BMC Plant Biology

RESEARCH

Open Access

ANAgdb: a multi-omics and taxonomy database for ANA-grade

Zhonglong Guo[1†], Shaoxuan Luo[1†], Qi Wang[1†], Yixiang Yang[1], Yawen Bai[1], Junrong Wei[1], Dong Wang[3], Yifan Duan[1*], Xiaozeng Yang[2*] and Yong Yang[1*]

Abstract

Background The ANA-grade, encompassing early-diverging angiosperm lineages, Amborellales, Nymphaeales, and Austrobaileyales, represents a fundamental phase in the evolutionary history of flowering plants. Since the completion of key assembly of the *Amborella* genome, the continuous influx of omics data from the lineage underscores the need for a specialized database.

Results Here, we introduce the ANA-grade Genome Database (ANAgdb, https://anagenome.cn/), which integrates multi-omics data including 11 genomes, 167 transcriptomes, and 10 miRNAomes, as well as extensive taxonomic details specific to the ANA-grade. Designed with an array of user-friendly tools, ANAgdb not only facilitates the effective storage, querying, and analysis of data but also enables the integration and dissemination of crucial genomic and taxonomic information.

Conclusion By integrating the comprehensive resources and tools, ANAgdb aims to significantly advance research in phylogenomics and taxonomic studies, providing a robust platform for researchers to explore the genetic and morphological diversities of these ancient plant lineages.

Keywords ANA-grade, Multi-omics, Taxonomy, Database

Background

The ANA-grade, comprising three basally-diverging groups of angiosperms, Amborellales (A), Nymphaeales (N), and Austrobaileyales (A), holds a crucial evolutionary position [1–4]. Amborellales consists of a monotypic genus of living plants, *Amborella*, which includes only one species, *Amborella trichopoda*. This species, native to Grande-Terre in New Caledonia, a Pacific island east of Australia, has sparked significant interest of botanists as it is considered the sister species to all other extant angiosperms [1]. Nymphaeales includes three families, Hydatellaceae, Cabombaceae, and Nymphaeaceae (water lilies), which collectively comprise eight genera and nearly 90 species [5]. Austrobaileyales is composed of Austrobaileyaceae (*Austrobaileya*), Schisandraceae (*Illicium*, *Kadsura*, and *Schisandra*), and Trimeniaceae

†Zhonglong Guo, Shaoxuan Luo and Qi Wang contributed equally to this work and share first authorship.

*Correspondence:
Yifan Duan
yifanduan@njfu.edu.cn
Xiaozeng Yang
yangxz@ibcas.ac.cn
Yong Yang
yangyong@njfu.edu.cn
[1]Co-Innovation Center for Sustainable Forestry in Southern China, College of Life Sciences, Nanjing Forestry University, Nanjing 210037, China
[2]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[3]WeiRan Biotech, Beijing 100085, China

(*Trimenia*), accounting for fewer than 100 species of trees, shrubs, and woody vines [6]. These species within the ANA-grade retain certain ancestral traits and developmental processes, providing a unique perspective on exploring the evolutionary trajectory of flowering plants [7]. Furthermore, research on the ANA-grade offers crucial insights into the genetic and morphological innovations that have driven the extensive diversity and adaptability of modern angiosperms across diverse ecological niches [8].

Since the completion of the first reference genome of *A. trichopoda* [9], several genomes within the ANA-grade have been assembled [10–13], significantly advancing our understanding of the early evolution of angiosperms. Meanwhile, the widespread adoption of next-generation sequencing technology has generated extensive RNA-seq and sRNA-seq datasets [14–18]. Additionally, given its unique phylogenetic position, detailed taxonomic information such as nomenclature, type specimens, and type locality is crucial for taxonomic studies. These vast datasets require a specialized database to effectively store, query, analyze, integrate, and disseminate the information.

Web-based databases that offer interactive data analysis and visualization tools have become increasingly popular in recent years, significantly promoting scientific research across various fields. A prime example of an impactful database in botany is MaizeGDB (https://www.maizegdb.org/) [19], which integrates diverse omics data, germplasm resource information, multiple analytical tools, and communication platforms, effectively facilitating the advancement of breeding practices into the Breeding 4.0 era. Databases like LettuceGDB [20] and HollyGTD [21] provide a range of analysis modules that enable researchers to thoroughly explore and visualize genomes, transcriptomes, miRNAomes, genotypes, and metabolomes, thereby providing valuable support to specialists dedicated to studying lettuces or hollies, respectively. However, there is still a lack of an integrative web-based database specifically focused on the ANA-grade.

Here, we have successfully constructed the ANA-grade genome database (ANAgdb, https://www.anagenome.cn), a comprehensive database that combines publicly available data with newly generated data from our group. ANAgdb hosts multi-omics data (genome, transcriptome, and miRNAome), and integrates extensive taxonomic information specific to the ANA-grade. This database is designed with multiple user-friendly interfaces that allow for easy navigation and display of distinct types of data. ANAgdb includes six online tools for data analysis and a data download page to enhance user accessibility. Consequently, we believe that ANAgdb will provide significant benefits to the research community of botany.

Construction and content

Hardware and software

The ANAgdb was deployed on a Linux server (CentOS 7.9) powered by Alibaba Cloud technology, utilizing Apache (2.4.6) as the web server software. The web application development and technical support were both conducted using PHP language. MySQL was employed for back-end server development. The website interfaces of ANAgdb were crafted using HTML5 (Hypertext Markup Language 5), CSS (Cascading Style Sheets), and JavaScript. For dynamic data visualizations, histograms and heatmaps were integrated using Highcharts (https://www.highcharts.com).

Genome sources

ANAgdb collected 10 publicly available genome assemblies across six species and 1 newly assembled genome (*Amborella trichopoda*, Amtr_2024) produced by our group, including *A. trichopoda* [9], *Brasenia schreberi* [13], *Euryale ferox* [11], *Nuphar advena*, *Nymphaea colorata* [12], and *Nymphaea thermarum* [10] (Table S1). These datasets represent all open-access genome assemblies in the ANA-grade.

The genome of Amtr_2024 was assembled using a combination of Illumina and PacBio HiFi data (~90 Gb, 100× coverage), with an average HiFi reads length of 15,114 bp. Whole genome Illumina sequencing reads were sourced from the SRA database (SRR7500283, ~15× coverage, 14 Gb). Following assembly, correction, and polishing, the final *Amborella* genome assembly (Amtr_2024) was completed, resulting in a 710 Mb assembly. This assembly consists of 37 contigs and 24 scaffolds, including 13 chromosome-level scaffolds, one chloroplast genome, seven mitochondrial contigs, and three contig-scale scaffolds. The assembly achieved a contig N50 of 44 Mb and a scaffold N50 of 54 Mb (Table S2).

Transcriptome sources and analysis

We collected 167 RNA-Seq datasets of five different species from the NCBI Sequence Read Archive (SRA) [22] (https://www.ncbi.nlm.nih.gov/sra) (Table S3). These datasets were initially in compressed form and were converted into Fastq format using the SRA toolkit Linux version 2.8.2. To ensure data quality, FastQC [23] was employed for quality control checks. Trim Galore (version 0.5.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was utilized, applying the parameters '-q 20 --stringency 3 --length 20' to remove adapter of reads. Only reads exceeding 100 bases were retained after trimming. The mapping of clean reads to the corresponding genomes was carried out using Hisat2 [24]. We

used StringTie v1.3.3 [25] to perform transcript assembly and quantification of each RNA-Seq dataset. Transcript expression levels were normalized using fragments per kilobase of transcript per million mapped read (FPKM).

sRNAome sources and analyses

We processed raw data from 10 sRNA-Seq libraries, obtained from the NCBI SRA [22] (Table S4). The original compressed files were converted to Fastq format using the SRA toolkit Linux version 2.8.2. We utilized Trim Galore (version 0.5.0) to trim adapter sequences, applying settings '–length 18 –max_length 28 –small_rna'. After quality control, these Fastq files were then transformed into Fasta format, with common reads merged using a custom Perl script. Reads that matched non-coding RNAs like tRNA, rRNA, snRNA, and snoRNA sequences from the Rfam database (version 13.0), with a tolerance of ≤1 mismatch, were filtered out to enhance annotation accuracy. The filtered sequences were mapped to the corresponding genomes with Bowtie [26]. The miRDeep-P2 software was performed to identify candidate miRNAs [27, 28]. To annotate these miRNAs, the predicted mature miRNA sequences, including±1 nucleotide flanking regions, were aligned against the mature miRNAs in PmiREN2.0 [17, 18] using Bowtie, allowing no more than two mismatches.

Target genes prediction of using psRNATarget

To predict miRNA target genes of *A. trichopoda* and *N. colorata*, we utilized psRNATarget [29] and RNAhybrid [30] independently. Mature miRNA sequences and transcripts were analyzed using the psRNATarget webserver, employing the updated default parameters of Schema V2 (2017 release). The specific parameters were set as: the number of top targets was set to 200, the expectation at 5, penalties for G: U pairing and other mismatches were set at 0.5 and 1 respectively, with extra weight in the seed region adjusted to 1.5. The seed region itself was defined from nucleotides 2 to 13, allowing up to two mismatches, with an HSP size of 19, gap opening penalty at 2, gap extension penalty at 0.5, and the translation inhibition range was defined between 10 and 11 nucleotides. Concurrently, RNAhybrid was employed to identify plausible miRNA: transcript duplexes under plant-specific parameters, maintaining a cut-off value for minimum free energy (MFE)/minimum duplex energy (MDE) at 0.70.

Gene annotation via InterProScan

We utilized InterProScan (version 5.30) [31] to identify and annotate functional domains in all protein sequences. Every protein-coding gene was provided a detailed page containing information about domains, homologues, families, repeats, and Gene Ontology (GO) terms.

Taxonomy sources

We retrieved the nomenclature for 527 scientific names in the ANA-grade from the Plants of the World Online (POWO) database (https://powo.science.kew.org/) (Table S5). Additionally, photos of *A. trichopoda* were supplied by the author, Yong Yang.

Literature retrieval

We employed a Python script to retrieve relevant literature of the ANA-grade from the PubMed database. The process involved the following steps. First, we utilized the Entrez tool to search the PubMed database, using the names of 208 species and 35 related keywords within the ANA-grade. Next, the "esearch" function was used to retrieve the unique identifiers (PMIDs) of these publications. Following this, the "efetch" function was employed to extract detailed information for each article, including the authors, publication year, title, journal, keywords, abstract, and DOI link. The extracted information was organized and saved in a TSV format file, which serves as the foundational data for MySQL of ANAgdb. The source code and keywords used for publication retrieval are available on GitHub (https://github.com/luosx0403/ANAgdb).

Utility and discussion

Database overview

ANAgdb integrates 11 annotated assemblies across six species, representing three orders of the early-diverging angiosperms (Fig. 1). It includes re-analyses of 167 RNA-Seq and 10 sRNA-Seq datasets, along with a comprehensive collection of taxonomic information including 527 scientific names. To enhance user accessibility, ANAgdb offers four hierarchically structured pages: Genome, Transcriptome, miRNA, and Taxonomy (Fig. S1). Additionally, ANAgdb provides six built-in tools including Blast, JBrowse, Search Gene, Gene Annotation, Primer Design, Literature for browsing, gene functional exploration and experimental practice. All data in ANAgdb are freely accessible on the Data page.

Genome

The ANAgdb includes a total of 11 assemblies from six species, including five assemblies from *A. trichopoda*, two assemblies from *E. ferox*., and one assembly each from four other species. Among the five assemblies of *A. trichopoda*, we present a near-gapless chromosome-level genome assembly with only 13 gaps, significantly surpassing the previous assemblies in terms of continuity and completeness. On the Genome page, users can access metadata for each assembly (Fig. 2A). All related information, including genome sequences and detailed genomic annotations, can be downloaded using the FTP.Additionally, the Blast tool has been developed to

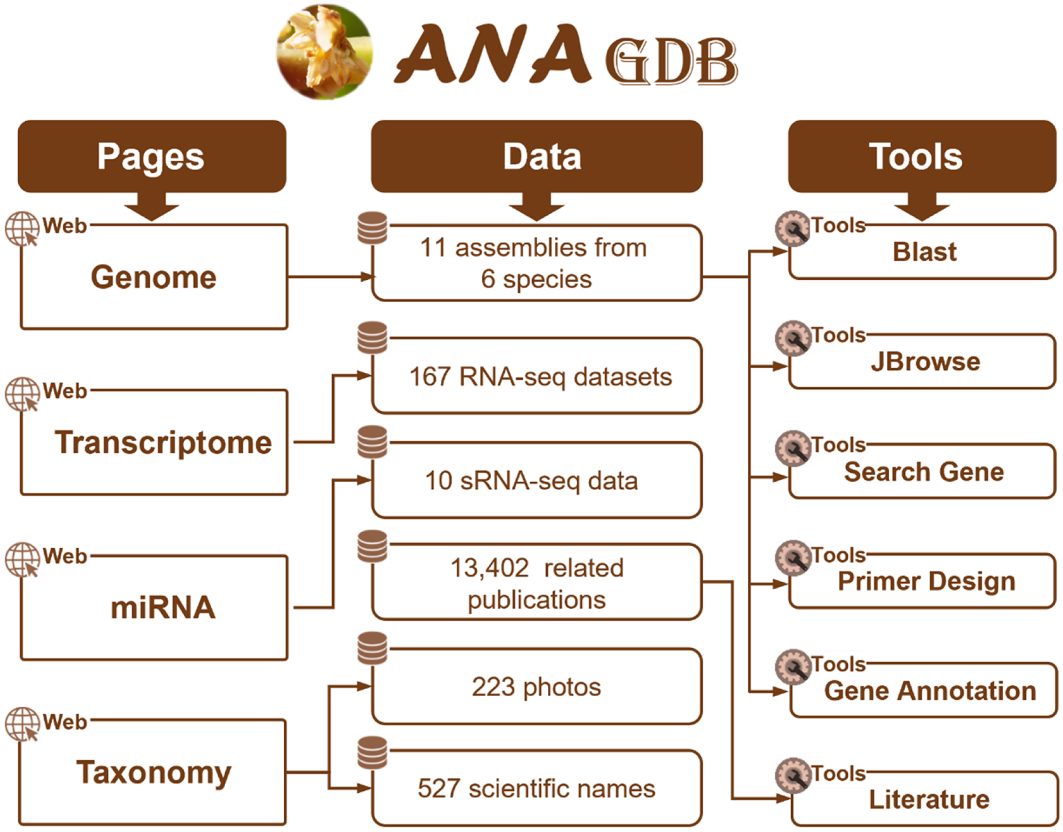

Fig. 1 Schematic of ANAgdb

facilitate homology searches for each gene, enabling users to search annotated genes efficiently.

Taxonomy

The Taxonomy page on ANAgdb offers a detailed and organized overview of 527 scientific names within the ANA-grade (Fig. 2B). Each entry in our summary table includes the scientific name, naming authority, references and taxon status for the nomenclature. By clicking on any scientific name, users are directed to a detailed page that includes information about the type specimen. Additionally, this page provides open-access images of the plants. This resource is designed to support both academic research and general botanical education.

Transcriptome

ANAgdb now includes re-analyzed results from 167 RNA-Seq libraries derived from various tissues of five ANA-grade species. These libraries encompass 14 tissues from *A. trichopoda*, 7 from *B. schreberi*, 2 from *E. ferox*, 7 from *N. colorata*, and 5 from *N. thermarum*. On the Transcriptome page, users can select a species and the specific tissues of interest, and enter a comma-separated list of genes to be queried, then click 'Search' (Fig. 3A). The expression patterns of these genes are displayed through an interactive heatmap, line chart, and a summary table. Furthermore, this page provides the FPKM values for all genes within individual RNA-seq library, making it a valuable resource for gene expression analysis.

miRNA

ANAgdb has collected sRNA-seq datasets for *A. trichopoda* and *N. colorata* from public databases. Utilizing the established miRDeep-P2 [27] pipeline, we identified 186 miRNAs belonging to 109 families in *A. trichopoda* and 141 miRNAs belonging to 88 families in *N. colorata*. The miRNA page offers a summary table of all miRNAs specific to each species, which can be easily accessed and switched via a drop-down list (Fig. 3B). Clicking on a miRNA entry directs users to a detailed information page that includes basic genomic information, cluster information, expression pattern, targets of miRNAs.

Tools

The Blast [32] tool enables users to search for homologous sequences within the genomes of ANA-grade species by either entering a sequence directly into a text box or uploading a file with Fasta format (Fig. 4A). Users can choose from five available Blast algorithms, blastn, blastp, blastx, tblastn, or tblastx, and set detailed parameters

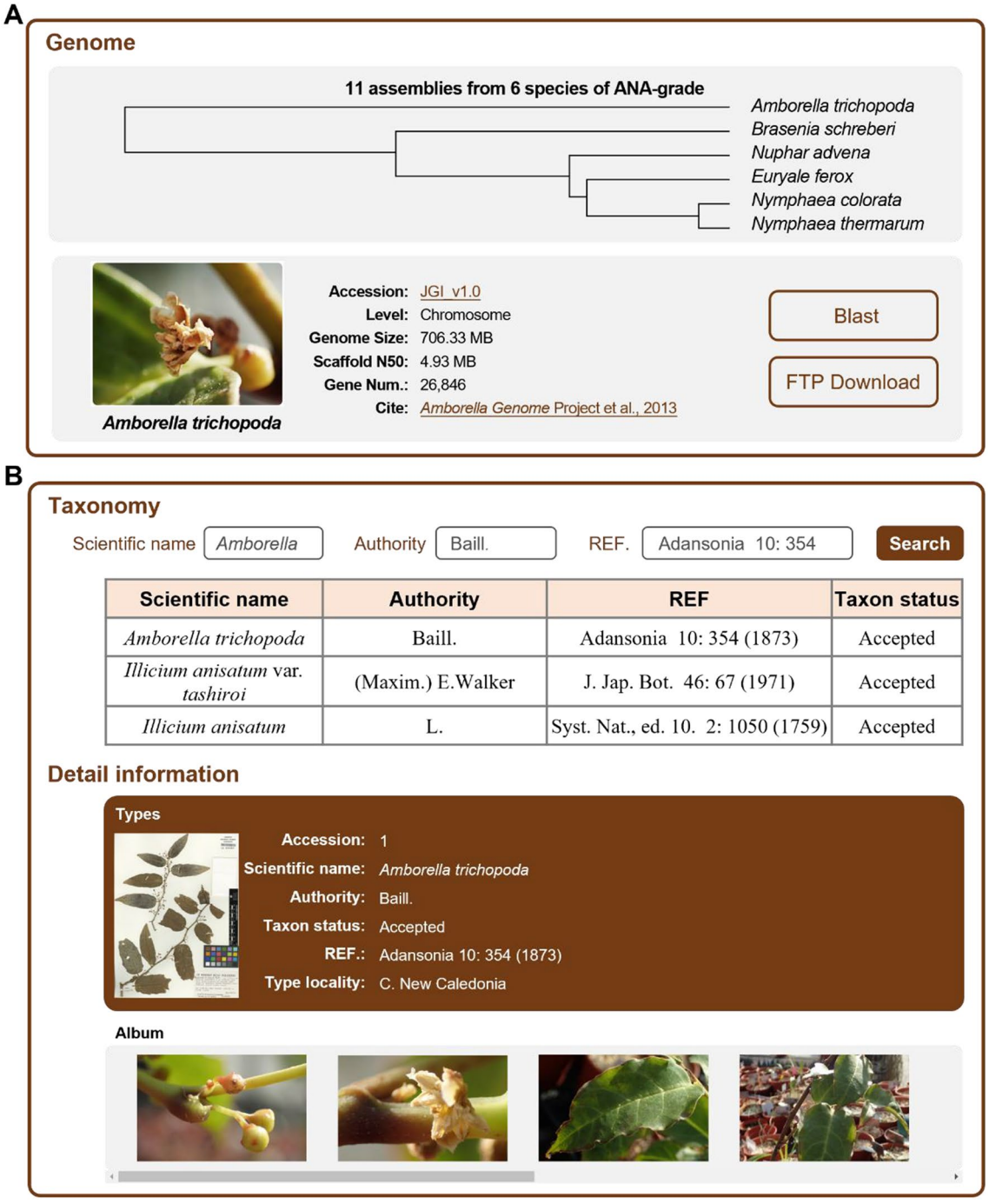

Fig. 2 Genome and taxonomy pages at ANAgdb. **A** Genome page. **B** Taxonomy page

using advanced options. ANAgdb hosts four Blast databases, genome, mRNA, coding sequences, and protein sequences. The results of Blast searches are displayed in a standard table format featuring collapsible fields for Query name, Target name, Score, Identities, Percentage, and Expect, allowing for detailed examination of each hit.

JBrowse is an open-source and pluggable and comprehensive bioinformatic tool designed to visualize and integrate multi-omics data [33]. In ANAgdb, JBrowse is utilized to display integrated genomic information and annotated genomic datasets of all assemblies (Fig. 4B). User can also upload their personal data to easily browse and explore specific information such as gene loci, expression levels of particular genes.

The Search Gene tool on ANAgdb is designed to efficiently retrieve sequences of specific genes (Fig. 4C).

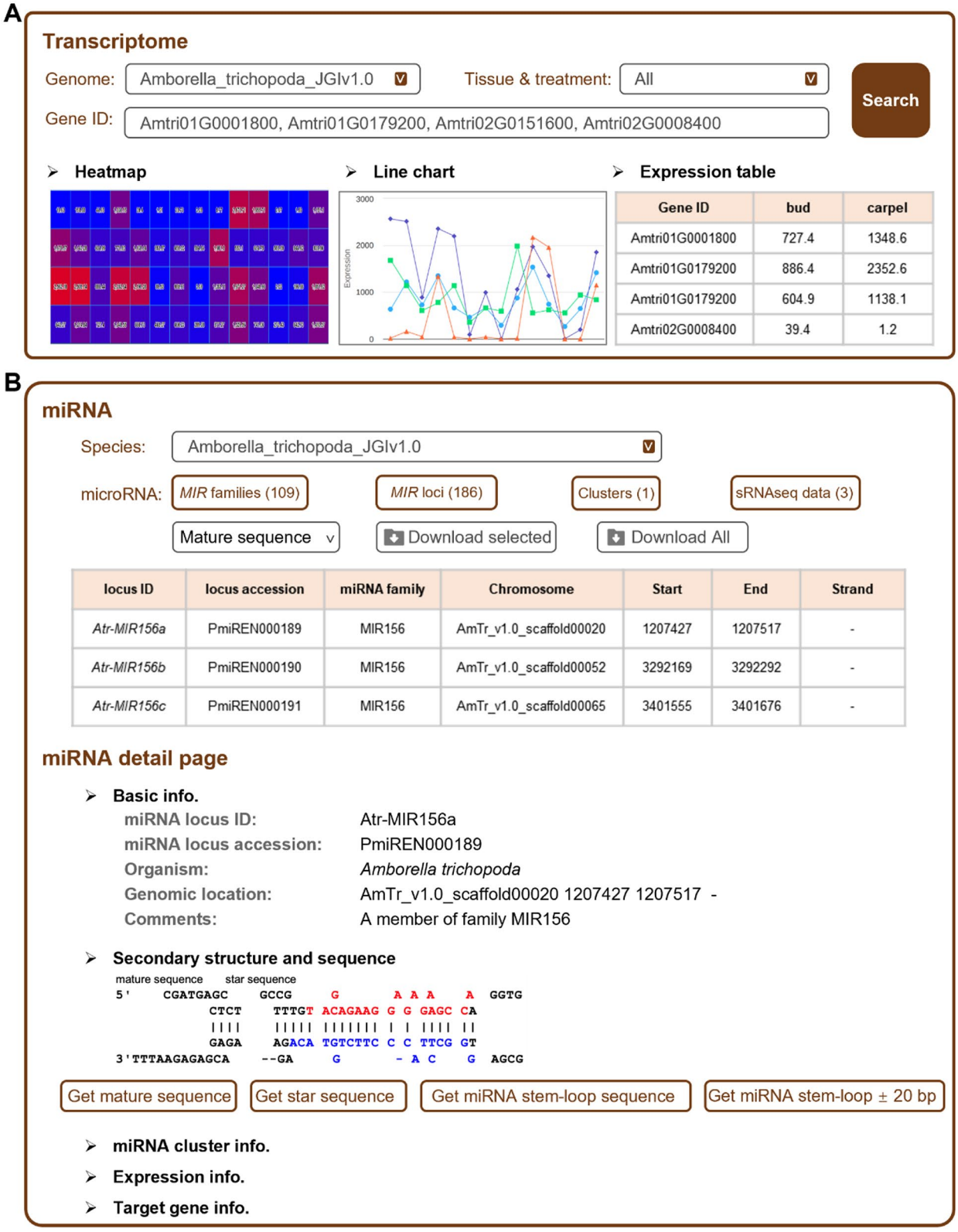

Fig. 3 Transcriptome and miRNA pages at ANAgdb. **A** Transcriptome page. **B** miRNA page

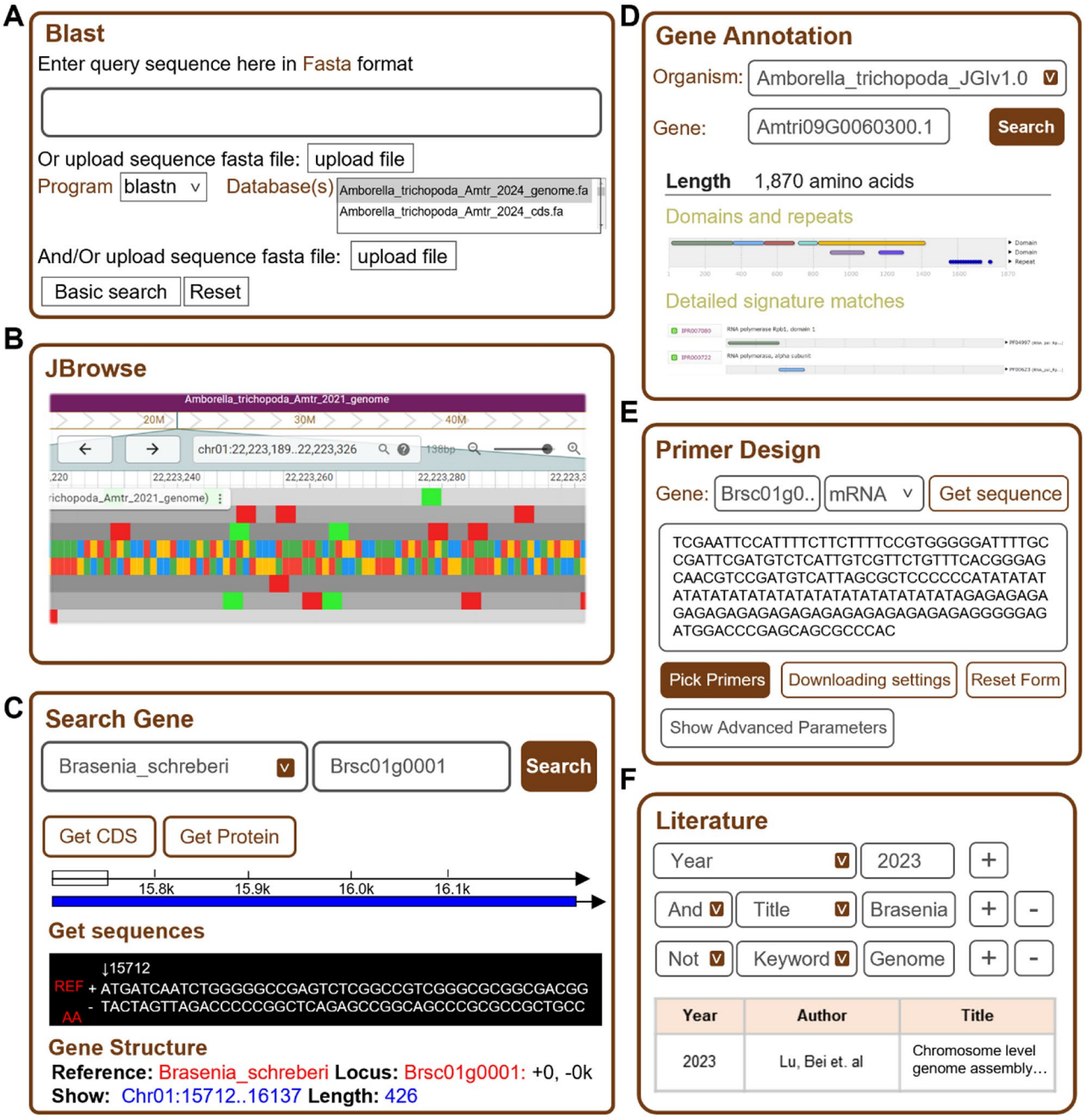

Fig. 4 Schematic diagram of tools and community at ANAgdb. **A** Blast. **B** JBrowse. **C** Search gene. **D** Gene annotation. **E** Primer design. **F** Literature

To use this tool, user first selects a genome assembly from a drop-down list. After selecting the assembly, the user then inputs a gene identifier into the text box. Subsequently, a pop-up window appears, displaying the requested gene sequences. Additionally, the tool also shows the gene structure, including exons, introns, and their corresponding sequences.

The Gene Annotation tool provides extensive functional annotations for each gene in ANAgdb (Fig. 4D). It offers detailed insights into the protein family, homologous superfamily, domains, repeats, and Gene Ontology (GO) terms associated with specific genes. These annotations are derived through similarity searches conducted using the InterPro database [31]. This process involves comparing the gene sequences to known genes in the database to identify similarities and classify the gene based on its functional and structural properties. This helps researchers better understand the potential roles and relationships of genes within broader biological contexts.

The Primer design tool on ANAgdb, powered by the primer3 core program [34], enhances user experimentation by facilitating web-based PCR primer design (Fig. 4E). This interface offers traditional primer design functions along with innovative features convenient for genetic experiments. For example, the genomic, mRNA, or CDS sequences can be automatically loaded into the input field by entering the gene ID. The interface allows users to customize a variety of primer design parameters.

The Literature tool on ANAgdb offers a professional search engine for accessing publications focusing on the ANA-grade, consisting a collection of 13,402 papers (Fig. 4F). This tool enhances the efficiency of literature triage and curation by allowing users to conduct keyword searches by year, author, title, journal, and other keywords. Additionally, the search results provide hyperlinks to the full texts of the publications, facilitating easy access to the relevant research.

Data

All data in ANAgdb are readily accessible for download on the Data page. To streamline storage and download, different data types are systematically organized into specific folders.

Conclusions

In this study, we presented the ANAgdb, the first database specifically dedicated to the ANA-grade, integrating genomic, transcriptomic, miRNAomic, and taxonomic data, all accessible through a user-friendly platform. Given the significance of the ANA-grade, which comprises early-diverging lineages within angiosperms, ANAgdb will serves as a useful resource for botanical research, specifically enhancing our understanding of the origins and evolutionary trajectory of flowering plants.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05613-4.

Supplementary Material 1: Fig. S1. Framework of ANAgdb

Supplementary Material 2: Table S1. Information of assemblies in ANAgdb. Table S2. Information of Amtr_2024 assembly. Table S3. Information of RNA-seq libraries in ANAgdb. Table S4. Information of sRNA-seq libraries in ANAgdb. Table S5. Information of taxonomy in ANAgdb

Acknowledgements

We thank all members in Dr Guo's and Dr. Yang's laboratories for their comments and suggestions on this study.

Author contributions

Z.G. and Y.Y. designed the project; Z.G., S.L., Q.W., D.W., and X.Y. designed and developed the database; S.L. and Q.W. performed the RNA-seq and sRNA-seq analysis; YX.Y. Y.B. and J.W collected taxonomic and phenotypic records; Z.G., Y.Y., and S.L. wrote the manuscript; Z.G., Y.D., X.Y. and Y.Y. revised the manuscript; All authors commented on the manuscript.

Funding

This work was supported by by the National Natural Science Foundation of China, grant number 32300449 to Z.G and 32270217 to Y.Y.

Data availability

ANA-GDB is freely available at https://anagenome.cn/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 20 May 2024 / Accepted: 23 September 2024
Published online: 28 September 2024

References

1. Group TAP, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
2. Endress PK, Doyle JA. Ancestral traits and specializations in the flowers of the basal grade of living angiosperms. Taxon. 2015;64(6):1093–116.
3. Scutt CP. The origin of angiosperms. Evolutionary developmental biology: a reference guide. Springer; 2021. pp. 663–82.
4. Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C, et al. Phylogenomics and the rise of the angiosperms. Nature. 2024;629(8013):843–50.
5. Borsch T, Löhne C, Wiersema J. Phylogeny and evolutionary patterns in Nymphaeales: integrating genes, genomes and morphology. Taxon. 2008;57(4):1052–E1054.
6. Simpson MG. 7 - Diversity and Classification of Flowering Plants: Amborellales, Nymphaeales, Austrobaileyales, Magnoliids, Monocots, and Ceratophyllales. In: *Plant Systematics (Third Edition)*. Edited by Simpson MG: Academic Press; 2019: 187–284.
7. Romanov MS, Bobrov AVFC, Iovlev PS, Roslov MS, Zdravchev NS, Sorokin AN, Romanova ES, Kandidov MV. Fruit and seed structure in the ANA-grade angiosperms: ancestral traits and specializations. 2024, 111(1):e16264.
8. Friedman WE. The meaning of Darwin's abominable mystery. 2009, 96(1):5–21.
9. Amborella Genome P, Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S, Sankoff D, et al. The amborella genome and the evolution of flowering plants. Science. 2013;342(6165):1241089.
10. Povilus RA, DaCosta JM, Grassa C, Satyaki PRV, Moeglein M, Jaenisch J, Xi Z, Mathews S, Gehring M, Davis CC et al. Water lily (*Nymphaea Thermarum*) genome reveals variable genomic signatures of ancient vascular cambium losses. 2020, 117(15):8649–56.
11. Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T, Zhang L, Shen X, Meng F, et al. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. Nat Plants. 2020;6(3):215–22.
12. Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X, et al. The water lily genome and the early evolution of flowering plants. Nature. 2020;577(7788):79–84.
13. Lu B, Shi T, Chen J. Chromosome-level genome assembly of watershield (Brasenia schreberi). Sci Data. 2023;10(1):467.
14. Käfer J, Bewick A, Andres-Robin A, Lapetoule G, Harkess A, Caïus J, Fogliani B, Gâteblé G, Ralph P, dePamphilis CW et al. A derived ZW chromosome system in Amborella trichopoda, representing the sister lineage to all other extant flowering plants. 2022, 233(4):1636–42.
15. Wu P, Zhu Y, Liu A, Wang Y, Zhao S, Feng K, Li L. EfABI4 transcription factor is involved in the regulation of Starch Biosynthesis in Euryale ferox Salisb Seeds. 2022, 23(14):7598.

16. Su Q, Wang H-Y, Tian M, Li C-N, Li X-M, Huang Z-W, Bu Z-Y, Lu J-s. Transcriptomic insight into Viviparous Growth in Water Lily. Biomed Res Int. 2022;2022:8445484.
17. Guo Z, Kuang Z, Wang Y, Zhao Y, Tao Y, Cheng C, Yang J, Lu X, Hao C, Wang T, et al. PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res. 2020;48(D1):D1114–21.
18. Guo Z, Kuang Z, Zhao Y, Deng Y, He H, Wan M, Tao Y, Wang D, Wei J, Li L, et al. PmiREN2.0: from data annotation to functional exploration of plant microRNAs. Nucleic Acids Res. 2022;50(D1):D1475–82.
19. Portwood JL 2nd, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, et al. MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res. 2019;47(D1):D1146–54.
20. Guo Z, Li B, Du J, Shen F, Zhao Y, Deng Y, Kuang Z, Tao Y, Wan M, Lu X, et al. LettuceGDB: the community database for lettuce genetics and omics. Plant Commun. 2023;4(1):100425.
21. Guo Z, Wei J, Xu Z, Lin C, Peng Y, Wang Q, Wang D, Yang X, Xu KW. HollyGTD: an integrated database for Holly (Aquifoliaceae) genome and taxonomy. Front Plant Sci. 2023;14:1220925.
22. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50(D1):D20–6.
23. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics. 2017;33(19):3137–9.
24. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
25. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.
26. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinf. 2010;32(1):11. 11-11.17. 14.
27. Kuang Z, Wang Y, Li L, Yang X. miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. Bioinformatics. 2019;35:2521–2.
28. Yang X, Li L. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics. 2011;27(18):2614–5.
29. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 2018;46(W1):W49–54.
30. Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 2006;34(suppl2):W451–4.
31. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
32. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos JS, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
33. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, et al. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 2016;17:66.
34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115–115.

Publisher's note