9):4365-4380.
22. Conrad C, Rauhut R. Ribonuclease III: new sense from nuisance. *Int J Biochem Cell Biol*. 2002;134(2):116-129.
23. Kim YK, Kim B, Kim VN. Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(13):E1881-1889.
24. Lim B, Sim M, Lee H, Hyun S, Lee Y, Hahn Y, et al. Regulation of Escherichia coli RNase III activity. *Journal of microbiology (Seoul, Korea)*. 2015;53(8):487-494.
25. Redko Y, Bechhofer DH, Condon C. Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Molecular microbiology*. 2008;68(5):1096-1106.
26. Grzechnik P, Szczepaniak SA, Dhir S, Pastucha A, Parslow H, Matuszek Z, et al. Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nat Commun*. 2018;9(1):1783.
27. Chanfreau G, Rotondo G, Legrain P, Jacquier A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *The EMBO journal*. 1998;17(13):3726-3737.
28. Iino Y, Sugimoto A, Yamamoto M. *S. pombe* pac1+, whose overexpression inhibits sexual development, encodes a ribonuclease III-like RNase. *The EMBO journal*. 1991;10(1):221-226.
29. Johanson TM, Lew AM, Chong MM. MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *Open Biol*. 2013;3(10):130144.
30. Hu Y, Stenlid J, Elfstrand M, Olson A. Evolution of RNA interference proteins dicer and argonaute in Basidiomycota. *Mycologia*. 2013;105(6):1489-1498.
31. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infect Genet Evol*. 2018;65:329-332.
32. Weinberg DE, Nakanishi K, Patel DJ, Bartel DP. The inside-out mechanism of Dicers from budding yeasts. *Cell*. 2011;146(2):262-276.
33. Bernstein DA, Vyas VK, Weinberg DE, Drinnenberg IA, Bartel DP, Fink GR. *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc Natl Acad Sci U*

- S A. 2012;09(2):523-528.
34. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, et al. RNAi in budding yeast. *Science*. 2009;326(5952):544-550.
 35. Liang YH, Lavoie M, Comeau MA, Abou Elela S, Ji X. Structure of a eukaryotic RNase III postcleavage complex reveals a double-ruler mechanism for substrate selection. *Molecular cell*. 2014;54(3):431-444.
 36. Lall S. Silence of the budding yeast. *Nat Struct Mol Biol*. 2009;16(10):1008.
 37. Yu X, Li X, Zheng L, Ma J, Gan J. Structural and functional studies of a noncanonical Dicer from *Entamoeba histolytica*. *Sci Rep*. 2017;7:44832.
 38. Mueller AK, Hammerschmidt-Kamper C, Kaiser A. RNAi in Plasmodium. *Current pharmaceutical design*. 2014;20(2):278-283.
 39. Pong SK, Gullerova M. Noncanonical functions of microRNA pathway enzymes - Drosha, DGCR8, Dicer and Ago proteins. *FEBS Lett*. 2018;592(17):2973-2986.
 40. Burger K, Gullerova M. Swiss army knives: non-canonical functions of nuclear Drosha and Dicer. *Nature reviews Molecular cell biology*. 2015;16(7):417-430.
 41. Bernstein DA, Vyas VK, Weinberg DE, Drinnenberg IA, Bartel DP, Fink GR. *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(2):523-528.
 42. Catalanotto C, Pallotta M, ReFalo P, Sachs MS, Vayssie L, et al. Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in *Neurospora crassa*. *Mol Cell Biol*. 2004;24(6):2536-2545.
 43. Kadotani N, Nakayashiki H, Tosa Y, Mayama S. One of the two Dicer-like proteins in the filamentous fungi *Magnaporthe oryzae* genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. *J Biol Chem*. 2004; 279(43):44467-44474.
 44. Kadotani N, Murata T, Quoc NB, Adachi Y, Nakayashiki H. Transcriptional control and protein specialization have roles in the functional diversification of two dicer-like proteins in *Magnaporthe oryzae*. *Genetics*. 2008; 180(2):1245-1249.
 45. Kwon SC, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, et al. Structure of Human DROSHA. *Cell*. 2016; 164(1-2):81-90.
 46. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018; 46(W1):W296-W303.
 47. DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*. 2002; 40: 82-92.
 48. Barraud P, Emmerth S, Shimada Y, Hotz HR, Allain FH, Buhler M. An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO J*. 2011; 30(20):4223-4235.

49. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 2012; 40(Database issue):D306-312.
50. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32(5):1792-1797.
51. Xu Z, Hao BL. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. *Nucleic Acids Research.* 2009; 37:W174-W178.
52. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14(6):1188-1190.

Figures

Figure 1

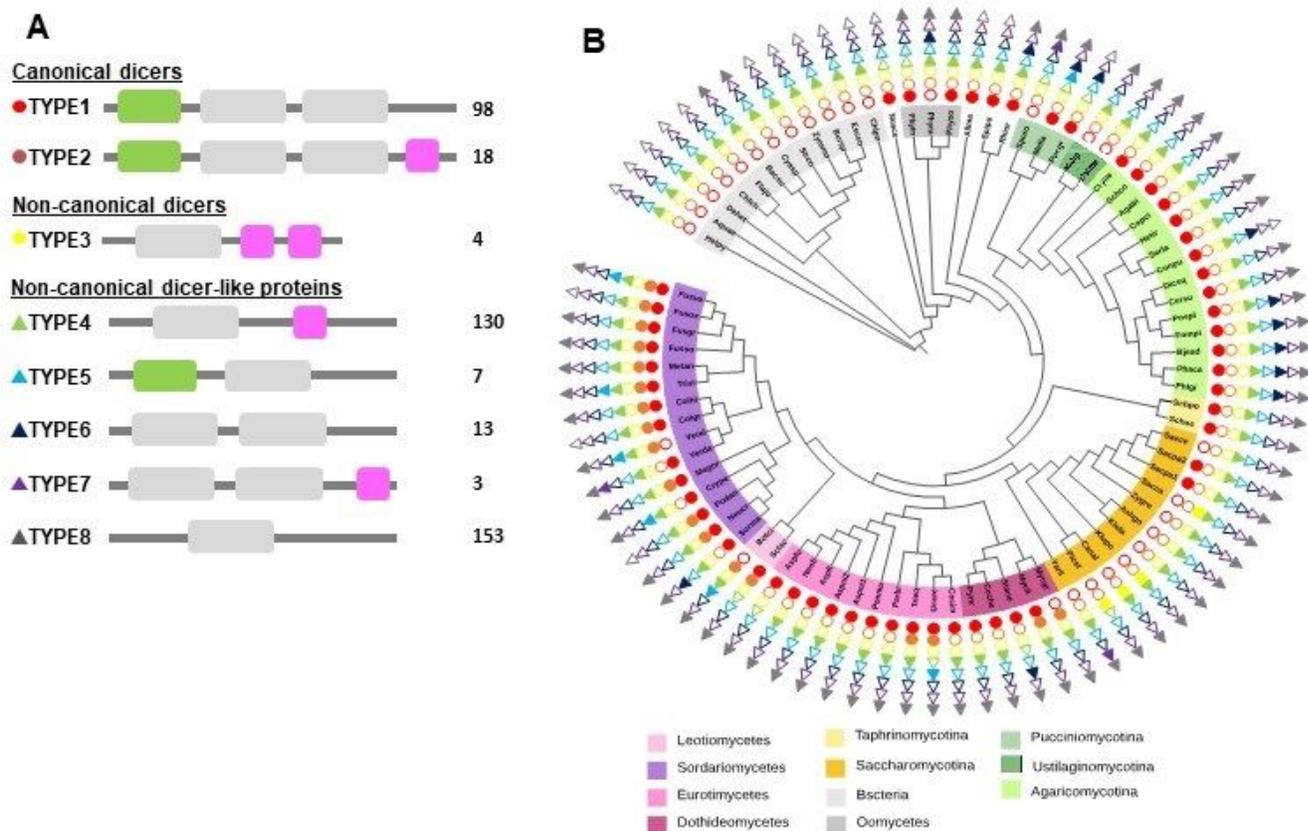


Figure 1

Taxonomic distribution of putative Dicers. (A) Domain structures of bacterial and fungal putative Dicer proteins. Gray boxes for RNase III domain, green for Dicer Dimerisation domain (IPR005034), and pink for other dsRNA-binding domains (IPR014720). Diagrams are not drawn to scale. (B) Distribution of the six

types of Dicer proteins across 83 proteomes, including 68 fungi, 3 Oomycetes, and 12 bacteria. Each circle and triangle indicates different type of proteins, filled sign for presence of the corresponding type and empty for absence. Abbreviations for species names are as follows (in alphabetical order): Agabi (*Agaricus bisporus*), Allma (*Allomyces macrogynus*), Aquae (*Aquifex aeolicus* VF5), Ashgo (*Ashbya gossypii* ATCC10895), Aspfl (*Aspergillus flavus*), Aspfu (*A. fumigatus* Af293), Aspni1 (*A. nidulans*), Aspni2 (*A. niger* CBS 513.88), Bacsu (*Bacillus subtilis* subsp. *subtilis* str. 168), Bjead (*Bjerkandera adusta*), Botci (*Botrytis cinerea*), Bursp (*Burkholderia* sp. CCGE1002), Canal (*Candida albicans*), Cersu (*Ceriporiopsis subvermisporea* B), Chlch (*Chlorobium chlorochromatii* CaD3), Chlpn (*Chlamydomophila pneumoniae* LPCoLN), Coche (*Cochliobolus heterostrophus* C5), Cocim (*Coccidioides immitis* RS), Colgr (*Colletotrichum graminicola* M1.001), Colhi (*C. higginsianum*), Conpu (*Coniophora puteana*), Copci (*Coprinus cinereus*), Cryne (*Cryptococcus neoformans* var. *grubii* H99), Crypa (*Cryphonectria parasitica*), Cyasp (*Cyanothece* sp. ATCC 51142), Dehet (*Dehalococcoides ethenogenes* 195), Dicsq (*Dichomitus squalens*), Escco (*Escherichia coli* K12 DH10B), Flajo (*Flavobacterium johnsoniae* UW101), Fompi (*Fomitopsis pinicola* FP-58527 SS1), Fusgr (*Fusarium graminearum*), Fusox (*F. oxysporum*), Fusso (*F. solani*), Fusve (*F. verticillioides*), Helpy (*Helicobacter pylori* 51), Hetir (*Heterobasidion irregulare* TC 32-1), Klula (*Kluyveromyces lactis*), Klupo (*K. polysporus*), Magor (*Magnaporthe oryzae* 70-15), Malgl (*Malassezia globosa*), Mella (*Melampsora laricis-populina*), Metan (*Metarhizium anisopliae* var. *acidum*), Mycfi (*Mycosphaerella fijiensis*), Mycgr (*M. graminicola*), Neofi (*Neosartorya fischeri*), Neucr (*Neurospora crassa*), Nosce (*Nosema ceranae*), Parbr (*Paracoccidioides brasiliensis* Pb01), Penma (*Penicillium marneffei*), Phaca (*Phanerochaete carnosae* HHB-10118-Sp), Phlgi (*Phlebiopsis gigantea*), Phyin (*Phytophthora infestans*), Phyra (*P. ramorum*), Physo (*P. sojae*), Picst (*Pichia stipitis*), Podan (*Podospora anserina*), Pospl (*Postia placenta*), Pucgr (*Puccinia graminis*), Pyrtr (*Pyrenophora tritici-repentis*), Rhior (*Rhizopus oryzae*), Sacca (*Saccharomyces castellii*), Sacce (*S. cerevisiae* S288C), Sacpa1 (*S. paradoxus* A12), Sacpa2 (*S. pastorianus* Weihenstephan 34/70), Schco (*Schizophyllum commune*), Schoc (*Schizosaccharomyces octosporus*), Schpo (*S. pombe* 132), Scpsc (*Sclerotinia sclerotiorum*), Serla (*Serpula lacrymans* S7.9), Sorma (*Sordaria macrospora*), Spipu (*Spizellomyces punctatus* DAOM BR117), Sporo (*Sporobolomyces roseus*), Stano (*Stagonospora nodorum*), Strco (*Streptomyces coelicolor* A3 Trieq (*Trichophyton equinum*), Trivi (*Trichoderma virens*), Uncre (*Uncinocarpus reesii*), Ustma (*Ustilago maydis* 521), Veral (*Verticillium albo-atrum* VaMs. 102), Verda (*V. dahliae* VdLs. 17), Yarli (*Yarrowia lipolytica*), Zygro (*Zygosaccharomyces rouxii*) and Zymmo (*Zymomonas mobilis* subsp. *mobilis* NCIMB 11163).

Figure 2

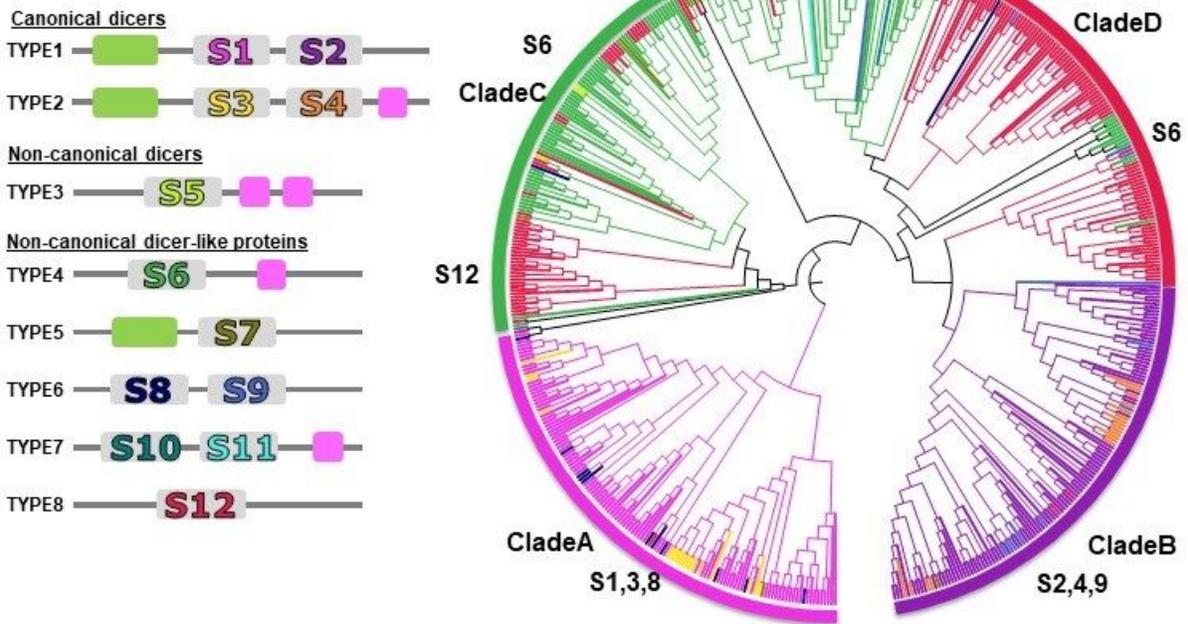


Figure 2

Phylogenetic tree for RNase III domains. (A) Every RNase III domain is assigned as S1 to S12. Colors for each RNase III domain are shown in the diagram. (B) Neighbor-joining tree of all RNase III domains.

Figure 3

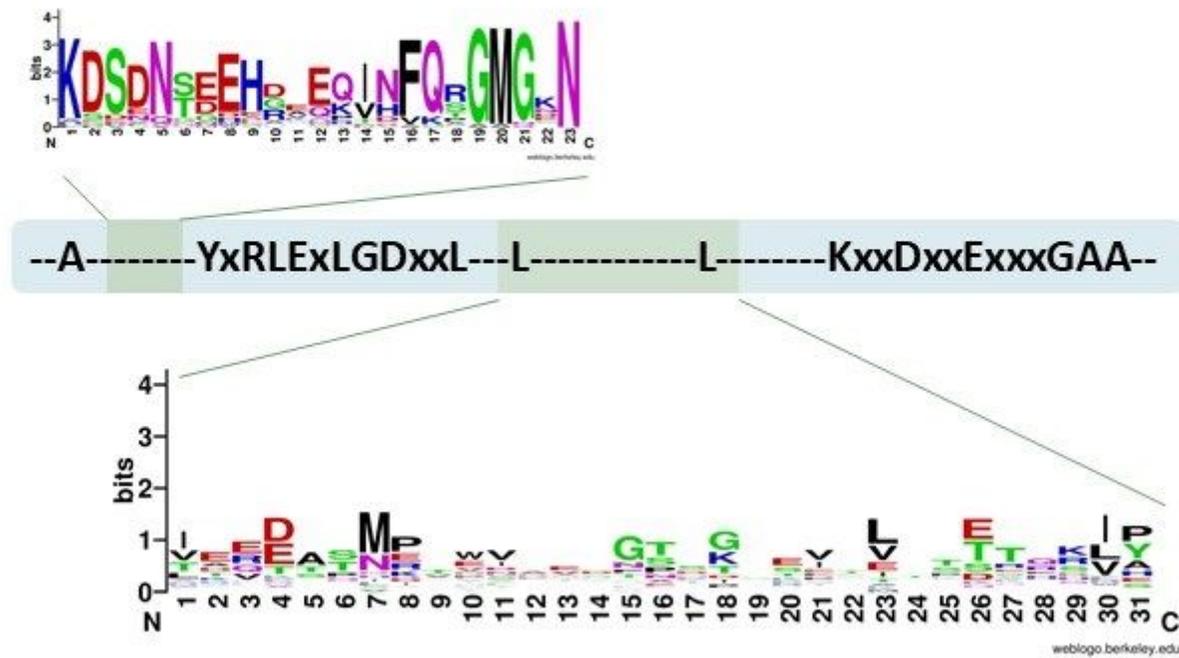


Figure 3

Sequences variations between the first and second RNase III Domains. S1, S3, are the first RNase III domains for TYPE1, TYPE2 canonical Dicers, S2, S4 are the second RNase III domains. Conserved site over 80% showed the regions of S13 pattern and S24 insertions in RNase III domains (indicated by green boxes), the sequences are showed by WEBLOGO [52].

Figure 4

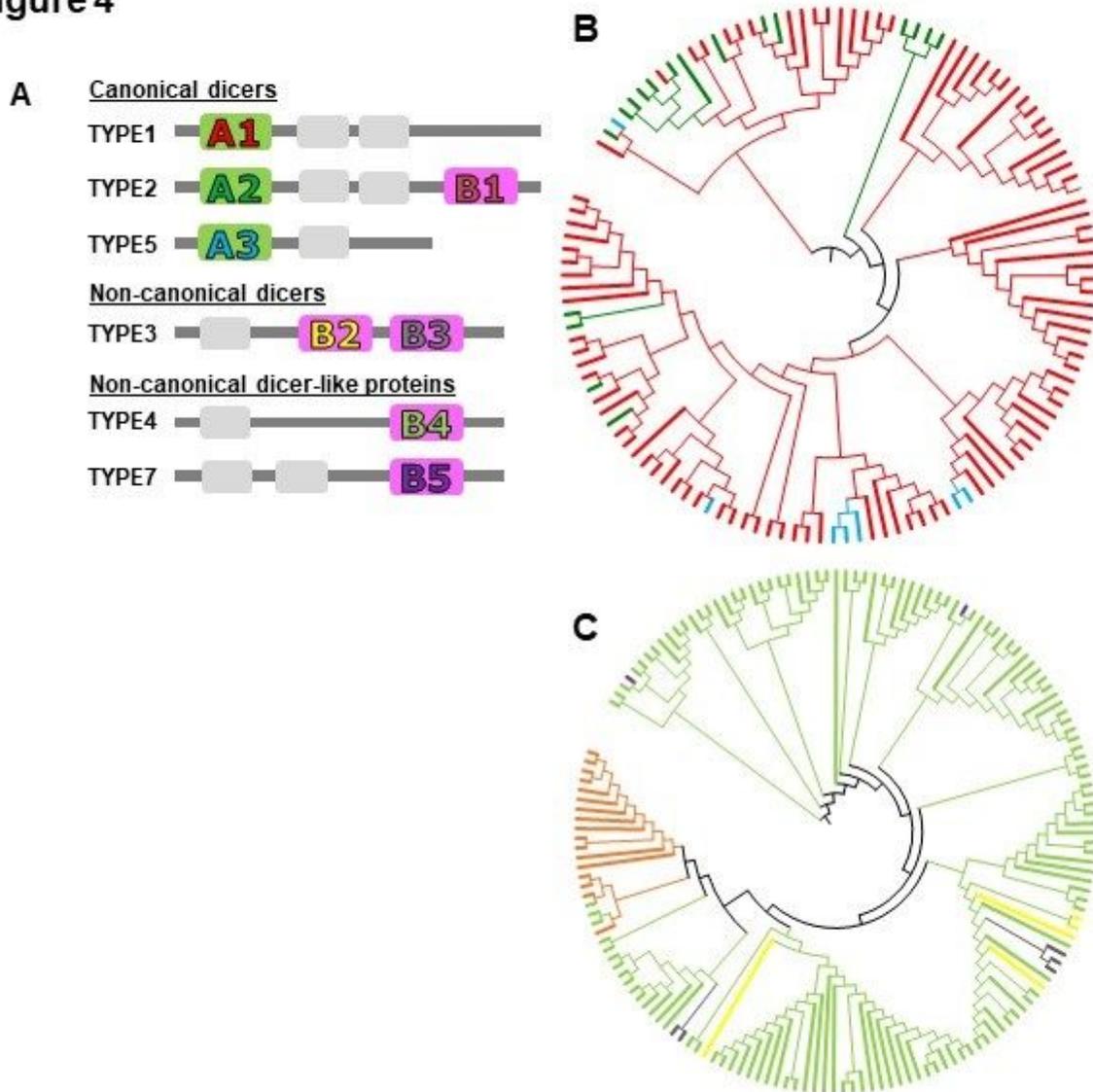


Figure 4

Phylogenetic trees for dsRNA-binding domains. (A) Color scheme is shown in the domain structure diagrams. (B) Neighbor-joining tree of dsRNA-binding domains (IPR005034), which is only found in canonical Dicers. (C) Neighbor-joining tree of dsRNA-binding domains (IPR014720).

Figure 5.

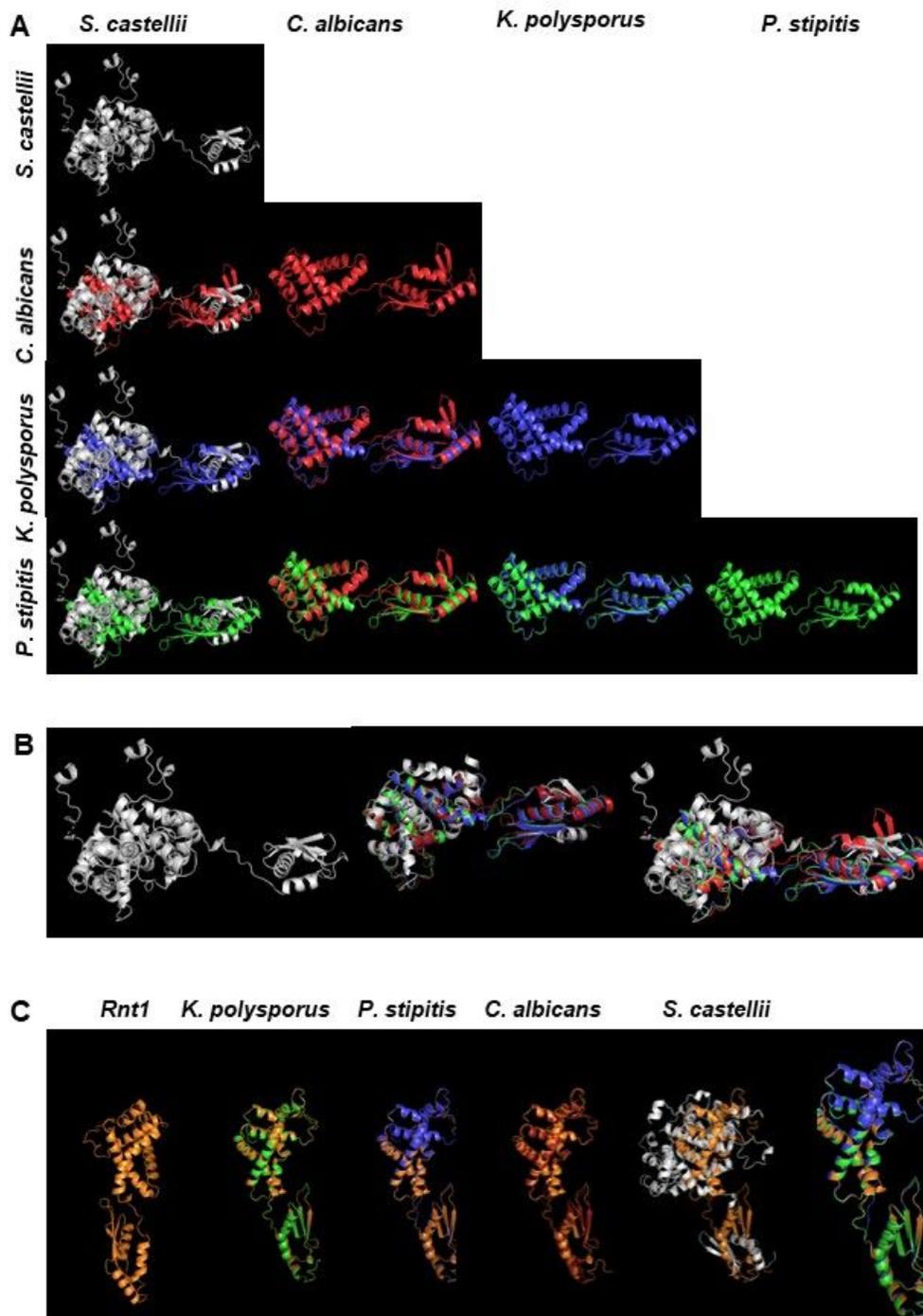


Figure 5

Structural alignments among non-canonical Dicers and between non-canonical Dicers and TYPE4 proteins. (A) Pairwise structural alignment of non-canonical dicers. For each column, the protein in the first row was used as template for comparing the others. (B) From the left, 3D structure of *S. castellii* non-canonical dicer, aligned regions from the four proteins, and full structural view of four proteins. (C)

Structural alignment between non-canonical dicers and Yeast Rnt1 proteins in which the species do not have any canonical/non-canonical dicers.

Figure 6.

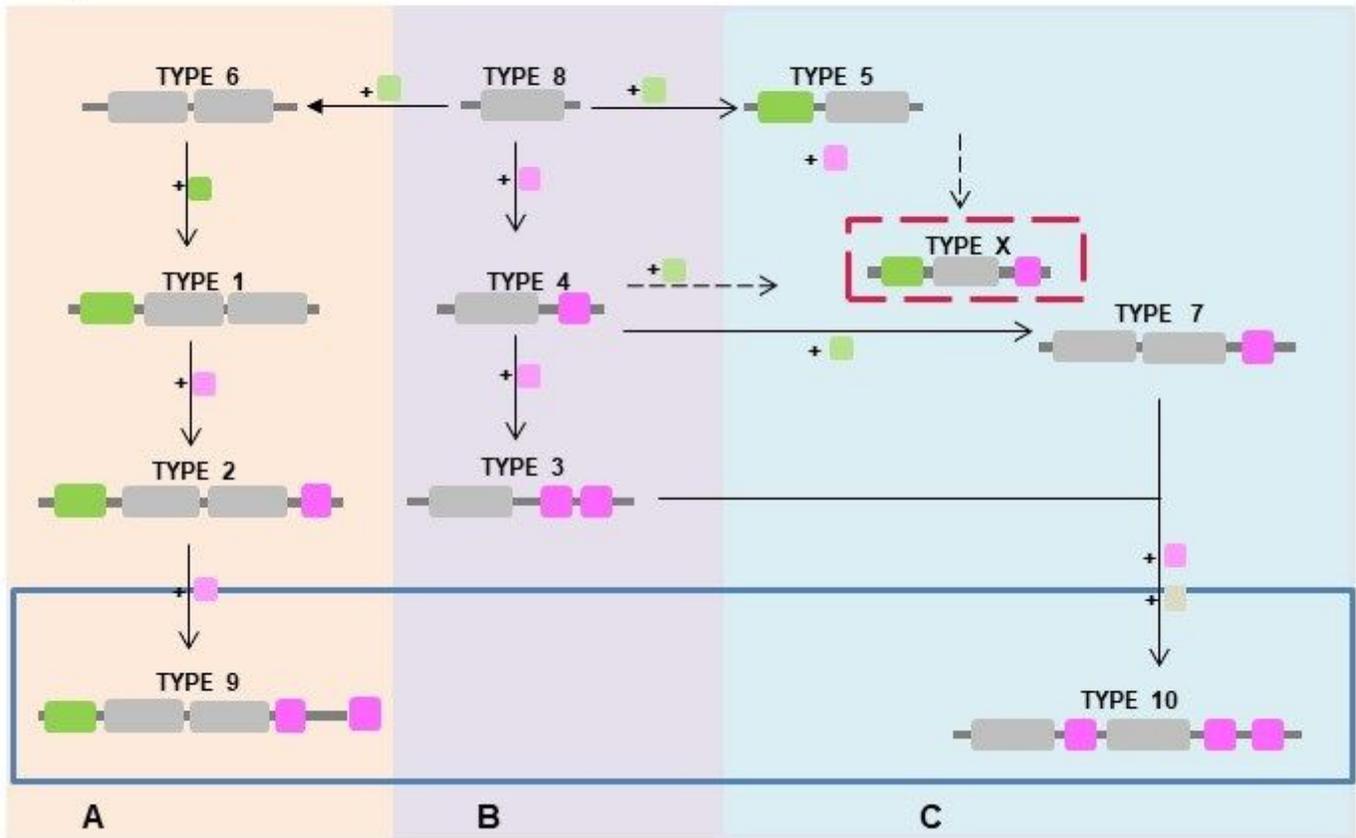


Figure 6

Possible evolution of Dicer family. (A) From Type6, proteins gaining RBD domains but did not gain Dicer function. These RNase III proteins are mostly found in animal and plant species but rare in fungi. (B) Type8 proteins became TYPE4 proteins by gaining RNA-binding domains, and TYPE3 with 2 RBD domains are non-canonical Dicers in budding yeast. TYPE8 and 4 are ubiquitous in eukaryotes, but not functional. (C) TYPE5 protein obtained Dicer Dimerisation domain and then another RNase III domain, finally became Type 1 canonical Dicers in fungi. From Type 1, they are canonical Dicers and differently distributes in kingdoms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DicerTABLESupplementary.xlsx](#)
- [SupplementaryFigure1.pptx](#)

- [SupplementaryFigure2.pptx](#)