

RESEARCH

Open Access



The burdock database: a multi-omic database for *Arctium lappa*, a food and medicinal plant

Yueyue Song[†], Yanyun Yang[†], Liang Xu^{*}, Che Bian, Yanping Xing, Hefei Xue, Wenjuan Hou, Wenxiao Men, Deqiang Dou and Tingguo Kang^{*}

Abstract

Background Burdock is a biennial herb of Asteraceae found in Northern Europe, Eurasia, Siberia, and China. Its mature dry fruits, called *Niu Bang Zi*, are recorded in various traditional Chinese medicine books. With the development of sequencing technology, the mitochondrial, chloroplast, and nuclear genomes, transcriptome, and sequence-related amplified polymorphism (SRAP) fingerprints of burdock have all been reported. To make better use of this data for further research and analysis, a burdock database was constructed.

Results This burdock multi-omics database contains two burdock genome datasets, two transcriptome datasets, eight burdock chloroplast genomes, one burdock mitochondrial genome, one *A. tomentosum* chloroplast genome, one *A. tomentosum* mitochondrial genome, 26 phenotypes of burdock varieties, burdock rhizosphere-associated microorganisms, and chemical constituents of burdock fruit, pericarp, and kernel at different growth stages (using UPLC-Q-TOF-MS). The wild and cultivation distribution of burdock in China was summarized, and the main active components and pharmacological effects of burdock currently reported were concluded. The database contains ten central functional modules: Home, Genome, Transcriptome, Jbrowse, Search, Tools, SRAP fingerprints, Associated microorganisms, Chemical, and Publications. Among these, the "Tools" module can be used to perform sequence homology alignment (Blast), multiple sequence alignment analysis (Muscle), homologous protein prediction (Gene-wise), primer design (Primer), large-scale genome analysis (Lastz), and GO and KEGG enrichment analyses (GO Enrichment and KEGG Enrichment).

Conclusions The database URL is <http://210.22.121.250:41352/>. This burdock database integrates molecular and chemical data to provide a comprehensive information and analysis platform for interested researchers and can be of immense help to the cultivation, breeding, and molecular pharmacognosy research of burdock.

Keywords *Arctium lappa*, Database, Genome, Transcriptome, Tools

[†]Yueyue Song and Yanyun Yang contributed equally to this work.

*Correspondence:

Liang Xu

861364054@qq.com

Tingguo Kang

kangtingguo@163.com

School of Pharmacy, Liaoning University of Traditional Chinese Medicine, Dalian 116600, China

Background

Arctium lappa L., a medicinal and edible plant of Asteraceae, is native to Northern Europe, Eurasia, Siberia, and China. Because of its rich nutritional value, it enjoys the reputation of being the king of vegetables in Europe. The burdock root is consumed as a vegetable in Europe and Japan, while its fruit has a long medicinal history in China. The dried, ripe fruit is piquant or bitter to the taste and has cold properties associated with the lungs




and stomach meridians. It is traditionally used to disperse wind–heat, detoxify and purify blood, and treat common cold, cough, sore throat, measles, and rash [1, 2]. Burdock was first described in the *Mingyibieliu* and is a commonly used traditional medicine since ancient times [3]. In *Taipingshenghuifang*, *Niu Bang Zi San* (burdock powder) is prescribed to treat febrile conditions and gastrointestinal disorders resulting from heat and poison attacks. Furthermore, *Fructus Arctii* is acrid and neutral in nature and serves as an adjunct medicine to *Yin Qiao San* (Lonicera and Forsythia powder), as stated in the *Wenbingtiaoli*. It moistens the lungs, disperses wind, clears heat, and relieves sore throat. Several chemical components are present in burdock, with lignans, volatile oils, fatty acids, terpenoids, and phenolic acids being the primary

components [4]. Therefore, *Fructus Arctii* has several antitumor, anti-inflammatory, antidiabetic, antibacterial, and antiviral properties. It is clinically used to treat cold, cough, phlegm, measles, rubella, sore throat, and other diseases [5]. Many Chinese patent medicines are made of burdock using modern pharmaceutical technology, such as Yanshu capsule, Yanshu oral liquid, Qingre Qudu pill, vitamin C Yinqiao granules and so on [6, 7].

Asteraceae is one of the largest angiosperm families, including more than 1000 genera. It is widely distributed in the world, with most species having important economic value as ornamental, food, and medicinal plants. The family has the highest evolutionary status among dicotyledons [8]. Therefore, with the continuous development and improvement of sequencing technology

Arctium lappa Database




Home
Genome ▾
Transcriptome ▾
Jbrowse ▾
Search ▾
Tools ▾
SRAP fingerprints
Associated microorganisms ▾
Chemical ▾
Publication ▾

Welcome to *Arctium lappa* Database


According to the Pharmacopoeia of the People's Republic of China (2020), burdock was first described in *Mingyibieliu*, an ancient medical book, and belongs to the family Asteraceae. The dried, ripe fruit tastes bitter and is cold in nature to the lungs and stomach. It is also used for wind-heat evacuation, lung through rash, detoxification and pharynx effect. According to the Chinese pharmacopoeia, the arctiin derived from *Arctium lappa* must have a concentration of at least 5.0% in dried products. Furthermore, the moisture content must be no more than 9.0%, and the total ash content must not exceed 7.0%. Diseases such as cold, phlegm-producing coughs, measles, rubella and sore throats can be treated using clinical methods. Recent research states that several chemical components are present abundantly in *A. lappa*, with lignans, volatile oils, fatty acids, terpenoids and phenolic acids being the primary components. Therefore, burdock has several anti-tumour, anti-inflammatory, anti-diabetic, anti-bacterial and anti-viral properties. *A. lappa* has been used as a medicine since ancient times. The *taipingshenghuifang* in burdock powder known as Niubangzi (*Fructus arctii*) is used to treat heat and poison attacks, as well as hot, rough faeces. Furthermore, Niubangzi is acrid in flavour and neutral in nature and serves as an adjunct medicine to Yin Qiao San (Lonicera and Forsythia Powder), as stated in the *wenbingtiaoli*. It moistens the lungs, resolves heat, dissipates binds, extinguishes wind and relieves sore throat. There are currently many Chinese patent medicines made of burdock using modern pharmaceutical technology, such as Yanshu capsule, Yanshu oral liquid, QingreQudu pill, vitamin C Yinqiao granules, Yinqiao granules, Yinqiao tablets, sand plum Xiaoke capsule, Yulan hypoglycaemia-lowering capsule, red grass anti-snoring granules, Yimingmu tablets and many more.

Search




Herbs to 2 m tall, biennial. Stem purplish, erect, stout, apically branched, sparsely cobwebby; branches ascending. Leaves abaxially grayish white or pale green and thinly felted, adaxially green, sparsely strigose, and yellow gland-dotted. Basal leaves with petiole ca. 32 cm; leaf blade broadly heart-shaped, ca. 30 × 21 cm, margin entire and repand to mucronulate denticulate. Cauline leaves similar to basal leaves or ovate; uppermost cauline leaves ovate to shallowly cordate, smaller. Capitula few to many, ± corymbose. Involucre ovoid, 1.5–2 cm in diam., glabrous except for minute glandular hairs. Phyllaries all with a hooked apex; outer phyllaries triangular to lanceolate-subulate, ca. 15 × 1 mm; middle and inner phyllaries lanceolate to linear-subulate, ca. 15 × 1.5–3 mm. Corolla purplish red, ca. 1.4 cm, tube ca. 8 mm. Achene pale brown variegated with dark brown or concolorous, narrowly sometimes obliquely obovoid, 5–7 mm, multistriate. Pappus bristles to 3.5 mm. Fl. and fr. Jun–Sep. 2n = 18, 34, 36.


Near villages, roadsides, near rivers, wet and waste places, forest margins, thickets, valleys, slopes; 700–3500 m. Throughout China except for Hainan, Taiwan, and Xizang [Afghanistan, Bhutan, India, Japan, Nepal, Pakistan; SW Asia, Europe].




Navigation




Browse




Jbrowse




Search



Blast



Muscle



Download

Copyright © Copyright 2021 Auto Parts All Right Reserved

Fig. 1 Burdock database homepage. Ten main modules displayed at the top of the interface include “Home”, “Genome”, “Transcriptome”, “Jbrowse”, “Search”, “Tools”, “SRAP fingerprints”, “Associated microorganisms”, “Chemical”, and “Publications”

in recent years, many Asteraceae plant genomes have been completely sequenced to study the evolution of biological groups. These include *Artemisia annua* [9], lettuce (*Lactuca sativa*) [10], sunflower (*Helianthus annuus*) [11], artichoke (*Cynara cardunculus*) [12], *Mikania micrantha* [13], *Carthamus tinctorius* [14], *Stevia rebaudiana* [15], chicory (*Cichorium intybus*), endive (*C. endivia*), great burdock (*A. lappa*), yacon (*Smallanthus sonchifolius*) [16], and *Erigeron breviscapus* [17]. Our laboratory also performed complete genome sequencing of burdock. Approximately 1.70 Gb (95.4%) of the contig sequences were anchored onto 18 chromosomes using Hi-C data [18].

Wei et al. (2020) sequenced the transcriptome of the burdock root, laying an experimental foundation for the identification of burdock functional genes, analysis of secondary metabolic pathways, and research on its regulatory mechanism [19]. Our laboratory performed transcriptome sequencing of 10 burdock samples (five growth stages of burdock fruits, leaves, stems, perianths, petioles, and roots) to study the biosynthetic pathway of arctiin and key enzyme genes in burdock [20]. Based on genome data, transcriptome analysis was also performed on the roots of burdock at different growth stages. By comparative transcriptome analysis, our research team preliminarily speculated on the regulatory mechanism of secondary metabolite biosynthesis of burdock and found three proteins highly related to the accumulation of arctiin: 4-coumarate-CoA ligase (4CL), dirigent protein (DIR), and hydroxycinnamoyl transferase (HCT) [18].

However, the synthetic pathway of arctiin has not been verified based on the current research results.

Accessing large and complex data efficiently and accurately to conduct research is a problem to be solved. At present, online databases include the Cucurbitaceae database (<http://www.gourdbase.cn/>) [21], Bayberry database (<http://www.bayberrybase.cn/>) [22] and Malvaceae database (<http://magen.whu.edu.cn/>) [23]. To further explore the biosynthetic pathway, breeding, and ecology-related research of burdock, we developed a burdock multi-omics database. For convenience, we compiled the phenotypic, genomic, transcriptomic, and chemical research data of burdock to develop a user-friendly burdock database. The database also contains tools for researchers, such as genome browser, sequence alignment, homologous protein prediction, and primer design.

Utility and discussion

Burdock database content

The burdock database contains ten central functional modules: Home, Genome, Transcriptome, Jbrowse, Search, Tools, SRAP fingerprints, Associated microorganisms, Chemical, and Publications (Fig. 1). It includes two genome datasets, two transcriptome datasets, eight burdock chloroplast genomes, one burdock mitochondrial genome, one *A. tomentosum* chloroplast genome, one *A. tomentosum* mitochondrial genome, 26 phenotypes of burdock varieties [24–27], distribution of burdock resources in China, burdock rhizosphere-associated microorganisms [28], 18 pairs of sequence-related

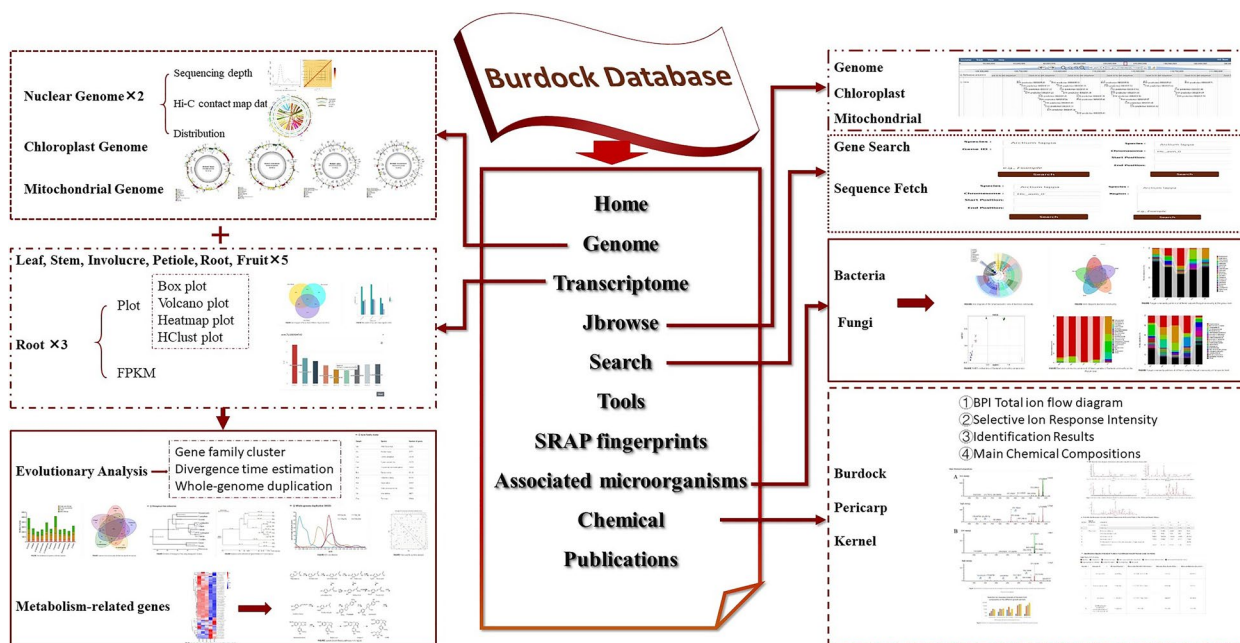


Fig. 2 Schematic of the burdock database

amplified polymorphism (SRAP) primers [27], and chemical constituents of burdock fruit, pericarp, and kernel at different growth stages (using UPLC-Q-TOF-MS) [20, 29]. In addition, the summary and conclusion on 9 reported pharmacological effects of burdock have been realized. In the “Publications” module, 120 relevant papers and a book, *The Study of Burdock in China*, can be downloaded for research purposes. The database also contains intuitive and convenient tools to facilitate data retrieval (Fig. 2).

Applications of the burdock database

Genome

This module contains the first high-quality chromosome-level draft genome of *A. lappa* obtained using Illumina and PacBio sequencing data. The assembled genome is approximately 1.79 Gb, with 1.12 Gb (68.46%) of repetitive sequences, 32,771 protein-coding genes, and 616 positively selected candidate genes [18]. Another

published burdock genome, with assembled the contigs of approximately 1.73 Gb, is also included [16]. This module includes the sequencing results of eight complete chloroplast genomes of burdock, with a full length between 152,767 bp and 163,851 bp, and 90 annotated protein-coding genes. It also includes the complete chloroplast genome of *A. tomentosum*, a close source plant of burdock, with a total length of 152,688 bp, and 89 annotated protein-coding genes [30]. A burdock chloroplast genome dataset from the NCBI database (accession number MH161419) is also present. The module includes one burdock and one *A. tomentosum* mitochondrial genome, with similar genome sizes of 312,598 bp and 312,609 bp, respectively, and 76 and 75 annotated protein-coding genes, respectively [31]. The module summarizes the phenotypic characteristics of 26 burdock cultivars reported thus far, and it contains the name, origin, root characteristics, leaf color, and maturity descriptions of these cultivars [24–27]. To better develop and utilize the medicinal

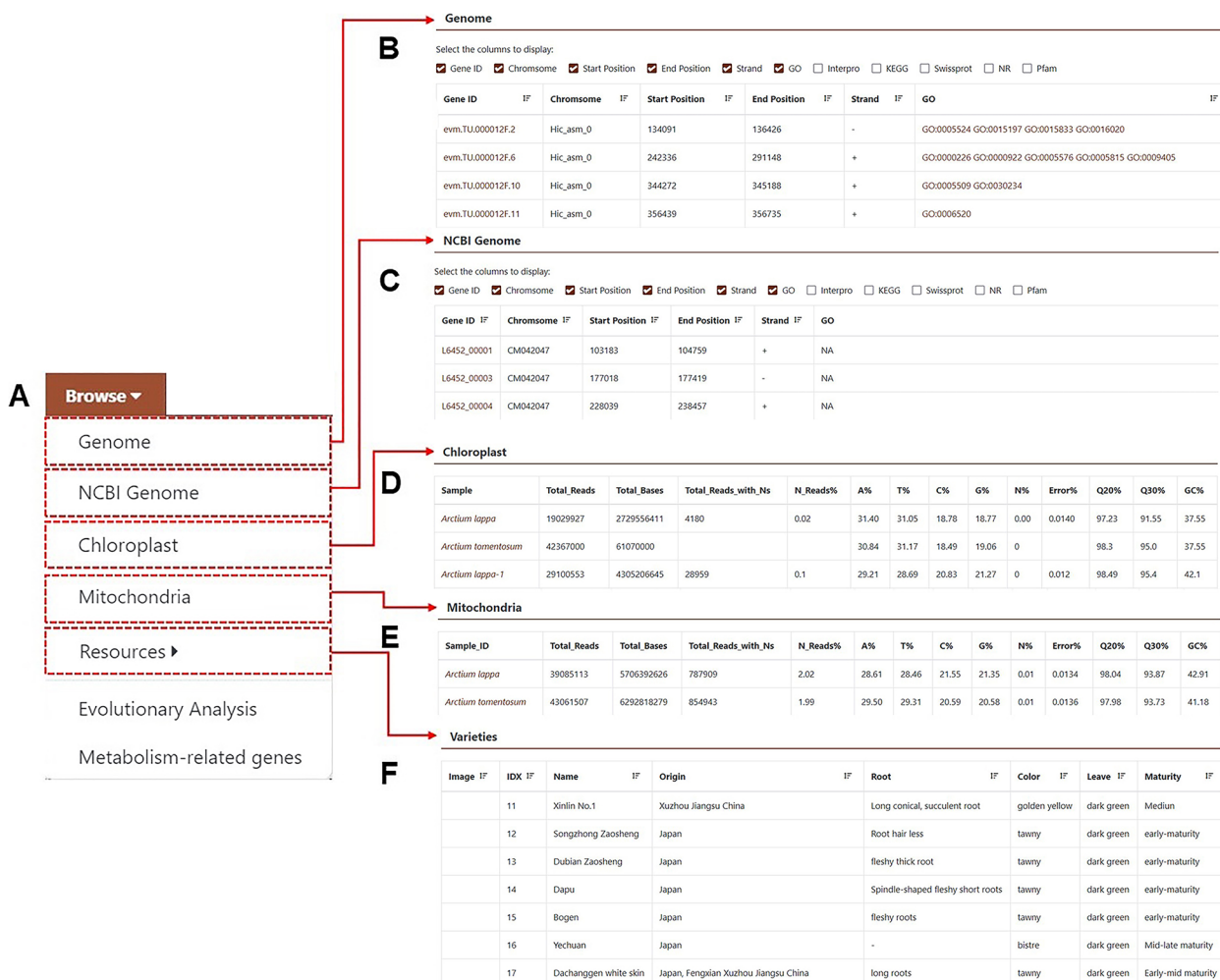
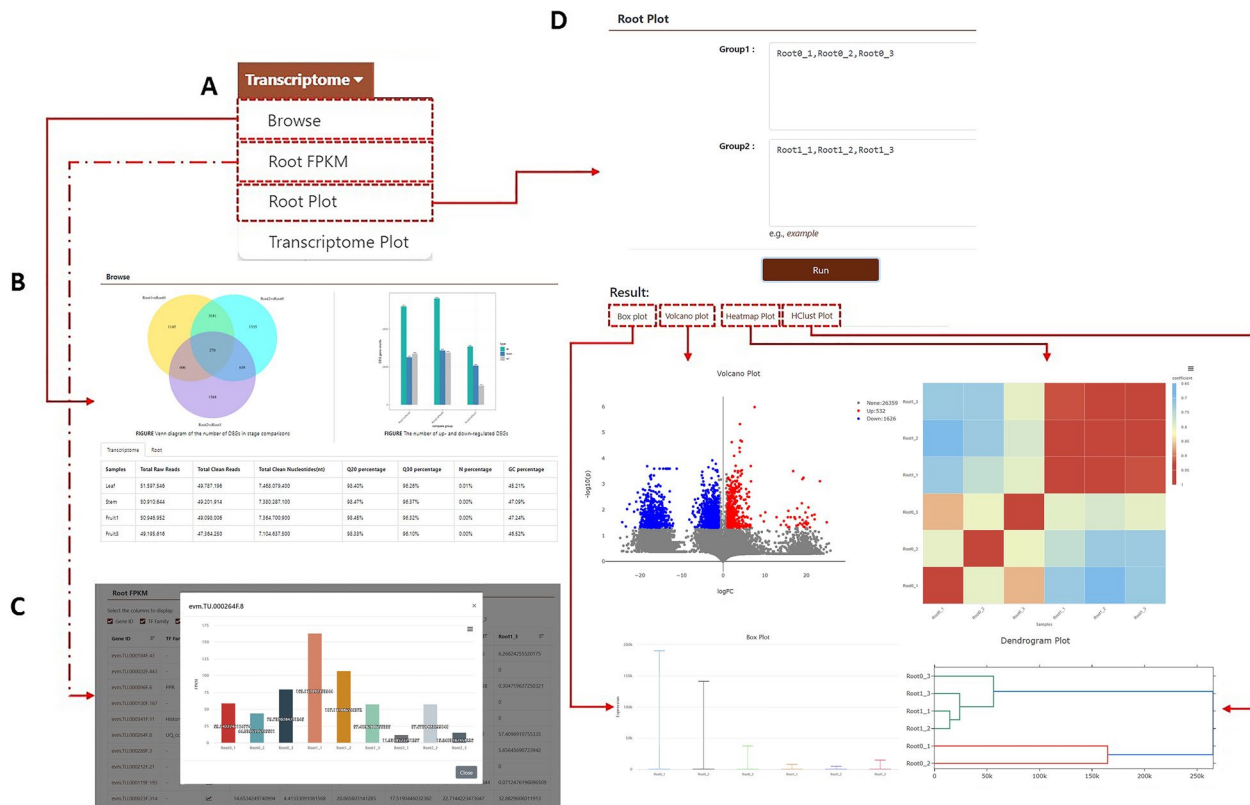


Fig. 3 Details of the “Browse” module. **A** Sub-modules include “Genome”, “NCBI Genome”, “Chloroplast”, “Mitochondria”, “Varieties”, “Evolutionary Analysis”, and “Metabolism-related genes”. **B–F** Functions of “Genome”, “NCBI Genome”, “Chloroplast”, and “Varieties” are presented from top to bottom



resources of burdock, the module further summarized the main wild and cultivation conditions in China.

To explore, users can click the “Genome” tab in the navigation bar, where “Genome”, “NCBI Genome”, “Chloroplast”, “Mitochondria”, “Resources”, “Evolutionary Analysis”, and “Metabolism-related genes” appear in the drop-down list (Fig. 3A). Users can click any label on this list according to their needs. After clicking “Genome”, the page shows the chromosome position, start and end positions of each gene, as well as the protein annotation and sequence number of genes in the GO, InterPro, KEGG, Swiss-Prot, NR, and Pfam databases (Fig. 3B). “NCBI Genome”, “Chloroplast”, and “Mitochondria” are similar (Fig. 3C-F). With this module, users can learn the basics of each gene in the burdock genome in detail. The gene ID is clickable and links to the “Browse” module for visualising gene information. Moreover, a click on the ID number of each protein database will link to the corresponding database. “Evolutionary Analysis” shows the results of phylogenetic estimation of the divergence time of protein-coding genes in 11 sequenced genomes (including burdock), and the expansion and contraction results of 11 plant gene families. “Metabolism-related genes” shows the predicted biosynthetic pathways

of lignan and lignin in burdock based on current research results. To verify this prediction, the module collected 18 kinds of enzymes related to lignans and lignin biosynthesis pathway recently reported. Through comparing, the module will have those with corresponding domains as candidate genes. The database showed the results of heat map analysis of candidate genes at five different growth stages of the burdock fruit. With the corresponding gene ID indicated in the figure, users can search for such an ID and then download the sequence if they want to perform qPCR validation for the results. All the gene sequences in this module can be downloaded, which can be used for analysis of individual or population variation information of burdock, such as single nucleotide polymorphism (SNP), insertion deletion (InDel) and copynumber variation (CNV); Obtain molecular genetic characteristics, further predict candidate genes for economic traits, and analyze genetic evolution.

Transcriptome

This module contains RNA-seq datasets from different tissues of burdock at different growth stages, including fruit (bud, early flowering, full flowering, late flowering, mature), perianth, stem, petiole, involucre, leaf, root (seedling, annual,

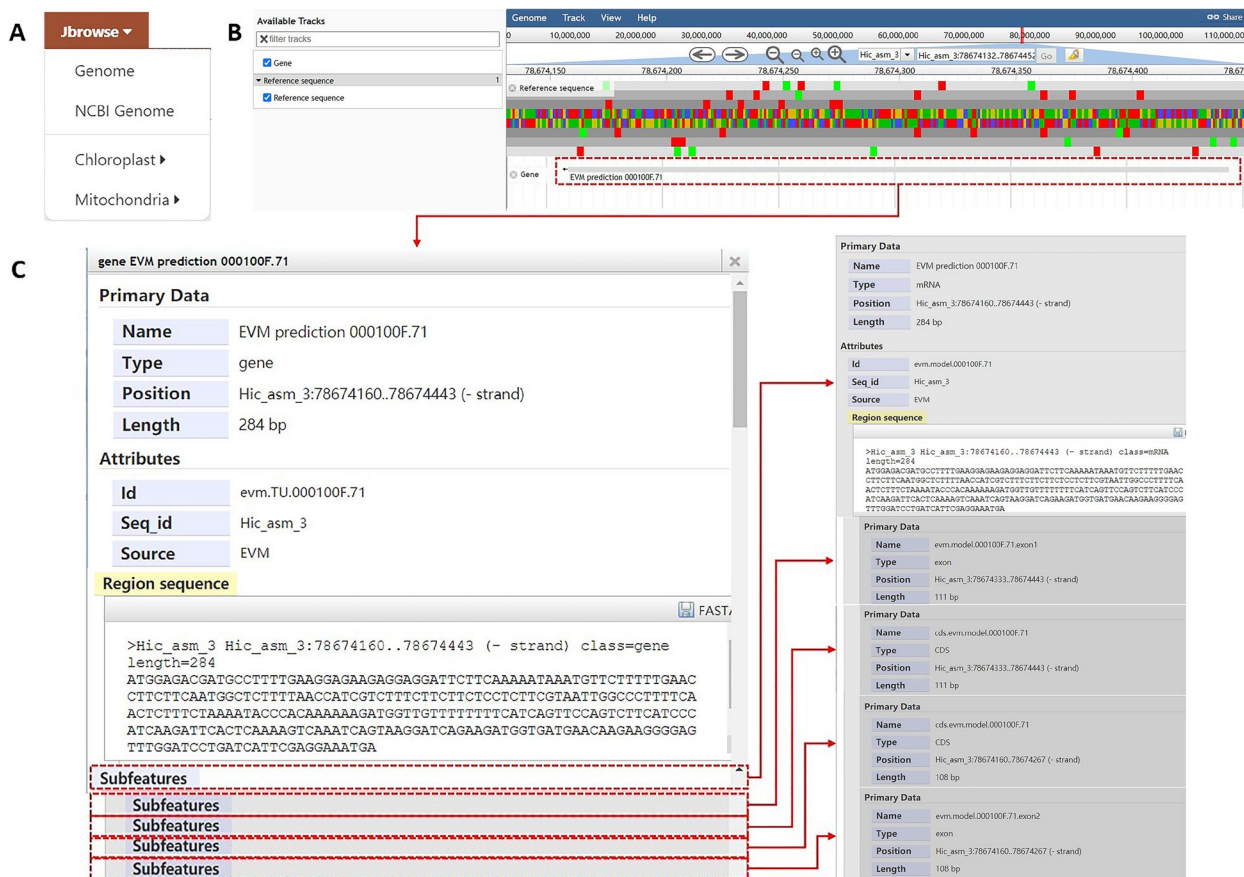


Fig. 5 Detailed description of the “Jbrowse” module. **A** Sub-modules include “Genome”, “NCBI Genome”, “Chloroplast”, and “Mitochondria”. **B** The “Genome” interface. **C** Example of “EVM prediction 000100F.71”

and biennial), and stalk (Fig. 4). The database presents the results of comparative transcriptome analysis and candidate differentially expressed genes (DEGs) for roots at different developmental stages as a bar chart (Fig. 4C) [18]. Gene ID under this module can also be directly linkable to the “Jbrowse” module and can be downloaded. The database also supports four visual analysis charts (“Box plot”, “Volcano plot”, “Heatmap plot”, and “HClust plot”) that show the gene expression levels of fragments per kilobase of transcript per million fragments mapped (FPKM) in the different tissues described above. Users must click the “Transcriptome” tab in the navigation bar and select “Root Plot” or “Transcriptome Plot” in the drop-down list. Taking “Root0 versus Root1” as an example, users can click the different options in the horizontal results bar to display the comparison results of the four charts (Fig. 4D). Users can visually identify the differences in expression at different sites.

Jbrowse

The “Jbrowse” module of the database uses JBrowse, a genome visualization tool, to display two nuclear

genomes of burdock and the chloroplast and mitochondrial genomes of burdock and its near source species *A. tomentosum* (Fig. 5A). Users can select a specific region of the desired chromosome in the search box above, and click the “Go” button to browse. For example, Fig. 5B shows the 78,674,134–78,674,445 bp interval on chromosome “Hic_asm_3” (Fig. 5B). As shown in Fig. 5C, users can click the “EVM prediction 000100F.71” gene, and the transcript of the gene will be displayed on the page, showing the transcript information of all exons and introns of the gene. A FASTA format file can also be downloaded (Fig. 5C). Clicking on gene IDs obtained from other modules can also link users to this tool.

Search

The burdock database contains two genomes of size 1.79 [18] and 1.73 Gb [16]. The search module contains two sub-pages: gene search and sequence acquisition (Fig. 6A). On the gene search page, users can search by gene ID or range. The search results in the following information: “Gene ID”, “Chromosome”, “Start Position”,

The diagram illustrates the structure of the Search module. Part A shows a navigation menu with 'Search' (dropdown), 'Gene Search', and 'Sequence Fetch'. Part B shows a 'Search' page with 'Search By Gene ID' and 'Search By Range' options. Part C shows a 'Sequence Fetch' page with 'Sequence Fetch' and 'Multiple Sequence Fetch' options. Three screenshots show search results for 'Arctium lappa' with columns for Gene ID, IF, Chromosome, IF, Start Position, IF, End Position, IF, Strand, IF, and GO.

Gene ID	IF	Chromosome	IF	Start Position	IF	End Position	IF	Strand	IF	GO
evm.TU.000012F.2		Hic_asm_0		134091		136426		-		GO:000524 GO:0015197 GO:0015833 GO:0016020
evm.TU.000012F.6		Hic_asm_0		242336		291148		+		GO:0000226 GO:0000922 GO:0005576 GO:0005815 GO:0009405
evm.TU.000012F.10		Hic_asm_0		344272		345188		+		GO:0005509 GO:0009234

The screenshots also show a 'Result:' section with 'Select the columns to display:' checkboxes and a 'Sequence:' section with a 'Download' button. The sequence text is shown in a text area.

Fig. 6 Detailed description of the “Search” module. A Sub-modules include “Gene Search” and “Sequence Fetch”. B Example of “Gene Search”. C Example of “Sequence Fetch”

“End Position”, “Strand” and the annotation of genes in six protein databases: “GO”, “InterPro”, “KEGG”, “Swiss-Prot”, “NR”, and “Pfam” (Fig. 6B). Users can click “Sequence Fetch” to download the nucleotide sequence in the specified region. They can enter the start and end points of the sequence into the search box, and click “Download” to download all sequences in that range. To download multiple sequences, click “Multiple Sequence Fetch” in the secondary navigation to enter all sequences

into the “Region” search box, and multiple sequences can be downloaded simultaneously (Fig. 6C).

Tools

The burdock database also has applications including “Blast”, “Muscle”, “Genewise”, “Primer”, “Lastz”, “GO Enrichment”, and “KEGG Enrichment” to meet most data processing requirements (Fig. 7A). The “Blast” module can perform homology alignment with the two burdock

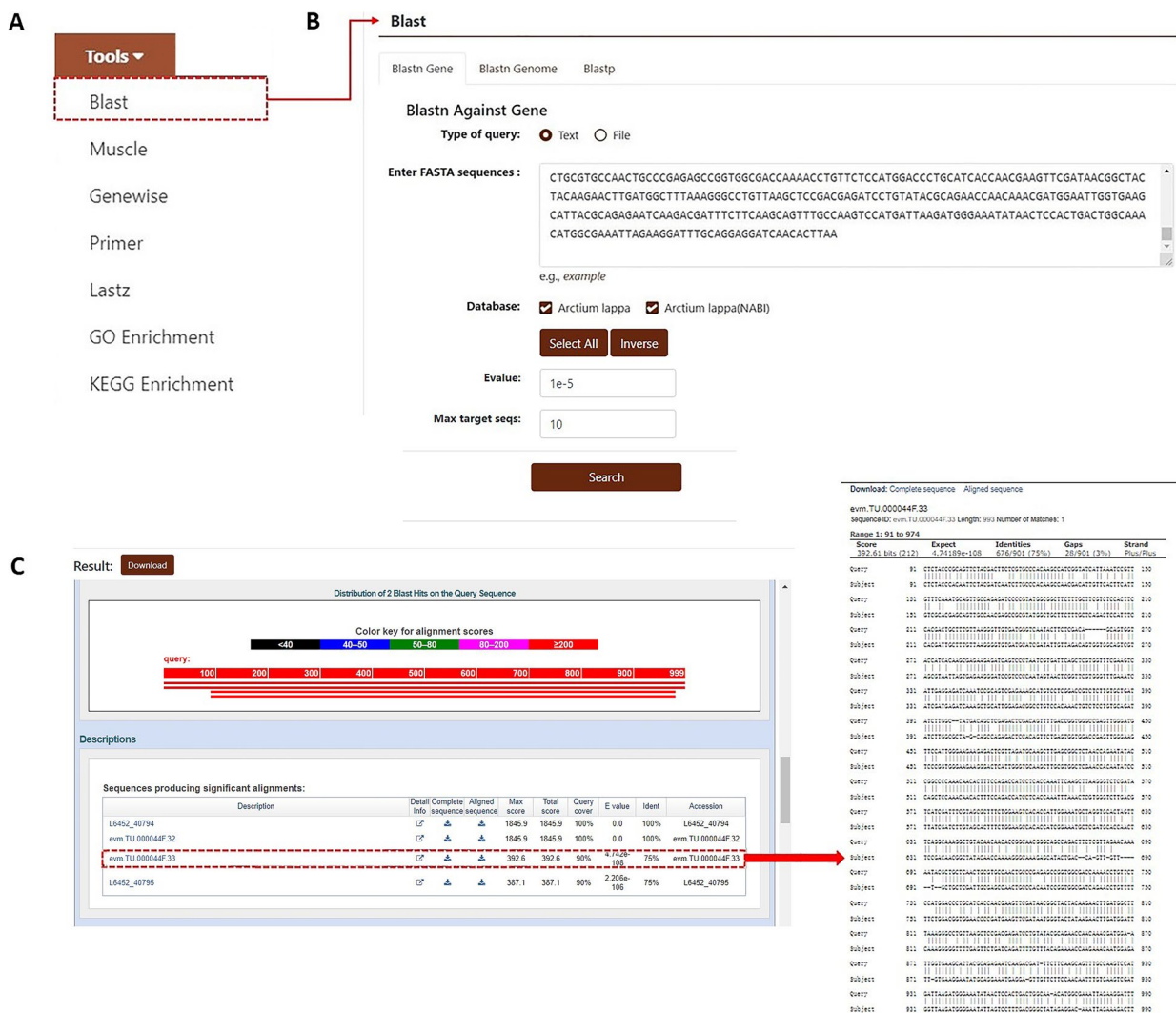


Fig. 7 Introduction to the “Blast” tool in the “Tools” module. **A** Sub-modules include “Blast”, “Muscle”, “Genewise”, “Primer”, “Lastz”, “GO Enrichment”, and “KEGG Enrichment”. **B** Nucleotide sequences in the database compared using “Blastn”. **C** Results of nucleotide sequence alignment

genomes in the database. The two kinds of alignment tools “Blastn” and “Blastp” can be used to perform nucleotide and protein sequence alignment. Users can click on “Blast”, and select “Blastn Gene”, “Blastn Genome”, and “Blastp” in the secondary navigation as required. They can enter the nucleotide or protein sequence in the FASTA format into the search box and click the “Search” button below to obtain the alignment results and download the sequence as required (Fig. 7). Users can obtain possible gene sequences by using this tool for comparison and screening if they want to research a certain type of burdock gene.

Users can click on “Muscle” to perform multiple sequence alignment analysis of nucleotide or protein sequences. The database also supports the direct

construction of the phylogenetic tree of alignment results using the maximum likelihood method (Fig. 8A). Users are not limited to burdock-related sequences included in the database but can also paste the sequences of other species into the search bar according to their experimental needs and then obtain the tree diagram results by following the above steps. For homologous protein prediction, users must click the “Genewise” option, enter the nucleotide and protein sequences in the FASTA format into the two search boxes, respectively, and click the “Run” button to obtain the results (Fig. 8B). This tool is also not limited to sequences in the database.

If users require to amplify the burdock sequence, they can directly use the “Primer” function in the burdock database to design primers. Users can fill in the scope

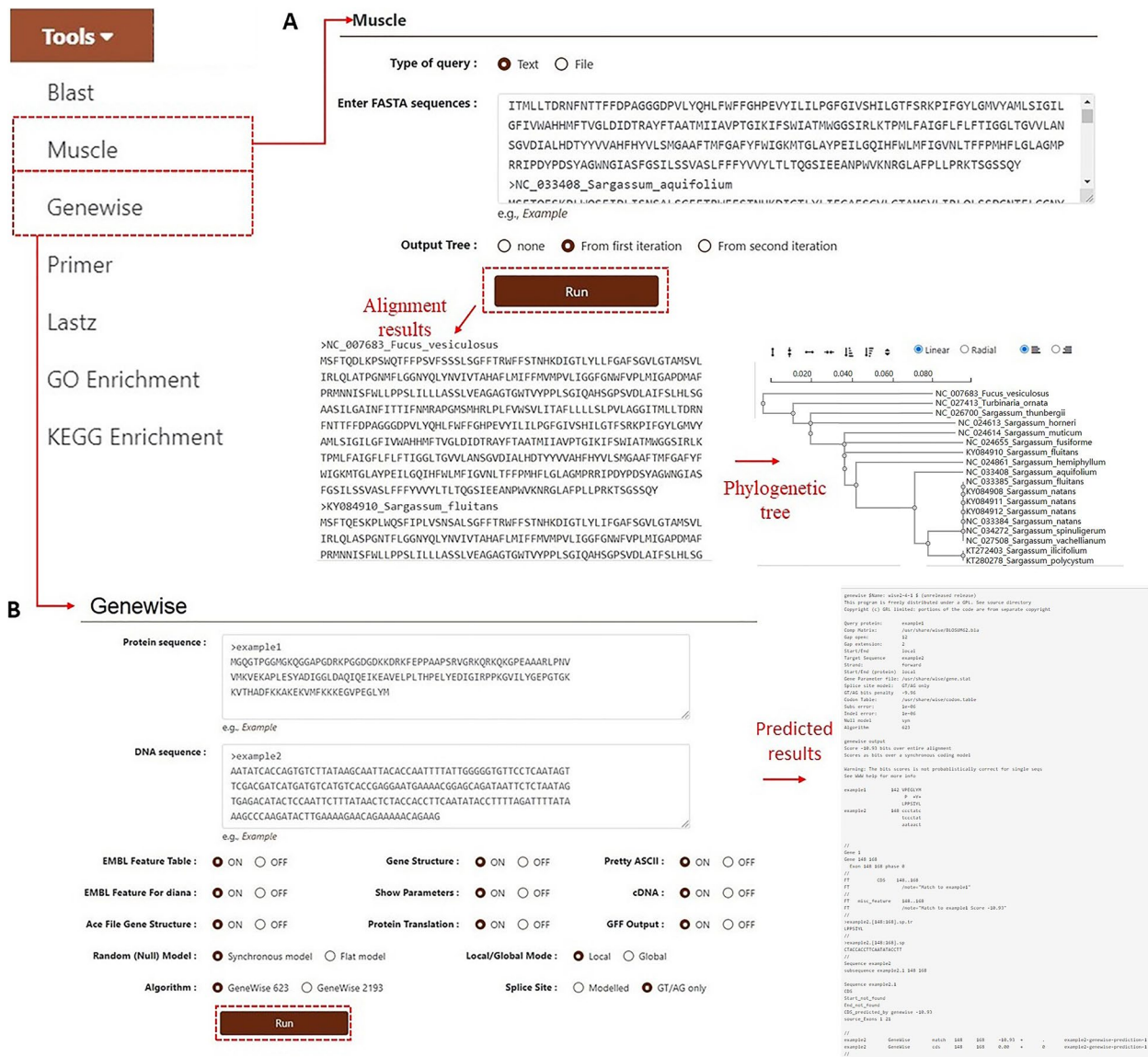


Fig. 8 A. “Muscle” displaying alignment results and the phylogenetic tree. B. “GeneWise” showing the prediction of homologous proteins

of the amplified sequence and the length and other requirements of the primer in the search box, and click the “Search” button (Fig. 9). If other species sequences require to be amplified, this function can also be used to design primers. Users can select the “Seq” radio button on the first line “Type of query”, and a FASTA search box will appear below. They can then enter the sequence, fill in the primer length and other requirements, and click the “Search” button. Users can only use this database to complete the search of the burdock target sequence, and then directly complete the comparison and primer design.

The burdock database can also meet the requirements of large-scale genome analysis. Users can click on “Lastz” in the navigation bar and enter the gene sequence in the FASTA format into the two search boxes. They can then click the “Run” button to obtain the results of the gene collinearity analysis (Fig. 10A). This module can fulfill the comparison task of users by using target sequences of other species. To study the biosynthetic pathways of secondary metabolites, the database also provides GO/KEGG enrichment analysis tools [27–29]. Users can enter the filtered DEGs into the search box and click the “Run” button. The locations of these annotated genes in

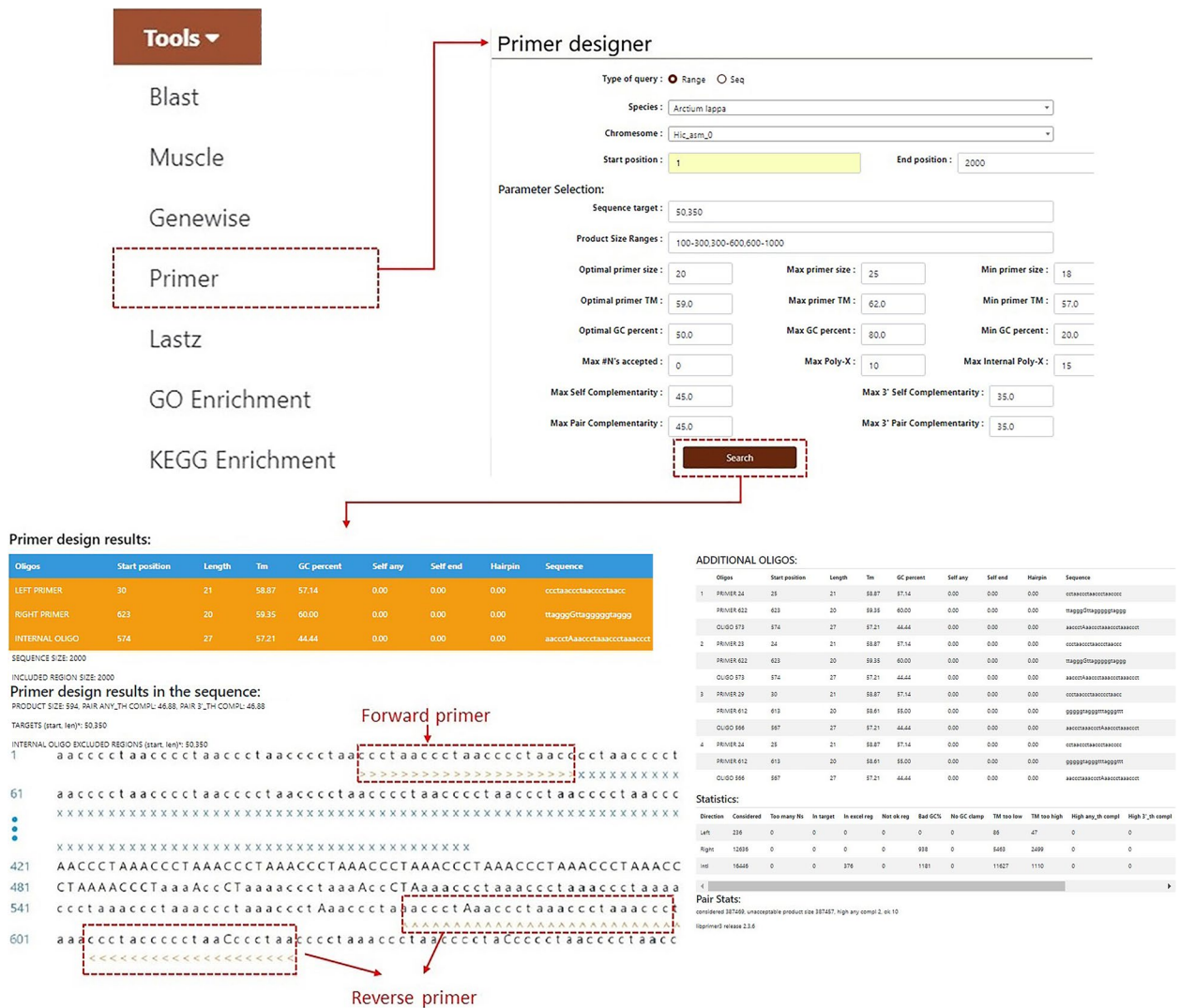


Fig. 9 Example of the use of the “Primer” tool

the GO/KEGG database are presented in both tabular and pictorial formats (Fig. 10B). Users can enter gene IDs obtained from other modules to get their analysis results in two protein databases. It can be used for the functional study of genes and the screening of key genes.

SRAP fingerprints, Associated microorganisms, Chemical, and Publications

In the “SRAP fingerprints” module, 18 pairs of primers with high polymorphism and clear bands were screened by Geng et al. (2019) for burdock varieties [32]. SRAP marker technology is widely used in many kinds of vegetables and horticultural species. Users can use this result to easily identify burdock germplasm, which facilitates the breeding of burdock. The “Associated microorganisms” module contains microbial information on burdock root, stem, leaf, fruit, and rhizosphere soil. By

high-throughput sequencing, 796,891 16S rRNA and 626,270 internal transcribed spacer reads were obtained from the samples. Using these reads, 922 bacterial species and 334 fungal species were detected. This page shows the distribution of bacteria and fungi in different tissues [33]. Users can use this data to study the formation mechanism of the secondary metabolites of burdock. The module of “Chemical” includes two sub-modules, “UPLC-Q-TOF-MS” and “Pharmaceutical Action”, of which, “UPLC-Q-TOF-MS” displays the identification results of chemical components in fruits, seeds, and pericarps of burdock by UPLC-Q-TOF-MS. From these, 31, 32, and 33 compounds were identified, and the content analysis of the key compounds in each part was included. This data can be used to understand the processes of burdock secondary metabolite accumulation [20, 34]. In the module of “Pharmacological Action”, the

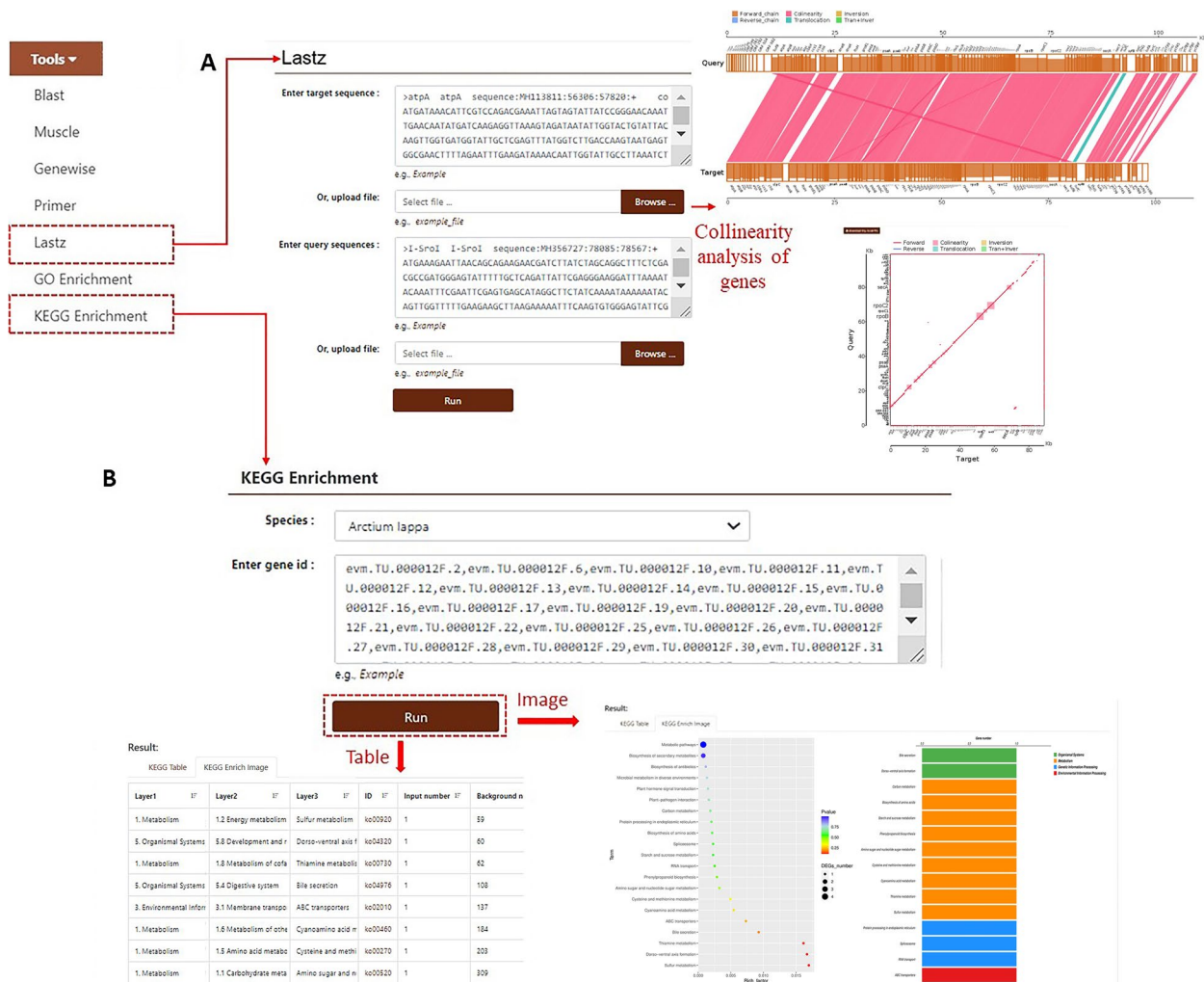


Fig. 10 A “Lastz” displaying the results of gene collinearity analysis. B Example of the use of “KEGG Enrichment”

pharmacological actions of various chemical components of burdock in recent years were summarized. The “Publications” module of the database provides 120 research papers on burdock microscopic identification, chemical composition, herbal research, molecular identification, ecological suitability studies, cultivation, and varieties. *The Study of Burdock in China*, a book edited by Professor Tinguo Kang, is available for easy perusal.

Conclusions

The burdock database is available at <http://210.22.121.250:41352/>. It aims to promote research of burdock in bioinformatics, molecular pharmacognosy, and genetic engineering. We hope to provide data support for research on the biosynthetic pathway of arctiin and other secondary metabolites as well as the genetic map of burdock. The multi-omics data in the database laid a foundation for the molecular breeding of burdock. It is

convenient for experts and scholars to discover the key enzyme genes for the regulation and metabolism of drug efficacy and quality traits of burdock. It significantly improved the operability of polygenic traits improvement and the efficiency of new variety breeding. The establishment of the multi-omics database is of great significance to the understanding of the molecular mechanism of burdock growth and development and its position in the evolutionary process. With the development of bioscience and various technologies, the burdock database will be gradually enriched and provide more reference value for scholars. For now, you are welcome to browse and use the database for burdock research.

Methods

The Burdock database was deployed in Ubuntu 16.04 operation system and developed by AKKA 2.6.5(web-server), MySQL 5.7(database server), Scala 2.13.2 and

SBT 1.3.9. All data in database were managed and stored by MySQL Database Management System. The query function was enforced based on Slick 3.3.2 middleware tier. To visualize the genome, we used the Jbrowser 1.16.6. The website interface components were designed and implemented by the Bootstrap 4.6.0 and Play Framework 2.8.7. The website has been tested in several popular web browsers, including Firefox, Google Chrome and Internet Explorer.

Acknowledgements

We thank Mr. Weiqi Xue and Mr. Zequn Zheng (VGsoft Team, China) for technical assistances in website construction.

Authors' contributions

LX, YYY, WXM, DQD and TGK conceived the idea of the study; YYS and CB collected the transcriptome and variety information; YYS and YPX sorted out genome and Associated microorganism data; HFX, YYY and WJH sorted out relevant paper data; YYS completed the article writing; LX, YYY, DQD and TGK revised the paper. all authors participated in the study as the result of their joint discussion. The author(s) read and approved the final manuscript.

Funding

This research was funded by 2019 Liaoning Provincial Department of Education Scientific Research Project, China [No. L201942], National Key Research and Development in the 13th Five-Year Plan [No. 2018YFC1708200], Major Special Fund for Science and Technology of Inner Mongolia Autonomous Region [No. 2019ZD004], Natural Science Fund Project of Liaoning Province [No. 2020-MS-224], Liaoning BaiQianWan Talents Program [No.2021A039], Key project at central government level: The ability establishment of sustainable use for valuable Chinese medicine resources [2060302] and the National Natural Science Foundation of China [General Program, Grant Numbers 81874338, 81773852].

Availability of data and materials

The database and web interface can be accessed at <http://210.22.121.250:41352/>.

Declarations

Ethics approval and consent to participate

This study was approved by the School of Pharmacy of Liaoning University of Traditional Chinese Medicine in May 2018. All operations were carried out following the guidelines in the Specification on Good Agriculture and Collection Practices for Medicinal Plants (GACP; Number: T/CCCMHPIE 2.1–2018). *A. lappa* is not a member of the national key protected wild plants, so collecting it is not against the Regulations of the People's Republic of China on Wild Plant Protection. Article 5 of the regulations stipulates that the state encourages and supports scientific research on wild plants and on-site and ex-situ protection of wild plants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 November 2022 Accepted: 27 January 2023

Published online: 09 February 2023

References

- Jaiswal R, Kuhnert N. Identification and characterization of five new classes of chlorogenic acids in burdock (*Arctium lappa* L.) roots by liquid chromatography/tandem mass spectrometry. *Food Funct.* 2011;2(1):63–71.
- Commission CP. Pharmacopoeia of People's Republic of China part 1. Beijing: China Medical Science Press; 2020. (in Chinese).
- Xu L, Dou D, Kang T. Textual Research and Genuine Changes of *Fructus arctii*. *Res Pract Chin Med.* 2010;24(01):24–6 (in Chinese).
- Ferracane R, Graziani G, Gallo M, Fogliano V, Ritieni A. Metabolic profile of the bioactive compounds of burdock (*Arctium lappa*) seeds, roots and leaves. *J Pharm Biomed Anal.* 2010;51(2):399–404.
- Gao Q, Yang M, Zuo Z. Overview of the anti-inflammatory effects, pharmacokinetic properties and clinical efficacies of arctigenin and arctiin from *Arctium lappa* L. *Acta Pharmacol Sin.* 2018;39(5):787–801.
- Yuan M. Literature study on medicinal properties and efficacy of *Arctium lappa* L. Master: Beijing University of Traditional Chinese Medicine; 2012. (in Chinese).
- Ma T, Chen Y, Cheng S, Sun X, Fu R, Dong H. Research Progress of *Fructus arctii*. *J Liaoning Univ TCM.* 2018;20(09):113–6 (in Chinese).
- Garnatje T, Canela MA, Garcia S, Hidalgo O, Pellicer J, Sanchez-Jimenez I, Siljak-Yakovlev S, Vitales D, Valles J. GSAD: a genome size in the Asteraceae database. *Cytometry A.* 2011;79(6):401–4.
- Shen Q, Zhang L, Liao Z, Wang S, Yan T, Shi P, Liu M, Fu X, Pan Q, Wang Y, et al. The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis. *Mol Plant.* 2018;11(6):776–88.
- Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikat S, Song C, Xia L, Froenicke L, Lavelle DO, Truco MJ, et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat Commun.* 2017;8:14953.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature.* 2017;546(7656):148–52.
- Acquadro A, Portis E, Valentino D, Barchi L, Lanteri S. "Mind the Gap": Hi-C Technology Boosts Contiguity of the Globe Artichoke Genome in Low-Recombination Regions. *G3 (Bethesda).* 2020;10(10):3557–64.
- Liu B, Yan J, Li W, Yin L, Li P, Yu H, Xing L, Cai M, Wang H, Zhao M, et al. *Mikania micrantha* genome provides insights into the molecular mechanism of rapid growth. *Nat Commun.* 2020;11(1):340.
- Wu Z, Liu H, Zhan W, Yu Z, Qin E, Liu S, Yang T, Xiang N, Kudrna D, Chen Y, et al. The chromosome-scale reference genome of safflower (*Carthamus tinctorius*) provides insights into linoleic acid and flavonoid biosynthesis. *Plant Biotechnol J.* 2021;19(9):1725–42.
- Xu X, Yuan H, Yu X, Huang S, Sun Y, Zhang T, Liu Q, Tong H, Zhang Y, Wang Y, et al. The chromosome-level *Stevia* genome provides insights into steviol glycoside biosynthesis. *Hortic Res.* 2021;8(1):129.
- Fan W, Wang S, Wang H, Wang A, Jiang F, Liu H, Zhao H, Xu D, Zhang Y. The genomes of chicory, endive, great burdock and yacon provide insights into Asteraceae palaeo-polyloidization history and plant inulin production. *Mol Ecol Resour.* 2022;22:3124–40.
- He S, Dong X, Zhang G, Fan W, Duan S, Shi H, Li D, Li R, Chen G, Long G, et al. High quality genome of *Erigeron breviscapus* provides a reference for herbal plants in Asteraceae. *Mol Ecol Resour.* 2021;21(1):153–69.
- Yang Y, Li S, Xing Y, Zhang Z, Liu T, Ao W, Bao G, Zhan Z, Zhao R, Zhang T, et al. The first high-quality chromosomal genome assembly of a medicinal and edible plant *Arctium lappa*. *Mol Ecol Resour.* 2022;22(4):1493–507.
- Wei J, Zhang S, Shi Y, Shan C, Ma K, Wu J. Transcriptome analysis reveals key enzyme genes involved in lignin biosynthesis pathway in *Arctium lappa*. *Chin Tradit Herb Drugs.* 2020;51(16):4300–7 (in Chinese).
- Li S. Accumulation of arctiin in burdock fruits at different growth stages and analysis of related metabolic genes. 2021. master. Liaoning University of Traditional Chinese Medicine. (in Chinese).
- Wang Y, Xu P, Wu X, Wu X, Wang B, Huang Y, Hu Y, Lin J, Lu Z, Li G. Gourd-Base: a genome-centered multi-omics database for the bottle gourd (*Lagenaria siceraria*), an economically important cucurbit crop. *Sci Rep.* 2018;8(1):3604.
- Ren H, He Y, Qi X, Zheng X, Zhang S, Yu Z, Hu F. The bayberry database: a multiomic database for *Myrica rubra*, an important fruit tree with medicinal value. *BMC Plant Biol.* 2021;21(1):452.
- Dehe W, Weiliang F, Xiaolong G, Kai W, Siyu Z, Zonggui C, Danyang L, Kun W, Yuxian Z, Yu Z. MaGenDB: a functional genomics hub for Malvaceae plants. *Nucleic Acids Res.* 2020;48(D1):D1076–84.
- Su Y, Fu J, Sun Y, Luo Y, Wang R, Li Y, Cao Q, Liu Y. Genome Size Estimation and Ploidy Analysis of *Arctium lappa* L. Germplasm Resources. *Molecular Plant Breeding:1–8*[2023-01-30]. (in Chinese).

25. Zhang A, Hu C, Liu P, Yan C. Breeding of New Burdock Variety Xinlin 1. *J Changjiang Veg.* 2004;08:48–9 (in Chinese).
26. He H, Gan X, Lu L, Wei B. Nutritional quality analysis of different burdock varieties roots. *Chin Veg.* 2011;06:73–5 (in Chinese).
27. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49(D1):D545–51.
28. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 2019;28(11):1947–51.
29. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
30. Xing YP, Xu L, Chen SY, Liang YM, Wang JH, Liu CS, Liu T, Kang TG. Comparative analysis of complete chloroplast genomes sequences of *Arctium lappa* and *A. tomentosum*. *Biologia plantarum.* 2019;63:565–74.
31. Zhang DC, Xing YP, Xu L, Zhao R, Yang YY, Zhang TT, Li SN, Kang T-G. The complete mitochondrial genome of *Arctium lappa* (Campanulales, Asteraceae). *Mitochondrial DNA Part B.* 2020;5(2):1722–3.
32. Geng G. Identification of Germplasm Resources and Development of High-efficiency Cultivation Technology in Burdock (*Arctium lappa* L.). 2019. master. Nanjing Agricultural University. (in Chinese).
33. Xing Y, Yang Y, Xu L, Hao N, Zhao R, Wang J, Li S, Zhang D, Zhang T, Kang T. The Diversity of Associated Microorganisms in Different Organs and Rhizospheric Soil of *Arctium lappa* L. *Curr Microbiol.* 2020;77(5):746–54.
34. Zhang D. Preliminary study on quality evaluation and mechanism of *Fructus arctii* by microscopic characteristic index method. 2021. master. Liaoning University of Traditional Chinese Medicine. (in Chinese).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

