


DATABASE

Open Access



# TCM-Blast for traditional Chinese medicine genome alignment with integrated resources

Zhao Chen<sup>1,2\*</sup> , Jing Li<sup>1,2</sup>, Ning Hou<sup>1,2</sup>, Yanling Zhang<sup>1,2</sup> and Yanjiang Qiao<sup>1,2\*</sup>

## Abstract

The traditional Chinese medicine (TCM) genome project aims to reveal the genetic information and regulatory network of herbal medicines, and to clarify their molecular mechanisms in the prevention and treatment of human diseases. Moreover, the TCM genome could provide the basis for the discovery of the functional genes of active ingredients in TCM, and for the breeding and improvement of TCM. The traditional Chinese Medicine Basic Local Alignment Search Tool (TCM-Blast) is a web interface for TCM protein and DNA sequence similarity searches. It contains approximately 40G of genome data on TCMs, including protein and DNA sequence for 36 TCMs with high medical value. The development of a publicly accessible TCM genome alignment database hosted on the TCM-Blast website (<http://viroblast.pungentdb.org.cn/TCM-Blast/viroblast.php>) has expanded to query multiple sequence databases to obtain TCM genome data, and provide user-friendly output for easy analysis and browsing of BLAST results. The genome sequencing of TCMs helps to elucidate the biosynthetic pathways of important secondary metabolites and provides an essential resource for gene discovery studies and molecular breeding. The TCMs genome provides a valuable resource for the investigation of novel bioactive compounds and drugs from these TCMs under the guidance of TCM clinical practice. Our database could be expanded to other TCMs after the determination of their genome data.

## Background

Whole-genome sequencing of the plants that form the basis of traditional Chinese medicine (TCM) is an important means for gene discovery and cultivation, synthetic biology, drug discovery and molecular breeding involving TCMs [1–4]. The genomic sequence provides a valuable resource not only for fundamental and applied research, but also for evolutionary and comparative genomics analyses, particularly in TCMs [5–9].

Experimental and clinical studies have demonstrated that TCMs have a wide range of pharmacological properties such as anti-inflammatory, antiviral, antimicrobial, antioxidative, antifungal, antithrombotic,

antihyperlipidemic, analgesic, antidiabetic, antidepressant, antiasthma and anticancer activities as well as immunomodulatory, antidiabetic, gastroprotective, hepatoprotective, neuroprotective and cardioprotective effects [10–18]. Genome sequencing and its annotations provide an essential resource for TCM improvement through molecular breeding [19–21] and for the discovery of useful genes for engineering bioactive compounds through synthetic biology approaches [1, 22–24]. The availability of these genomic resources will facilitate the discovery of medicinally and nutritionally important genes, the genetic improvement of TCMs [7, 21, 25] and the identification of novel drug candidates [26].

The Herbal Medicine Omics Database (<http://herbalplant.ynau.edu.cn/html/Genomes/>) has collected only 23 published genomes of medicinal herbs and there has been no continued update of the increased data since 2019. Only 14 kinds of medicinal plant genome data were provided in the Medicinal Plant Genomics Resource (<http://medicinalplantgenomics.msu.edu>).

\*Correspondence: zhaochen@bucm.edu.cn; yjqiao@bucm.edu.cn

<sup>1</sup> School of Chinese Materia Medica, Beijing University of Chinese Medicine, Yangguang South Avenue, Fangshan District, Beijing 102488, China

<sup>2</sup> Research Center of TCM-Information Engineering, State Administration of Traditional Chinese Medicine of The Peoples Republic of China, Yangguang South Avenue, Fangshan District, Beijing 102488, China



BLAST against plant genomes data ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&\\_TYPE=BlastSearch&BLAST\\_SPEC=Plants\\_MV&LINK\\_LOC=blasttab&LAST\\_PAGE=blastp](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&_TYPE=BlastSearch&BLAST_SPEC=Plants_MV&LINK_LOC=blasttab&LAST_PAGE=blastp)) included few types of medicinal plants, and the genome comparison of the most common edible plants was provided).

### Construction and content

Genome data of TCMs were originated from the Herbal Medicine Omics Database (<http://herbalplant.ynau.edu.cn/html/Genomes/>), the Medicinal Plant Genomics Resource (<http://medicinalplantgenomics.msu.edu>), and the BIG Data Center in Beijing Institute of Genomics, Chinese Academy of Sciences (<http://bigd.big.ac.cn/gsa/statistics>).

The genome data of Chinese medicinal materials originating from unlabeled references are from <http://medicinalplantgenomics.msu.edu/>, <http://bigd.big.ac.cn/gsa/statistics>.

The deployment strategy for TCM-Blast involves instantiating a provided Viroblast [27] that bundles the core components for TCM genome alignment. A user-friendly web interface to search the database has been implemented in PHP 7.0.32 (<http://www.php.net>) and deployed on an Apache 2.4.18 web server (<http://www.apache.org/>) and MySQL database server (<https://www.mysql.com/>) with Ubuntu 16.04 server (<http://mirrors.aliyun.com/ubuntu-releases/16.04/>). TCM-Blast had 36 TCMs genome datasets.

The information regarding TCM genome datasets is summarized in an online at the TCM-Blast website. The TCM genome data used in TCM-Blast were collected from the Herbal Medicine Omics Database (<http://herbalplant.ynau.edu.cn/html/Genomes/>), the Medicinal Plant Genomics Resource (<http://medicinalplantgenomics.msu.edu>), and the BIG Data Center in Beijing Institute of Genomics (<http://bigd.big.ac.cn/gsa/statistics>) (the further details on the genome data sources for the thirty-six TCMs, see Table 1). These data resources have been published in professional journals and plant gene databases by academic institutions or government departments merged with plant gene databases, with abundant data sources and reliable data quality. In addition to other data resources, this database in our study has the following advantages: 1) this database is currently the largest Chinese medicine genome database; 2) this database includes the plant genetic data of Chinese medicine sources; and 3) this database provides support for the TCM breeding, cultivation of TCMs and the discovery of active ingredients in TCMs.

## Utility and discussion

### Overview of TCM-Blast

We have developed TCM-Blast, a web-based database for TCM genome alignment (Fig. 1). TCM-Blast offers an interface to choose from TCM genome databases including TCM protein and DNA sequence datasets, which provide query functions with BLAST implementation [40]. TCM-Blast currently contains approximately 40 GB of TCM genome data, including the proteins and DNA sequences of 36 TCMs.

### The main functions of TCM-Blast

The user can directly enter the query sequence directly by pasting into the query box or by uploading the sequence as a FASTA file from a local file. TCM-Blast provides multiple TCM sequence databases. Users can then select specific TCM genome databases to run different programs (blastn, blastp, blastx, tblastn, tblastx). TCM-Blast consists of five general BLAST form types [27, 41–43] for TCM genome data:

blastn: search TCM nucleotide databases using a nucleotide query.

blastp: search TCM protein databases using a protein query.

blastx: search TCM protein databases using a translated nucleotide query

tblastn: search TCM translated nucleotide databases using a protein query.

tblastx: search TCM translated nucleotide databases using a translated nucleotide query

TCM-Blast provides an optional search function for advanced users who need to collect more specific information (Fig. 2) with the ability to set different parameters, such as the expected threshold, word size, max target sequences, etc., to glean more specific information for users. The TCM-Blast sequence alignment results of the TCM genome sequence are displayed in the summary table, which contains the query sequence name, subject sequence name, subject source database, position score, identity percentage, and E value (Fig. 3).

### A case study of this database

For example, the user can select the *Salvia Miltiorrhiza* protein database with the programs blastp and obtain their expected BLAST results by inputting the protein sequence. In Fig. 4, the user has input the protein sequence fragment:

**Table 1** Data sources of thirty-six TCM genomes

Latin name	Pin Yin	Genome sequencing method	Reference
<i>Dendrobium Officinale</i>	Tiepishihu	combining the second-generation Illumina Hiseq 2000 and third-generation PacBio sequencing technologies	Ref [8]
<i>Ginkgo Biloba</i>	Yinxing	Hiseq 2000/4000 platform	Ref [5]
<i>Erigeron Breviscapusd</i>	Dengzhanhua	Illumina sequencing and PacBio single-molecular real-time sequencing on the Illumina HiSeq platform	Ref [24]
<i>Panax Ginseng</i>	Sanqi	Illumina paired-end libraries for the whole-genome sequencing	Ref [26]
<i>Eucommia Ulmoides</i>	Duzhong	Illumina HiSeq, MiSeq short-read sequencing, and PacBio single molecular long-read sequencing	Ref [28]
<i>Punica Granatum</i>	Shiliu	Illumina paired-end reads of libraries	Ref [29]
<i>Dioscorea Routundata</i>	Shanyao	Illumina MiSeq platform, HiSeq 2500 platform	Ref [30]
<i>Ginseng</i>	Renshen	paired-end sequencing on the HiSeq X-Ten platform (Illumina)	Ref [21]
<i>Boea Hygrometrica</i>	Niuercao	whole-genome shotgun approach (Illumina HiSeq and Roche 454 platforms)	Ref [31]
<i>Jatropha Curcas</i>	Mafengshu	Illumina GAll and HiSeq	Ref [7]
<i>Glycyrrhiza Uralensis</i>	Gancao	short reads from Illumina and long reads from Pacific Biosciences sequencing	Ref [1]
<i>Moringa Oleifera</i>	Lamu	Illumina Hiseq2500TM	Ref [32]
<i>Salvia Miltiorrhiza</i>	Danshen	Illumina sequencing and PacBio sequencing,	Ref [33]
<i>Cannabis Sativa</i>	Dama	Illumina mate-pair library construction and sequencing	Ref [34]
<i>Mentha Longifoli</i>	Bohe	Illumina sequencing, Pacific Biosystems sequencing	Ref [22]
<i>Macleaya Cordata</i>	Boluohui	paired-end sequences on HiSeq 2000	Ref [35]
<i>Calotropis Gigantea</i>	Niuguajiao	Illumina HiSeq 2500	Ref [36]
<i>Rhodiola Rosea</i>	Hongjingtian	Illumina HiSeq 2000/4000 platform using a whole genome shotgun sequencing (WGS) strategy	Ref [37]
<i>Capsicum annum</i>	Lajiao	Illumina HiSeq 2500	
<i>Lilium</i>	Baihe	Illumina HiSeq X Ten	
<i>Tupaia belangeri</i>	Baihuabaihe	Illumina HiSeq 2000	
<i>Arctium lappa</i>	Niubang	Illumina HiSeq X Ten	
<i>Anemone flaccida</i>	Ezhangcao	Illumina HiSeq 2000	
<i>Atropa belladonna</i>	Dianqie	RNA-seq for expression abundances	
<i>Digitalis purpurea</i>	Zihuayangdihuang	RNA-seq for expression abundances	
<i>Dioscorea villosa</i>	Changroumaoshuyu	RNA-seq for expression abundances	
<i>Echinacea purpurea</i>	Zizhuiyu	RNA-seq for expression abundances	
<i>Hoodia gordonii</i>	Hutieyaxianrenzhang	RNA-seq for expression abundances	
<i>Hypericum perforatum</i>	Guanyejiinsitao	RNA-seq for expression abundances	
<i>Panax quinquefolius</i>	Xiyangshen	RNA-seq for expression abundances	
<i>Rauvolfia serpentina</i>	yinduluofumu	RNA-seq for expression abundances	
<i>Rosmarinus officinalis</i>	Midiexiang	RNA-seq for expression abundances	
<i>Valeriana officinalis</i>	Xiecao	RNA-seq for expression abundances	
<i>Camptotheca acuminata</i>	Huaxishu	Illumina sequencing platform	Ref [38]
<i>Catharanthus roseus</i>	Changchunhua	whole genome shotgun sequencing approach	Ref [39]
<i>Lepidium Meyenii</i>	Maca	Illumina HiSeq 2500 platform yielded 1.88 billion reads in ten paired-end libraries	Ref [40]

“MEKKQEDEKKT KLQGLPVDTS PYTQYKDLLD-DYKKQAYGTEGHLQPNPGRGAAASTDAPTTTAAAD-DPNKQLSSTDAINRQGV” in the “Enter query sequences” box; selected the *Salvia Miltiorrhiza* protein database; and obtained the BLAST result by clicking the “Basic Search” button. The top score of this search was “evm.model.C153610.1” subject, indicating

that the input sequence fragment has high similarity to the *Salvia Miltiorrhiza* protein. For more detailed use cases for this database, please refer to the [Supplementary file](#).

In the future, we will collect more Chinese medicine genome data to provide data support for Chinese medicine research.

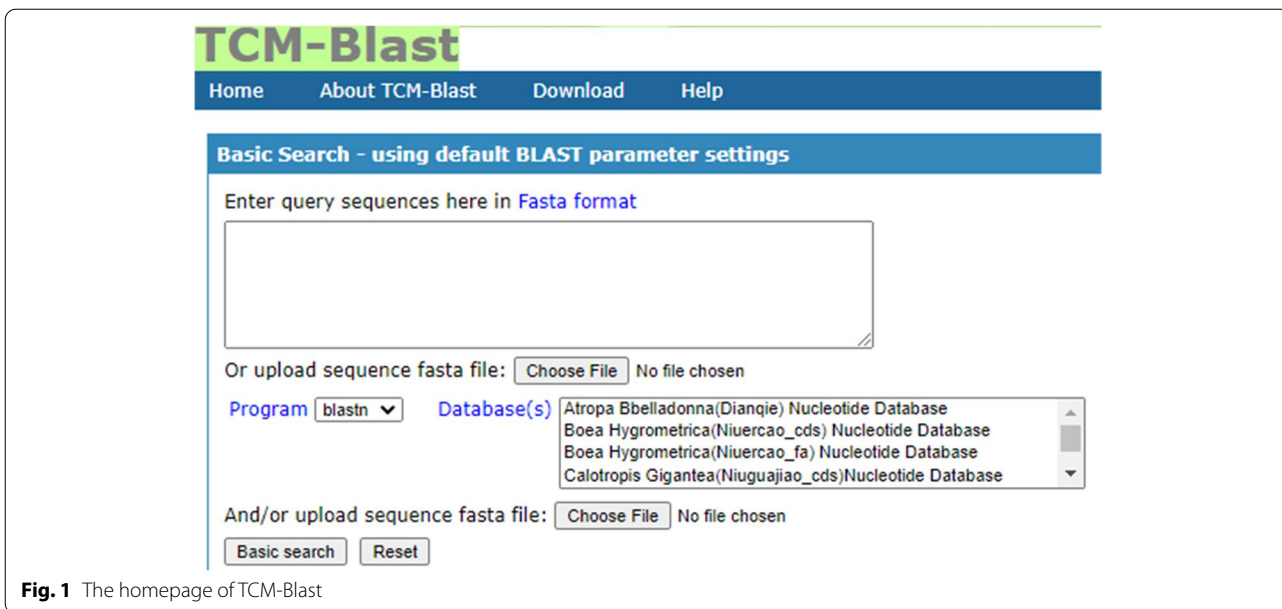


Fig. 1 The homepage of TCM-Blast

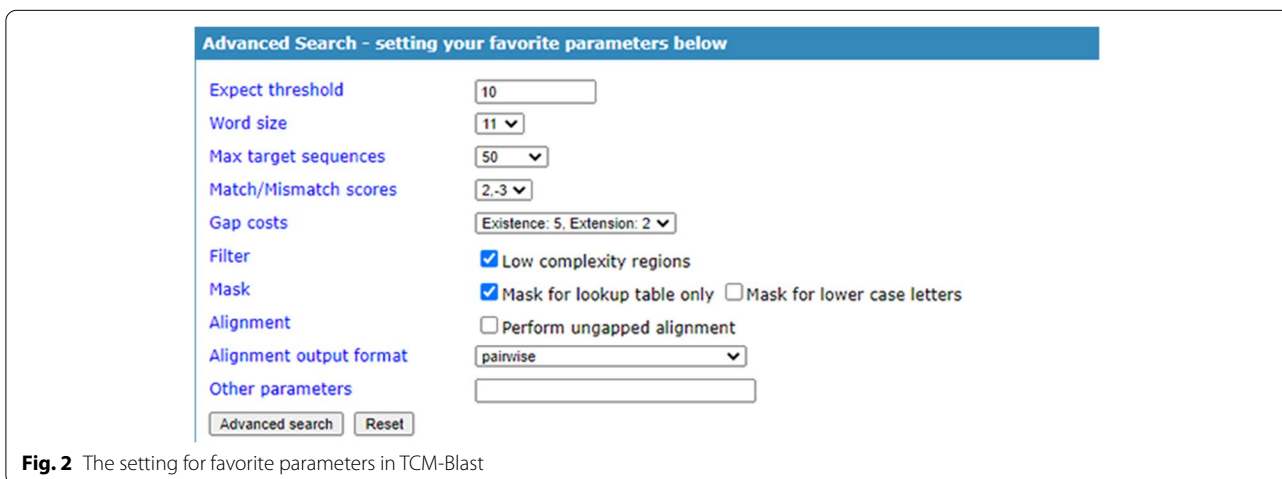


Fig. 2 The setting for favorite parameters in TCM-Blast

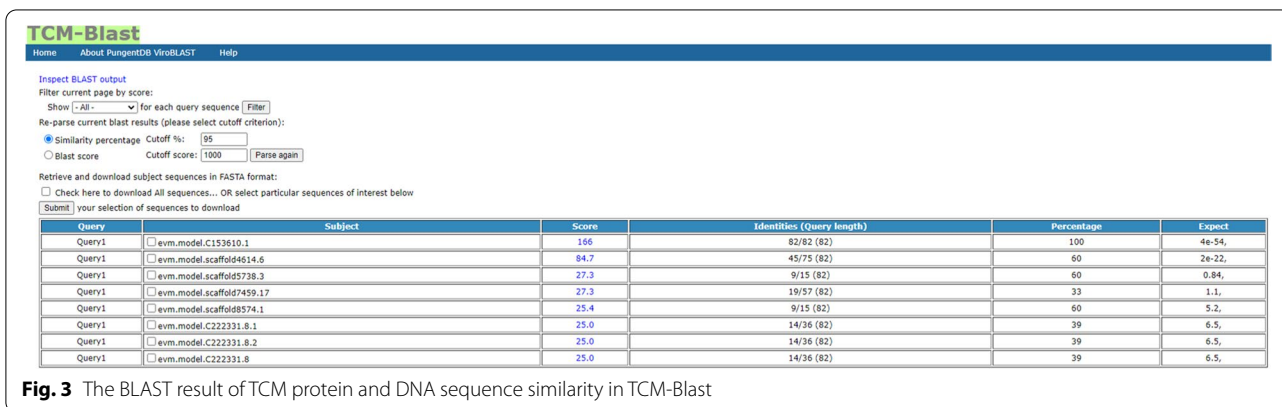
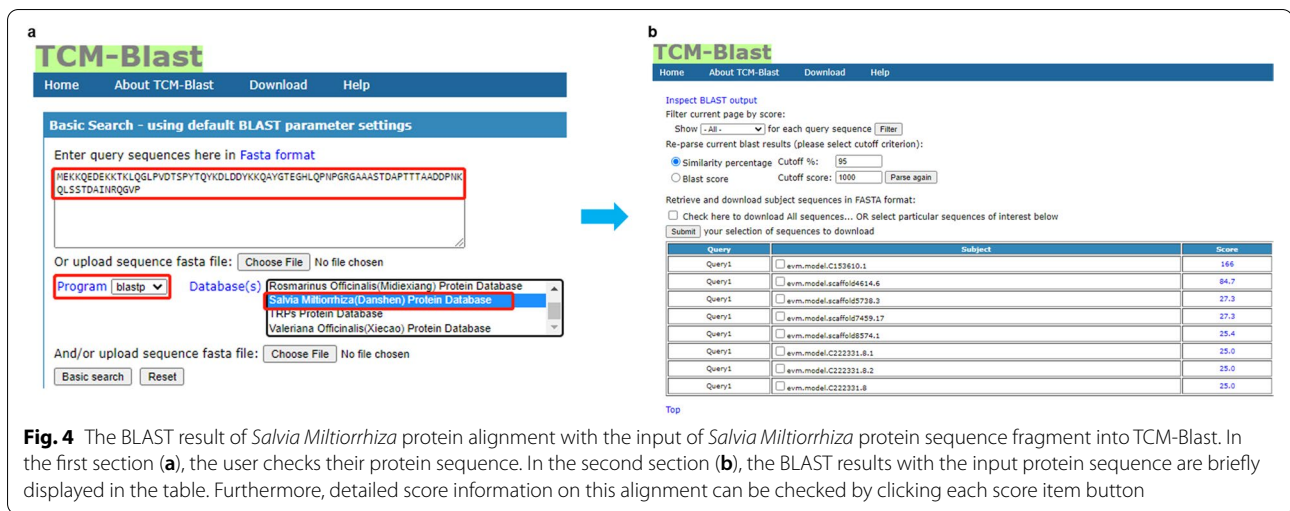


Fig. 3 The BLAST result of TCM protein and DNA sequence similarity in TCM-Blast



**Fig. 4** The BLAST result of *Salvia Miltiorrhiza* protein alignment with the input of *Salvia Miltiorrhiza* protein sequence fragment into TCM-Blast. In the first section (a), the user checks their protein sequence. In the second section (b), the BLAST results with the input protein sequence are briefly displayed in the table. Furthermore, detailed score information on this alignment can be checked by clicking each score item button

## Conclusions

Here, we reported a database of TCM-Blast database that integrates several database resources and markedly improves the efficiency of TCM genomic research. This database will allow users to perform batch sequence searches against integrated TCM genomic sequence databases. Therefore, TCM-Blast provided comprehensive Chinese medicine genome resource data on TCM scientific research and eliminates the latent redundancy occurring in other platforms.

## Abbreviations

TCM: Traditional Chinese medicine; DNA: Deoxyribonucleic acid; TCM-Blast: Traditional Chinese medicine Basic Local Alignment Search Tool.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-03096-1>.

**Additional file 1: Figure S1.** Setting of protein sequence alignment options with *Glycyrrhiza Uralensis* protein database through the program of 'blastp'. **Figure S2.** BLAST result of protein sequence alignment with *Glycyrrhiza Uralensis* protein database by inputting the query protein sequence. **Figure S3.** Setting of protein sequence alignment options with *Glycyrrhiza Uralensis* Nucleotide Database by the program of 'tblastn'. **Figure S4.** BLAST result of protein sequence alignment with *Glycyrrhiza Uralensis* protein database by the program of 'tblastn'. **Figure S5.** Setting of nucleotide sequence alignment options with *Glycyrrhiza Uralensis* Nucleotide Database through the program of 'blastn'. **Figure S6.** BLAST result of nucleotide sequence alignment with *Glycyrrhiza Uralensis* nucleotide Database via the program of 'blastn'. **Figure S7.** Setting of nucleotide sequence alignment options with *Glycyrrhiza Uralensis* Protein (Gancao) Database through the program of 'blastx'. **Figure S8.** BLAST result of nucleotide sequence alignment with *Glycyrrhiza Uralensis* Protein (Gancao) Database via the program of 'blastx'.

## Acknowledgements

We would like to thank Xingde Ren, Xiaosa Shi valuable suggestions. This work was supported by the National Natural Science Foundation

of China (No.81430094) and China Postdoctoral Science Foundation (No.2020M670236).

## Authors' contributions

Y.Z. and Y.Q. conceived and designed the experiments; Z.C., J.L. and N.H. collected the data; Z.C. contributed reagents/materials/analysis tools; Z.C. constructed the database and wrote this manuscript, Y.Z. and Y.Q. revised the manuscript. All authors read and approved the final manuscript.

## Funding

Funding for open access charge: National Natural Science Foundation of China (No.81430094) and China Postdoctoral Science Foundation (No.2020M670236). The National Natural Science Foundation of China (No.81430094) substantial contributions to the conception or design of the work. China Postdoctoral Science Foundation (No.2020M670236) supports the analysis and interpretation of data for the work and the writing of this manuscript.

## Availability of data and materials

TCM-Blast is a free database and visualization tool open to all users with no login requirements and can be accessed at the following URL: <http://viroblast.pungentdb.org.cn/TCM-Blast/viroblast.php>. The web tool is functional on all modern web browsing environments including Google Chrome, Mozilla Firefox and Safari. All related species genomes data can be downloaded from <http://viroblast.pungentdb.org.cn/TCM-Blast/db>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

Received: 13 December 2020 Accepted: 14 June 2021  
Published online: 17 July 2021

## References

- Mochida K, Sakurai T, Seki H, Yoshida T, Takahagi K, Sawai S, et al. Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *Plant J*. 2017;89(2):181–94. <https://doi.org/10.1111/tpj.13385>.



2. Rehman F, Gong H, Li Z, Zeng S, Yang T, Ai P, et al. Identification of fruit size associated quantitative trait loci featuring SLAF based high-density linkage map of goji berry (*Lycium* spp.). *BMC Plant Biol.* 2020;20(1):1–18. <https://doi.org/10.1186/s12870-020-02567-1>.
3. Chen X, Li J, Wang X, Zhong L, Tang Y, Zhou X, et al. Full-length transcriptome sequencing and methyl jasmonate-induced expression profile analysis of genes related to patchouli biosynthesis and regulation in *Pogostemon cablin*. *BMC Plant Biol.* 2019;19(1):1–18. <https://doi.org/10.1186/s12870-019-1884-x>.
4. Chen S, Song J, Sun C, Xu J, Zhu Y, Verpoorte R, et al. Herbal genomics: examining the biology of traditional medicines. *Science.* 2015; 347(6219):S27–S29. <https://doi.org/10.1126/science.1276601>.
5. Guan R, Zhao Y, Zhang H, Fan G, Liu X, Zhou W, et al. Draft genome of the living fossil *Ginkgo biloba*. *Gigascience.* 2016;5(1):13742-016-0154-1. <https://doi.org/10.1186/s13742-016-0154-1>.
6. Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, et al. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Mol Plant.* 2017;10(10):1293–306. <https://doi.org/10.1186/s13742-016-0154-1>.
7. Wu P, Zhou C, Cheng S, Wu Z, Lu W, Han J, et al. Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant J.* 2015;81(5):810–21. <https://doi.org/10.1111/tpj.12761>.
8. Yan L, Wang X, Liu H, Tian Y, Lian J, Yang R, et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol Plant.* 2015;8(6):922–34. <https://doi.org/10.1016/j.molp.2014.12.011>.
9. Liu Y, Zeng S, Sun W, Wu M, Hu W, Shen X, et al. Comparative analysis of carotenoid accumulation in two goji (*Lycium barbarum* L. and *L. ruthenicum* Murr.) fruits. *BMC Plant Biol.* 2014;14(1):1–14. <https://doi.org/10.1186/s12870-014-0269-4>.
10. Chen Z, Cao Y, Zhang Y, Qiao Y. A novel discovery: holistic efficacy at the special organ level of pungent flavored compounds from pungent traditional Chinese medicine. *Int J Mol Sci.* 2019;20(3):752. <https://doi.org/10.3390/ijms20030752>.
11. Cheung F. TCM: made in China. *Nature.* 2011;480(7378):S82–3. <https://doi.org/10.1038/480S82a>.
12. Hosseinzadeh H, Nassiri-Asl M. Pharmacological effects of *Glycyrrhiza* spp. and its bioactive constituents: update and review. *Phytother Res.* 2015;29(12):1868–86. <https://doi.org/10.1002/ptr.5487>.
13. Jiang W-Y. Therapeutic wisdom in traditional Chinese medicine: a perspective from modern science. *Trends Pharmacol Sci.* 2005;26(11):558–63. <https://doi.org/10.1016/j.tips.2005.09.006>.
14. Qiu J. China plans to modernize traditional medicine. *Nature.* 2007;446:590–1. <https://doi.org/10.1038/446590a>.
15. Science AftAo. The art and science of traditional medicine part 1: TCM today—a case for integration. *Science.* 2014;346(6216):1569. <https://doi.org/10.1126/science.1262161>.
16. Xiong X. Integrating traditional Chinese medicine into Western cardiovascular medicine: an evidence-based approach. *Nat Rev Cardiol.* 2015;12(6):374–374. <https://doi.org/10.1038/nrcardio.2014.177-c1>.
17. Tian P. Convergence: where west meets east. *Nature.* 2011;480(7378):S84–6. <https://doi.org/10.1038/480S84a>.
18. Zhao J, Jiang P, Zhang W. Molecular networks for the study of TCM pharmacology. *Brief Bioinform.* 2010;11(4):417–30. <https://doi.org/10.1093/bib/bbp063>.
19. Song C, Liu Y, Song A, Dong G, Zhao H, Sun W, et al. The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of chrysanthemum flowers and medicinal traits. *Mol Plant.* 2018;11(12):1482–91. <https://doi.org/10.1016/j.molp.2018.10.003>.
20. da Silva JAT, Jin X, Dobránszki J, Lu J, Wang H, Zolt G, et al. Advances in *Dendrobium* molecular research: applications in genetic variation, identification and breeding. *Mol Phylogenet Evol.* 2016;95:196–216. <https://doi.org/10.1016/j.ympev.2015.10.012>.
21. Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, et al. *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience.* 2017;6(11):gix093. <https://doi.org/10.1093/gigascience/gix093>.
22. Vining KJ, Johnson SR, Ahkami A, Lange I, Parrish AN, Trapp SC, et al. Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Mol Plant.* 2017;10(2):323–39. <https://doi.org/10.1016/j.molp.2016.10.018>.
23. Shen Q, Zhang L, Liao Z, Wang S, Yan T, Shi P, 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. <https://doi.org/10.1016/j.molp.2018.03.015>.
24. Yang J, Zhang G, Zhang J, Liu H, Chen W, Wang X, et al. Hybrid de novo genome assembly of the Chinese herbal fleabane *Erigeron breviscapus*. *Gigascience.* 2017;6(6):gix028. <https://doi.org/10.1093/gigascience/gix028>.
25. Zhang L, Li X, Ma B, Gao Q, Du H, Han Y, et al. The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol Plant.* 2017;10(9):1224–37. <https://doi.org/10.1016/j.molp.2017.08.013>.
26. Chen W, Kui L, Zhang G, Zhu S, Zhang J, Wang X, et al. Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. *Mol Plant.* 2017;10(6):899–902. <https://doi.org/10.1016/j.molp.2017.02.010>.
27. Deng W, Nickle DC, Learn GH, Maust B, Mullins JL. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics.* 2007;23(17):2334–6. <https://doi.org/10.1093/bioinformatics/btm331>.
28. Wuyun T-N, Wang L, Liu H, Wang X, Zhang L, Bennetzen JL, et al. The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis. *Mol Plant.* 2018;11(3):429–42. <https://doi.org/10.1016/j.molp.2017.11.014>.
29. Qin G, Xu C, Ming R, Tang H, Guyot R, Kramer EM, et al. The pomegranate (*Punica granatum* L.) genome and the genomics of punicalagin biosynthesis. *Plant J.* 2017;91(6):1108–28. <https://doi.org/10.1111/tpj.13625>.
30. Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu M, et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* 2017;15(1):1–20. <https://doi.org/10.1186/s12915-017-0419-x>.
31. Xiao L, Yang G, Zhang L, Yang X, Zhao S, Ji Z, et al. The resurrection genome of *Boea hygrometrica*: a blueprint for survival of dehydration. *Proc Natl Acad Sci.* 2015;112(18):5833–7. <https://doi.org/10.1073/pnas.1505811112>.
32. Tian Y, Zeng Y, Zhang J, Yang C, Yan L, Wang X, et al. High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. *Sci China Life Sci.* 2015;58(7):627–38. <https://doi.org/10.1007/s11427-015-4872-x>.
33. Zhang G, Tian Y, Zhang J, Shu L, Yang S, Wang W, et al. Hybrid de novo genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). *GigaScience.* 2015;4(1):s13742-015-0104-3. <https://doi.org/10.1186/s13742-015-0104-3>.
34. Van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, et al. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 2011;12(10):1–18. <https://doi.org/10.1186/gb-2011-12-10-r102>.
35. Liu X, Liu Y, Huang P, Ma Y, Qing Z, Tang Q, et al. The genome of medicinal plant *Macleaya cordata* provides new insights into benzyloisoquinoline alkaloids metabolism. *Mol Plant.* 2017;10(7):975–89. <https://doi.org/10.1016/j.molp.2017.05.007>.
36. Hoopes GM, Hamilton JP, Kim J, Zhao D, Wiegert-Rininger K, Crisovan E, et al. Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. *G3: Genes, Genomes, Genetics.* 2018;8(2):385–91. <https://doi.org/10.1534/g3.117.300331>.
37. Fu Y, Li L, Hao S, Guan R, Fan G, Shi C, et al. Draft genome sequence of the Tibetan medicinal herb *Rhodiola crenulata*. *Gigascience.* 2017;6(6):gix033. <https://doi.org/10.1093/gigascience/gix033>.
38. Zhao D, Hamilton JP, Pham GM, Crisovan E, Wiegert-Rininger K, Vaillancourt B, et al. De novo genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *GigaScience.* 2017;6(9):gix065. <https://doi.org/10.1093/gigascience/gix065>.
39. Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, Vaillancourt B, et al. Genome-guided investigation of plant natural product biosynthesis. *Plant J.* 2015;82(4):680–92. <https://doi.org/10.1111/tpj.12827>.
40. Zhang J, Tian Y, Yan L, Zhang G, Wang X, Zeng Y, et al. Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high-altitude adaptation in the central Andes. *Mol Plant.* 2016;9(7):1066–77. <https://doi.org/10.1016/j.molp.2016.04.016>.

41. Jones DT, Swindells MB. Getting the most from PSI-BLAST. *Trends Biochem Sci.* 2002;27(3):161–4. [https://doi.org/10.1016/S0968-0004\(01\)02039-4](https://doi.org/10.1016/S0968-0004(01)02039-4).
42. Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, et al. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 2001;29(14):2994–3005. <https://doi.org/10.1093/nar/29.14.2994>.
43. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(suppl\_2):W5–W9. <https://doi.org/10.1093/nar/gkn201>

### 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](https://biomedcentral.com/submissions)

