

# Establishing a Genomic Database for the Traditional Medicinal Plants in the Brazilian Pharmacopoeia

GUANRU ZHOU (✉ [guanru\\_zhou@hotmail.com](mailto:guanru_zhou@hotmail.com))

Hubei University of Chinese Medicine <https://orcid.org/0000-0001-5016-8424>

**Baosheng Liao**

China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

**Qiushi Li**

China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

**Jiang Xu**

China Academy of Chinese Medical Sciences

**Shilin Chen**

China Academy of Chinese Medical Sciences

---

## Research

**Keywords:** Genome, Traditional medicinal plants, Brazilian Pharmacopoeia, DNA barcodes, Chloroplast genome, Herbgenomics

**Posted Date:** June 24th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-627096/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Chinese Medicine on August 5th, 2021. See the published version at <https://doi.org/10.1186/s13020-021-00484-5>.

# Abstract

## Background

Brazil is exceptionally **abundant** in traditional medicinal plant resources and has a rich ethnopharmacological history. Brazilian Pharmacopoeia (BP) acts as a national standard regulating drugs' quality and has six published editions. Recent genomic approaches have led to a resurgence of interest in herbal drugs. Plants' genomic data have been used for pharmaceutical applications, protecting natural resources, and efficiently regulating the market. However, there are few genomic databases specifically on medicinal plants, and establishing one focusing on the herbs of BP is urgently needed.

## Methods

The BP editions' medicinal plant species were analyzed to understand the evolution of the herbal drugs in Brazil. A new database, BPGD (Brazilian Pharmacopoeia Genomic Database), was constructed based on a typical web server architecture. A BLAST server for species identification and sequence searching with internal transcribed spacer 2 (ITS2), intergenic region (*psbA-trnH*), and chloroplast genome (cp-G) of Brazilian traditional medicinal plants was also embedded in BPGD. Data of 82 plants in BP were collected and categorized into four parts: DNA barcodes, super-barcodes, genomes, and sequencing data. Further, the cp-Gs of *Aloe* genus in the database were analyzed as an illustration.

## Results

BPGD (V1.0) has been tested and opened for public users. The database provides a comprehensive set of data, including the description and identification criteria of medicinal plants, and allows sequence-based search using BLAST. The database has 753 ITS2 of 76 species, 553 *psbA-trnH* and 190 genomes (whole genome and chloroplast genome) of 57 species, and 37 genome sequence data sets of 24 species and 616 transcriptome sequence data sets of 34 species. The data includes 187 cp-Gs representing 57 medicinal species in BP. Analysis of the six cp-Gs of three *Aloe* species identified the variable regions in cp-Gs, which could be used to identify species and understand the intraspecific relationship.

## Conclusions

This study presents the first genomic database for traditional medicinal plants listed in the latest BP. It serves as an efficient platform to obtain genomic data, specifically on medical plants listed in the BP (<http://bpgenome.com/>).

## Background

Traditional medicinal plants are sources of phytochemicals that play vital roles in disease prevention and treatment. These sources are inexpensive and readily available and have been used in developed and least developed countries [1, 2]. The discovery of Salicin (analgesic and antipyretic) by Raffaele Piria, in 1832, from *Salix alba* [3] is considered a milestone in developing the global pharmaceutical industry.

Since then, traditional medicinal plants have gained considerable importance as sources of bioactive phytochemicals for drug discovery [4, 5]. It has been estimated that about 30% of therapeutic drugs are derived from natural resources, particularly plants and microorganisms [6, 7]. Besides, the completion of the Human Genome Project has opened a new chapter in understanding and treating human diseases, initiating and gradually deepening research on herbs to the genomic level [8]. Further, Tu Youyou's discovery of artemisinin from the plant *Artemisia annua* to treat malaria also exhibited the potential role of medicinal plants [9].

Traditional herbal medicines in the Latin America came into contact with other medical traditions at the beginning of the 16th century, introduced throughout the Conquest and European colonial expansion. Within a dominant sociopolitical framework, the folk herb traditions were fused with other medical cultures and syncretized by religious doctors and other various social workers of the certain regional public health system [10]. In the next 300 years, the historical relations between the native medical traditions in the Latin America and the medical cultures of other continents, created a particularly rich ethnomedicinal foundation [11] and continually impacted the national health system of Latin America countries.

Brazil is an ideal country in Latin America that has established itself in public health, emphasizing the application and development of traditional medicinal plants and their derivatives. Brazil has the world's largest share of biodiversity (15–20%), and most of this biodiversity has not been explored, offering plenty of scope for herbal medicine development [3, 12]. The diverse Brazilian culture contributed mainly to the use of herbal medicines. Besides, approximately 305 ethnic groups, speaking 274 languages [13], have thousand years of ethnopharmacological history [14]; around 1000 plant species have been used as Amazon's traditional medicine [15, 16]. Additionally, the Portuguese, who colonized Brazil between 1500 and 1822, brought herbs from other parts of the world instead of exploring the native medicinal plants [17]. The traditional herbal medical system with the combined knowledge of the indigenous people, Europeans, and Africans has led to the development of botanical medicines [18]. Consequently, Brazil has become the biggest pharmaceutical market and the only country in Latin America ranked amongst the top pharmaceutical markets worldwide [19]; it is one of the world's most profitable pharmaceutical markets [20]. As the standard publications regulating the quality of drugs, the pharmacopoeias make the quality standards obligatory, ensuring consistency in medicines approved by representatives of specific political units and representing the local progress in related scientific fields [21]. The latest (sixth) edition of BP has revoked all other editions; serves as the nucleus of future editions through constant review, seeking to emerge as an international standard. At present, the Brazilian Pharmacopoeia Commission is an observer for European and International pharmacopoeias and has a mutual acknowledgement with the Argentine Pharmacopoeia [22]. It will also help guide proposals for the joint development of pharmacopoeias with countries on the South America continent. Meanwhile, as a national pharmacopoeia in the Portuguese language, the Brazilian pharmacopoeia also has an influence on other regions, including Macau, in where could be a blossom of traditional medicinal plants' development, integrating Brazilian herbal knowledge and Chinese.

New proteomic and genomic technologies have led to a resurgence of interest in natural products in academia and pharmaceutical organizations [23, 24]. DNA barcode is a short DNA fragment different between species [25], which provides a practical solution for identifying species. In addition, the optimal combination of single-locus barcodes with chloroplast genome (super-barcodes) provides a new method for efficient plant identification [26]. Thus, a database integrating DNA barcodes and organelle genomes may solve the increasing challenges in plant identification, not just in Brazil but also globally. With the development of sequencing technology and synthetic biology, the transcriptome and genome of plants have been sequenced and used to synthesize the desired compounds by bacterial engineering [27]. For example, ingredients with high medicinal activity, such as artemisinin [28, 29] and paclitaxel [30], have been extracted from medicinal plants. With the help of omics data, the decomposition of the biosynthetic pathway of drug compounds has also got into the fast lane [31, 32].

Quality germplasm resources are the key to the generation of omics data. However, various wild resources of plant species have been endangered due to habitat destruction and extensive exploitation and utilization [33–35]. Molecular-marker-assisted breeding based on genomic data can enrich germplasm resources and protect wild resources efficiently [36]. Cultivation of wild medicinal plants has become an inevitable trend of sustainable development [37]. Representative examples of comprehensive omics databases assisting crop breeding are available for rice [38], maize [39], and wheat [40], but none in herbs. A database collecting omics data of medicinal plants could be significant.

A genomic database is a warehouse that organizes, stores, and manages a variety of genomic data. There are three public, comprehensive genomic databases available: the National Centre for Biotechnology Information (NCBI), the European Institute of Bioinformatics (EBI), and the DNA Database of Japan (DDBJ). With the rapid increase in the volume and complexity of biological data, these databases play important roles in advancing molecular research [41]. Several studies have published genomic data of medicinal plants; however, problems such as different research team hosts, inconsistent data formats, and unstable web services bring challenges to utilizing the herb genomic data. Besides, due to the continuous improvement in genome assembly and species sequencing, the number of genome assemblies within a single species is also increasing [31, 42]. The use of these multiple versions of a genome assembly can be confusing and time-consuming. Therefore, genomic data need to be organized and displayed for further use.

In the present study, BPGD (Brazilian Pharmacopoeia Genomic Database), a database of Brazilian Pharmacopoeia (BP) medicinal plants containing genetic information, including genome, transcriptome, cp-G, and DNA barcodes, was built. The BPGD will provide a valuable resource for accelerating genome research and molecular breeding of traditional medicinal plants in Brazil.

## Methods

### Data collection and curation workflow

Schematic overview of the BPGD database construction pipeline is provided (Fig. 1). The BP editions' medicinal plant species were analyzed to understand the evolution of the herbal drugs in Brazil. Statistical data of medicinal plants were sorted through all six editions of BP, such as Latin names, Chinese names, families, relative monographs, occurrence numbers, native or exotic status and so on. The details on the plant species in the first four editions were collected from previous work [43], while those in the last two editions were collected and categorized. Latin names appearing in BP adapted according to monograph description while the native geographical distribution of the plant taxa according to several modern floras, such as the Flora Reipublicae Popularis Sinicae (<http://www.iplant.cn/>), the Flora do Brasil 2020 (<http://floradobrasil.jbrj.gov.br/>).

Genomic data were collected from public databases. The ITS2 and *psbA-trnH* sequences available in the NCBI database were linked to BPGD. The nuclear genomes were assembled from the NCBI genome database. The majority of raw data were collected from the Sequence Read Archive (SRA) database and a few from the China National GeneBank DataBase (CNGB). The uploaded data were categorized into four parts: DNA barcodes, super-barcodes, genomes, and sequencing data sets.

Data on 82 medicinal plants with definite Latin names, compiled from the plant monographs in the latest (sixth) edition of the BP, were incorporated into BPGD. Currently, BPGD (v1.0) has 753 ITS2 of 76 species, 553 *psbA-trnH* and 187 cp-Gs of 57 species, the optimal nuclear genomes of 3 species, and 37 raw genome sequencing data of 24 species, whose assembled nuclear genome data have not been published, and 616 raw transcriptome sequencing data of 34 species.

The cp-Gs were assembled, following the methods described in a previous study [44]. All Aloe cp-Gs were aligned using the MAFFT program [45]. Sequence identity in a 50 bp window of the alignment was calculated. The same genes from different cp-Gs were pairwise aligned using Clustal Omega (<http://www.clustal.org/omega/>), and pairwise genetic distance was calculated using the EMBOSS program distmat (<http://emboss.sourceforge.net/>) with default parameters.

## Database implementation

BPGD was constructed based on the typical web server architecture pattern, LNMP, including Linux operating system, Nginx web server with high-performance HTTP and reverse proxy servers, MySQL database server and H2 data management system, and PHP programming language. A BLAST server was also set up to identify the Brazilian medicinal plants based on ITS2, *psbA-trnH*, and cp-G.

## Results

### Medicinal plants recorded in the Brazilian Pharmacopoeia (BP)

This study explored the history and development of traditional medicinal plants in the BP (Fig. 2). A total of 200 plants have been recorded only in the first edition; 44 in the first and second editions, and 43 in the fifth and sixth editions, which are not included in any other editions. Meanwhile, only three medicinal plant species have been recorded in all the six BP editions and 14 in five editions (Fig. 2A). Meanwhile, consistent with the earlier editions [22, 43, 46, 47], the exotic species were more than native species in the sixth edition (Fig. 2B), indicating the broader applications of non-native herbal drugs in Brazil due to the excellent documentation of quality control [48]. However, combined the previous work [22, 43, 46, 47] and the data in the last two editions, the number of native plants has been on the rise since the fourth edition of the BP. Analysis of 22 plant families, with most species recorded in the BP history (Fig. 2C), pointed out that eight families alone (Fig. 2C; Eight from left) comprise more than 50% of the species in the BP. Thus, the BP analysis confirms that the research on traditional medicinal plants in Brazil is constantly improving.

## Structure of the database

The public has open and free access to the BPGD over the internet (<http://bpggenome.com/>). The database and its web pages are user-friendly (Fig. 3). The BPGD homepage provides the latest information on the compiled BP medicinal plants. The database homepage has a navigation menu consisting of four entries: "Home", "Blast", "Species", and "Contact us". The homepage also displays a brief introduction and the data collection status. A BLAST tool compatible with ITS2, *psbA-trnH*, and cp-G sequences provided under the "Blast" entry on the homepage enables search against the genomes and transcriptomes. Users can view and download the hits from the "Results" page; it shows the species with the closest identity to the query sequence. Through the "Species" page, users can browse the species list and search specific plants using the Latin name keyword. Each plant comes with "details" tab; the detailed data consist of four parts: DNA barcodes, super barcodes, genomes, and sequencing data sets. Under the "ITS2" and "*psbA-trnH*" gateways, two kinds of DNA barcoding nucleotide sequences are available in the Fasta format. Meanwhile, the optimal nuclear genome data and the links to the original and other assembled versions of the genome can be obtained under the "Genome" gateway.

## Characteristics of collected cp-Gs in BPGD

Chloroplast genome (cp-G) has a suitable length and sufficient variable sites between species, which has been used as super-barcodes to identify plant species [26]. In total, 187 cp-Gs representing 57 medicinal plant species of the BP were collected and stored in BPGD (Table1, Supplementary Table1). Babosa (*Aloe vera* L.) is common in northeastern Brazil. Its leaves, extracts, and resins have antibacterial, anti-inflammatory, and healing properties and are used to treat liver and stomach diseases [49]. Six cp-Gs of three *Aloe* species (*A. vera*, *A. maculata*, and *A. barbadensis*; two cp-Gs from each) in BPGD were analyzed to identify the variable regions in the cp-Gs of these species. The most variable region in the cp-Gs was located in the large single-copy (LSC) region, and the most variable gene was NADH dehydrogenase F (*ndhF*; Fig.4). The variable region could be used to identify species and understand the intraspecific relationship of the *Aloe* genus.

Table 1  
 Statistics of chloroplast genomes stored in BPGD

<b>Family</b>	<b>Genera</b>	<b>Cp-count</b>
Malvaceae	<i>Althaea, Gossypium, Theobroma</i>	21
Rutaceae	<i>Citrus</i>	15
Zingiberaceae	<i>Curcuma, Zingiber</i>	14
Asteraceae	<i>Helianthus, Matricaria</i>	13
Fabaceae	<i>Glycyrrhiza, Stryphnodendron</i>	12
Solanaceae	<i>Atropa, Datura, Hyoscyamus</i>	11
Myrtaceae	<i>Corymbia, Eucalyptus, Eugenia, Melaleuca, Psidium, Syzygium</i>	11
Caprifoliaceae	<i>Sambucus, Valeriana</i>	9
Apiaceae	<i>Anethum, Centella, Coriandrum, Foeniculum</i>	8
Lauraceae	<i>Cinnamomum, Persea</i>	8
Rosaceae	<i>Crataegus, Prunus</i>	7
Oleaceae	<i>Olea</i>	6
Passifloraceae	<i>Passiflora</i>	6
Amaryllidaceae	<i>Allium</i>	5
Quillajaceae	<i>Quillaja</i>	4
Poaceae	<i>Cymbopogon</i>	3
Ranunculaceae	<i>Hydrastis</i>	3
Phyllanthaceae	<i>Phyllanthus</i>	3
Plantaginaceae	<i>Plantago</i>	3
Orchidaceae	<i>Vanilla</i>	3
Asphodelaceae	<i>Aloe</i>	2
Schisandraceae	<i>Illicium</i>	2
Lamiaceae	<i>Mentha, Thymus</i>	2
Sapindaceae	<i>Paullinia</i>	2
Apocynaceae	<i>Rauvolfia</i>	2
Polygonaceae	<i>Rheum</i>	2
Hippocastanaceae	<i>Aesculus</i>	1

Family	Genera	Cp-count
Ericaceae	<i>Arctostaphylos</i>	1
Sterculiaceae	<i>Cola</i>	1
Hamamelidaceae	<i>Hamamelis</i>	1
Pedaliaceae	<i>Harpagophytum</i>	1
Convolvulaceae	<i>Operculina</i>	1
Monimiaceae	<i>Peumus</i>	1
Anacardiaceae	<i>Schinus</i>	1
Loganiaceae	<i>Strychnos</i>	1

## Discussion

### Evolution of Brazilian traditional medicinal plants

Brazil is endowed with unique medicinal plant resources, which play a crucial role in sustaining human health. The history of the BP has been explained well and contains data to understand the evolution of traditional medicinal plants. The first edition of the BP, containing the list of plant species used in conventional and traditional medicine, is still considered outstanding [50]. This edition also includes 196 monographs about the native herbs not in any other Pharmacopoeia [51] and 42% of medicinal products related to plants [52]. In the early 20th century, the name of traditional Brazilian medicine was generated based on a combination of indigenous knowledge and understanding of those brought by Portuguese, Spaniards, and Africans [53]. The Brazilian society transformed severely [54] after the release of the first edition. At around 1940, due to the growth of the global pharmaceutical industry and lack of pharmacological knowledge on the efficacy and toxicity of medicinal plants [55, 56], the Brazilians reorganized the pharmaceutical industry to produce drug formulas using imported materials instead of native ones [57, 58]. Later, during World War II, synthetic medicines strongly substituted the traditional medicinal plants [57, 59, 60], which is evident in the subsequent two BP editions. The monographs on plants reduced from 713 in the first edition to 205 in the second edition and 23 in the third edition [43, 61]. Several studies have shown that deforestation affecting the natural ecosystem and devaluation of phytotherapy have threatened traditional medicinal plants and medical knowledge [18, 62, 63]. Further, the number of medicinal plants began to grow as the public started favoring phytotherapy. Consequently, the monographs for medicinal plants increased to 17% in the fourth edition [43, 50, 64] and 25% in the fifth edition [51]; finally, the sixth edition (2019) included a section dedicated to medicinal plants with 149 monographs of plants and plant derivatives. The increase in public interest in herbal drugs also indicated the recovery of phytotherapy. However, considering the Brazilian biodiversity and the rich phytotherapy history, the patents and medicines based on BP are few [3]. Millions of people in Brazil would benefit from the use of medicinal plants in health care [65]; the expansion of phytotherapy based on advanced

approaches can improve the quality of life of people and contribute to the economic and technological development of the country.

### **The era of “herbgenomics”**

Pathogenic microbes, toxic compounds [66, 67], adulterations, and counterfeit drugs [68–71] have raised significant concerns about the quality and safety of traditional medicines. The classical analytical procedures provided by monographs from BP have projected the difficulties in solving these complex problems related to both exotic and native herbs. Additionally, the natural ecosystem destruction observed in Brazil [60, 72, 73] demands the conservation of Brazilian medicinal plant resources. With the global impact of genomics and due to the increasing damage to the domestic ecological environment, the local government and scientists also made efforts to develop genomics. The Brazilian government started building a genetic database in 1984. The Brazilian government organization for Agricultural Research (EMBRAPA) created the National Genetic Resources and Biotechnology Research Center (CENARGEN), aiming at managing the activities related to genetic resources in Brazil [15].

Research on medicinal plants has recently entered the “herbgenomics” era with the development of advanced technologies, especially genomics, transcriptomics, and proteomics. Omics-based analysis has helped understand the genetic makeup of traditional medicines revealing their origin, quality, synthesis, safety, and conservation requirements. These technologies will promote the sustainable development of medicinal plants, progressing in human health [74]. The emergence of herbgenomics enabled systematic research on plant-based medicines from a genetic perspective [8]. A publicly accessible database will act as the key hub to efficiently access and interpret the available data sets. Therefore, to support researchers in handling the rapidly changing data, the development of databases is crucial and challenging, which requires user-friendly design, accurate and comprehensive data updates, and continuous maintenance.

## **Conclusions**

The database, BPGD, acts as a bridge connecting the cutting-edge genetic approaches and findings with the traditional Brazilian medicinal plants and ethnopharmacy. BPGD is the first database established collecting genomic data (DNA barcoding, transcriptome, nuclear genome, and cp-G) of medicinal plants listed in the BP. The database has a ready-to-use data storage function and includes description, molecular identification, and functions of the medicinal plants listed in the BP. BPGD provides a platform for biological researchers to access the published genetic data quickly and effectively, accelerating research on traditional Brazilian medicinal plants and facilitating the rational development on their market regulation. BPGD has covered more than 70% medicinal plant species in the last edition of BP and will be regularly updated and maintained to provide more comprehensive data and a better user experience.

## **Abbreviations**

BPGD: Brazilian Pharmacopoeia Genomic Database

BLAST: Basic Local Alignment Search Tool

BP: Brazilian Pharmacopoeia

NCBI: National Centre for Biotechnology Information

EBI: European Institute of Bioinformatics

DDBJ: DNA Database of Japan

LNMP: Nginx, PHP, MySQL, phpMyAdmin

HTTP: HyperText Transfer Protocol

PHP: Hypertext Preprocessor

ITS: Internal transcribed spacer

NCBI SRA: National Centre for Biotechnology Information, Sequence Read Archive database

CNGB: China National GeneBank DataBase

LSC: [Large single-copy](#)

ndhF: NADH dehydrogenase F

MAFFT: Multiple Alignment using Fast Fourier Transform

EMBOSS: European Molecular Biology Open Software Suite

## **Declarations**

### **Availability of data and materials**

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Funding

Not applicable.

## Acknowledgments

This work was supported by the National Key R&D Program of China from the Ministry of Science and Technology of China (grant No. 2019YFC1711100), the National Natural Science Foundation of China and Karst Science Research Center of Guizhou Province (U1812403-1), and the Special Foundation for National Science and Technology Basic Research Program of China (2018FY100701).

## Authors' contributions

Guanru Zhou and Baosheng Liao contributed equally to this work.

Conception and design: Jiang Xu and Shilin Chen. Acquisition, analysis and interpretation of data: Guanru Zhou, Baosheng Liao, Qiushi Li. Writing of the manuscript: Guanru Zhou, Baosheng Liao. All authors read and approved the final manuscript

## References

1. Fürst R, Zündorf I. Evidence-based phytotherapy in Europe: where do we stand. *Planta Med.* 2015;81(12–13):962–7.
2. De Oliveira DR. *A consolidação das práticas integrativas e complementares no século 21.* VITTALE-Revista de Ciências da Saúde, 2018. 30(1): p. 7–8.
3. Dutra RC, et al. Medicinal plants in Brazil: Pharmacological studies, drug discovery, challenges and perspectives. *Pharmacol Res.* 2016;112:4–29.
4. Ricardo L, Goulart E, Brandão M. Plantas medicinais da Bacia do Rio das Velhas: avaliação das condições para produção e uso em saúde pública. *Revista Brasileira de Plantas Medicinais.* 2015;17(3):398–406.
5. Azwanida N. A review on the extraction methods use in medicinal plants, principle, strength and limitation. *Med Aromat Plants.* 2015;4(196):2167–0412.
6. Newman DJ, Cragg GM. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J Nat Prod.* 2012;75(3):311–35.
7. Cragg GM, Newman DJ. *Natural products: A continuing source of novel drug leads.* *Biochimica et Biophysica Acta (BBA) - General Subjects,* 2013. 1830(6): p. 3670–3695.
8. Hu H, et al. Herbgenomics: A stepping stone for research into herbal medicine. *Sci China Life Sci.* 2019;62(7):913–20.

9. Shi-lin C, et al. Highlights on the progress of traditional Chinese medicine and natural drugs during 2015–2020[J]. *Acta Pharmaceutica Sinica*. 2020;55(12):2751–76.
10. Pedersen D, Baruffati V. Health and traditional medicine cultures in Latin America and the Caribbean. *Soc Sci Med*. 1985;21(1):5–12.
11. Harrison IE, Cosminsky S. Traditional medicine: implications for ethnomedicine, ethnopharmacology, maternal and child health, mental health, and public health: an annotated bibliography of Africa, Latin America, and the Caribbean. Vol. 147: Taylor & Francis; 1976.
12. Gurib-Fakim A. Medicinal plants: Traditions of yesterday and drugs of tomorrow. *Mol Aspects Med*. 2006;27(1):1–93.
13. IBGE, *Censo 2010: população indígena é de 896,9 mil, tem 305 etnias e fala 274 idiomas*. 2012.
14. Calixto JB. Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). *Brazilian Journal of Medical Biological Research*. 2000;33(2):179–89.
15. Vieira RF, Skorupa LA. BRAZILIAN MEDICINAL PLANTS GENE BANK. Leuven: International Society for Horticultural Science (ISHS); 1993.
16. Wolters B. Jarhtausend vor Kolumbus: Indianische kulturpflanzen und Arzneidroge. *Deutsche Apotheker Zeitung*. 1992;40:1–10.
17. Marchant A. Dom Joao's botanical garden. *Hispanic American Historical Review*. 1961;41(2):259–74.
18. Bruning MCR, Mosegui GBG, Vianna CMdM, *A utilização da fitoterapia e de plantas medicinais em unidades básicas de saúde nos municípios de Cascavel e Foz do Iguaçu-Paraná: a visão dos profissionais de saúde*. *Ciência & saúde coletiva*, 2012. **17**: p. 2675–2685.
19. Ríos AM, *Pharmaceutical industry in Brazil - Statistics & Facts*. 2020.
20. Lapere Y, Pavan A, Knutsson N, *THE HEALTH INDUSTRY IN BRAZIL*. 2018.
21. Urdang G. The development of pharmacopoeias; a review with special reference to the pharmacopoea Internationalis. *Bull World Health Organ*. 1951;4(4):577–603.
22. Anvisa ANd.V.S.-. *Farmacopeia Brasileira 6ª edição*. 2019.
23. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov*. 2015;14(2):111–29.
24. Sheridan C. Recasting natural product research. *Nat Biotechnol*. 2012;30(5):385–7.
25. Hebert P, Cywinska A, *Biological identifications through DNA barcodes*. 2003.
26. Li X, et al. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90(1):157–66.
27. Wang X, et al. HMOD: An Omics Database for Herbal Medicine Plants. *Mol Plant*. 2018;11(5):757–9.
28. Callaway E, Cyranoski D. Anti-parasite drugs sweep Nobel prize in medicine 2015. *Nature News*. 2015;526(7572):174.
29. Klayman DL. Qinghaosu (artemisinin): an antimalarial drug from China. *Science*. 1985;228(4703):1049–55.

30. Weaver BA. How Taxol/paclitaxel kills cancer cells. *Molecular biology of the cell*. 2014;25(18):2677–81.
31. Xu J, et al. *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience*. 2017;6(11):1–15.
32. Guo L, et al. The opium poppy genome and morphinan production. *Science*. 2018;362(6412):343–7.
33. Zhibin G, Zailin L, Er L. A Rare and Endangered Medicinal Plant: *Dendrobium nobile* [J]. *Chinese Wild Plant Resources*. 2002;4:011.
34. Bantawa P, et al. Status and conservation threats of *Picrorhiza scrophulariiflora* Pennell. (Scrophulariaceae): An endangered high valued medicinal plant of Indo-China Himalayan region. 3: *Bioremediation, Biodiversity, Bioavailability*; 2009. pp. 15–22.
35. Chik W-I, et al. *Saussurea involucreata*: A review of the botany, phytochemistry and ethnopharmacology of a rare traditional herbal medicine. *J Ethnopharmacol*. 2015;172:44–60.
36. Bevan MW, et al. Genomic innovation for crop improvement. *Nature*. 2017;543(7645):346–54.
37. Cordell GA. Ecopharmacognosy and the responsibilities of natural product research to sustainability. *Phytochem Lett*. 2015;11:332–46.
38. Woldegiorgis ST, et al. Rice Stress-Resistant SNP Database. *Rice*. 2019;12(1):1–12.
39. Portwood JL, et al. MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic acids research*. 2019;47(D1):D1146–54.
40. Alaux M, et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome biology*. 2018;19(1):1–10.
41. Gelbart WM. Databases in Genomic Research. *Science*. 1998;282(5389):659.
42. Kim NH, et al. Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant biotechnology journal*. 2018;16(11):1904–17.
43. Maria G. Lins Brandão<sup>1\*</sup> GPC. Raquel Assis Moreira<sup>1</sup>, Roberto L.M. Monte-Mor<sup>2</sup>, *Medicinal plants and other botanical products from the Brazilian Official Pharmacopoeia*. *Brazilian Journal of Pharmacognosy*; 2006.
44. Shen X, et al., *Complete Chloroplast Genome Sequence and Phylogenetic Analysis of the Medicinal Plant Artemisia annua*. *Molecules*, 2017. 22(8).
45. Nakamura T, et al. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*. 2018;34(14):2490–2.
46. Anvisa ANd.V.S.-. *The Second Supplement of the Brazilian Pharmacopoeia, 5th edition*. 2017.
47. Tabajara deO, Martins D, et al. The historical development of pharmacopoeias and the inclusion of exotic herbal drugs with a focus on Europe and Brazil. *J Ethnopharmacol*. 2019;240:111891.
48. Carvalho ACB, et al. The Brazilian market of herbal medicinal products and the impacts of the new legislation on traditional medicines. *J Ethnopharmacol*. 2018;212:29–35.

49. Matos FdA, et al. Constituintes químicos ativos e propriedades biológicas de plantas medicinais brasileiras. Fortaleza: Editora UFC; 2004.
50. Brandão MGL, et al. Traditional uses of American plant species from the 1st edition of Brazilian Official Pharmacopoeia. *Revista Brasileira de Farmacognosia*. 2009;19:478–87.
51. Anvisa ANd.V.S.-. *Brizilian Pharmacopoeia Volume 1 5th Edition*. 2010.
52. Leite PM, Camargos LM, Castilho RO. Recent progress in phytotherapy: A Brazilian perspective. *European Journal of Integrative Medicine*. 2021;41:101270.
53. Santos AC, et al., *Current use of native medicinal species recorded by European Naturalists in the 19th century along the Royal Road, Minas Gerais, Brazil*.
54. De Carvalho ACD. Feiticeiros, burlões e mistificadores: criminalidade e mudança das práticas populares de saúde em São Paulo, 1950 a 1980. Unesp; 2003.
55. Brandão MGL, et al. Influence of Brazilian herbal regulations on the use and conservation of native medicinal plants. *Environ Monit Assess*. 2010;164(1):369–77.
56. Carvalho AC, et al. Situação do registro de medicamentos fitoterápicos no Brasil. *Revista Brasileira de Farmacognosia*. 2008;18(2):314–9.
57. Fernandes TM. Plantas medicinais: memória da ciência no Brasil. Editora Fiocruz; 2004.
58. Lucchesi G, *Dependência e autonomia no setor farmacêutico: um estudo da CEME*. 1991.
59. Manhã EM, et al. PLANT: A bibliographic database about medicinal plants. *Revista Brasileira de Farmacognosia*. 2008;18:614–7.
60. Brandão M, Montemor R. Sabedoria antiga em risco. *Cienc Hoje*. 2008;42:77–9.
61. Brandão MGL, et al. Other medicinal plants and botanical products from the first edition of the Brazilian Official Pharmacopoeia. 18: *Revista Brasileira de Farmacognosia*; 2008. pp. 127–34.
62. Maziero M, Teixeira MP, *A EXPANSÃO DA UTILIZAÇÃO DE FITOTERÁPICOS NO BRASIL*. Anais do Salão Internacional de Ensino, Pesquisa e Extensão, 2017. **9**(2).
63. Ricardo LM, et al., *Plants from the Brazilian traditional medicine: species from the books of the Polish physician Piotr Czerniewicz (Pedro Luiz Napoleão Chernoviz, 1812–1881)*. *Revista Brasileira de Farmacognosia*, 2017. **27**(3): p. 388–400.
64. Brandao MG, et al. Brazilian medicinal plants described by 19th century European naturalists and in the Official Pharmacopoeia. *J Ethnopharmacol*. 2008;120(2):141–8.
65. Alencar NL, Santoro FR, Albuquerque UP. What is the role of exotic medicinal plants in local medical systems? A study from the perspective of utilitarian redundancy. *Revista Brasileira de Farmacognosia*. 2014;24(5):506–15.
66. Chen S, et al., *A renaissance in herbal medicine identification: From morphology to DNA*. *Biotechnology advances*, 2014. 32.
67. Gilbert, Natasha. Herbal medicine rule book. *Nature*. 2011;480(7378):22–9.
68. Poon WT, et al. *Aconite poisoning in camouflage*. *Hong Kong medical journal = Xianggang yi xue za zhi /*. Hong Kong Academy of Medicine. 2006;12(6):456–9.

69. Lowe L, Matteucci MJ, Schneir AB. Herbal aconite tea and refractory ventricular tachycardia. *N Engl J Med*. 2005;353(14):1532.
70. Chan T. *Incidence of Herb-Induced Aconitine Poisoning in Hong Kong*. 2002. 25(11): p. 823–828.
71. Han J, et al. An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Sci Rep*. 2016;6:18723.
72. Dean W, *A ferro e fogo: a história e a devastação da Mata Atlântica brasileira*, in *A ferro e fogo: a história e a devastação da Mata Atlântica brasileira*. 1996. p. 484–484.
73. Shanley P, Rosa NdA, *Conhecimento em Erosão: Um Invetário Etnobotânico na Fronteira de Exploração da Amazônia Oriental*. 2005.
74. Chen S, et al. Herbal genomics: examining the biology of traditional medicines. *Science*. 2015;347(6219):S27–9.

## Figures

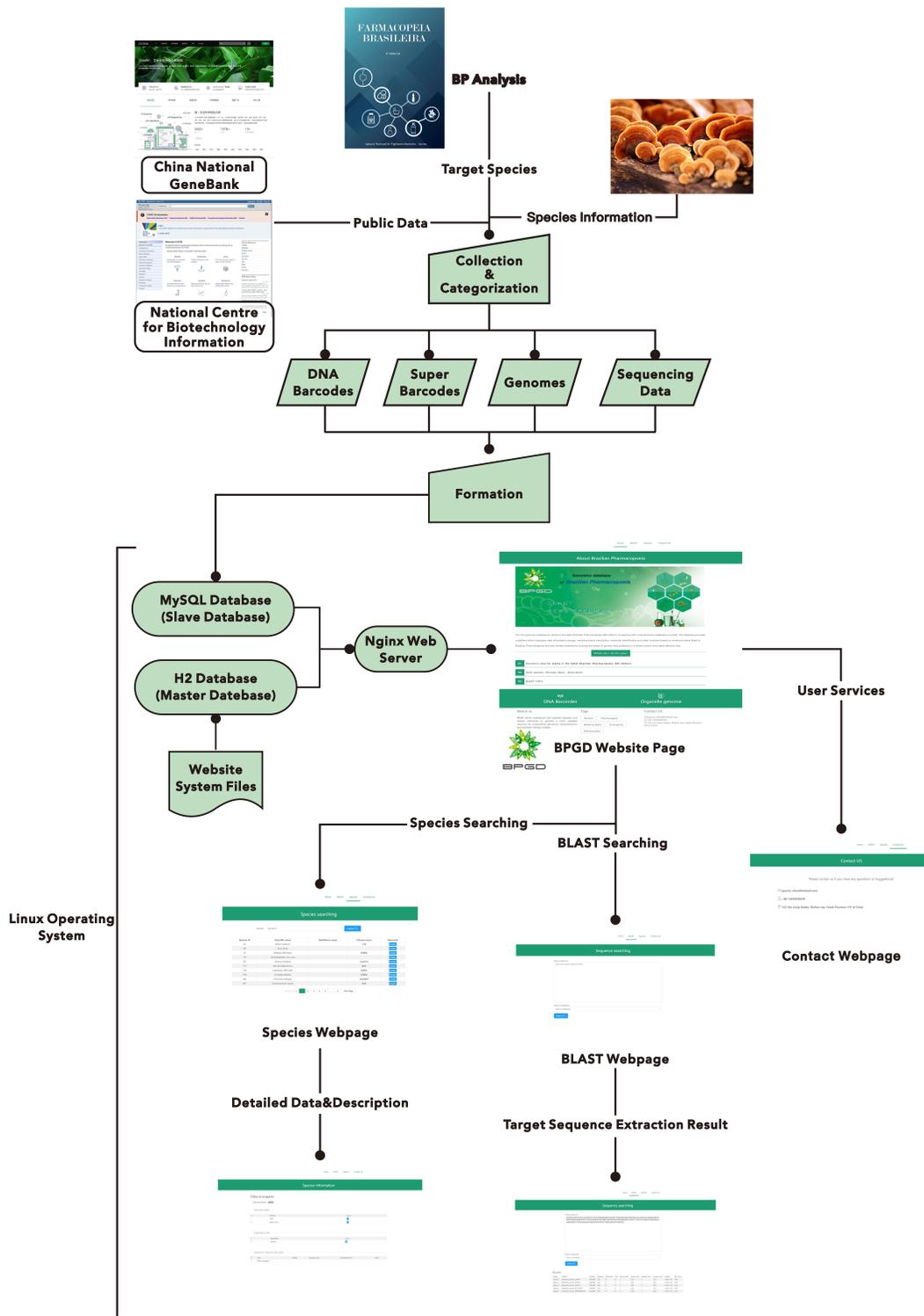
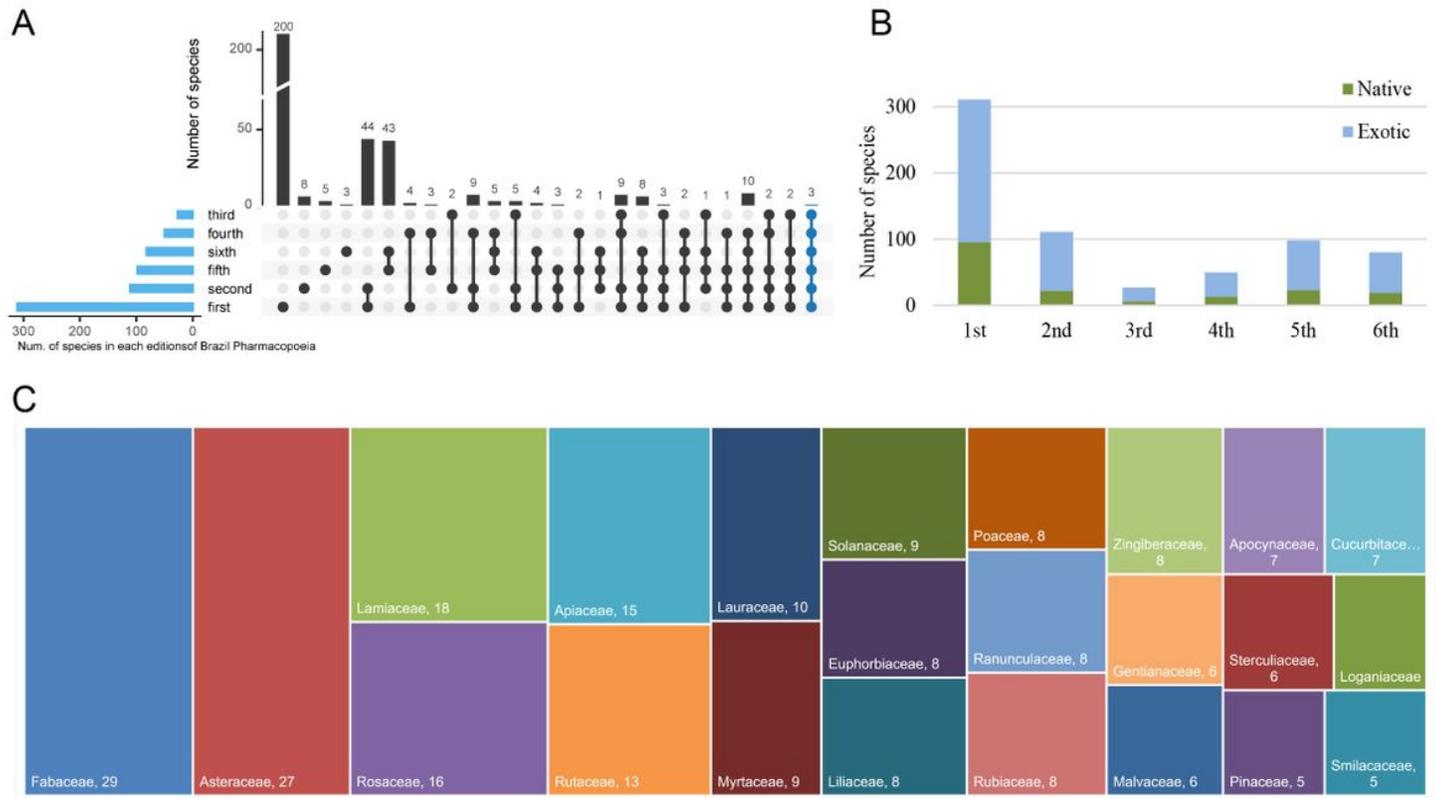


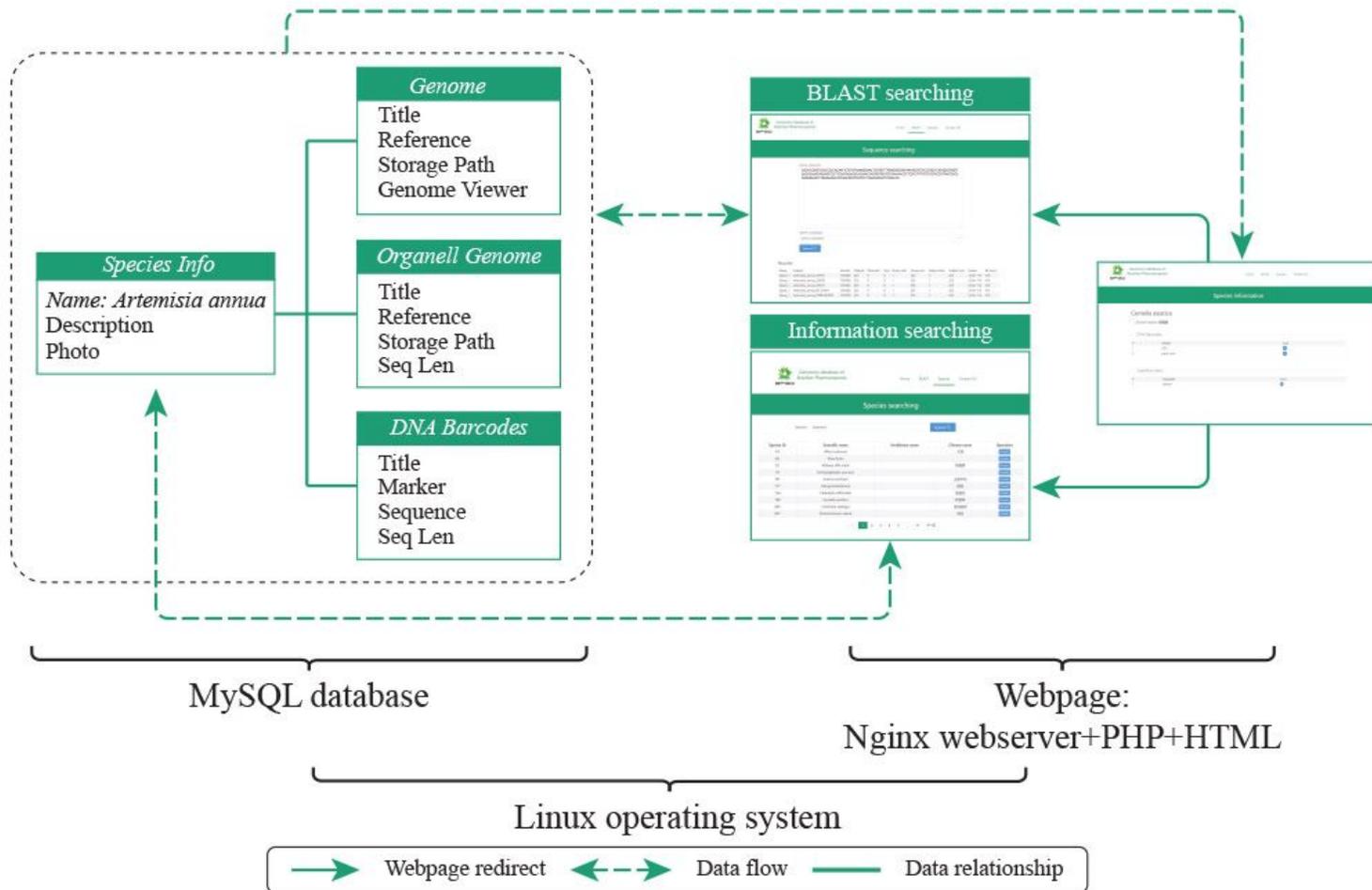
Figure 1

Schematic overview of the BPGD database construction pipeline



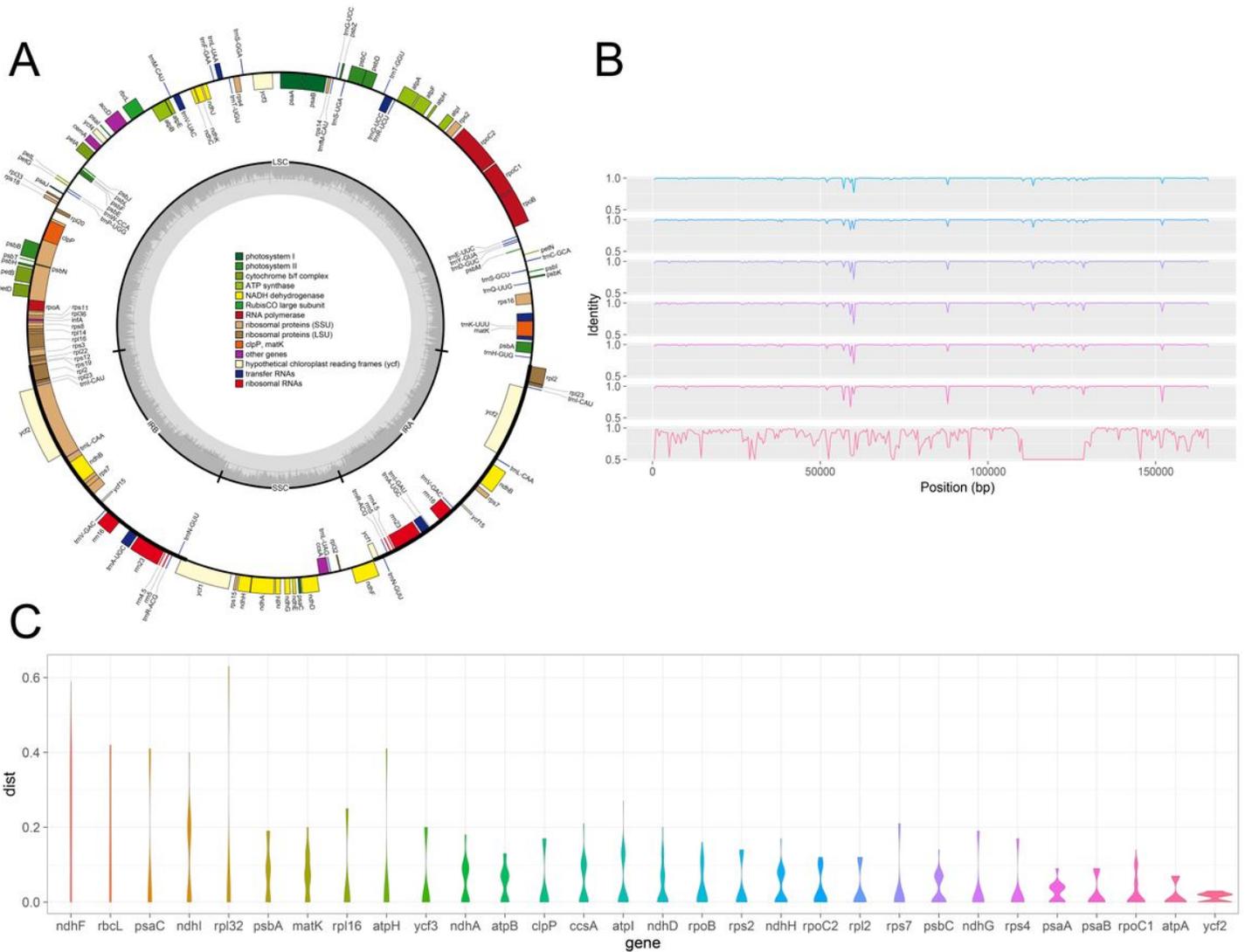
**Figure 2**

Summary of medicinal species recorded in Brazilian Pharmacopoeia. (A) Medicinal species information of different editions of Brazilian Pharmacopoeia, dot represents its horizontally aligned edition of BP, line connecting dots represents species intersection of different editions of BP. (B) Native and exotic species in six editions of Brazilian Pharmacopoeia. (C) Twenty-two families with most species recorded in Brazilian Pharmacopoeia, the area of rectangle represents number of species in the family.



**Figure 3**

BPGD structure and core web pages.



**Figure 4**

Chloroplast genome analysis. (A) Chloroplast genome of *Aloe vera*. (B) Whole-sequence alignment of three *Aloe* species, using *Belamcanda chinensis* as the outgroup. (C) Pairwise distance for each gene in the chloroplast genome of three *Aloe* species.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)