Food Control 183 (2026) 111893

Contents lists available at ScienceDirect

Food Control

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

Comparison of DNA barcoding and metabarcoding approaches for the authentication of herbal infusions in the Portuguese market

M. Alice Pinto [a], Joana Santos [a], Andreia Quaresma [a], Mónica Honrado [a,b], José Rufino [c], Joana S. Amaral [a,*]

[a] CIMO, La SusTEC, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253, Bragança, Portugal
[b] LAQV-REQUIMTE, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal
[c] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança, Portugal

ARTICLE INFO

Keywords:
Authenticity
Infusions
matK
rbcL
ITS2 metabarcoding

ABSTRACT

The authentication of commercial herbal products such as herbal teas and infusions is essential to ensure consumer safety and product integrity. In this study, 100 commercial herbal infusions, including 94 single-species products and six polyherbal formulations, were analysed using DNA barcoding (*matK* and *rbcL*) and ITS2 metabarcoding. Moreover, the performance of two custom reference databases (NCBI and BOLD) was compared in ITS2 metabarcoding. DNA extraction was successful for 94 samples, while six single-species products failed to amplify any of the tested barcodes. Among the 88 remaining single-species samples, ITS2 showed the highest amplification success (100 %), outperforming the barcodes *rbcL* (94 %) and *matK* (84 %). Sanger sequencing confirmed the labelled species in 69.3 % of cases with *rbcL* and 48.9 % with *matK*. While 63 samples would be considered authentic solely based on barcoding (i.e., if either *rbcL* or *matK* matched the label), ITS2 metabarcoding revealed that many of these contained additional undeclared species, indicating that barcoding alone overestimated product authenticity. Of the 85 samples successfully analysed by ITS2 metabarcoding, only 27 (32 %) fully matched their label, while 58 (68 %) contained either additional undeclared species or complete substitutions. Several products contained undeclared species in significant proportions, indicating potential economic adulteration. Analysis of mock-mixtures demonstrated that ITS2 metabarcoding can approximate ingredient proportions, though recovery bias may occur. Overall, the results revealed (i) the importance of curated and comprehensive databases, with a higher number of species being identified by NCBI database, (ii) the superior sensitivity of ITS2 metabarcoding, and (iii) the widespread mislabelling in commercial herbal products.

1. Introduction

Medicinal plants have played a fundamental role in human health since ancient times, being widely used in folk medicine and traditional medical systems such as Chinese medicine and Ayurveda (Heinrich & Anagnostou, 2017). Currently, medicinal plants remain essential in healthcare, particularly in regions with limited access to conventional medical care. At the same time, their use is growing in developed countries mainly due to increased consumer interest in natural remedies for health promotion, disease prevention, and as complementary treatments alongside conventional medicine (Ibrahim et al., 2024). Therefore, the global herbal medicine market has experienced a considerable expansion in recent years, with an estimated value of approximately USD 70.57 billion in 2023 and projected to grow by circa 20 % from 2024 to 2030 (Grand View Research, 2024).

Herbal products can be considered medicine or food, with regulatory frameworks varying significantly by region. In the European Union (EU), medicinal plants can be used in herbal products classified either as Traditional Herbal Medicinal Products (THMPs), which are regulated under Directive 2004/24/EC (amending Directive 2001/83/EC), or as food products. This latter category includes herbal infusions, teas, and food supplements, all regulated under food legislation, specifically by Regulations EC 178/2002 and EC 1924/2006, and Directive 2002/46/EC. THMPs must adhere to strict standards regarding pharmaceutical quality, safety, and efficacy, whereas herbal infusions marketed as foods are prohibited from making explicit therapeutic claims. Therefore,

* Corresponding author.
E-mail address: jamaral@ipb.pt (J.S. Amaral).

https://doi.org/10.1016/j.foodcont.2025.111893
Received 17 July 2025; Received in revised form 1 October 2025; Accepted 3 December 2025
Available online 5 December 2025

products containing the same medicinal plant may be sold either as medicine or food, depending on their intended use and the claims made on their labels, which may create ambiguity. This sometimes-unclear boundary can complicate regulatory supervision and quality assurance. In THMP, quality control involves detailed testing of raw materials and/or final products for identity, strength, purity, and contaminants using validated methods. On the other hand, when the same medicinal plant is traded as food (such as herbal infusions or food supplements), the requirements are generally less strict, as quality assurance and quality control mainly focus on food hygiene and safety in accordance with food law.

The rapid expansion of the herbal products market and its growing global value have raised concerns about fraud, mainly due to botanical substitution (Gafner et al., 2023; Grazina et al., 2020a; Ichim, 2019). The lower level of regulatory control over herbal products sold as foods makes them especially vulnerable to botanical adulteration, encompassing the deliberate substitution of high-value plants with lower-cost ones or the addition of undeclared fillers. Mislabelling can also occur when morphologically similar plants are inadvertently confused. In either case, the expected bioactive compounds may be absent or present in lower amounts, reducing the product's effectiveness. At the same time, compounds from undeclared or substituted plant species may be present, posing potential health risks to consumers. Therefore, reliable methods are needed to authenticate herbal products.

Several approaches have been proposed for the authentication of herbal products, including microscopic examination, chemical analysis, and DNA-based methods, the latter being the most effective for accurate species identification (Grazina et al., 2020a). Among the DNA-based methods, DNA barcoding, which relies on the amplification of short, standardised DNA regions for species identification, has been broadly used for the authentication of medicinal plants (Chen et al., 2023; Srivastava et al., 2016; Yu et al., 2021). While the cytochrome *c* oxidase I (COI) gene is the barcode of choice for animals, identifying a universal barcode for plants has proven more challenging. The chloroplast markers ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) and maturase K (*matK*) were initially proposed by the Consortium for the Barcode of Life (CBOL) Plant Working Group (Hollingsworth et al., 2009) as DNA barcodes for land plants due to their complementary characteristics. While *rbcL* exhibits high universality across a broad taxonomic range, it provides low resolution at the species level. In contrast, *matK* offers greater discriminatory power, although its amplification is more prone to failure (Hollingsworth et al., 2009; Letsiou et al., 2024). Therefore, other markers have been proposed, among which the internal transcribed spacer 2 (ITS2) stands out. ITS2 offers high discriminatory power across a wide range of plant taxa (Chen et al., 2010; Quaresma et al., 2021), carries conserved flanking regions that facilitate the design of universal primers, and its relatively short sequence is suitable for different Next-Generation Sequencing (NGS) platforms, enabling high-throughput analysis (Yao et al., 2010). Moreover, unlike traditional DNA barcoding, which relies on Sanger sequencing, ITS2 is well-suited for metabarcoding applications, allowing the identification of multiple species within complex herbal mixtures (Travadi et al., 2023). This makes it appropriate for quality control and surveillance of herbal products.

Although DNA barcoding is widely used for plant authentication and metabarcoding has been successfully employed to detect botanical adulteration in herbal medicinal products (Ivanova et al., 2016; Mück et al., 2024; Pandit et al., 2021; Travadi et al., 2023), few studies to date have directly compared the two approaches on the same medicinal plant samples, and those available are based on limited sample sizes. This study addresses this gap by systematically comparing the effectiveness of DNA barcoding and metabarcoding on a larger sample set, with a particular focus on medicinal plants traditionally employed to support brainhealth (including cognitive function, mood, and sleep). To this end, DNA barcoding was performed using the *rbcL* and *matK* markers, while metabarcoding targeted the ITS2 region. For identification of ITS2 amplicons, a stepwise approach was implemented based on custom-built reference databases, comprising sequences of (i) medicinal plants traditionally used for brain health, (ii) medicinal plants in general, and (iii) land plants. Additionally, separate databases were built using sequences retrieved from GenBank (National Center for Biotechnology Information, NCBI) and the Barcode of Life Data Systems (BOLD), and the classifications from each source were compared. Finally, a survey of the Portuguese market was conducted to assess labelling compliance.

2. Materials and methods

2.1. Samples

A total of 100 commercial samples of medicinal plants, sold in plastic or paper bags (50 g–100 g) containing fragmented herbal material for infusions, were purchased online from multiple suppliers in Portugal (Ervanário Portuense, Chás do Mundo, Ervas e Mezinhas, Antiga ervanária, and Ervanária Fragária). Of these, 94 were labelled as containing a single plant species (hereafter referred to as monoherbal), while the remaining six were labelled as mixtures of multiple species (hereafter referred to as polyherbal). All labels included the scientific name of the declared plant species. The samples were selected to contain plant species traditionally used for their activity on the brain, such as improving sleep, mood or cognition, or described as an adaptogen. Twelve additional samples with known origin were obtained from Botanical Gardens or Herbaria, including Jardim Botânico d'Ajuda (Lisbon), Universidade Tecnológica Federal do Paraná (Brazil), and Herbário da Universidade de Coimbra (Portugal). Finally, three mock mixtures containing known mass proportions of different species were prepared to test the accuracy of the pipeline. For each mixture, the medicinally relevant parts of each plant species were individually ground into powder and weighed as follows: (i) equal masses (25 % each) of *Crataegus monogyna*, *Matricaria chamomilla*, *Rosmarinus officinalis,* and *Verbena* officinalis; (ii) 30 % *Centella asiatica*, 30 % *Foeniculum vulgare*, and 40 % *Ilex paraguariensis*; (iii) 30 % *Rosmarinus officinalis*, 30 % *Valeriana officinalis,* and *40 % Angelica archangelica*.

2.2. Sample preparation and DNA extraction

For each commercial sample, the whole leaves, roots or stem pieces were ground in a blender until a fine powder was obtained. Before using the blender, the container and blade were thoroughly washed and treated with a DNA decontaminator solution to avoid cross-contamination. To minimise the risk of airborne pollen contamination, all DNA extractions were performed in a laminar airflow biosafety cabinet. Approximately 100 mg of the obtained powder was transferred to a 2.0 mL screwcap tube containing two 3 mm zirconia beads, and the DNA was extracted using the Macherey-Nagel NucleoSpin Plant II Kit (Düren, Germany). Briefly, after adding 400 μL of PL1 lysis buffer, the samples were ground in a Precellys ® Evolution (Bertin Instruments) set up for three cycles at 6200 rpm for 5 s. Subsequently, 10 μL of RNase A was added to the 2.0 mL tubes, which were incubated at 65 °C for 10 min. The remainder of the protocol was carried out according to the manufacturer's instructions. The quality and yield of the extracts were assessed by UV spectrophotometry using a SPECTROstar® Nano microplate reader (BMG Labtech, Offenburg, Germany) with an LVis plate accessory. To estimate DNA content and purity, the absorbance was measured at 260, 280 and 230 nm using the Multi-user Reader Control and MARS Data Analysis Software (LVis) (BMG Labtech, Offenburg, Germany). The obtained extracts were kept at −20 °C until further analysis and diluted to 10 ng/μL before PCR.

2.3. MatK and rbcL amplification and sequencing

DNA extracts obtained from the samples labelled as containing a single species were amplified targeting the *matK* region using the

primers matK-KIM1R: 5′-ACCCAGTCCATCTGGAAATCTTGGTTC-3′ and matK-KIM3F: 5′-CGTACAGTACTTTTGTGTTTACGAG-3′ and targeting the *rbcL* region using the primers rbcLa-F: 5′-ATGTCACCACAAACAGAGACTAAAGC-3′ and rbcLa-R: 5′-GTAAAATCAAGTCCACCRCG-3′ (Fazekas et al., 2012; Hollingsworth et al., 2009). PCR was carried out in a 25 μL total volume using 12.5 μL of Q5 High-Fidelity 2X Master Mix (New England Biolabs), 1.25 μL of each primer (10 μM), 9 μL of nuclease-free molecular biology grade water, and 1 μL of DNA extract (10 ng/μL). For both markers, reactions were performed in a T100 Thermal Cycler (BioRad™) using the temperature profile consisting of an initial denaturation at 98 °C for 3 min, followed by 40 cycles of amplification at 98 °C for 10 s, 55 °C for 30 s, and 72 °C for 40 s, and a final extension at 72 °C for 2 min. The amplicons were sent to STABVIDA Inc. (Portugal) for purification and Sanger sequencing in both directions and then analysed using BioEdit v7.2.5.0. The sequences were then identified using the Basic Local Alignment Search Tool (BLAST) (July 2024).

2.4. ITS2 metabarcoding

2.4.1. ITS2 amplification and sequencing

DNA metabarcoding was performed using the primers ITS-S2F (Chen et al., 2010) and ITS-S4R (White et al., 1990), as described by Quaresma et al. (2021). The primers were modified to contain the overhang adapters (in bold) proposed by Illumina (Illumina, 2025), namely ITS-S2F-adp1 5′-**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**ATGCGATA CTTGGTGTGAAT-3′ and ITS-S4R-adp1 5′-**GTCTCGTGGGCTCGGAGATGTGTAT**.

AAGAGACAGTCCTCCGCTTATTGATATGC-3′. PCR was carried out in triplicate for each sample in a two-stage PCR process. The first-stage PCR was performed on a T100 Thermal Cycler (BioRadTM) using a total volume of 10 μL, containing 5 μL of Q5 High-Fidelity 2X Master Mix (New England Biolabs), 0.5 μL of each primer at 10 μM, and 1 μL of DNA extract at 10 ng/μL. Thermal cycling conditions consisted of 98 °C for 3 min, followed by 35 cycles at 98 °C for 10 s, 52 °C for 30 s, and 72 °C for 40 s, followed by a final extension at 72 °C for 2 min.

The amplicons were sent to the Centre for Molecular Analyses of the Research Centre in Biodiversity and Genetic Resources (CIBIO, Vairão, Portugal), where a purification step of the amplicons was performed using 0.8 × reversible immobilisation paramagnetic beads (Agencourt AMPure XP) per microlitre of PCR product. The second-stage PCR was then performed to incorporate the custom-made unique indexes developed by CIBIO based on previous literature (Gansauge et al., 2020; Kircher et al., 2012). The reaction consisted of a 10-μL total volume containing 0.5 μL of each oligonucleotide at 1 μM, 5 μL of KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems), and 2 μL of a 1:10 dilution of the purified amplicons. Thermal cycling conditions were 95 °C for 3 min, followed by 10 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 5 min. As before, a purification step of the indexed amplicons was performed using the paramagnetic beads. The purified indexed products were quantified using an Epoch Microplate Spectrophotometer (BioTek Instruments), normalised to a final concentration of 10 nM, and then pooled. For each pool, the amplicon size distribution was determined on a TapeStation 2200 using the HS D1000 kit (Agilent Technologies) and quantified by a SYBR green quantitative PCR assay using the KAPA Library Quantification kit (Kapa Biosystems). The pools were combined equimolarly into a single sequencing library containing all samples. The sequencing library was diluted to 2 nM, spiked with 10 % Illumina-generated PhiX control library, and then sequenced on the Illumina MiSeq using the 2 × 250 cycles v2 nano chemistry (MiSeq reagent Nano Kit v2), according to the manufacturer's instructions. The samples were analysed in two distinct runs, comprising 56 and 53 samples.

2.4.2. Construction of ITS2 databases

The sequence reads were classified using a stepwise approach (see 2.4.3). First, the reads were aligned against a custom database containing only ITS2 sequences of plant species traditionally used for brain health support. If identification was unsuccessful, the reads were then aligned against a broader custom database of medicinal plants and finally against a global database comprising 307,977 sequences, representing 534 families, 11,034 genera, and 111,382 species of vascular plants (Quaresma et al., 2024).

To create the medicinal plants custom database, a list of scientific names was compiled from multiple sources, including the 8th Edition of the European Pharmacopoeia (Ph. Eur. 8), herbal monographs from the European Medicines Agency (EMA), the Indian Medicinal Plants, Phytochemistry and Therapeutics 2.0 (IMPPAT 2.0, https://cb.imsc.res.in/imppat/home) and the Diseases Plants Eliminate (DISPEL, https://compbio.iitr.ac.in/dispel/). The collected names were consolidated into a single file, and a script was run to remove duplicates. To prepare a second list of plant species for the custom database focused on brain health, information was retrieved using the "Explore" option in the DISPEL database by searching for diseases with the following keywords: "brain disorders", "brain injuries", "sleep disorders", "anxiety disorders", "depression", "memory disorders", "mental disorders", "Alzheimer disease", and "Parkinson disease".

The scientific names included in these two lists were used to construct four ITS2 databases. Two were constructed using sequences obtained from BOLD Systems: one focused on brain health plants (BaseDeDadosHeadBOLD) and the other on medicinal plants in general (BaseDeDadosGeralBOLD). The other two were constructed using sequences retrieved from NCBI for the same plant species: one focused on brain health plants (BaseDeDadosHeadNCBI) and the other on medicinal plants in general (BaseDeDadosGeralNCBI). Finally, after the custom databases were created, two additional refinement steps were undertaken: (i) ITS2 sequences obtained via NGS from authentic samples were incorporated into each database; (ii) manual curation was performed to remove bacterial or fungal sequences that had been misidentified as plant species. The final custom BOLD and NCBI databases are publicly available as FASTA files on Zenodo (https://doi.org/10.5281/zenodo.15794189).

2.4.3. Bioinformatics

Pools were de-multiplexed in BaseSpace Sequence Hub based on their unique indexes incorporated in the second-stage PCR. Raw sequence reads (Fastq files) were processed using VSEARCH v2.14.1 (Rognes et al., 2016). Reads R1 and R2 were merged using the fastq_mergepairs. Merged reads were filtered (length <200 bp and >500 bp, ambiguous base pairs), dereplicated, denoised and chimaeras were removed. Amplicon sequence variants (ASVs) were generated and then classified at the species level against the custom NCBI and BOLD databases separately, using VSEARCH v2.14.1 with a sequence similarity threshold of 97 %. Following the previously described stepwise approach, reads that could not be classified using either the brain-health or general medicinal plant databases were subsequently analysed using the broader land plant database (Quaresma et al., 2024). These unclassified reads were then subjected to hierarchical classification beginning at the family level with SINTAX (Edgar, 2016), applying a 97 % similarity threshold.

A community matrix format table, with columns representing samples and rows the identified species, and a file containing the taxonomic lineage for identified species were created and imported into R-Studio v4.5.3 (Team, 2018). The package Phyloseq v1.46.0 (McMurdie & Holmes, 2013) was then used to merge the files and normalise the data. To account for the possibility of naturally occurring pollen in the raw plant material, any taxa with a relative abundance below 1 % were excluded from further analysis.

3. Results

3.1. DNA isolation and PCR amplification

DNA was successfully extracted from 94 of the 100 samples, including 88 labelled as single species (monoherbal) and six as mixtures (polyherbal). Extract concentrations ranged from 2.8 to 221.2 ng/μl, with purity (A260/A280 ratio) mostly lying between 1.6 and 2.0. An overview of the PCR amplification success is shown in Fig. 1. All 88 single-species samples were successfully amplified with the ITS2 marker (100 %), while *rbcL* and *matK* amplicons were obtained for only 83 (94 %) and 74 (84 %) samples, respectively. A total of 71 samples were successfully amplified for both *matK* and *rbcL*. Six samples failed to yield positive amplification for any of the three tested markers.

3.2. DNA barcoding using matK and rbcL markers

Fig. 2 and Table S1 (supplementary material) show the results obtained for DNA barcoding using *matK* and *rbcL*. Of the 83 *rbcL* and 74 *matK* sequences generated by Sanger sequencing, 60 (72.3 %) and 49 (66.2 %), respectively, yielded high-quality reads with electropherograms showing clear, unambiguous base calls. Double peaks were observed in at least one sequencing direction in 15 *rbcL* and 11 *matK* amplicons; however, reliable consensus sequences were still obtained through careful manual editing. For the remaining eight *rbcL* and 14 *matK* amplicons, the sequences contained excessive overlapping peaks, preventing the generation of a reliable consensus. In these cases, both the raw forward and reverse sequences were analysed separately for identification purposes.

When analysing *rbcL* and *matK* separatelly, of the 88 monoherbal samples, 61 (69.3 %) and 43 (48.9 %) matched the label, 21 (23.8 %) and 29 (33.0 %) differed from the label, and 6 and 16 were inconclusive—either due to amplification failure (5.7 % and 15.8 %) or the absence of reference sequences in GenBank (1.1 % and 2.3 %), respectively (Fig. 2). Overall, *rbcL* outperformed *matK*, demonstrating both a higher authentication success rate and superior amplification efficiency. The discriminatory power of the markers and the coverage of the reference databases may account for this finding. Although *rbcL* is less variable than *matK* and therefore potentially less effective in resolving closely related species, it may be more tolerant of degraded DNA, which is often obtained from commercial products. In contrast, while *matK* offers higher taxonomic resolution, it suffers from poor primer universality and a longer amplicon size, frequently resulting in amplification failure with degraded or processed samples (Molina et al., 2018; A. C. Raclariu, Heinrich, Ichim, & de Boer, 2018). Moreover, due to its higher amplification success, *rbcL* tends to have a broader representation in reference databases like GenBank, as compared to *matK* (Hollingsworth et al., 2009).

When combining *matK* and *rbcL*, 47 monoherbal samples showed consistent taxonomic identification between the two markers (Fig. 2 and Table S1). Of these, 40 samples matched the label, while seven consistently corresponded to a different species. Ten additional samples were considered mislabelled according to both markers, while 14 yielded conflicting results: two were correctly identified by *matK* but not by *rbcL*, while 12 were correctly identified by *rbcL* but not by *matK* (Table S1). In several cases, the sequences that did not match the labelled species were of low to moderate quality, displaying multiple overlapping peaks in the electropherogram, suggesting that more than one species was present in the commercial product. This finding was confirmed by ITS2 metabarcoding (see section 3.3.2). For example, sample P51 was identified by *rbcL* as *Pistacia lentiscus*, in accordance with its label, while *matK* assigned it to *Carica papaya*. Subsequent ITS2 metabarcoding confirmed the coexistence of both species in the sample.

3.3. DNA metabarcoding using the ITS2 marker

3.3.1. Construction of ITS2 reference databases

As detailed in 2.4.2, sequence data were mined from NCBI and BOLD Systems to construct the first two hierarchical levels of the custom databases. The first level contained sequences of plant species traditionally used for brain health support: the BOLD database contains 2900 sequences corresponding to 593 species, nine of which are absent from the NCBI database; the NCBI database comprises 28,174 sequences representing 920 species, including 336 present in the NCBI database but not in the BOLD database. The second level contained sequences of general medicinal plant species: the BOLD database included 6628 sequences representing 1719 species, with 15 unique to this database and absent from the NCBI counterpart; the NCBI database contained 64,812 sequences from 3155 medicinal plants, of which 1451 species are unique to the NCBI database. To further enrich the two-level databases, voucher

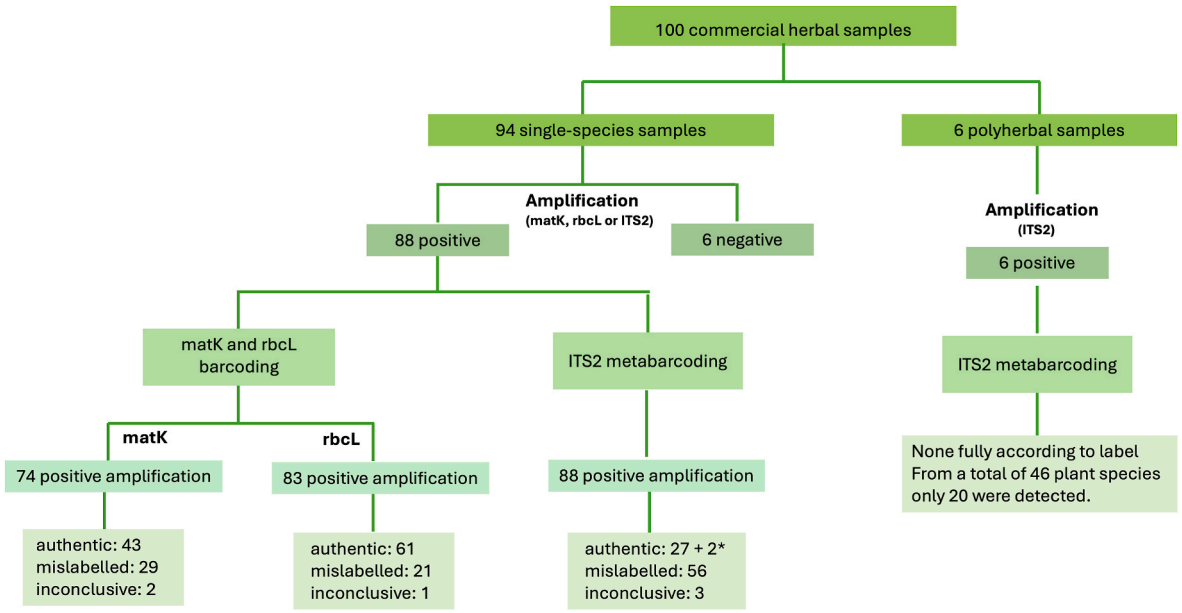

Fig. 1. Overview of the results obtained using matK and rbcL DNA barcoding and ITS2 metabarcoding (*: samples considered authentic according to Pharmacopoeias criteria).

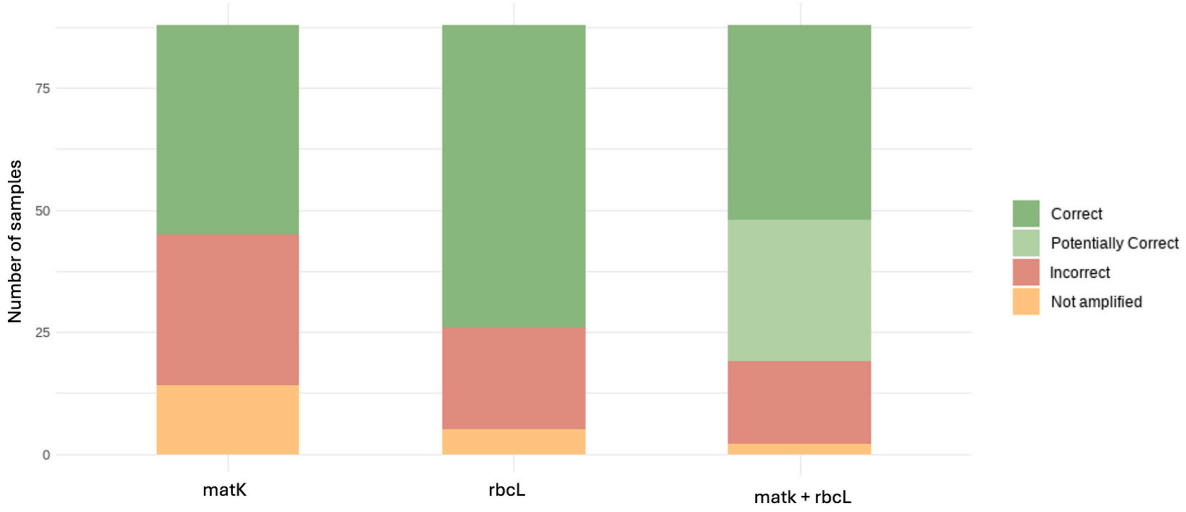

Fig. 2. Classification of 88 monoherbal samples by DNA barcoding using matK and rbcL markers.

samples were requested from botanical gardens and herbaria. Of the 12 provided, only five were successfully amplified for ITS2, including *Ginkgo biloba*, *Pfaffia glomerata*, *Vinca minor*, *Crataegus monogyna*, and *Cola nitida*. The corresponding sequences were added to both BOLD and NCBI custom databases. The third level of the custom databases consisted of a global curated database containing 111,382 species, represented by 307,977 sequences, previously constructed by Quaresma et al. (2024).

The ASVs obtained from the ITS2 high-throughput sequencing of the 88 monoherbal samples were initially queried separately against the NCBI and BOLD custom databases, using the stepwise approach in both cases. When the ASVs could not be assigned to the species level in the first two-level databases, identification in the third-level database was implemented at the genus level to allow performance comparison between the custom NCBI and BOLD databases.

Upon analysing the classification results generated by the bioinformatic pipeline using the two-level custom NCBI databases, several unexpected species were repeatedly identified across independent samples, i.e. from different suppliers and labelled species. These species included *Astragalus membranaceus, Prunus persica, Senna alexandrina*, and *Huperzia serrata* (Table S1). To verify whether these were spurious assignments, a closer inspection was undertaken by manually analysing the corresponding ASVs and performing BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST). This analysis revealed that 281 sequences corresponded to fungi but were misidentified as plant species in NCBI. As a result, these misidentified sequences were removed from the custom NCBI database during the curation process. It is important to note that although the primers used for ITS2 sequencing (ITS-S2F/ITS4) were designed for plant identification purposes, due to their universal target regions within eukaryotes, these can also amplify fungal DNA, which is often found infecting plant tissues. This explains the presence of fungal sequences deposited in NCBI identified as plant entries (Quaresma et al., 2024).

3.3.2. *ITS2 metabarcoding of mock mixtures*

To test the ITS2 metabarcoding quantitative capacity, three plant mock mixtures were prepared containing different proportions of randomly selected plant species. The percentage of each species in the mixture was calculated based on the total number of reads and compared to the corresponding proportions in the initial mass used to

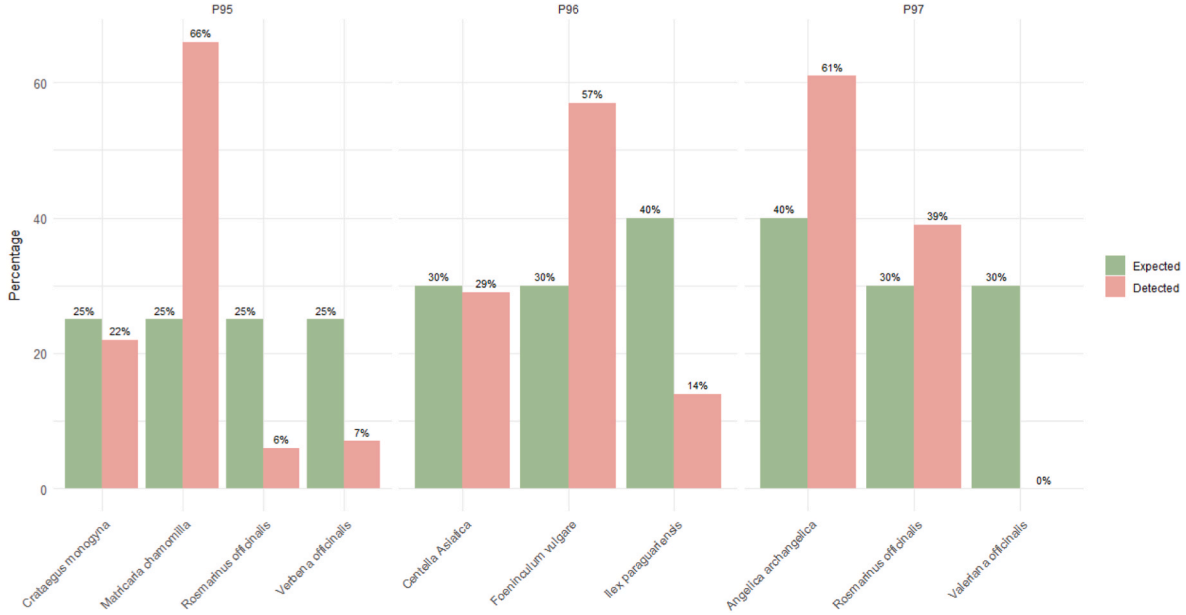

Fig. 3. Comparison of expected (% plant mass) and detected (% of total reads) species.

prepare the mock mixture (Fig. 3). In each sample, one of the species presented a close match between the percentage of reads and its proportion in the mock mixture, namely *Crataegus monogyna* in P95, *Centella asiatica* in P96, and *Rosmarinus officinalis* in P97. However, the other species were either over- or underrepresented. For example, in P95, *Matricaria chamomilla* was overrepresented (66 % instead of 25 %), and both *R. officinalis* and *Verbena officinalis* were underrepresented (each ~6 % instead of 25 %). *R. officinalis* was also an ingredient of P97, but in this case, the detected percentage (39 %) was similar to the expected one (30 %). Moreover, in this sample, ITS2 metabarcoding failed to detect *Valeriana officinalis* in the mixture, even though this species was included in the databases and had been successfully identified in single-species samples. This suggests that the failure to identify the species in the mock mixtures was not due to its absence in the reference databases but instead to potential DNA extraction and/or PCR bias, which were not investigated in this study.

3.3.3. ITS2 metabarcoding of monoherbal samples

Of the 94 monoherbal samples, only 88 were sequenced since 6 samples failed to amplify, namely those labelled as *Leonurus cardiaca*, *Persicaria bistorta*, *Erythrina* sp., *Papaver rhoeas* samples, and two *Angelica archangelica* samples. High-throughput sequencing of the 88 samples produced 2,219,198 and 2,015,430 raw reads in the first and second runs, respectively. Species classification was conducted using the described bioinformatic pipeline with the two custom databases independently: NCBI and BOLD. The results for each sample are detailed in Table S2 (supplementary material), including the identified species, the number of reads for the labelled species, and the total number of reads assigned at the species, genus, and family levels. Percentages presented in Table S2 were calculated based on the number of reads at the species level. Of the 88 samples, three labelled as *Moringa oleifera* had very few total reads (<40) and were therefore excluded from further analyses. The remaining 85 samples had total read counts ranging from 1373 (P44) to 21,117 (P18).

3.3.3.1. Classification using the custom NCBI databases. Using the pipeline with the custom NCBI databases to classify the sequence reads, 27 of the 85 (31.8 %) monoherbal samples were classified in agreement with the labelled species, confirming their authenticity (Tables 1 and S2). The remaining 58 were mislabelled. Of these, 15 did not contain the declared species and were therefore classified as adulterated. Although the absence of the labelled species could be explained by methodological limitations, such as preferential primer annealing to other plant sequences, the 15 samples were still considered mislabelled, given the presence of DNA from other species or genera (ranging from 1 to 11 taxa). The other 43 samples were also classified as adulterated, as their composition did not fully match the label declarations despite the presence of the labelled species. In 38 of these, the labelled species was the most abundant taxon identified (i.e., with the highest number of reads). However, while 11 samples (P02, P22, P26, P29, P31, P40, P41, P47, P50, P52 and P70) contained only one additional species beyond the labelled one, the remaining 27 contained between two and nine additional species. In five samples (P03, P10, P15, P20 and P25), the labelled species was detected but only in minor proportions.

For most monoherbal samples, the number of reads assigned to the species, genus, and family levels was nearly identical, with differences lower than 1 %. However, this was not the case in 11 samples (P21, P32, P48, P58, P64, P66, P69, P77, P78, P80, and P85; Table S2), which unexpectedly contained sequences assigned to the genus *Clusia*. This result was further investigated using BLAST, which revealed these sequences as additional misidentified and corresponding to fungi, most of which were *Aspergillus* spp. Moreover, two other samples (P12, P22) showed a higher number of reads assigned to the family level as compared to the species level, and these were identified via BLAST as *Penicillium* spp. The detection of these fungal genera may raise potential concerns regarding product quality and public health, as they include species known to produce mycotoxins. Nevertheless, chromatographic methods would be required to confirm the presence of toxins. Harvested plants may naturally contain fungal contaminants, such as spores or even mycelium, which may proliferate during plant drying or storage if conditions are not adequately controlled. Such contamination has been reported in previous studies on medicinal plants (Ivanova et al., 2016; Rizzo et al., 2004). For instance, Ivanova et al. (2016) used Sanger and NGS for taxonomic authentication of 15 herbal supplements and detected a diverse fungal community commonly associated with plants. Similarly, Rizzo et al. (2004) reported fungal contamination in 152 dried medicinal and aromatic herbs used as raw material for drugs, concluding that 52 % of the samples were contaminated with species of the *Aspergillus* genus while 16 % were contaminated with *Fusarium* spp.

3.3.3.2. Classification using the custom BOLD databases. Using the pipeline with the custom BOLD databases to classify the sequence reads, 28 of the 85 (32.9 %) monoherbal samples were authenticated as they contained only the labelled species (Table S2). However, two of these samples, labelled as *Crataegus monogyna* (P18) and *Hypericum perfuratum* (P22), also included reads assigned only at the family and genus levels, respectively. In the case of sample P18, these additional reads matched the genus *Crataegus*, consistent with the labelled species. In contrast, P22 showed a substantial number of reads assigned to a different family (Convolvulaceae). A BLAST search identified these reads as belonging to *Calystegia sepium*, which agrees with the classification using the custom NCBI databases. Therefore, the final number of authenticated samples was 27 (31.8 %), as sample P22 was considered adulterated due to the presence of undeclared plant material.

The remaining 57 samples were classified as non-conforming with the label, indicating potential cases of adulteration. Of these, 28 showed no detectable traces of the labelled species and instead revealed the presence of undeclared taxa. The remaining did contain the labelled plant species but were still considered adulterated due to the presence of undeclared species.

Moreover, a notable discrepancy between the number of reads assigned at species and genus or family levels was observed in 33 samples. While in 7 samples the reads classified by the global database for vascular plants were attributed to *Clusia* genus and later confirmed to be fungi, in the majority of the cases (26 samples), this discrepancy was attributed to the absence of reference sequences in the custom BOLD databases for specific species of several genera, including *Aristolochia*, *Calystegia*, *Carduus*, *Cyperus*, *Dinebra*, *Grevilea*, *Hesperocyparis*, *Lavandula*, *Marsdenia*, *Mikania*, *Monteverdia*, *Paulina*, *Poa* and *Senna*. As a result, the corresponding sequences could only be classified at the genus level.

3.3.4. ITS2 metabarcoding of polyherbal samples

Finally, the six polyherbal samples contained a total of 46 plant species according to their labelsBoth the NCBI and the custom BOLD databases yielded comparable results since both enabled the detection of only 20 of these species (43.5 %) (Table S2).

4. Discussion

4.1. Single-species commercial products: classification using NCBI versus BOLD custom databases

The accurate classification of herbal products, whether containing single or multiple species, is influenced by several factors, including the quality and type of the raw material, the yield and purity of extracted DNA, the choice of DNA barcode(s), and the availability of curated and comprehensive reference databases. Beyond using curated databases, a common strategy to enhance resolution involves restricting the reference database to species expected to occur within the specific study

context (Richardson et al., 2019). For example, Quaresma et al. (2024) constructed a global ITS2 database for vascular plant taxa that included a reference database for European crops and individual subsets for each of the 27 EU countries. A similar approach was adopted here, consisting of a stepwise classification, initially restricting taxonomic assignment to a custom database containing only species traditionally used for brain health; sequences that remained unassigned were subsequently compared against a broader medicinal plant database, and those still unresolved were finally evaluated against a comprehensive Viridiplantae database to capture all potential plant taxa present.Moreover, in this study, the custom databases constructed based on information mined from NCBI and from BOLD Systems were compared. Currently, BOLD is considered the most reliable reference sequence database for DNA barcoding, as sequence submissions undergo rigorous review, specimens must be taxonomically identified by an expert, and each entry includes extensive metadata to ensure accurate identification (Palhares et al., 2015). As a result of these stringent requirements, a lower number of sequences are available in BOLD than in NCBI. Consequently, as noted in section 3.3.1, the number of included sequences and the diversity of plant species represented in the custom BOLD databases were lower than in the custom NCBI databases.

Fig. 4 shows two-way Venn diagrams comparing the total number of expected species, as declared on the labels of single-species products (green group), with the species detected by ITS2 metabarcoding (pink group). As can be observed, classification using the pipeline with the custom NCBI databases detected 70 of the 88 labelled species, whereas the custom BOLD databases identified only 59. On the other hand, the NCBI databases detected 242 plant species across all samples, while the BOLD databases identified 258 species. Comparing the results obtained using the custom BOLD and NCBI databases separately (Table S2 and Table 1), both confirmed the correct labelling of 27 samples. However, they differed in the classification of two samples: P36 and P69. In sample P36, classification using the custom BOLD databases identified all

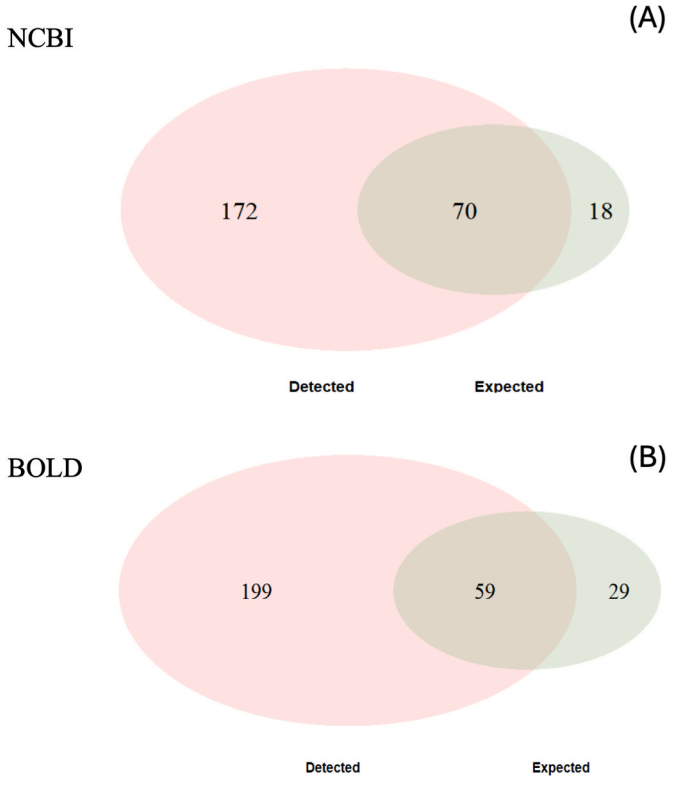

Fig. 4. Expected versus detected two-way Venn diagrams representing the detection of individual species via ITS2 metabarcoding in the monoherbal commercial samples (n = 88) using NCBI (A) and BOLD (B) custom databases.

12,923 reads as *Citrus aurantium,* whereas the custom NCBI databases assigned only 7798 reads to this species, with the remainder identified as *Citrus medica* and *Citrus aurantiifolia*. Although the custom BOLD database contains three ITS2 sequences for *C. medica* and two for *C. aurantiifolia*, these species may not have been detected due to limited representation in the database and the intraspecific variability characteristic of the genus. Different studies have documented substantial genetic diversity within the *Citrus* genus, highlighting the challenges of reliable species identification based on a small number of reference sequences. For instance, Herrero et al. (1996) observed high heterozygosity in limes, including *C. aurantiifolia*, suggesting an origin through interspecific hybridisation and explaining elevated intraspecific variability. Similarly, polymorphisms have also been documented for *C. medica* (Barbhuiya et al., 2016; Pessina et al., 2011). Therefore, the combination of limited reference sequences in the custom BOLD databases and potentially high variability within *C. medica* and *C. aurantiifolia* likely contributed to the observed classification discrepancies in sample P36. For sample P69, classification using the custom NCBI databases assigned all 10,201 reads to *Lavandula latifolia*, in agreement with the label, while the custom BOLD databases assigned 6507 reads to *Lavandula stoechas*. This discrepancy may be due to the limited number of ITS2 sequences for *Lavandula* species deposited in the BOLD database, which includes only two sequences of *Lavandula angustifolia* and one of *L. stoechas*, as well as to the sequence similarity between species.

Classification with both databases also showed that 15 samples did not match the label, failing to contain the declared species. According to the classification using both databases, samples P06, P72, P74 and P86, labelled as *Jasminum grandiflorum*, were mixtures of *Jasminum sambac* with *Camellia sinensis*, except for P72, which additionally showed a minor amount of *Melissa officinalis*. Considering that reference sequences of *J. grandiflorum* are included in both custom NCBI and BOLD databases, we can conclude that the presence of *J. sambac* constitutes a plant substitution. This may result from the similar morphology of the two species, leading to unintentional swap, or from economically motivated adulteration. Still, the consistent presence of a substantial amount of *C. sinensis* across all four samples, each acquired from distinct sellers, suggests an intentional substitution. Furthermore, this pattern of adulteration may indicate a broader problem in the value chain, potentially originating at the supplier level.

Sample P51 labelled as *C. sempervirens* exhibited complete substitution by *Platycladus orientalis*. Sample P08, also labelled as *C. sempervirens*, was revealed to be a mixture of species with no trace of the declared one. Regarding samples P34, P45, P62, and P79 labelled *Ptychopetalum olacoides,* we verified the absence of reference sequences in both databases, which would hinder their taxonomic assignment. However, distinct conclusions can be drawn from the classification results. In sample P34, all 14,094 reads were assigned to the *Croton* genus (family Euphorbiaceae), suggesting the presence of a single species and indicating full substitution of the labelled species. Multiple species (5–11) were identified in the remaining three samples, indicating the addition of extraneous plant material. Therefore, all four samples, each acquired from a different seller, were adulterated. Interestingly, *P. olacoides* is a species native to the Amazon region, where it is used in traditional medicine as a neurostimulant and aphrodisiac (Piato et al., 2010). Given its restricted native range and the challenges associated with sustainable harvesting, the observed adulteration may be linked to its limited availability in international markets.

Reference sequences were also unavailable for the labelled species in samples P11 (*Aristolochia esperanzae*), P13 (*Vinca minor*), and P53 (*Hypericum undulatum*). For P11 and P53, species from the same genera were assigned, namely *Aristolochia trilobata* and *Hypericum perforatum*. While definitive species confirmation would require sequencing of authenticated vouchers for comparison, both samples contained additional undeclared species. Thus, they were also considered adulterated due to the presence of multiple species. Regarding sample P13, a

Table 1
Comparison between Sanger sequencing of *matK* and *rbcL* markers and ITS2 metabarcoding of single-species commercial herbal products.

Internal ID	Labelled species	Supplier	Agreement with label				Result	Observations
			matK	*rbcL*	ITS2			
					NCBI [a]	BOLD [a]		
P01	*Hypericum androsaemum*	1	✗	✓✓	✓✓	✓✓	Authentic	*matK* misclassified as adulterated
P02	*Marsdenia condurango*	1	inc[b]	—— [c]	✗ (major species[d]; ≥95 %)	✗ (family ✓✓)	Mislabelled*	*: possible cross-contamination as >95 % of reads were assigned to this species (NCBI); BOLD database does not include sequences of this species; barcoding inconclusive.
P03	*Passiflora incarnata*	1	✗	✗	✗ (contains the species)	✗	Mislabelled	Potential adulteration; major ingredient (*C. papaya*) detected also by barcoding
P04	*Melissa officinalis*	1	✗	✓✓	✓✓	✓✓	Authentic	*matK* misclassified as adulterated
P05	*Artemisia vulgaris*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P06	*Jasminum grandiflorum*	1	✗	✗	✗	✗	Mislabelled	Full substitution by other species
P07	*Matricaria chamomilla*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P08	*Cupressus sempervirens*	1	✗	✗	✗	✗	Mislabelled	Full substitution by other species
P09	*Mikania hirsutissima*	1	inc[b]	inc[b]	✗ (contains the genus)	✗	Mislabelled	Mixture of plants containing a species of the *Mikania* genus
P10	*Pistacia lentiscus*	1	✗	✗	✗ (contains the species)	✗ (contains the species)	Mislabelled	Potential adulteration; major ingredient (*M. chamomilla*) detected also by barcoding
P11	*Aristolochia esperanzae*	1	✓✓	–	✗ (contains the genus)	✗	Mislabelled	Mixture of plants containing a species of the *Aristolochia* genus; *matK* misclassified as authentic
P12	*Eleutherococcus senticosus*	1	✓✓	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Contains the labelled species as major ingredient but mixed with other species; *matK* and *rbcL* misclassified as authentic
P13	*Vinca minor*	1	✓✓	✓✓	✗	✗	Inconclusive	Authentic according to barcoding; a different *Vinca* species according to metabarcoding. Further sequences of both species should be obtained
P14	*Rosmarinus officinalis*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P15	*Solanum nigrum*	1	✓✓	✓✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *matK* and *rbcL* misclassified as authentic
P16	*Melissa officinalis*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P17	*Foeniculum vulgare*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P18	*Crataegus monogyna*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P19	*Verbena officinalis*	1	✓✓	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as major ingredient, but mixed with other species; *matK* and *rbcL* misclassified as authentic
P20	*Ginkgo biloba*	1	✗	✓✓	✗ (contains the species)	✗	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassifies as authentic
P21	*Equisetum arvense*	1	—— [c]	✓✓	✗ (major species)	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P22	*Hypericum perforatum*	1	✗	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; BOLD only identifies the species, but a

(*continued on next page*)

Table 1 (*continued*)

Internal ID	Labelled species	Supplier	Agreement with label				Result	Observations
			matK	*rbcL*	ITS2			
					NCBI [a]	BOLD [a]		
								significant number of reads assigned to other genera; *rbcL* misclassified as authentic
P23	*Valeriana officinalis*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P24	*Angelica archangelica*	1	✓✓	✓✓	✓✓	✓✓	Authentic	
P25	*Panax ginseng*	1	✗	✗	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species
P26	*Paullinia cupana*	1	✗	✗	✗ (major species)[d]	✗	Authentic**	** Pharmacopoeias allows for 2 % foreign matter. Contains a minor amount of *Passiflora incarnata*, possibly a cross-contamination in the value chain/processing. Absence of *matK* sequences and only one unverified *rbcL* sequence in GenBank
P27	*Cola nitida*	1	—— [c]	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P28	*Ginkgo biloba*	2	—— [c]	✓✓	✗ (contains the species)	✗	Mislabelled	Mixture of plants; the traditional use of *E. senticosus* and its amounts suggests adulteration by species substitution; *rbcL* misclassified as authentic.
P29	*Eleutherococcus senticosus*	2	✓✓	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient but mixed with another species; *matK* and *rbcL* misclassified as authentic
P30	*Ilex paraguariensis*	2	✓✓	✓✓	✓✓	✓✓	Authentic	
P31	*Hypericum perforatum*	2	✗	✓✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P32	*Passiflora incarnata*	2	–	✓✓	✓✓	✓✓	Authentic	
P33	*Moringa oleifera*		✓✓	✓✓	inconclusive	inconclusive	Potentially correct	Authentic according to barcoding; inconclusive according to metabarcoding
P34	*Ptychopetalum olacoides*	2	✗	✗	✗	✗	Mislabelled	Substitution with a species of the *Croton* genus
P35	*Myrcia citrifolia*	2	—— [c]	✗	✗ (contains the species)	✗	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species
P36	*Citrus aurantium*	2	✓✓	✓✓	✗ (major species)[d]	✓✓	Inconclusive	Potentially adulterated by admixture with other two *Citrus* species
P37	*Tilia europaea*	2	✓✓	✓✓	✗	✗	Inconclusive	Authentic according to barcoding; substitution by other species of the same genus according to metabarcoding
P38	*Centella asiatica*	2	✓✓	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *matK* and *rbcL* misclassified as authentic
P39	*Cymbopogon citratus*	2	✗	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *rbcL* misclassified as authentic
P40	*Valeriana officinalis*	2	✓✓	✓✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *matK* and *rbcL* misclassified as authentic

(*continued on next page*)

Table 1 (*continued*)

Internal ID	Labelled species	Supplier	Agreement with label				Result	Observations
			matK	*rbcL*	ITS2			
					NCBI [a]	BOLD [a]		
P41	*Citrus limon*	2	✓	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other *Citrus* species; *matK* and *rbcL* misclassified as authentic
P42	*Eleutherococcus senticosus*	3	✓	✓	✓	✓	Authentic	
P43	*Passiflora incarnata*	3	✗	✗	✗ (contains the species)	✗ (contains the genus)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species.
P44	*Ginkgo biloba*	3	✗	✓	✗ (contains the species)	✗	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P45	*Ptychopetalum olacoides*	3	✗	✗	✗	✗	Mislabelled	Mixture of plants
P46	*Moringa oleifera*		✓	✓	inconclusive	inconclusive	Potentially correct	Authentic according to barcoding; inconclusive according to metabarcoding
P47	*Ilex paraguariensis*	3	✓	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *matK* and *rbcL* misclassified as authentic
P48	*Cymbopogon citratus*	3	✗	✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled*	*: possible cross-contamination as >95 % of reads were assigned to this species (NCBI); Mixture of plants; contains the labelled species, which is the major ingredient, but mixed with another species
P49	*Carica papaya*	3	✓	✓	✓	✓	Authentic	
P50	*Melissa officinalis*	3	✗	✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *rbcL* misclassified as authentic
P51	*Cupressus sempervirens*	3	—— [c]	✗	✗	✗	Mislabelled	Full substitution by other species
P52	*Pistacia lentiscus*	3	✗	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with another species; *rbcL* misclassified as authentic
P53	*Hypericum undulatum*	3	✗	✗	✗	✗	Mislabelled	Mixture of plants
P54	*Hypericum perforatum*	3	✗	✓	✓	✓	Authentic	*MatK* misclassified as adulterated
P55	*Carica papaya*	3	✓	✓	✓	✓	Authentic	
P56	*Carica papaya*	4	✓	✓	✓	✓	Authentic	
P57	*Valeriana officinalis*	4	✓	✓	✓	✓	Authentic	
P58	*Centella asiatica*	4	✓	✓	✗ (contains the species)	✗ (major species)	Mislabelled	Mixture of plants; contains the labelled species as the major ingredient, but mixed with another species; *matK* and *rbcL* misclassified as authentic
P59	*Bacopa monnieri*	4	✓	✓	✓	✓	Authentic	
P60	*Moringa oleifera*	4	✓	✓	inconclusive	inconclusive	Potentially correct	Authentic according to barcoding; inconclusive according to metabarcoding due to few sequences obtained (<40)
P61	*Citrus sinensis*	4	✓	✗	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species;
P62	*Ptychopetalum olacoides*	4	✗	✗	✗	✗	Mislabelled	Mixture of plants
P63	*Matricaria chamomilla*	4	✓	✓	✓	✓	Authentic	
P64	*Ginkgo biloba*	4	—— [c]	✓	✗ (major species)[d]	✗	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic

(*continued on next page*)

Table 1 (*continued*)

Internal ID	Labelled species	Supplier	Agreement with label				Result	Observations
			matK	*rbcL*	ITS2			
					NCBI [a]	BOLD [a]		
P65	*Hypericum perforatum*	4	✗	—— [c]	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species
P66	*Eschscholzia californica*	4	✓	✗	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *matK* misclassified as authentic
P67	*Passiflora incarnata*	4	✗	✓	✗ (contains the species)	✗ (contains the genus)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P68	*Rosmarinus officinalis*	4	✓	✓	✓	✓	Authentic	
P69	*Lavandula latifolia*	4	✓	✓	✓	✗	Authentic	Authentic according to both metabarcoding (NCBI classification) and barcoding; no sequences of this species included in the BOLD database
P70	*Lavandula stoechas*	4	✓	✓	✗ (major species[d]; ≥98 %)	✗	Authentic**	** Pharmacopoeias allow for 2 % foreign matter. Contains a minor amount of *Senna alexandrina*, possibly a cross-contamination in the value chain/processing
P71	*Melissa officinalis*	4	✓	✓	✓	✓	Authentic	
P72	*Jasminum grandiflorum*	4	✗	✗	✗	✗	Mislabelled	Species substitution
P73	*Passiflora incarnata*	5	—— [c]	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *rbcL* misclassified as authentic
P74	*Jasminum grandiflorum*	5	✗	✗	✗	✗	Mislabelled	Species substitution
P75	*Cymbopogon citratus*	5	✗	✗	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as major ingredient, but mixed with another species; *rbcL* misclassified as authentic; *matK* detects one of the adulterants
P76	*Hypericum perforatum*	5	—— [c]	✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *rbcL* misclassified as authentic
P77	*Verbena officinalis*	5	✓	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *matK* and *rbcL* misclassified as authentic
P78	*Valeriana officinalis*	5	✓	✓	✗ (major species)[d]	✗ (major species)[d]	Mislabelled	Mixture of plants; contains the labelled species as a major ingredient, but mixed with another species; *matK* and *rbcL* misclassified as authentic
P79	*Ptychopetalum olacoides*	5	—— [c]	✗	✗	✗	Mislabelled	Mixture of plants
P80	*Crataegus monogyna*	5	✓	✓	✓	✓	Authentic	
P81	*Centella asiatica*	5	✓	✓	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species; *matK* and *rbcL* misclassified as authentic
P82	*Matricaria chamomilla*	5	—— [c]	✓	✓	✓	Authentic	
P83	*Artemisia vulgaris*	5	—— [c]	✓	✓	✓	Authentic	
P84	*Eleutherococcus senticosus*	5	✓	✓	✓	✓	Authentic	
P85	*Eschscholzia californica*	6	—— [c]	—— [c]	✗ (contains the species)	✗ (contains the species)	Mislabelled	Mixture of plants; contains the labelled species but mixed with

(*continued on next page*)

Table 1 (*continued*)

Internal ID	Labelled species	Supplier	Agreement with label				Result	Observations
			matK	*rbcL*	ITS2			
					NCBI [a]	BOLD [a]		
								other species; *matK* and *rbcL* misclassified as authentic
P86	*Jasminum grandiflorum*	6	✕	✕	✕	✕	Mislabelled	Species substitution
P87	*Bacopa monnieri*	6	—— [c]	—— [c]	✕ (major species)[d]	✕ (major species)[d]	Mislabelled*/ Authentic**	*Possible cross-contamination as >95 % of reads were assigned to this species (NCBI); **Pharmacopoeias allow for 2 % foreign matter, and according to BOLD contains 1 %–2 % of other species; overall, results suggest cross-contamination over the value chain/processing
P88	*Passiflora incarnata*	6	✕	✕	✕ (contains the species)	✕	Mislabelled	Mixture of plants; contains the labelled species but mixed with other species

[a] Classification was performed using the custom curated NCBI databases and custom BOLD databases (as shown in Table S2).
[b] Inconclusive due to the lack of reference sequences available in NCBI.
[c] Did not provide amplification with the used primer set.
[d] Contains different species but the labeled species is the major one.

voucher specimen of *V. minor* was obtained from a botanical garden (Jardim Botânico d'Ajuda, Lisbon), and the corresponding ITS2 sequence was added to both custom databases. Despite this effort, the sample was identified as *Vinca difformis*, a species mainly cultivated as an ornamental, with limited documented use in traditional or medicinal applications. Nevertheless, including additional reference sequences would be advisable to account for possible intraspecific variability.

Regarding sample P09 (*Mikania hirsutissima*), only three ITS2 sequences are available in NCBI, all originating from the same study (Godoy et al., 2017). Nevertheless, in addition to reads assigned to the *Mikania* genus, sample P09 also revealed the presence of several other species, with *Artemisia vulgaris* accounting for 62 % of the reads at the species level, thus being considered adulterated.

Sample P37 was identified as a mixture of *Tilia cordata* and *Tilia tomentosa*, whereas the label referred to *Tilia europaea*. *Tilia*×*europaea*, commonly known as common lime or linden, is a hybrid of *T. cordata* and *Tilia platyphyllos*. Interestingly, *T. cordata* is also widely known as linden. Moreover, in traditional medicine, several *Tilia* species (including the mentioned ones) are often used interchangeably due to similar morphology and therapeutic properties. Therefore, this case likely reflects a taxonomic confusion or unintentional mislabelling rather than deliberate adulteration. This highlights the need for improved traceability, good harvesting practices, and stricter quality control across the herbal supply chain.

The remaining 43 samples were also classified as adulterated by both custom databases, as their composition did not fully match the label information despite the detection of the labelled plant species. Among these, classification using the custom NCBI databases identified the labelled species in ≥98 % of reads for two samples (P26, P70). According to the BOLD classification, sample P87 also reached ≥98 %, yet the NCBI classification resulted only in 96 % of the reads attributed to the labelled species. According to pharmacopeial monographs for herbal raw material, such as those from the European and United States Pharmacopoeias, the content of up to 2 % of foreign matter is usually allowed (Parveen et al., 2016). Given this criterion, samples P26 and P70 can also be considered authentic. Moreover, if we consider that intentional adulteration is typically associated with the presence of a substitute species in higher proportions, as low-level substitution offers limited economic benefit, the presence of <5 % of non-labelled species in samples may be attributed to unintentional contamination. Cross-contamination can occur during various stages of the value chain, including harvesting (due to co-collection of similar species), drying (from shared surfaces/areas with other plants), and grinding and packaging (via equipment carryover). Accordingly, the presence of extraneous plant material in samples P02, P48, and P87, which contained ≥95 % of the labelled species, may result from incidental contamination rather than deliberate adulteration.

Among the remaining 13 samples, discrepancies were observed between NCBI and BOLD classifications. These are primarily attributed to the limited representation of certain species in the BOLD databases, as mentioned in section 3.3.2.2. Nevertheless, in several cases, classification at the genus or family level was still consistent with the labelled species. For example, in sample P02, BOLD assigned only 251 reads to the species level (corresponding to 14 different species, thus likely representing cross-contamination), whereas 11,005 reads were classified within the family Apocynaceae, which includes *Marsdenia condurango*, the labelled species, which was identified using the custom NCBI databases.

4.2. Monoherbal samples: DNA barcoding versus DNA metabarcoding

Overall, ITS2 metabarcoding demonstrated superior amplification success compared to *rbcL* and *matK*, with the latter showing the lowest success rate across samples. As expected, a key distinction between the two approaches was the ability of metabarcoding to detect multiple species within a single sample. This feature proved crucial in the assessment of the samples, as several samples classified as authentic monoherbal products by DNA barcoding were, in fact, polyherbal (Table 1).

Using *matK* barcoding, 43 samples were consistent with their label, while 29 were mislabelled. In comparison, *rbcL* barcoding identified 61 samples as matching the label and 21 as mislabelled. A total of 40 samples yielded consistent results across both markers, with the species identified by both *matK* and *rbcL* matching the label. If a less stringent criterion were applied—considering a sample as matching the label if either of the two markers (*matK* or *rbcL*) identified the labelled species—the number of correct or potentially correct samples increases to 62 (Tables 1 and S1). However, the higher sensitivity of ITS2 metabarcoding revealed that many of the samples categorised as "authentic" by DNA barcoding actually contained undeclared plant species in addition to the labelled one. In some cases, the labelled species was the predominant taxon, while in others, it was present only in minor

proportions. For instance, in sample P25, labelled *Panax ginseng*, the declared species accounted for just 16 % of the total reads. Similarly, sample P28, labelled *Ginkgo biloba*, contained only ~44 % of the reads matching this species, with a comparable proportion identified as *Eleutherococcus senticosus* (Brazilian ginseng). Given the known therapeutic uses of *E. senticosus* and its substantial presence in the sample, this likely represents intentional adulteration. Both *P. ginseng* and *G. biloba* are high-demand botanicals in the global market as they are widely used in food supplements, thus making these species frequent targets of economically motivated adulteration (Grazina et al., 2020b, 2021).

The comparison between DNA barcoding and ITS2 metabarcoding also revealed that, in cases of plant mixtures, the predominant species was frequently detectable by DNA barcoding, particularly when using the *rbcL* marker. This included not only samples in which the labelled species was the dominant taxon (highlighted as major species in Table 1) but also samples in which the adulterant was present in higher abundance. For instance, in sample P08, *rbcL* identified *Hesperocyparis stephens*onii, and ITS2 corroborated the presence of multiple reads classified within the *Hesperocyparis* genus. Similarly, in sample P03, both *matK* and *rbcL* detected *Carica papaya*, which was identified as the predominant species by ITS2 metabarcoding. Another interesting observation was that *matK* barcoding could identify the adulterant species in several samples even when these were not the dominant taxa in the mixture (Tables S1 and S2). For example, in sample P65, *matK* classified it as *Calystegia hederacea*, which was detected in only ~30 % of the reads by ITS2 metabarcoding. Similar findings were observed for samples P25, P31, P43, P50, P51, P39, and P53. Conversely, when comparing the results obtained with *matK* and *rbcL* to those obtained by ITS2 metabarcoding (Table 1), several samples that matched the label according to DNA barcoding were mixtures of different plant species. Thus, these samples were mislabelled and could not be considered authentic. Specifically, 12 and 23 samples matched the product label when analysed using *matK* and *rbcL*, respectively, and would have been incorrectly classified as authentic if only these markers were used. These findings underscore the limitations of barcoding approaches in detecting complex plant mixtures and highlight the discriminatory power of ITS2 metabarcoding, particularly in the detection of cases of adulteration or unintentional admixture.

As for the polyherbal samples, which were analysed exclusively by ITS2 metabarcoding, similar results were obtained with both custom reference databases (Table S2). Of the six samples analysed, only one (P91) contained all the species listed on the label. However, *Citrus limon*, which was not declared on the P91 label, was also identified but at very low abundance (2 % and 1 % of the total reads identified by NCBI and BOLD, respectively). This minor amount likely reflects incidental cross-contamination during production rather than intentional adulteration. Two samples (P92 and P93) contained most of the labelled species, except for *Olea europaea*, which was not detected in either sample. In addition, a small number of reads of *Solanum americanum* were identified in P92, which may also indicate cross-contamination. Although we cannot exclude the possibility that certain species, such as *O. europaea*, were not detected due to methodological limitations, a considerable proportion of reads from *Tilia* sp. (14–16 %) were identified in sample P93, thus indicating mislabelling. The most discrepant case was found in sample P90, which declared eight species, and only one was detected. Similarly, in P89, only four out of the 14 labelled species were identified. However, in this case, some of the ingredients may have been added solely for their aromatic properties or in the form of extracts (e.g., cinnamon, red pepper, cardamom), which would explain their non-detection. Sample P100 contained only labelled species, yet three expected species (*O. europaea*, *P. incarnata* and *Valeriana officinalis*) were not detected. Overall, the six polyherbal samples listed a total of 46 labelled plant ingredients, of which 20 were successfully identified using either BOLD or NCBI custom databases (Fig. 5).

Previous studies have also shown discrepancies between the labelled and detected plant species in herbal products such as teas, infusions and

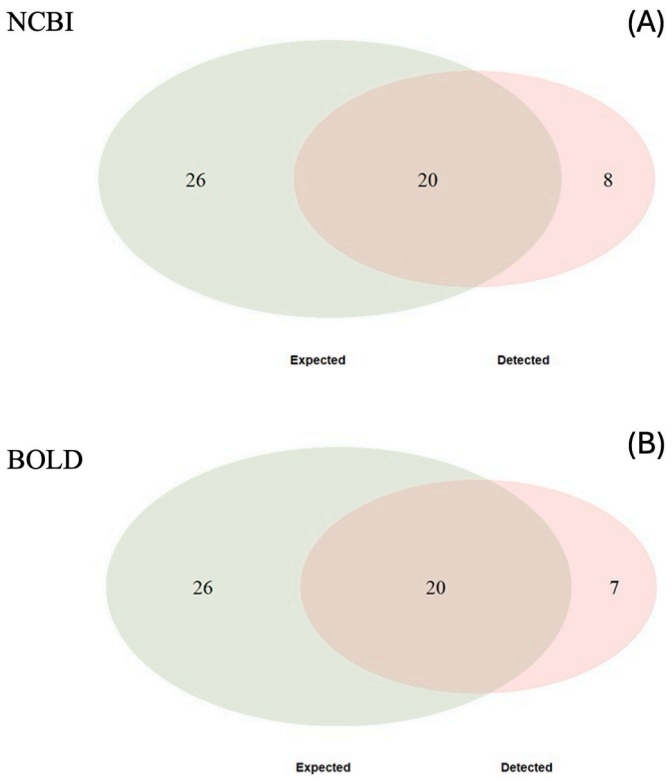

Fig. 5. Expected versus detected two-way Venn diagrams representing the detection using ITS2 metabarcoding of the individual ingredients of six polyherbal commercial samples, using NCBI (A) and BOLD (B) custom databases.

food supplements. In a comprehensive global survey, Ichim (2019) compiled and analysed the data from multiple studies that employed DNA-based methods to evaluate the authenticity of commercial herbal products. The analysis of 5957 products sold in 37 countries revealed that 27 % were adulterated, highlighting the widespread nature of species substitution and mislabelling in the herbal product market. In this study, ITS2 metabarcoding revealed that only 32 % (27/85) of the products matched the labelled species. The remaining samples either contained additional undeclared species or lacked the labelled one entirely. While standard barcoding with *rbcL* and *matK* suggested higher authenticity rates, this approach often failed to detect mixed or substituted species, particularly when the labelled ingredient was predominant and preferentially amplified. So far, only a few studies have compared the effectiveness of barcoding and metabarcoding to authenticate medicinal plants. Ivanova et al. (2016) tested 15 herbal supplements using ITS2 and *rbcL* Sanger sequencing alongside ITS2 metabarcoding. The authors found that DNA from the manufacturer-listed medicinal plants was successfully identified in four samples by ITS2 barcoding, seven by *rbcL* barcoding, while ITS2 metabarcoding succeeded in eight. Mück et al. (2024) evaluated 51 Traditional Chinese Medicine products using Sanger sequencing and metabarcoding of ITS1 and ITS2. ITS2 barcoding successfully identified the labelled species in six products, while four were resolved only to the genus level, seven contained unexpected species, and no identification was possible for 32. Similar to our study, ITS2 metabarcoding confirmed a low number of authentic samples with only three samples matching the label; 25 were successfully identified, but only at the genus level, and 19 contained unexpected species.

The higher number of mislabelled samples identified by ITS2 metabarcoding, many of which could not be identified by traditional barcoding due to their mixed-species composition, highlights the superior effectiveness of metabarcoding in analysing complex herbal products.

Although metabarcoding is widely used in food authentication (Nehal et al., 2021), pollen analysis (Sickel et al., 2015), environmental monitoring (Ruppert et al., 2019), and microbiome profiling (Abdelfattaha et al., 2018), its application to herbal product authentication remains limited. Among the few studies, several focus on specific taxa, such as herbal products of *Hypericum perforatum* (Raclariu, Paltinean, et al., 2017), *Veronica officinalis* (Raclariu, Mocan, et al., 2017), *Echinacea* (Booker et al., 2018), and *Bacopa monnieri* (Kumari et al., 2023), or on herbal products with specific ingredients, such as those listed on Traditional Chinese Medicine (Mück et al., 2024) or the Thai National List of Essential Medicines (Urumarudappa et al., 2020). Only a few studies adopted a broader approach by applying metabarcoding to analyse a wide variety of herbal products, both monoherbal and polyherbal formulations, encompassing multiple plant taxa (Anthoons et al., 2021; Frigerio et al., 2021; Travadi et al., 2023; Seethapathy et al., 2019).

Our results evidence a high level of adulteration due to the addition of extraneous species, consistent with the rates reported by other studies. Anthoons et al. (2021) analysed 71 herbal products from the Greek market using ITS2 metabarcoding and found that only ~40 % matched the labelled composition. Frigerio et al. (2021) used ITS2 and psbA-trnH metabarcoding on 15 herbal teas, detecting on average 70 % of the labelled ingredients but also several undeclared species, with only one product fully matching its label. Travadi et al. (2023) applied newly designed ITS2 and *rbcL* primers to 32 commercial products, detecting the declared species in most samples but also numerous undeclared plants, reflecting a high adulteration rate, consistent with our findings.

5. Conclusions

Medicinal plants used in traditional herbal products and food supplements are typically traded as crude drugs, with their quality frequently assessed through botanical, physicochemical, and chemical tests outlined in official Pharmacopoeias. However, these methods often fail to ensure species-level identification, especially when materials are fragmented or powdered. Therefore, DNA-based approaches offer a promising alternative for authenticating medicinal plants, with DNA barcoding and metabarcoding being the most widely applied. DNA barcoding is widely recognised across different fields, including food authentication, as a robust, rapid, cost-effective and broadly applicable method for species identification. While Sanger sequencing of standard barcodes is effective when the sample contains DNA from a single species, it is not suitable for samples with mixed DNA, whether from multispecies formulations or contaminated/adulterated single-species products, owing to preferential amplification of the dominant template due to primer biases and its inability to resolve multiple sequences in a single read. In this study, both *rbcL* and *matK* could detect some adulterated samples, particularly when the labelled species was absent or present in low abundance. However, when the labelled species was the dominant ingredient in the mixture, it tended to be preferentially amplified, often resulting in false classification as authentic, despite the presence of undeclared plant species. In many such cases, overlapping peaks in the Sanger electropherograms were indicative of mixed DNA templates, a finding later confirmed by ITS2 metabarcoding.

Although ITS2 metabarcoding requires specialised equipment, bioinformatic expertise and is generally more costly than conventional barcoding, it has the major advantage of simultaneous detection of multiple species within a sample, enabling the identification of both undeclared and substitute plant species, as well as potential contaminants.

Overall, the results of this study highlight the limitations of *rbcL* and *matK* barcoding in detecting the presence of multiple species in products labelled as single species, particularly when the declared species is dominant. In contrast, metabarcoding is a more powerful and reliable approach for comprehensively assessing medicinal plant products across the entire value chain of herbal products. Its ability to detect labelled and unlabelled components makes it a valuable tool for strengthening regulatory oversight and safeguarding consumer health.

CRediT authorship contribution statement

M. Alice Pinto: Writing – review & editing, Supervision, Conceptualization. **Joana Santos:** Methodology, Investigation. **Andreia Quaresma:** Writing – review & editing, Methodology, Investigation. **Mónica Honrado:** Investigation. **José Rufino:** Methodology. **Joana S. Amaral:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through the project "POIROT: novel methods and approaches for detecting the illegal addition of Pharmaceutical drugs and bOtanIcal adulteRatiOn in plant food supplements" (PTDC/SAU-PUB/3803/2021, DOI: 10.54499/PTDC/SAU-PUB/3803/2021) and through national funds FCT/MCTES (PIDDAC): CIMO, UIDB/00690/2020 (DOI: 0.54499/UIDB/00690/2020) and UIDP/00690/2020 (DOI: 10.54499/UIDP/00690/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). M. Honrado and A. Quaresma acknowledge the Ph.D. scholarship from Fundação para a Ciência e Tecnologia (FCT 2021.08119.BD and 2020.05155.BD, respectively). The authors are thankful to Jardim Botânico d'Ajuda (Lisbon), Universidade Tecnológica Federal do Paraná (Brazil), and Herbário da Universidade de Coimbra for supplying vouchers of plant species.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2025.111893.

Data availability

Data will be made available on request.

References

Abdelfattaha, A., Malacrinò, A., Wisniewski, M., Cacciola, S. O., & Schena, L. (2018). Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. *Biological Control, 20*, 1–10. https://doi.org/10.1016/j.biocontrol.2017.07.009

Anthoons, B., Karamichali, I., Schrøder-Nielsen, A., Drouzas, A. D., de Boer, H., & Madesis, P. (2021). Metabarcoding reveals low fidelity and presence of toxic species in short chain-of-commercialization of herbal products. *Journal of Food Composition and Analysis, 97*, Article 103767. https://doi.org/10.1016/j.jfca.2020.103767

Barbhuiya, A. R., Khan, M. L., & Dayanandan, S. (2016). Genetic structure and diversity of natural and domesticated populations of *Citrus medica* L. in the Eastern Himalayan region of Northeast India. *Ecology and Evolution, 6*, 3898–3911. https://doi.org/10.1002/ece3.2174

Booker, A., Johnston, D., & Heinrich, M. (2018). Value chains of herbal medicines—Research needs and key challenges in the context of ethnopharmacology. *Phytomedicine, 50*, 147–160. https://doi.org/10.1016/j.phymed.2018.03.058

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). doi.org/journal.pone.0008613.

Chen, S., Yin, X., Han, J., Sun, W., Yao, H., Song, J., & Li, X. (2023). DNA barcoding in herbal medicine: Retrospective and prospective. *Journal of Pharmaceutical Analysis, 13*(5), 431–441. https://doi.org/10.1016/j.jpha.2023.03.008

Edgar, R. C. (2016). SINTAX: A simple Non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*. , Article 074161. https://doi.org/10.1101/074161

Fazekas, A. J., Kuzmina, M. L., Newmaster, S. G., & Hollingsworth, P. M. (2012). DNA barcoding methods for land plants. In W. Kress, & D. Erickson (Eds.), *DNA barcodes. Methods in molecular biology, 858*. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-61779-591-6_11.

Frigerio, J., Agostinetto, G., Mezzasalma, V., De Mattia, F., Labra, M., & Bruno, A. (2021). DNA-based herbal teas' authentication: An ITS2 and psbA-trnH multi-marker DNA metabarcoding approach. *Plants, 10*(10), 2120. https://doi.org/10.3390/plants10102120

Gafner, S., Blumenthal, M., Foster, S., Cardellina, J. H., Khan, I. A., & Upton, R. (2023). Botanical ingredient forensics: Detection of attempts to deceive commonly used analytical methods for authenticating herbal dietary and food ingredients and supplements. *Journal of Natural Products, 86*(2), 460–472. https://doi.org/10.1021/acs.jnatprod.2c00929

Gansauge, M. T., Aximu-Petri, A., Nagel, S., & Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols, 15*(8), 2279. https://doi.org/10.1038/s41596-020-0338-0

Godoy, S. M., Silva, J. F. M., Paula, G. B. N., Ruas, P. M., Góes, B. D., & Ruas, C. F. (2017). Phylogenetic relationships of Brazilian Mikania species (Asteraceae, Eupatorieae) based on multilocus DNA markers. *Botanical Journal of the Linnean Society, 184*(3), 326–346. https://doi.org/10.1093/botlinnean/box030

Grand View Research. (2024). Herbal medicine market size, Share & trends analysis report by intervention (Ayurveda, traditional Chinese medicine), by product form (Tablet/Capsules, powder), by source, by distribution channel, by Region, and segment forecasts, 2024 - 2030. Available from: https://www.grandviewresearch.com/industry-analysis/herbal-medicine-market-report. May 2025.

Grazina, L., Amaral, J. S., Costa, J., & Mafra, I. (2020). Authentication of Ginkgo biloba herbal products by a novel quantitative real-time PCR approach. *Foods, 9*(9), 1233. https://doi.org/10.3390/foods9091233

Grazina, L., Amaral, J. S., Costa, J., & Mafra, I. (2021). Towards authentication of Korean ginseng-containing foods: Differentiation of five Panax species by a novel diagnostic tool. *Food Control, 123*, Article 107833. https://doi.org/10.1016/j.foodcont.2020.107833

Grazina, L., Amaral, J. S., & Mafra, I. (2020). Botanical origin authentication of dietary supplements by DNA-based approaches. *Comprehensive Reviews in Food Science and Food Safety, 19*(3), 1080–1109. https://doi.org/10.1111/1541-4337.12551

Heinrich, M., & Anagnostou, S. (2017). From pharmacognosia to DNA-Based medicinal plant authentication - Pharmacognosy through the centuries. *Planta Medica, 83*(14–15), 1110–1116. https://doi.org/10.1055/s-0043-108999

Herrero, R., Asíns, M. J., Carbonell, E. A., et al. (1996). Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. *Theoretical and Applied Genetics, 92*, 599–609. https://doi.org/10.1007/BF00224564

Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., … Grp, C. P. W. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Ibrahim, M., Detroja, A., Sheth, B. P., Bhadja, P., Sanghvi, G., & Bishoyi, A. K. (2024). Existing status and future advancements of adulteration detection techniques in herbal products. *Molecular Biology Reports, 51*(1). doi.org/s11033-023-09122-9.

Ichim, M. C. (2019). The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. *Frontiers in Pharmacology, 10*, 1227. https://doi.org/10.3389/fphar.2019.01227

Illumina. (2025). 16S metagenomic sequencing Library preparation. *Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System. Part # 15044223 Rev. B.* https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. (Accessed 20 January 2025).

Ivanova, N. V., Kuzmina, M. L., Braukmann, T. W. A., Borisenko, A. V., & Zakharov, E. V. (2016). Authentication of herbal supplements using next-generation sequencing. *PLoS One, 11*(5), Article e0156426. https://doi.org/10.1371/journal.pone.0156426

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research, 40*(1). https://doi.org/10.1093/nar/gkr771

Kumari, P., Shah, A. P., Travadi, T., Sharma, S., Pandit, R., Joshi, C., & Joshi, M. (2023). Comprehensive analysis using DNA metabarcoding, SCAR marker based PCR assay, and HPLC unveils the adulteration in Brahmi herbal products. *Molecular Biology Reports, 50*, 7605–7618. https://doi.org/10.1007/s11033-023-08653-5

Letsiou, S., Madesis, P., Vasdekis, E., Montemurro, C., Grigoriou, M. E., Skavdis, G., Moussis, V., Koutelidakis, A. E., & Tzakos, A. G. (2024). DNA barcoding as a plant identification method. *Applied Sciences, 14*(4), 1415. https://doi.org/10.3390/app14041415

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One, 8*(4). https://doi.org/10.1371/journal.pone.0061217

Molina, J., Sherpa, C., Ng, J., Sonam, T., & Stuhr, N. (2018). DNA barcoding of online herbal supplements: Crowd-sourcing pharmacovigilance in high school. *Open Life Sciences, 13*(1), 48–55. https://doi.org/10.1515/biol-2018-0007

Mück, F., Scotti, F., Mauvisseau, Q., Thorbek, B. L. G., Wangensteen, H., & de Boer, H. J. (2024). Three-tiered authentication of herbal traditional Chinese medicine ingredients used in women's health provides progressive qualitative and quantitative insight. *Frontiers in Pharmacology, 15*, Article 1353434. https://doi.org/10.3389/fphar.2024.1353434

Nehal, N., Choudhary, B., Nagpure, A., & Gupta, R. K. (2021). DNA barcoding: A modern age tool for detection of adulteration in food. *Critical Reviews in Biotechnology, 41*(5), 767–791. https://doi.org/10.1080/07388551.2021.1874279

Palhares, R.M., Drummond, M.G., Brasil, B.D.S.A.F., Cosenza, G.P., Brandão, M.G.L., & Oliveira, G. (2015). Medicinal plants recommended by the World Health Organization: DNA Barcode identification associated with chemical analyses guarantees their quality. *PLoS One*, *10*(5), Article e0127866. 10.1371/journal.pone.0127866.

Pandit, R., Travadi, T., Sharma, S., Joshi, C., & Joshi, M. (2021). DNA meta-barcoding using *rbcL* based mini-barcode revealed presence of unspecified plant species in Ayurvedic polyherbal formulations. *Phytochemical Analysis, 32*, 804–810. https://doi.org/10.1002/pca.3026

Parveen, I., Gafner, S., Techen, N., Murch, S., & Khan, I. (2016). DNA barcoding for the identification of botanicals in herbal medicine and dietary supplements: Strengths and limitations. *Planta Medica, 82*(14), 1225–1235. https://doi.org/10.1055/s-0042-111208

Pessina, D., Gentili, R., & Rossi, G. (2011). DNA content, morphometric and molecular marker analyses of *Citrus limonimedica*, *C. limon* and *C. medica* for determining variability and genetic relationships within the genus *Citrus*. *Scientia Horticulturae, 129*, 663–673. https://doi.org/10.1016/j.scienta.2011.05.012

Piato, A. L., Detanico, B. C., Linck, V. M., Herrmann, A. P., Nunes, D. S., & Elisabetsky, E. (2010). Anti-stress effects of the "tonic" Ptychopetalum olacoides (Marapuama) in mice. *Phytomedicine, 17*(3–4), 248–253. https://doi.org/10.1016/j.phymed.2009.07.001

Quaresma, A., Ankenbrand, M. J., Garcia, C. A. Y., Rufino, J., Honrado, M., Amaral, J., … Keller, A. (2024). Semi-automated sequence curation for reliable reference datasets in ITS2 vascular plant DNA (meta-)barcoding. *Scientific Data, 11*(1). https://doi.org/10.1038/s41597-024-02962-5

Quaresma, A., Brodschneider, R., Gratzer, K., Gray, A., Keller, A., Kilpinen, O., … Pinto, M. A. (2021). Preservation methods of honey bee-collected pollen are not a source of bias in ITS2 metabarcoding. *Environmental Monitoring and Assessment, 193*(12). https://doi.org/10.1007/s10661-021-09563-4

Raclariu, A. C., Heinrich, M., Ichim, M. C., & de Boer, H. (2018). Benefits and limitations of DNA barcoding and metabarcoding in herbal product authentication. *Phytochemical Analysis, 29*(2), 123–128. https://doi.org/10.1002/pca.2732

Raclariu, A. C., Mocan, A., Popa, M. O., Vlase, L., Ichim, M. C., Crisan, G., Brysting, A. K., & de Boer, H. (2017). Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with Veronica chamaedrys. *Frontiers in Pharmacology, 8*, 378. https://doi.org/10.3389/fphar.2017.00378

Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., Crisan, G., Brysting, A. K., & de Boer, H. (2017). Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. *Scientific Reports, 7*, 1291. https://doi.org/10.1038/s41598-017-01389-w

Richardson, R. T., Curtis, H. R., Matcham, E. G., Lin, C. H., Suresh, S., Sponsler, D. B., … Johnson, R. M. (2019). Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Molecular Ecology, 28*(3), 686–697. https://doi.org/10.1111/mec.14975

Rizzo, I., Vedoya, M. P., Maurutto, M. C., Haidukowski, M. C., & Pascale, A. (2004). Assessment of toxigenic fungi on Argentinean medicinal herbs. *Microbiological Research, 159*(2), 113–120. https://doi.org/10.1016/j.micres.2004.01.013

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ, 4*. https://doi.org/10.7717/peerj.2584

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation, 17*, Article e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Seethapathy, G. S., Raclariu-Manolica, A.-C., Anmarkrud, J. A., Wangensteen, H., & de Boer, H. J. (2019). DNA metabarcoding authentication of Ayurvedic herbal products on the European market raises concerns of quality and fidelity. *Frontiers in Plant Science, 10*, 68. https://doi.org/10.3389/fpls.2019.00068

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., & Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology, 15*, 20. https://doi.org/10.1186/s12898-015-0051-y

Srivastava, S., Sanchita, Bhargava, M., & Sharma, A. (2016). DNA barcoding of medicinal plants. In H. S. Tsay, L. F. Shyur, D. Agrawal, Y. C. Wu, & S. Y. Wang (Eds.), *Medicinal plants - Recent advances*. Singapore: Research and Development. Springer. https://doi.org/10.1007/978-981-10-1085-9_4.

Team, R. C. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Travadi, T., Shah, A. P., Pandit, R., Sharma, S., Joshi, C., & Joshi, M. (2023). A combined approach of DNA metabarcoding collectively enhances the detection efficiency of medicinal plants in single and polyherbal formulations. *Frontiers in Plant Science, 14*, Article 1169984. https://doi.org/10.3389/fpls.2023.1169984

Urumarudappa, S. K. J., Tungphatthong, C., Prombutara, P., & Sukrong, S. (2020). DNA metabarcoding to unravel plant species composition in selected herbal medicines on the National List of Essential Medicines (NLEM) of Thailand. *Scientific Reports, 10*, 18259. https://doi.org/10.1038/s41598-020-75305-0

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR Protocols: a guide to methods and applications* (pp. 315–322). Academic Press. https://doi.org/10.1016/B978-0-12-372180-8.50042-1.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One, 5*(10), Article e13102. https://doi.org/10.1371/journal.pone.0013102

Yu, J., Wu, X., Liu, C., Newmaster, S., Ragupathy, S., & Kress, W. J. (2021). Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety, 208*, Article 111691.