Chinese Herbal Medicines 17 (2025) 87–93

Contents lists available at ScienceDirect

Chinese Herbal Medicines

journal homepage: www.elsevier.com/locate/chmed

Original Article

TPMGD: A genomic database for the traditional medicines in Pakistan

Rushuang Xiang [a,b,c,1], Huihua Wan [a,b,1], Wei Sun [a,b], Baozhong Duan [c], Weiqian Chen [a,b], Xue Cao [a,b], Sifan Wang [a,b], Chi Song [b], Shilin Chen [b], Yan Wang [d], Atia-tul Wahab [e], M. Iqbal Choudhary [d], Xiangxiao Meng [a,b,*]

[a] State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[b] Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[c] College of Pharmaceutical Science, Dali University, Dali 671000, China
[d] H.E.J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan
[e] Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan

ARTICLE INFO

Article history:
Received 31 October 2023
Revised 31 January 2024
Accepted 22 March 2024
Available online 18 May 2024

Keywords:
chloroplast genome
database
DNA barcodes
Pakistan
traditional medicine

ABSTRACT

Objective: In Pakistan, traditional medicines are an important component of the medical system, with numerous varieties and great demands. However, due to the scattered resources and the lack of systematic collection and collation, adulteration of traditional Pakistani medicine (TPM) is common, which severely affects the safety of their medicinal use and the import and export trades. Therefore, it is urgent to systematically organize and unify the management of TPM and establish a set of standards and operable methods for the identification of TPM.

Methods: We collected and organized the information on 128 TPMs with regard to their medicinal parts, efficacy, usage, and genetic material, based on Pakistan *Hamdard Pharmacopoeia of Eastern Medicine: Pharmaceutical Codex*. The genetic information of TPM is summarized from national center for biotechnology information (NCBI) and global pharmacopoeia genome database (GPGD). Furthermore, we utilized bioinformatics technology to supplement the chloroplast genome (cp-genome) data of 12 TPMs. To build the web server, we used the Linux + Apache + MySQL + PHP (LAMP) system and constructed the webpage on a PHP: Hypertext Preprocessor (PHP) model view controller (MVC) framework.

Results: We constructed a new genomic database, the traditional Pakistani medicine genomic database (TPMGD). This database comprises five entries, namely homepage, medicinal species, species identification, basic local alignment search tool (BLAST), and download. Currently, TPMGD contains basic profiles of 128 TPMs and genetic information of 102 TPMs, including 140 cytochrome *c* oxidase subunit I (COI) sequences and 119 mitochondrial genome sequences from *Bombyx mori*, 1 396 internal transcribed spacer 2 (ITS2) sequences and 1 074 intergenic region (*psb*A-*trn*H) sequences specific to 92 and 83 plant species, respectively. Additionally, TPMGD includes 199 cp-genome sequences of 82 TPMs.

Conclusion: TPMGD is a multifunctional database that integrates species description, functional information inquiry, genetic information storage, molecular identification of TPM, etc. The database not only provides convenience for TPM information queries but also establishes the scientific basis for the medication safety, species identification, and resource protection of TPM.

1. Introduction

Traditional medicine has been utilized by humans for thousands of years to prevent and treat various illnesses (Mukherjee et al., 2017; Liu, 2021; Li et al., 2022; Rehman et al., 2022). It has gained widespread acceptance and trust from populations worldwide and continues to play a crucial role in modern healthcare systems (Huang et al., 2021; Wei et al., 2023; Xu et al., 2023). Pakistan is home to a diverse range of plant species (Zhan & Chen, 1997; Zhang, Bao, & Yang, 2019), including numerous medicinal plants that are widely used for disease prevention and treatment by the majority of the population (Bibi et al., 2015; Yaseen et al., 2015;

* Corresponding author.
E-mail address: xxmeng@icmm.ac.cn (X. Meng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.chmed.2024.03.004
1674-6384/

Zahoor et al., 2017). In Pakistan, traditional medicine holds a significant position in healthcare system, with the prevalent system being the Unani system (Wu et al., 2022). The Unani system of medicine is among the oldest systems of healthcare and has a unique theoretical framework for understanding human disease and health (Yasmeen, Sofi, & Khan, 2020). They employ various therapeutic agents to treat diseases, including both monomeric constituent and polypill drugs derived from botanical, mineral, or animal sources, among which plant-based medicines are the most commonly used (Zahoor et al., 2017; Yasmeen, Sofi, & Khan, 2020; Wu et al., 2022). The *Hamdard Pharmacopoeia of Eastern Medicine* is a valuable reference book for traditional medicines in Pakistan. It originates from the Unani system and is widely used by practitioners of traditional medicine in Pakistan. The pharmacopeia includes information on various traditional medicines, including herbal medicines, animal medicines, and mineral medicines. It aims to standardize the safety, quality, and efficacy of traditional Pakistani medicine (TPM) by providing guidelines for their preparation, quality control, and usage.

Although the species characteristics of herbs have been recorded in the *Hamdard Pharmacopoeia*, the prevalence of adulteration and counterfeiting of medicinal materials due to the lack of effective identification methods is still a serious concern, as it undermines medication safety and affects the trade of medicinal resources (Zhan & Chen, 1997; Huang, Chen, & Awais, 2017). As globalization continues to deepen and there is an increasing number of people are turning to traditional medicines, establishing a standardized and easy-to-use method for identifying traditional medicinal plants is crucial for the safe usage of TPM (Ahmad et al., 2017). DNA barcoding is a widely used molecular biology technique that allows for accurate and rapid identification of species, regardless of sample morphology or tissue type. Its speed, accuracy, and versatility have made it a popular tool among researchers worldwide. In particular, the use of internal transcribed spacer 2 (ITS2) as a universal sequence for herb medicine identification (Sun, Xu, Song, & Chen, 2022; Chen et al., 2023) has been widely adopted. However, the limited genetic variation provided by short universal DNA barcodes can pose a challenge in identifying complex groups of species. To address this issue, the chloroplast genome (cp-genome) has been proposed as a "super barcode" due to its abundance of genetic information and high-resolution capability (Daniell, Lin, Yu, & Chang, 2016; Cheng et al., 2017; Lee et al., 2019). The cp-genome has been shown to be highly effective in identifying complex groups of species. Additionally, clade-specific markers can be developed based on the cp-genome (Liu et al., 2021; Wang et al., 2022). As a result, cp-genomes and DNA barcodes hold great potential for advancing the accurate identification and responsible development and utilization of medicinal plants.

With the reduction of sequencing costs and technological advancements, a large amount of genomic data has been generated in recent years and is often scattered across different media, hindering its effective utilization. Specialized genomic databases have become important platforms for effectively organizing and utilizing genomic data. Traditional Chinese medicines (TCM) Barcode System (https://www.tcmbarcode.cn/en/) has been developed for the identification of TCM, providing strong molecular evidence for the authenticity identification of traditional medicines (Chen et al., 2010; Chen et al., 2014). In addition, comprehensive genomic databases for medicinal plant research have been established, such as the Brazilian Pharmacopeia Medicinal Plant Genomic Database (Zhou, Liao, Li, Xu, & Chen, 2021) and the global pharmacopoeia genome database (GPGD) (Liao et al., 2022), which records whole-genome, organelle-genome, and other relevant genomic data of medicinal species from eight major pharmacopeias worldwide, greatly accelerating the modernization of traditional medicine research.

In this study, we developed traditional Pakistani medicine genomic database (TPMGD) by collecting genomic data from public databases as well as through our own sequencing efforts. The TPMGD aims to provide a comprehensive resource for the safe use, conservation of resources, and scientific research of TPM. The database is a reliable and user-friendly "omics" tool with several data mining functions. The TPMGD is publicly available at https://www.tpmgd.com/.

2. Materials and methods

2.1. Collection and organization of TPM data

The species directory in TPMGD is summarized from Part III of the *Hamdard Pharmacopoeia of Eastern Medicine*, which consists of 128 TPMs, comprising 124 botanical medicines, three animal medicines, and one mineral medicine. TPMs were primarily collected and organized based on their scientific name, a common name in Chinese, medicinal part used, and therapeutic functions. To address the issue of synonymy and scientific name revisions in the species names, we searched and collected synonyms and former names for species from databases such as Tropicos (https://www.tropicos.org/), Flora of China (https://www.iplant.cn/), and eFloras (https://www.efloras.org). This directory will assist users in navigating the database more efficiently and retrieving relevant information about TPMs. Fig. 1 illustrates a schematic overview of the pipeline utilized for constructing the TPMGD.

To construct TPMGD database, we collected ITS2, *psb*A-*trn*H, cytochrome *c* oxidase subunit I (COI), and organellar genome sequences of TPMs from public databases such as national center for biotechnology information (NCBI) and GPGD. Additionally, this study supplemented cp-genome data of 12 TPMs using second-generation sequencing technology. Uploaded data were classified into two sections: organellar genomes and DNA barcodes. These sections will enable users to search for and compare TPM DNA sequences and retrieve relevant genomic information related to TPM.

2.2. Supplement of cp-genome

2.2.1. Plant and DNA sources

Fresh leaves from 12 TPM species were collected and then dried using silica gel for DNA extraction. Table S1 and Fig. S1 provide detailed sample information. Genomic DNA was extracted using the Plant Genomic DNA kit (Tiangen, Beijing, China). DNA quantity was determined using Qubit® 3.0 Fluorometer (Life Technologies, CA, USA).

2.2.2. Assembly and annotation

We used 0.5 μg of DNA per sample as input material and generated sequencing libraries using Annoroad® Universal DNA Library Prep Kit V2.0 (AN200101-L). Sequencing was performed on the Novaseq 6000 S4 platform. The cp-genomes were assembled using GetOrganelle (Jin et al., 2020) and annotated using CpGAVAS2 and GeSeq (Wick, Schultz, Zobel, & Holt, 2015). The annotated cp-genome sequences were deposited in the Genome Warehouse of the National Genomics Data Center (Table 1). Finally, the maps of the cp-genomes were generated using online tools, OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) and IRscope (https://irscope.shinyapps.io/Chloroplot/).

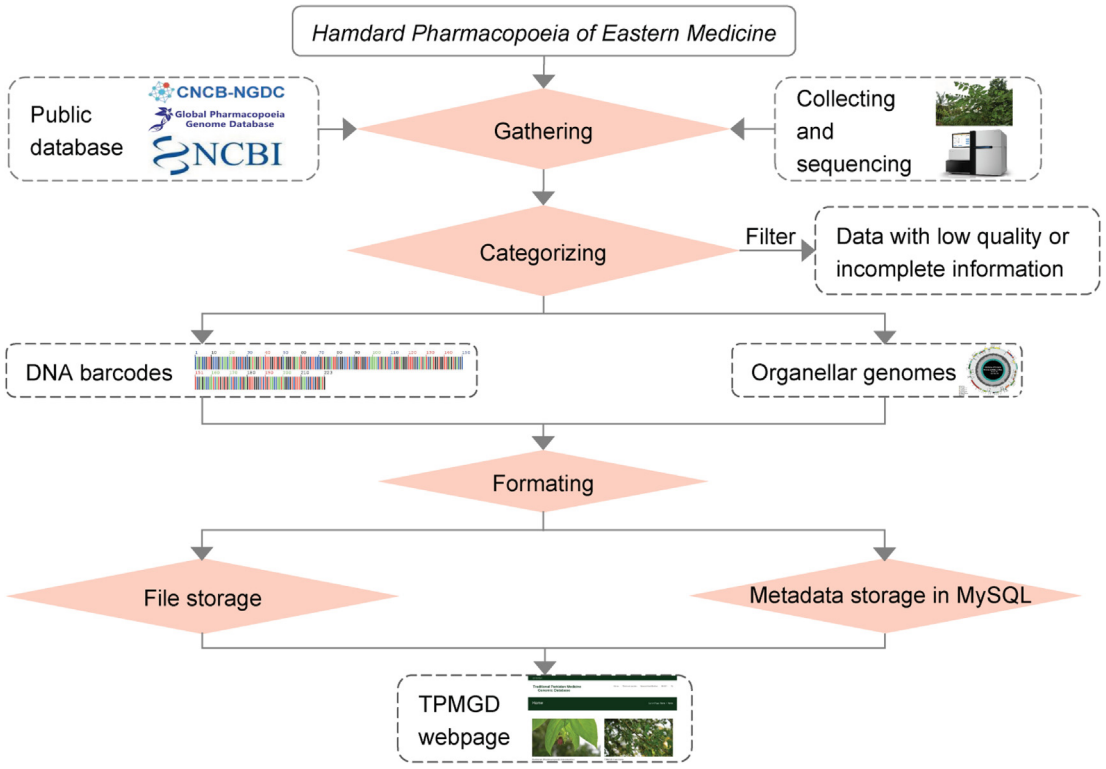

Fig. 1. Data collection and categorization process.

Table 1
Cp-genomes features of 12 TPMs.

Sample ID	Species	Genome size (bp)	LSC size (bp)	SSC size (bp)	IRs size (bp)	Total GC content (%)	GC content in LSC (%)	GC content in IRs (%)	GC content in SSC (%)	Number of genes	Protein-coding genes	tRNA genes	rRNA genes	GWH accession
S01	*M. officinalis*	127 677	–	–	–	33.63	–	–	–	108	75	29	4	GWHBPAK01000000
S02	*Ipomoea mauritiana*	161 926	88 105	11 951	30 935	37.52	36.09	40.59	32.22	126	81	37	8	GWHBOZV01000000
S03	*Malva cathayensis*	158 794	87 216	20 766	25 406	37.12	34.94	42.96	32.02	129	85	36	8	GWHBOZW01000000
S04	*Calophyllum inophyllum*	161 181	89 061	17 430	27 345	36.37	33.94	42.17	30.57	125	82	35	8	GWHBPAC01000000
S05	*G. odorata*	163 784	91 097	18 935	26 876	36.28	33.90	42.47	30.18	129	84	37	8	GWHBPAP01000000
S06	*M. communis*	159 342	88 028	18 490	26 412	37.0	34.83	42.76	30.69	127	83	36	8	GWHBPAD01000000
S07	*A. costus*	152 709	83 671	18 640	25 199	37.66	35.80	43.08	31.37	127	83	36	8	GWHBPAE01000000
S08	*A. augustum*	161 200	90 065	20 043	25 546	36.88	34.64	42.99	31.35	130	86	36	8	GWHBPAH01000000
S09	*Valeriana jatamansi*	155 392	85 691	15 341	27 180	38.32	36.68	42.52	32.63	125	79	38	8	GWHBPAG01000000
S10	*Operculina turpethum*	152 683	87 884	19 537	22 631	37.67	35.99	43.25	32.31	124	79	37	8	GWHBPAL01000000
S11	*Lactuca serriola*	152 730	84 066	18 598	25 033	37.55	35.73	43.02	31.05	125	81	36	8	GWHBPAJ01000000
S12	*Phyla nodiflora*	154 341	85 185	17 222	25 967	39.19	37.34	44.03	33.71	128	83	37	8	GWHBPAI01000000

2.2.3. Repeat analysis and codon usage

Simple sequence repeats (SSRs) and long repeat analysis were performed using the MIcroSAtellite (MISA) identification tool (https://pgrc.ipk-gatersleben.de/misa/) (Beier, Thiel, Munch, Scholz, & Mascher, 2017) and REPuter software (https://bibi serv.techfak.uni-bielefeld.de/reputer/) (Kurtz et al., 2001). Furthermore, the codon usage values and relative synonymous codon usage (RSCU) were analyzed using CodonW v.1.4.2. TBtools software was utilized to produce and enhance numerical heat maps (Chen et al., 2020).

2.3. Construction of TPMGD

We first processed and filtered all basic profiles and genetic information data to ensure uniformity. Then, we organized metadata from the organellar genome and DNA barcode datasets, as well as species information, and stored them in a MySQL database (Fig. 2). To build the web server, we used the LAMP system and constructed the webpage on a PHP MVC framework (Laravel v9.50.2; https://laravel.com/). Additionally, to enable the identification of TPM based on their organellar genomes, we first built a

BLAST database using the organellar genome sequences. Subsequently, we established a BLAST server that utilized this database for identifying TPM.

3. Results

3.1. Compilation of TPM

The objective of this study is to systematically categorize the fundamental descriptions of the TPMs contained in Part III of the *Hamdard Pharmacopoeia*. The recorded information includes the scientific names and Chinese names of the TPMs, their medicinal parts, therapeutic effects, and recommended usage. A total of 128 TPMs were analyzed, consisting of 124 varieties of herbal medicines from 52 different families and 115 genera, along with three types of animal medicines from three families and one mineral medicine.

The TPMGD contains the essential details for all 128 TPMs, with 80 % (102 species) are provided with genetic information. The genetic data for these 102 TPMs consists of 140 COI sequences and 119 mitochondrial genome sequences of *B. mori*. It also includes 1 396 ITS2 sequences and 1 074 *psb*A-*trn*H sequences specific to 92 and 83 plant species, respectively. Moreover, the database includes 199 cp-genome sequences of 82 TPMs. These resources provide valuable information for researchers and users interested in traditional medicinal plant species, facilitating analysis and comparison of genomic data across different species.

3.2. Structure of database

3.2.1. TPMGD homepage

The TPMGD homepage provides users with the latest information about TPM. The navigation menu on the homepage consists of five entries: "Home", "Medicinal Species", "Species Identifications", "Blast", and "Download". In addition, the homepage also briefly introduces the data composition and database functions (Fig. 2).

3.2.2. Medicinal species

The "Medicinal species" page provides users with the option to browse the list of species and search for specific medicines using the scientific name. For each medicinal species, detailed data is available on the "details" tab, which is divided into four sections. These sections contain comprehensive information related to the medicinal species, including part used, therapeutic functions, DNA barcodes, and organellar genomes (Fig. 2).

3.2.3. Species identification

Currently, species identification in the TPMGD uses organellar genome and DNA barcodes. The TPMGD contains 2 611 DNA barcodes covering 93 species (https://www.tpmgd.com/species). The species identification tool in our database allows users to compare their submitted sequences with organellar genomes or DNA barcodes (ITS2、*psb*A-*trn*H、COI) using BLAST. To use this tool, users simply input their sequence and select a database. The tool then returns the best-hit queries and top 50 hits, enabling users to achieve accurate species identification (Fig. 2).

3.2.4. BLAST tool and download

The homepage offers a BLAST tool under the "Blast" entry, which contains organellar genome sequences that allows users to search against them. After the search is completed, users can access the "Results" page, which displays the species exhibiting the highest degree of resemblance to the query sequence. From this page, users are able to view and retrieve the matches (Fig. 2). There is also an entry "Downloads" on the TPMGD homepage, which enables users to download relevant sequences.

3.3. Supplement of cp-genome

3.3.1. Genome atructure and annotation

We collected 12 TPMs and conducted next-generation sequencing (Fig. S1). These 12 TPMs have been used for treating various diseases as traditional medicines in multiple countries including China and India. Some of these species also have important biological significance, such as *Aucklandia costus* Falc. is a rare and endangered species and *Melilotus officinalis* (L.) Pall. has an atypical chloroplast genome structure (Fig. S2). Subsequently, we processed the raw data from the 12 TPMs, including removal of adapters and low-quality reads. This quality control step resulted in an approximate output of 10 Gb data for each species. Twelve gapless circular cp-genomes were assembled using the GetOrganelle software. The datasets presented in this study can be found in online repositories (https://ngdc.cncb.ac.cn/gwh). Table 1 provides detailed information regarding the size of each cp-genome, as well as a comprehensive list of gene content for each of the 12 TPMs. Worthy of notes, *M. officinalis* had the smallest cp-genome size (127 677 bp), while *Gynocardia odorata* R. Br. had the largest (163 784 bp). The majority of species exhibited a typical quadripartite structure (Fig. 3A and Fig. S2). However, due to the loss of the internal inverted repeat (IR) region, the cp-genome of *M. officinalis* has undergone structural changes and size reduction, lacking the typical quadripartite structure found in most cp-genomes (Fig. S2). In addition, the GC content in the IR region of these 12 TPMs (40.59%–44.03%) was higher than that of the large single-copy (LSC) (33.90%–37.34%) and small single-copy (SSC) (30.18%–33.71%) (Table 1).

Through annotation, we obtained 108–130 genes from each of the 12 TPMs, comprising 75–86 protein-coding genes, 4–8 ribosomal ribonucleic acid (rRNAs), and 29–38 transfer ribonucleic acid (tRNAs) (Table 1). In addition, we found that 101 genes were shared among the 12 TPMs, and based on their function, they could be divided into four categories (Table S2).

3.3.2. Repeat analysis and codon usage

Repetitive sequences are known to play a critical role in cp-genome rearrangements and evolution (Nazareno, Carlsen, & Lohmann, 2015; Zheng et al., 2020; Wang et al., 2022). We conducted an oligonucleotide repeat analysis of reverse (R), palindromic (P), forward (F), and complement (C) repeats using REPuter. Interestingly, we found the abundance of R and C repeats was lower than that of F and P repeats, and that the majority of repeat sequences were within the range of 30–39 bp (Fig. 3B and C). *Myrtus communis* L. had the minimum number of repeats (36), whereas *M. officinalis* had the maximum (598). Complete details are listed in Table S3. SSRs in chloroplasts are often used for phylogenetic analysis or population genetics. In our study, a total of 38–138 SSRs were identified among all 12 cp-genomes (Table S4). Over half of these SSRs consisted of mononucleotide A/T motifs, followed by dinucleotide motifs with a predominant AT/TA motif (Fig. 3D and Table S4).

For codon usage, the coding genes of these 12 TPMs include 64 codons, of which 61 encode for 20 amino acids and three stop codons (Table S5). The findings indicated that the utilization of A or T nucleotides in the third codon position was more prevalent than that of other nucleotides. The RSCU values of 27 codons were higher than 1, the RSCU values of 2 codons were equal to 1, and others were lower than 1 (Fig. S3).

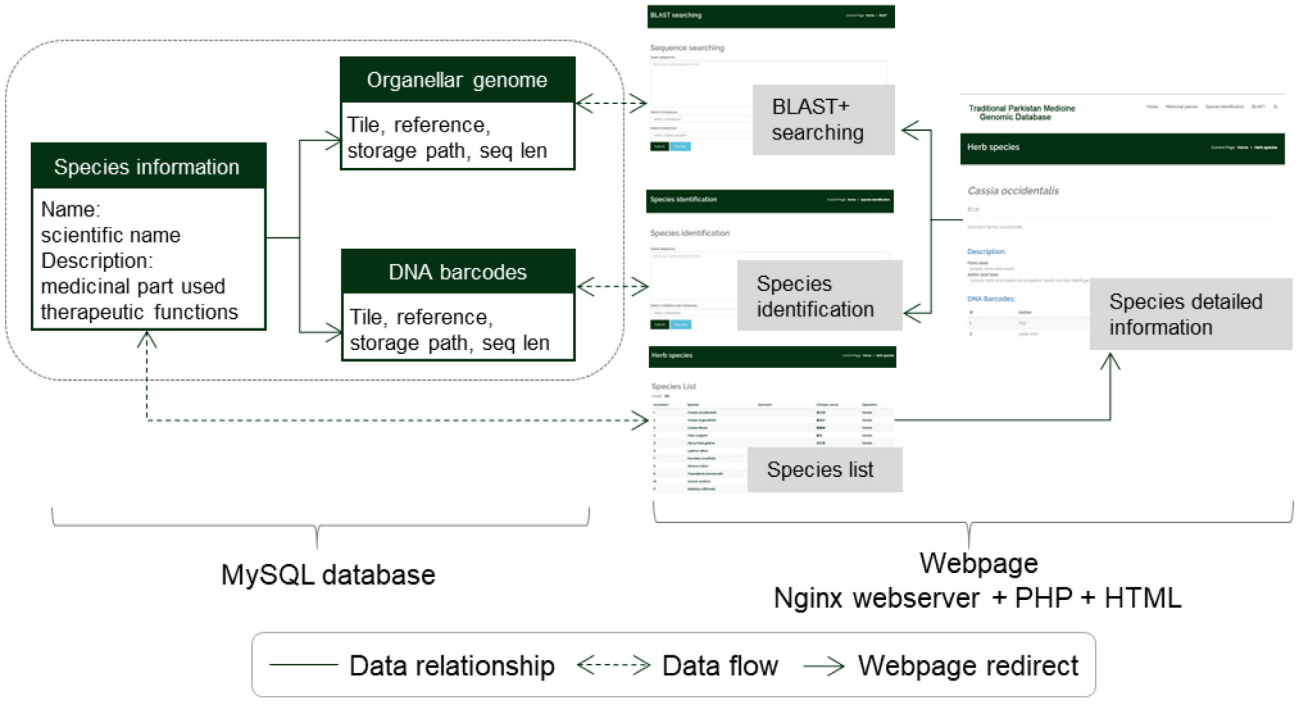

Fig. 2. TPMGD database architecture model.

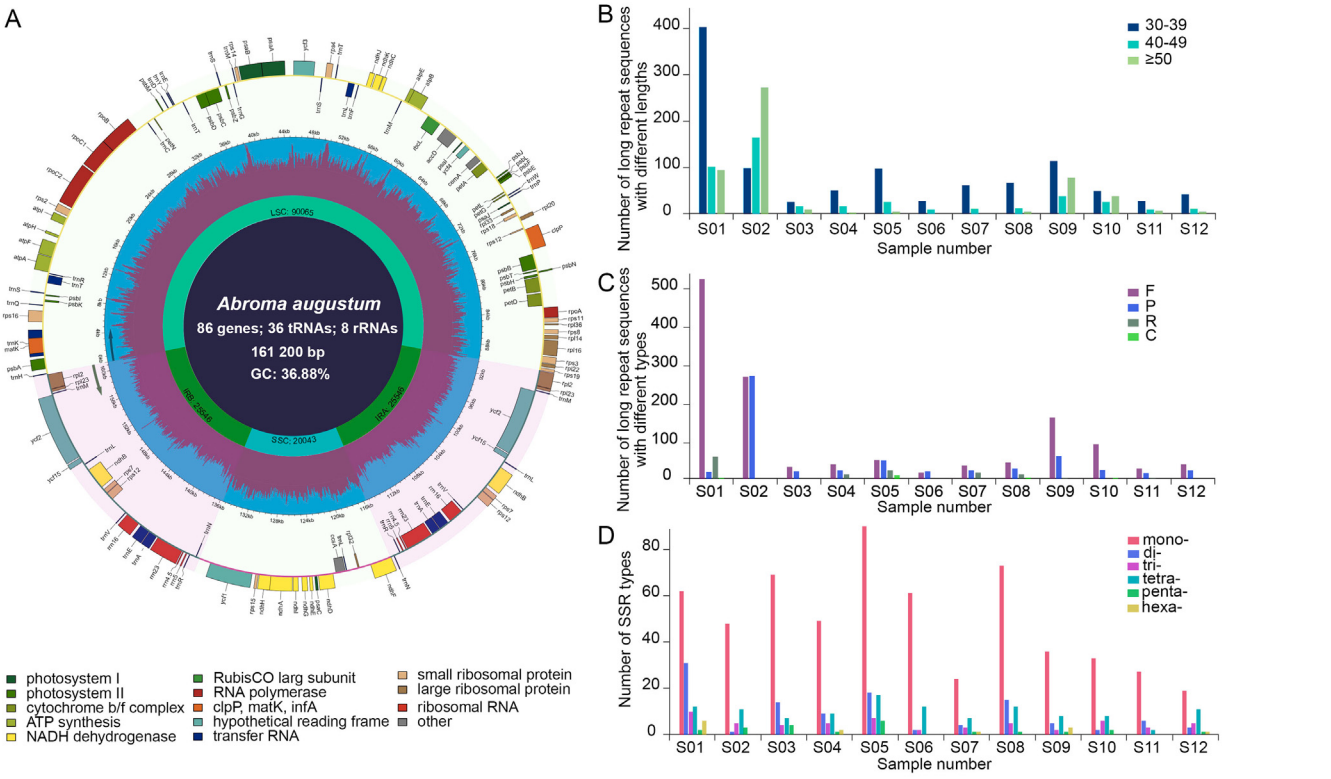

Fig. 3. Morphological characteristics and chloroplast genome analysis of 12 TPMs. (A) Cp-genome map of *A. augustum*. (B) Number of long repeat sequences with different lengths in 12 TPMs. (C) Number of long repeat sequences with different types in 12 TPMs. R, P, F and C indicate repeat types: R (Reverse), P (Palindromic), F(Forward), and C (Complement). (D) Number of SSR types in chloroplast genomes of 12 TPMs.

4. Discussion

4.1. TPMGD

Pakistan has rich traditional medicinal resources, which have made significant contributions to the health of local and global populations (Shinwari, 2010; Faheem et al., 2022; Rehman et al., 2022). However, for a long time, improper use, mixing, and even intentional adulteration have seriously affected the efficacy of TPM and even caused many safety issues (Atif et al., 2020; Haq, 2004; Zahra, Shinwari, & Qaiser, 2016). TPM resources urgently require scientific sorting and standardized naming using modern

technologies and effective identification methods such as molecular biology (Chen & Huang, 2006; Zahra, Shinwari, & Qaiser, 2016; Ahmad et al., 2017). At the same time, an in-depth exploration of traditional medicine using multi-omics methods can elucidate the mechanisms of disease treatment, promote rational use, and drive the modernization of TPM development (Sun, Xu, Song, & Chen, 2022).

With the rapid advancement of sequencing technologies, sequencing read lengths have increased while sequencing costs have decreased. As a result, a growing amount of genomic data on TPM has been released. Specialized genomic databases have emerged in response to the rapidly growing amount of genomic data. TPMGD has provided complete, organized, and easily accessible genetic information on TPM for traditional medicine practitioners in Pakistan. TPMGD is expected to aid in molecular identification, facilitate the discovery of new drug materials, foster the cultivation of superior varieties, and preserve invaluable medicinal plant genetic resources. TPMGD offers researchers valuable resources to facilitate their studies and contributes to the advancement of global healthcare.

4.2. Cp-genome

This study generated the first complete cp-genome sequences for the 12 TPMs. In this study, 11 cp-genomes exhibited typical quadripartite structures, similar to most of reported, such as *Polygonatum* and *Paris* (Jiang et al., 2022; Wang et al., 2022). In *M. officinalis*, the absence of the IR region led to the loss of the typical quadripartite structure in the cp-genome. This phenomenon mainly occurs in the subfamily Papilionoideae, and these species are referred to as the inverted repeat-lacking clade (Sabir et al., 2014). In addition, we also conducted analyses on the long repeats, SSR, and codon usage bias (CUB) of these species.

These data greatly enhance the genetic information available for TPMs, which can be effectively employed not only for molecular identification but also for phylogenetic analysis, chloroplast function analysis, and the development of specific molecular markers.

5. Conclusion

Traditional medicine has been a guardian of human health for a long time. It is affordable, easily accessible, and still widely used today. However, due to issues such as unstable quality of medicinal materials and misuse or adulteration, the effectiveness and safety of traditional medicine have been seriously affected, damaging the reputation of traditional medicine. The emergence of modern technologies, such as molecular biology, provides opportunities for the accurate identification and selection of high-quality varieties of traditional medicine. TPMGD is the first specialized genomic database that collects and organizes large-scale TPM genetic data. The database has a user-friendly website that integrates data storage, retrieval, acquisition, and species identification functions. TPMGD not only provides convenience for TPM information queries but also lays the scientific foundation for medication safety, molecular identification, and resource protection of TPM. To ensure the usefulness of TPMGD, it will be regularly updated and maintained. The database is not limited to current TPM species and data types and will continue to enrich the variety of species in the future, gradually adding whole-genome, transcriptome, metabolome, proteome, and other data, striving to build a comprehensive TPM multi-omics database, and effectively promoting the modernization of TPM.

CRediT authorship contribution statement

Rushuang Xiang: Conceptualization, Writing – original draft. **Huihua Wan:** Conceptualization, Writing – review & editing. **Wei Sun:** Formal analysis. **Baozhong Duan:** Formal analysis. **Weiqian Chen:** Formal analysis. **Xue Cao:** Formal analysis. **Sifan Wang:** Formal analysis. **Chi Song:** Conceptualization, Methodology. **Shilin Chen:** Conceptualization, Methodology. **Yan Wang:** Resources. **Atia-tul Wahab:** Resources. **M. Iqbal Choudhary:** Resources. **Xiangxiao Meng:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing finacial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Scientific and technological innovation project of the China Academy of Chinese Medical Sciences (No. CI2021A03712, CI2021A03710), the Fundamental Research Funds for the Central Public Welfare Research institutes (No. ZZ15-YQ-039), the National Key R&D Program of China (No. 2021YFE0100900), and the National Natural Science Foundation of China (No. 82204579).

We thank Dr. Baosheng Liao (Second Clinical College, Guangzhou University of Chinese Medicine) for his technical support and database construction.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chmed.2024.03.004.

References

Ahmad, W., Muhammad, K., Hussain, A., Ahmad, H., Kahn, K., Qarshi, I. A., ... Iqbal, J. (2017). DNA fingerprinting of essential commercialized medicinal plants from Pakistan. *American Journal of Plant Sciences, 8*(9), 2119–2132.

Atif, M., Malik, I., Asif, M., Qamar-Uz-Zaman, M., Ahmad, N., & Scahill, S. (2020). Drug safety in Pakistan. In *Drug safety in developing countries* (pp. 287−325). Academic Press.

Beier, S., Thiel, T., Munch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics, 33*(16), 2583–2585.

Bibi, T., Ahmad, M., Tareen, N. M., Jabeen, R., Sultana, S., Zafar, M., & Zain-ul-Abidin, S. (2015). The endemic medicinal plants of Northern Balochistan, Pakistan and their uses in traditional medicine. *Journal of Ethnopharmacol, 173*, 1–10.

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., & Xia, R. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular plant, 13*(8), 1194–1202.

Chen, S. L., Yao, H., Han, J. P., Liu, C., Song, J. Y., Shi, L. C., ... Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLos One, 5*(1), e8613.

Chen, S. L., Pang, X. H., Song, J. Y., Shi, L. C., Yao, H., Han, J. P., & Leon, C. (2014). A renaissance in herbal medicine identification: From morphology to DNA. *Biotechnology Advances, 32*(7), 1237–1244.

Chen, S. L., Yin, X. M., Han, J. P., Sun, W., Yao, H., Song, J. Y., & Li, X. W. (2023). DNA barcoding in herbal medicine retrospective and prospective. *Journal of Pharmaceutical Analysis, 13*(5), 431–441.

Chen, S. T., & Huang, L. Q. (2006). A brief review of traditional theories of medicine in Pakistan. *Chinese Journal of Information on TCM, 5*, 104–105.

Cheng, H., Li, J. F., Zhang, H., Cai, B. H., Gao, Z. H., Qiao, Y. S., & Mi, L. (2017). The complete chloroplast genome sequence of strawberry (*Fragaria* x *ananassa* Duch.) and comparison with related species of Rosaceae. *PeerJ, 5*, e3919.

Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology, 17*(1), 134.

Faheem, M., Ameer, S., Khan, A. W., Haseeb, M., Raza, Q., Shah, F. A., ... Alsiwiehri, N. O. (2022). A comprehensive review on antiepileptic properties of medicinal plants. *Arabian Journal of Chemistry, 15*(1), 103478.

Haq, I. (2004). Safety of medicinal plants. *Pakistan Journal of Medical Research, 43*(4), 203–210.
Huang, K., Zhang, P., Zhang, Z. H., Youn, J. Y., Wang, C., Zhang, H. C., & Cai, H. (2021). Traditional Chinese Medicine (TCM) in the treatment of COVID-19 and other viral infections: Efficacies and mechanisms. *Pharmacology & Therapeutics, 225*, 107843.
Huang, X. J., Chen, H., & Awais, I. (2017). Comparison of traditional Chinese and Brazilian medicines studies. *Asia-Pacific Traditional Medicine, 13*(13), 1–2.
Jiang, Y., Miao, Y. J., Qian, J., Zheng, Y., Xia, C. L., Yang, Q. S., ... Duan, B. Z. (2022). Comparative analysis of complete chloroplast genome sequences of five endangered species and new insights into phylogenetic relationships of *Paris*. *Gene, 833*, 146572.
Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology, 21*(1), 241.
Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research, 29*(22), 4633–4642.
Lee, H. O., Joh, H. J., Kim, K., Lee, S. C., Kim, N. H., Park, J. Y., ... Yang, T. J. (2019). Dynamic chloroplast genome rearrangement and DNA barcoding for three Apiaceae species known as the medicinal herb "Bang-Poong". *International Journal of Molecular Sciences, 20*(9), 2196.
Li, X. L., Wu, L., Wu, R. X., Sun, M., Fu, K., Kuang, T. T., & Wang, Z. (2022). Comparison of medicinal preparations of Ayurveda in India and five traditional medicines in China. *Journal of Ethnopharmacology, 284*, 114775.
Liao, B. S., Hu, H. Y., Xiao, S. M., Zhou, G. R., Sun, W., Chu, Y., ... Chen, S. L. (2022). Global Pharmacopoeia Genome Database is an integrated and mineable genomic database for traditional medicines derived from eight international pharmacopoeias. *Science China Life Sciences, 65*(4), 809–817.
Liu, C. X. (2021). Overview on development of ASEAN traditional and herbal medicines. *Chinese Herbal Medicines, 13*(4), 441–450.
Liu, K., Wang, R., Guo, X. X., Zhang, X. J., Qu, X. J., & Fan, S. J. (2021). Comparative and phylogenetic analysis of complete chloroplast genomes in Eragrostideae (Chloridoideae, Poaceae). *Plants (Basel), 10*(1), 109.
Mukherjee, P. K., Harwansh, R. K., Bahadur, S., Banerjee, S., Kar, A., Chanda, J., ... Katiyar, C. K. (2017). Development of Ayurveda-tradition to trend. *Journal of Ethnopharmacol, 197*, 10–24.
Nazareno, A. G., Carlsen, M., & Lohmann, L. G. (2015). Complete chloroplast genome of *Tanaecium tetragonolobum*: The first Bignoniaceae plastome. *PLos One, 10*(6), e129930.
Rehman, S., Iqbal, Z., Qureshi, R., Rahman, I. U., Khan, M. A., Elshaer, M. M. A., ... Abu Bakr Elsaid, N. M. (2022). Ethnogynaecological knowledge of traditional medicinal plants used by the indigenous communities of north Waziristan, Pakistan. *Evidence-Based Complementary and Alternative Medicine, 2022*, 6528264.
Sabir, J., Schwarz, E., Ellison, N., Zhang, J., Baeshen, N. A., Mutwakil, M., ... Ruhlman, T. (2014). Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnology Journal, 12*(6), 743–754.
Shinwari, Z. K. (2010). Medicinal plants research in Pakistan. *Journal of Medicinal Plants Research, 4*(3), 161–176.
Sun, W., Xu, Z. C., Song, C., & Chen, S. L. (2022). Herbgenomics: Decipher molecular genetics of medicinal plants. *Innovation, 3*(6), 100322.
Wang, J., Qian, J., Jiang, Y., Chen, X. C., Zheng, B. J., Chen, S. L., ... Duan, B. Z. (2022). Comparative analysis of chloroplast genome and new insights into Phylogenetic relationships of *Polygonatum* and tribe Polygonateae. *Frontiers in Plant Science, 13*, 882189.
Wei, Z. C., Chen, J., Zuo, F., Guo, J. L., Sun, X. D., Liu, D. M., & Liu, C. H. (2023). Traditional Chinese medicine has great potential as candidate drugs for lung cancer: A review. *Journal of Ethnopharmacology, 300*, 115748.
Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics, 31*(20), 3350–3352.
Wu, L., Qiao, L. Y., Li, X. L., Xiao, H. X., Li, H., & Wang, Z. (2022). Varieties, medicinal materials and indications of Unani preparations in India. *Pharmacy and Clinics of Chinese Materia Medica, 13*(2), 74–79.
Xu, Z. Y., Dong, M., Yin, S. P., Dong, J., Zhang, M., Tian, R., ... Chen, J. (2023). Why traditional herbal medicine promotes wound healing: Research from immune response, wound microbiome to controlled delivery. *Advanced Drug Delivery Reviews, 195*, 114764.
Yaseen, G., Ahmad, M., Sultana, S., Alharrasi, A. S., Hussain, J., Zafar, M., & Shafiq-Ur-Rehman. (2015). Ethnobotany of medicinal plants in the Thar Desert (Sindh) of Pakistan. *Journal of Ethnopharmacology, 163*, 43–59.
Yasmeen, A., Sofi, G., & Khan, K. (2020). Aamar-e-Advia (shelf-lives) of drugs in Unani system of medicine: A conceptual review. *Journal of Integrative Medicine, 18*(2), 114–124.
Zahoor, M., Yousaf, Z., Aqsa, T., Haroon, M., Saleh, N., Aftab, A., ... H. (2017). An ethnopharmacological evaluation of Navapind and Shahpur Virkanin district Sheikupura, Pakistan for their herbal medicines. *Journal of Ethnobiology and Ethnomedicine, 13*(1), 27.
Zahra, N. B., Shinwari, Z. K., & Qaiser, M. (2016). DNA barcoding: A tool for standardization of herbal medicinal products (HMPS) of Lamiaceae from Pakistan. *Pakistan Journal of Botany, 48*(5), 2167–2174.
Zhan, Y. H., & Chen, K. L. (1997). Survey of studies on Pakistany medicinal plants. *Journal of Wuhan Botanical Research, 15*(3), 274–276.
Zhang, K. X., Bao, F., & Yang, Z. G. (2019). Medicinal plant resources of Pakistan. *Asia-Pacific Traditional Medicine, 15*(2), 1–8.
Zheng, G., Wei, L. L., Ma, L., Wu, Z. Q., Gu, C. H., & Chen, K. (2020). Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: Identification of highly divergent regions and inference of phylogenetic relationships. *Plant Molecular Biology, 102*(6), 659–676.
Zhou, G. R., Liao, B. S., Li, Q. S., Xu, J., & Chen, S. L. (2021). Establishing a genomic database for the medicinal plants in the Brazilian Pharmacopoeia. *Chinese Medicine, 16*(1), 71.