DATABASE
The Journal of Biological Databases and Curation

AGODB: a comprehensive domain annotation database of argonaute proteins

Bowen Li[1,†], Shanshan Yang[1,†], Jinjin Long[1], Xue Chen[1], Qianyue Zhang[1], Lin Ning[2], Bifang He[1,*], Heng Chen[1,*] and Jian Huang[3,*]

[1]Medical College, Guizhou University, Guiyang, Guizhou 550025, China
[2]School of Health Care Technology, Chengdu Neusoft University, Chengdu, Sichuan 611844, China
[3]Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

[*]Corresponding author: Tel: +0851-88274016; Fax: +0851-83621956; Email: bfhe@gzu.edu.cn
Correspondence may also be addressed to Heng Chen. Tel: +0851-88274016; Fax: +0851-83621956; Email: hchen13@gzu.edu.cn and Jian Huang.
Tel: +86-28-6183-0670; Fax: +86-28-8320-8238; Email: hj@uestc.edu.cn
[†]These authors contributed equally to this work.

Abstract

Argonaute (Ago) proteins are widely expressed in almost all organisms. Eukaryotic Ago (eAgo) proteins bind small RNA guides forming RNA-induced silencing complex that silence gene expression, and prokaryotic Ago (pAgo) proteins defend against invading nucleic acids via binding small RNAs or DNAs. pAgo proteins have shown great potential as a candidate 'scissors' for gene editing. Protein domains are fundamental units of protein structure, function and evolution; however, the domains of Ago proteins are not well annotated/curated currently. Therefore, full functional domain annotation of Ago proteins is urgently needed for researchers to understand the function and mechanism of Ago proteins. Herein, we constructed the first comprehensive domain annotation database of Ago proteins (AGODB). The database curates detailed information of 1902 Ago proteins, including 1095 eAgos and 807 pAgos. Especially for long pAgo proteins, all six domains are annotated and curated. Gene Ontology (GO) enrichment analysis revealed that Ago genes in different species were enriched in the following GO terms: biological processes (BPs), molecular function and cellular compartment. GO enrichment analysis results were integrated into AGODB, which provided insights into the BP that Ago genes may participate in. AGODB also allows users to search the database with a variety of options and download the search results. We believe that the AGODB will be a useful resource for understanding the function and domain components of Ago proteins. This database is expected to cater to the needs of scientific community dedicated to the research of Ago proteins.

Database URL: http://i.uestc.edu.cn/agodb/

Introduction

Historically, the argonaute (Ago) protein family was discovered in a plant mutagenesis screen (1). Eukaryotic Ago (eAgo) proteins are grouped into four subfamilies: Ago-like, PIWI-like, WAGO and *Trypanosoma* Ago subfamily (2). We focused on the Ago-like subfamily in this study. eAgo proteins, as the core component of the RNA-induced silencing complex, have become a key player in RNA interference (RNAi) pathways (3). The eAgo proteins mainly form binary complexes with their guide RNAs and recognize complementary target mRNAs for subsequent site-specific cleavage or silencing, which then leads to the degradation and inhibition of protein translation (4). However, due to lack of RNAi pathways in archaea and bacteria, the physiological function of prokaryotic Ago (pAgo) proteins has remained elusive for a long time (5). In recent years, studies have gradually proved that pAgo proteins can bind small RNAs or DNAs to defend against foreign invasive genomes and showed the potential as candidate 'scissors' for gene editing (6–9).

eAgo proteins consist of four domains: N-terminal (N), P-element-induced wimpy testis (PIWI)–Ago–Zwille (PAZ), Middle (MID) and PIWI, along with two domain linkers: Linker 1 (L1) and Linker 2 (L2) (10). Each domain is involved in different steps of the protein's enzyme activity. The N-domain plays an important role in loading and unwinding the small RNA duplex, while the PAZ and MID domains provide binding pockets to anchor the 3′ and 5′ ends of the microRNA, respectively (11). The PIWI domain is structurally and functionally similar to RNase-H and contributes to endonuclease activity for the target strand (12). However, the endonuclease activity is not a property of all Ago proteins (13, 14). The L1 links the N and PAZ domains, and L2 links the PAZ and MID domains (15). Protein domains are critical in classifying proteins, understanding their biological functions, annotating their evolutionary mechanisms and protein design (16). They are considered to be homologous portions of sequences encoded in different gene contexts, kept intact by evolution at the sequence level (17). Therefore, comprehensive

domain annotation of Ago proteins will help researchers to understand the function and mechanism of Ago proteins.

According to the domain architecture, pAgo proteins can be divided into long pAgo proteins, short pAgo proteins and PIWI-RE proteins [PIWI with conserved R (Arg) and E (Glu) residues] (5). The overall domain information of most long pAgo proteins is essentially similar to eAgo proteins, retaining all four domains and two linkers, whereas short pAgo proteins only contain the MID and PIWI domains (18). PIWI-RE proteins with unknown functions appear in several major bacterial lineages, and their domain constituent is similar to short pAgo proteins with two conserved MID and PIWI domains (19). But unlike short pAgo proteins, PIWI-RE proteins have two conserved residues: arginine (R) and glutamate (E) in PIWI and MID domains, which are essential for nucleic acid-binding (4, 19). Currently, although the function and mechanism of eAgo and several long pAgo proteins have been studied in detail, little is known about those of short pAgo and PIWI-RE proteins (20–23).

Protein domains are fundamental units of protein structure, function and evolution (24). At the sequence level, domains are homologous segments in evolution, and at the structural level, domains are units that can fold and work independently (16). Proteins can usually be decomposed into domains according to their similar sequence or structural characteristics. The traditional method of domain annotation is to analyze the structure of Ago proteins by nuclear magnetic resonance, X crystal diffraction and cryo-electron microscopy three-dimensional reconstruction techniques and then divide the protein domains artificially. Using these methods for domain annotation is time-consuming and labor-intensive. Fortunately, Jiang *et al.* (25) proposed a computational method, called AGONOTES, for comprehensive domain annotation of Ago proteins. However, there is no specific database available for storing domain annotation data of Ago proteins currently. Existing public protein databases, such as the Universal Protein Resource (UniProt) (26), protein families database (Pfam) (27), simple modular architecture research tool (SMART) (28) and Conserved Domain Database (CDD) (29), enable the domain annotation of various proteins, whereas the domains of eAgo proteins have not been well annotated/curated, and only the relatively conservative PIWI and PAZ domains have been annotated and curated. Since the domains of pAgo proteins are not as conserved as that of eAgo proteins, the domain annotation of pAgo protein is more incomplete. Therefore, a comprehensive domain annotation database for Ago proteins is desperately needed.

Herein, we constructed the first comprehensive domain annotation database of Ago proteins (AGODB). The online resource not only provides complete domain annotation but also allows users to search the database with a variety of options and download the search results. In addition, an Ago protein prediction tool named AGOPredict (30) has also been integrated into AGODB to help users to discover novel Ago proteins. Users can access our database at http://i.uestc.edu.cn/agodb/ freely. The database will be of great significance for promoting the research of Ago proteins and related fields.

Materials and methods

Data source

We first searched UniProt with 'argonaute protein' as the keyword. We collected all Ago proteins that were manually reviewed and belong to Ago subfamily (namely Ago-like subfamily). Meanwhile, we manually curated the Ago proteins from the published literature. Ago proteins collected from UniProt and published literature with three-dimensional structures resolved experimentally were considered as experimentally validated Ago proteins. Then, we collected the general information corresponding to each protein from UniProt and National Center for Biotechnology Information (NCBI) (31) and the domain information corresponding to each protein from published literature. For those Ago proteins with no domain information in the literature, we used the domain annotation tool of Ago proteins named AGONOTES to make comprehensive domain prediction and collected the results into our database. To find more candidate Ago proteins, we then put the sequences of the experimentally verified Ago proteins into the blastp tool in BLAST 2.12.0+ for similarity search with default parameters (32). Based on the similarity search results, the protein was identified as a potential Ago protein when the identity was greater than 90%, the e-value was less than 1e – 5 and the coverage identity was greater than 85%. After confirming that the protein was not in the Ago protein data set already collected, we curated it into AGODB. To provide a non-redundant Ago protein data set, we utilized the CD-HIT tool to remove redundant sequences with a sequence identity cutoff of 80% (33). Users can download this data set at http://i.uestc.edu.cn/agodb/download.html.

GO and KEGG pathway analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and Gene Ontology (GO) functional annotation can preliminarily analyze the biological process (BP) or signal pathway that genes may participate in. We used KOBAS-intelligence (KOBAS-i) (34) to perform gene annotation and pathway enrichment analysis. Significant GO terms and pathways were identified after multiple testing adjustments using the false discovery rate (FDR) method; FDR-corrected *P*-value < 0.05 indicated a statistically significant difference.

Database implementation

The front end of the AGODB website was built with the languages HTML and PHP. MySQL, which is one of the most popular relational database management systems, was used for data storage and processing in the backend. The architecture of AGODB is shown in Figure 1. Browsing and searching panels were developed for visiting and retrieving data from the database efficiently. AGODB has been tested in three popular web browsers, including the Google Chrome, Internet Explorer and Mozilla Firefox browsers.

Tools integration

AGOPredict, which is a support vector machine-based predictor specifically for identifying Ago proteins based on sequence information (30), was integrated into AGODB. This bioinformatics tool will help users to discover more potential Ago proteins, and users can use it for Ago protein prediction by clicking 'AGOPREDICT' at the top of AGODB. In addition, we also provide links to AGONOTES and AGO3D on the homepage of AGODB to facilitate users to study Ago proteins. AGONOTES (25) is an online service dedicated to annotating the comprehensive domains of Ago proteins, and AGO3D (35)
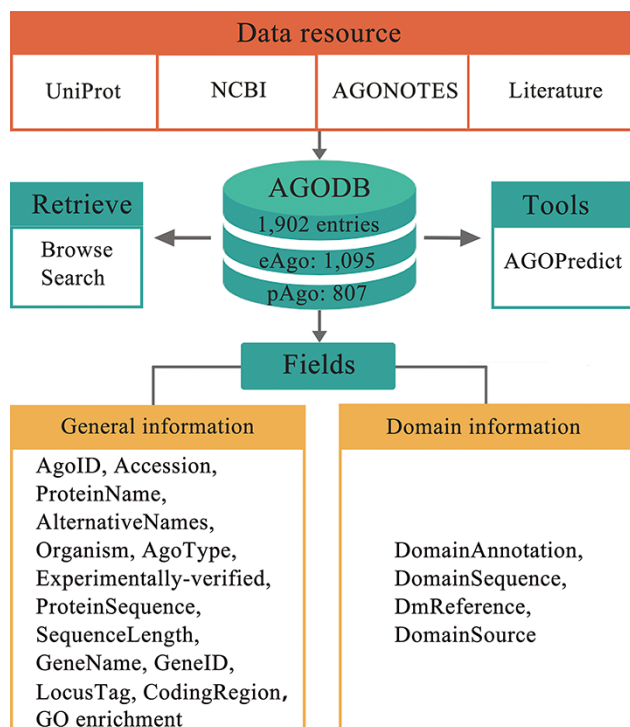
Figure 1. Architecture of AGODB. AGODB consists of 1902 entries, including 1095 eAgo proteins and 807 pAgo proteins. The relevant information of each entry in AGODB was collected from UniProt, NCBI, AGONOTES and published articles. The development of AGODB includes the design of fields, the development of the retrieval system and the integration of tool(s).

is an online service tailored for constructing three-dimensional structures of Ago proteins.

Results

Data content and summary

In AGODB, the collected information was divided into two modules: general information and domain annotation.

General information

Information on Ago proteins was manually obtained from UniProt or NCBI and stored as general information. Each entry included the following information: (i) AgoID—the unique number of the Ago protein in the database; (ii) Accession—the accession number of the Ago protein in NCBI, with a link to the NCBI database. If the accession number of the Ago protein cannot be linked to NCBI, then linked to UniProt; (iii) ProteinName—the name of the Ago protein; (iv) AlternativeNames—synonyms of the Ago protein name; (v) Organism—the name of organism that is the source of the protein sequence; (vi) AgoType—eAgo or pAgo; (vii) Experimentally verified—whether it has been experimentally verified; (viii) ProteinSequence—the amino acid sequence of the Ago protein named after the accession number, with a link to fasta format sequence in the NCBI protein database. If the accession number of the Ago protein cannot be linked to NCBI, then linked to UniProt. (ix) SequenceLength—the number of amino acids; (x) GeneName—the name of the

gene that codes for the protein sequence, with its synonyms; (xi) GeneID—a unique number for each gene, with a link to the NCBI gene database; (xii) LocusTag—the identifier in the genome; (xiii) CodingRegion—coding region; (xiv) GO enrichment—GO terms which provide insights into the BP that Ago genes may participate in.

Domain information

On the basis of general information, the important domain information of Ago proteins was collected. Domain information from the literature or annotated by AGONOTES was manually curated. This information consisted of four fields, including (i) DomainAnnotation: the domain annotation information of the Ago protein displayed in the form of a figure. (ii) DomainSequence: a download button to retrieve the complete sequence of each domain. For parts of Ago proteins, due to the scope of each domain from the literature is not clear, the domain sequence is not available. (iii) DmReference: reference for domain annotation. (iv) DomainSource: the source of the domain annotation information (AGONOTES or literature).

The current release of AGODB holds related information about Ago proteins, with a total of 1902 entries. As shown in Figure 2A, 70 Ago proteins are experimentally validated, and the others are putative Ago proteins that have yet to be
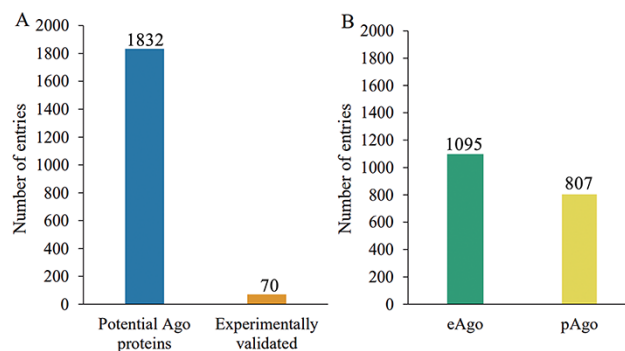

Figure 2. Schematic of Ago protein entries at various categories in the AGODB database. (A) Number of potential and experimentally validated Ago proteins. (B) Number of eAgo and pAgo proteins.
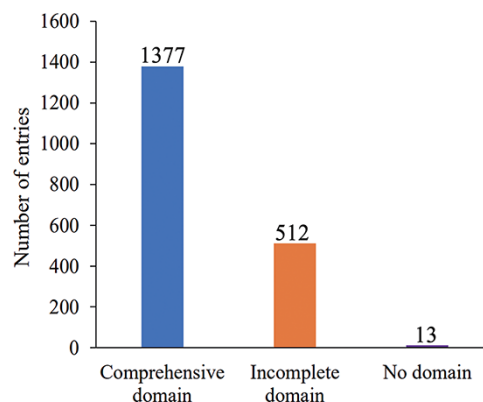

Figure 3. Number distribution of Ago protein entries according to domain annotation information integrity.

Figure 4. Browsing page of AGODB. (A) A brief browsing table displays fields, including AgoID, Accession, ProteinName, Organism, AgoType, Experimentally verified and links directing to the 'Detail' page. (B) The 'Detail' page shows detailed descriptions of each Ago protein.

further confirmed. As shown in Figure 2B, 1095 Ago proteins are derived from eukaryotes and 807 are derived from prokaryotes. As depicted in Figure 3, 1377 Ago proteins have complete domain annotation information, of which 1366 are annotated by AGONOTES and 11 are from published articles; 512 Ago proteins have incomplete domain information which was provided by Ryazansky *et al.* (18) and 13 Ago proteins had no domain information.

Figure 5. Search interface of AGODB. For example, if users want to retrieve all the experimentally validated eAgo proteins from mouse, they should first select the search field 'Organism' from the pull-down list box, input 'mouse' into the corresponding blank text form and select the 'and' button. Then, users should select the 'Experimentally-verified' search field, input 'Yes' and select the 'and' button. Finally, they should select the 'AgoType' field, input 'eAgo' and press the 'Submit Query' button.

Data browsing and searching

A user-friendly browsing interface has been developed. By clicking 'BROWSE' at the top of the AGODB website, users first access a brief browsing table that displays fields, including AgoID, Accession, ProteinName, Organism, AgoType, Experimentally verified and links guiding to the 'Detail' page (Figure 4A). As shown in Figure 4B, detailed descriptions are provided in the 'Detail' page, which can be reached by selecting 'more' in the browsing page.

AGODB also provides a search system that allows users to easily retrieve data by clicking 'SEARCH' at the top of website or 'Search System' on the home page. As shown in Figure 5, users can submit a single query against most fields of the database, such as AgoID, Accession, ProteinName, Organism, AgoType, Experimentally verified, GeneID, etc., and also can submit multiple queries simultaneously with Boolean expressions (e.g. AND and OR). The search results are displayed on the search result page in tabular form, similar to browsing the database. More detailed descriptions are displayed in the 'Detail' page, which can be reached by clicking 'more'. Simultaneously, the search results can be conveniently downloaded in batch in *.xml or *.csv format.

GO and KEGG pathway analysis

In AGODB, we collected a total of 1902 proteins, which are from 1068 species. 'Gene ID' can be found in 842 Ago proteins. We performed GO functional annotation and KEGG pathway enrichment analysis using KOBAS-i (34). We kept the results with P-value < 0.05 (FDR corrected) and finally got the results of GO functional annotation for 11 species. Ago genes in species *Arabidopsis thaliana, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Homo sapiens, Mus musculus, Rattus norvegicus, Sus scrofa* and *Xenopus tropicalis* were enriched in the following GO terms: BPs, molecular function and cellular compartment (Supplementary Tables S1–S11), and the top five subclasses of GO enrichment terms in species *Arabidopsis thaliana* and *Bos taurus* are shown in Figure 6. The top five subclasses of GO enrichment

terms in the other nine species are depicted in Supplementary Figures 1–5. We integrated the results of GO function annotation into AGODB. However, no results were obtained from the KEGG pathway enrichment analysis.

Discussion and conclusion

During the past 20 years, significant progress has been made in understanding the structural and biochemical function of the Ago superfamily (4). In recent years, Ago proteins have received extensive attention for their potential applications in gene editing. Researchers are still looking for Ago proteins that can work at moderate temperatures and are suitable for gene editing. Hence, AGODB not only can provide candidates for developing gene editing tools but also has important significance for further promoting the data mining of Ago proteins. In AGODB, comprehensive domain annotation of Ago proteins is mainly provided, followed by general information of Ago proteins and the BPs that Ago genes may participate in.

We encountered some challenges when collecting domain information of Ago proteins. There were only 11 Ago proteins with comprehensive domain annotation that can be manually curated from the literature, and incomplete domain information of 512 Ago proteins was provided by Ryazansky *et al.* (18). Therefore, we took advantage of a well-performing domain annotation tool called AGONOTES to predict the domain of the remaining Ago proteins. AGONOTES annotated comprehensive domain of 1366 Ago proteins. However, there are still 13 Ago proteins' domains that cannot be annotated by AGONOTES.

In conclusion, we provided the first database special for Ago proteins, AGODB, with data retrieval capabilities and browsing, searching and downloading options to facilitate data access. This bioinformatics resource contains detailed information of Ago proteins, including 1095 eAgos and 807 pAgos. AGODB is freely available at http://i.uestc.edu.cn/agodb/. In the future, we will continue to collect the related information of Ago proteins and update AGODB.
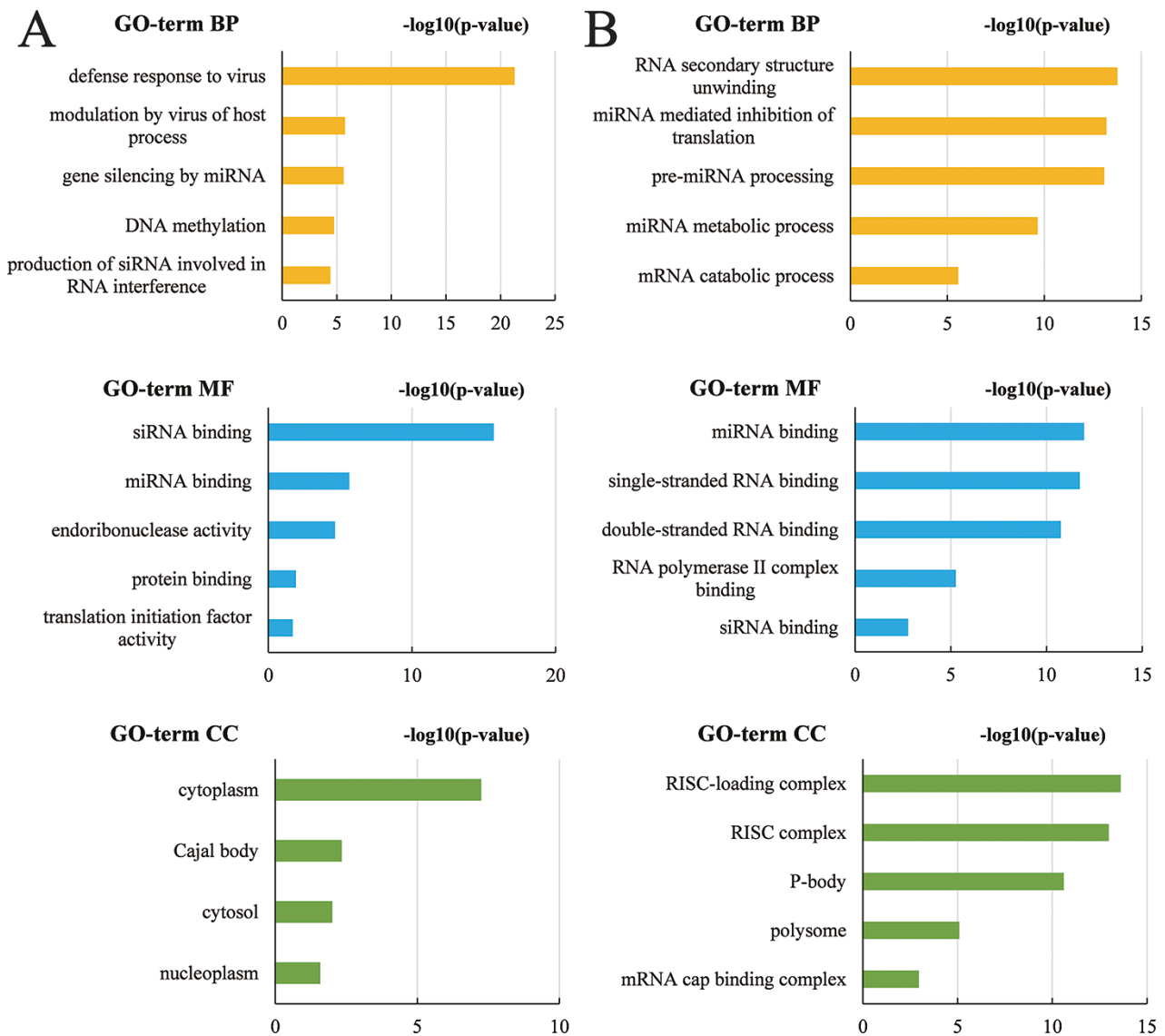
Figure 6. The top five subclasses of GO enrichment terms. (A) Arabidopsis thaliana species. (B) Bos taurus species. BP: biological processes; MF: molecular function; CC: cellular component.

Supplementary data

Supplementary data are available at *Database* Online.

Acknowledgements

Funding

Author's contributions

B.L., S.Y., J.L., X.C., Q.Z. and L.N. developed the web interface of the database. B.L. and S.Y. collected and curated the data. B.L., S.Y. and B.H. wrote the manuscript. B.L., S.Y., B.H., H.C. and J.H. conceived the idea and coordinated the project.

Conflict of interest

There have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

http://i.uestc.edu.cn/agodb/index.html.

References

1. Bohmert,K., Camus,I., Bellini,C. *et al.* (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *Embo J.*, **17**, 170–180. 10.1093/emboj/17.1.170.
2. Swarts,D.C., Makarova,K., Wang,Y. *et al.* (2014) The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.*, **21**, 743–753. 10.1038/nsmb.2879.
3. Niaz,S. (2018) The AGO proteins: an overview. *Biol. Chem.*, **399**, 525–547. 10.1515/hsz-2017-0329.
4. Jin,S., Zhan,J. and Zhou,Y. (2021) Argonaute proteins: structures and their endonuclease activity. *Mol. Biol. Rep.*, **48**, 4837–4849. 10.1007/s11033-021-06476-w.
5. Hegge,J.W., Swarts,D.C. and van der Oost,J. (2018) Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat. Rev. Microbiol.*, **16**, 5–11. 10.1038/nrmicro.2017.73.
6. Olovnikov,I., Chan,K., Sachidanandam,R. *et al.* (2013) Bacterial argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell*, **51**, 594–605. 10.1016/j.molcel.2013.08.014.
7. Swarts,D.C., Jore,M.M., Westra,E.R. *et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, **507**, 258–261. 10.1038/nature12971.
8. Kuzmenko,A., Oguienko,A., Esyunina,D. *et al.* (2020) DNA targeting and interference by a bacterial Argonaute nuclease. *Nature*, **587**, 632–637. 10.1038/s41586-020-2605-1.
9. Jolly,S.M., Gainetdinov,I., Jouravleva,K. *et al.* (2020) *Thermus thermophilus* Argonaute functions in the completion of DNA replication. *Cell*, **182**, 1545–1559 e18. 10.1016/j.cell.2020.07.036.
10. Wu,J., Yang,J., Cho,W.C. *et al.* (2020) Argonaute proteins: structural features, functions and emerging roles. *J. Adv. Res.*, **24**, 317–324. 10.1016/j.jare.2020.04.017.
11. Ye,Z., Jin,H. and Qian,Q. (2015) Argonaute 2: a novel rising star in cancer research. *J. Cancer*, **6**, 877–882. 10.7150/jca.11735.
12. Song,J.J., Smith,S.K., Hannon,G.J. *et al.* (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, **305**, 1434–1437. 10.1126/science.1102514.
13. Meister,G., Landthaler,M., Patkaniowska,A. *et al.* (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197. 10.1016/j.molcel.2004.07.007.
14. Liu,J., Carmell,M.A., Rivas,F.V. *et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441. 10.1126/science.1102513.
15. Höck,J. and Meister,G. (2008) The Argonaute protein family. *Genome Biol.*, **9**, 210. 10.1186/gb-2008-9-2-210.
16. Wang,Y., Zhang,H., Zhong,H. *et al.* (2021) Protein domain identification methods and online resources. *Comput. Struct. Biotechnol. J.*, **19**, 1145–1153. 10.1016/j.csbj.2021.01.041.
17. Copley,R.R., Doerks,T., Letunic,I. *et al.* (2002) Protein domain analysis in the era of complete genomes. *FEBS Lett.*, **513**, 129–134. 10.1016/S0014-5793(01)03289-6.
18. Ryazansky,S., Kulbachinskiy,A. and Aravin,A.A. (2018) The expanded universe of prokaryotic Argonaute proteins. *mBio.*, **9**, 6. 10.1128/mBio.01935-18.
19. Burroughs,A.M., Iyer,L.M. and Aravind,L. (2013) Two novel PIWI families: roles in inter-genomic conflicts in bacteria and mediator-dependent modulation of transcription in eukaryotes. *Biol. Direct.*, **8**, 13. 10.1186/1745-6150-8-13.
20. Sheng,G., Gogakos,T., Wang,J. *et al.* (2017) Structure/cleavage-based insights into helical perturbations at bulge sites within *T. thermophilus* Argonaute silencing complexes. *Nucleic Acids Res.*, **45**, 9149–9163. 10.1093/nar/gkx547.
21. Hegge,J.W., Swarts,D.C., Chandradoss,S.D. *et al.* (2019) DNA-guided DNA cleavage at moderate temperatures by *Clostridium butyricum* Argonaute. *Nucleic Acids Res.*, **47**, 5809–5821. 10.1093/nar/gkz306.
22. Swarts,D.C., Hegge,J.W., Hinojo,I. *et al.* (2015) Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res.*, **43**, 5120–5129. 10.1093/nar/gkv415.
23. Doxzen,K.W. and Doudna,J.A. (2017) DNA recognition by an RNA-guided bacterial Argonaute. *PLoS One*, **12**, e0177097. 10.1371/journal.pone.0177097.
24. Ochoa,A., Llinas,M. and Singh,M. (2011) Using context to improve protein domain identification. *BMC Bioinform.*, **12**, 90. 10.1186/1471-2105-12-90.
25. Jiang,L., Yu,M., Zhou,Y. *et al.* (2020) AGONOTES: a robot annotator for argonaute proteins. *Interdiscip Sci.*, **12**, 109–116. 10.1007/s12539-019-00349-4.
26. UniProt, C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res*, **43**, D204–D212. 10.1093/nar/gku989.
27. El-Gebali,S., Mistry,J., Bateman,A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432. 10.1093/nar/gky995.
28. Schultz,J., Milpetz,F., Bork,P. *et al.* (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U S A*, **95**, 5857–5864. 10.1073/pnas.95.11.5857.
29. Marchler-Bauer,A., Bo,Y., Han,L. *et al.* (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.*, **45**, D200–D203. 10.1093/nar/gkw1129.
30. Yang,S., Long,J., Chen,H. *et al.* (2020) AGOPredict: a predictor for identifying argonaute proteins. In: *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*. Changsha, Hunan, China. New Jersey: IEEE Computer Society Conference Publishing Services (CPS), pp. 184–188.
31. Schoch,C.L., Ciufo,S., Domrachev,M. *et al.* (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database (Oxford)*. 10.1093/database/baaa062.
32. Camacho,C., Coulouris,G., Avagyan,V. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinform.*, **10**, 1471–2105. (Electronic), 421. 10.1186/1471-2105-10-421.
33. Huang,Y., Niu,B., Gao,Y. *et al.* (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics (Oxford, England)*, **26**, 680–682. 10.1093/bioinformatics/btaq003.
34. Bu,D., Luo,H., Huo,P. *et al.* (2021) KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.*, **49**, W317–W25. 10.1093/nar/gkab447.
35. AGO3D: An online service tool for Ago three-dimensional structure construction. (2018) http://i.uestc.edu.cn/ago3d/.