Development of a DNA barcode library of plants in the Thai Herbal Pharmacopoeia and Monographs for authentication of herbal products

Santhosh Kumar J. Urumarudappa1,2, Chayapol Tungphatthong1, Jirayut Jaipaew3, Natapol Pornputtapong3,4, Duangkamol Pakdeesattayapong4, Sornkanok Vimolmangkang3 & Suchada Sukrong1,2*

Traditional herbal medicine has long been practiced as a method of health care in many countries worldwide. The usage of herbal products has been increasing and is expected to continue to do so in the future. However, admixture and adulteration are concerns regarding the quality of herbal medicine, including its safety and efficacy. We aimed to develop a reference DNA barcode library of plants listed in the Thai Herbal Pharmacopoeia (THP) and Monographs of Selected Thai Materia Medica (TMM) (n = 101 plant species) using four core barcode regions, namely, the ITS2, matK, rbcL and trnH-psbA intergenic spacer regions, for authentication of the plant origin of raw materials and herbal products. Checking sequences from samples obtained from local markets and the Thai Food and Drug Administration (Thai FDA) against our digital reference DNA barcode system revealed the authenticity of eighteen out of twenty tested samples as claimed on their labels. Two samples, no. 3 and 13, were not Cyanthillium cinereum (L.) H.Rob. and Pueraria candollei Wall. ex Benth. as claimed, respectively. They were recognized as Emilia sonchifolia (L.) DC. and Butea superba (Roxb.), respectively. Hence, it is important for the Thai FDA or regulatory agencies to immediately initiate strict enforcement for the development of pharmacopoeial standards as well as revisions or modifications of available regulatory guidelines and to implement close monitoring for the quality control of herbal products in terms of authentication before they enter the herbal market. The centralized digital reference DNA barcode database developed here could play a very important role in monitoring or checking the authenticity of medicinal plants.

Traditional herbal medicine has long been practiced in health care systems in many countries worldwide. The global trade of herbal remedies and supplements is estimated to increase every year and is expected to reach approximately USD$ 117.02 billion by 20241. The usage of herbal products has gained significant momentum in the recent past and is expected to continue to increase in the near future. In Thailand, traditional Thai medicine (TTM) was the most conventional healthcare system until the establishment of modern healthcare systems. Consequently, as a result of many social and economic status changes, the use of TTM became limited to indigenous Thai people. However, the government has been trying to rejuvenate TTM to benefit the Thai medical system, especially in rural areas4,5. The quality parameters of herbal products are generally documented in the Thai Herbal Pharmacopoeia (THP) and Monographs of Selected Thai Materia Medica (TMM), the two reference textbooks endorsed by the Thai government.

The THP, currently in its 2021 edition6, was first established in 1989 by the Bureau of Drug and Narcotics, Department of Medical Sciences, Ministry of Public Health, to set forth quality standards for plants or herb-based

1Center of Excellence in DNA Barcoding of Thai Medicinal Plants, Chulalongkorn University, Bangkok 10330, Thailand. 2Department of Pharmacognosy and Pharmaceutical Botany, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand. 3Department of Biochemistry and Microbiology, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand. 4Herbal Products Division, Food and Drug Administration, Ministry of Public Health, Nonthaburi 11000, Thailand. *email: suchada.su@chula.ac.th
drugs and herbal product preparations marketed in Thailand to ensure their identity, quality, safety, and efficacy. Intentional or unintentional adulteration of herbs leads to lower efficacy and affects herbal trade1–10. The common traditional authentication process of herbal products includes methods of botanical identification such as plant taxonomy, microscopic and macroscopic examination, and advanced chemical methods11. However, each method has advantages and limitations. The most frequent approaches are macroscopic and microscopic identification, which are fast and cost-effective qualitative techniques. However, macroscopic analysis requires the whole plant, and it is difficult to apply to forms where plant morphology cannot be determined, for instance, mixtures of multiple herbs or extracted samples12. Phytochemical approaches or metabolomics profiling has been used for the identification of botanical drugs, dietary or food supplements and plant extracts13. Generally, phytochemical authentication depends on the selection of chemical markers that are unique to the selected plant species and is not always successful due to variation in geographical location and environmental conditions, including soil type, plant age, plant part, processing, storage conditions and other factors8. In addition, phytochemical analysis requires more reference samples from multiple populations to account for natural variability14. Among all recently developed methods, DNA-based methods are well-established for identifying plants in mixtures of herbal medicine products15,16.

A precise assessment is of foremost significance for purchasers, customers, patients and researchers along herbal product value chains15 including collectors, processors, harvesters, producers, regulators, traders, distributors, retailers, and traditional and medical practitioners16. Today, the armamentarium of prescription treatment is communicated through ‘pharmacopoeias’, which are standard collections of information on the quality of pharmaceutical drugs, excipients and flavoring correctives. The pharmacopoeia includes information on testing methodologies, purity, storage guidelines, composition and concentration for drugs. Pharmacopoeias ensure the consistency of cures endorsed by delegates of a particular unit and outline required quality principles. However, the regulatory affairs or policies for natural herbal products differ among nations. In some countries, for example, Canada, the United States and countries in the European Union (EU), governing regulatory agencies assess the quality and safety of herbal drugs/medicines before they enter the herbal market, but in practice, activities to control the authenticity and quality of herbal products in the herbal market appear to be limited19. The European Medicines Agency (EMA) updates the European Pharmacopoeias, including the monographs and testing methods in their database20, and the databases provide the most recent monographs and suitable methods for quality estimations of particular herbal drug products21,22. With accurate and rapid DNA-based techniques, DNA barcoding is now officially recognized as a method for identifying herbal drugs method23. DNA barcoding for quality control of herbal drugs is included in the British Pharmacopoeia (BP)22,23, Pharmacopoeia of the People’s Republic of China24 and Korean Pharmacopoeia25, which includes plant sampling, DNA isolation, PCR amplification and development of standard reference sequence databases9.

Herein, we aimed to develop a digital reference DNA barcode library of plants listed in the THP and TMM using the nuclear and chloroplast DNA regions and to test for species adulteration in selected herbal products obtained from local markets and the Thai FDA. The centralized digital DNA barcode database developed here will also aid in the identification of any botanicals or herbal products in registration or regulatory processes.

Results

DNA barcoding of selected plants in the THP and TMM. Genomic DNA was successfully extracted from all 101 plant species belonging to 89 genera and 51 families (Table S1). The core DNA barcode regions, namely, the ITS2, matK, rbcL and trnH-psbA intergenic spacer regions, were amplified. In the PCRs, positive and negative control amplifications gave accurate results. All PCR amplicons were clearly segregated and visible as single bands of the expected size. The partial sequence lengths ranged between 228 and 278 bp (average 258) for ITS2, 424 and 478 bp (average 450) for matK, 540 and 580 bp (average 550) for rbcL and 420 and 458 bp (average 428) for the trnH-psbA intergenic spacer. All nucleotide sequences were submitted to NCBI GenBank, and their accession numbers are listed in Table 1.

Authentication of herbal products. Genomic DNA was successfully isolated from all twenty different dosage forms of herbal products (Fig. 1; Table S2) and amplified for four barcode regions, namely, the ITS2, matK, rbcL and trnH-psbA intergenic spacer regions. Furthermore, the authenticity of all twenty samples of single-herb formulation products was tested using our reference DNA barcode database and nucleotide Basic Local Alignment Search Tool (BLAST) analysis of available NCBI GenBank sequences (Table S3). The results confirmed the authenticity of eighteen out of the twenty samples tested. The sequences obtained from the other two samples, no. 3 and 13, which were purchased from local markets, did not match the name on their labels (Table 2). Sample no. 3 was labeled as Cyanthillium cinereum and sample no. 13 was labeled as Pueraria candolleti. However, our nucleotide BLAST results showed that sample no. 3 and 13 were Emilia sonchifolia and Butea superba, respectively. All samples provided by the Thai FDA were correct according to their claims. The NCBI GenBank nucleotide blast results of these samples are provided in Table 2.

Maximum likelihood phylogenetic analysis. Maximum likelihood (ML) phylogenetic analysis of all reference plant species was performed using the ITS2, matK, rbcL, and psbA-trnH regions. The unrooted phylogenetic tree of the rbcL region showed clear clades, and each cluster represented a specific group of plant species (Fig. 2). Each color represents a monophyletic clade based on plant genera and families, indicating their close phylogenetic relationships. A large number of plant species clusters belonged to the Asteraceae, Fabaceae, Lamiaceae, Rutaceae, and Zingiberaceae families. The bootstrap values were estimated with 1000 replicates with support values. These findings showed that the rbcL region-based phylogenetic tree can be used as an efficient
| No | Botanical name                  | Family name | Monograph | Voucher number/ID | Collection location | GenBank accession numbers |
|----|--------------------------------|-------------|-----------|------------------|---------------------|--------------------------|
| 1  | Syzygium aromaticum (L.) Merr. & L.M. Perry | Myrtaceae   | I         | FPSCU SS-043     | FPSCU               | LC435390, LC435391, LC435392, LC435393 |
| 2  | Cinnamomum camphora (L.) Presl | Lauraceae   | I         | FPSCU SS-044     | FPSCU               | LC435394, LC435395, LC435396, LC435397 |
| 3  | Strychnos nux-vomica L.     | Loganiaceae | I         | FPSCU SS-109     | FPSCU               | LC461741, LC461742, LC461743, LC461744 |
| 4  | Terminalia chebula Retz     | Combretaceae| I         | FPSCU SS-017     | FPSCU               | LC435434, LC435435, LC435436, LC435437 |
| 5  | Carica papaya L.            | Caricaceae  | I         | FPSCU SS-002     | FPSCU               | LC461717, LC461718, LC461719, LC461720 |
| 6  | Zingiber officinale Rosc.   | Zingiberaceae| I         | FPSCU SS-033     | FPSCU               | LC461745, LC461746, LC461747, LC461748 |
| 7  | Cassia fistula L.           | Fabaceae    | I         | FPSCU SS-051     | FPSCU               | LC435398, LC435399, LC435400, LC435401 |
| 8  | Santalum album L.           | Santalaceae | I         | FPSCU SS-025     | FPSCU               | LC435402, LC435403, LC435404, LC435405 |
| 9  | Pierocarpus santalinus Lf.  | Fabaceae    | I         | FPSCU SS-026     | Bangkok             | LC461725, LC461726, LC461727, LC461728 |
| 10 | Plumbago zeylanica L.       | Plumbaginaceae | I      | FPSCU SS-054     | Bangkok             | LC435406, LC435407, LC435408, LC435409 |
| 11 | Aegle marmelos L.           | Rutaceae    | I         | FPSCU SS-055     | FPSCU               | LC435410, LC435411, LC435412, LC435413 |
| 12 | Senna alata (L.) Roxb.      | Fabaceae    | I         | FPSCU SS-004     | FPSCU               | LC435426, LC435427, LC435428, LC435429 |
| 13 | Centella asiatica (L.) Urb | Apiaceae    | I         | FPSCU SS-012     | FPSCU               | LC461753, LC461754, LC461755, LC461756 |
| 14 | Vetiveria zizanioides (L.) Nash ex Small | Poaceae   | I         | FPSCU SS-110     | FPSCU               | LC461921, LC461922, LC461923, LC461924 |
| 15 | Piper nigrum L.             | Piperaceae  | I         | FPSCU SS-005     | FPSCU               | LC461757, LC461758, LC461759, LC461760 |
| 16 | Anogeissus latifolius (Burm. f.) Wall. ex Nees | Fabaceae | I         | FPSCU SS-007     | FPSCU               | LC461761, LC461762, LC461763, LC461764 |
| 17 | Phyllanthus emblica L.      | Euphorbiaceae| I         | FPSCU SS-015     | FPSCU               | LC435430, LC435431, LC435432, LC435433 |
| 18 | Dracaena cochinchinensis (Lour.) S.C.Chen | Dracaenaceae | I      | FPSCU SS-032     | FPSCU               | LC461765, LC461766, LC461767, LC461768 |
| 19 | Terminalia bellirica (Gaertn.) Roxb | Combretaceae | I       | FPSCU SS-018     | FPSCU               | LC438866, LC438867, LC438868, LC438869 |
| 20 | Aegle marmelos L.           | Rutaceae    | I         | FPSCU SS-093     | FPSCU               | LC461769, LC461770, LC461771, LC461772 |
| 21 | Cyperus rotundus L.         | Cyperaceae  | I         | FPSCU SS-149     | Bangkok             | LC461773, LC461774, LC461775, LC461776 |
| 22 | Boscia albitrunca (L.) Manel | Zingiberaceae| I         | FPSCU SS-037     | FPSCU               | LC461777, LC461778, LC461779, LC461780 |
| 23 | Oxicecum tenellaform L.     | Lamiaceae   | I         | FPSCU SS-001     | FPSCU               | LC461781, LC461782, LC461783, LC461784 |
| 24 | Pluca indica (L.) Less      | Asteraceae  | I         | FPSCU SS-047     | FPSCU               | LC438882, LC438883, LC438884, LC438885 |
| 25 | Alpinia galanga (L.) Wild  | Zingiberaceae| I         | FPSCU SS-112     | FPSCU               | LC461785, LC461786, LC461787, LC461788 |
| 26 | Senna siamea (Lam.) H.S.Irwin & Barneby | Fabaceae | I         | FPSCU SS-010     | FPSCU               | LC438886, LC438887, LC438888, LC438889 |
| 27 | Aristolochia pinnata Less  | Aristolochiaceae| I       | MUS-5409        | Sakon Nakhon       | KP998796, KP998782, KP998768, KP998810 |
| 28 | Capparis micracantha DC    | Capparaceae | I         | FPSCU SS-060     | FPSCU               | LC438890, LC438891, LC438892, LC438893 |
| 29 | Myristica fragrans Houtt.   | Myristicaceae| I         | FPSCU SS-052     | FPSCU               | LC461925, LC461926, LC461927, LC461928 |
| 30 | Piper retrofractum Vahl     | Piperaceae  | I         | FPSCU SS-011     | FPSCU               | LC461929, LC461930, LC461931, LC461932 |
| 31 | Nelumbo nucifera Gaertn.    | Nymphaeaceae| I         | FPSCU SS-027     | Bangkok             | LC438878, LC438879, LC438880, LC438881 |
| 32 | Mesua ferrea L.             | Calophyllaceae| I         | FPSCU SS-028     | FPSCU               | LC461789, LC461790, LC461791, LC461792 |
| 33 | Kaempferia galanga L.       | Zingiberaceae| I         | FPSCU SS-068     | FPSCU               | LC461793, LC461794, LC461795, LC461796 |
| 34 | Mimusops elengi L.          | Sapotaceae  | I         | FPSCU SS-029     | FPSCU               | LC438870, LC438871, LC438872, LC438873 |
| 35 | Citrus hystrix DC           | Rutaceae    | I         | FPSCU SS-014     | FPSCU               | LC438898, LC438899, LC438900, LC438901 |
| 36 | Tomarindus indica L.        | Fabaceae    | I         | FPSCU SS-076     | FPSCU               | LC461733, LC461734, LC461735, LC461736 |
| 37 | Ficus racemosa L.           | Moraceae    | I         | FPSCU SS-077     | FPSCU               | LC461797, LC461798, LC461799, LC461800 |
| 38 | Aegle marmelos (L.) Correa  | Rutaceae    | I         | FPSCU SS-078     | FPSCU               | LC461801, LC461802, LC461803, LC461804 |
| No | Botanical name | Family name | Monograph | THP (year) | Voucher number/ID | Collection location | GenBank accession numbers |
|----|----------------|-------------|-----------|------------|------------------|---------------------|--------------------------|
| 43 | Jasminum sambac (L.) Sol | Oleaceae | II | – | FPSCU SS-114 | FPSCU | LC461805 LC461806 LC461807 LC461808 |
| 44 | Clerodendrum indicum (L.) Kurz | Lamiaceae | II | 2021 | FPSCU SS-082 | HPMSHG | LC461737 LC461738 LC461739 LC461740 |
| 45 | Tibacera triandra Diels | Menispermaceae | II | 2021 | FPSCU SS-083 | HPMSHG | LC438894 LC438895 LC438896 LC438897 |
| 46 | Brucea javanica (L.) Merr | Simaroubaceae | II | – | FPSCU SS-086 | HPMSHG | LC438902 LC438903 LC438904 LC438905 |
| 47 | Acacia concinna (L.) Merr | Fabaceae | II | – | FPSCU | HPMSHG | LC438902 LC438903 LC438904 LC438905 |
| 48 | Clerodendrum indicum (L.) Kuntze | Lamiaceae | II | 2021 | FPSCU SS-082 | HPMSHG | LC461933 LC461934 LC461935 LC461936 |
| 49 | Hibiscus sabdariffa L | Malvaceae | III | 2021 | FPSCU SS-036 | FPSCU | LC461809 LC461810 LC461811 LC461812 |
| 50 | Cananga odorata (Lam.) Hook. f. et Thomson var. Odorata | Annonaceae | III | – | FPSCU SS-038 | FPSCU | LC438906 LC438907 LC438908 LC438909 |
| 51 | Alocasia macrorrhiza (L.) G. Don | Araceae | III | – | FPSCU SS-116 | FPSCU | LC461937 LC461938 LC461939 LC461940 |
| 52 | Euphorbia antiquorum L | Euphorbiaceae | III | – | FPSCU SS-042 | FPSCU | LC389010 LC389011 LC389012 LC389013 |
| 53 | Artocarpus heterophyllus Lam | Moraceae | III | – | FPSCU SS-120 | FPSCU | LC461813 LC461814 LC461815 LC461816 |
| 54 | Aquilaria crassna Pierre ex Lecomte | Thymelaeaceae | III | – | FPSCU SS-040 | FPSCU | LC438910 LC438911 LC438912 LC438913 |
| 55 | Pueraria candollei Wall. ex Benth. var. mirifica (Airy Shaw et Suvat.) Niyomdham | Fabaceae | III | – | FPSCU SS-041 | FPSCU | LC456342 LC456343 LC456344 LC456345 |
| 56 | Eclipta prostrata (L.) L | Asteraceae | III | – | FPSCU SS-043 | FPSCU | LC456346 LC456347 LC456348 LC456349 |
| 57 | Streblus asper Lour | Moraceae | III | – | FPSCU SS-044 | FPSCU | LC456350 LC456351 LC456352 LC456353 |
| 58 | Jatropha multifida L | Euphorbiaceae | III | – | FPSCU SS-045 | FPSCU | LC456354 LC456355 LC456356 LC456357 |
| 59 | Clinacanthus nutans (Burm.f.) Lindau | Acanthaceae | III | 2021 | FPSCU SS-046 | FPSCU | LC456358 LC456359 LC456360 LC456361 |
| 60 | Piper betle L | Piperaceae | III | – | FPSCU SS-047 | FPSCU | LC456362 LC456363 LC456364 LC456365 |
| 61 | Garcinia mangostana L | Clusiaceae | III | – | FPSCU SS-048 | FPSCU | LC456366 LC456367 LC456368 LC456369 |
| 62 | Salacia chinensis L | Celastraceae | IV | – | FPSCU | FPSCU | LC456370 LC456371 LC456372 LC456373 |
| 63 | Cryptolepis dubia (Burm.f.) M.R.Almeida | Apocynaceae | IV | – | FPSCU SS-049 | FPSCU | LC456374 LC456375 LC456376 LC456377 |
| 64 | Annona squamosa L | Annonaceae | IV | – | FPSCU SS-050 | FPSCU | LC456378 LC456379 LC456380 LC456381 |
| 65 | Caesalpinia sappan L | Fabaceae | IV | – | FPSCU | FPSCU | LC456382 LC456383 LC456384 LC456385 |

Continued
resource for species authentication of Thai medicinal plants. Our unrooted ML phylogenetic tree mirrored the taxonomic classification of Thai plants listed in the THP and TMM (Fig. 2).

Development of a centralized reference DNA barcode database. In this study, a centralized digital reference DNA barcode system for regulating herbal products was developed. The reference DNA barcode database incorporates voucher numbers, scientific names, common names, Thai names, plant habitats, collection forms, plant photographs, herbarium images and other information, such as collection dates, collection locations, collectors, and taxonomists, along with geocoordinates (Fig. 3). All DNA barcode marker information, including genes, gene sequences, and GenBank accession numbers, will be included in the database. Using the scientific name or Thai name in the search option, the end user can obtain all the information for a particular plant. An attempt to establish a digital database system is made, and the database is found to be an efficient tool with which to systematically assess traditional medicine and its herbal products and connect it with both national and international herbal trade regulators. This database system is a novel concept in Thai herbal development, and its availability to the industry as well as consumers and researchers will bring a noticeable change in the regulation of herbal trade.

Discussion

The global markets of herbal drugs are large and increasing every year. However, increasing demand leads to adulteration or substitution in the raw materials. Many reports of adverse reactions may often be due to the consumption of unintended herbs, which has directly affected the marketing or campaign of herbal products. Various identification methods, including taxonomic, genomic, and phytochemistry methods, have been used to authenticate herbal products. However, each method has advantages and limitations. Recently, DNA-based methods have been widely established for the authentication of herbal products.
In this study, DNA barcodes of 101 highly traded medicinal plants listed in the THP and TMM of Thailand were developed. The highly traded samples of single-herb formulations that are not restricted to closely related plant species obtained from a local market and the Thai FDA were tested for their authenticity. Irrespective of the herbal samples, DNA analysis has been done using our own reference database along with available NCBI nucleotide blast analysis. Due to the inherent limitations of single-locus of DNA barcoding, an emerging DNA-based, or phylogenetic method is needed for the identification of closely related plant species. The utilization of DNA as a source of information for identifying inaccurate plant ingredients on herbal product labels is starting to be explored. Four core DNA barcode regions, namely, the ITS2, matK, rbcL, and trnH-psbA intergenic spacer regions, were used to develop a reference DNA barcode library for testing the authenticity of twenty single formulation herbal products. Our analysis indicated that all twenty samples tested for their authenticity were correct according to their labels, except samples no. 3 and no. 13, which were from powder and capsules labeled *Cyanthillium cinereum* and *Pueraria candollei*, respectively. Nucleotide BLAST results revealed that the *Cyanthillium cinereum* (sample no. 3) powder was replaced by *Emilia sonchifolia*, and the *Pueraria candollei* (sample no. 13) capsules contained instead *Butea superba*. Similar morphologies and confusion of vernacular names could explain this replacement. *Cyanthillium cinereum* has high antioxidant activity and is used in Thai medicine to reduce smoking withdrawal symptoms and treat skin ailments, as well as asthma, bronchitis, cough, cancer, malaria, gastrointestinal conditions, diuresis, pain, and diabetes. *Emilia sonchifolia* is used for the treatment of anti-inflammatory stomach tumors, ophthalmia, diarrhea, wounds, intestinal worm infections and bleeding piles. *Pueraria candollei* is used to relieve menopausal symptoms, including vasomotor symptoms, reproductive symptoms, depression, and musculoskeletal pain, in estrogen-deficient women. *Butea superba* has been used for rejuvenation, for sexual arousal, and to prevent erectile dysfunction. These results clearly indicate the extent of the problem that might occur due to the use of unauthentic raw drugs in Thai medicine. There were no *rbcL* reference sequences of sample no. 13 in NCBI GenBank; hence, the Barcode of Life Data System (BOLD) database was used to analyze this sample. Both of these samples were obtained from a local market. Herbal products purchased from local marketplaces could be more likely to obtain adulterations or admixtures, especially powder samples. It is very difficult to differentiate mixed powdered forms. Previously, many reports showed that the powdered form of samples had a greater chance of admixture than other forms, for example, the powdered form of ginger (*Zingiber officinale* Roscoe) admixed with chili powder (*Capsicum annuum* L.) and the powdered form of black pepper (*Piper nigrum* L.) admixed with chili powder (*Capsicum annuum* L.).
For the purpose of this study, an ML phylogenetic tree of our reference plant species was constructed using all four DNA barcode regions. Among the markers, \( rbcL \) is highly conserved, and its sequence query revealed the highest identity with plant species or closely related plant species. However, identification by this marker will not be reliable if the taxonomic identity of the nucleotide sequence in the GenBank database is incorrect. These issues can be resolved by using a phylogenetic tree wherein the incorrectly identified samples are highly likely to be located in unexpected clades. Our \( rbcL \) region phylogenetic tree showed the arrangement of all the plant species in appropriate clades or plant groups, as would be expected based on phylogenetic relationships among the plant species (Fig. 2). Therefore, taxonomic identification using the \( rbcL \) region at the species level is more reliable than other regions tested in this study. These results were consistent with those of previous reports that the \( rbcL \) region is a suitable candidate region for plant species identification. Previously, the utility of the \( rbcL \) region in discriminating land plants was successfully validated. The use of \( rbcL \) has increased due to its high discrimination proportions at low taxonomic levels. In this study, the \( matK \) and \( trnH-psbA \) regions were unable to differentiate the plant species, and the ITS2 region showed similar results, with a few of the plants of the same genus clustered with different groups of plants (Fig. S1). Therefore, this study was restricted to the \( rbcL \) region-based ML phylogenetic tree; however, multilocus DNA barcode techniques could be used as advanced tools for the accurate identification of medicinal plants.

Numerous adulteration and substitution studies of herbal products have been reported worldwide, including in Thailand. In addition, the international herbal product supply chain repeatedly lacks botanical expertise to provide suitable documentation for the identification of raw herbal materials. Unfortunately, in Thailand, there is no systematic regulatory mechanism for the quality control of herbal drugs before entering the market. It is very important to use appropriate analytical techniques for herbal products. Through this study, we propose a centralized digital DNA barcode database to aid in the regulatory step of identifying the plants used in herbal products. This reference database incorporates voucher numbers, scientific and common names, Thai names, plant habitats, collection forms, and plant photographs, including herbarium images, and other information such as collection dates, collection locations, and geocoordinates. By using scientific or common names, one can obtain all the information on a particular plant species or herbal product. This database could play a very important role in monitoring or checking medicinal plants or herbal trade and could ensure that all essential information is freely accessible to consumers and regulatory authorities in Thailand. Herbal testing centers and certification facilities will enhance the quality control of herbal products and help regulate the national and international herbal trade. Further, we are planning to extend the test to medicinal and non-medicinal plants available in Thailand. Future research will continue to validate and update the reference DNA barcode library and protocol or procedure for analyzing herbal samples. Furthermore, certification of the ingredients mentioned on herbal product labels using our reference DNA barcode database will continue.

### Table 2. Nucleotide sequence BLAST results of herbal products

| Sample code | Corresponding scientific names as per their label claim | Dosage form | NCBI BLAST result | Species identified using our reference DNA barcode library |
|-------------|--------------------------------------------------------|-------------|-------------------|---------------------------------------------------|
| 1           | Bacopa monnieri                                        | Powder      | BK                 | Bacopa monnieri                                   |
| 2           | Aristolochia pteri                                    | Crude drug  | AK                 | Aristolochia pteri                                |
| 3           | Cyanthillium cinereum                                 | Powder      | SC                 | Aristolochia pteri                                |
| 4           | Thunbergia laurifolia                                 | Powder      | SL                 | Thunbergia laurifolia                             |
| 5           | Phyllanthus emblica                                   | Powder      | PE                 | Phyllanthus emblica                               |
| 6           | Andrographis paniculata                               | Powder      | AP                 | Andrographis paniculata                           |
| 7           | Pueraria candollei ver. mirifica                       | Powder      | PC                 | Pueraria candollei                                |
| 8           | Senna alata                                           | Powder      | SA                 | Senna alata                                       |
| 9           | Boesenbergia rotunda                                  | Powder      | BR                 | Boesenbergia rotunda                              |
| 10          | Clinacanthus nutans                                   | Powder      | CN                 | Clinacanthus nutans                               |
| 11          | Curcuma longa                                          | Tablet      | CL                 | Curcuma longa                                     |
| 12          | Centella asiatica                                     | Capsule     | CA                 | Centella asiatica                                 |
| 13          | Pueraria candollei                                    | Capsule     | PC                 | Pueraria candollei                                |
| 14          | Centella asiatica                                     | Tablet      | CA                 | Centella asiatica                                 |
| 15          | Curcuma longa                                          | Capsule     | CL                 | Curcuma longa                                     |
| 16          | Kaempferia parviflora                                 | Capsule     | KP                 | Kaempferia parviflora                             |
| 17          | Centella asiatica                                     | Cream       | CA                 | Centella asiatica                                 |
| 18          | Curcuma longa                                          | Powder      | CL                 | Curcuma longa                                     |
| 19          | Phyllanthus emblica                                   | Powder      | PE                 | Phyllanthus emblica                               |
| 20          | Zingiber montanum                                     | Powder      | ZM                 | Zingiber montanum                                 |

*Indicates that the Barcode of Life Data System (BOLD) database was used for sample analysis.
Conclusion
Admixture or adulteration in herbal products is one of the main problems in herbal trade because the identification of herbal ingredients is challenging. Hence, there is an important requirement to develop a reference DNA barcode library or centralized digital database system that could serve as a regulatory database for ensuring the safety and quality of traded herbs. It is very important that the Thai FDA immediately begin to strictly enforce the development of pharmacopeial standards as well as revisions or modifications of existing regulatory guidelines to check or monitor the authenticity of raw materials or herbal products before they enter the herbal market. For quality assessment of herbal products, we strongly recommend incorporating DNA-based methods into the THP and TMM to maintain the safety, quality and efficacy of herbal medicines prior to them entering the market.

Materials and methods

Plant materials and herbal products. Multiple accessions of plant species mentioned in the THP and TMM were collected from several locations in Thailand (Table 1). The procedures for plant collection and field studies were conducted by following standard guidelines of Chulalongkorn University, Thailand. Those collections including samples from Thai FDA are permitted and legal. A total of 101 plant species and their voucher numbers were prepared as herbarium specimens and deposited at the Museum of Natural Medicine, Chulalongkorn University, Bangkok, Thailand. All plant species were identified by an independent expert taxonomist, Associate Professor Thatree Phadungcharoen of the Faculty of Pharmaceutical Sciences, Chulalongkorn University.

Figure 2. Maximum likelihood tree showing the phylogenetic relationships of reference Thai medicinal plants based on the Kimura-2-parameter (K2P) model using the rbcL region. The bootstrap support values were estimated with 1000 replicates. The respective family names are shown to the right.
DNA BARCODE OF SELECTED THAI MEDICINAL PLANTS
The Research Unit of The DNA Barcoding of Thai Medicinal Plants, Chulalongkorn University

Figure 3. Overview of the proposed digital reference DNA barcode database.

DNA isolation and PCR amplification. Genomic DNA was isolated from leaves using a DNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer’s protocol. Further PCR amplification was carried out in a 25 µL reaction volume that consisted of 1X PCR buffer, 1.5 mM MgCl₂, 0.2 mM dNTPs mix, 0.2 mM each forward and reverse primer, 0.5 U of Platinum Taq polymerase (Invitrogen, USA) and 30–40 ng of genomic DNA. Amplification was performed with an Eppendorf Master Cycler Gradient (Hamburg, Germany). PCR amplification with primers was carried out by using universal barcode regions⁴⁰. the ITS2 nuclear region (ITS2F-ATT CCC GGA CCA CGC CTG GCTGA⁴²; ITS4-TCC TCC  GCT TAT TGA TAT GC⁴³) and three chloroplast regions: matK (matK_xF-TAA TTT ACG ATC AAT TCA TTC⁴⁴; matK-MALPR1- ACA AGA AAG TCG AAG TAT⁴⁵), the trnH-psbA intergenic spacer (trnHf_05– CGC GCA TGG TGG ATT CAC AATCC⁴⁶; psbA3_f–GTT ATG CAT GAA CGT AAT GCTC⁴⁷) and rbcL (rbcLa-F-ATGTCACCACAAACAGACCTAAAGC⁴⁸; rbcLa-R-GTAAAATCAAGTCCACCGC⁴⁹) were used. PCR amplification of the ITS and psbA-trnH intergenic spacer regions was performed at 95 °C for 4 min followed by 30 cycles of 94 °C for 45 s, 58 °C for 45 s, and 72 °C for 90 s, with a final extension at 72 °C for 7 min. The amplification profiles for matK and rbcL consisted of 94 °C for 4 min followed by 30 cycles of initial denaturation at 94 °C for 60 s, 55 °C for 45 s, and 72 °C for 90 s, with a final extension step at 72 °C for 10 min. The obtained PCR amplicons were sequenced bidirectionally using their respective primers on an ABI3500 sequencer (Applied Biosystem, USA).

Genomic DNA of different dosage forms of the herbal product was extracted using a DNeasy Plant Mini Kit (Qiagen, Germany) and further purified using a GENECLEAN Kit (MP Biomedicals, France). The DNA isolation of herbal samples required multiple attempts to obtain good PCR amplification against the ITS2, matK, rbcL and psbA-trnH intergenic spacer regions. Subsequently, all those PCR products were sequenced as described above.

DNA sequencing and phylogenetic analysis. The sequences were edited using BioEdit software (version 5.0.6). BLAST analysis was conducted with the sequences as queries to determine the similarity of the nucleotide sequences in NCBI GenBank. The sequences with the maximum query coverage, highest homology, and maximum score were downloaded in FASTA format from the database and included in our analysis. The ML method was used to construct the relationships among plant samples with an appropriate model of nucleotide evolution. The final alignment file was imported into MEGA 7 to determine the character information prior to phylogenetic analysis using the Kimura 2-parameter molecular evolution model with 1,000 rapid bootstrapping replicates⁵⁰.
Data availability
The datasets generated during and/or analyzed during the current study are available in the NCBI GenBank repository, accession no: LC214981, LC214982, LC214984, LC214987, LC435390-LC435437, LC438866-LC456409, LC461717-LC461952, LC503563-LC503566, KP998768, KP998782, KP998796, KP998810.

Received: 13 December 2021; Accepted: 10 May 2022
Published online: 10 June 2022

References
1. ReportBuyer. Herbal Medicine Market Size, Forecast And Trend Analysis, 2014–2024. https://www.prnewswire.com/news-releases/the-global-herbal-medicine-market-is-expected-to-reach-usd-117-02-billion-by-2024--300726926.html. Accessed 10th Aug 2021
2. Chotchoungchatchai, S., Saralamp, P., Jenjitikul, T., Pornsiripongse, S. & Pranthanturarug, S. Medicinal plants used with Thai Traditional Medicine in modern healthcare services: A case study in Kaibhoeng hospital, Surin Province, Thailand. J. Ethnopharmacol. 141, 193–205 (2012).
3. Chompun, P. History of Thai Traditional Medicine (Matichon Printing, 2002).
4. Thai Traditional Medicine at Glance. Department for Development of Thai Traditional and Alternative Medicine (Ministry of Public Health, 2020).
5. Subcharoen, P. & Pechpray, D. International seminar on traditional medicine 10–13 (Queen Sirikit National Convention Center, 1995).
6. Thai Herbal Pharmacopoeia 2021, Bureau of Drug and Narcotic, Department of Medical Sciences, Ministry of Public Health, Thailand (2021).
7. Ichim, M. C. & de Boer, H. J. A review of authenticity and authentication of commercial ginseng herbal medicines and food supplements. Front. Pharmacol. 11, 2185 (2020).
8. Lo, Y. T. & Shaw, P. C. Application of next-generation sequencing for the identification of herbal products. Biotechnol. Adv. 37, 107450 (2019).
9. Ichim, M. C. The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. Front. Pharmacol. 10, 1227 (2019).
10. Sriram, R. et al. Species adulteration in the herbal trade: Causes, consequences and mitigation. Drug Saf. 40, 651–661 (2017).
11. Thongkhoa, K. et al. Differentiation of Cynanchum cineus, a smoking cessation herb, from its adulterant Emilia sonchifolia using macroscopic and microscopic examination, HPTLC profiles and DNA barcodes. Sci. Rep. 10, 1–11 (2020).
12. Urumaradappa, S. K. J., Tungphathong, C. & Sukrong, S. Mitigating the impact of admixtures in Thai herbal products. Front. Pharmacol. 10, 1205 (2019).
13. Wang, M. et al. An integrated approach utilizing chemometrics and GC/MS for classification of chamomile flowers, essential oils and commercial products. Food Chem. 152, 391–398 (2014).
14. Grazina, L., Amaral, J. S. & Mafra, I. Botanical origin authentication of dietary supplements by DNA-based approaches. Compr. Rev. Food Sci. Food Saf. 19, 1080–1109 (2020).
15. Raclariu-Manolica, A. C. et al. DNA metabarcoding for quality control of basil, Oregano, and Paprika. Front. Plant Sci. 12, 2 (2021).
16. Tungphathong, C., Urumaradappa, S. K. J., Awachai, S., Sooksawate, T. & Sukrong, S. Differentiation of Mitragyna speciosa, a narcotic plant, from allied Mitragyna species using DNA-barcoding-high-resolution melting (Bar-HRM) analysis. Sci. Rep. 11, 1–9 (2021).
17. Booker, A., Johnston, D. & Heinrich, M. Value chains of herbal medicines-research needs and key challenges in the context of ethnopharmacology. J. Ethnopharmacol. 140, 624–633 (2012).
18. Seethapathy, G. S., Raclariu-Manolica, A. C., Anmarkrud, J. A., Wangensteen, H. & de Boer, H. J. DNA metabarcoding authentication of ayurvedic herbal products on the European market raises Concerns of quality and fidelity. Front. Plant Sci. 10, 68 (2019).
19. Urumaradappa, S. K. J., Tungphathong, C., Prombutara, P. & Sukrong, S. DNA metabarcoding to unravel plant species composition in selected herbal medicines on the National List of Essential Medicines (NLEM) of Thailand. Sci. Rep. 10, 1–11 (2020).
20. European Pharmacopoeia. 8th Edition. Council of Europe: Strasbourg.
21. Dhanya, K. & Sasikumar, B. Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. J. Ethnopharmacol. 141, 114517 (2021).
22. Mishra, P. et al. DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol. J. 14, 8–21 (2016).
23. Novoa, E. M., Mouindipa, P. F. & Niedermyer, T. H. Codiaeum variegatum (L.) Rumph. ex A. Juss. (Euphorbiacea): An overview of its botanical diversity, traditional uses, phytochemistry, pharmacological effects and perspectives toward developing its plant-based products. J. Ethnopharmacol. 2, 114244 (2021).
24. Lee, H.-J. & Lee, J.-H. Effects of medicinal herb tea on the smoking cessation and reducing smoking withdrawal symptoms. J. Chin. Med. 33, 127–138 (2005).
25. Prasopthum, A., Pouyfung, P., Saraput, S., Srisook, E. & Rongnoparut, P. Inhibition effects of Vernonia cinerea (L.) Rumph. ex A. Juss. (Euphorbiaceae): An overview. Evid. Based Complement. Alternat. Med. 20, 2 (2021).
26. Dhanya, K. & Sasikumar, B. Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. Curr. Trends Biotechnol. Pharm. 4, 454–489 (2010).
27. Parvathy, V. et al. DNA barcoding to detect chilli adulteration in traded black pepper powder. Food Biotechnol. 28, 25–40 (2014).
Acknowledgements
This research was supported by the Ratchadaphisek Somphot Fund for Postdoctoral Fellowship, Chulalongkorn University. We acknowledge the Department of Thai Traditional and Alternative Medicine, Ministry of Public Health, Thailand, and Chulalongkorn University for providing samples and laboratory facilities. This research was supported and partially financed by the Thai FDA under the project “Developing a medicinal plant identity verification tool to support permission” (Project No. RES_60_388_33_007). Additionally, we thank Associate Professor Thatree Phadungcharoen at the Faculty of Pharmaceutical Sciences, Chulalongkorn University, Thailand, for the taxonomic identification of the plant samples.

Author contributions
S.K.U., C.T., J.J., N.P., D.P., S.V., S.S. Conceived and designed the experiments: S.S.; analyzed data: S.K.U., J.J., and C.T.; carried out molecular analysis: J.J., N.P., D.P., S.K.U., and S.V.; wrote the manuscript: S.K.U.; reviewed and edited the manuscript: S.K.U., C.T., N.P., and S.S. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-13287-x.

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022