RAISING THE BAR

MOLECULAR IDENTIFICATION OF PLANTS

Vincent Manzanilla

University of Oslo, 2018

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1986

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Raising the bar

Molecular identification of plants

Vincent Manzanilla

Dissertation presented for the degree of

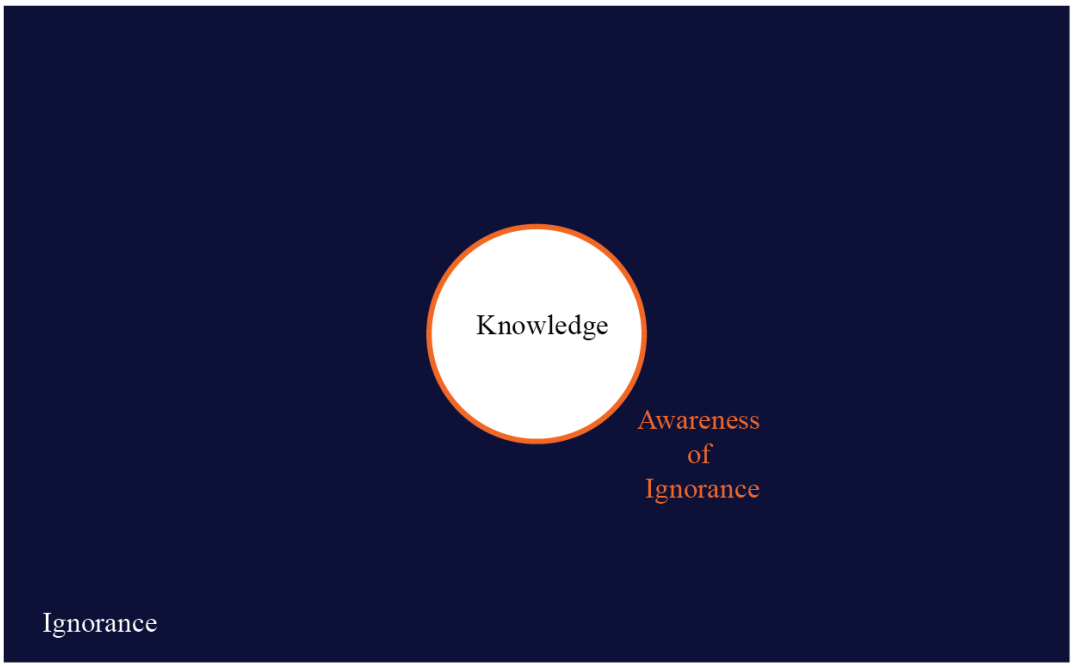
Philosophiae Doctor

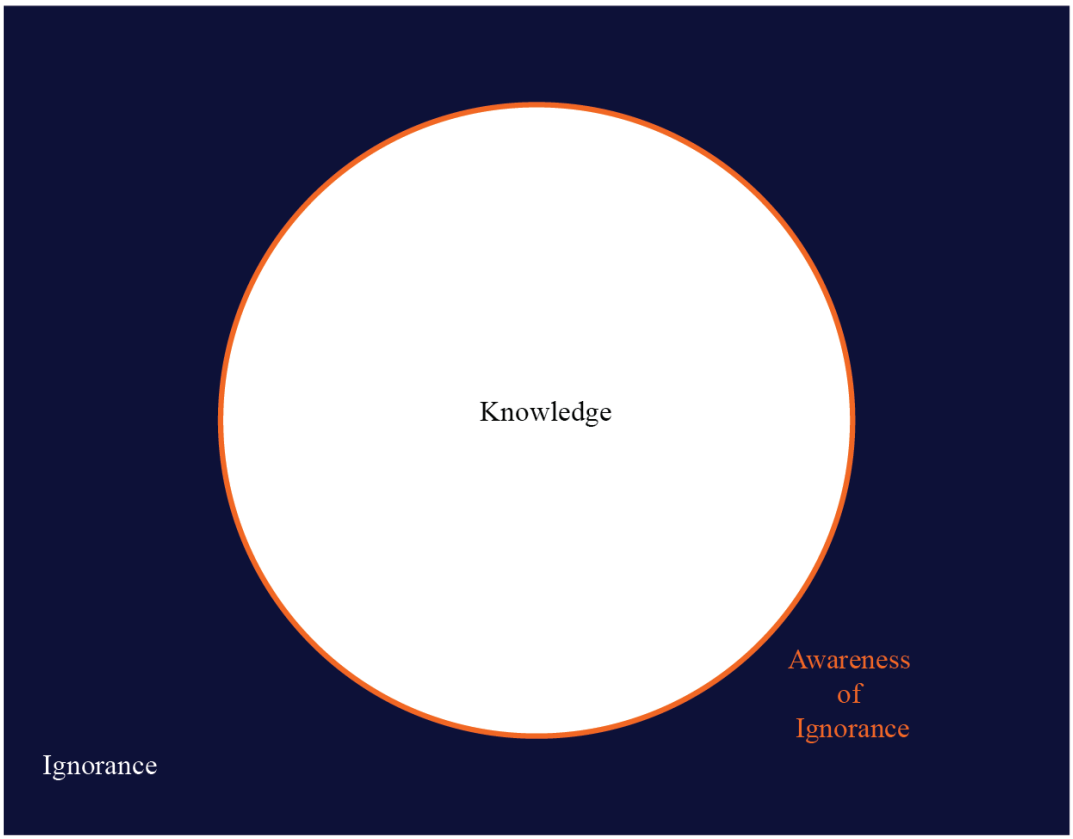
Natural History Museum

Faculty of Mathematics and Natural Sciences

University of Oslo

2018





“The greater is the circle of light, the greater is the boundary of the darkness by which it is confined.”

Priestley, Joseph. 1790. Experiments and observations on different kinds of air, and other branches of natural philosophy, connected with the subject, vol. 3, 2nd ed. Birmingham: Thomas Pearson.

SUMMARY

Plant identification using DNA markers, essential to address global issues of biodiversity conservation, safe use and legal trade, has proven to be a challenge for the past 15 years. Recent methodological developments in DNA sequencing as well as in data analyses offer new opportunities for molecular plant identification. The goal of this thesis is, first, to explore new technologies for traditional barcoding and metabarcoding, and then to develop alternative approaches to achieve better resolution for the identification of species. The new molecular identification approaches developed here take into account plant evolution allowing species identification to be anchored in an evolutionary framework.

In this thesis, I aim to identify mixed or single-ingredient traded plant products using different methodologies: from standard barcodes and plastid genomes to combinations of hundreds of nuclear genes. The thesis chapters are organised as a progression in complexity of molecular identification approaches. **Paper I** shows illegal plant trade of orchids using standard barcoding markers and distance-based methods to identify traded species. These tools are also used in **Paper II**, which aims to identify adulteration and substitution of *Hypericum perforatum* L. herbal supplements. **Paper III** focuses on ginseng, one of the oldest medicinal plants in trade worldwide. This paper uses species delimitation approaches based on the phylogenomics of plastid genomes for the identification of ginseng species. The last two chapters address the evolution and conservation of the genus *Anacyclus*. By using this recently diverged, hybrid species complex as a case study, in these two chapters I present a new method for molecular

identification successful in cases of high evolutionary complexity. First, an in-depth understanding of the evolutionary history of the genus is necessary, including evaluation of hybridisation events that often hamper traditional molecular identification. **Paper IV** presents the evolution of *Anacyclus* and tests for hybridization using hundreds of nuclear genes. Then, in **Paper V**, traded market samples from Morocco and India are used to understand the value chain of *Anacyclus pyrethrum* (L.) Lag. and the implications for conservation. Molecular identification of traded samples presented is based on a phylogenomic framework using hundreds of nuclear markers and coalescent-based multispecies methods. In conclusion, the thesis makes steps forward for molecular identification covering from evolutionary baseline analyses of hybridization to the application of these methods in biodiversity conservation and product authentication.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my official and unofficial supervisors, **Anneleen Kool** and **Hugo de Boer**, to whom I am endlessly thankful for the freedom they gave me to explore the making of science during my PhD. I am also thankful for the opportunities that they offered me by allowing me to engage in side projects and getting involved in the research environments of Oslo and Uppsala. **Gary Martin**, my third supervisor, hosted me at the Global Diversity Foundation, where I had a unique experience learning about ethnobotany.

In the field, I seriously needed the support of a local specialist with skills to facilitate social entry and resolve practical issues. **Mohamed El Haouzi** was an incredible host in Marrakech: I am thankful for his introduction to Moroccan culture as well as the logistic support during my field work. In between two exhausting trips, **Irene**'s riad was an oasis of tranquillity (despite of the fleas) where I ate the best salads and omelettes in Marrakech. Prof. **Ahmed Ouhammou** was of invaluable help in the maze of the administration in Marrakech.

Back to the North, the support team at the NHM's lab, **Audun Schrøder-Nielsen**, **Lisbeth Thorbek** and **Jarl Andreas Anmarkrud**, was extraordinary. I thank them for their patience with my impatience, their advice and dedicated work. Also, the PET group at the NHM Oslo that gathers a great crowd of scientists, humbly named "the cool kids", provided invaluable support: **Abel Gizaw Seid**, **Ancuta Cristina Raclariu**, **Aurelie Labarre**, **Brecht Verstraete**, **Christian Brochmann**, **Dimitar Stefanov Dimitrov**, **Filip Kolár**, **Lovisa Gustafsson**, **Maria Bello**, **Maxime Borry**, **Michael Nowak**, **Sarina Veldman**, **Seethapathy Gopalakrishnan Saroja**, and **Siri Birkeland**. The students at the NHM are a nice addition

to this crowd, specifically, the indispensable **José Cerca de Oliveira** and his childish sense of humour.

This PhD was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under the Grant agreement no. 606895 to the FP7-MCA-ITN MedPlant, "Phylogenetic Exploration of Medicinal Plant Diversity". I would like to thank all the “MadPlanters” for their discussions, joy, terrific sense of humour and impressive drinking habits: **Andrea Salm, Astrid Henz**, **Irene Teixidor**, **Karen Martinez**, **Madeleine Ernst**, **Maonian Xu**, **Marco Kreuzer**, **Matthias Geck**, **Natalie Iwanycki**, **Paul Wennekes, Peter Staub**, **Rosa Buonfiglio** and **Sujay Guha**.

During my PhD, I spent several months at the University of Reading, where I took time to explore my data and other crazy scientific ideas with **Marco Kreuzer**. I am thankful to **Julie Hawkins** for being an amazing host during that time. In the UK, Fisherman’s pub was a blessing after work, where we spend hours with **Marco Kreuzer**, **Andrea Marks** and **Irene Texidor** continuing the discussions on science and beyond.

I need to acknowledge the Dutch for the invention **Gin**, and the British for the **Tonic**: without it I would not have survived the seemingly endless days powered by the **Coffee Machine** (thank you **Hugo**, for this major equipment acquisition).

Thanks to my **family** for their unconditional support during these years of (not always comprehensible, but always intense) labour.

Finally, I am thankful to **Irene Texidor** with whom I share a beautiful friendship developed during the PhD, scientific interests and projects, and soon, a kid.

CONTENTS

INCLUDED PAPERS

This thesis is based on five papers. They will be referred to in the text by their Roman numerals (I-V).

I. De Boer, H.J., Ghorbani, A., **Manzanilla, V.**, Raclariu, A.C., Kreziou, A., Ounjai, S., Osathanunkul, M., Gravendeel, B. 2017. DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade. *Proc. R. Soc. B*. 284: 20171182. DOI:10.1098/rspb.2017.1182

II. Raclariu, A.C., Paltinean, R., Vlase, L., Labarre, A., **Manzanilla, V.**, Ichim, M.C., Crisan, G., Brysting, A.K., de Boer, H.J. 2017. Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding, TLC and HPLC-MS. *Sci. Rep.* 7: 1291. DOI:10.1038/s41598-017-01389-w

III. **Manzanilla, V.**, Kool, A., Nguyen Nhat, L., Nong Van, H., Le Thi Thu, H., de Boer, H.J. 2018. Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species. *BMC Evol. Biol. In press.*

IV. **Manzanilla, V.**, Ouhammou, A., Martin, G.J., Kolár, Schrøder-Nielsen, A., de Boer, H.J., Kool, A. Hybridisation and adaptive radiation of the genus *Anacyclus* L. (Anthemidae, Compositae). *In prep.*

V. **Manzanilla, V.**, Teixidor-Toneu, I., Martin, G.J., de Boer, H.J., Kool, A. Raising the bar of molecular plant identification sheds light on complex trade in red-listed species. *In prep.*

INTRODUCTION

Biodiversity identification underpins all biological studies and is key to address global issues of safe use and conservation through the authentication of traded biodiversity products. Distinguishing the millions of species that exist on Earth requires a large community of taxonomists (Thomson et al. 2018). By using morphological approaches, phenotypic and genotypic variation can hamper correct identification, cryptic taxa are overlooked and not all species can be discriminated from each other in all life stages (Taberlet et al. 2007). Genetics and more recently genomics provide powerful tools to face these 21st century challenges in systematics (Wen et al. 2015).

1. PLANT MOLECULAR IDENTIFICATION

1.1. DEFINITIONS

The term "barcoding" was first used in the early 2000s to describe molecular approaches to species identification based on DNA sequences from short and standardised genome regions (Hebert et al. 2003). This approach has revolutionised species identification methods (Hebert et al. 2003), overcoming some of the challenges of morphological and chemical identification. In animals, the mitochondrial marker COI is known to be a suitable barcode providing species-level resolution (Hebert et al. 2003), but for plants no single universal barcode region is able to distinguish all species because of the low mutation rates of the plastid genome and nrDNA (Fazekas et al. 2008). Instead, combinations of several regions of the plastid genome have been tested as barcodes for plants (Kress et al. 2005; Kress and Erickson 2007; Fazekas et al. 2008; Hollingsworth et al. 2009). In 2009,

the CBOL plant working group proposed the two-marker combination of *rbcL* and *matK* as the core barcode for land plants (CBOL et al. 2009), but their conclusions were based on a relatively small sample biased towards some specific clades (CBOL et al. 2011). Since 2009, other markers have been proposed as barcodes for plants, specifically the plastid genome region *trnH-psbA* and the nuclear ribosomal internal transcribed spacer (ITS) (Kress and Erickson 2007; Chen et al. 2010; Yao et al. 2010; Hollingsworth et al. 2011). Further studies found that the combination of ITS and any plastid marker have the highest discriminatory power for plants and it was proposed to incorporate ITS into the core barcode of plants (CBOL et al. 2011).

Following recent methodological developments, "DNA metabarcoding" is increasingly used in addition of traditional DNA barcoding. DNA metabarcoding is based on high-throughput multi-taxa sequencing technology using extracellular or total DNA extracted from complex DNA samples (Taberlet et al. 2007, 2012; Staats et al. 2016). Many DNA metabarcoding studies focusing on plants have used the P6 loop of the *trnL* intron (plastid marker), as it has high primer universality, short amplicon length and high sequence variation (Taberlet et al. 2007). The combination of these three characteristics has made the *trnL* intron P6 loop the marker of choice for ancient DNA and sediment DNA metabarcoding studies. However, Taberlet et al. (2007) do point out that the *trnL* intron (254–767 bp) has relatively low resolution at the species level, and that the P6 loop (10–143 bp) has even lower resolution.

1.2. Recent methodological developments

Technologies to sequence DNA underwent enormous improvement when next-generation sequencing (NGS) methods emerged (Shendure and Ji 2008; Glenn 2011). High-throughput sequencing (HTS) offers new

possibilities for plant molecular identification that were not available when barcoding was first proposed (Lemmon and Lemmon 2013; Coissac et al. 2016; Hollingsworth et al. 2016). Three sequencing and library preparation methods have been developed from HTS with applications for molecular identification: amplicon sequencing (Bybee et al. 2011; O'Neill et al. 2013), genome 'skimming' by shotgun sequencing (Straub et al. 2012; Malé et al. 2014) and target capture (Mamanova et al. 2010). The barcoding vocabulary had followed these recent developments of the sequencing technology and introduce at least two new terms: "Extended barcodes" (Coissac et al. 2016; Hollingsworth et al. 2016), "ultra-barcodes" (Kane et al. 2012) and "super-barcode" (Li et al. 2015). The term "extended barcode" consists of an entire organelle genome and nuclear ribosomal DNA, along with numerous single-copy nuclear genes, whereas "ultra-barcode" and super-barcode" stands only for whole organellar genomes.

Amplicon sequencing methods allow to sequence amplified traditional barcode markers in parallel, which enables the analysis of multi-species samples (Taberlet et al. 2012). The processing of large numbers of samples is furthermore made feasible and cost-effective by using uniquely tagged primers to pool several PCR products in a single sequencing run (Valentini et al. 2009; Coissac 2012). This method has quickly been adopted for molecular identification. Ultimately, amplicon sequencing generates the same type of data as traditional barcoding, hence it uses existing reference databases and bioinformatics pipelines (Lammers et al. 2014; Zepeda-Mendoza et al. 2016). Amplicon sequencing has been applied in studies of plant identification for health and safety (Raclariu et al. 2017c, 2017b; Schmiderer et al. 2017; Sgamma et al. 2017), illegal trade monitoring and biodiversity conservation (Sgamma et al. 2017; Veldman et al. 2017), as well as ecology and biodiversity (Willerslev et al. 2014; Taberlet et al. 2018; Zobel et al. 2018).

Shotgun sequencing yields low coverage genomic data, enabling an approach aptly termed ‘genome skimming’, which is mainly used to retrieve plastid and mitochondrial genomes, although it can also be used to retrieve traditional nrDNA markers. This method generates three to six millions of reads per sample and allows multiplexing of libraries from hundreds of individuals. The bioinformatics workflow needed for data analysis is well described today (Hahn et al. 2013; Dierckxsens et al. 2016; Ankenbrand et al. 2018). This method is overall more expensive than amplicon sequencing, but plant identification at species level is more efficient and accurate (Parks et al. 2009; Nock et al. 2011; Kane et al. 2012; Ruhsam et al. 2015; Manzanilla et al. 2018). However, extensive plastome or mitochondrial genome reference databases are not yet available. Large scale genome skimming projects are underway (Coissac et al. 2015; NorBOL 2018), but this method has barely been applied for plant identification.

Both traditional barcodes and whole plastome molecular methods are able to identify plants at species level. However, plants from recently diverged groups, those that result from hybridisation events or that belong to introgressed clades with extensive incomplete lineage sorting (ILS) are not easily identified by single markers, combinations of a few markers or whole plastome (Hollingsworth 2011; Coissac et al. 2016; Hollingsworth et al. 2016). Recent speciation, hybridisation and introgression with extensive ILS are all frequent evolutionary scenarios among plants (Wood et al. 2009; Soltis et al. 2015), and traditional barcoding approaches cannot accommodate such cases of complex evolution.

Target capture methods can overcome these challenges by simultaneously sequencing hundreds of low-copy nuclear loci using RNA baits (Mamanova et al. 2010). Target capture can successfully sequence

degraded DNA similar to shotgun sequencing (Särkinen et al. 2012; Staats et al. 2013), which is common in traded plant material used in conservation and pharmacovigilance studies. Along with targeted low-copy nuclear genes, traditional barcodes can be retrieved with this method (Bybee et al. 2011; O'Neill et al. 2013; Weitemier et al. 2014; Schmickl et al. 2016)(Särkinen et al. 2012; Staats et al. 2013)Several bait sets targeting low-copy nuclear genes have been created in recent years for *Asclepia* (Williams et al. 2016), *Fragaria* (Kamneva et al. 2017) and *Oxalis* (Schmickl et al. 2016) amongst other plant groups. Target capture has been advocated as a powerful tool for molecular identification (Pillon et al. 2013; Ruhsam et al. 2015; Coissac et al. 2016), but up till now it had not yet been applied to identify real samples.

2. EVOLUTIONARY CONSIDERATIONS FOR PLANT MOLECULAR IDENTIFICATION

Polyploidy is a common characteristic of vascular plants, and there is substantial evidence that most, if not all, plant species have polyploid ancestry (Soltis and Soltis 2009; Wood et al. 2009; Soltis et al. 2015). Allopolyploids arise from interspecific hybridization and doubling of non-homologous genomes and frequently result in plant speciation. Examples among crops of major economic importance, are wheat (Marcussen et al. 2014) and cotton (Paterson et al. 2012). This reticulated nature of the evolution of plants impedes standard molecular identification of plants.

A large body of empirical data suggests that homoploid hybridization is common in plants yet large controversy remains how frequently homoploid hybridization results in speciation. Schumer et al. (2014) identify homoploid speciation when three criteria are met: (1) reproductive isolation

is strong between the parents and the hybrid species, (2) there is genetic evidence of hybridization, and (3) an isolating mechanism derived from hybridization itself exists. However, these criteria are not always present. Several examples support the idea that reproductive isolation is not necessary for homoploid speciation (Feliner et al. 2017; Thompson et al. 2017). In addition, under homoploid speciation models simulations based on sympatric populations that have low genetic isolation support the possibility of homoploid speciation with a more relaxed reproductive isolation (Buerkle et al. 2000; Seehausen 2004; Feliner et al. 2017).

Hybridization between two species may initiate an adaptive radiation event by providing new genetic variation (Meier et al. 2017). Specifically, hybridization between closely related lineages can generate genotypes that allow to reach fitness peaks that were previously unoccupied, what is known as the syngameon hypothesis (Lotsy 1931; Stanford 1995; Seehausen 2004; Boecklen 2017). Such events can be common in cases of secondary contact of allopatric lineages, and selection against hybrids may be weak during colonization of new niches. When this coincides with new ecological opportunities, hybrid swarms can accelerate adaptive radiation. For plant groups that underwent rapid speciation, species relationships can be obscured by ancestral polymorphisms retained as a consequence of incomplete lineage sorting (ILS) (Maddison 1997).

Polyploidization, hybridization and rapid speciation blur the evolutionary delimitation of plant species (i.e., plants are not discrete genetic entities). Single or multi-locus barcodes from a single origin (i.e., the plastome) do not reflect plants' reticulated evolutionary histories, and have limited application for molecular identification (Fazekas et al. 2009). On average, the resolution of barcodes combining plastome and nrDNA markers can discriminate 70% of plant species (Rieseberg et al. 2006; Kress and

Erickson 2007; Fazekas et al. 2008, 2009). Since hybridization, polyploidization and rapid speciation are challenging for accurate plant molecular identification, identification methods should take these evolutionary events into account.

3. MEDICINAL PLANTS IN TRADE

At least 30,000 plants have some recorded use, and more than half of these are documented as medicines (Royal Botanic Gardens Kew 2016). However, in total it is estimated that up to 50,000 plants could be used as medicine globally, representing 20% of the world's vascular flora (Schippmann et al. 2002; Hamilton 2004). Plants are used as traditional medicines by cultures across the world and they are also important raw materials for the pharmaceutical, perfume and cosmetic industries. The demand for medicinal plants is expected to continue to grow. In 2006, the herbal medicine industry was valued in $14 billion, but its value could reach, according to the WHO, $5 trillion by 2050 (Booker et al. 2012). Most of these plant resources continue to be harvested from the wild, and their trade is an essential element of the livelihoods of harvesters who are mostly in developing countries (Schippmann et al. 2002; Hamilton et al. 2006). Collection for commercial trade is an overwhelming conservation problem (Hamilton 2003). Medicinal plants are easily traded nationally or across borders outside the CITES regulation (CITES and Medicinal Plants 2018) and many are threatened with extinction not only due to overharvesting and illegal international trade, but also habitat loss and climate change (Hawkins 2008). However, the trade of medicinal plants does not only raise conservation concerns, but also important health and safety issues (de Boer et al. 2015).

Quality and identification are a concern for medicinal plants traded as raw materials, processed herbal medicines and food supplements (Booker et al. 2014; de Boer et al. 2017; Raclariu et al. 2017c, 2017b). Quality issues arise from the deliberate addition of adulterants to increase product efficacy or business revenues (Raclariu et al. 2017a and references therein), from plant misidentification along trade chains to stocking practices that lead to the contamination of the product (Booker et al. 2016). Many plants are traded as powder, processed in various ways or mixed with other ingredients. All these factors hamper the identification of traded medicinal plants by their morphology (Veldman et al. 2014a; de Boer et al. 2015, 2017; Ghorbani et al. 2017; Raclariu et al. 2017c, 2017b). Moreover, vernacular names do not easily identify plants at species level due to ethnotaxonomic challenges (Berlin, Breedlove et al. 1973) and because names may change along the value chain (Otieno et al. 2015). Even traders may not able to accurately identify traded materials (Kool et al. 2012; Ouarghidi et al. 2012, 2013; de Boer et al. 2014).

So far, the identification and authentication of both raw medicinal plant materials and final herbal products relies on chemical analyses that detect species-specific compounds (European Medicines Agency (EMA) 2011; World Health Organization 2011). In Europe, the European Pharmacopoeia is responsible for herbals' quality control and bases its assessment recommendations on both morphological and chemical analyses (EDQM 2018). However, complex species chemistry as well as presence of many species in a product challenge the identification and authentication of species-specific target compounds (Khan and Smillie 2012). International conventions including the Convention on Biological Diversity (CBD) and the Convention on International Trade in Endangered Species (CITES) are increasingly putting in place effective strategies for the conservation and sustainable use of medicinal plants, as well as the regulation of their trade.

For example, botanical gardens have developed conservation plans based on the Global Strategy for Plant Conservation (GSPC) and organisations such as FairWild promote the sustainable collection of wild ingredients (World Health Organization 2011; FairWild 2018). Guidelines for monitoring safety of use of herbal medicines in pharmacovigilance systems are also in place (WHO 2004), but implementing these is challenged by accurate identification. Of particular interest to monitor in trade are plants for which economical value fuels both adulteration and overharvesting. Plants for which the roots and other underground parts are used are especially threatened by overharvesting (Schippmann et al. 2002; Hamilton 2004; Ticktin 2004). Roots are particularly challenging to identify and are easily adulterated (Kool et al. 2012; Ouarghidi et al. 2012, 2013; de Boer et al. 2014; Ghorbani et al. 2017). This is the case for four out of five of the plants or groups of plants presented in detail below.

3.1. ORCHIDS: HIGHLY DIVERSE AND WIDELY THREATENED

Orchidaceae is the most diverse and widespread family of all vascular plants with about 25,000 species in more than 750 genera (Christenhusz and Byng 2016). Many orchid species are used in traditional preparations, especially as food, medicines and dietary supplements (Figure 1) (Bulpitt 2005; Bulpitt et al. 2007; Chinsamy et al. 2011; Hossain 2011). For example, the traditional snack *chikanda* prepared in Tanzania, Zambia and Malawi is made from processed *Disa*, *Satyrium* and *Habenaria* tubers (Veldman et al. 2014a), and in the eastern Mediterranean, *salep*, made from dried orchid tubers, is used to prepare a warming drink in winter and ice cream in summer (Figure 1) (Kasparek and Grimm 1999; Ece Tamer et al. 2006; Starin 2012). These tubers represent an important source of nutrients for their consumers (Arditti 1992). An increased popularity of orchid-based food products and traditional medicines results in higher harvesting pressure on wild orchid populations and poses conservation issues (Ghorbani et al. 2014; Veldman et al. 2014b; de Boer et al. 2017).

Together with overharvesting, habitat loss is a threat to orchid diversity conservation, as most orchids have very limited areas of distribution. Monitoring trade is challenging because the plant part used are the tubers, which have no morphological characteristics that would allow species-level macroscopic or microscopic identification (Figure 1). Molecular identification is starting to be used as a successful identification tool to identify orchid products in trade (Ghorbani et al. 2014, 2016; Veldman et al. 2014b, 2017; de Boer et al. 2017).

Figure 1. (a) Plants of Orchis simia *thrown away after harvesting fresh tubers (Ghorbani et al. 2014); (b) Tuber samples of different morphology purchased from the markets. Samples of* Orchis/Anacamptis *type tubers; (c) Samples of* Dactylorhiza *type tubers (Ghorbani et al. 2017); Individual dose of (d)* salep *powder and (e)* salep *drink.*

3.2. A TOP-SELLING HERB: *HYPERICUM PERFORATUM* L.

Hypericum perforatum L. is a medicinal plant native to parts of Europe and Asia, with use in European herbal medicine documented since the time of Dioscorides (c. 40-90 AD; Figure 2) (Robert 1993; De Vos 2010). Traditionally used for a broad range of ailments, today it has become one of the most popular herbal remedies in complementary and alternative medicine to treat mild and moderate depressions (Figure 2) (Linde 2009). *H. perforatum* is one of the best investigated medicinal plants from a pharmacological perspective and its pharmacological applications are still being developed (Galeotti 2017). This species is among the top-selling herbs in Europe and is sold over-the-counter in pharmacies, supermarkets, health shops as well as online (Borrelli and Izzo 2009). Its consumption has increased dramatically in recent years to become one of the most highly demanded medicinal plants (Galeotti 2017).

H. perforatum is commonly found in temperate regions across the world. In North America, it has spread as an invasive weed and in the last decades its cultivation has gradually expanded in Western Europe (Büter et al. 1998). Although no conservation threats to *H. perforatum* are known to date, the low quality of processed herbal products is alarming (Raclariu et al. 2017c). The highly competitive market of herbal products, together with the lack of standardised methods for quality assessment incentives, fuels the use of substitutes and unlabelled fillers (Coghlan et al. 2012; Newmaster et al. 2013; de Boer et al. 2015).

Figure 2. (a) Hypericum perforatum *L.; (b)* Hypericum perforatum *herbal supplement; and, (c) grounded* Hypericum *sp.*

3.3. AN OLD AND BOOMING MARKET: GINSENG (*PANAX* L.)

Ginseng is one of the most popular traditional Chinese medicinal herbs. *Panax ginseng* C.A. Mey. and other *Panax* species have been used in Asia for thousands of years (Robbins 1998). In Asia, it is considered a powerful tonic of the vital energy *qi* that restores the pulse, improves weak conditions, benefits several internal organs and calms the mind (Jaiswal et al. 2016). In Europe and America, it is used in complementary and alternative medicine. The roots of ginseng are used, and this complicates sustainable use, as the whole plant is damaged during harvesting (Schippmann et al. 2002; Hamilton 2004; Ticktin 2004).

An increase in the demand for ginseng in the 18th century almost drove the wild *P. ginseng* populations to extinction and triggered the collection and trade of American ginseng (*P. quinquefolius* L.), which subsequently decimated its wild populations in North America (Millspaugh 1892; Kimmens 1975). Together with *P. ginseng* and *P. quinquefolius*, many other Asian ginseng species are currently endangered, including at least *P. assamicus* R.N. Banerjee, *P. japonicas* (T.Nees) C.A.Mey., and *P. pseudoginseng* Wall. (Joshi et al. 1991; Jain 1994; Basnet and Dey 2008; Zhang et al. 2015).

Both *P. ginseng* and *P. quinquefolius* are now widely cultivated in Asia and America, respectively. Although differences in ginsenoside content between wild and cultivated plants are insignificant (Schlag and McIntosh 2006), pressure on wild populations still exists because consumers often prefer wild-harvested ginseng as it is considered to be more potent (Booker et al. 2015). Between 1999 and 2009, the price for wild or semi-wild ginseng doubled whilst the price for the cultivated crop dropped by 75% (Booker et al. 2015). Despite global threats of overharvesting of wild ginseng populations, and the fact that several other *Panax* species are

harvested from the wild in Asia (Booker et al. 2015), only the international trade of *P. ginseng* and *P. quinquefolius* is to some extent regulated through CITES (CITES 2016). The only *Panax* species with an IUCN Red-List assessment is *P. zingiberensis* C.Y.Wu & Feng, and this has been assessed as Endangered (EN) due to habitat loss and indiscriminate overharvesting (China Plant Specialist Group 2004). As with orchids, the absence of morphological characters for unambiguous species identification of the traded parts, makes it impossible to assess the conservation and harvest pressure on different *Panax* species (Figure 3). A myriad of analytical methods, including chemical, pharmacological and molecular, have been developed to identify samples in trade, but these have either little resolution at species level or have not included sufficient sampling to assess this. Molecular identification is the most promising approach for *Panax* species identification (Komatsu et al. 2001; Shi et al. 2015), but more data and novel methods are needed to find suitable markers.

Figure 3. Ginseng in a Corean market stall.

3.4. FROM WEST TO EAST: THE INTERNATIONAL TRADE OF THE ATLAS DAISY, *ANACYCLUS PYRETHRUM* (L.) LAG.

Anacyclus pyrethrum (L.) Lag. is one of few herbal remedies currently traded from West to East. The plant is an endemic of north-eastern Africa and southern Spain (Humphries 1979; Rosato et al. 2017) and is an important ingredient in Ayurvedic medicine (Pittle 2005). Hence, it is traded internationally mostly from north-eastern Africa to India (Ghosh and General 2013; Jaiswal et al. 2016) and Nepal (Tiwari et al. 2004), where it is used mostly for dental care preparations (Ministry of Health and Family Welfare 2007). *A. pyrethrum* has also been used historically in Islamo-Arabic and European medicine (Pittle 2005; De Vos 2010; Adams et al. 2011; Staub et al. 2016) and continues to be used nowadays as a potent painkiller and to treat inflammations in North Africa and the Middle East (Merzouki et al. 2000; Pittle 2005; Ouarghidi et al. 2012, 2013; Jamila and Mostafa 2014; Rhafouri et al. 2014; Benarba et al. 2015; Benarba 2016; Ouelbani et al. 2016). Similar to orchids and ginseng, underground parts are traded, and this complicates conservation as well as identification in trade (Figure 4). *A. pyrethrum* is overharvested in Morocco (Ouarghidi et al. 2012, 2017; Rhafouri et al. 2014; Taleb 2017) and international demand is likely the most important factor driving the decline of wild populations. *A. pyrethrum* is increasingly unavailable locally in Moroccan markets (Ouarghidi et al. 2012, 2013), but paradoxically, it can be purchased online, mostly from Asian companies.

Moroccan harvesters (Figure 4) can distinguish the two varieties of *A. pyrethrum* (var. *pyrethrum* and var. *depressus*) and refer to them with different vernacular names (*iguendez* and *tiguendizt*, respectively). *A. pyrethrum* var. *pyrethrum* is considered more potent and can be ten times more expensive than var. *depressus* (Ouarghidi et al. 2012); V.M. *pers.*

obs.). However, while collectors can successfully identify this plant to the variety level, the species is misidentified by other stakeholders including middlemen and herbalists in its regional and national value chain, and it is likely adulterated in Moroccan trade (Kool et al. 2012; Ouarghidi et al. 2012, 2013; de Boer et al. 2014).

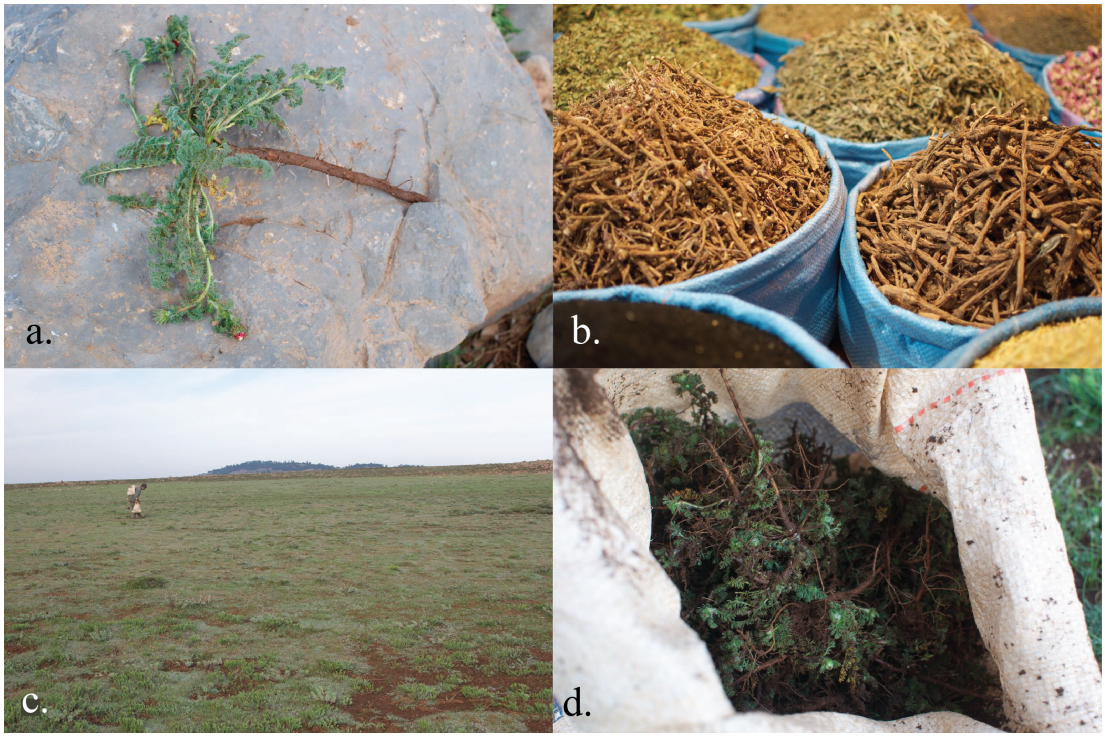


Figure 4. (a) Specimen of A. pyrethrum *var.* pyrethrum *showing the medicinal root; (b) two baskets of* A. pyrethrum *at an herbalist's market stall in Meknes; (c)* A. pyrethrum *collector in the High Atlas; and, (d) bag of the collector with only* A. pyrethrum.

4. AIM AND OBJECTIVES OF THIS THESIS

The aim of this thesis is to apply and develop plant molecular identification methods to successfully identify plants regardless of the complexity of their evolutionary history.

Papers I and **II** use metabarcoding methods to identify mixtures of plants with traditional barcode markers. **Paper I** evaluates sequencing barcoding marker efficacy and investigates species diversity in *salep* (an orchid-based food product) by identifying its common species, adulterants and substitutes. This paper highlights the prevalence of endangered species in *salep*. **Paper II** focuses on herbal supplements that contain *Hypericum perforatum* and assesses the efficacy of amplicon metabarcoding compared to HPLC-MS and TLC methods to detect possible adulteration and substitution in herbal medicines. **Paper III** first provides new insights into the evolutionary history of the *Panax* genus from the full plastid genome phylogeny. Based on this phylogeny, a species delimitation approach is used to evaluate the discrimination power of selected plastid markers. **Paper IV** focuses on the hybridization and adaptive radiation of the genus *Anacyclus* using hundreds of low-copy nuclear markers. Based on this evolutionary framework and on hundreds of low-copy nuclear markers, **Paper V** identifies internationally traded *Anacyclus pyrethrum* samples and evaluates the plant's value chain. Importantly, this paper compares nrDNA, plastid genome and low-copy nuclear marker approaches in discriminating species with complex evolutionary history.

MATERIALS AND METHODS

1. SAMPLE COLLECTION

1.1. *SALEP*, AN ORCHID-BASED PRODUCT (**PAPER I**)

Fifty-five processed *salep* samples were purchased from supermarkets, herbal stores, pharmacies, and markets in Iran (n=19), Germany (n=15), Greece (n=12) and Turkey (n=9) to represent the commercially available *salep* products as well as its different producers and vendors. Traded *salep* samples included bulk powder (n=29), packed commercial powders (n=23), processed beverages (n=2) and ice cream (n=1). Four of the labelled products claimed to contain only *salep* flavouring, whereas the rest were sold as genuine *salep*.

1.2. *HYPERICUM PERFORATUM* (**PAPER II**)

In total, 77 herbal products labelled as *Hypericum perforatum* were acquired from pharmacies (n=44), herbal shops (n=25), supermarkets (n=2) or via e-commerce (n=7), in Romania (n=51), Germany (n=5), Poland (n=4), Turkey (n=4), Slovakia (n=3), Spain (n=2), UK (n=2), Austria (n=2), Czech Republic (n=1), France (n=1), Italy (n=1), Sweden (n=1) and the Netherlands (n=1). Thirty-eight samples were single ingredient products, 33 contained between two and ten ingredients, and seven products contained more than ten ingredients, according to the information presented in the products' label. A range of herbal teas (n=44), capsules (n=15), tablets (n=14) and extracts (n=5) were bought.

1.3. *PANAX* SPP. (**PAPER III**)

Fresh material of *Panax bipinnatifidus*, *P. stipuleanatus*, and *P. vietnamensis* (n=2) was collected in Vietnam to complement plastid

genomes from open data repositories for 57 selected plastid genomes from across the Araliaceae family. Minimally, two individuals or species were selected per genus across the Araliaceae family. Thirty-eight *Panax* sp. plastid genomes were sampled in total, representing eight of the 12 accepted *Panax* species (The Plant List 2018). *Hydrocotyle verticillata* was used as outgroup due to its early divergence within the family.

1.4. *ANACYCLUS* SP. (**PAPERS IV** AND **V**)

In total, 65 vouchered samples of *Anacyclus* sp. were acquired. Fifty nine were collected in Morocco and Spain from wild populations of *Anacyclus* and identified at the NHM Oslo. Eleven herbarium voucher specimens were acquired from internationally registered herbaria for those species occurring elsewhere in the Mediterranean. Vouchers or living collection specimens were collected for two species of *Matricaria*, two species of *Achillea*, one of *Otanthus* and two of *Heliocauta* to be used as outgroups.

Fifty trade samples consisting of 100 g of roots were bought in Morocco and India. Using the local vernacular names for the two *A. pyrethrum* varieties (Ouarghidi et al. 2012) in Morocco, *tiguendizt* and *iguendez* roots were acquired from collectors, middle men, whole sellers, export companies, and herbal shops. The two varieties are not distinguished in India, and the local name *akarkara* (Ved and Goraya 2007) was used to buy products from herbal shops. In Morocco, semi-structured interviews were conducted following the International Society of Ethnobiology Code of Ethics (2018) with thirty-nine vendors from whom samples were purchased to enquire about various aspects of the trade of *Anacyclus*. The total quantity of *Anacyclus* in herbalists' shops was weighted and estimates were provided by interviewed employees for export companies.

2. AMPLICON SEQUENCING, SHOTGUN SEQUENCING AND TARGET CAPTURE

2.1. AMPLICON SEQUENCING (PAPERS I AND II)

Total DNA was extracted from the *salep* samples and *Hypericum perforatum* herbal products using the CTAB protocol (Doyle 1987) and together with extraction blanks. Primers pairs for the plant-specific nrITS1 and nrITS2 markers were used to amplify the extracted DNA (Sun et al. 1994). To determine the suitability of the primer pairs in amplifying the target orchid species as well as common expected adulterants, in-silico amplification with EcoPCR (Ficetola et al. 2010) of GenBank nrITS data was used. nrITS amplicons were sequenced on an Ion-Torrent Personal Genome Machine with Ion 316 v2 Chips.

2.2. LIBRARY PREPARATION WITH METHYLATION ENRICHMENT (PAPER III)

A new plastid enrichment method was applied to improve the shotgun sequencing efficacy, exploiting the low methylation of the organellar genomes compared to the nuclear genome (Feng et al. 2010). This method uses the methyl-CpG-binding domain (MBD2) to partition fragments of genomic DNA into a methylation-poor fraction (enriched for plastid and mitochondrial DNA) and a methylation-rich fraction (depleted in organellar DNA) (Yigit et al. 2014). This method has the advantage of useing a small quantity of dry material (below 40 mg) and is suitable for non-model organisms. We used a NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, Massachusetts, USA) with IgG1 fused to the human methyl-CpG-binding domain (together "MBD2-Fc") to separate a methyl-CpG-enriched fraction from a bead-associated element from a methyl-depleted fraction in the supernatant. About 400 ng template

DNA extract was used per sample with the DNeasy Plant Mini Kit (Qiagen) and the manufacturers recommendations were respected with the following modifications. The non-methylated DNA fractions were purified using 0.9X AMpure XP beads (Beckman Coulter, Brea, CA, USA) and eluted in 40 µl 1X TE buffer. To capture the methylated DNA, we followed the manufacturer's protocol. The DNA was sheared to ~400 bp fragments using a M220 Focused Ultrasonicator (Covaris Inc., Woburn, MA, USA) and microTUBES-50 (Covaris Inc.). We used the NEBNext Fast DNA Library Prep Set for Ion Torrent (NEB) for end repair and adapter ligation of the sheared DNA. The samples were indexed using the IonXpress Barcode Adapter kit (ThermoFischer, Waltham, MA, USA). For each of the four samples both fractions, methyl-CpG-enriched and methyl-CpG-depleted, were indexed and sequenced. After adapter ligation, the four methyl-CpG-enriched fractions were pooled in one library and the four methyl-CpG-depleted fractions were pooled in another library. The adapter-ligated libraries were size selected (450-540 bp) using a BluePippin (Sage Science, Beverly, MA, USA), and subsequently amplified using the NEBNext Fast DNA Library Prep Set for Ion Torrent kit using 12 PCR cycles.

2.3. LIBRARY PREPARATION FOR SHOTGUN SEQUENCING AND TARGET ENRICHMENT (**PAPERS IV** AND **V**)

DNA from *Anacyclus* reference and traded samples were extracted from approximately 40 mg of dry leaf or root material using the DNeasy Plant Mini Kit (Qiagen). Material from all 72 fieldwork and herbarium reference samples were used, as well as 110 selected individual roots from the 65 trade samples. Total DNA (0.2-1.0 µg) was sheared to 500 bp fragments using a Covaris S220 sonicator (Woburn, MA, USA) and dual indexed libraries were prepared using the Meyer and Kircher protocol (Meyer and

Kircher 2010). For the shotgun sequencing dataset, we normalized 149 libraries and sequenced them on one lane on the HiSeq 3000.

2.4. TARGET CAPTURE (**PAPERS IV** AND **V**)

2.4.1 SKIMMING DATA

Low-copy nuclear markers (600-1000 bp in length) were identified using the Hyb-Seq pipeline based on the skimming assembly of *A. radiatus* subsp. *radiatus* (MV54) and the transcriptome assembly of a close relative outgroup, *Matricaria matricarioides* (Less.) Porter (voucher 132745) (Matasci et al. 2014). The original Hyb-Seq pipeline was adapted to identify introns as well as exons (Schmickl et al. 2016).

The total DNA of *A. radiatus* subsp. *radiatus* (MV54) was sequenced on an Illumina NextSeq 500 with a paired-end library using a TruSeq DNA PCR-Free library kit. After sequencing, library adapter sequences and low quality reads were removed with Trimmomatic v. 0.32 (Bolger et al. 2014) with a quality threshold set at Q20 with a sliding window of 10 bp.

Prior to the *denovo* nuclear genome assembly, the plastid genome was assembled with the trimmed reads using MITObim v. 1.8 (Hahn et al. 2013) with the plastid genome of *Chrysanthemum indicum* L. (NC_020320) as a reference. Protein-coding genes in the chloroplast genome were annotated with DOGMA (Wyman et al. 2004), and after visual inspection, its gene map was drawn using OGDRAW v. 1.2 (Organellar Genome Draw (Lohse et al. 2007). The annotated plastid genome is deposited on NCBI. Prior to *denovo* nuclear genome assembly, we removed the reads belonging to the organelle genomes and the nrDNA using BWA v. 0.7.5a (Langmead and Salzberg 2012). The previously assembled plastid and mitochondrial genomes of *Helianthus annuus* L. (NC_023337.1) and the nrDNA from *Anacyclus valentinus* (GU818490) were used as references. The nuclear

genome of *A. radiatus* subsp. *radiatus* was assembled using SOAPdenovo2 v. r223 (Xie et al. 2014) with nine kmer values between 20 and 100. We evaluated the best genome assembly with Quast v. 2.3 (Gurevich et al. 2013).

2.4.2 Low copy Nuclear marker design and sequencing

Transcriptome and skimming data were pre-processed to ensure selection of sufficiently long nuclear regions as markers using the Hyb-Seq pipeline by filtering out plastid and mitochondrial sequences using *Helianthus annuus* NC_023337.1, as well as nrDNA using the *A. radiatus* subsp. *radiatus* assembly. Subsequently a length threshold was applied and transcripts below 120 bp (RNA probe size) and contigs from the skimming data below 600 bp were discarded. Subsequently, the contigs were mapped against the *M. matricarioides* transcriptomes using Blat v. 3.5 (Kent 2002), and alignments were selected with a minimum length of 80% of the contig size. Alignments with more than 10% divergence and contigs with more than one match against the *M. matricarioides* transcriptomes were discarded. The obtained preliminary set of markers was mapped with Burrows-Wheeler Aligner (BWA) version 0.7.5a-r405 (Li and Durbin 2010) against the reads from the *A. radiatus* nuclear genome assembly. We extracted the coverage from this alignment using BEDtools v. 2.17 (Quinlan 2014), and contigs with a higher coverage than average were discarded because they were suspected to be multiple copy genes or contain transposable elements. A total of 872 putative low-copy nuclear markers were retained for which RNA probes were ordered from Arbor Bioscience (Ann Arbor, Michigan, USA). The baits were designed to cover each base per marker four times (i.e., 4x tiling). To ensure that the probes targeted only the nuclear genome, we mapped the probes against the previously

assembled organelles and nrDNA with BWA and discarded those that matched.

We prepared twelve equimolar pools of libraries for target capture enrichment with ten to 24 samples and an average 300 ng of input DNA per pool. The RNA probes were hybridized for 16 hours before target baiting, and 14 PCR cycles were carried out after enrichment following the MyBates v.3 manual. A calculation based on previous studies was used to estimate the sequencing coverage for the targeted loci and the plastomes. The equimolar pooled and enriched libraries were sequenced with 150 bp paired-end reads on a single Illumina HiSeq 3000 lane.

3. BIOINFORMATICS

3.1. DNA METABARCODING (PAPERS I AND II)

For **Paper I** and **Paper II**, the FASTQ read files from the amplicon sequencing runs were processed using the HTS-barcode-checker pipeline (Lammers et al. 2014) available as a Galaxy pipeline at the Naturalis Biodiversity Centre (http://145.136.240.164:8080/). PRINSEQ (Schmieder and Edwards 2011) was used to inspect read lengths, Phred base qualities and mean quality scores. Reads were selected with a minimum length of 300 bp in order to filter out short reads below the target amplicon length. Reads were trimmed to a maximum length of 360 bp as base quality scores dropped sharply beyond that point. Reads with mean Phred quality scores below 25 were filtered to avoid selecting reads with errors or poor base calling. CD-HIT-EST (Li and Godzik 2006) was used to cluster reads into molecular operational taxonomic units (MOTUs) defined by a sequence similarity of more than 99% and a minimum number of two reads. The consensus sequences of non-singleton MOTUs were queried using

BLAST+ (Camacho et al. 2009) against a local copy of the NCBI/GenBank nucleotide data- base, with a maximum e-value of 0.05, a minimum hit length of 100 bp and sequence identity of more than 97%.

3.2. *PANAX* SPP. PLASTID GENOMES AND METHYLATION ENRICHMENT (**PAPER III**)

Sequencing reads were demultiplexed into FASTQ files using Flexbar version 3.0.3. Trimmomatic version 0.36 (Bolger et al. 2014) was used for adapter trimming and quality filtering of reads using a sliding window of 15 bp and an average Phred threshold of 20. Low-end quality bases below a Phred score of 20 were removed, and only reads longer than 100 bp were retained. For each enriched library, MITOBim version 1.7 (Hahn et al. 2013) was used for assembly of the single-end Ion Torrent reads using iterative mapping with in silico baiting using the following reference plastomes, *P. vietnamensis* (KP036470) and *P. stipuleanatus* (KX247147).

Inverted repeats and ambiguous portions of the assemblies were resequenced using Sanger sequencing. Specific primers were designed and used for DNA amplification of interest regions and sequenced on a Sanger sequencing machine.

In order to test the efficacy of the NEBNext Microbiome DNA Enrichment Kit, the proportion of reads belonging to the plastome was estimated for both the methylated and the non-methylated fraction. The *P. ginseng* whole genome sequencing SRR19873 experiment was used to estimate the starting proportion of plastome reads, by mapping the reads against the plastid genome of *P. ginseng* (NC_006290) using Bowtie 2 (Langmead and Salzberg 2012). Association of reads to their taxonomic identification and organelles, was made using a tailored database of *Panax* plastome data representing the same data as that downloaded from public repositories for the phylogenetic analyses. For mitochondrial data, all angiosperm

mitochondrion genomes available on NCBI were used, and for the microbiome all remaining reads were blasted against the full NCBI database. Taxonomic identifications were retrieved using the lowest common ancestor (LCP) algorithm in Megan version 5.11.3, with minimum read lengths of 150 bp and at least 10 reads for each taxon identified with an e-value of 1e-20 or less. The proportion of plastid DNA in the gDNA was estimated using Bowtie2 by mapping the proportion of reads belonging to the plastid genome for *P. ginseng* (following SRR experiment SRR1181600).

Plastid genomes were annotated using Geneious version 6.1, and annotations of exons and introns were manually checked by alignment with their respective genes in the same annotated species genome. Representative maps of the chloroplast genomes were created using OGDraw (Lohse et al. 2007).

3.3. Retrieving Plastome, nrDNA and low copy nuclear genes from raw sequence data (**Papers IV** and **V**)

3.3.1 Nuclear loci dataset

The workflow to retrieve the low-copy nuclear genes dataset is described in Figure 5. We first trimmed and quality filtered the data with Trimmomatic (Bolger et al. 2014) using a sliding window of 10 bp for an average of Phred score of Q20 with a minimum length of 100 bp. For each sample, the quality of the reads was assessed using FastQC (Andrews 2010) and individual samples were plotted with MultiQC (Ewels et al. 2016) to visualize different quality indices. Low-copy nuclear markers and their alleles were retrieved for each sequenced sample using the following steps. First, the reads from the 149 enriched samples were mapped using BWA *mem* (Li and Durbin 2010) with a mismatch penalty of 10, a gap penalty of 20, a minimum seed length of 50, and a clipping penalty of 25. A minimum

of threshold for mapping quality score was set to 40 using SAMtools (Li et al. 2009). Duplicate reads were removed using Picard version 2.17.6 (Wysoker et al. 2015). Alleles were phased for each marker and individual using SAMtools *phase* and BFCtools version 1.1 and VCFtools version 0.1.13 (Danecek et al. 2011; Narasimhan et al. 2016), and saved as separate FASTQ files for each of the two alleles per marker and individual. These files were converted into FASTA using seqtk (Li 2012). Allele indels were masked using BEDtools by generating *bed* files in which positions with zero depth were masked. The final script creates single gene matrices with all the allele sequences for each sample.

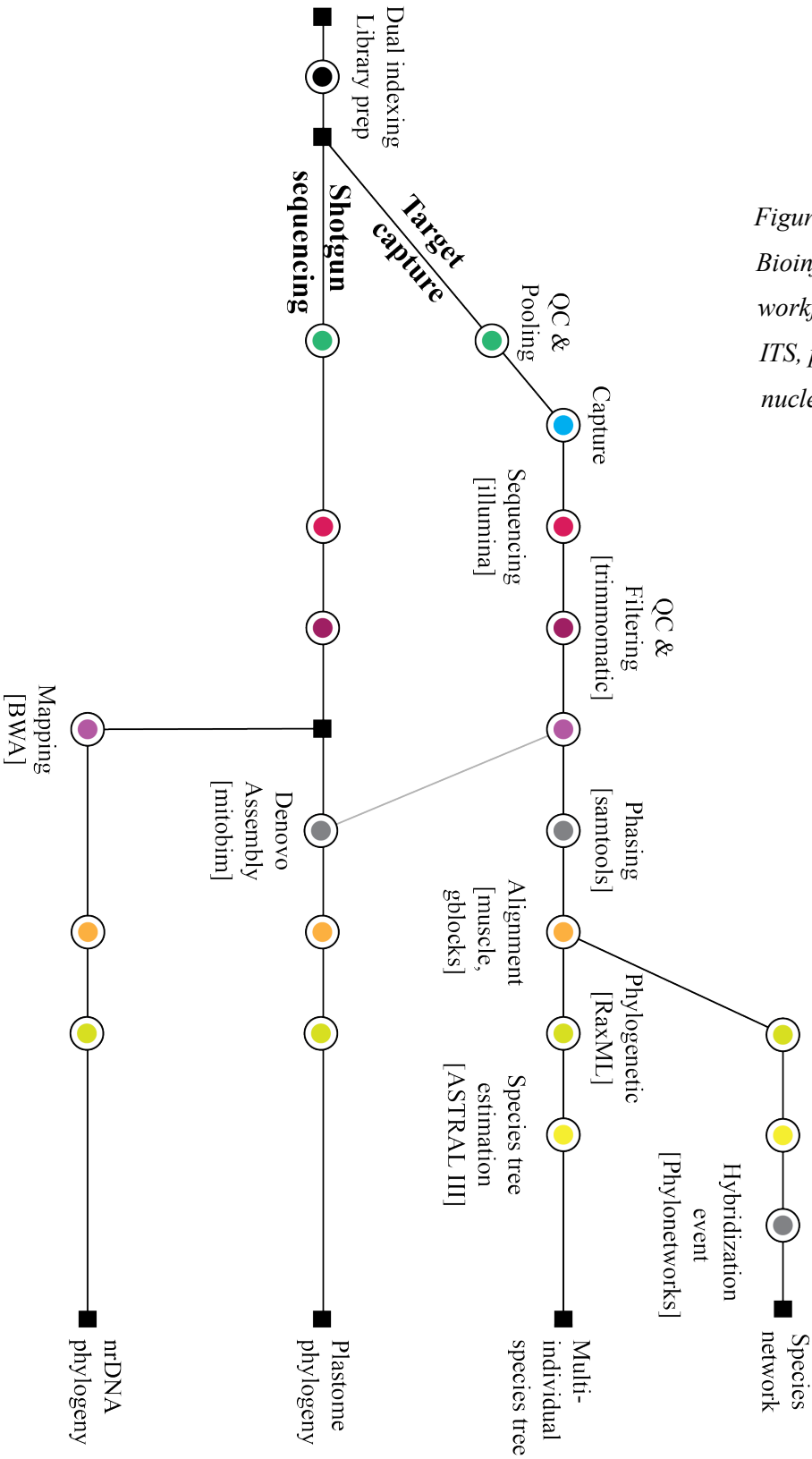


Figure 5. Bioinformatics workflow to retrieve the ITS, plastome and nuclear datasets.

3.3.2 PLASTOME AND NRDNA DATA

The shotgun sequencing run was filtered using the same procedure and parameters as the target capture data. In target capture, hybridization and cleaning have limited stringency and plastid reads are obtained mixed with the targeted loci, thus shotgun sequencing and target enrichment data were merged before retrieving the plastid and nrDNA (Figure 5). Plastid genomes were retrieved using MITOBim with mapping steps set to 5% divergence. nrDNA sequences were recovered using BWA by mapping the reads to the reference nrDNA of *Anacyclus pyrethrum* (KY397478) for *Anacyclus* species and traded samples, to the reference *Achillea pyrenaica* Sibth. ex Godr. (AY603247) for *Otanthus* and *Achillea*, and to the reference *Matricaria aurea* (Loefl.) Sch.Bip. (KT954177) for *Matricaria* samples.

4. PHYLOGENOMICS: SPECIES DELIMITATION AND MSC

4.1. PANAX PHYLOGENOMICS AND SPECIES DELIMITATION (PAPER III)

The matrix for phylogenomic analyses of *Panax sp.* consisted of complete aligned plastid genomes, and the global alignment was done using MAFFT version 7.3 (Katoh and Standley 2013) with local re-alignment using MUSCLE version 3.8.31 (Edgar 2004), and manual adjustments where necessary. Aligned DNA sequences have been deposited in the Open Science Framework (OSF) directory (https://osf.io/ryuz6). The final matrix has a total length of 163,499 bp for a total of 61 individuals with no missing data. Single nucleotide polymorphisms (SNPs) were visualized using Circos version 0.69 (Krzywinski et al. 2009). Relationships from the nucleotide matrix were inferred using Maximum Likelihood (ML) and Bayesian inference. The data were partitioned in coding regions, introns and intergenic spacers, and a best-fit partitioning scheme for the combined

dataset was determined using PartitionFinder version 2.1.1 (Lanfear et al. 2012) using the Bayesian Information Criterion. Branch lengths were linked across partitions.

The dataset was analyzed using RAxML version 8.2.10 (Stamatakis 2006) and mrBayes version 3.2.6 (Ronquist et al. 2012). RAxML and Bayesian searches used the partition model determined by PartitionFinder. For the ML analyses, tree searches and bootstrapping were conducted simultaneously with 1000 bootstrap replicates. Bayesian analysis were started using a random starting tree and were run for a total of ten million generations, sampling every 1000 generations. Four Markov runs were conducted with eight chains per run. We used AWTY to assess the convergence of the analyses (Nylander et al. 2008). Conflicting data within ML and Bayesian analyses were visualized and explored using the R package *phangorn* using the *consensusNet* function (Schliep 2011).

Suitable barcoding markers were selected by extracting the SNP density over the plastid genome alignment of all *Panax* species and individuals included in this study (matrix available as supplementary data on OSF). We used SNP-sites version 2.3.2 (Page et al. 2016) to extract the SNP positions from the alignment of a matrix containing only the *Panax* species, and created bins every 800 bp using Bedtools version 2.26.0 (Quinlan 2014) (script available on OSF) and plotted the SNP density using Circos (Krzywinski et al. 2009). The coordinates of each annotation on the aligned *Panax* species matrix were found using a reference consisting of the four annotated genomes produced in this study, and subsequently exported to Circos. We selected the most variable regions and designed suitable primers for these regions. From the matrix used for the Aralioideae, we extracted 15 plastid markers and download ITS sequences for the *Aralia-Panax* group. We performed maximum likelihood analyses on individual and

concatenated matrices using RAxML. Species delimitation analyses were performed with the multi-rate Poisson Tree Processes (mPTP) package using the ML trees from the individual and concatenated markers, and using the Markov chain Monte Carlo (MCMC) algorithm with two chains and the Likelihood Ratio Test set to 0.01.

4.2. PLASTID AND NRDNA TREE RECONSTRUCTION (**PAPERS IV** AND **V**)

We recovered plastid genomes for 54 *Anacyclus* traded samples and for the nrDNA from 102 *Anacyclus* traded samples. The recovered matrices were aligned using MAFFT (Katoh and Standley 2013) and the alignment refined with MUSCLE (Edgar 2004). Plastome and ITS phylogenies were inferred using RAxML version 8.0.26 (Stamatakis 2006), with 1000 bootstrap replicates under the GTRGAMMA model.

4.3. NUCLEAR GENE TREE RECONSTRUCTION (**PAPERS IV** AND **V**)

Nuclear gene trees were reconstructed for each individual nuclear locus. Samples with >7% missing data across markers were removed from the entire dataset, and markers with >5% missing data were removed as well as these were considered to have insufficient enrichment success. Retained matrices were re-aligned using MUSCLE, and Gblock (Talavera and Castresana 2007). For the phylogenetic analyses, only matrices were used with more than 400 bp and no missing samples were used. In brief, the final set of matrices consisted of 443 matrices with two alleles per individual, with a minimum length of 400bp, no missing samples and less than 5% missing data. For each of these low-copy nuclear markers, we inferred a gene tree using RAxML version 8.0.26 with 1000 bootstrap replicates under the GTRGAMMA model. A species tree was inferred from the individual nuclear gene trees using ASTRAL-III v5.5.9 (Mirarab and Warnow 2015). The multi-alleles option in ASTRAL-III was used for reconciliation of the independent evolutionary history of the alleles.

4.4. HYBRIDISATION ANALYSES

In order to understand hybridization events in the evolution of *Anacyclus*, a network approach using SNaQ (Species Networks applying Quartets) that generates explicit networks with reticulated nodes reflecting gene flow between taxa (Solís-Lemus and Ané 2016) was used. Under incomplete lineage sorting (ILS) or gene flow resulting from hybridization and introgression, concatenation of matrices is statistically inconsistent (Kubatko and Degnan 2007) and multispecies coalescence methods are not robust (Solís-Lemus et al. 2016). The advantage of the explicit network-based model is the incorporation of uncertainty into the estimated gene trees as well as gene tree discordance due to ILS. SNaQ is implemented in the package phylonetworks (Solís-Lemus et al. 2017). Because of computation limitations of the quartet frequencies (CF) (Solís-Lemus et al. 2017), we only tested hybridization hypotheses on a reduced dataset. Based on the CF from the RaxML gene trees, we ran five different analyses with the SNaQ algorithm using 20 random starting points. We allowed a range of maximum of possible hybridization events from zero (null hypothesis) to five (*hmax*).

5. FLOW CYTOMETRY (PAPER IV)

Silica gel-dried tissue of 57 *Anacyclus* sp. samples from all but one of the sequenced taxa (*A. monanthos*) was used to carry out flow cytometry analyses. Relative nuclear DNA content was determined by flow cytometry following the simplified two-step protocol of Doležel and Bartoš (2007). For each sample separately, part of the leaf blade was chopped together with an appropriate volume of the internal reference standard (*Bellis perennis* L., 2C = 3.38 pg; (Schönswetter et al. 2007)) using a sharp razor blade in a Petri-dish containing 0.5 mL of ice-cold Otto I buffer (0.1 M citric acid, 0.5 % Tween 20). The suspension was filtered through a 42-µm nylon mesh and incubated for 10 min at room temperature. The relative fluorescence intensity of 3,000 particles was recorded using a Partec ML flow cytometer (Partec GmbH, Münster, Germany) equipped with a UV LED chip as the excitation source. Histograms were evaluated using FloMax software, ver. 2.4d (Partec).

RESULTS: SUMMARY OF PAPERS

The results of this thesis are presented in three published papers (**Papers I**, **II** and **III**) and two manuscripts ready for submission (Papers IV and V).

PAPER I. DNA METABARCODING OF ORCHID-DERIVED PRODUCTS REVEALS WIDESPREAD ILLEGAL ORCHID TRADE

Orchids are some of the most vulnerable plants and its trade is regulated by national and international legislation (CITES). Accurate identification of orchid material in trade is paramount to target conservation efforts. In this paper, orchids used in Eastern Mediterranean food products traded internationally are identified using nrITS1 and nrITS2 DNA metabarcoding. Fifty-five commercial orchid products from Iran, Turkey, Greece and Germany were purchased and amplicon sequencing retrieved DNA sequences for 30 samples.

A total of 161 plant taxa, as well as the co-occurrence of the most popular ingredients in *salep* mixtures, were identified (Figure 6). Ten species of terrestrial orchid species with tuberous bulbs were identified in 13 samples. Whilst some samples contained only adulterants and substitutes, the international trade of *salep* poses threatening pressure to wild orchid populations. According to the interspecific genetic distance analyses results, species level identification of the most commonly traded orchids in Eastern Mediterranean can be made with a high level of confidence.

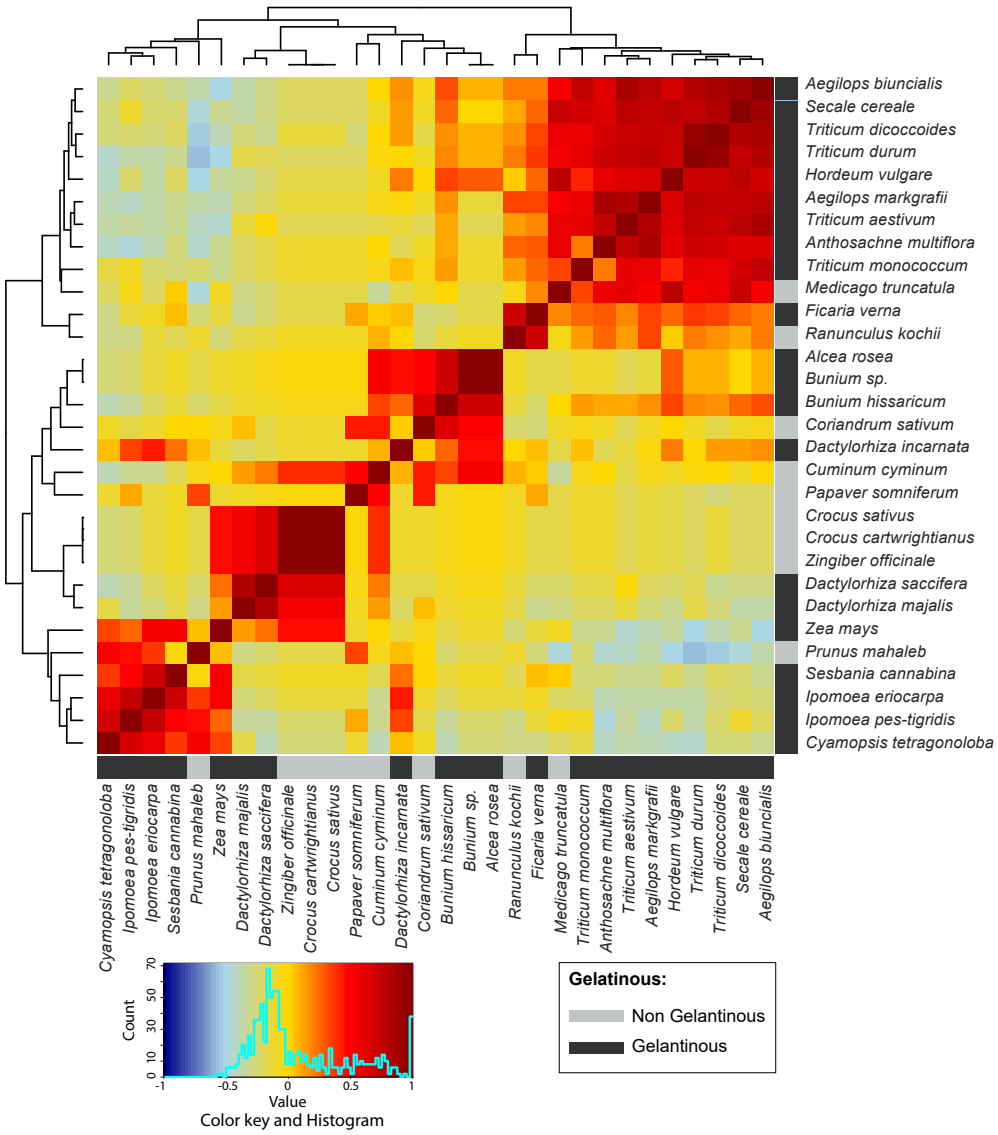


Figure 6. Pearson's correlation heat map showing correlation between gelatinous taxa across the salep samples. Dark red denotes high correlation ($r \rightarrow 1$), dark blue high anti-correlation ($r \rightarrow -1$), and yellow a lack of correlation ($r \cong 0$). The histogram in the color key represents the density of the Pearson's correlation coefficients across the matrix.

PAPER II. COMPARATIVE AUTHENTICATION OF *HYPERICUM PERFORATUM* HERBAL PRODUCTS USING DNA METABARCODING, TLC AND HPLC-MS

Correct identification of herbal products is key to ensure health and safety issues related to their commercialisation and consumption. This paper compares chemical and DNA metabarcoding approaches for plant identification and quality control of single or multi-species herbal products. Seventy-eight products containing *Hypericum perforatum* alone or mixed with other plants were tested to certify presence of the species mentioned on the product labels, as well as identify possible adulterants.

Standard chemical methods were not able to accurately distinguish between *Hypericum perforatum* and other *Hypericum* species. DNA was successfully sequenced by HTS amplicon methods from 38 out of the 78 products. Forty of the samples yielded no MOTUs for either nrITS1 or nrITS2, and were excluded from further analysis after applying quality criteria. ITS sequences identified a total of 219 species from thirty-eight samples using BLAST. This method identified *Hypericum perforatum* only in 68% of the investigated herbal products (Figure 7). Adulterants were found in all products. Several multi-species products did not contain all the species listed on the label, including the target *Hypericum perforatum*.

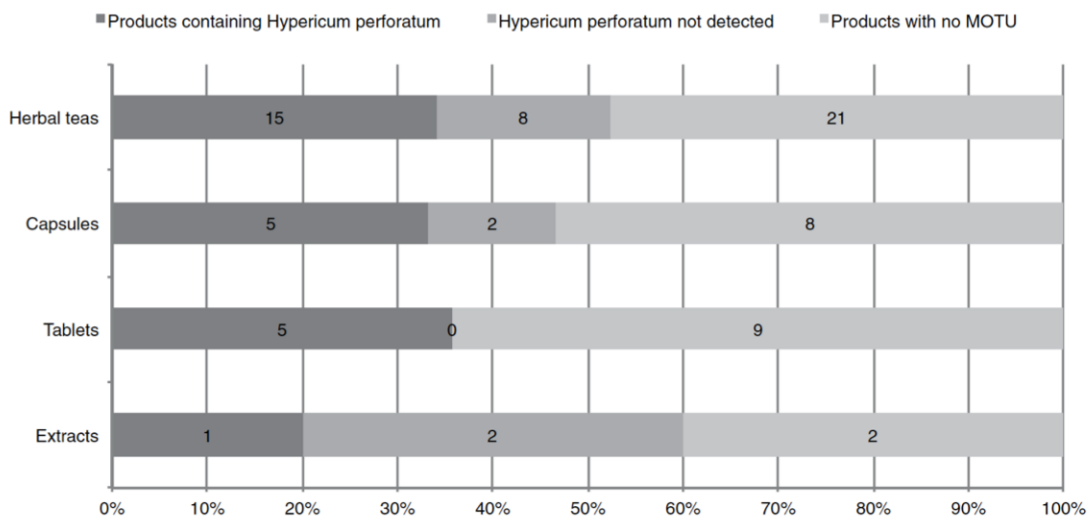


Figure 7. Presence of Hypericum perforatum *in amplicon-sequenced products.*

PAPER III. PHYLOGENOMICS AND BARCODING OF *PANAX*: TOWARD THE IDENTIFICATION OF GINSENG SPECIES

Ginseng is one of the most important globally traded medicinal plants, with a trade estimated in 2.1 billion USD. Ginseng products are commonly adulterated and chemical methods are easily deceived. In this paper, molecular methods based on a full plastome phylogenomic approach are developed to resolve the evolutionary history of *Panax* as well as point out markers for identification at species level.

HTS methods are used to sequence MBD2-depleted total DNA for plastome assembly of *Panax bipinnatifidus, P. stipuleanatus,* and *P. vietnamensis*. Supplementing these new plastomes with publically available plastid genomes, a phylogeny was built based on 60 fully assembled plastomes from eight different species. Analyses of the plastome matrix show that a combination of the markers *trnC-rps16, trnS-trnG,* and *trnE-trnM* can be used to distinguish all studied *Panax* species (Figure 8). Fractioning the plastid genome with MBD2 depletion reduces the cost of plastome sequencing, which makes it an alternative to traditional DNA barcoding approaches that use only a few markers.

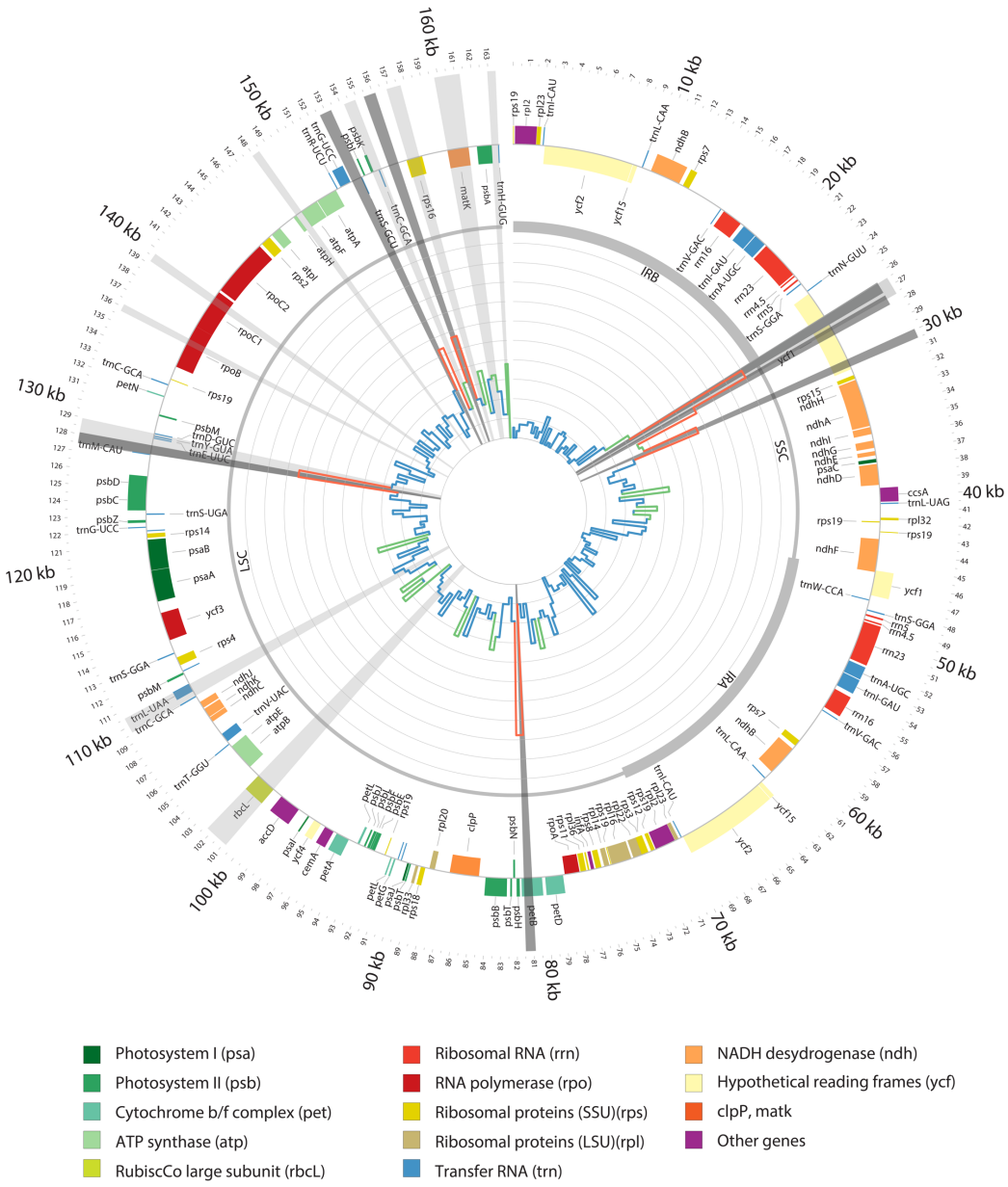


Figure 8. Plastid genome representation of the 38 aligned Panax *genomes. The internal histogram plot represents the SNPs density over the alignment. The colors indicate when the standard deviation of the bin falls in different intervals compare to the average standard deviation, between 0 and 1 in blue (low variation), between 1 to 2 in green (moderate variation) and over two in red (high variation). Inverted repeats A and B (IRA and IRB), large single copy (LSC) and small single copy (SSC) are shown in the inner circle by different line weights. Genes shown outside the outer circle are transcribed clockwise, and those inside are transcribed counter clockwise. Genes belonging to different functional groups are color-coded. Radial grey highlights show the regions in focus of the study, light grey previously used barcodes, in dark grey newly developed barcodes.*

PAPER IV. HYBRIDISATION AND ADAPTIVE RADIATION OF THE GENUS *ANACYCLUS* L. (ANTHEMIDAE, COMPOSITAE)

Paper IV describes the evolutionary history of the genus *Anacyclus*. Hybridisation often leads to speciation when it coincides with the emergence of new ecological niches in allopatric plant lineages. Hybridised lineages pose a challenge for molecular identification, and in-depth knowledge of a group's evolutionary history is necessary to design accurate molecular identification methods.

This paper presents a well-supported phylogeny of *Anacyclus* and sister groups in the Matricariinae tribe (Asteraceae) based on 443 low-copy nuclear markers. Multispecies coalescent (MSC) methods are used to estimate the species trees from single gene trees under strong incomplete lineage sorting (ILS) (Figure 9). Moreover, genome sizes of the different species are estimated using flow cytometry (Figure 9).

By analyzing gene tree topologies, two hybridization events are identified. An early hybridization event coincides with the establishment of the Mediterranean climatic rhythm (3.2 Mya) and the Quaternary-type Mediterranean climatic fluctuations (2.3 Mya). *Anacyclus* exemplifies the syngameon hypothesis: hybridization between distinct lineages may seed the onset of an entire adaptive radiation. A later second hybridization within *Anacyclus* confirms previous hypotheses based on morphological characters.

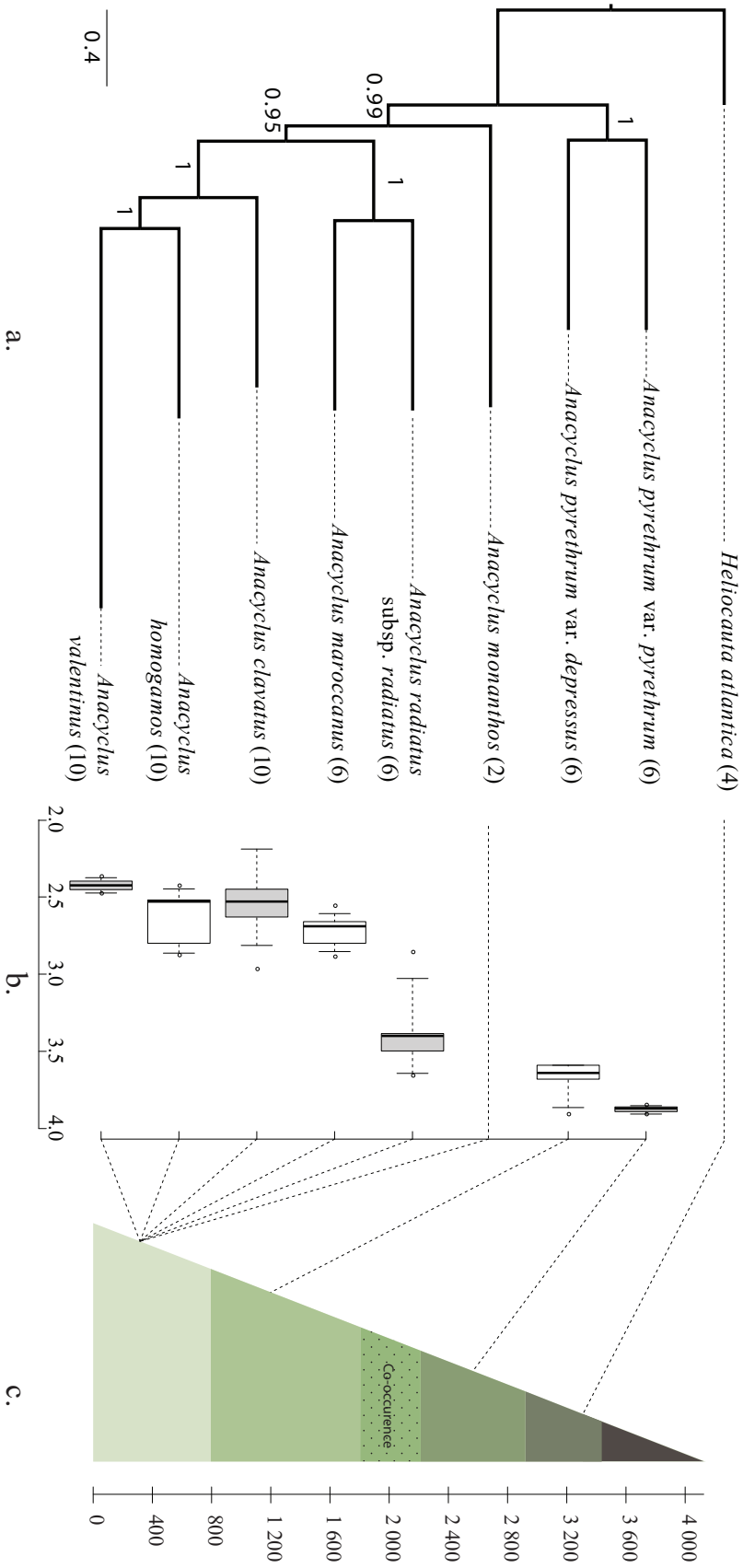


Figure 9. (a) Estimated species tree of a reduced dataset of 443 loci, 30 individuals and 2 alleles per individual. The numbers on the branches represent the posterior probabilities from the quartet analyses. The numbers in parentheses represent the number of sequences used for each species; (b) Relative nuclear DNA content of the different species in Anacyclus*, inferred with flow cytometry. The scale represents the genome size relative to the genome size of the internal standard* Bellis perennis*; (c) Ecological niches for the different species of* Anacyclus *(Humphries 1977, 1979; Rankou et al. 2015), the scale on the left represent the altitude in meters. .*

PAPER V. RAISING THE BAR OF MOLECULAR PLANT IDENTIFICATION SHEDS LIGHT ON COMPLEX TRADE IN RED-LISTED SPECIES

This paper uses a novel approach for molecular plant identification based on target enrichment sequencing and the selection of hundreds of low-copy nuclear markers. This approach has been suggested to address challenges posed by standard DNA barcoding methods. First, it provides accurate identification at species level in plant groups with a complex evolutionary history. Second, it allows successful identification from degraded samples.

In this paper, these methods are used for the first time to track the national and international trade of a red-listed medicinal species, *Anacyclus pyrethrum*. These methods are compared with partial plastid and nrITS approaches to identification. Fifty samples of *Anacyclus pyrethrum* in trade were collected in Morocco and India. Adulteration was assessed using both morphological and molecular characters. Traded samples were included in a phylogeny inferred from 443 nuclear loci and identified at species level.

Target capture sequencing and the selection of low-copy nuclear markers proves to be the most successful method to recover DNA even from degraded samples and to identify them to the species and population levels (Figure 10). This approach allowed to distinguish and track the national and international value chains of *A. pyrethrum*, with important implications for biodiversity conservation. Molecular plant identification based on large sets of nuclear genes retrieved from target capture approaches may be the way forward to guarantee the quality, safety, conservation and legal trade of use of both food and medicinal plant products.

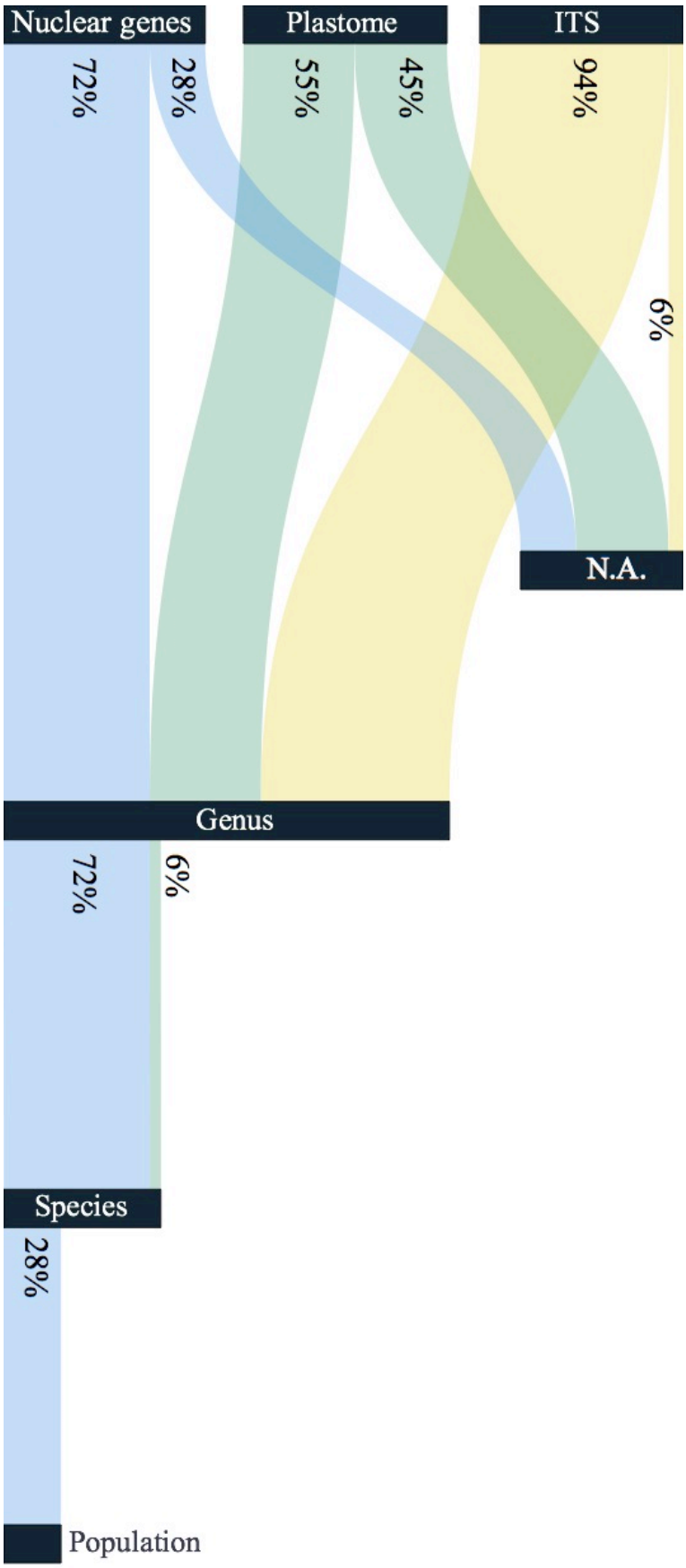


Figure 10: Success of the molecular identification at the genus, species and population level based on the ML phylogeny of ITS, plastome and nuclear markers.

DISCUSSION AND CONCLUSION

This thesis applies and evaluates all state-of-the-art biodiversity molecular identification tools, focusing on particularly challenging plant case studies. Amplicon-based barcoding uses the traditional barcoding approach with a few DNA markers, whereas shotgun and target sequencing techniques are based on the plastid and nuclear genomes for molecular identification, respectively. Each sequencing method has implications for the quantity and quality of molecular data retrieved, which impacts the success of identification. **Papers I** and **II** use distance-based methods to identify species based on nrDNA markers. This approach renders possible the identification of plant products that would otherwise be unidentifiable by using morphological or chemical techniques, but does not take into account the evolutionary history of the identified organisms. This can potentially bias identification, especially below genus level (Meier et al. 2008; Ross et al. 2008; Collins and Cruickshank 2013). False positive identifications could occur if reference sequences are missing from the reference database or when molecular variation of the markers is low and does not discriminate between closely related species (Meyer and Paulay 2005; Ross et al. 2008). In **Papers III** and **V**, molecular identification is grounded on the evolutionary frameworks of the plant groups studied, *Panax* and *Anacyclus* (this further developed in **Paper IV**). **Paper III** and **V** implement either a species delimitation approach or, for the first time in plants, a multi-species coalescence method (MSC).

Overall, molecular identification methods shed light on both alarming adulteration rates as well as illegal international trade of endangered plants. **Paper I** underscores the continuous role of terrestrial orchids in *salep* products, as well as the ubiquitous presence of substitutes with similar

gelatinous properties such as guar gum, and to a lesser extent common wheat, emmer wheat, durum wheat, rye, barley, and maize. Whilst adulteration does not seem to involve products with implications for human health, the growing demand for this product increases pressure on wild orchid populations. In **Paper II** high levels of substitution, adulteration, admixture or a combination of these are observed in herbal medicine. These results corroborate those from previous studies on the authentication of herbal medicine and show the potential of metabarcoding to complement traditional methods of quality control for food and safety agencies (Stoeckle et al. 2011; Baker 2012; Little and Jeanson 2013; Seethapathy et al. 2015; Raclariu et al. 2017a). However, amplicon sequencing has a relatively low success rate in generating sequence reads per product, possibly due to the degraded quality of plant DNA in this type of samples. This could make amplicon sequencing DNA metabarcoding unrewarding if implemented for routine screening. In **Paper I**, DNA sequences were retrieved for 55% of the sampled products. In **Paper II** only 49% of the tested products yielded sequences as compared to 95% of products that could be used for TLC and HPLC-MS analyses. This, together with limitations of identification at the species level (Hollingsworth 2011; Coissac et al. 2016; Hollingsworth et al. 2016), pointed towards the need to develop "extended barcodes" (Coissac et al. 2016; Hollingsworth et al. 2016), "super-barcode" (Li et al. 2015) or "ultrabarcodes" (Kane et al. 2012). So far, except for the "ultrabarcodes" used in Kane et al. (2012) , "extended barcodes" and "super-barcodes" had been proposed as solutions to barcoding plants but had not yet implemented to identify real samples.

To move forward with this next generation of molecular identification techniques, the evolutionary background of the targeted organisms has to be taken into account. Two case studies were selected here for the development of "extended barcodes" based on evolutionary complexity,

commercial value and threat to conservation. First, ginsengs were chosen as a case study to develop a whole plastid genome identification approach. The understanding of the evolution of the genus was improved by using whole plastid genome phylogeny for the first time in **Paper III**. **Paper III** provides a new hypothesis for the evolution of the *Panax* genus that contrasts with previous studies (e.g., Choi and Wen 2000; Shi et al. 2015). Discrepancies between nrDNA and plastome phylogenies are common in plants (Álvarez and Wendel 2003), and this could explain differences between **Paper III** and previous nrDNA-based *Panax* phylogenies. The phylogeographical implications of **Paper III** are unclear due to missing taxa, and the addition of further population data as well as additional species will improve our understanding of the evolutionary and biogeographical history of the genus. Since *Panax* is a complex of recently diverged and hybrid species (Kim and Lee 2004), traditional barcoding markers are most likely inaccurate (Hollingsworth 2011). As a result of **Paper III**, it is now possible to identify ginsengs to species level using plastid genome DNA sequence data.

As background work for yet a step further in molecular identification, **Paper IV** uses cutting-edge sequencing technologies and bioinformatic tools in evolutionary biology to study the evolution of the genus *Anacyclus* based on low-copy nuclear markers. This approach is able to provide insights in the past and recent hybridization events within the genus, which corresponds to common biogeographic and evolutionary patterns of many plant species in the Mediterranean region (Oberprieler 2004). Finally, **Paper V** shows that the implementation of identification methods based on low-copy nuclear genes associated with MSC can provide species- and population-level resolution for plant identification. This has important implications in the regulation of trade and conservation biology, for example by protecting specific species or populations from trade through

molecular monitoring. For example, the trade of *Aloe vera* is legal, but all other *Aloe* species are protected by the Convention on International Trade in Endangered Species (Grace 2011; CITES 2016). Another example are the various ginseng species and populations with different levels of protection presented here. The ability of these new molecular identification methods applied and developed in this thesis to detect traded plants at species level will enable regulatory agencies (e.g. custom offices, the European Medicines Agency, CITES authorities and environmental agencies amongst others) to monitor trade of endangered, protected or legitimate plants and plant products by authenticating traded materials. Only with accurate identification, national and international legislation can be enforced.

REFERENCES

Adams M., Alther W., Kessler M., Kluge M., Hamburger M. 2011. Malaria in the Renaissance: remedies from European herbals from the 16th and 17th century. J. Ethnopharmacol. 133:278–288.

Álvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ankenbrand M., Pfaff S., Terhoeven N., Qureischi M., Gündel M., L Weiß C., Hackl T., Förster F. 2018. chloroExtractor: extraction and assembly of the chloroplast genome from whole genome shotgun data. J. Open Source Softw. 3:464.

Arditti J. 1992. Fundamentals of orchid biology. John Wiley & Sons.

Baker D.A. 2012. DNA Barcode Identification of Black Cohosh Herbal Dietary Supplements. J. AOAC Int. 95:1023–1034.

Basnet D., Dey K. 2008. Studies on seed germination of an Indian ginseng (*Panax assamicus* Ban. spec. nov.) for successful cultivation and conservation. Indian J. For.

Benarba B. 2016. Medicinal plants used by traditional healers from South-West Algeria: An ethnobotanical study. J. Intercult. Ethnopharmacol. 5:320.

Benarba B., Belabid L., Righi K., amine Bekkar A., Elouissi M., Khaldi A., Hamimed A. 2015. Ethnobotanical study of medicinal plants used by traditional healers in Mascara (North West of Algeria). J. Ethnopharmacol. 175:626–637.

Boecklen W.J. 2017. Topology of syngameons. Ecol. Evol. 7:10486–10491.

de Boer H.J., Ghorbani A., Manzanilla V., Raclariu A.-C., Kreziou A., Ounjai S., Osathanunkul M., Gravendeel B. 2017. DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade. 284:20171182.

de Boer H.J., Ichim M.C., Newmaster S.G. 2015. DNA barcoding and pharmacovigilance of herbal medicines. Drug Saf. 38:611–620.

de Boer H.J., Ouarghidi A., Martin G., Abbad A., Kool A. 2014. DNA barcoding reveals limited accuracy of identifications based on folk taxonomy. PLoS ONE. 9:e84291.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Booker A., Frommenwiler D., Johnston D., Umealajekwu C., Reich E., Heinrich M. 2014. Chemical variability along the value chains of turmeric (*Curcuma longa*): a comparison of nuclear magnetic resonance spectroscopy and high performance thin layer chromatography. J. Ethnopharmacol. 152:292–301.

Booker A., Johnston D., Heinrich M. 2012. Value chains of herbal medicines-Research needs and key challenges in the context of ethnopharmacology. J. Ethnopharmacol. 140:624–633.

Booker A., Johnston D., Heinrich M. 2015. Value chains of herbal medicines—ethnopharmacological and analytical challenges in a globalizing world. Evidence-Based Validation of Herbal Medicine. Elsevier. p. 29–44.

Booker A., Johnston D., Heinrich M. 2016. The welfare effects of trade in phytomedicines: a multi-disciplinary analysis of turmeric production. World Dev. 77:221–230.

Borrelli F., Izzo A.A. 2009. Herb–drug interactions with St John's wort (*Hypericum perforatum*): an update on clinical observations. AAPS J. 11:710.

Buerkle C.A., Morris R.J., Asmussen M.A., Rieseberg L.H. 2000. The likelihood of homoploid hybrid speciation. Heredity. 84:441–451.

Bulpitt C. 2005. The uses and misuses of orchids in medicine. Qjm. 98:625–631.

Bulpitt C.J., Li Y., Bulpitt P.F., Wang J. 2007. The use of orchids in Chinese medicine. J. R. Soc. Med. 100:558–563.

Büter B., Orlacchio C., Soldati A., Berger K. 1998. Significance of genetic and environmental aspects in the field cultivation of *Hypericum perforatum*. Planta Med. 64:431–437.

Bybee S.M., Bracken-Grissom H., Haynes B.D., Hermansen R.A., Byers R.L., Clement M.J., Udall J.A., Wilcox E.R., Crandall K.A. 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol. Evol. 3:1312–1323.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421.

CBOL, Hollingsworth P.M., Forrest L.L., Spouge J.L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M.W., Cowan R.S., Erickson D.L., Fazekas A.J., Graham S.W., James K.E., Kim K.-J., Kress W.J., Schneider H., van AlphenStahl J., Barrett S.C.H., van den Berg C., Bogarin D., Burgess K.S., Cameron K.M., Carine M., Chacón J., Clark A., Clarkson J.J., Conrad F., Devey D.S., Ford C.S., Hedderson T.A.J., Hollingsworth M.L., Husband B.C., Kelly L.J., Kesanakurti P.R., Kim J.S., Kim Y.-D., Lahaye R., Lee H.-L., Long D.G., Madriñán S., Maurin O., Meusnier I., Newmaster S.G., Park C.-W., Percy D.M., Petersen G., Richardson J.E., Salazar G.A., Savolainen V., Seberg O., Wilkinson M.J., Yi D.-K., Little D.P. 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. 106:12794–12797.

CBOL, Li D.-Z., Gao L.-M., Li H.-T., Wang H., Ge X.-J., Liu J.-Q., Chen Z.-D., Zhou S.-L., Chen S.-L., Yang J.-B., Fu C.-X., Zeng C.-X., Yan H.-F., Zhu Y.-J., Sun Y.-S., Chen S.-Y., Zhao L., Wang K., Yang T., Duan G.-W. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc. Natl. Acad. Sci. 108:19641–19646.

Chen S., Yao H., Han J., Liu C., Song J., Shi L., Zhu Y., Ma X., Gao T., Pang X., Luo K., Li Y., Li X., Jia X., Lin Y., Leon C. 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE. 5:e8613.

China Plant Specialist Group. 2004. *Panax zingiberensis*. .

Chinsamy M., Finnie J.F., Van Staden J. 2011. The ethnobotany of South African medicinal orchids. South Afr. J. Bot. 77:2–9.

Choi H.-K., Wen J. 2000. A phylogenetic analysis of *Panax* (Araliaceae): integrating cpDNA restriction site and nuclear rDNA ITS sequence data. Plant Syst. Evol. 224:109–120.

Christenhusz M.J., Byng J.W. 2016. The number of known plants species in the world and its annual increase. Phytotaxa. 261:201–217.

CITES. 2016. CITES (Convention on International Trade in Endangered Species) Handbook: convention on species of wild fauna and flora, July 2016. CITES Secretariat de la Convention sur le commerce international des espèces de faune et de flore sauvages menacées d'extinction.

CITES and Medicinal Plants. 2018. CITES and Medicinal Plants. Available from https://www.cites.org/eng/prog/medplants.

Coghlan M.L., Haile J., Houston J., Murray D.C., White N.E., Moolhuijzen P., Bellgard M.I., Bunce M. 2012. Deep Sequencing of Plant and Animal DNA Contained within Traditional Chinese Medicines Reveals Legality Issues and Health Safety Concerns. PLoS Genet. 8:e1002657.

Coissac E. 2012. OligoTag: a program for designing sets of tags for next-generation sequencing of multiplexed samples. Methods Mol Biol. 888.

Coissac E., Hollingsworth P.M., Lavergne S., Taberlet P. 2016. From barcodes to genomes: extending the concept of DNA barcoding. Mol. Ecol. 25:1423–1428.

Coissac E., Taberlet P., Roquet C., Boleda M., Gielly L., Alberti A., Wincker P., Alsos I., Boyer F., Lavergne S. 2015. Towards an universal genome-based DNA barcode-The PhyloAlps project. 58:206–206.

Collins R., Cruickshank R. 2013. The seven deadly sins of DNA barcoding. Mol. Ecol. Resour. 13:969–975.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

De Vos P. 2010. European materia medica in historical texts: longevity of a tradition and implications for future use. J. Ethnopharmacol. 132:28–47.

Dierckxsens N., Mardulyn P., Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18–e18.

Doležel J., Greilhuber J., Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nat. Protoc. 2:2233.

Doyle J.J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull Bot Soc Am. 19:11–15.

Ece Tamer C., Karaman B., Utku Copur O. 2006. A traditional Turkish beverage: salep. Food Rev. Int. 22:43–50.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

EDQM. 2018. European Pharmacopoeia 9th Edition | EDQM. Available from https://www.edqm.eu/en/european-pharmacopoeia-ph-eur-9th-edition.

European Medicines Agency (EMA). 2011. Guideline on specifications: test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products /traditional herbal medicinal products. .

Ewels P., Magnusson M., Lundin S., Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:3047–3048.

FairWild. 2018. FairWild. Available from http://www.fairwild.org/.

Fazekas A.J., Burgess K.S., Kesanakurti P.R., Graham S.W., Newmaster S.G., Husband B.C., Percy D.M., Hajibabaei M., Barrett S.C. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One. 3:e2802.

Fazekas A.J., Kesanakurti P.R., Burgess K.S., Percy D.M., Graham S.W., Barrett S.C., Newmaster S.G., Hajibabaei M., Husband B.C. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol. Ecol. Resour. 9:130–139.

Feliner G.N., Álvarez I., Fuertes-Aguilar J., Heuertz M., Marques I., Moharrek F., Piñeiro R., Riina R., Rosselló J., Soltis P. 2017. Is homoploid hybrid speciation that rare? An empiricist's view. Heredity. 118:513–516.

Feng S., Cokus S.J., Zhang X., Chen P.-Y., Bostick M., Goll M.G., Hetzel J., Jain J., Strauss S.H., Halpern M.E. 2010. Conservation and

divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. 107:8689–8694.

Ficetola G.F., Coissac E., Zundel S., Riaz T., Shehzad W., Bessière J. 2010. An in silico approach for the evaluation of DNA barcodes. BMC Genom. 11.

Galeotti N. 2017. *Hypericum perforatum* (St John's wort) beyond depression: A therapeutic perspective for pain conditions. J. Ethnopharmacol. 200:136–146.

Ghorbani A., Gravendeel B., Naghibi F., de Boer H. 2014. Wild orchid tuber collection in Iran: a wake-up call for conservation. Biodivers. Conserv. 23:2749–2760.

Ghorbani A., Gravendeel B., Selliah S., Zarré S., Boer H. 2017. DNA barcoding of tuberous Orchidoideae: a resource for identification of orchids used in Salep. Mol. Ecol. Resour. 17:342–352.

Ghorbani A., Gravendeel B., Selliah S., Zarre S., de Boer H.J. 2016. DNA barcoding of tuberous Orchidoideae: A resource for identification of orchids used in Salep. Mol Ecol Resour.

Ghosh S., General F.D.D. 2013. Promotion of Medicinal and Aromatic Plants in the Asia-Pacific Region. Expert Consult. Promot. Med. Aromat. Plants Asia-Pac. Reg. Proc.:33.

Glenn T.C. 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11:759–769.

Grace O.M. 2011. Current perspectives on the economic botany of the genus Aloe L.(Xanthorrhoeaceae). South Afr. J. Bot. 77:980–987.

Gurevich A., Saveliev V., Vyahhi N., Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.

Hahn C., Bachmann L., Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41:e129–e129.

Hamilton A. 2003. Medicinal plants and conservation: issues and approaches. Int. Plants Conserv. Unit WWF-UK. 51.

Hamilton A. 2004. Medicinal plants, conservation and livelihoods. Biodivers. Conserv. 13:1477–1517.

Hamilton A., Dürbeck K., Lawrence A. 2006. Towards a sustainable herbal harvest: a work in hand. Plant Talk. 43:32–35.

Hawkins B. 2008. Plants for life: Medicinal plant conservation and botanic gardens. Botanic Gardens Conservation International,.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270:313–321.

Hollingsworth M.L., Andra Clark A., Forrest L.L., Richardson J., Pennington R., Long D.G., Cowan R., Chase M.W., Gaudeul M., Hollingsworth P.M. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol. Ecol. Resour. 9:439–457.

Hollingsworth P.M. 2011. Refining the DNA barcode for land plants. Proc. Natl. Acad. Sci. 108:19451–19452.

Hollingsworth P.M., Graham S.W., Little D.P. 2011. Choosing and Using a Plant DNA Barcode. PLoS ONE. 6:e19254.

Hollingsworth P.M., Li D.-Z., van der Bank M., Twyford A.D. 2016. Telling plant species apart with DNA: from barcodes to genomes. Phil Trans R Soc B. 371:20150338.

Hossain M.M. 2011. Therapeutic orchids: traditional uses and recent advances—an overview. Fitoterapia. 82:102–140.

Humphries C.J. 1979. A revision of the genus *Anacyclus* L. (Compositae: Anthemideae). Bull. Br. Mus. Nat. Hist. Hist. Ser.:83–142.

ISE Code of Ethics Online2018. Available from http://www.ethnobiology.net/what-we-do/core-programs/ise-ethics-program/code-of-ethics/code-in-english/.

Jain A. 1994. Vulnerable and threatened plants of economic value: *Panax pseudo-ginseng* Wall. Himal. Ginseng MFP News. 4:21.

Jaiswal Y., Liang Z., Zhao Z. 2016. Botanical drugs in Ayurveda and traditional Chinese medicine. J. Ethnopharmacol. 194:245–259.

Jamila F., Mostafa E. 2014. Ethnobotanical survey of medicinal plants used by people in Oriental Morocco to manage various ailments. J. Ethnopharmacol. 154:76–87.

Joshi G., Tiwari K., Tiwari R., Uniyal M. 1991. Conservation and large scale cultivation strategy of Indian ginseng- *Panax pseudoginseng* Wall. Indian For. 117:131–134.

Kamneva O.K., Syring J., Liston A., Rosenberg N.A. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. BMC Evol. Biol. 17:180.

Kane N., Sveinsson S., Dempewolf H., Yang J.Y., Zhang D., Engels J.M.M., Cronk Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. Am. J. Bot. 99:320–329.

Kasparek M., Grimm U. 1999. European trade in Turkish salep with special reference to Germany. Econ. Bot. 53:396–406.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kent W.J. 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Khan I.A., Smillie T. 2012. Implementing a "quality by design" approach to assure the safety and integrity of botanical dietary supplements. J. Nat. Prod. 75:1665–1673.

Kim K.-J., Lee H.-L. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 11:247–261.

Kimmens A.C. 1975. Tales of the ginseng. Morrow.

Komatsu K., Zhu S., Fushimi H., Qui T.K., Cai S., Kadota S. 2001. Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of *Panax vietnamensis* and five related species. Planta Med. 67:461–465.

Kool A., de Boer H.J., Krüger Å., Rydberg A., Abbad A., Björk L., Martin G. 2012. Molecular identification of commercialized medicinal plants in southern Morocco. PLoS ONE. 7:e39459.

Kress W.J., Erickson D.L. 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One. 2:e508.

Kress W.J., Wurdack K.J., Zimmer E.A., Weigt L.A., Janzen D.H. 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U. S. A. 102:8369–8374.

Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J., Marra M.A. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lammers Y., Peelen T., Vos R.A., Gravendeel B. 2014. The HTS barcode checker pipeline, a tool for automated detection of illegally traded species from high-throughput sequencing data. BMC Bioinformatics. 15:44.

Lanfear R., Calcott B., Ho S.Y., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth. 9:357–359.

Lemmon E.M., Lemmon A.R. 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44:99–121.

Li H. 2012. seqtk Toolkit for processing sequences in FASTA/Q formats. .

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N. 2009. Genome project data processing S. The sequence alignment/Map format and SAMtools. Bioinformatics. 25.

Li W., Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658–1659.

Li X., Yang Y., Henry R.J., Rossetto M., Wang Y., Chen S. 2015. Plant DNA barcoding: from gene to genome. Biol. Rev. 90:157–166.

Linde K. 2009. St. John's wort–an overview. Complement. Med. Res. 16:146–155.

Little D.P., Jeanson M.L. 2013. DNA barcode authentication of saw palmetto herbal dietary supplements. Sci. Rep. 3:3518.

Lohse M., Drechsel O., Bock R. 2007. Organellar Genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr. Genet. 52:267–274.

Lotsy J.P. 1931. On the species of the taxonomist in its relation to evolution. Genetica. 13:1–16.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Malé P.G., Bardon L., Besnard G., Coissac E., Delsuc F., Engel J., Lhuillier E., Scotti-Saintagne C., Tinaut A., Chave J. 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour. 14:966–975.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., Turner D.J. 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods. 7:111.

Manzanilla V., Kool A., Nhat L.N., Van H.N., de Boer H. 2018. Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species. bioRxiv.:244780.

Marcussen T., Sandve S.R., Heier L., Spannagl M., Pfeifer M., Jakobsen K.S., Wulff B.B., Steuernagel B., Mayer K.F., Olsen O.-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science. 345:1250092.

Matasci N., Hung L.-H., Yan Z., Carpenter E.J., Wickett N.J., Mirarab S., Nguyen N., Warnow T., Ayyampalayam S., Barker M. 2014. Data access for the 1,000 Plants (1KP) project. Gigascience. 3:17.

Meier J.I., Marques D.A., Mwaiko S., Wagner C.E., Excoffier L., Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8.

Meier R., Zhang G., Ali F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Syst. Biol. 57:809–813.

Merzouki A., Ed-derfoufi F., Molero Mesa J. 2000. Contribution to the knowledge of Rifian traditional medicine. II: Folk medicine in Ksar Lakbir district (NW Morocco). Fitoterapia. 71:278–307.

Meyer C.P., Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3:e422.

Meyer M., Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb. Protoc. 2010:pdb.prot5448.

Millspaugh C.F. 1892. American medicinal plants: an Illustrated and descriptive guide to plants indigenous to and naturalized in the United States which are used in medicine. Dover Publications.

Ministry of Health and Family Welfare. 2007. The ayurvedic pharmacopoeia of India. Govt. of India, Ministry of Health and Family Welfare, Dept. of ISM & H.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.

Narasimhan V., Danecek P., Scally A., Xue Y., Tyler-Smith C., Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 32:1749–1751.

Newmaster S., Grguric M., Shanmughanandhan D., Ramalingam S., Ragupathy S. 2013. DNA barcoding detects contamination and substitution in North American herbal products. BMC Med. 11:222.

Nock C.J., Waters D.L., Edwards M.A., Bowen S.G., Rice N., Cordeiro G.M., Henry R.J. 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol. J. 9:328–333.

NorBOL. 2018. Norwegian Barcode of Life (NorBOL). Available from http://www.norbol.org/.

Nylander J.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24:581–583.

Oberprieler C. 2004. On the taxonomic status and the phylogenetic relationships of some unispecific Mediterranean Genera of

Compositae-Anthemideae I. *Brocchia*, *Endopappus* and *Heliocauta*. Willdenowia. 34:39–57.

O'Neill E.M., Schwartz R., Bullock C.T., Williams J.S., Shaffer H.B., Aguilar-Miguel X., Parra-Olea G., Weisrock D.W. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Mol. Ecol. 22:111–129.

Otieno J., Abihudi S., Veldman S., Nahashon M., van Andel T., de Boer H.J. 2015. Vernacular dominance in folk taxonomy: a case study of ethnospecies in medicinal plant trade in Tanzania. J. Ethnobiol. Ethnomedicine. 11:10.

Ouarghidi A., Martin G., Powell B., Esser G., Abbad A. 2013. Botanical identification of medicinal roots collected and traded in Morocco and comparison to the existing literature. J. Ethnobiol. Ethnomedicine. 9:59.

Ouarghidi A., Powell B., Martin G.J., Abbad A. 2017. Traditional Sustainable Harvesting Knowledge and Distribution of a Vulnerable Wild Medicinal Root (*A. pyrethrum* var. *pyrethrum*) in Ait M'hamed Valley, Morocco. Econ. Bot. 71:83–95.

Ouarghidi A., Powell B., Martin G.J., De Boer H., Abbad A. 2012. Species substitution in medicinal roots and possible implications for toxicity of herbal remedies in Morocco. Econ. Bot. 66:370–382.

Ouelbani R., Bensari S., Mouas T.N., Khelifi D. 2016. Ethnobotanical investigations on plants used in folk medicine in the regions of Constantine and Mila (North-East of Algeria). J. Ethnopharmacol. 194:196–218.

Page A.J., Taylor B., Delaney A.J., Soares J., Seemann T., Keane J.A., Harris S.R. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genomics. 2.

Parks M., Cronn R., Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7:84.

Paterson A.H., Wendel J.F., Gundlach H., Guo H., Jenkins J., Jin D. 2012. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 492:423.

Pillon Y., Johansen J., Sakishima T., Chamala S., Barbazuk W.B., Roalson E.H., Price D.K., Stacy E.A. 2013. Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. BMC Evol. Biol. 13:35.

Pittle K.D. 2005. Continuity and Change in Islamic Ethnopharmacological Practice: New Methods for Cognitive Dialectometry. .

Quinlan A.R. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinforma.:11–12.

Raclariu A.C., Heinrich M., Ichim M.C., Boer H. 2017a. Benefits and limitations of DNA barcoding and metabarcoding in herbal product authentication. Phytochem. Anal.

Raclariu A.C., Mocan A., Popa M.O., Vlase L., Ichim M.C., Crisan G., Brysting A.K., de Boer H. 2017b. *Veronica officinalis* Product Authentication Using DNA Metabarcoding and HPLC-MS Reveals Widespread Adulteration with *Veronica chamaedrys*. Front. Pharmacol. 8.

Raclariu A.C., Paltinean R., Vlase L., Labarre A., Manzanilla V., Ichim M.C., Crisan G., Brysting A.K., de Boer H. 2017c. Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding, TLC and HPLC-MS. Sci. Rep. 7:1291.

Rhafouri R., Aafi A., Zair T., Strani B., El Omari M., Ghanmi M. 2014. Ethnobotanical study of medicinal plants in Ifran's National Park (Morocco). .

Rieseberg L.H., Wood T.E., Baack E.J. 2006. The nature of plant species. Nature. 440:524.

Robbins C.S. 1998. American ginseng: the root of North America's medicinal herb trade. Traffic North America.

Robert T. 1993. Gunther. The Greek Herbal of Dioscorides. .

Ronquist F., Teslenko M., Mark P., Ayres D.L., Darling A., Hohna S. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61.

Rosato M., Álvarez I., Feliner G.N., Rosselló J.A. 2017. High and uneven levels of 45S rDNA site-number variation across wild populations of a diploid plant genus (*Anacyclus*, Asteraceae). PloS One. 12:e0187131.

Ross H.A., Murugan S., Sibon Li W.L. 2008. Testing the reliability of genetic methods of species identification via simulation. Syst. Biol. 57:216–230.

Royal Botanic Gardens Kew. 2016. State of the World's Plants, 2016. Royal Botanic Gardens Kew.

Ruhsam M., Rai H.S., Mathews S., Ross T.G., Graham S.W., Raubeson L.A., Mei W., Thomas P.I., Gardner M.F., Ennos R.A. 2015. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? Mol. Ecol. Resour. 15:1067–1078.

Särkinen T., Staats M., Richardson J.E., Cowan R.S., Bakker F.T. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PloS One. 7:e43808.

Schippmann U., Leaman D.J., Cunningham A. 2002. Impact of cultivation and gathering of medicinal plants on biodiversity: global trends and issues. Biodivers. Ecosyst. Approach Agric. For. Fish.

Schlag E.M., McIntosh M.S. 2006. Ginsenoside content and variation among and within American ginseng (*Panax quinquefolius* L.) populations. Phytochemistry. 67:1510–1519.

Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593.

Schmickl R., Liston A., Zeisek V., Oberlander K., Weitemier K., Straub S.C.K., Cronn R.C., Dreyer L.L., Suda J. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). Mol. Ecol. Resour. 16:1124–1135.

Schmiderer C., Lukas B., Ruzicka J., Novak J. 2017. What Else Is in *Salviae officinalis folium*? Comprehensive Species Identification of Plant Raw Material by DNA Metabarcoding. Planta Med.

Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 27.

Schönswetter P., Suda J., Popp M., Weiss-Schneeweiss H., Brochmann C. 2007. Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. Mol. Phylogenet. Evol. 42:92–103.

Schumer M., Rosenthal G.G., Andolfatto P. 2014. How common is homoploid hybrid speciation? Evolution. 68:1553–1560.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Seethapathy G.S., Ganesh D., Kumar J.U.S., Senthilkumar U., Newmaster S.G., Ragupathy S., Shaanker R.U., Ravikanth G. 2015. Assessing product adulteration in natural health products for laxative yielding plants, *Cassia*, *Senna*, and *Chamaecrista*, in Southern India using DNA barcoding. Int. J. Legal Med. 129:693–700.

Sgamma T., Lockie-Williams C., Kreuzer M., Williams S., Scheyhing U., Koch E., Slater A., Howard C. 2017. DNA barcoding for industrial quality assurance. Planta Med.

Shendure J., Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol. 26.

Shi F.-X., Li M.-R., Li Y.-L., Jiang P., Zhang C., Pan Y.-Z., Liu B., Xiao H.-X., Li L.-F. 2015. The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). BMC Plant Biol. 15:297.

Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12:e1005896.

Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298.

Solís-Lemus C., Yang M., Ané C. 2016. Inconsistency of Species Tree Methods under Gene Flow. Syst. Biol. 65:843–851.

Soltis P.S., Marchant D.B., Van de Peer Y., Soltis D.E. 2015. Polyploidy and genome evolution in plants. Curr. Opin. Genet. Dev. 35:119–125.

Soltis P.S., Soltis D.E. 2009. The Role of Hybridization in Plant Speciation. Annu. Rev. Plant Biol. 60:561–588.

Staats M., Arulandhu A.J., Gravendeel B., Holst-Jensen A., Scholtens I., Peelen T., Prins T.W., Kok E. 2016. Advances in DNA metabarcoding for food and wildlife forensic species identification. Anal. Bioanal. Chem. 408:4615–4630.

Staats M., Erkens R.H., van de Vossenberg B., Wieringa J.J., Kraaijeveld K., Stielow B., Geml J., Richardson J.E., Bakker F.T. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS One. 8:e69189.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.

Stanford P.K. 1995. For pluralism and against realism about species. Philos. Sci. 62:70–91.

Starin D. 2012. Salepi extinction, salepi survival: how a change in ingredients could help safeguard orchids. Am Orch Soc Bull. 81:490–494.

Staub P., Casu L., Leonti M. 2016. Back to the roots: A quantitative survey of herbal drugs in Dioscorides' De Materia Medica (ex Matthioli, 1568). Phytomedicine. 23:1043–1052.

Stoeckle M.Y., Gamble C.C., Kirpekar R., Young G., Ahmed S., Little D.P. 2011. Commercial Teas Highlight Plant DNA Barcode Identification Successes and Obstacles. Sci Rep. 1.

Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99:349–364.

Sun Y., Skinner D., Liang G., Hulbert S. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor. Appl. Genet. 89:26–32.

Taberlet P., Bonin A., Zinger L., Coissac E. 2018. Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Taberlet P., Coissac E., Pompanon F., Brochmann C., Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21:2045–2050.

Taberlet P., Coissac E., Pompanon F., Gielly L., Miquel C., Valentini A., Vermat T., Corthier G., Brochmann C., Willerslev E. 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 35:e14.

Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Taleb M.S. 2017. Aromatic and Medicinal Plants in Morocco: Diversity and Socio-Economic Role. World Acad. Sci. Eng. Technol. Int. J. Biol. Biomol. Agric. Food Biotechnol. Eng. 11:742–746.

The Plant List. 2018. The Plant List. Available from http://www.theplantlist.org/.

Thompson J.D., Gauthier P., Papuga G., Pons V., Debussche M., Farris E. 2017. The conservation significance of natural hybridisation in Mediterranean plants: from a case study on *Cyclamen* (Primulaceae) to a general perspective. Plant Biol.

Thomson S.A., Pyle R.L., Ahyong S.T., Alonso-Zarazaga M., Ammirati J., Araya J.F., Ascher J.S., Audisio T.L., Azevedo-Santos V.M., Bailly N., Baker W.J., Balke M., Barclay M.V.L., Barrett R.L., Benine R.C., Bickerstaff J.R.M., Bouchard P., Bour R., Bourgoin T., Boyko C.B., Breure A.S.H., Brothers D.J., Byng J.W., Campbell D., Ceríaco L.M.P., Cernák I., Cerretti P., Chang C.-H., Cho S., Copus J.M., Costello M.J., Cseh A., Csuzdi C., Culham A., D'Elía G., d'Udekem d'Acoz C., Daneliya M.E., Dekker R., Dickinson E.C., Dickinson T.A., van Dijk P.P., Dijkstra K.-D.B., Dima B., Dmitriev D.A., Duistermaat L., Dumbacher J.P., Eiserhardt W.L., Ekrem T., Evenhuis N.L., Faille A., Fernández-Triana J.L., Fiesler E., Fishbein M., Fordham B.G., Freitas A.V.L., Friol N.R., Fritz U., Frøslev T., Funk V.A., Gaimari S.D., Garbino G.S.T., Garraffoni A.R.S., Geml J., Gill A.C., Gray A., Grazziotin F.G., Greenslade P., Gutiérrez E.E., Harvey M.S., Hazevoet C.J., He K., He X., Helfer S., Helgen K.M., van Heteren A.H., Hita Garcia F., Holstein N., Horváth M.K., Hovenkamp P.H., Hwang W.S., Hyvönen J., Islam M.B., Iverson J.B., Ivie M.A., Jaafar Z., Jackson M.D., Jayat J.P., Johnson N.F., Kaiser H., Klitgård B.B., Knapp D.G., Kojima J., Kõljalg U., Kontschán J., Krell F.-T., Krisai-Greilhuber I., Kullander S., Latella L., Lattke J.E., Lencioni V., Lewis G.P., Lhano M.G., Lujan N.K., Luksenburg J.A., Mariaux J., Marinho-Filho J., Marshall C.J., Mate J.F., McDonough M.M., Michel E., Miranda V.F.O., Mitroiu M.-D., Molinari J., Monks S., Moore A.J., Moratelli R., Murányi D., Nakano T., Nikolaeva S., Noyes J., Ohl M., Oleas N.H., Orrell T., Páll-Gergely B., Pape T., Papp V., Parenti L.R., Patterson D., Pavlinov I.Y., Pine R.H., Poczai P., Prado J., Prathapan D., Rabeler R.K., Randall J.E., Rheindt F.E., Rhodin A.G.J., Rodríguez S.M., Rogers

D.C., Roque F. de O., Rowe K.C., Ruedas L.A., Salazar-Bravo J., Salvador R.B., Sangster G., Sarmiento C.E., Schigel D.S., Schmidt S., Schueler F.W., Segers H., Snow N., Souza-Dias P.G.B., Stals R., Stenroos S., Stone R.D., Sturm C.F., Štys P., Teta P., Thomas D.C., Timm R.M., Tindall B.J., Todd J.A., Triebel D., Valdecasas A.G., Vizzini A., Vorontsova M.S., de Vos J.M., Wagner P., Watling L., Weakley A., Welter-Schultes F., Whitmore D., Wilding N., Will K., Williams J., Wilson K., Winston J.E., Wüster W., Yanega D., Yeates D.K., Zaher H., Zhang G., Zhang Z.-Q., Zhou H.-Z. 2018. Taxonomy based on science is necessary for global conservation. PLOS Biol. 16:e2005075.

Ticktin T. 2004. The ecological implications of harvesting non-timber forest products. J. Appl. Ecol. 41:11–21.

Tiwari N.N., Poudel R.C., Uprety Y. 2004. Study on domestic market of medicinal and aromatic plants (MAPs) in Kathmandu Valley. Kathmandu Winrock Int.

Valentini A., Pompanon F., Taberlet P. 2009. DNA barcoding for ecologists. Trends Ecol. Evol. 24:110–117.

Ved D., Goraya G. 2007. Demand and supply of medicinal plants in India. NMPB New Delhi FRLHT Bangalore India. 18.

Veldman S., Gravendeel B., Otieno J.N., Lammers Y., Duijm E., Nieman A., Bytebier B., Ngugi G., Martos F., van Andel T.R. 2017. High-throughput sequencing of African chikanda cake highlights conservation challenges in orchids. Biodivers. Conserv. 26:2029–2046.

Veldman S., Otieno J., van Andel T., Gravendeel B., de Boer H.J. 2014a. Efforts urged to tackle thriving illegal orchid trade in Tanzania and Zambia for chikanda production. Traffic Bull. 26:47–50.

Veldman S., Otieno J., Gravendeel B., van Andel T., de Boer H. 2014b. Conservation of endangered wild harvested medicinal plants: use of DNA barcoding. Nov Plant Bioresour Appl Food Med Cosmet.:81–88.

Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston A. 2014. Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics. Appl. Plant Sci. 2:1400042.

Wen J., Ickert-Bond S.M., Appelhans M.S., Dorr L.J., Funk V.A. 2015. Collections-based systematics: Opportunities and outlook for 2050. J. Syst. Evol. 53:477–488.

Willerslev E., Davison J., Moora M., Zobel M., Coissac E., Edwards M.E., Lorenzen E.D., Vestergard M., Gussarova G., Haile J., Craine J., Gielly L., Boessenkool S., Epp L.S., Pearman P.B., Cheddadi R., Murray D., Brathen K.A., Yoccoz N., Binney H., Cruaud C., Wincker P., Goslar T., Alsos I.G., Bellemain E., Brysting A.K., Elven R., Sonstebo J.H., Murton J., Sher A., Rasmussen M., Ronn R., Mourier T., Cooper A., Austin J., Moller P., Froese D., Zazula G., Pompanon F., Rioux D., Niderkorn V., Tikhonov A., Savvinov G., Roberts R.G., MacPhee R.D.E., Gilbert M.T.P., Kjaer K.H., Orlando L., Brochmann C., Taberlet P. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature. 506:47–51.

Williams M., McDonnell A., Coburn F., Fuller D., Bagazinski M., Fishbein M. 2016. Determining The Asclepias Phylogenetic Tree Through Combined Target Enrichment & Genome Skimming. Res. Rep. Life Sci. Freshmen Res. Sch. 2.

Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H. 2009. The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci U A. 106.

World Health Organization. 2011. Quality control methods for herbal materials. .

Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252–3255.

Wysoker A., Tibbetts K., Fennell T. 2015. Picard tools. Broad Institute.

Xie Y., Wu G., Tang J., Luo R., Patterson J., Liu S., Huang W., He G., Gu S., Li S., Zhou X., Lam T.-W., Li Y., Xu X., Wong G.K.-S., Wang J. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 30:1660–1666.

Yao H., Song J., Liu C., Luo K., Han J., Li Y., Pang X., Xu H., Zhu Y., Xiao P. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PloS One. 5:e13102.

Yigit E., Hernandez D.I., Trujillo J.T., Dimalanta E., Bailey C.D. 2014. Genome and metagenome sequencing: using the human methyl-

binding domain to partition genomic DNA derived from plant tissues. Appl. Plant Sci. 2:1400064.

Zepeda-Mendoza M.L., Bohmann K., Baez A.C., Gilbert M.T.P. 2016. DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC Res. Notes. 9:255.

Zhang S., Wang R., Zeng W., Zhu W., Zhang X., Wu C., Song J., Zheng Y., Chen P. 2015. Resource investigation of traditional medicinal plant *Panax japonicus* (T. Nees) CA Mey and its varieties in China. J. Ethnopharmacol. 166:79–85.

Zobel M., Davison J., Edwards M.E., Brochmann C., Coissac E., Taberlet P., Willerslev E., Moora M. 2018. Ancient environmental DNA reveals shifts in dominant mutualisms during the late Quaternary. Nat. Commun. 9:139.

PUBLICATIONS AND SUBMITTED MANUSCRIPTS (I-V)

I

II

SCIENTIFIC REPORTS

OPEN

Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding, TLC and HPLC-MS

Received: 8 December 2016
Accepted: 29 March 2017
Published online: 02 May 2017

Ancuta Cristina Raclariu[1,2], Ramona Paltinean[3], Laurian Vlase[4], Aurélie Labarre[1], Vincent Manzanilla[1], Mihael Cristin Ichim[2], Gianina Crisan[3], Anne Krag Brysting[5] & Hugo de Boer[1]

Many herbal products have a long history of use, but there are increasing concerns over product efficacy, safety and quality in the wake of recent cases exposing discrepancies between labeling and constituents. When it comes to St. John's wort (*Hypericum perforatum* L.) herbal products, there is limited oversight, frequent off-label use and insufficient monitoring of adverse drug reactions. In this study, we use amplicon metabarcoding (AMB) to authenticate 78 *H. perforatum* herbal products and evaluate its ability to detect substitution compared to standard methods using thin-layer chromatography (TLC) and high performance liquid chromatography coupled with mass spectrometry (HPLC-MS). *Hypericum perforatum* was detected in 68% of the products using AMB. Furthermore, AMB detected incongruence between constituent species and those listed on the label in all products. Neither TLC nor HPLC-MS could be used to unambiguously identify *H. perforatum*. They are accurate methods for authenticating presence of the target compounds, but have limited efficiency in detecting infrageneric substitution and do not yield any information on other plant ingredients in the products. Random post-marketing AMB of herbal products by regulatory agencies could raise awareness among consumers of substitution and would provide an incentive to manufacturers to increase quality control from raw ingredients to commercialized products.

St. John's wort (*Hypericum perforatum* L.) herbal products are popular in complementary and alternative medicine, and are widely used to treat mild to moderate depression but have a much broader traditional use[1]. These products play an important role in primary healthcare, and their popularity is determined by consumer health concerns, cultural habits and by the belief that they are natural and thus safe[2]. *Hypericum perforatum* is among the top-selling herbs and is sold as over-the-counter (OTC) products in pharmacies, supermarkets, health shops and through e-commerce[3]. Products are typically labeled as natural foods or dietary supplements, and claims regarding their possible health benefits appear on labels and in associated advertising. In 2011, the global market for herbal products was estimated to be US$83 billion[4] with Europe being the largest market.

Lack of standardized methods for quality assessment and the highly competitive market of herbal products has increased the incentive to use substitutes and unlabeled fillers[5–7]. However, adulteration is not necessarily intentional, and herbal products may be altered due to accidental adulteration, misidentification[8] and confusion resulting from vernacular names[9,10]. In any case, the use of unreported ingredients is a serious safety concern as adverse drug reactions cannot be associated to the product label and ingredients[11,12]. Signal detection in herbal adverse drug reactions is greatly impeded if ingredients in the products are not reported on the labels[10].

[1]Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318, Oslo, Norway. [2]NIRDBS/"Stejarul" Research Centre for Biological Sciences, Alexandru cel Bun Street, 6, 610004, Piatra, Neamt, Romania. [3]Department of Pharmaceutical Botany, University of Medicine and Pharmacy "Iuliu Haţieganu", Faculty of Pharmacy, Gheorghe Marinescu Street, 23, 400337, Cluj-Napoca, Romania. [4]Department of Pharmaceutical Technology and Biopharmaceutics, "Iuliu Hatieganu" University of Medicine and Pharmacy, Ion Creanga Street, 8-10, 400010, Cluj-Napoca, Romania. [5]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, P.O. Box 1066 Blindern, 0316, Oslo, Norway. Correspondence and requests for materials should be addressed to H.d.B. (email: hugo.deboer@nhm.uio.no)

The European Medicines Agency (EMA) is the European Union agency that is responsible for the evaluation of medicinal products. However, differences in herbal medicine classification exist between EU/EEA member states and this complicates quality monitoring of these products[11]. EMA does not test the composition of herbal products or verify whether ingredients included on the label are included in the product but delegates this responsibility to the manufacturers of these products. EMA requires quality assurance of herbal substances, preparations and products and specifies the use of macroscopic and microscopic characterization, phytochemical analysis of therapeutic target compounds and markers and assays for toxic constituents such as heavy metals and toxins[13]. EMA suggests that identification tests specific for substitute and adulterant detection either use a combination of separate chromatographic approaches (e.g., HPLC with TLC-densitometry) or combine different approaches into a single procedure (e.g., HPLC-UV, HPLC-MS or GC-MC)[13]. Herbal products are usually highly processed and have numerous ingredients, and applying these methods might not enable the accurate identification of all plant ingredients, especially if target species are admixed with other species within the same genera. To complement traditional identification methods, the EMA, as well as the United States Food and Drug Administration (FDA), support the use of innovative analytical technologies such as DNA barcoding.

DNA barcoding is a validated molecular identification method that can provide species-level resolution that is commonly used in authentication of taxonomic provenance of herbal products[5–7,14]. Sanger sequencing based DNA barcoding studies have revealed widespread levels of substitution: 6% in saw palmetto herbal dietary supplements[15], 16% in ginkgo products[16], 25% in black cohosh[17], 33% in herbal teas[18], and 50% in ginseng[19]. A blind test of 44 herbal products sold in North America using DNA barcoding[7] found that 59% contained species not listed on the labels, and only two out of twelve screened companies had products free of substitution, contamination or unreported fillers[7]. High-throughput sequencing based amplicon metabarcoding (AMB)[20] studies can provide insights into species composition of complex mixtures of DNA such as processed herbal products. For example, in a study by Coghlan *et al.*[5], the species composition of 15 highly processed traditional Chinese medicines (TCM) were evaluated using high-throughput sequencing and found that these contained species and genera included on CITES appendices I and II. Other AMB studies have shown similar concerns of varying quality and product label-content fidelity. For instance, Ivanova *et al.*[21] found that 15 tested herbal supplements contained non-listed, non-filler plant DNA and Cheng *et al.*[22] showed that the quality of 27 tested herbal preparations was highly affected by the presence of contaminants.

In this study, we investigated complex herbal products containing St. John's wort *(Hypericum perforatum*), marketed in the EU/EEA both as herbal food supplement and herbal drug. St. John's wort has traditional indications in nervous system, psychiatric, gastrointestinal, hepatobiliary, renal and urinary, respiratory, thoracic, endocrine, musculoskeletal, metabolism and nutritional disorders, as well as in infections and infestations[23]. The mode of action of the major responsible bioactive compounds of *H. perforatum* is still not completely known, but it seems that they act in a synergetic manner to achieve the clinical effectiveness[24]. However, several studies showed that the antidepressant activity is associated mainly to the phloroglucinol derivative hyperforin[25] and the naphthodianthrones hypericin and pseudohypericin[26]. The quantity and quality of active constituents in *Hypericum* herbal medicines are highly affected by the manufacturing process[1]. Treatments involving St. John's wort are generally safe, but several studies show that use in combination with other drugs can cause potentially life-threatening adverse drug reactions due to pharmacokinetic interactions[3]. These adverse interactions are still not fully understood despite extensive studies in its mechanisms. *Hypericum* extracts can cause serious side effects when administrated simultaneously with some antidepressants[27]. Other studies reported adverse reactions due to high doses of hypericin in *Hypericum* products leading to a phototoxic effect that may induce photodermatitis and can also decrease or nullify the effect of other drugs when administrated simultaneously[28]. Hall *et al.*[29] reported exacerbated vaginal breakthrough bleeding, and Murphy *et al.*[30] observed evidence of follicle growth and probable ovulation when simultaneously administrating *H. perforatum* with low-dose oral contraceptives.

The aim of this study was to evaluate the ability of using AMB to detect substitution in single and multi-ingredient *Hypericum* herbal products compared with standard identification approaches suggested by the European Pharmacopoeia and the European Medicine Agency, such as HPLC-MS and TLC.

Results

Thin Layer Chromatography. The European Pharmacopoeia sets a minimum concentration of 0.08% total hypericins in the dried drug[31], and includes a TLC based identification assay intended to distinguish *H. perforatum* from other species, including other species within the same genus. The European Pharmacopoeia 8.0[31] *Hyperici herba* monograph TLC test yields four zones, corresponding to rutin, hyperoside, pseudohypericin and hypericin, calibrated by two reference solution compounds, rutin and hyperoside (Fig. 1). This can be used to distinguish between presence of rutin in *H. perforatum* and small quantities of rutin in *H. maculatum* Crantz. Our test control samples included *H. elegans* Stephan ex Willd., *H. maculatum*, *H. olympicum* L., *H. patulum* Thunb., *H. perforatum*, and *H. polyphyllum* Boiss. & Balansa. This testing revealed some remarkable challenges associated with this authentication method (Fig. 1). The chromatograms of *H. olympicum*, *H. patulum*, and *H. polyphyllum* were indistinguishable from those of *H. perforatum*.

The product samples were all tested using the recommended test. Sample extraction for four samples (a tincture, two oils, and one juice) out of 78 failed. The test results could be grouped into three categories: (1) rutin, hyperoside, pseudohypericin and hypericin present. This could indicate presence of *H. perforatum* or other *Hypericum* species with indistinguishable chromatograms, e.g. *H. olympicum*, *H. patulum*, and *H. polyphyllum*; (2) hyperoside, hypericin, and pseudohypericin present, but rutin in low or undetectable concentrations. This made presence of *H. perforatum* or other *Hypericum* species with indistinguishable chromatograms unlikely, but could indicate presence of *H. maculatum* or other *Hypericum* species with indistinguishable chromatograms, for example *H. elegans*; (3) rutin detected or not, and hyperoside, pseudohypericin and hypericin not detected. This ruled out presence of *Hypericum* species. Fifty-five out of 74 (74%) samples contained *H. perforatum* or an

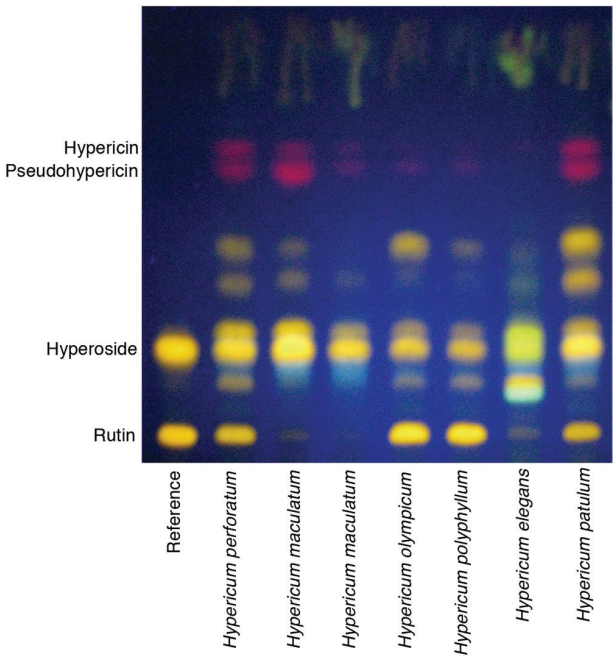


Figure 1. Thin layer chromatogram (TLC) of *Hypericum perforatum* and other *Hypericum* species. The yellow-orange fluorescent bands from the lower third of the chromatogram correspond to rutin and hyperoside, and are used for the identification of *H. perforatum*. This distinguishes between presence of rutin in *H. perforatum* and absence or only small quantities of rutin in *H. maculatum*. The bands corresponding to rutin and hyperoside are found also in *H. olympicum*, *H. patulum*, and *H. polyphyllum*.

indistinguishable adulterant, or a mixture of *Hypericum* species including *H. perforatum* or an indistinguishable adulterant. Nine out of 74 (12%) samples contained *Hypericum* species other than *H. perforatum* or indistinguishable adulterants, or a mixture of those other *Hypericum* species. Ten out of 74 (14%) samples did not contain *Hypericum* species in detectable amounts (Fig. 2A; Supplementary Fig. S1; Supplementary Table S2).

High-Performance Liquid Chromatography-Mass Spectrometry. Detection of *H. perforatum* using HPLC-MS is based on the presence of the two main bioactive compounds in *Hyperici herba*, hyperforin and hypericin. The *Hyperici herba* monograph in the European Pharmacopoeia 8.0[31] states that the dried drug should have a minimum content of total hypericins, expressed as hypericin, corresponding to 0.08% to verify the presence of *H. perforatum*. A literature review of 88 taxa of *Hypericum* with qualitative (74 taxa[32]) and quantitative (27 taxa[33]) measures of hyperforin and hypericin, together with the six test control species measured in this study (Supplementary Table S3), revealed that the hyperforin and hypericin content was not an accurate predictor of *H. perforatum* presence. *Hypericum perforatum* and its main adulterant *H. maculatum* had hypericin contents within similar ranges, but *H. maculatum* had very low content of hyperforin, <0.018%. However, several of the tested or reviewed species had both hyperforin and hypericin contents similar to those of *H. perforatum*, most notably *H. olympicum* and *H. polyphyllum*.

All 78 product samples were analyzed using HPLC-MS, except for four samples that could not be extracted (a tincture, two oils, and one juice). The tested products showed presence of rutin, hyperoside, pseudohypericin and hypericin in 55 out of 74 (74%) samples, which could indicate *H. perforatum* or other *Hypericum* species with indistinguishable chromatograms. In nine out of 74 (12%) samples hyperoside, hypericin, and pseudohypericin were present, but rutin was present in low or undetectable concentrations, which made presence of *H. perforatum* or other *Hypericum* species with indistinguishable chromatograms unlikely, but could indicate presence of *H. maculatum* or other *Hypericum* species with indistinguishable chromatograms, such as *H. elegans*, or a mixture of *Hypericum* species. In ten out 74 (14%) samples rutin was detected or not, and hyperoside, pseudohypericin and hypericin were not detected, which ruled out presence of *Hypericum* species. The test results could further be grouped into three categories: (1) hyperforin content higher than hypericin content. This could indicate presence of *H. perforatum* or other *Hypericum* species with high levels of hyperforin, e.g. *H. elegans*, *H. olympicum* or *H. polyphyllum*; (2) hyperforin content equal to or lower than hypericin content. This could indicate presence of *H. maculatum* or other species with low levels of hyperforin, such as *H. barbatum* Jacq., *H. hirsutum* L., *H. humifusum* L., *H. linarioides* Bosse, *H. richeri* Vill., *H. rumeliacum* Boiss. or *H. tetrapterum* Fr.; (3) no hyperforin or hypericin detected. This indicates that material from *Hypericum* species is absent or that these compounds are present in undetectable amounts. It should be noted that low levels of hyperforin could also indicate low content of *H. perforatum* in the tested product. Sixty-one out of 74 samples (82%) had hyperforin content higher than

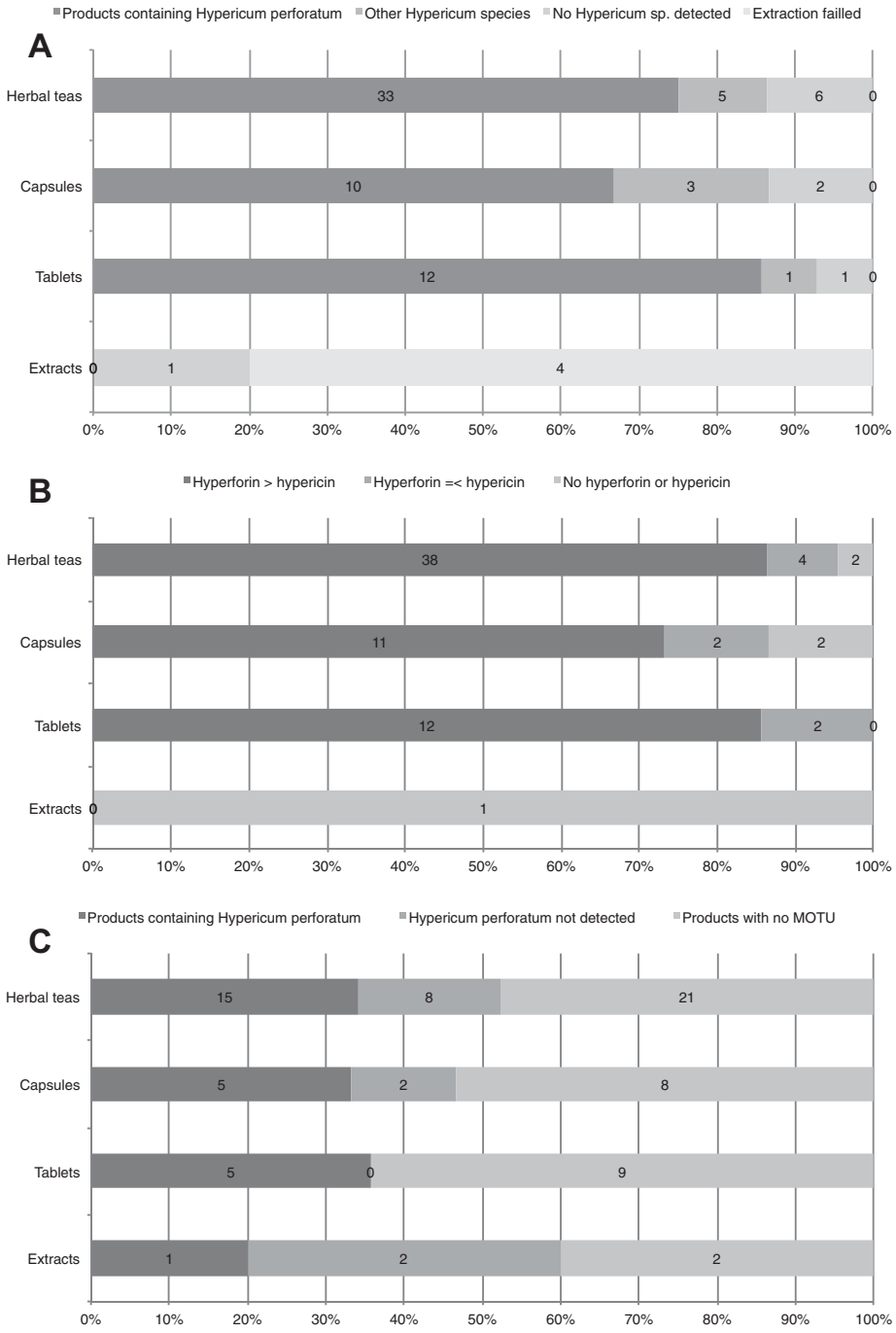


Figure 2. Presence of *Hypericum perforatum* within the products. (**A**) Detection using thin layer chromatogram (TLC). (**B**) Detection using high-performance liquid chromatography-mass spectrometry (HPLC-MS). (**C**) Detection using amplicon metabarcoding (AMB). Detection between methods is not fully comparable as the resolution of the approaches differs.

hypericin content; eight out of 74 (11%) had hyperforin content equal to or lower than hypericin content; and five out of 74 (7%) had no detectable hyperforin or hypericin (Fig. 2B; Supplementary Table S4).

DNA metabarcoding. The DNA extracted from 78 samples was highly variable in quantity and quality. Fragment Analyzer measurements gave results for 47 (60%) samples, with DNA concentration ranging from 0.01 to 140 ng/μl. Thirty-one samples (40%) did not contain measurable DNA concentrations: 11 capsules, eight tablets, eight herbal teas and four extracts. PCR amplification reactions were performed for all 78 samples and amplicons were obtained from 76% of the samples for nrITS1 and 73% for nrITS2. The highest amplification rate was

obtained for herbal teas (97% for nrITS1 and 88% for nrITS2), followed by capsules (62% for nrITS1 and 62% for nrITS2), tablets (43% for nrITS1 and 45% for nrITS2) and extracts (27% for nrITS1 and 40% for nrITS2). There was no significant correlation between sample total DNA concentration and nrITS amplicon concentration or between amplicon concentration and sequenced reads or bases (Supplementary Table S5; Supplementary Fig. S6).

The raw data consisted of 9,416,033 sequences, with an average of 60,359 sequences per sample for each marker. Sequencing success rates were 49%, respectively 44% (34/78 samples) for nrITS1 and 47% (37/78 samples) for nrITS2. A dataset consisting of 1,511,356 reads, fulfilling our trimming and filtering quality criteria, was obtained, including 737,010 nrITS1 and 774,346 nrITS2 reads (on average 19,395 nrITS1 and 20,377 nrITS2 reads per sample). Forty samples out of 78 samples (51%) yielded no molecular operational taxonomic units (MOTUs) for either nrITS1 or nrITS2 and are excluded from the results and discussion (2, 7, 9, 11–14, 16, 18–20, 22, 23, 25, 26, 28, 31, 33, 34, 36, 40, 41, 45, 47, 49, 52–54, 56–58, 60, 62, 63, 64, 66–70). These included 21 herbal teas, eight capsules, nine tablets and two extracts (Supplementary Table S7). The MOTU yielding samples included 23 herbal teas, seven capsules, five tablets and three extracts.

A total of 219 different species were identified using BLAST from the retained MOTUs (Supplementary Table S7; Fig. 3 for MOTUs with >1% normalized reads). For nrITS1 we detected a total of 143 different species and for nrITS2 137 species (Supplementary Fig. S8, Supplementary Fig. S9). Analysis of the BLAST identified MOTUs showed that a total of 34 genera (24%) and 76 species (36%) were detected exclusively with nrITS1. Analysis of nrITS2 showed that 82 (39%) species, but no unique genera, were detected exclusively with this marker. A total of 106 genera (76%) and 55 (26%) species were detected with both markers. Identified MOTUs and their read numbers were merged for both markers per sample for further analysis.

The number of species detected per sample ranged from one to 57, with an average of 18.7 species per sample. Out of the 17 successfully analyzed single ingredient samples, those containing only *H. perforatum* according to the label, none contained only *H. perforatum*, 13 contained more than one species, and four did not contain *H. perforatum*. Out of the 21 successfully analyzed multiple ingredient samples, those containing *H. perforatum* and other species according to the label, none contained only the ingredients listed on the label, two contained all species listed on the label plus additional species not listed on the label, 19 contained fewer species than listed on the label plus additional species, and eight did not contain *H. perforatum* (Fig. 2). The fidelity for *H. perforatum* in single ingredient products was 76% (13 out of 17), and for multi ingredient products, 62% (13 out of 21). The overall ingredient fidelity (detected species from product label/total number of species on label) for multi ingredient products was 41% and for all products 57%. The following four species were found in more than 50% of the samples: *Hypericum perforatum* (68%), *Convolvulus arvensis* L. (63%), *Achillea millefolium* L. (53%) and *Urtica dioica* L. (50%). Plant taxa present in more than 20% of the samples are listed in Table 1. Eleven out of the 78 products were commercialized and registered as herbal medicines, and 4 of these yielded AMB results. *Hypericum perforatum* was found in all four products, and in addition an average of 20.3 other species.

In addition to the target species, *H. perforatum*, that was detected in 68% (26) of the MOTU yielding samples, several other *Hypericum* species were also detected: *H. humifusum* in 21% (8), *H. tetrapterum* in 13% (5) and *H. hirsutum* in 3% (1). Other *Hypericum* species were never detected without *H. perforatum* suggesting that adulteration by admixture is more widespread than complete substitution. The five most common species detected with AMB, but which were not present on the label of the products were: *Agrostis gigantea* Roth (Poaceae) detected in 24%, *Centaurea cyanus* L. (Compositae) in 21%, *Vaccinium myrtillus* L. (Ericaceae) in 21%, *Lolium perenne* L. (Poaceae) in 18%, and *Apera spica-venti* (L.) P. Beauv. (Poaceae) in 16% of the samples. A total of 34 anemophilous (wind-pollinated) species were detected, including 31 grasses and the woody species *Fraxinus excelsior* L., *Humulus lupulus* L., and *Juglans regia* L. (Supplementary Table S7).

Comparative results. In Fig. 2 the detection rates of the three authentication methods are summarized. The detection rate of AMB was much lower than that of the other approaches, but neither TLC nor HPLC-MS could be used to unambiguously identify *H. perforatum*, and it was thus difficult to make an overall comparison of the three methods. Considering only the 26 samples in which *H. perforatum* was detected using AMB, TLC could be used to detect rutin, hyperoside, hypericin and pseudohypericin in 21 samples (81%), and HPLC-MS could be used to detect higher hyperforin than hypericin content in 23 samples (88%). All three methods were in agreement for 19 samples (73%). For the 12 samples that yielded MOTUs using AMB but in which *H. perforatum* was not detected, TLC could be used to detect rutin, hyperoside, hypericin and pseudohypericin in six samples (50%), and HPLC-MS could be used to detect higher hyperforin than hypericin content in six samples (50%). For five samples (42%), the results from TLC and HPLC-MS were in agreement.

Discussion

Misidentification of *H. perforatum* and/or adulteration of products containing *H. perforatum* with other *Hypericum* species, and the common and hyperforin-less *H. maculatum* in particular, requires the use of accurate analytical methods for the quality control of herbal products of *H. perforatum*. Identification of *Hypericum* species is feasible using taxonomic identification keys, but recognition of species in the field is challenging and many *Hypericum* species have superficially similar morphology. Positive identification requires the researchers to study flower, leaf and stem morphology including diagnostic characters that might be absent early or late in the flowering season. Identification is further complicated as *H. perforatum* may occasionally hybridize with other species in the genus, resulting in hybrids that have intermediate morphology and secondary metabolite spectra[34,35]. Eastern Europe is a significant source of *Hypericum* material for European herbal products[36], and here material is mostly wild-harvested and several different species co-occur, which could easily lead to intentional and/or accidental picking of the different species.

Using the TLC based identification assay included in the European Pharmacopoeia on a limited number of reference species included to check the accuracy of the test, show that several species have indistinguishable

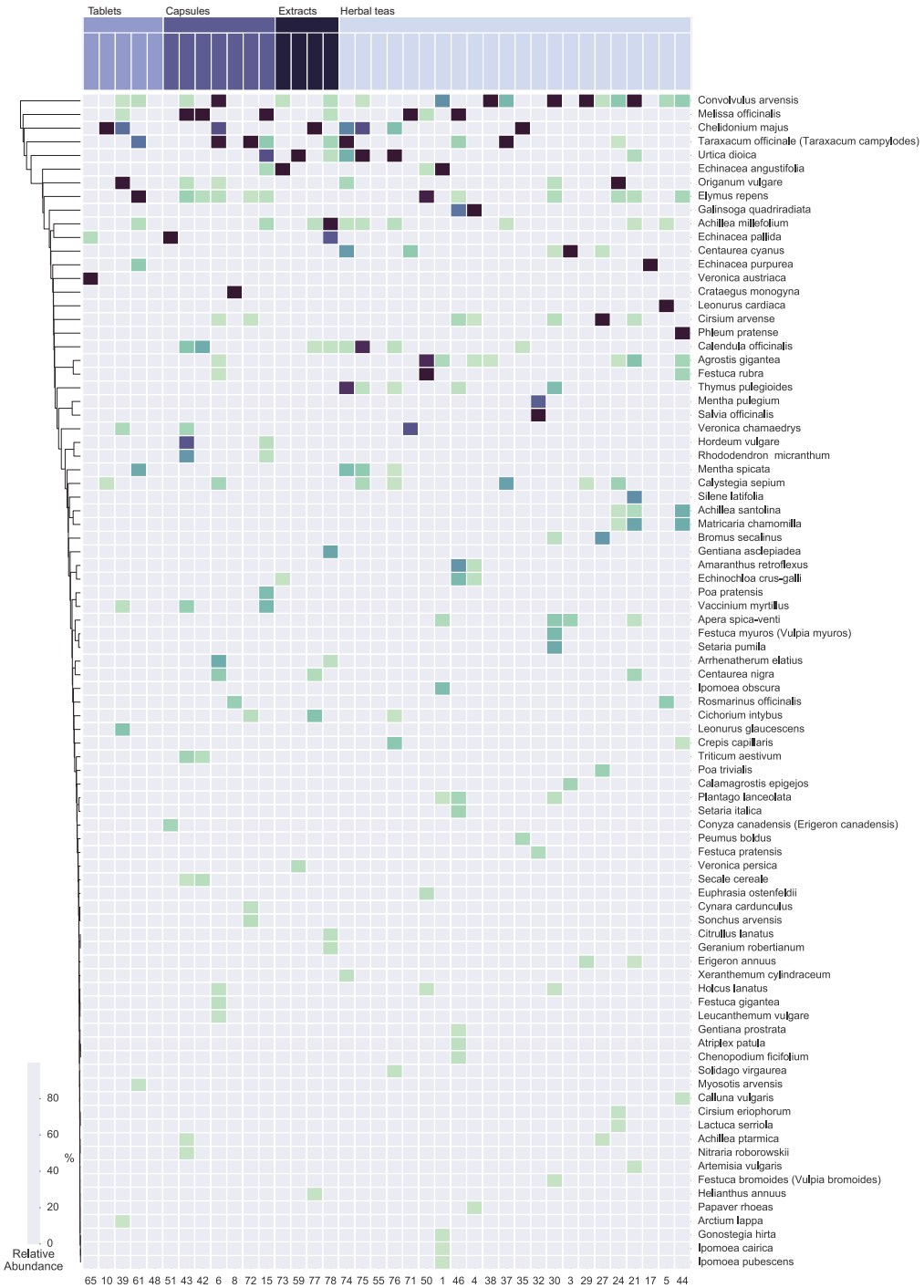


Figure 3. Species identified within the products using amplicon metabarcoding (AMB). Only MOTUs with >1% normalized read numbers per sample are shown. Species are colored according to relative abundance of normalized read numbers. Products are grouped by product form: herbal teas, capsules, tablets/pills/pastilles and extracts/tinctures/oils.

chromatograms from that of *H. perforatum* (Fig. 1). In addition to the ambiguous results with regard to different *Hypericum* species, the test would show the same chromatogram for an admixture of *H. perforatum* and *H. maculatum* as for an unadulterated *H. perforatum* sample. In addition, the TLC assay does not provide useful information about the concentrations of the main bioactive compounds, hyperforin and hypericin.

An alternative to testing for indicator compounds using TLC is to use HPLC-MS to detect and measure the content of the main bioactive secondary metabolites. This method does not claim to distinguish *H. perforatum*

Species	Family	Occurrence
Hypericum perforatum L.	Hypericaceae	68%
Convolvulus arvensis L.	Convolvulaceae	63%
Achillea millefolium L.	Compositae	53%
Urtica dioica L.	Urticaceae	50%
Elymus repens (L.) Gould	Poaceae	47%
Taraxacum campylodes G.E.Haglund	Compositae	47%
Melissa officinalis L.	Lamiaceae	39%
Calendula officinalis L.	Compositae	37%
Chelidonium majus L.	Papaveraceae	37%
Cirsium setidens (Dunn) Nakai	Compositae	34%
Mentha spicata L.	Lamiaceae	29%
Plantago lanceolata L.	Plantaginaceae	29%
Sambucus nigra L.	Adoxaceae	29%
Calystegia sepium (L.) R. Br.	Convolvulaceae	26%
Origanum vulgare L.	Lamiaceae	26%
Agrostis gigantea Roth	Poaceae	24%
Thymus pulegioides L.	Lamiaceae	24%
Viola tricolor L.	Violaceae	24%
Centaurea cyanus L.	Compositae	21%
Cichorium intybus L.	Compositae	21%
Cynara cardunculus L.	Compositae	21%
Echinacea angustifolia DC.	Compositae	21%
Hypericum humifusum L.	Hypericaceae	21%
Vaccinium myrtillus L.	Ericaceae	21%

Table 1. Plant species detected by amplicon metabarcoding (AMB) in more than 20% of the products.

from other *Hypericum* species. Reviews state that the hyperforin content in *H. perforatum* is 2.0–4.5%, and the hypericin content 0.1–0.15% in fresh material[37,38]. This is supported by several recent studies that have found hyperforin and hypericin contents of 1.4% and 0.5%, respectively, in the flowers and of 0.4–1.4% and 0.015–0.17%, respectively, in the vegetative parts[39,40]. Both the hyperforin and hypericin contents vary greatly between studies, probably based on the nature and quality of the analyzed plant material[41]. To check the accuracy of HPLC-MS for distinguishing *H. perforatum*, we analyzed seven samples representing six *Hypericum* species, and compiled literature on hyperforin and hypericin values from 27 species[33,39,40,42]. The results show that the hyperforin and hypericin contents vary considerably between and within species (Supplementary Table S3). *Hypericum perforatum* and its main adulterant *H. maculatum* have hypericin contents within similar ranges, but *H. maculatum* has very low content of hyperforin, 0.004–0.018%. This compilation highlights the low predictive value of hyperforin and hypericin content in determining substitution in *H. perforatum* preparations. Several species have similar hyperforin and hypericin contents to *H. perforatum*, most notably *H. olympicum* and *H. polyphyllum*.

The categorization of hyperforin and hypericin content per product into three groups made it possible to conclude that five products out of 74 (7%) have no detectable hyperforin or hypericin, eight out of 74 (11%) have low content or absence of *H. perforatum*, and 61 out of 74 samples (82%) have spectra typical of *H. perforatum* but also several other species. The 74 analyzed samples had hypericin contents ranging from 0–0.03%, and none passed the minimum threshold set by the European Pharmacopoeia (Supplementary Table S4). Wurglics *et al.*[43] showed that commercial products bought in Germany have total hypericin contents ranging from 0.16–0.30%, and thus exceeding the European Pharmacopoeia minimum. The reason for the discrepancy between these and our results is not entirely clear, especially as both have been measured using similar methodology, but could be a result of lower product quality in our study material. In summary, HPLC-MS is not an accurate method for detection of substitution, adulteration or admixture, but suitable for control of bioactive compound content in products, and thus important in quality control for consumer safety. A superior, but more cost-intensive approach, than TLC and HPLC-MS is NMR metabolomics that enables chemical fingerprinting encompassing a huge range of target molecules[44].

Several studies have shown the resolution and efficacy of DNA metabarcoding[20] for identifying plant species diversity in a range of products[5,6,21,22,45–47]. Comparative identification of processed food and pharmaceutical products is challenging as compared to substrates that can be used for morphological identification, such as pollen clumps[46,48] and pollen in honey[45]. The advantage of pharmaceutical products and traditional and complementary alternative medicines is that these have their putative contents printed on the package[5,6,21,22]. Galimberti *et al.*[48], Richardson *et al.*[46], and Hawkins *et al.*[45] used *rbc*L and *trn*H-*psb*A, and nrITS2 and *rbc*L, respectively, to analyze DNA from pollen in pollen grains and honey to investigate honey bee foraging preferences. Cheng *et al.*[22] used amplicon metabarcoding to analyze nine traditional Chinese medicines (TCMs) and detected on average 4.8

species using nrITS2 and 2.8 using *trn*L. Coghlan *et al.*[5,6] analyzed TCMs for presence of both animal and plant ingredients and found over 68 plant families and eight vertebrate genera in these products. Ivanova *et al.*[21] used universal nrITS primers and found a host of plant species in eight herbal supplements, as well as many fungi due to specificity of these primers in amplifying fungal nrITS. In this study, plant specific nrITS primers were used to amplify nrITS1 and nrITS2, and an average of 18.7 species were detected per sample. In addition, the presence of *H. perforatum* was detected in 26 out of 38 sequenced samples (68%). These findings corroborate previous results that amplicon metabarcoding is an effective way to investigate species composition in products that contain a mixture of DNA from different species. Looking at the subset of registered herbal medicines that yielded AMB data, all four were found to contain *H. perforatum*, and this was supported by TLC results for all four, and by HPLC results in only two cases.

The relatively low success rate (49%) after applying strict read quality and filtering criteria makes this method challenging to use for routine screening at this time. Ivanova *et al.*[21] reported a slightly higher success rate with eight out of 15 samples (53%) for nrITS2 AMB of herbal products from *Echinacea*, *Gingko*, *Hypericum*, *Trigonella*, and *Valeriana*. Cheng *et al.*[22] reported a 100% success rate for 30 individual samples of TCM, all of which were unprocessed crude drugs. The varying degrees of success probably reflect the quality and type raw material, but also the many details in the analysis that can be varied to optimize the results, roughly in order of significance: extraction procedures and purification, primers, markers, identification approach, clustering, MOTU thresholds, sequencing platform, filtering, quality thresholds and chimera removal, library preparation, and amplification protocols.

Quantifying contamination is important when focusing on the tolerated levels of foreign matter in herbal pharmaceuticals. Quantifying relative species abundances based on sequence read numbers from samples with unknown ingredients is hampered by several factors. Firstly, AMB relies on the availability of DNA, but plant DNA can be removed or highly degraded during the harvesting, drying, storage, transportation, and processing (e.g., mode of extraction, irradiation, ultraviolet light exposure, heat or pressure, filtration, extractive distillation or supercritical fluid extraction)[49]. Secondly, AMB is a PCR-based method, and variation in nrITS copy number, primer annealing, and amplification bias all influence the number of taxon-specific reads[50]. Thirdly, incomplete reference databases and sequences with incorrect species names, can render taxonomic identifications prone to uncertainties.

AMB detected the target species, *H. perforatum*, in 68% of the samples, but in addition other *Hypericum* species were detected, *H. humifusum* in 21% (8), *H. tetrapterum* in 13% (5) and *H. hirsutum* in 3% (1) of the samples. *Hypericum tetrapterum* belongs to the same taxonomic section as *H. perforatum* but is more closely related to *H. maculatum*, whereas *H. humifusum* and *H. hirsutum* belong to sections Oligostema and Hirtella s.l., respectively[51,52]. These species can co-occur in natural habitats from which *H. perforatum* is wild crafted. These other *Hypericum* species were never detected without *H. perforatum* suggesting that adulteration by admixture is more widespread than complete substitution.

The overlooked species diversity through poor primer fit and amplification bias is difficult to quantify but some diversity is likely missed[53]. The detection of additional plant species, other that the ones from the label or those that can be expected as substitutes, contaminants or fillers, may be explained by (1) amplified PCR chimeras; (2) false-positive BLAST identifications due to incomplete or error-prone reference databases; or (3) presence of pollen from anemophilous (wind-pollinated) species. Thirty-four of the MOTUs belong to anemophilous species, including 31 grasses of which *Agrostis gigantea* Roth (Poaceae) was detected in 24% of samples, *Lolium perenne* L. (Poaceae) in 18%, and *Apera spica-venti* (L.) P.Beauv. (Poaceae) in 16%. DNA from pollen from these species can end up in the products through co-occurrence with the collected material in natural habitats as well as during other steps in the process chain.

Conclusions

The metabarcoding results confirm that AMB can be used to test for the presence of *H. perforatum* and simultaneously to detect substitution, adulteration and/or admixture of other species. These results corroborate with previous results that show the usefulness of metabarcoding for use in complementing traditional methods of quality control for consumer safety[5,6,21,22]. It should be emphasized however that the relatively low success rate of generating sequence reads per product makes this method challenging to use for routine screening, as only 49% of the tested products yielded sequences as compared to 95% of products that could be used for TLC and HPLC-MS analyses. Moreover, the high sensitivity of AMB in detecting everything from grass pollen on field-collected plants to plant dust left in production equipment requires a careful consideration of the concept of contamination. AMB results for herbal pharmaceutical authentication should be interpreted with a focus on presence and or absence of target species, i.e. the labeled ingredients, but alarms need not be raised over trace contaminations from species plausibly present in the cultivation, transport or production chain. However, the TLC and HPLC-MS results show that these methods are of limited applicability with regard to detecting species substitution, but may be used efficiently to detect target compounds. The clear advantage of HPLC-MS over TLC is the ability to quantify constituents and to screen for a vastly larger number of compounds. If product safety relies on threshold levels of specific bioactive compounds, absence of toxins, allergens and admixed pharmaceuticals, then chemical analysis methods are more relevant than DNA based composition analysis, but if product fidelity, species substitution or adulteration is suspected then the latter method outperforms in terms of resolution. Development of novel molecular markers and approaches for genomic barcoding are likely to increase the resolution of DNA barcoding in species-level identification. Several others have advocated the use of DNA barcoding and metabarcoding in herbal product authentication and herbal pharmacovigilance[5,7,9,21,22,54–56] and adoption of standards for quality control by regulatory agencies could raise product quality and increase consumer confidence.

Methods

Sample collection. Seventy-eight herbal products that included *Hypericum perforatum* according to the label were randomly purchased in European countries (and Turkey), including Romania (51), Germany (5), Poland (4), Turkey (4), Slovakia (3), Spain (2), UK (2), Austria (2), Czech Republic (1), France (1), Italy (1), Sweden (1) and the Netherlands (1). The samples were bought from pharmacies (44), herbal shops (25), super markets (2) or via e-commerce (7), and were sold as herbal teas (44), capsules (15), tablets (14) and extracts (5). According to the label information, the products included 38 single ingredient products, 33 products contained between two and ten ingredients and seven products contained more than ten ingredients. These medicinal products for scientific analyses were imported into Norway under Norwegian Medicines Agency license no. 16/04551–2. An overview of the samples including label information, but not the producer/importer name, lot number, expiration date or any other information that could lead to the identification of that specific product can be found in Supplementary Table S10.

For the phytochemical analysis, aerial parts of *H. elegans* (voucher ARNHM01He), *H. maculatum* (ARNHM01Hm), *H. maculatum* (92151 UMF CLUJ), *H. olympicum* (ARNHM01Ho), *H. patulum* (ARNHM01Hpa), *H. perforatum* (92141 UMF CLUJ) and *H. polyphyllum* (ARNHM01Hpo) were used as references for the identification and quantification of the main compounds. The reference species were selected to include the main adulterant, *H. maculatum* (two geographically isolated samples from Romania and Norway), species with known high levels of hyperforin and hypericin, *H. polyphyllum* and *H. olympicum*, as well as random species collected during fieldwork in Romania, *H. patulum* and *H. elegans*. All voucher specimens are deposited in the Herbarium of the Alexander Borza Botanical Garden (CL) of Babes-Bolyai University, Cluj-Napoca, Romania.

Phytochemical analysis. *Thin layer chromatography (TLC).* Samples were processed according to the *Hyperici herba* monograph in the European Pharmacopoeia 8.0[31]. Control solutions were prepared by mixing 500 mg of ground identified and vouchered material of selected *Hypericum* species with 10 ml of methanol (Analytical grade, Chimreactiv SRL, Romania) for 10 min at 60 °C. After cooling, the obtained solution was filtered. The reference solution was obtained by dissolving 5 mg of hyperoside (Analytical grade, Sigma-Aldrich) and 5 mg of rutin (Analytical grade, Sigma-Aldrich) in 5 ml of methanol. The test solutions of the *Hypericum* products were prepared in the same way as the control solution. Herbal products were processed depending on the pharmaceutical formulation and following the principles of the same extraction procedure, which were adapted to each pharmaceutical formulation. The analysis of the samples was performed in triplicate on SilicaGel plates (60 G F_{254}, 20 × 20 cm, Merck), in twin bands of 10 mm, consisting of a 10 µl test sample and a 5 µl reference solution. As a mobile phase, a mixture of formic acid (Analytical grade, Nordic):distilled water:ethyl acetate (Analytical grade, Lachner) (6:9:90 V/V/V) was used. After migration of the principal components, the plates were dried at 100–105 °C for 10 min. Detection was performed in UV light, at 365 nm, after spraying with a mixture of 1% methanolic diphenylboryloxyetylamine (Sigma-Aldrich) and 5% methanolic polyethylene glycol 400 (Sigma-Aldrich) and 30 min incubation.

High-Performance Liquid Chromatography-Mass Spectrometry (HPLC-MS). *Hypericum* extracts were prepared by adding 750 mg of powdered plant material of *Hypericum* species, *H. perforatum* and *H. maculatum*, to 15 ml methanol in a glass tube[31,57]. The tubes were capped and agitated in the dark at 25 °C for 3 hours on a digital ceramic hotplate stirrer (Arec. X. Velp Scientifica). The extracts were filtered and diluted (1:100) in the mobile phase consisting of a mixture of 1 mM acetonitrile (Merck)/ammonium acetate (Merck) 45/50 (V/V) in double distilled, deionised water (Infusion Solution Laboratory of the University of Medicine and Pharmacy Cluj-Napoca - Romania), and 1 µl of the mixture was injected into the HPLC chromatographic system. Quantities of the herbal product test samples were individually adapted in order to identify and quantify the reference compounds and the extraction procedure was subsequently followed as described above for the references. The HPLC system used was an 1100 series Agilent Technologies model (Darmstadt, Germany) consisting of a G1312A binary pump, an in-line G1379A degasser, a G1329A autosampler, a G1316A column thermostat and an Agilent Ion Trap Detector 1100 SL. Chromatographic separation was performed on a Zorbax SB-C18 (50 mm × 2.1 mm i.d., 3.5 µm) column (Agilent Technologies) equipped with a Zorbax SB-C18 precolumn with the mobile phase above, at 45 °C with a flow rate of 0.6 ml/min. The detection of analytes was performed in triplicate in non-reactive MS^2 mode for the quantification of hypericin (Hwi Analytik Gmbh) or in reactive MS^2 mode for hyperforin (Sigma), negative ion ionisation, using an ion trap mass spectrometer equipped with an electrospray ionisation ion source (ESI): capillary +2500 V, nebulizer 40 psi (nitrogen), dry gas nitrogen at 8 l/min, dry gas temperature 350 °C.

Standard calibration curves were obtained by plotting the peak areas of standard concentrations of hypericin (10, 20, 50, 100, 200 and 500 ng/ml) and hyperforin (2, 4, 10, 20, 40 and 100 ng/ml) against their nominal concentrations. Two linear regression equations ($R^2 > 0.998$) were obtained. Positive identification of the target compounds was performed by mass-spectrometry, and quantification of hypericin and hyperforin was based on peak area (RT, retention time of 1.1 and 2.3 min, respectively) in comparison with the standard curves.

Genetic analysis. *DNA extraction and quantification.* Total DNA was extracted using a modified CTAB extraction method from small amounts of each herbal product (about 300 mg)[58]. The substrate was homogenized using 2–3 zirconium grinding beads in a Mini-Beadbeater-1 (Biospec Products Inc., USA). The final elution volume was 100 µl, and extracted DNA was quantified using a Fragment Analyzer™ (Advanced Analytical Technologies, Inc., USA) and a DNF-488-33 HS Genomic DNA Reagent Kit (50 bp–40,000 bp).

Amplicon library preparation. All amplicon libraries were prepared using fusion PCR based on two nuclear ribosomal target sequences, internal transcribed spacers nrITS1 and nrITS2. PGM fusion primers were based

on 17SE and 5.8 I1, and 5.8 I2 and 26 SE, respectively[59]. The forward primers were labeled with unique 10 bp multiplex identifier (MID) tags and the reverse primers with uniform truncated P1 (trP1) tags. Thermal cycling was carried out in 25 μl reaction volumes, and each reaction contained 5 μl 5X Q5 reaction buffer (New England Biolabs Inc, UK), 1.5 μl 10 μM of each primer (Biolegio, the Netherlands), 0.5 μl 10 mM dNTPs, 0.25 μl 20 U/μl Q5 High-Fidelity DNA Polymerase (New England Biolabs Inc, UK), 5 μl 5X Q5 High GC enhancer, 10.75 μl of Milli-Q ultrapure water and 0.5 μl of template DNA. The following thermocycling protocol was used: 30 s of initial denaturation at 98 °C, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 30 s, and elongation at 72 °C for 30 s, followed by a final elongation step at 72 °C for 2 min. The annealing temperature was 56 °C for nrITS1, and 71 °C for nrITS2.

Equimolar pool preparation. The size, purity and the molar concentration (nmol/l) of each amplicon library was measured using a Fragment Analyzer™ (Advanced Analytical Technologies, Inc., USA) and a DNF-910 dsDNA Reagent Kit (35 bp–1,500 bp). An equimolar pool (2 ng/μl/library) was prepared from the amplicon libraries using the Biomek 4000 Laboratory Automation Workstation (Beckman Coulter, USA). Agencourt AMPure XP (Beckman Coulter, USA) was used for removal of unincorporated primers and nucleotides using the manufacturer's instructions (Agencourt AMPure XP v. B37419AA). The total concentration of the purified pooled amplicon library stock and three serial dilutions (undiluted, 1/5, 1/10) were analyzed using the Fragment Analyzer™ (Advanced Analytical Technologies, Inc., USA) and DNF-488 High Sensitivity Genomic DNA Analysis Kit in order to identify the optimum concentration range for the template preparation.

High throughput sequencing. An Ion Chef (Life Technologies (LT), Thermo-Fisher Scientific, USA) was used to prepare pooled Ion AmpliSeq libraries (LT) for emulsion PCR and to load the sequencing chips. The input DNA template concentration was adjusted to the number of Ion Sphere Particles (ISPs) and added to the emulsion PCR master mix. The emulsion PCR was done using the Ion Chef, and template-positive ISPs were enriched and loaded on an Ion 318 v2 Chip (LT) and sequenced on an Ion Torrent Personal Genome Machine (LT) using an Ion PGM Sequencing 400 kit (LT). Sequencing read data was analyzed and demultiplexed into FASTQ files per sample using Torrent Suite version 5.0.4 (LT).

Bioinformatics analysis. FASTQ read files were processed using the HTS-barcode-checker pipeline[60] available as a Galaxy pipeline at the Naturalis Biodiversity Center (http://145.136.240.164:8080/). Using the HTS pipeline, nrITS1 and nrITS2 primer sequences were used to demultiplex the sequencing reads per sample and to filter out reads that did not match any of the primers. PRINSEQ[61] was used to determine filtering and trimming values based on read lengths and Phred read quality[62]. All reads with a mean Phred quality score of less than 26 were filtered out, as well as reads with a length of less than 300 bp. Remaining reads were trimmed to a maximum length of 440 bp for nrITS1 and 350 bp for nrITS2. CD-HIT-EST[63] was used to cluster reads into molecular operational taxonomic units (MOTUs) defined by a sequence similarity of >99% and a minimum number of ten reads. The consensus sequences of non-singleton MOTUs were queried using BLAST[64] against a reference nucleotide sequence database, with a maximum e-value of 0.05, a minimum hit length of 100 bp and sequence identity of >97%. The number of reads per MOTU, as well as the BLAST results per MOTU, were compiled using custom scripts from the HTS Barcode Checker pipeline[60]. The reference sequence database consisted of a local copy of the NCBI/GenBank nucleotide database.

Data availability. Ion-Torrent amplicon read data is deposited in DRYAD: doi:10.5061/dryad.32j7r.

References

1. Linde, K. St. John's Wort – an overview. *Forsch. Komplementärmedizin Res. Complement. Med.* **16**, 146–155 (2009).
2. Lynch, N. & Berry, D. Differences in perceived risks and benefits of herbal, over-the-counter conventional, and prescribed conventional, medicines, and the implications of this for the safe and effective use of herbal products. *Complement. Ther. Med.* **15**, 84–91, doi:10.1016/j.ctim.2006.06.007 (2007).
3. Borrelli, F. & Izzo, A. A. Herb–drug Interactions with St John's wort (*Hypericum perforatum*): an update on clinical observations. *AAPS J.* **11**, 710–727, doi:10.1208/s12248-009-9146-8 (2009).
4. WHO. The World medicines situation 2011—Traditional medicines: global situation, issues and challenges. WHO Press, Geneva, Switzerland (2011).
5. Coghlan, M. L. *et al*. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLOS Genet.* **8**, e1002657, doi:10.1371/journal.pgen.1002657 (2012).
6. Coghlan, M. L. *et al*. Combined DNA, toxicological and heavy metal analyses provides an auditing toolkit to improve pharmacovigilance of traditional Chinese medicine (TCM). *Sci. Rep.* **5**, 17475, doi:10.1038/srep17475 (2015).
7. Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S. & Ragupathy, S. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.* **11**, 222, doi:10.1186/1741-7015-11-222 (2013).
8. Saslis-Lagoudakis, C. H. *et al*. Identification of common horsetail (*Equisetum arvense* L.; Equisetaceae) using thin layer chromatography versus DNA barcoding. *Sci. Rep* **5**, 11942, doi:10.1038/srep11942 (2015).
9. De Boer, H. J., Ichim, M. C. & Newmaster, S. G. DNA barcoding and pharmacovigilance of herbal medicines. *Drug Saf.* **38**, 611–620, doi:10.1007/s40264-015-0306-8 (2015).
10. Farah, M. H. *et al*. Botanical nomenclature in pharmacovigilance and a recommendation for standardisation. *Drug Saf.* **29**, 1023–1029, doi:10.2165/00002018-200629110-00002 (2006).
11. Gilbert, N. Regulations: Herbal medicine rule book. *Nature* **480**, S98–S99, doi:10.1038/480S98a (2011).
12. Heubl, G. New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Planta Med.* **76**, 1963–1974, doi:10.1055/s-0030-1250519 (2010).
13. European Medicines Agency. *EMEA/CVMP/815/00 - Guideline on specifications: test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products*, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003393.pdf (2006).

14. De Boer, H. J., Ouarghidi, A., Martin, G., Abbad, A. & Kool, A. DNA barcoding reveals limited accuracy of dentifications based on folk taxonomy. *PLOS ONE* **9**, e84291, doi:10.1371/journal.pone.0084291 (2014).
15. Little, D. P. & Jeanson, M. L. DNA barcode authentication of Saw Palmetto herbal dietary supplements. *Sci. Rep.* **3**, 3518, doi:10.1038/srep03518 (2013).
16. Little, D. P. Authentication of *Ginkgo biloba* herbal dietary supplements using DNA barcoding. *Genome* **57**, 513–516, doi:10.1139/gen-2014-0130 (2014).
17. Baker, D. A., Stevenson, D. W. & Little, D. P. DNA barcode identification of black cohosh herbal dietary supplements. *J. AOAC Int.* **95**, 1023–1034, doi:10.5740/jaoacint.11-261 (2012).
18. Stoeckle, M. Y. *et al*. Commercial teas highlight plant DNA barcode identification successes and obstacles. *Sci. Rep.* **1**, 42, doi:10.1038/srep00042 (2011).
19. Wallace, L. J. *et al*. DNA barcodes for everyday life: Routine authentication of natural health products. *Food Res. Int.* **49**, 446–452, doi:10.1016/j.foodres.2012.07.048 (2012).
20. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050, doi:10.1111/j.1365-294X.2012.05470.x (2012).
21. Ivanova, N. V., Kuzmina, M. L., Braukmann, T. W. A., Borisenko, A. V. & Zakharov, E. V. Authentication of herbal supplements using next-generation sequencing. *PLOS ONE* **11**, e0156426, doi:10.1371/journal.pone.0156426 (2016).
22. Cheng, X. *et al*. Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for Liuwei Dihuang Wan. *Sci. Rep.* **4**, 5147, doi:10.1038/srep05147 (2014).
23. European Medicines Agency. *EMA/HMPC/101303/2008 - Assessment report on Hypericum perforatum L., herba* http://www.ema.europa.eu/docs/en_GB/document_library/Herbal_-_HMPC_assessment_report/2010/01/WC500059144.pdf (2009).
24. Butterweck, V. & Schmidt, M. St. John's wort: Role of active compounds for its mechanism of action and efficacy. *Wien. Med. Wochenschr.* **157**, 356–361, doi:10.1007/s10354-007-0440-8 (2007).
25. Wonnemann, M., Singer, A., Siebert, B. & Müller, W. E. Evaluation of synaptosomal uptake inhibition of most relevant constituents of St. John's wort. *Pharmacopsychiatry* **34**, 148–151, doi:10.1055/s-2001-15465 (2001).
26. Butterweck, V., Korte, B. & Winterhoff, H. Pharmacological and endocrine effects of Hypericum perforatum and hypericin after repeated treatment. *Pharmacopsychiatry* **34**, 2–7, doi:10.1055/s-2001-15508 (2001).
27. Bonetto, N., Santelli, L., Battistin, L. & Cagnin, A. Serotonin syndrome and rhabdomyolysis induced by concomitant use of triptans, fluoxetine and hypericum. *Cephalalgia Int. J. Headache* **27**, 1421–1423, doi:10.1111/j.1468-2982.2007.01430.x (2007).
28. Parker, V., Wong, A. H., Boon, H. S. & Seeman, M. V. Adverse reactions to St John's wort. *Can. J. Psychiatry Rev. Can. Psychiatr.* **46**, 77–79, doi:10.1177/070674370104600112 (2001).
29. Hall, S. D. *et al*. The interaction between St John's wort and an oral contraceptive. *Clin. Pharmacol. Ther.* **74**, 525–535, doi:10.1016/j.clpt.2003.08.009 (2003).
30. Murphy, P. A., Kern, S. E., Stanczyk, F. Z. & Westhoff, C. L. Interaction of St. John's Wort with oral contraceptives: effects on the pharmacokinetics of norethindrone and ethinyl estradiol, ovarian activity and breakthrough bleeding. *Contraception* **71**, 402–408, doi:10.1016/j.contraception.2004.11.004 (2005).
31. Council of Europe. *European Pharmacopoeia. 8th Edition*. Strasbourg: Council Of Europe (2014).
32. Crockett, S. L., Schaneberg, B. & Khan, I. A. Phytochemical profiling of new and old world *Hypericum* (St. John's Wort) species. *Phytochem. Anal.* **16**, 479–485, doi:10.1002/(ISSN)1099-1565 (2005).
33. Smelcerovic, A., Zuehlke, S., Spiteller, M., Raabe, N. & Özen, T. Phenolic constituents of 17 *Hypericum* species from Turkey. *Biochem. Syst. Ecol.* **36**, 316–319, doi:10.1016/j.bse.2007.09.002 (2008).
34. Mártonfi, P., Repčák, M. & Mihoková, L. *Hypericum maculatum* Crantz subsp. *maculatum* x *H. perforatum* L. (Hypericaceae): Corroboration of natural hybridization by secondary metabolite analysis. *Folia Geobot.* **31**, 245–250, doi:10.1007/BF02812067 (1996).
35. Noack, K. L. Geschlechtsverlust und Bastardierung beim Johanniskraut. *Forsch Fortschr* **17**, 13–15 (1941).
36. Kathe, W., Honnef, S. & Heym, A. *Medicinal and aromatic plants in Albania, Bosnia-Herzegovina, Bulgaria, Croatia and Romania*. Bundesamt für Naturschutz-Skripten 91. Bonn: Federal Agency for Nature Conservation (2003).
37. Barnes, J., Anderson, L. A. & Phillipson, J. D. St John's wort (*Hypericum perforatum* L.): a review of its chemistry, pharmacology and clinical properties. *J. Pharm. Pharmacol.* **53**, 583–600, doi:10.1211/0022357011775910 (2001).
38. Greeson, J. M., Sanford, B. & Monti, D. A. St. John's wort (*Hypericum perforatum)*: a review of the current pharmacological, toxicological, and clinical literature. *Psychopharmacology (Berl.)* **153**, 402–414, doi:10.1007/s002130000625 (2001).
39. Smelcerovic, A. & Spiteller, M. Phytochemical analysis of nine *Hypericum* L. species from Serbia and the F.Y.R. Macedonia. *Int. J. Pharm. Sci.* **61**, 251–252 (2006).
40. Umek, A., Kreft, S., Kartnig, T. & Heydel, B. Quantitative phytochemical analyses of six Hypericum species growing in Slovenia. *Planta Med.* **65**, 388–390, doi:10.1055/s-2006-960798 (1999).
41. Hölzl, J. & Ostrowski, H. Johanniskraut (*Hypericum perforatum* L.) HPLC-Analyse der wichtigen Inhaltsstoffe und deren Variabilität in einer Population. *Dtsch. Apoth.-Ztg.* **127**, 1227–1230 (1987).
42. Smelcerovic, A. *et al*. Phytochemical analysis and genetic characterization of six *Hypericum* species from Serbia. *Phytochemistry* **67**, 171–177, doi:10.1016/j.phytochem.2005.10.021 (2006).
43. Wurglics, M. *et al*. Comparison of German St. John's wort products according to hyperforin and total hypericin content. *J. Am. Pharm. Assoc. Washington DC 1996* **41**, 560–566, doi:10.1016/S1086-5802(16)31280-3 (2001).
44. Kim, H. K., Choi, Y. H. & Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **5**, 536–549, doi:10.1038/nprot.2009.237 (2010).
45. Hawkins, J. *et al*. Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLOS ONE* **10**, e0134735, doi:10.1371/journal.pone.0134735 (2015).
46. Richardson, R. T. *et al*. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Appl. Plant Sci.* **3**, 1400066, doi:10.3732/apps.1400066 (2015).
47. Galimberti, A. *et al*. DNA barcoding as a new tool for food traceability. *Food Res. Int.* **50**, 55–63, doi:10.1016/j.foodres.2012.09.036 (2013).
48. Sickel, W. *et al*. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecol.* **15**, 20, doi:10.1186/s12898-015-0051-y (2015).
49. Novak, J., Grausgruber-Gröger, S. & Lukas, B. DNA-based authentication of plant extracts. *Food Res. Int.* **40**, 388–392, doi:10.1016/j.foodres.2006.10.015 (2007).
50. Pawluczyk, M. *et al*. Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Anal. Bioanal. Chem.* **407**, 1841–1848, doi:10.1007/s00216-014-8435-y (2015).
51. Meseguer, A. S., Aldasoro, J. J. & Sanmartín, I. Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (*Hypericum*). *Mol. Phylogenet. Evol.* **67**, 379–403, doi:10.1016/j.ympev.2013.02.007 (2013).
52. Nürk, N. M., Madriñán, S., Carine, M. A., Chase, M. W. & Blattner, F. R. Molecular phylogenetics and morphological evolution of St. John's wort (*Hypericum*; Hypericaceae). *Mol. Phylogenet. Evol.* **66**, 1–16, doi:10.1016/j.ympev.2012.08.022 (2013).
53. Aylagas, E., Borja, Á., Irigoien, X. & Rodríguez-Ezpeleta, N. Benchmarking DNA Metabarcoding for biodiversity-based monitoring and assessment. *Front. Mar. Sci.* **3**, 679–688, doi:10.3389/fmars.2016.00096 (2016).

54. Palhares, R. M. *et al*. Medicinal plants recommended by the World Health Organization: DNA barcode Identification associated with chemical analyses guarantees their quality. *PLOS ONE* **10**, e0127866, doi:10.1371/journal.pone.0127866 (2015).
55. Techen, N., Crockett, S. L., Khan, I. A. & Scheffler, B. E. Authentication of medicinal plants using molecular biology techniques to compliment conventional methods. *Curr. Med. Chem.* **11**, 1391–1401, doi:10.2174/0929867043365206 (2004).
56. Techen, N., Parveen, I., Pan, Z. & Khan, I. A. DNA barcoding of medicinal plant material for identification. *Curr. Opin. Biotechnol.* **25**, 103–110, doi:10.1016/j.copbio.2013.09.010 (2014).
57. Cossuta, D. *et al*. Extraction of hyperforin and hypericin from St. John's wort (*Hypericum perforatum* L.) with different solvents. *J. Food Process Eng.* **35**, 222–235, doi:10.1111/jfpe.2012.35.issue-2 (2012).
58. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
59. Sun, Y., Skinner, D. Z., Liang, G. H. & Hulbert, S. H. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theor. Appl. Genet.* **89**, 26–32, doi:10.1007/BF00226978 (1994).
60. Lammers, Y., Peelen, T., Vos, R. A. & Gravendeel, B. The HTS barcode checkerpipeline, a tool for automated detection of illegally traded species from high-throughput sequencing data. *BMC Bioinformatics* **15**, 44, doi:10.1186/1471-2105-15-44 (2014).
61. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864, doi:10.1093/bioinformatics/btr026 (2011).
62. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194, doi:10.1101/gr.8.3.186 (1998).
63. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659, doi:10.1093/bioinformatics/btl158 (2006).
64. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410, doi:10.1016/S0022-2836(05)80360-2 (1990).

Acknowledgements

The authors acknowledge funding from the Romanian - EEA Research Programme operated by the MECS-ANCSI PO under the EEA Financial Mechanism 2009–2014 and project contract number 2SEE/2014.

Author Contributions

A.C.R., A.K.B., M.C.I., H.d.B. conceived the experiment. A.C.R. carried out the molecular lab work. R.P., L.V., G.C. carried out the chemical lab work and analysis. A.C.R., A.L., V.M. carried out the analysis of the molecular lab work. A.C.R. and H.d.B. wrote the manuscript, and all authors have read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-01389-w

Competing Interests: The authors declare that they have no competing interests.

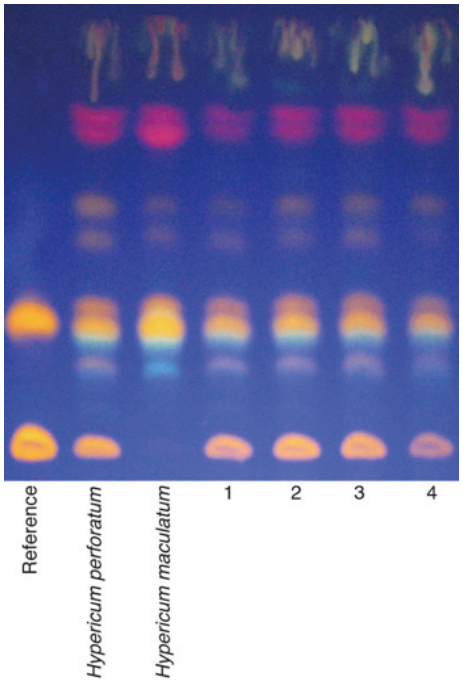
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

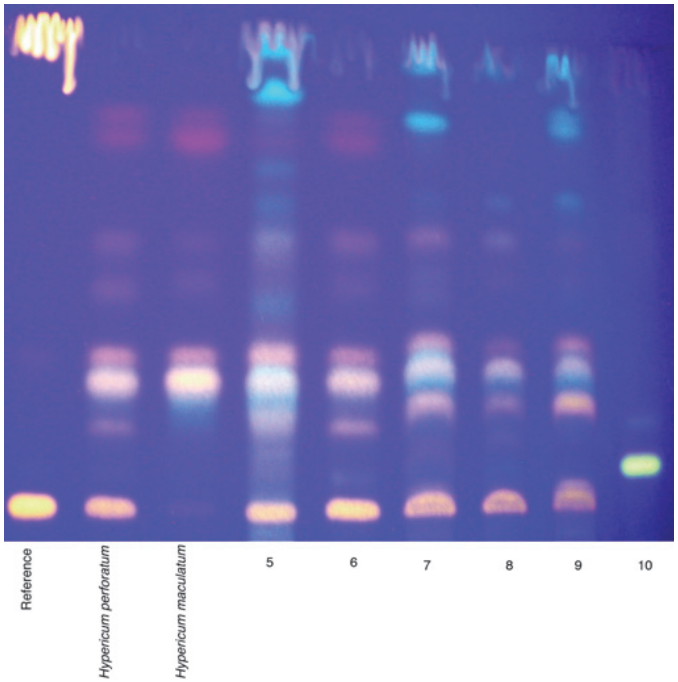
Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS

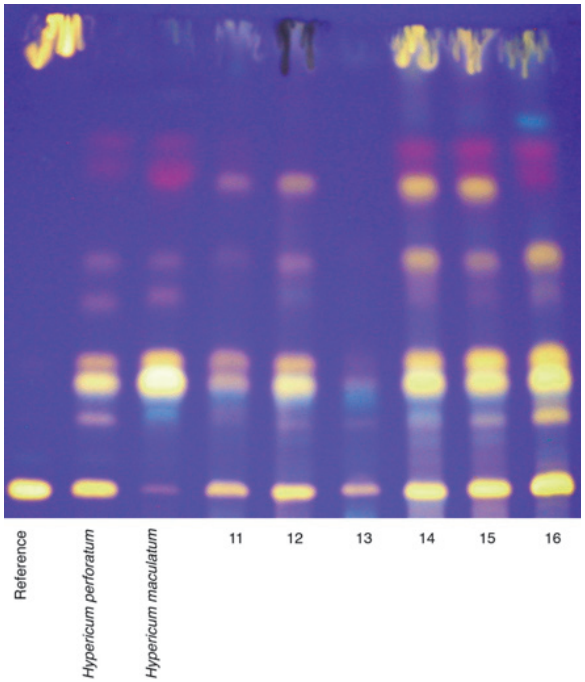
Ancuta Cristina Raclariu, Ramona Paltinean, Laurian Vlase, Aurélie Labarre, Vincent Manzanilla, Mihael Cristin Ichim, Gianina Crisan, Anne Krag Brysting, Hugo de Boer

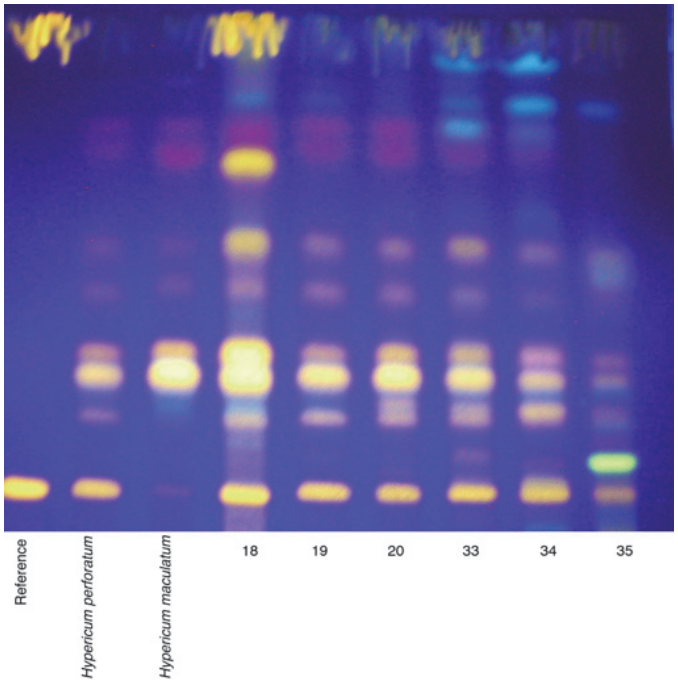
Supplementary Information

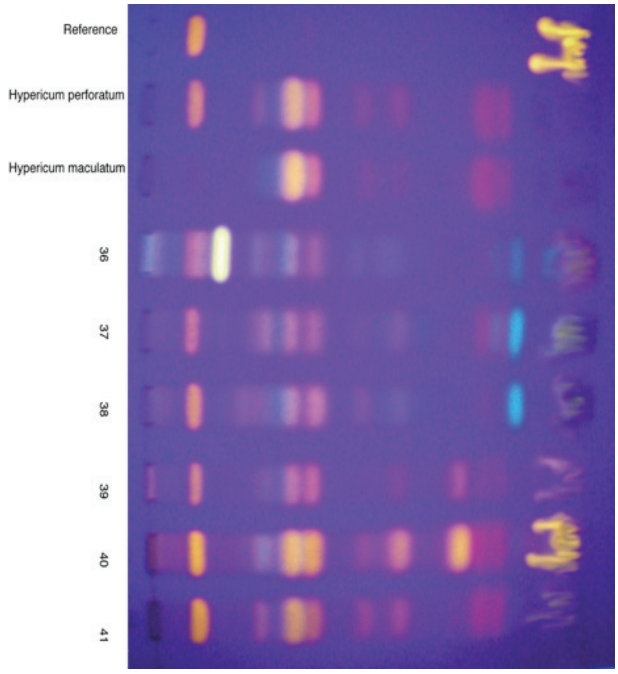
Supplementary Figure S1. TLC chromatograms

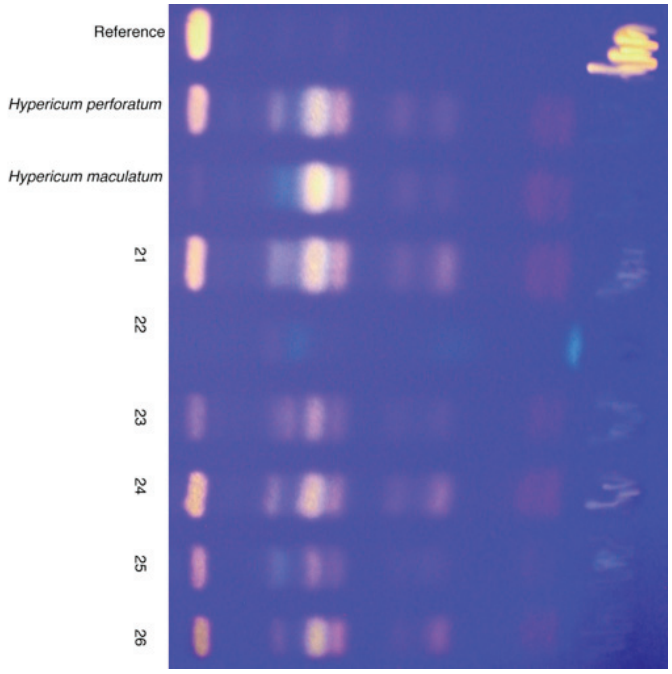


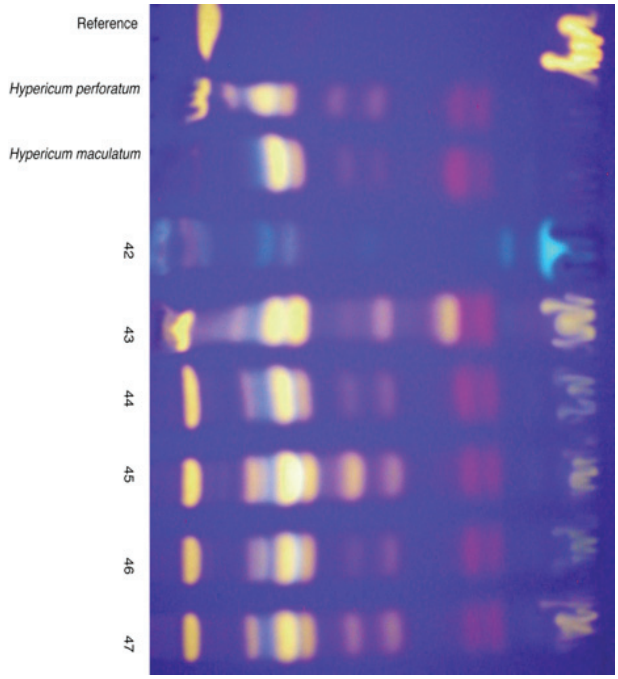


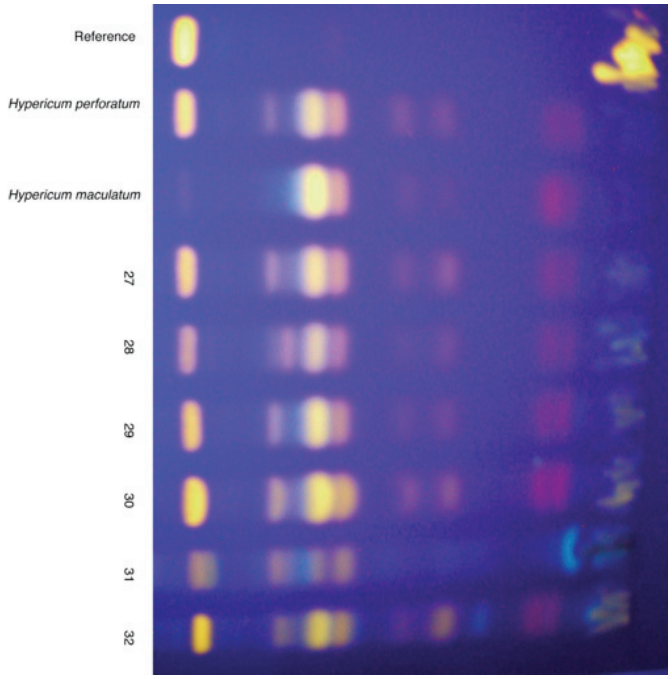


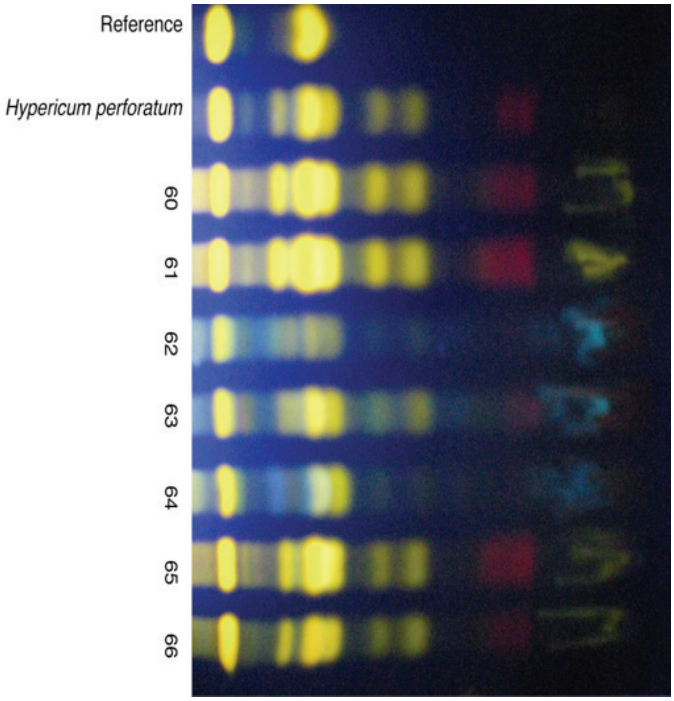


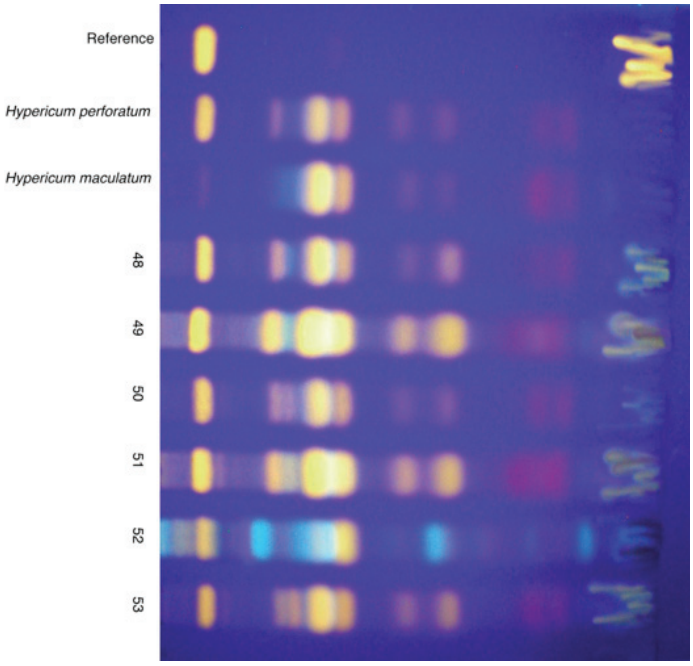

Reference
Hypericum perforatum
Hypericum maculatum
36
37
38
39
40
41

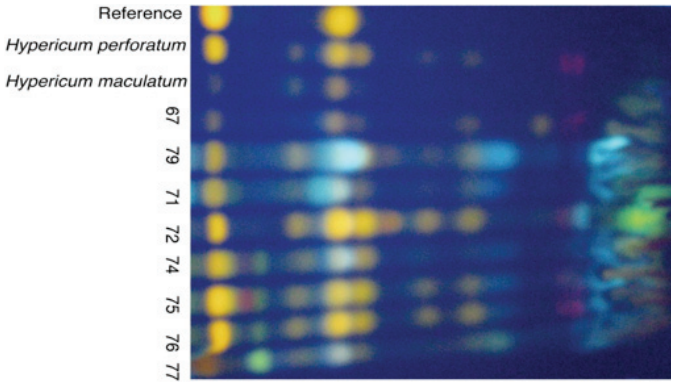

Reference
Hypericum perforatum
Hypericum maculatum
21
22
23
24
25
26

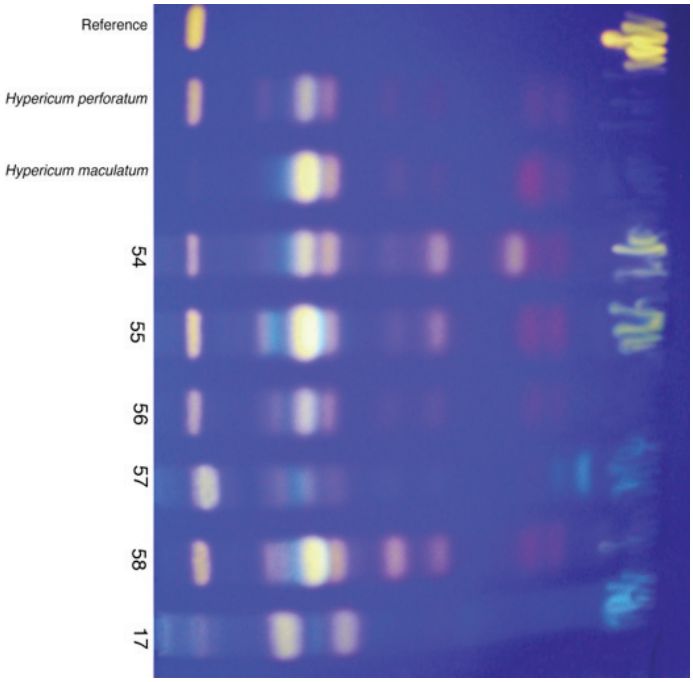

Reference
Hypericum perforatum
Hypericum maculatum
42
43
44
45
46
47


Reference
Hypericum perforatum
Hypericum maculatum
27
28
29
30
31
32


Reference
Hypericum perforatum
60
61
62
63
64
65
66


Reference
Hypericum perforatum
Hypericum maculatum
48
49
50
51
52
53


Reference
Hypericum perforatum
Hypericum maculatum
67
79
71
72
74
75
76
77


Reference
Hypericum perforatum
Hypericum maculatum
54
55
56
57
58
17


Reference
73

Supplementary Table S2. TLC results

Sample no.	Product type	Rutin	Hyperoside	Hypericin	Pseudohypericin	*Hypericum perforatum*	Other *Hypericum species*	No *Hypericum* detected	Extraction failed
1	Herbal tea	+	+	+	+	+			
2	Herbal tea	+	+	+	+	+			
3	Herbal tea	+	+	+	+	+			
4	Herbal tea	+	+	+	+	+			
5	Herbal tea	+	+	±	+	+			
6	Capsules	+	+	+	+	+			
7	Capsules	+	+	-	-		+		
8	Capsules	+	+	-	-		+		
9	Capsules	+	+	-	-		+		
10	Tablets	-	-	-	-			+	
11	Tablets	+	+	+	+	+			
12	Tablets	+	+	-	-		+		
13	Capsules	-	-	-	-			+	
14	Capsules	+	+	+	+	+			
15	Capsules	+	+	+	+	+			
16	Tablets	+	+	+	+	+			
17	Herbal tea	-	-	-	-			+	
18	Tablets	+	+	+	+	+			
19	Herbal tea	+	+	+	+	+			
20	Herbal tea	+	+	+	+	+			
21	Herbal tea	+	+	+	+	+			
22	Herbal tea	-	-	-	-			+	
23	Herbal tea	+	+	+	+	+			
24	Herbal tea	+	+	+	+	+			
25	Herbal tea	+	+	±	±	+			
26	Herbal tea	+	+	+	+	+			
27	Herbal tea	+	+	+	+	+			
28	Herbal tea	+	+	+	+	+			
29	Herbal tea	+	+	+	+	+			
30	Herbal tea	+	+	+	+	+			
31	Herbal tea	+	+	-	±		+		
32	Herbal tea	+	+	+	+	+			
33	Herbal tea	+	+	±	+	+			
34	Herbal tea	+	+	±	+	+			
35	Herbal tea	+	+	±	±		+		
36	Herbal tea	+	±	±	±		+		
37	Herbal tea	+	±	±	+	+			
38	Herbal tea	+	±	+	±	+			
39	Tablets	+	±	+	+	+			
40	Tablets	+	+	+	+	+			
41	Capsules	+	+	+	+	+			
42	Capsules	+	-	-	-			+	
43	Capsules	+	+	+	+	+			
44	Herbal tea	+	+	+	+	+			
45	Herbal tea	+	+	+	+	+			
46	Herbal tea	+	+	+	+	+			
47	Capsules	+	+	+	+	+			
48	Tablets	+	+	+	+	+			
49	Tablets	+	+	+	+	+			
50	Herbal tea	+	+	+	+	+			
51	Capsules	+	+	+	+	+			
52	Herbal tea	+	±	+	±	+			
53	Tablets	+	+	+	+	+			
54	Capsules	+	+	+	+	+			
55	Herbal tea	+	+	+	+	+			
56	Herbal tea	+	+	+	+	+			
57	Herbal tea	+	±	-	±	+			
58	Herbal tea	+	+	+	+	+			
59	Extract								+
60	Tablets	+	+	+	+	+			
61	Tablets	+	+	+	+	+			
62	Herbal tea	+	+	-	-		+		
63	Herbal tea	+	+	±	±	+			
64	Herbal tea	+	+	±	±		+		
65	Tablets	+	+	+	+	+			
66	Capsules	+	+	+	+	+			
67	Tablets	+	+	+	+	+			
68	Extract								+
69	Extract								+
70	Herbal tea	+	-	-	-			+	
71	Herbal tea	+	-	-	-			+	
72	Capsules	+	+	±	±	+			
73	Extract	+	-	-	-			+	
74	Herbal tea	+	-	-	-			+	
75	Herbal tea	+	+	±	±	+			
76	Herbal tea	+	+	+	+	+			
77	Herbal tea	-	-	-	-			+	
78	Extract								+

Supplementary Table S3. Hyperforin and hypericin concentrations measured in different species of *Hypericum* in µg/g dry weight

Species	Hyperforin (µg/g)	Hypericin (µg/g)	Source
H. androsaemum L.	90	0	Smelcerovic et al. 2008
H. aviculariifolium Jaub.& Spach	20	660	Smelcerovic et al. 2008
H. barbatum Jacq.	70	300	Smelcerovic et al. 2006
H. barbatum Jacq.	70	660	Smelcerovic and Spiteller, 2006
Average *H. barbatum*	70	300-660	
H. bithynicum Boiss.	150	1050	Smelcerovic et al. 2008
H. elegans Stephan ex Willd.	82.41	29	This study
H. heterophyllum Vent.	80	510	Smelcerovic et al. 2008
H. hirsutum L.	60	40	Smelcerovic et al. 2006
H. hirsutum L.	50	250	Smelcerovic and Spiteller, 2006
H. hirsutum L. flowers	0	2020	Umek et al. 1999
H. hirsutum L. herb	0	440	Umek et al. 1999
H. hirsutum L.	200	540	Smelcerovic et al. 2008
Average *H. hirsutum*	0-540	40-2020	
H. humifusum L. flowers	0	1640	Umek et al. 1999
H. humifusum L. herb	0	1180	Umek et al. 1999
Average *H. humifusum*	0	117-1640	
H. hyssopifolium Vill.	40	520	Smelcerovic et al. 2008
H. linarioides Bosse	20	20	Smelcerovic et al. 2006
H. linarioides Bosse	0	40	Smelcerovic and Spiteller, 2006
H. linarioides Bosse	0	340	Smelcerovic et al. 2008
Average *H. linarioides*	0-349	20-40	
H. maculatum Crantz	50	30	Smelcerovic et al. 2006
H. maculatum Crantz (RO)	0	142	This study
H. maculatum Crantz	180	70	Smelcerovic and Spiteller, 2006
H. maculatum Crantz (NO)	35	46	This study
H. maculatum Crantz flowers	0	1870	Umek et al. 1999
H. maculatum Crantz herb	0	460	Umek et al. 1999
Average *H. maculatum*	0-180	30-1870	
H. montbretii Spach	3450	740	Smelcerovic et al. 2008
H. montanum L.	0	1130	Smelcerovic et al. 2008
H. montanum L. herb	0	450	Umek et al. 1999
Average *H. montanum*	0-1130	0-450	
H. nummularioides Trautv.	250	200	Smelcerovic et al. 2008
H. olympicum L.	3955	11	This study
H. olympicum L.	20	50	Smelcerovic and Spiteller, 2006
Average *H. olympicum*	20-3955	11-50	
H. origanifolium Willd.	0	0	Smelcerovic et al. 2008
H. patulum Thunb.	8532	153	This study
H. perfoliatum L.	140	290	Smelcerovic et al. 2008
H. perforatum L.	4925	151	This study
H. perforatum L.	3550	170	Smelcerovic and Spiteller, 2006
H. perforatum L. flowers (2)	13590	5200	Umek et al. 1999
H. perforatum L. herb (1)	6010	1690	Umek et al. 1999
H. perforatum L.	5460	3470	Smelcerovic et al. 2008
Average *H. perforatum*	3470-13590	151-5200	
H. orientale L.	30	20	Smelcerovic et al. 2008
H. polyphyllum Boiss. & Balansa	6030	18	This study
H. pruinatum Boiss.& Balansa	50	360	Smelcerovic et al. 2008
H. richeri Vill.	360	450	Smelcerovic and Spiteller, 2006
H. rumeliacum Boiss.	70	180	Smelcerovic et al. 2006
H. rumeliacum Boiss.	150	230	Smelcerovic and Spiteller, 2006
Average *H. rumeliacum*	70-150	180-230	
H. scabrum L.	20	40	Smelcerovic et al. 2008
H. tetrapterum Fr.	110	90	Smelcerovic et al. 2006
H. tetrapterum Fr.	270	150	Smelcerovic and Spiteller, 2006
H. tetrapterum Fr. flowers	0	2230	Umek et al. 1999
H. tetrapterum Fr. herb	0	1100	Umek et al. 1999
Average *H. tetrapterum*	0-270	90-2230	
H. triquetrifolium Turra	50	4560	Smelcerovic et al. 2008

(1) An average of 22 samples. (2) An average of 21 samples.

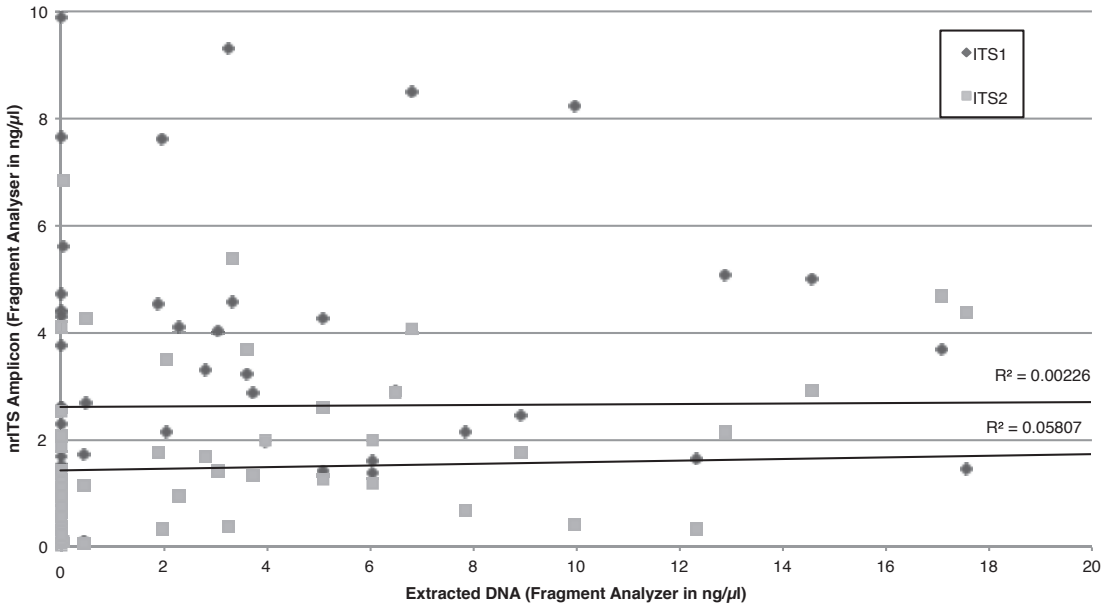
Supplementary Table S4. HPLC-MS results

Sample no.	Product type	Hyperforin (µg/g of herbal product)	Hypericin (µg/g of herbal product)	Total hypericins (%)	Hyperforin absent, but hypericin present (=Hypericum sp.)	Hyperforin and hypericin present (=Hypericum sp.)	Extraction failled	Hyperforin larger than hypericin
1	Herbal tea	444.09	53.14	0.005	0	1	0	1
2	Herbal tea	1333	81.25	0.008	0	1	0	1
3	Herbal tea	2491.83	109.96	0.011	0	1	0	1
4	Herbal tea	1398.09	88.1	0.009	0	1	0	1
5	Herbal tea	109.41	6.59	0.001	0	1	0	1
6	Capsules	1980.92	56.13	0.006	0	1	0	1
7	Capsules	2.79	1.8	0.000	0	1	0	1
8	Capsules	3.65	0	0.000	0	0	0	1
9	Capsules	7	1.3	0.000	0	1	0	1
10	Tablets	16.86	0	0.000	0	0	0	1
11	Tablets	17	4.12	0.000	0	1	0	1
12	Tablets	104.18	1.366	0.000	0	1	0	1
13	Capsules	0	0	0.000	2	0	0	0
14	Capsules	4005.73	287.26	0.029	0	1	0	1
15	Capsules	2421.38	126.15	0.013	0	1	0	1
16	Tablets	580.35	40.82	0.004	0	1	0	1
17	Herbal tea	4.79	0	0.000	0	0	0	1
18	Tablets	342.41	164.82	0.016	0	1	0	1
19	Herbal tea	3868.03	87.95	0.009	0	1	0	1
20	Herbal tea	270.47	90.33	0.009	0	1	0	1
21	Herbal tea	1855.18	82.59	0.008	0	1	0	1
22	Herbal tea	0	0	0.000	2	0	0	0
23	Herbal tea	972.99	43.43	0.004	0	1	0	1
24	Herbal tea	4650.58	162.25	0.016	0	1	0	1
25	Herbal tea	1331.71	45.41	0.005	0	1	0	1
26	Herbal tea	6077.32	90.31	0.009	0	1	0	1
27	Herbal tea	6691.28	63.94	0.006	0	1	0	1
28	Herbal tea	1034.96	74.02	0.007	0	1	0	1
29	Herbal tea	5043.67	106.77	0.011	0	1	0	1
30	Herbal tea	1044.15	97.23	0.010	0	1	0	1
31	Herbal tea	198.07	1.78	0.000	0	1	0	1
32	Herbal tea	63.48	24.87	0.002	0	1	0	1
33	Herbal tea	46.86	27.48	0.003	0	1	0	1
34	Herbal tea	94.68	12.75	0.001	0	1	0	1
35	Herbal tea	83.75	1.71	0.000	0	1	0	1
36	Herbal tea	364.5	6.29	0.001	0	1	0	1
37	Herbal tea	165.38	9.29	0.001	0	1	0	1
38	Herbal tea	51.49	4.52	0.000	0	1	0	1
39	Tablets	17.57	5.23	0.001	0	1	0	1
40	Tablets	56.81	42.53	0.004	0	1	0	1
41	Capsules	0	61.47	0.006	1	0	0	0
42	Capsules	0	0	0.000	2	0	0	0
43	Capsules	674.21	85.15	0.009	0	1	0	1
44	Herbal tea	237.28	106.82	0.011	0	1	0	1
45	Herbal tea	98.81	86.4	0.009	0	1	0	1
46	Herbal tea	31.09	109.02	0.011	0	1	0	0
47	Capsules	0	129.84	0.013	1	0	0	0
48	Tablets	0	13.25	0.001	1	0	0	0
49	Tablets	428.1	53.07	0.005	0	1	0	1
50	Herbal tea	80.48	118.53	0.012	0	1	0	0
51	Capsules	2797.5	298.34	0.030	0	1	0	1
52	Herbal tea	318.52	10.24	0.001	0	1	0	1
53	Tablets	116.69	96.4	0.010	0	1	0	1
54	Capsules	758.06	68.13	0.007	0	1	0	1
55	Herbal tea	0	80.23	0.008	1	0	0	0
56	Herbal tea	1108.58	79.29	0.008	0	1	0	1
57	Herbal tea	80.61	12.3	0.001	0	1	0	1
58	Herbal tea	4925.43	151.97	0.015	0	1	0	1
59	Extract			0.000	0	0	1	0
60	Tablets	31.03	42.61	0.004	0	1	0	0
61	Tablets	89.71	85.74	0.009	0	1	0	1
62	Herbal tea	72	3.58	0.000	0	1	0	1
63	Herbal tea	219.68	15.02	0.002	0	1	0	1
64	Herbal tea	84.6	1.99	0.000	0	1	0	1
65	Tablets	683.63	109.75	0.011	0	1	0	1
66	Capsules	2713.27	92.4	0.009	0	1	0	1
67	Tablets	52	24.78	0.002	0	1	0	1
68	Extract			0.000	0	0	1	0
69	Extract			0.000	0	0	1	0
70	Herbal tea	0	1.83	0.000	1	0	0	0
71	Herbal tea	51.21	1.39	0.000	0	1	0	1
72	Capsules	19.03	6.33	0.001	0	1	0	1
73	Extract	0	0	0.000	2	0	0	0
74	Herbal tea	44.48	2.39	0.000	0	1	0	1
75	Herbal tea	147.92	8.8	0.001	0	1	0	1
76	Herbal tea	689.36	14.66	0.001	0	1	0	1
77	Herbal tea	223	2.43	0.000	0	1	0	1
78	Extract			0.000	0	0	1	0

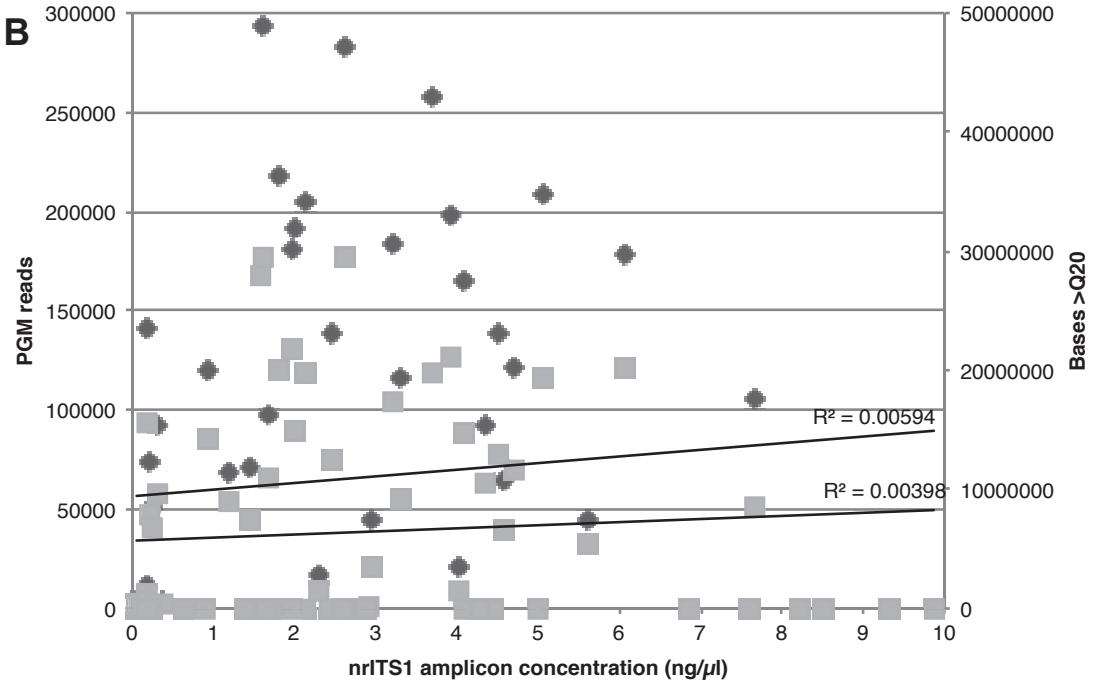
Supplementary Table S5. HTS success

Product	Reads		nrITS1			nrITS2		
No.	No. of bases before demultiplexing	No. of reads before demultiplexing	Amplicon concentration (ng/ul)	No. of reads	No. of bases (>Q20)	Amplicon concentration (ng/ul)	No. of reads	No. of bases (>Q20)
1	62722750	285184	3.70	258020	19830698	4.69	16022	1578942
2	47358	307	4.10	6	514	0.95	238	12464
3	521886	2161	2.91	652	58148	2.88	1361	118469
4	47455333	208055	3.30	115965	9196791	1.68	79050	7101592
5	22231491	145827	2.96	44775	3356973	1.46	80022	3183064
6	12558328	63808	4.02	21329	1564485	1.42	35834	2711535
7	17804	64	1.72	24	2925	1.12	32	2779
8	1386042	5492	0.39	3331	381846	0.25	1724	103955
9	8988	55	1.55	27	1662	0.89	26	1065
10	75383619	278105	6.06	178262	20295668	4.33	83862	0
11	2370	8	0.89	8	891	0.88	0	0
12	35056989	151616	0.15	0	0	0.13	0	0
13	4389	14	0.17	10	1424	0.13	0	0
14	0	0	0.12	0	0	0.16	0	0
15	27614658	86515	1.18	68820	8868686	0.91	15439	1711591
16	0	0	0.25	0	0	0.13	0	0
17	1544147	5029	0.06	3140	336506	0.21	734	86450
18	1903	10	0.12	9	599	0.30	0	0
19	10037	34	1.43	29	2997	2.59	4	513
20	5212	23	2.66	18	1582	4.27	3	9
21	138559130	566977	2.14	205015	19729261	3.50	320385	30508847
22	2028	8	6.84	0	0	1.86	3	293
23	1402	4	8.51	2	272	4.08	2	291
24	83221168	319287	4.71	121339	11588669	1.16	181237	18954323
25	6077	21	1.39	16	1683	1.99	2	266
26	9830	35	4.25	27	2806	1.28	2	37
27	71059267	278718	4.53	138667	13045974	1.74	125391	12717263
28	6294	32	2.14	19	1868	0.69	7	77
29	56684931	226221	3.22	183332	17339018	3.68	29203	2452257
30	53107442	242925	2.00	192134	14919716	3.10	34767	3221126
31	1350	7	2.43	0	0	2.30	0	0
32	29831458	114441	4.57	64443	6636146	5.37	42496	4057305
33	3667	13	4.08	2	116	5.72	1	4
34	22082	95	1.70	71	7040	3.34	12	334
35	22344764	73322	5.61	44708	5422673	6.83	26203	3230857
36	6503	21	4.99	10	1283	2.92	8	838
37	98800823	441576	5.05	208647	19367123	2.14	192205	15149014
38	62730698	259246	4.09	165272	14664712	3.37	77438	7002425
39	43379160	45620	1.67	97249	11054541	1.26	43372	5214116
40	459	2	1.61	0	0	1.17	0	0
41	15171	69	2.62	0	0	1.12	2	303
42	80345560	299829	3.91	198256	21106677	4.52	80502	7903473
43	46950181	168677	4.36	92104	10485352	1.87	67209	6991979
44	164237026	672962	3.76	591566	54135759	4.11	48363	2761743
45	1282	4	9.87	2	244	0.23	2	237
46	102410827	376539	2.62	283073	29390005	2.54	78775	7969442
47	125	2	0.08	1	17	0.04	0	0
48	44196732	150787	0.18	141146	15649176	0.07	4419	299653
49	27291	100	0.64	69	7625	0.19	0	0
50	76087488	287856	7.65	105680	8446677	0.74	171362	19150153
51	25615264	87625	0.18	12048	1326009	0.14	18	8462273
52	4911	33	8.22	8	355	0.41	13	990
53	337	2	0.06	2	101	0.11	0	0
54	2913	13	4.43	0	0	2.05	13	941
55	27503996	94631	0.30	91606	9700289	0.18	103	2256
56	2545	10	2.59	4	339	1.41	4	586
57	21533	154	2.89	0	0	1.33	104	4579
58	8840	36	4.28	24	2557	0.36	6	439
59	5273773	22446	0.20	6584	566608	0.17	15042	1226366
60	1235	6	0.19	1	1	0.39	3	254
61	29752850	118302	0.22	74241	7862582	0.50	39185	2473510
62	8383	28	9.32	6	722	0.39	21	2455
63	5413	17	1.63	8	1154	0.35	1	1
64	7966	32	7.60	0	0	0.33	31	3036
65	17529659	52196	0.25	50670	6663535	0.49	261	28162
66	0	0	0.12	0	0	0.19	0	0
67	0	0	0.12	0	0	0.05	0	0
68	0	0	0.20	0	0	0.15	0	0
69	0	0	0.17	0	0	0.21	0	0
70	10179	41	1.97	29	2816	1.99	0	0
71	55062199	220984	1.45	70947	7534569	4.36	131918	12071501
72	114616133	435823	1.97	180484	21776940	0.29	220087	20829439
73	26683841	106119	2.30	17391	1481069	0.57	83635	8025957
74	265805068	1044831	1.58	303233	28037295	2.21	654831	70358140
75	66354277	298951	2.47	139169	12433535	1.75	131723	10373161
76	86083702	377066	1.82	217576	19943312	0.84	119966	10397432
77	150477835	596181	1.61	292909	29389441	1.40	253969	24631492
78	60778605	202803	0.92	120092	14223598	0.95	72348	8575923

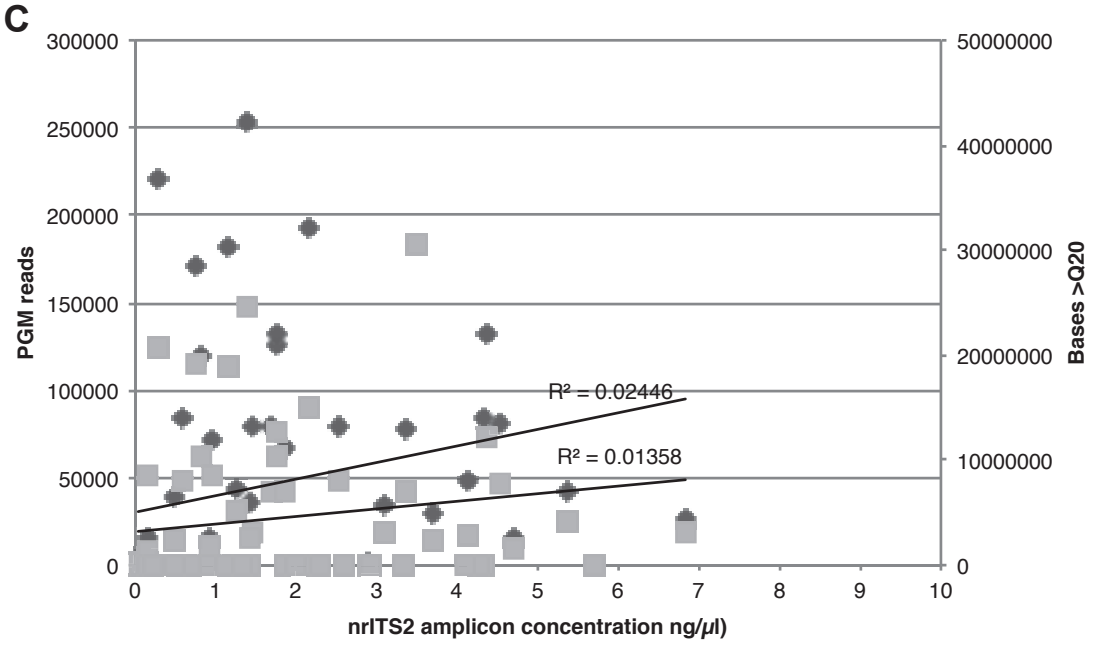
A Supplementary Information Figure S6



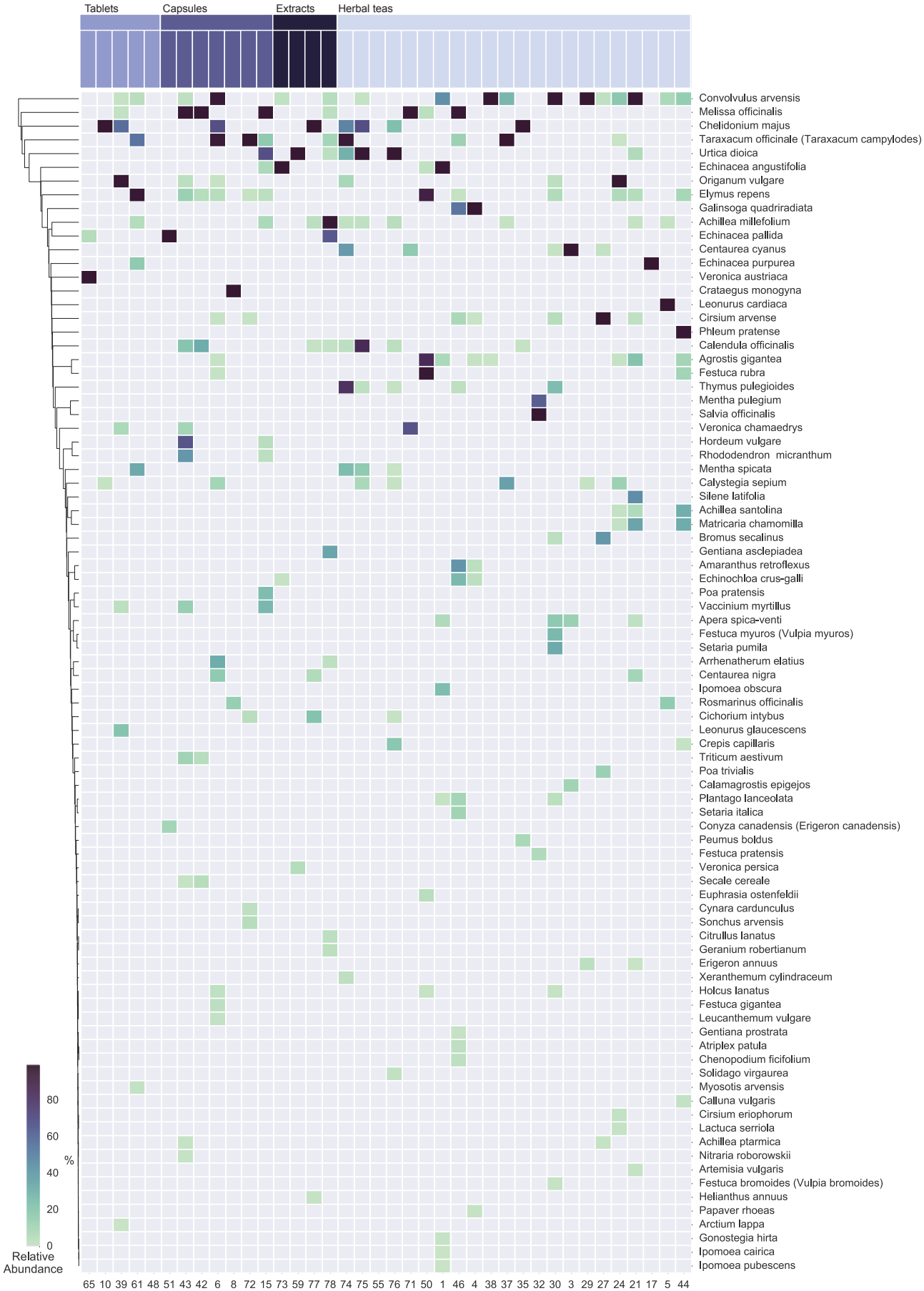
B



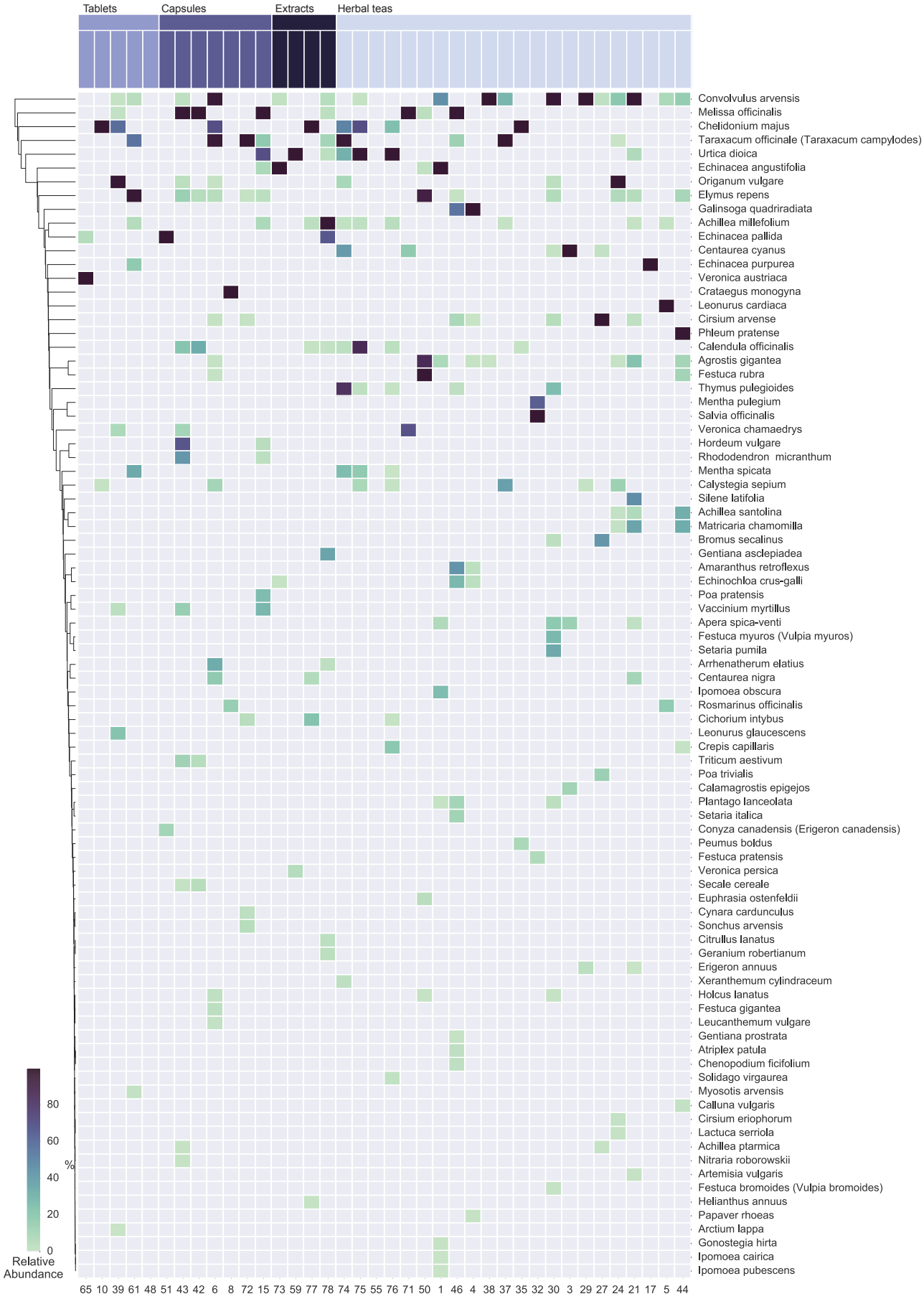
C



Supplementary Figure S8. nrITS1 heatmap of relative abundances of normalized read numbers



Supplementary Figure S9. nrITS2 heatmap of relative abundance of normalized read numbers



Supplementary Table S10. Information about products

Sample no.	Number of species on label	Scientific names of the plant ingredients	Product type	Country of origin	Country of acquisition	Vendor type	Product clasification
1	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Unknown
2	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
3	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
4	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Crataegus monogyna* Jacq.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Glycyrrhiza glabra* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Lavandula angustifolia subsp. angustifolia (syn. Lavandula officinalis* Chaix)	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Leonurus cardiaca* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Rosmarinus officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Sambucus nigra* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
5	10	*Valeriana officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
6	2	*Hypericum perforatum* L.	Capsules	Romania	Romania	E-commerce	Food supplement
6	2	Panax ginseng	Capsules	Romania	Romania	E-commerce	Food supplement
7	7	*Arctium lappa* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Elaeagnus rhamnoides* (L.) A.Nelson *(Hippophae rhamnoides* L.)	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Lycopodium clavatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Salix alba* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Taraxacum officinale* Webb	Capsules	Romania	Romania	Pharmacy	Food supplement
7	7	*Thymus serpyllum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Acorus calamus* var. americanus Raf.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Allium ursinum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Elaeagnus rhamnoides* (L.) A.Nelson *(Hippophae rhamnoides* L.)	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Lycopodium clavatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Mentha pulegium* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
8	7	*Salvia officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Achillea millefolium* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Acorus calamus* var. americanus Raf.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Calendula officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Foeniculum vulgare* Mill.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Mentha × piperita* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Robinia pseudoacacia* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Rubus idaeus* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
9	9	*Taraxacum officinale* Webb	Capsules	Romania	Romania	Pharmacy	Food supplement
10	6	*Chelidonium majus* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Cynara scolymus* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Epilobium hirsutum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Humulus lupulus* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Humulus lupulus* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
10	6	*Valeriana officinalis* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
11	5	*Calendula officinalis* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
11	5	*Foeniculum vulgare* Mill.	Tablets	Romania	Romania	Pharmacy	Food supplement
11	5	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
11	5	*Origanum vulgare* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
11	5	*Plantago* spp.	Tablets	Romania	Romania	Pharmacy	Food supplement
12	2	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
12	2	*Spirulina platensis* (Gomont) Geitler	Tablets	Romania	Romania	Pharmacy	Food supplement
13	5	*Humulus lupulus* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
13	5	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
13	5	*Leonurus cardiaca* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
13	5	*Tilia cordata* Mill.	Capsules	Romania	Romania	Pharmacy	Food supplement
13	5	*Valeriana officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
14	2	*Griffonia simplicifolia* (DC.) Baill.	Capsules	China	Romania	Pharmacy	Food supplement
14	2	*Hypericum perforatum* L.	Capsules	China	Romania	Pharmacy	Food supplement
15	1	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Unknown
16	1	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
17	5	*Foeniculum vulgare* Mill.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
17	5	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
17	5	*Medicago sativa* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
17	5	*Origanum vulgare* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
17	5	*Salvia sclarea* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
18	1	*Hypericum perforatum* L.	Tablets	USA	Romania	Pharmacy	Food supplement
19	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
20	1	*Hypericum perforatum* L.	Herbal teas	Poland	Romania	E-commerce	Food supplement
21	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
22	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
23	1	*Hypericum perforatum* L.	Herbal teas	Poland	Romania	Supermarket	Unknown
24	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Unknown
25	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Supermarket	Unknown
26	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Unknown
27	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
28	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
29	1	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
30	1	*Hypericum perforatum* L.	Herbal teas	Unknown	Romania	E-commerce	Food supplement
31	14	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Alchemilla xanthochlora* Rothm. (syn. *Alchemilla vulgaris* L.)	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Capsella bursa-pastoris* (L.) Medik.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Cichorium intybus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Galium verum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Geranium robertianum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Lamium album* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Leonurus cardiaca* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Matricaria chamomilla* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Salvia officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
31	14	*Thymus serpyllum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
32	7	*Acorus calamus* var. americanus Raf.	Herbal teas	Romania	Romania	E-commerce	Food supplement

32	7	*Allium ursinum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
32	7	*Elaeagnus rhamnoides (L.)* A.Nelson *(Hippophae rhamnoides* L.)	Herbal teas	Romania	Romania	E-commerce	Food supplement
32	7	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
32	7	*Lycopodium clavatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
32	7	*Mentha pulegium* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Salvia officinalis* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Arctium lappa* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Elaeagnus rhamnoides (L.)* A.Nelson *(Hippophae rhamnoides* L.)	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Lycopodium clavatum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Salix alba* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
33	7	*Taraxacum officinale* Webb	Herbal teas	Romania	Romania	E-commerce	Food supplement
33		*Thymus serpyllum* L.	Herbal teas	Romania	Romania	E-commerce	Food supplement
34	9	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Foeniculum vulgare* Mill.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Mentha × piperita* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Origanum vulgare* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Robinia pseudoacacia* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Taraxacum officinale* Webb	Herbal teas	Romania	Romania	Pharmacy	Food supplement
34	9	*Urtica dioica* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
35	8	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Chelidonium majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Cichorium intybus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
35		*Melilotus officinalis* subsp. *alba* (Medik.) H.Ohashi & Tateishi	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Mentha × piperita* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Peumus boldus* Molina	Herbal teas	Romania	Romania	Health shop	Food supplement
35	8	*Silybum marianum* (L.) Gaertn. (syn. *Carduus marianus* L.)	Herbal teas	Romania	Romania	Health shop	Food supplement
36	10	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Agrimonia eupatoria* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Chelidonium majus* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Convolvulus arvensis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Frangula dodonei* Ard.(*Rhamnus frangula* L.)	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Mentha × piperita* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Rosa canina* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Taraxacum officinale* Webb	Herbal teas	Romania	Romania	Pharmacy	Food supplement
36	10	*Zea mays* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
37	4	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
37	4	*Convolvulus arvensis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
37	4	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
37	4	*Taraxacum officinale* Webb	Herbal teas	Romania	Romania	Pharmacy	Food supplement
38	4	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
38	4	*Althaea officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
38	4	*Convolvulus arvensis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
38	4	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
39	5	*Calendula officinalis* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
39	5	*Foeniculum vulgare* Mill.	Tablets	Romania	Romania	Pharmacy	Food supplement
39	5	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
39	5	*Origanum vulgare* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
39	5	Plantago spp.	Tablets	Romania	Romania	Pharmacy	Food supplement
40	2	*Hypericum perforatum* L.	Tablets	Romania	Romania	Pharmacy	Food supplement
40	2	*Lavandula latifolia* Medik. (syn. *Lavandula angustifolia* Moench)	Tablets	Romania	Romania	Pharmacy	Food supplement
41	1	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Aloe vera* (L.) Burm.f. (syn. *Aloe barbadensis* Mill.)	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Calendula officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Citrus paradisi* Macfad.	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Lithothamnion calcareum* (Pallas)	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Melissa officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
42	7	*Passiflora edulis* Sims (syn. *Passiflora incarnata* L.)	Capsules	Romania	Romania	Pharmacy	Food supplement
43	5	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
43	5	*Melissa officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
43	5	*Passiflora edulis* Sims (syn. *Passiflora incarnata* L.)	Capsules	Romania	Romania	Pharmacy	Food supplement
43	5	*Trigonella foenum-graecum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
43	5	*Valeriana officinalis* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
44	1	*Hypericum perforatum* L.	Herbal teas	Unknown	Spain	Health shop	Unknown
45	1	*Hypericum perforatum* L.	Herbal teas	Germany	Germany	Health shop	Herbal medicine
46	1	*Hypericum perforatum* L.	Herbal teas	Germany	Germany	Health shop	Herbal medicine
47	1	*Hypericum perforatum* L.	Capsules	Germany	Germany	Health shop	Herbal medicine
48	1	*Hypericum perforatum* L.	Tablets	Germany	Germany	Health shop	Herbal medicine
49	1	*Hypericum perforatum* L.	Tablets	France	France	Pharmacy	Herbal medicine
50	1	*Hypericum perforatum* L.	Herbal teas	Poland	Poland	E-commerce	Food supplement
51	2	*Hypericum perforatum* L.	Capsules	Poland	Poland	Health shop	Food supplement
51	2	Panax ginseng	Capsules	Poland	Poland	Health shop	Food supplement
52	11	*Avena sativa* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Cirsium oleraceum* (L.) Scop.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Citrus limon* (L.) Osbeck	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	Elsholtzia spp.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Galium verum* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Heracleum mantegazzianum* Sommier & Levier	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Humulus lupulus* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Hypericum perforatum* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Stachys officinalis* (L.) Trevis.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Verbena officinalis* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
52	11	*Veronica officinalis* L.	Herbal teas	Poland	Poland	Health shop	Herbal medicine
53	1	*Hypericum perforatum* L.	Tablets	Poland	Poland	Health shop	Herbal medicine
54	2	*Hypericum perforatum* L.	Capsules	Italy	Italy	Health shop	Food supplement
54	2	Tilia spp.	Capsules	Italy	Italy	Health shop	Food supplement
55	1	*Hypericum perforatum* L.	Herbal teas	USA	Czech Republic	Health shop	Food supplement
56	1	*Hypericum perforatum* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Achillea millefolium* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Elymus repens* (L.) Gould	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Glycyrrhiza glabra* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Hypericum perforatum* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Juniperus communis* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Mentha × piperita* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown

57	9	*Pimpinella anisum* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Taraxacum officinale* Webb	Herbal teas	Austria	Austria	Pharmacy	Unknown
57	9	*Viola tricolor* L.	Herbal teas	Austria	Austria	Pharmacy	Unknown
58	1	*Hypericum perforatum* L.	Herbal teas	Sweeden	Sweeden	Health shop	Herbal medicine
59	1	*Hypericum perforatum* L.	Extracts	Germany	Germany	Health shop	Unknown
60	1	*Hypericum perforatum* L.	Tablets	USA	Nederland	Health shop	Food supplement
61	1	*Hypericum perforatum* L.	Tablets	Unknown	UK	Health shop	Herbal medicine
62	5	*Citrus limon* (L.) Osbeck	Herbal teas	Slovakia	United Kingdom	Health shop	Food supplement
62	5	*Citrus sinensis* (L.) Osbeck	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
62	5	*Hypericum perforatum* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
62	5	Lavandula spp.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
62	5	*Passiflora edulis* Sims (syn. *Passiflora incarnata* L.)	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
63	4	*Humulus lupulus* L.	Herbal teas	Czech Republic	Slovakia	Health shop	Food supplement
63	4	*Hypericum perforatum* L.	Herbal teas	Czech Republic	Slovakia	Health shop	Food supplement
63	4	*Matricaria chamomilla* L.	Herbal teas	Czech Republic	Slovakia	Health shop	Food supplement
63	4	Mentha spp.	Herbal teas	Czech Republic	Slovakia	Health shop	Food supplement
63	4	*Valeriana officinalis* L.	Herbal teas	Czech Republic	Slovakia	Health shop	Food supplement
64	10	*Achillea millefolium* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Acorus calamus* var. americanus Raf.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Calendula officinalis* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Centaurium erythraea* Rafn	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Cichorium intybus* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Galium verum* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Helichrysum arenarium* (L.) DC.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Hypericum perforatum* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Taraxacum officinale* Webb	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
64	10	*Urtica dioica* L.	Herbal teas	Slovakia	Slovakia	Health shop	Food supplement
65	1	*Hypericum perforatum* L.	Tablets	United Kingdom	United Kingdom	Health shop	Herbal medicine
66	1	*Hypericum perforatum* L.	Capsules	Spain	Spain	Pharmacy	Herbal medicine
67	1	*Hypericum perforatum* L.	Tablets	USA	Turkey	Pharmacy	Food supplement
68	2	*Hypericum perforatum* L.	Extracts	Turkey	Turkey	Pharmacy	Unknown
68	2	*Olea europaea* L.	Extracts	Turkey	Turkey	Pharmacy	Unknown
69	1	*Hypericum perforatum* L.	Extracts	Turkey	Turkey	Pharmacy	Unknown
70	13	*Achillea millefolium* L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Calluna vulgaris* (L.) Hull	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Camellia sinensis* (L.) Kuntze	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Foeniculum vulgare* Mill.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Frangula dodonei* Ard.(*Rhamnus frangula* L.)	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	Hibiscus spp.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Hypericum perforatum* L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Paullinia cupana* Kunth	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Pimpinella anisum* L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Prunus avium* (L.) L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Rosmarinus officinalis* L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Satureja hortensis* L.	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
70	13	*Stevia rebaudiana* (Bertoni) Bertoni	Herbal teas	Turkey	Turkey	Pharmacy	Unknown
71	8	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Gentiana asclepiadea* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Melissa officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Silybum marianum (L.) Gaertn. (*syn. *Carduus marianus L.)*	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Stevia rebaudiana* (Bertoni) Bertoni	Herbal teas	Romania	Romania	Pharmacy	Food supplement
71	8	*Veronica officinalis* L.	Herbal teas	Romania	Romania	Pharmacy	Food supplement
72	6	*Cichorium intybus* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
72	6	*Cynara scolymus* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
72	6	*Gentiana lutea* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
72	6	*Hypericum perforatum* L.	Capsules	Romania	Romania	Pharmacy	Food supplement
72	6	*Silybum marianum* (L.) Gaertn.	Capsules	Romania	Romania	Pharmacy	Food supplement
72	6	*Taraxacum officinale* Webb	Capsules	Romania	Romania	Pharmacy	Food supplement
73	11	*Achillea millefolium* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Calendula officinalis* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Carthamus tinctorius* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Chelidonium majus* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Cynara scolymus* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Gentiana lutea* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Hypericum perforatum* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Lavandula angustifolia subsp. angustifolia* (syn. *Lavandula officinalis* Chaix)	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Melissa officinalis* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Silybum marianum* (L.) Gaertn.	Extracts	Romania	Romania	Pharmacy	Food supplement
73	11	*Viscum album* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
74	10	*Agrimonia eupatoria* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Centaurium erythraea* Rafn	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Chelidonium majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Cichorium intybus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Cynara scolymus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Gentiana lutea* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
74	10	*Lycopodium clavatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Agrimonia eupatoria* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Artemisia absinthium* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Centaurium erythraea* Rafn	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Cotinus coggygria* Scop.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Echinacea* spp.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Gentiana lutea* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Geranium robertianum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Linaria vulgaris* Mill.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Lythrum salicaria* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Robinia pseudoacacia* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Salvia officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
75	18	*Symphytum officinale* L.	Herbal teas	Romania	Romania	Health shop	Food supplement

75	18	*Zea mays* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Agrimonia eupatoria* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Antirrhinum majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Aruncus dioicus* (Walter) Fernald	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Centaurium erythraea* Rafn	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Chelidonium majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Cichorium intybus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Cynara scolymus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	Echinacea spp.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Gentiana lutea* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Linaria vulgaris* Mill.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Lycopodium clavatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Potentilla anserina* L. (syn. *Argentina anserina* (L.) Rydb.)	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Salvia officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
76	18	*Verbascum phlomoides* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Achillea millefolium* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Agrimonia eupatoria* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Antirrhinum majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Asplenium scolopendrium* var. *americanum* (Fernald) Kartesz & Gandhi	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Calendula officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Centaurium erythraea* Rafn (syn. *Centaurium umbellatum* Gilib.)	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Chelidonium majus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Cichorium intybus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Cynara scolymus* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Echinacea purpurea* (L.) Moench	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Equisetum arvense* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Filipendula ulmaria* (L.) Maxim.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Frangula dodonei* Ard.(*Rhamnus frangula* L.)	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Gentiana lutea* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Gentiana punctata* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Hypericum perforatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Linaria vulgaris* Mill.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Lycopodium clavatum* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Lysimachia vulgaris* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Potentilla anserina* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Salvia officinalis* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
77	23	*Verbascum phlomoides* L.	Herbal teas	Romania	Romania	Health shop	Food supplement
78	7	*Achillea millefolium* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Calendula officinalis* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Gentiana lutea* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Hypericum perforatum* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Melissa officinalis* L.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Silybum marianum* (L.) Gaertn.	Extracts	Romania	Romania	Pharmacy	Food supplement
78	7	*Veronica officinalis* L.	Extracts	Romania	Romania	Pharmacy	Food supplement

III

Manzanilla et al. BMC Evolutionary Biology (2018) 18:44
https://doi.org/10.1186/s12862-018-1160-y

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species

V. Manzanilla[1*], A. Kool[1], L. Nguyen Nhat[2], H. Nong Van[2], H. Le Thi Thu[2†] and H. J. de Boer[1†]

Abstract

Background: The economic value of ginseng in the global medicinal plant trade is estimated to be in excess of US$2.1 billion. At the same time, the evolutionary placement of ginseng (*Panax ginseng*) and the complex evolutionary history of the genus is poorly understood despite several molecular phylogenetic studies. In this study, we use a full plastome phylogenomic framework to resolve relationships in *Panax* and to identify molecular markers for species discrimination.

Results: We used high-throughput sequencing of MBD2-Fc fractionated *Panax* DNA to supplement publicly available plastid genomes to create a phylogeny based on fully assembled and annotated plastid genomes from 60 accessions of 8 species. The plastome phylogeny based on a 163 kbp matrix resolves the sister relationship of *Panax ginseng* with *P. quinquefolius*. The closely related species *P. vietnamensis* is supported as sister of *P. japonicus*. The plastome matrix also shows that the markers *trnC-rps16*, *trnS-trnG*, and *trnE-trnM* could be used for unambiguous molecular identification of all the represented species in the genus.

Conclusions: MBD2 depletion reduces the cost of plastome sequencing, which makes it a cost-effective alternative to Sanger sequencing based DNA barcoding for molecular identification. The plastome phylogeny provides a robust framework that can be used to study the evolution of morphological characters and biosynthesis pathways of ginsengosides for phylogenetic bioprospecting. Molecular identification of ginseng species is essential for authenticating ginseng in international trade and it provides an incentive for manufacturers to create authentic products with verified ingredients.

Keywords: Barcoding, Genome, Ginseng, Marker, mPTP, NGS, *Panax*, Phylogenomics, Plastid

Background

Ginseng has been used in traditional medicine in China for thousands of years [1], but it was not until early 18th century that long-term, intense harvest nearly extirpated *Panax ginseng* C.A.Mey. from the wild [2]. Demand for ginseng roots in the 18th century also fuelled a subsequent boom in wild-harvesting American ginseng (*P. quinquefolius* L.) that decimated wild populations in North America [3]. Today wild *P. ginseng* occurs in only a few localities in Russia and China, with the largest distribution in the southern part of the Sikhote-Alin mountain range [4]. *P. ginseng* is Red-Listed in Russia, and roots and parts thereof from Russian populations are CITES Appendix II/NC listed [5]. Many other Asian ginseng species are also endangered but preliminary data is only available for wild-harvesting and conservation of *P. assamicus* R.N. Banerjee (synonym of *P. bipinnatifidus* var. *angustifolius* (Burkill) J.Wen) [6], *P. japonicus* (T.Nees) C.A. Mey. [7] and *P. pseudoginseng* Wall. [8, 9].

Elucidating the evolutionary relationships among species in the genus is essential to understand evolution of this Holarctic disjunct genus, but also evolution of derived secondary metabolite pathways. In addition, a phylogenetic framework can be used to develop accurate molecular identification of *Panax*, and enable identification of ginseng material in trade, both crude drugs and derived products, which is essential for conservation efforts and protection of the remaining wild populations of *P. ginseng* and related *Panax* species,

* Correspondence: vincent.manzanilla@nhm.uio.no
†Equal contributors
[1]The Natural History Museum, University of Oslo, Oslo, Norway
Full list of author information is available at the end of the article

since all may be under the pressure of illegal harvesting and international trade [10]. Furthermore, identification of *Panax* species and authentication of derived products is of great commercial importance as authentic ginseng is costly and the incentive for substitution is significant.

The phylogeny of *Panax* has been studied using several molecular markers, but lack of variation in the most commonly used markers highlight an important limitation of the method. The nuclear ribosomal ITS yields insufficient resolution for accurate species assignment [11] and even using multiple markers in combination, *matK, trnD, psbK-psbI, rbcL* and *ycf1* have a limited accuracy in identification of *Panax* species [12, 13]. The mutation rate of the studied markers does not allow a fine scale resolution, and is insufficient for identification of all *Panax* species and cultivars. The question of what species are in trade remains a mystery. Aside from phylogenetic approaches, a multitude of molecular and chemical analysis approaches have been developed and applied, including Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) [14], PCR-Restriction Fragment Length Polymorphism (PCR-RFLP) and Mutant Allele Specific Amplification (MASA) [15], Random Amplified Polymorphic DNA (RAPD) and High Performance Liquid Chromatography [16], Fourier Transformed-Infrared Spectroscopy (FT-IR) [17], Two-Dimensional Correlation Infrared Spectroscopy (2D-IR) [17], Multiplex Amplification Refractory Mutation System-PCR (MARMS) [18, 19], Microchip Electrophoresis Laser-Induced Fluorescence Detection [20], and microsatellite markers [21]. Most methods have focused on either positive identification of *P. ginseng*, or distinguishing *P. ginseng* and *P. quinquefolius* L., but most have limited resolution in detecting infraspecific or interspecific substitution, especially with poorly known congeneric species.

Suitability of molecular markers is often measured in interspecific distance using distance methods to estimate the number of variable sites or pairwise distances between sequences. Most current methods are based on the Refined Single Linkage (RESL) algorithm implemented in Barcode of Life Database (BOLD) [22] or clustering on distance matrices (Crop [23], OBITools [24], UCLUST [25], and Vsearch [26]) and ideally set a threshold to distinguish between intraspecific and interspecific variation, sometimes referred to as the "barcoding gap" [27]. Several programs and software packages determine and visualize barcoding gaps, including Automatic Barcode Gap Discovery (ABGD) [28] and Spider [29]. These distance-based methods are fast and suitable for large datasets, but they are not always biologically meaningful, especially when the species groups have complex evolutionary histories, including incomplete lineage sorting, and hybridization [30, 31]. As an alternative, tree-based methods offer several advantages compared to distance based methods. First, these methods do not work with a specified threshold (% variation, no barcoding gap) and second, these accommodate evolutionary processes, making them particularly suitable for species delimitation and identification. Several studies have shown that these methods are also more sensitive and more powerful for accurate species discrimination [32]. Recently proposed methods include the Generalized Mixed Yule Coalescent (GMYC) [33], Bayesian species identification using the multispecies coalescent (MSC) model [34], and Poisson Tree Processes (PTP, mPTP) [25, 32]. Despite constant methodological improvements, there is no silver bullet for species delimitation and concerns have been raised that species delimitation approaches are sensitive to the structure of the data tested [35]. Species delimitation methods assess speciation and coalescent processes but also the data structure of the selected markers [35]. From a marker development perspective, tree based methods provide an opportunity to increase the quality of the selection process of the barcoding markers. Here we use the mPTP approach [32] to test if speciation processes are supported by the barcoding markers and accordingly choose the best markers for delimitation of *Panax* species. mPTP method has the advantage of being computationally efficient, while at the same time accommodating better to population-specific and sampling characteristics than PTP and GYMC [32].

Evolution and phylogenetics of *Panax*

Previous phylogenetic studies of the Araliaceae family have identified four monophyletic groups (the Asian Palmate group, the Polyscias-Pseudopanax group, the Aralia-Panax group, and the greater Raukaua group) [36, 37]. However deep nodes are not well-supported to date [36, 37], and a broad sampling within Aralioideae is necessary to obtain an accurate placement of the Aralia-Panax group. Monophyly of the genus *Panax* (Araliaceae) is well supported by morphological synapomorphies, such as palmately compound leaves, a whorled leaf arrangement, a single terminal inflorescence, valvate petals in floral buds, and a bi- or tricarpellate ovary, as well as by several molecular phylogenies [12, 38]. A number of species have emerged from the complex of subspecies of *P. pseudoginseng* in the 1970s, and taxonomic studies have resulted in the description of various new species [38–40]. Currently 13 species of ginseng are recognized with broad consensus [38, 41], but publication of new taxa at species, subspecies and variety level are common [42, 43].

Previous molecular phylogenies support *P. stipuleatus* H.T.Tsai & K.M.Feng and *P. trifolius* L. as the sister group of all other ginseng species. Nevertheless the

placement of several other species still remains unclear (e.g., *P. binnatifidus*, *P. ginseng, P. japonicus, P. quinquefolius, P. vietnamensis* Ha & Grushv., *P. wangianus* S.C. Sun, *P. zingiberensis* C.Y.Wu & Feng). Species delimitation within the genus is problematic due to species of tetraploid origin (e.g., *P. bipinnatifidus, P. ginseng, P. japonicus,* and *P. quinquefolius* [44]), recent speciation events [12], high intraspecific morphological variation (e.g., *P. pseudoginseng* Wall.) and ancient genome duplication events [41, 45].

Phylogenetic studies have explored evolutionary relationships in Araliaceae with standard phylogenetic markers, such as the nuclear ribosomal ITS [11, 36, 38, 41, 44–46] and several plastid markers [11–13, 41]. More recently, an attempt with seven nuclear genes was tested with moderate results (*PGN7, W8, W28, Z7, Z14, Z15, Z16*) [12]. The topologies obtained were conflicting and non-consistent with previous evolutionary inferences of the genus, which is likely a result of multiple copies of nuclear genes and ancient whole genome duplication events [47]. Whole genome data have also been used to design microsatellites for species identification, but these have found limited application [21, 48–52]. Extensive population genetic studies have been done only on *P. quinquefolius* [53–59] and *P. ginseng* [60, 61] due to their major economic importance.

Developments in high throughput sequencing have provided new approaches for genome sequencing: increasing outputs and decreasing costs have made this a cost-effective alternative to Sanger-based amplicon sequencing [62, 63]. Full plastid genome sequencing, i.e. plastome sequencing, has been proposed as an augmented approach to DNA barcoding [64, 65], and is a straightforward method that recovers all standard barcodes plus the full plastome. The limited costs of shotgun sequencing and the availability of a number of Araliaceae reference plastomes facilitates the study of relationships in the family. Plastome phylogenies have helped disentangle evolutionary relationship in a number of plant clades [66], including Poales [67], magnoliids [68], *Pinus* [69], *Amborella* [67], *Equisetum* [70], and *Camellia* [71]. Single-copy nuclear genes have corroborated the robustness of plastome phylogenies [72–75], however plastome phylogenies reflect only maternal inheritance, and as such will not always be representative species trees. An advantage of plastome data for phylogenetic studies is the low mutation rate of plastid sequences, the abundance of plastid DNA in most material [76] and the low cost of generating whole plastid genomes with high throughput sequencing.

In total DNA, the proportion of plastid DNA typically constitutes only ~ 0.01–13% depending on the size of the nuclear genome, tissue and season [77–79]. Shotgun sequencing studies might have relatively low efficacy in plastid genome recovery due to the small proportion of plastid DNA in the total DNA. Ginseng species have a large genome size of 5–10 Gb [80, 81], and one can expect a proportion of plastid DNA of 1–5% in the gDNA [79], which makes shotgun sequencing relatively ineffective in obtaining full plastome data. Several methods have been developed for enriching plastid content prior to sequencing (for a discussion see Du et al. [82]. We apply a new plastid enrichment method to improve the shotgun sequencing efficacy, that utilizes the low methylation of the plastid genome compared to the nuclear genome [83]. The method uses the methyl-CpG-binding domain (MBD2) to partition fragments of genomic DNA into a methylation-poor fraction (e.g. enriched for plastid) and a methylation-rich fraction (e.g. depleted in plastid) [84]. This method has the advantage that it uses a small quantity of dry material (below 40 mg) and is suitable for non-model organisms.

This study has four main aims: (1) to construct a well-supported phylogeny of the genus *Panax*, while testing if the full plastome data yield sufficient variation to support and resolve phylogenetic relations in *Panax*, and specifically the position of the economically important *P. ginseng*; (2) to test if MBD2 can be used to fractionate *Panax* DNA into eukaryotic nuclear (methyl-CpG-rich) vs. organellar (methyl-CpG-poor) elements, and subsequently sequence the MBD2 depleted DNA to optimize plastome read yield; (3) to determine if the plastid genome can be used for molecular identification of traded species; and 4) to make a case for the need of a resolved plastome phylogeny to be used to design short markers for *Panax* species identification from processed ginseng products.

Methods

Sampling

Fresh material of three species, *P. bipinnatifidus, P. stipuleanatus,* and *P. vietnamensis (2)*, was sampled in Vietnam (Table 1, Additional file 1: Table S1) and 57 selected Araliaceae plastid genomes from across the Araliaceae family were downloaded from open data repositories (Additional file 2: Table S2) [12, 85–97]. Plant samples were collected in public land and no

Table 1 Summary information for the four assembled plastome genomes

Taxon	Number of reads	Plastome coverage	Length (bp)	NCBI Reference
P. vietnamensis (1)	292,401	16.90	156,022	MF377621
P. bipinnatifidus	405,910	133.38	156,248	MF377620
P. vietnamensis (2)	845,962	253.04	156,099	MF377623
P. stipuleanatus	423,538	91.31	156,090	MF377622

specific permits were required. At least two individuals or species were selected per genus, but for *Panax* we used 38 plastid genomes from eight species. *Hydrocotyle verticillata* was selected as outgroup based on its early divergence within Araliaceae [44].

Library preparation and sequencing

We extracted total DNA from two individuals of those sample collected in Vietnam, using a Qiagen DNeasy plant extraction kit with the provided protocol. The total DNA was quantified prior to library preparation to assess DNA quantity, fragmentation and fragment length distribution on a Fragment Analyzer (Advanced Analytical Technologies, Inc., Ankeny, USA) using the High Sensitivity genomic DNA Reagent Kit (50–40,000 bp) (Additional file 3: Figure S1). We selected one individual per extracted sample based on the yield and fragment size of the total DNA. The selected samples had average fragment sizes in excess of 10 kbp and a minimum DNA concentration of 4.77 ng/μl (Additional file 3: Figure S1).

We used a NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, Massachusetts, USA) that uses IgG1 fused to the human methyl-CpG-binding domain (together "MBD2-Fc") to pull down a methyl-CpG-enriched fraction from a bead-associated element, leaving a methyl-depleted fraction in the supernatant. About 400 ng template DNA extract was used per sample and the manufacturers recommendations were respected with the following exceptions. The non-methylated DNA fractions were purified using 0.9X AMpure XP beads (Beckman Coulter, Brea, CA, USA) and eluted in 40 μl 1X TE buffer. To capture the methylated DNA, we followed the manufacturer's protocol. Quality control in terms of size, purity and molar concentration (nmol/l) of both the methylated and the non-methylated fractions were measured using a Fragment Analyzer (Advanced Analytical Technologies Inc., USA) with a DNF-488-33 HS dsDNA Reagent Kit. The DNA was subsequently sheared to ~ 400 bp fragments using a M220 Focused Ultrasonicator (Covaris Inc., Woburn, MA, USA) using microTUBES-50 (Covaris Inc.). We used the NEBNext Fast DNA Library Prep Set for Ion Torrent (NEB) for end repair and adapter ligation of the sheared DNA. The samples were indexed using the IonXpress Barcode Adapter kit (ThermoFischer, Waltham, MA, USA). For each of the four samples both fractions, methyl-CpG-enriched and methyl-CpG-depleted, were indexed and sequenced. After adapter ligation, the four methyl-CpG-enriched fractions were pooled in one library and the four methyl-CpG-depleted fractions were pooled in another library. The adapter-ligated libraries were size selected (450–540 bp) using a BluePippin (Sage Science, Beverly, MA, USA), and subsequently amplified using the NEBNext Fast DNA Library Prep Set for Ion Torrent kit using 12 PCR cycles. The amplified libraries were purified twice using 0.7X AMpure XP beads. The purified amplified libraries were loaded on the sequencing chips using an Ion Chef (LT) and sequenced on an Ion Torrent Personal Genome Machine (LT) using Ion 318 v2 chips (LT) and the Ion PGM Sequencing 400 kit (LT).

Bioinformatic analyses and assembly

Sequencing reads were demultiplexed into FASTQ files using Flexbar version 3.0.3. Trimmomatic version 0.36 [98] was used for adapter trimming and quality filtering of reads using a sliding window of 15 bp and an average Phred threshold of 20. Low-end quality bases below a Phred score of 20 were removed, and only reads longer than 100 bp were retained. MITOBim version 1.7 [99] was used for assembly of the single-end Ion Torrent reads using iterative mapping with in silico baiting using the following reference plastomes, *P. vietnamensis* (KP036470) and *P. stipuleanatus* (KX247147).

Inverted repeats and ambiguous portions of the assembly were resequenced using Sanger sequencing. Specific primers were designed and used for DNA amplification of interest regions. PCR was performed on a Mastercycler® Pro (Eppendorf, USA) in a 20 μl final volume containing 2.5 μM of each primer, 1 mM of each dNTP, 10X DreamTaq Buffer, 0.75 U DreamTaq DNA polymerase (ThermoFisher Scientific, USA) and deionized water. The PCR cycling conditions included a sample denaturation step at 94 °C for 2 min followed by 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 50–55 °C for 30 s and primer extension at 72 °C for 1 min, followed by a final extension step at 72 °C for 5 min. PCR products were then purified using GeneJET PCR Purification Kit (ThermoFisher Scientific, USA). Sanger sequencing was performed on an ABI 3500 Genetic Analyzer system using BigDye Terminator v3.1 Cycle Sequencing Kit. Cycle sequencing was performed on a Veriti Thermal Cycler (Applied Biosystems, USA) using 3.2 μM of each primer, 200 ng purified PCR product, 5X BigDye Sequencing Buffer, 2.5X Ready Reaction Premix and deionized water in a 20 μl final volume. The thermocycling conditions included 1 min at 96 °C followed by 25 cycles of denaturation at 96 °C for 1 min, primer annealing at 50 °C for 5 s and primer extension at 60 °C for 4 min, followed by a holding step at 4 °C. Extension products were purified using ethanol/EDTA precipitation with 5 μl of EDTA 125 mM, 60 μl of absolute ethanol. Purified products were denatured at 95 °C for 5 min using 10 μl Hi-Di Formamide. DNA electrophoresis was performed in 80 cm × 50 μ capillary with POP-4 polymer (Applied Biosystems, USA).

In order to test the efficacy of the NEBNext Microbiome DNA Enrichment Kit the proportion of reads belonging to the plastome was estimated for both the methylated and the non-methylated fraction. The *P.*

ginseng whole genome sequencing SRR19873 experiment was used to estimate the starting proportion of plastome reads, by mapping the reads against the plastid genome of *P. ginseng* (NC_006290) using Bowtie 2. Association of reads to their taxonomic identification and organelles, was made using a tailored database of *Panax* plastome data representing the same data as that downloaded from public repositories for the phylogenetic analyses. For the mitochondrial data, all angiosperm mitochondrion genomes available on NCBI were used, and for the microbiome all remaining reads were blasted against the full NCBI database. Taxonomic identifications were retrieved using the lowest common ancestor (LCP) algorithm in Megan version 5.11.3, with minimum read length of 150 bp and at least 10 reads for each taxon identified with an e-value of 1e-20 or less. The proportion of plastid DNA in the gDNA was estimated using Bowtie2 by mapping the proportion of reads belonging to the plastid genome for *P. ginseng* (following SRR experiment SRR1181600).

The plastid genomes were annotated using Geneious version 6.1, and annotations of exons and introns were manually checked by alignment with their respective genes in the same annotated species genome. Representative maps of the chloroplast genomes were created using OGDraw (Organellar Genome Draw, [100]).

Phylogenomics

The matrix for phylogenomic analyses consisted of complete aligned plastid genomes, and the global alignment was done using MAFFT version 7.3 [101] with local re-alignment using MUSCLE version 3.8.31 [102], and manual adjustments where necessary. Aligned DNA sequences have been deposited in the Open Science Framework (OSF) directory (https://osf.io/ryuz6). The final matrix has a total length of 163,499 bp for a total of 61 individuals with no missing data. Single nucleotide polymorphisms (SNPs) were visualized using Circos version 0.69 [103]. Relationships from the nucleotide matrix were inferred using Maximum Likelihood (ML) and Bayesian inference. First, an un-partitioned phylogenetic analysis was performed to estimate a single nucleotide substitution model and branch length parameters for all characters. Next, the data was partitioned in coding regions, introns and intergenic spacers, and a best-fit partitioning scheme for the combined dataset was determined using PartitionFinder version 2.1.1 [104] using the Bayesian Information Criterion (Additional file 4: Table S3). Branch lengths were linked across partitions.

The dataset was analyzed using RAxML version 8.2.10 [105] and mrBayes version 3.2.6 [106]. RAxML and Bayesian searches used the partition model determined by PartitionFinder. For the ML analyses, tree searches and bootstrapping were conducted simultaneously with 1000 bootstrap replicates. Bayesian analysis were started using a random starting tree and were run for a total of ten million generations, sampling every 1000 generations. Four Markov runs were conducted with eight chains per run. We used AWTY to assess the convergence of the analyses [107]. Conflicting data within ML and Bayesian analyses were visualized and explored using the R package phangorn using the *consensusNet* function [108].

Barcoding - mPTP

Suitable barcoding markers were selected by extracting the SNP density over the plastid genome alignment of all *Panax* species and individuals included in this study (matrix available as supplementary data on OSF). We used SNP-sites version 2.3.2 [109] to extract the SNP positions from the alignment of a matrix containing only the *Panax* species, and created bins every 800 bp using Bedtools version 2.26.0 [110] (script available on OSF) and plotted the SNP density using Circos [103] (Fig. 1). The coordinates of each annotation on the aligned *Panax* species matrix were found using a reference consisting of the four annotated genomes produced in this study, and subsequently exported to Circos. We selected the most variable regions and designed suitable primers for these regions (Fig. 5, Additional file 5: Table S4). From the matrix used for the Aralioideae, we extracted 15 plastid markers (Fig. 5) and download ITS sequences for the *Aralia-Panax* group (Figs. 3 and 5) (Additional file 2: Table S2). We performed maximum likelihood analyses on individual and concatenated matrices using RAxML. mPTP analyses were performed using the ML trees from the individual and concatenated markers, and using the Markov chain Monte Carlo (MCMC) algorithm with two chains and the Likelihood Ratio Test set to 0.01.

Results

Ion torrent sequencing

After filtering out low-quality reads, 1.9 out of 3.3 and 3.3 out of 4.9 million reads were retained for the pooled MDB2 depleted and enriched fractions respectively. The chloroplast assemblies covered the entire circular plastid genome for all four accessions for the MDB2 depleted fraction (Additional file 6: Figure S2, Additional file 7: Figure S3, Additional file 8: Figure S4, Additional file 9: Figure S5; Table 1). The Sanger generated plastid sequences confirmed the genome assemblies in 18 regions, and also confirmed sequences of the inverted repeat. Complete lengths of the four plastid genomes ranged from 156,036 bp to 156,302 bp (Table 1). All four plastid genomes had the same genome structure and gene arrangement as that of the already assembled *Panax* plastid genomes.

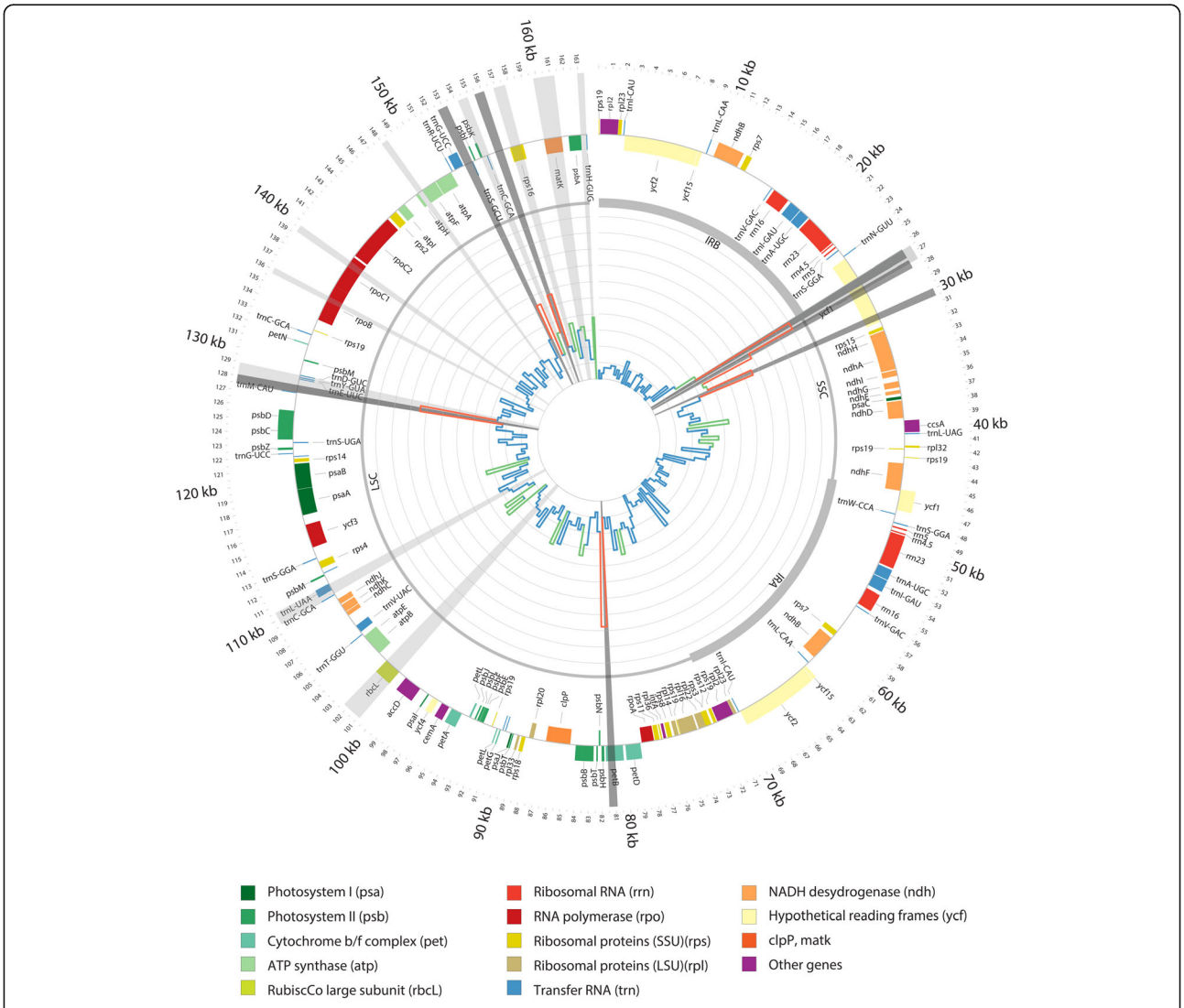


Fig. 1 Plastid genome representation of the 38 aligned *Panax* genomes. The internal histogram plot represents the SNPs density over the alignment of the plastid genomes of *Panax* genus. The colours indicate when the standard deviation of the bin falls in different interval compare to the average standard deviation, between 0 and 1 in blue (low variation), between 1 to 2 in green (moderate variation) and over two in red (highly variable). Inverted repeats A and B (IRA and IRB), large single copy (LSC) and small single copy (SSC) are shown in the inner circle by different line weights. Genes shown outside the outer circle are transcribed clockwise, and those inside are transcribed counter clockwise. Genes belonging to different functional groups are color-coded. Radial grey highlights show the regions in focus of study, light grey previously used barcodes, in dark grey newly developed barcodes

Methylation enrichment

The Fragment Analyzer results showed that DNA quantity and fragmentation differed for the four DNA samples (Additional file 3: Figure S1), and the results were used to normalize concentrations for subsequent capture. DNA concentrations after capture and fragment size selection are much lower for the methyl-depleted fraction compare to the methyl-CpG-enriched fraction (Fig. 2). The success of the fragment size selection was relatively poor for one of the *P. vietnamensis*. and resulted in a poorer quality in the sequencing and enrichment due to the excessive abundance of short DNA fragments. The shorter reads for *P. vietnamensis*. yielded a lower coverage for its genome assembly (16.9 X) (Table 1).

The enrichment and depletion of methylated DNA by pulling down a methyl-CpG-enriched fraction and leaving a methyl-depleted fraction drastically increased the proportion of organellar DNA within the depleted fraction. *P. ginseng* SRR experimental data had 5.63% plastid genome reads. In the methylation-depleted fraction, we found a variation of plastome reads ranging from 6 to 33%. In the methylation-enriched fraction, less than 1% of the reads are from the plastome. The enrichment also increased microbiome contamination in the depleted

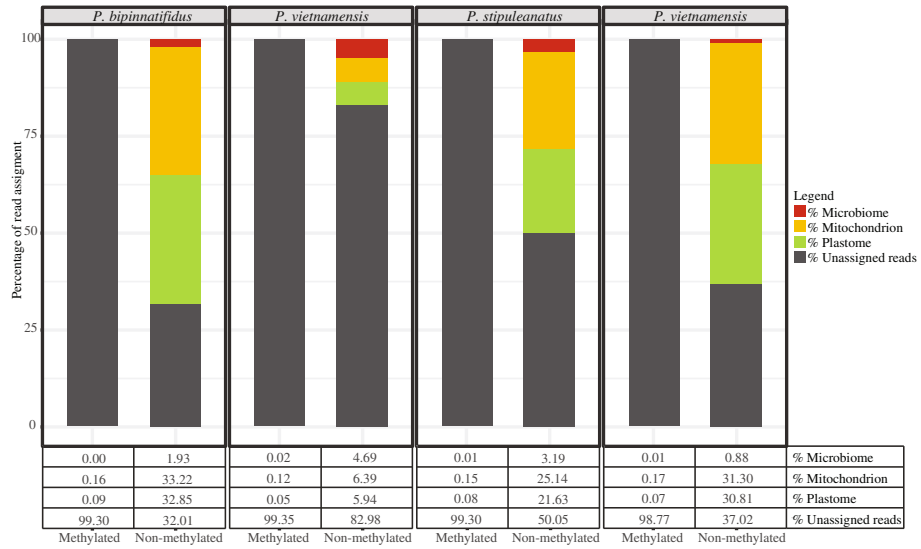


Fig. 2 Proportion of mitochondrial, plastid and microbiome reads from the methylated and non-methylated fraction for the four enriched libraries. These results show that the enrichment procedure successfully capture more than 99.5% of the plastid sequences contained in the total DNA

fraction from 0.8 to 4%. Overall, one of the *P. vietnamensis* samples was the least successful sample in the enrichment and yielded fewer and shorter reads.

Phylogenetic analyses

Alignment of the plastid genomes for phylogenetic analyses were consistent in length throughout the dataset. Based on the alignment, average plastome pairwise identity for the Araliaceae family is 83% and 99.2% for the *Panax* clade. The percentage of identical sites is 83.9% and 96.8% respectively. The global plastome alignment has a matrix length of 163,499 bp. Coding regions, introns and intergenic spacers represented 259 original partition schemes, and the best-fit partitioning scheme from PartitionFinder divided the data into 73 partitions (Additional file 4: Table S3).

Inspection of the posterior probabilities calculated using AWTY, yielded an estimated burnin of 10% for the Bayesian analysis. Phylogenetic analyses revealed significant divergence between major clades of the Araliaceae family. The ML and Bayesian trees showed strongly supported clades for all genera of the family (Fig. 3). Furthermore, the tree shows maximum support for each species of *Panax* included in the analyses. All intergeneric and infrageneric relationships were strongly supported (Fig. 3).

The basal node segregates two clades, one clade includes two genera, *Aralia* and *Panax*. The second clade includes *Schefflera*, *Fatsia*, *Eleutherococcus*, *Kalopanax*, *Metapanax*, *Brassaiopsis*, and *Dendropanax*. All species included in the study are monophyletic and have maximum support in both Bayesian and ML analyses.

The Araliaceae clade

The Araliaceae clade showed maximum support in the phylogeny except for the *Fatsia* clade, where the support is 99.6%. *Schefflera* is sister to the rest of the clade, followed by *Dendropanax*, then a clade with *Brassaiopsis*/*Eleutherococcus* and finally a clade with *Fatsia*/

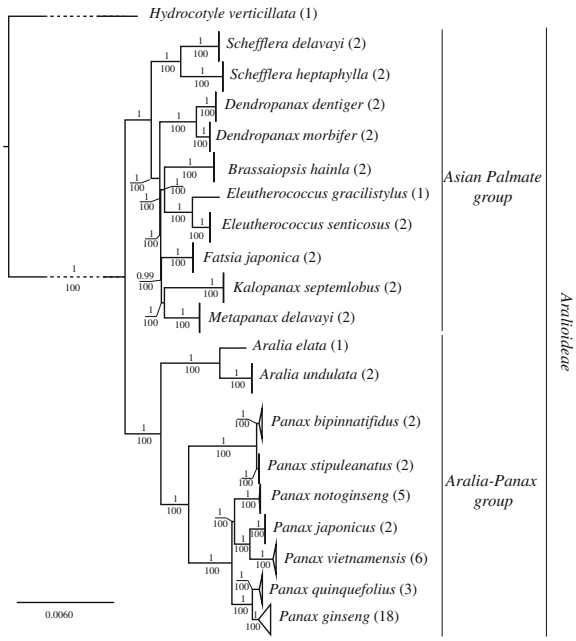


Fig. 3 Bayesian tree inferred from the 61 plastid genomes from the Aralioideae tribe. Numbers above branches indicate posterior probabilities and below the ML bootstrap values

Kalopanax/*Metapanax*. A comparison of the partitioned and non-partitioned analyses shows no differences in topology and support in the *Aralia-Panax* clade, but does in the remaining Araliaceae clade.

The Aralia-Panax clade

The genus *Panax* is monophyletic and *Aralia*, represented by two species, *A. elata* and *A. undulata*, is the sister group to the genus *Panax*. *Panax stipuleatus* and *P. binnatifidus* form a distinct clade sister to a clade consisting of *P. notoginseng* and its sister group of *P. vietnamensis* and *P. japonicus*, which as a whole is sister to *P. quinquefolius* and *P. ginseng*.

The consensus network was computed from the two Bayesian runs after discarding 10% burnin (Fig. 4). The network analysis shows two main conflicts in the data, one within the *P. ginseng* clade and another within the *P. vietnamensis* clade. Both clades have very little intraspecific variation (soft incongruence), and more variable markers are needed to segregate the different individuals correctly for these two species.

Barcoding analyses

The SNP density analyses retrieved 2052 SNPs over the full plastid alignment. We identified three regions (Figs. 1 and 5) that are suitable barcoding markers. Each of these regions has on average of 83 SNPs within *Panax* (Fig. 5). Individual marker phylogenies of these regions are suitable to segregate most of the species clades (Additional file 10: Figure S7, Additional file 11: Figure S8, Additional file 12: Figure S9, Additional file 13: Figure S10, Additional file 14: Figure S11). The exceptions are the two sister pairs, *P. quinquefolius* and *P. ginseng*, and *P. binnatifidus* and *P. stipuleatus*, where the bootstrap supports are weaker, leading to inference of single clades. The ML phylogeny of the concatenated markers, fully supports all species clades, except *P. binnatifidus* and *P. stipuleatus* (Fig. 5, Additional file 14: Figure S11).

In the mPTP analysis for the full plastid dataset, the Average Support Value (ASV) assesses the congruence of support values with the ML delimitation. The analyses return an ASV of 97.9%, suggesting a high confidence for the given species delimitation scheme. Species delimitation recognized 21 distinct entities out of 20 species (Additional file 15: Figure S6). Over-representation and intraspecific variation of the *P. ginseng* samples has resulted in oversplitting this clade into two discrete entities. The *P. stipuleatus* / *P. binnatifidus* clade has lower data structure and the analyses does not strongly support the group as two independent mPTP entities (PP = 0.68). *P. quinquefolius* has been also divided into two subgroups, but the posterior probability of the subdivision is low (PP = 0.4).

The result of mPTP analyses for all previously used and the newly proposed markers are described in Fig. 5 and the supported nodes for the speciation events have been added to the phylogenetic tree (Additional file 10: Figure S7, Additional file 11: Figure S8, Additional file 12: Figure

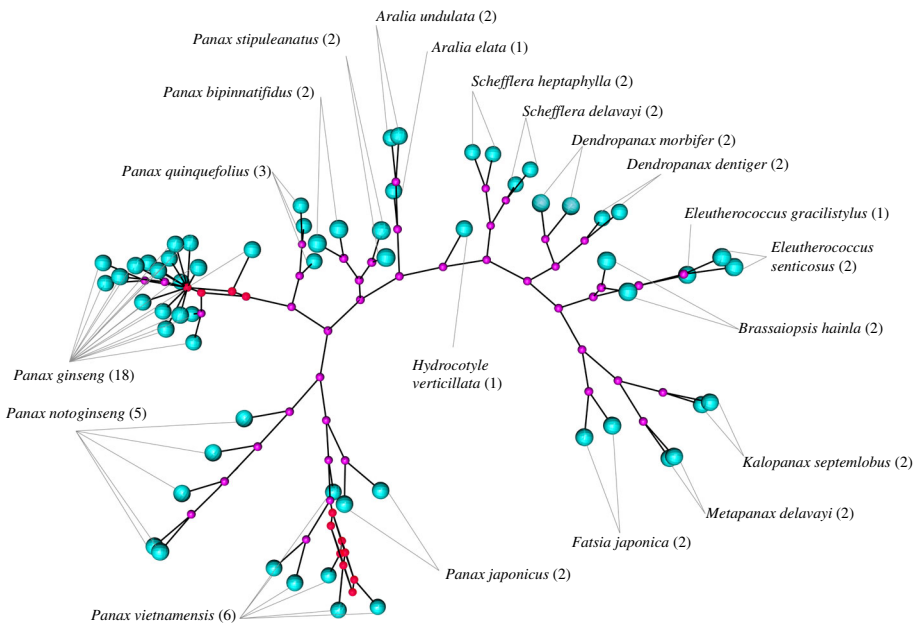


Fig. 4 Consensus network from Bayesian runs with 10% burn in. This network shows two main conflicts in the data, one on the *P. ginseng* clade, where there is very little intraspecific variation and one with the clade of *P. vietnamensis*. The nodes are represented in pink and the tips in blue. The red nodes show the two splits within the data

Marker name	size (bp)	Parsimony informative sites (PI)	Variable sites (%)	Correctly identified species mPTP
rpoC1	524	3	0.57	1
rpoB	489	6	1.23	2
atpF-H	427	24	5.62	2
psbK-L	716	39	5.45	2
rbcL	1465	59	4.03	2
matK	1512	50	3.31	2
ITS	619	84	13.57	2
ycf1	1084	126	11.62	2
trnH-psbA	457	47	10.28	3
trnL-F	898	44	4.9	3
petB	584	51	9.2	4
rps16	841	36	4.28	5
(4) trnS-trnG	673	54	8.02	5
(3) psbM-trnD	1340	238	17.76	5
(2) trnE-trnM	603	145	24.05	6
(1) trnC-rps16	657	151	22.98	6
1 + 2	1260	296	23.49	6
1 + 2 + 3	1933	350	18.11	6
1 + 2 + 3 +4	3273	588	17.97	6

Fig. 5 Percentage of variable site (orange) and successful species identified with the mPTP analyses (blue), for each marker and the concatenated matrices

S9, Additional file 13: Figure S10, Additional file 14: Figure S11). Out of the 15 analysed markers only four can be used to discriminate most species. Figure 5 also shows that regions with the highest density of parsimony informative sites are not necessarily the most efficient for species discrimination, and both skewed aggregated mutations as well as homoplasy can obscure phylogenetic patterns.

Discussion

Evolution of Araliaceae and ginsengs

The evolution of the Asian palmate group (Fig. 3) is concordant with previously published articles that show *Schefflera* at the base of the group. The paraphyletic genus *Dendropanax* was usually the most divergent in the group, but is now basal to the rest of the group. This position might be due to low sampling within the Asian palmate group. Results for *Brassaiopsis*, *Eleutherococcus*, *Fatsia*, *Kalopanax* and *Metapanax*, correspond with previously published phylogenies. Early radiations with interlineage hybridizations and genome doubling have been reported in the group [111] and this could explain the short internal branches. Further phylogenomic and biogeographical studies should be conducted to better understand the radiation of the Araliaceae.

In the Aralia-Panax group, Aralia is sister to *Panax*, and we find that *P. stipuleatus* forms a well-supported clade with *P. binnatifidus*, whereas previous studies have often reported that *P. binnatifidus* groups with *P. omeiensis*, *P. wangianus*, *P. zingiberensis* and *P. major* [11, 12, 38, 41], all four of which are however missing here. Due to the difficulty in obtaining material of *P. vietnamensis*, only three studies have included *P. vietnamensis* in a phylogeny [13, 96, 112]. The study by Lee et al. [112] using the plastid marker trnC–trnD does not resolve the position of *P. vietnamensis* in the phylogeny, but does identify a distinct clade consisting of *P. notoginseng*, *P. japonicus* and *P. vietnamensis*, which is also supported by our data. Komatsu et al. [13] recover a clade consisting of *P. vietnamensis*

along with *P. japonicus* and *P. pseudoginseng* subsp. *himalaicus*, a synonym of *P. bipinnatifidus*. Inferring *P. japonicus* to belong to this clade is contradictory to previous studies that have found a clade consisting of *P. quinquefolius*, *P. ginseng* and *P. japonicus* [12, 38, 41, 112]. The plastome phylogeny supports a sister-relationship of *P. ginseng* and *P. quinquefolius*, the two economically most important species of ginseng. Although this full plastome phylogeny significantly differs from previously published molecular phylogenies, the new evolutionary pattern is strongly supported by bootstrap values and posterior probabilities.

Incongruence between markers from different origin

Full length plastid genome data are a major improvement for the *Panax* phylogeny, and the addition of a bigger dataset has a strong influence on the phylogenetic hypothesis. However, discrepancies between full-length plastid genome phylogenies and nrDNA phylogenies are common in plants. nrDNA has been widely used for phylogenetic studies of *Panax* [11, 38, 41, 46], but the limitations of this approach have been extensively reviewed in [113]. Drawbacks of nrDNA include difficulties in aligning, and its limited use for phylogenetic inference between closely related and/or recently diverged taxa. It is also a challenge to determine the orthology and the paralogy of nrDNA sequences in the case of hybridization events or incomplete lineage sorting [114–116]. Bailey et al. [114] emphasise that despite valuable phylogenetic information from nrDNA, it might not the optimal choice to assess species trees, especially in case of allopolyploids or tetrapolypoids. Since this is also the case in *Panax*, we argue that nrDNA may be inappropriate to reconstruct the evolutionary history of this genus.

Phylogenetic congruence as well as incongruence of nuclear genomic and plastid marker data is well documented [117–119]. In the case of *Panax*, two of the nuclear markers used by [12] support the clade of *P. ginseng* and *P. quinquefolius* (Z14, Z8). However, our topology is incongruent for the remaining clades. Incongruences between the maternally inherited plastid genome and the biparentally inherited nuclear genes can be expected in genera with allopolyploid hybrids, like *Panax* [12]. Plastid phylogenies are not always representative of the species tree and might conflict with hypotheses of parsimonious morphological evolution [116, 120, 121]. Incongruences between plastome and nuclear gene trees have been reported in wide ranging groups of plants, such as *Asclepia* [72], *Helianthus* [122] and *Silene* [120].

Enrichment

The novel method based on methylation-based enrichment increased the concentration of plastid DNA by 30% which is in the range found by a previous pilot study [84]. It is a suitable method for enriching the organellar genome before sequencing. The methylated fraction shows extremely low amounts of organellar DNA, meaning that we removed more than 99% of the non-methylated DNA from the total DNA. The *P. vietnamensis* sample had originally more degraded DNA and as a result shows a less successful enrichment. Using MBD2 to increase the concentration of organellar DNA in the total DNA allows multiplexing a larger number of samples. This method is appropriate for building plastid reference genome databases for barcoding projects. In case of degraded samples, we recommend removal of shorter DNA fragments before the enrichment.

Selecting markers for molecular *Panax* identification

In DNA barcoding and plant product identification and authentication projects it is common to work with degraded DNA substrates for which it might be difficult to use methylation enrichment or the full plastid genome as a barcoding strategy. However, alternatives such as target enrichment and amplicon sequencing are possible [64, 123–125]. Here we have identified four variable regions that possess sufficient variation and genetic structure to discriminate most ginseng species. The identification of ginseng species is relatively complex because of the recent evolution and hybridization events. *P. ginseng* and *P quinquefolius* have recently diverged plastid genomes, and so do *P. binnatifidus* and *P. stipuleatus* [47]. Species delimitation using mPTP shows that for such species complexes traditional barcoding markers do not have enough structure for delimiting species. However, if carefully selected, some regions highlight specific structural patterns that enable the discrimination of species. The *trnC-rps16* region seems to be particularly promising, as it has enough variation to discriminate most species (Additional file 15: Figure S6). If plastid markers are to be used for barcoding, it is more relevant to use a combination of markers because mPTP analyses are better suited for multi-marker analyses [32]. A concatenated matrix with two, three or four markers combined improves the efficacy in segregating all the *Panax* species and specifically also those in closely related complexes. Our results suggest that a combination of the following markers: *trnC-rps16*, *trnE-trnM* and *psbM-trnD* (Fig. 5) enables confident identification of the main traded species *P. ginseng*, *P. quinquefolius* and *P. vietnamensis*. For further development, a complete sampling of all *Panax* species with multiple accessions per taxon should be made to confirm the observed variation in the selected markers.

In order to design accurate markers to monitor the trade of the medicinal species, it is necessary to understand the evolution of the targeted group. Many studies are based on the generic barcodes suggested by iBOL (International Barcode of Life) (*rbcL* and *matK)* without

having strong evidence for the evolutionary hypotheses of the targeted group and a limited idea *a fortiori* of the discriminatory power of the used markers. Nonetheless, when a barcoding study targets a specific plant group or genus, and the barcode markers fail to yield a supported phylogeny, then one should aim to construct robust phylogenies with new markers to achieve species discrimination. If the phylogenetic hypothesis is not robust, or if the data are weak in structure as it is often the case with the standard barcoding markers, *rbcL* and *matK*, the resulting identifications might be misleading because of inaccurate species delimitation hypotheses [31].

Conclusion

The addition of genomic data for the phylogeny of *Panax* radically changes what is known about the evolution of the genus. The implications in terms of phylogeography are still unclear due to missing taxa, and the addition of population data and additional species should improve our insight into the evolutionary history of the genus. The development of species delimitation methods changes perspectives in molecular identification and DNA barcoding by incorporating evolution hypotheses at the species level. The newly proposed molecular markers allow for accurate identification of *Panax* species and enable authentication of ginseng and derived products and monitoring of the ginseng trade, while ultimately aiding conservation of wild ginseng.

Additional files

Additional file 1: Table S1. Voucher specimens. (DOCX 14 kb)

Additional file 2: Table S2. Araliaceae species used for this study and their accession numbers. (PDF 385 kb)

Additional file 3: Figure S1. Fragment analyzer DNA report of *P. bipinnatifidus, P.* sp. *(puxailaileng), P. stipuleanatus, P. vietnamensis* samples, for the genomic DNA (gDNA), for the non-methylated and methylated fractions. (PDF 410 kb)

Additional file 4: Table S3. Partition finder scheme. (DOCX 29 kb)

Additional file 5: Table S4. selected markers and their primer sequences. (DOCX 19 kb)

Additional file 6: Figure S2. Annotated plastid genome for *P. binnatifidus (PDF 442 kb)*

Additional file 7: Figure S3. Annotated plastid genome for *P.* sp. (*puxailaileng*). *(PDF 359 kb)*

Additional file 8: Figure S4. Annotated plastid genome for *P. vietnamensis. (PDF 456 kb)*

Additional file 9: Figure S5. Annotated plastid genome for *P. stipuleanatus. (PDF 440 kb)*

Additional file 10: Figure S7. ML phylogeny for marker *trnC-rps16*. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation. (PDF 108 kb)

Additional file 11: Figure S8. ML phylogeny for marker *trnE-trnM*. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation. (PDF 85 kb)

Additional file 12: Figure S9. ML phylogeny of marker *trnS-trnG*. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation. (PDF 116 kb)

Additional file 13: Figure S10. ML phylogeny of the marker *psbM-trnD*. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation. (PDF 102 kb)

Additional file 14: Figure S11. ML phylogeny for the concatenated matrix with the four markers, *trnC-rps16, trnS-trnG, trnE-trnM* and *psbM-trnD*. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation. (PDF 108 kb)

Additional file 15: Figure S6. Results of the mPTP species delimitation analysis on the full plastid genome matrix. The red lines illustrate the branches representing speciation and the brown lines the branches representing coalescence processes. The numbers on the branches represent the Bayesian posterior probabilities for the delimited species. (PDF 145 kb)

Abbreviations

2D-IR: Two-Dimensional Correlation Infrared Spectroscopy; ABGD: Automatic Barcode Gap Discovery; AP-PCR: Arbitrarily Primed Polymerase Chain Reaction; ASV: Average Support Value; AWTY: Are We There Yet; BOLD: Barcode of Life Database; dNTP: Deoxynucleotide; FT-IR: Fourier Transformed-Infrared Spectroscopy; gDNA: Genomic DNA; GMYC: Generalized Mixed Yule Coalescent; iBOL: International Barcode Of Life; LCP: Lowest Common ancestor; MARMS: Multiplex Amplification Refractory Mutation System-PCR; MASA: Mutant Allele Specific Amplification; MBD2: Methyl-CpG-Binding Domain; MBD2-Fc: Methylated CpG-specific Binding protein (MBD2), fused to the Fc fragment of human IgG; MCMC: Markov Chain Monte Carlo; ML: Maximum Likelihood; mPTP: Multi-rate Poisson Tree Processes; MSC: Multispecies Coalescent; nrDNA: Nuclear Ribosomal DNA; OSF: Open Science Framework; PCR: Polymerase Chain Reaction; PCR-RFLP: PCR-Restriction Fragment Length Polymorphism; PTP: Poisson Tree Processes; RAPD: Random Amplified Polymorphic DNA; RESL: Refined Single Linkage; SNP: Single-Nucleotide Polymorphism

Acknowledgments

The authors wish to thank the following people and organizations, Jarl Andreas Anmarkrud for his assistance for the enrichment, members of de Boer group for their helpful discussions and feedback during manuscript preparation. This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacentre for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. We would like to acknowledge the support of Vietnam Academy of Science and Technology (VAST).

Funding

This project was supported by Vietnam Academy of Science and Technology (grant No. VAST02.01/16–17) and the European Union's Seventh Framework Programme for research, technological development and demonstration under the Grant agreement no. 606895 to the FP7-MCA-ITN MedPlant, "Phylogenetic Exploration of Medicinal Plant Diversity". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The raw sequence data from the *P. bipinnatifidus, P. stipuleanatus,* and *P. vietnamensis* samples have been to submitted to GenBank on the following accessions: SRR5725242, SRR5725240, SRR5725505, SRR5725492, SRR5738925, SRR5738922, SRR5738920, SRR5738920, SRR5738927.The DNA matrix used for the phylogenomic analyses are available on Open Science Framework, (https://doi.org/10.17605/OSF.IO/Z7RWE). The plastome sequences of *P. bipinnatifidus, P. stipuleanatus,* and *P. vietnamensis* (2) have been submitted to NCBI GenBank, (MF377620, MF377621, MF377622, MF377623).

Authors' contributions

The project was conceived and designed by HdB, HLTT, NVH, and VM. NNL performed the laboratory work. VM performed data analysis. AK, VM and HdB drafted the manuscript. All other authors gave useful contribution on the

analysis of data and text of the manuscript. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]The Natural History Museum, University of Oslo, Oslo, Norway. [2]Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.

Received: 26 June 2017 Accepted: 21 March 2018
Published online: 03 April 2018

References

1. Robbins CS. American ginseng: the root of North America's medicinal herb trade: Traffic North America; 1998.
2. Millspaugh CF. American medicinal plants: an illustrated and descriptive guide to plants indigenous to and naturalized in the United States which are used in medicine: Dover Publications; 1892.
3. Kimmens AC. Tales of the ginseng: Morrow; 1975.
4. Zhuravlev YN, Koren OG, Reunova GD, Muzarok TI, Gorpenchenko TY, Kats IL, Khrolenko YA. *Panax ginseng* natural populations: their past, current state and perspectives. Acta Pharmacol Sin. 2008;29(9):1127–36.
5. programme SotcoitiesowffUNe: CITES (convention on international trade in endangered species) handbook: convention on species of wild fauna and flora, July 2016: CITES Secretariat de la Convention sur le commerce international des espèces de faune et de flore sauvages menacées d'extinction; 2016.
6. Basnet D, Dey K. Studies on seed germination of an Indian ginseng (*Panax assamicus* Ban. spec. nov.) for successful cultivation and conservation. Indian J For. 2008.
7. Zhang S, Wang R, Zeng W, Zhu W, Zhang X, Wu C, Song J, Zheng Y, Chen P. Resource investigation of traditional medicinal plant *Panax japonicus* (T. Nees) CA Mey and its varieties in China. J Ethnopharmacol. 2015;166:79–85.
8. Joshi G, Tiwari K, Tiwari R, Uniyal M. Conservation and large scale cultivation strategy of Indian ginseng- *Panax pseudoginseng* wall. Indian Forester. 1991; 117(2):131–4.
9. Jain A. Vulnerable and threatened plants of economic value: *Panax pseudo-ginseng* wall. (The Himalayan Ginseng) MFP News. 1994;4:21.
10. Blundell AG, Mascia MB. Discrepancies in reported levels of international wildlife trade. Conserv Biol. 2005;19(6):2020–5.
11. Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S. DNA barcoding of *Panax* species. Planta Med. 2011;77(02):182–7.
12. Shi F-X, Li M-R, Li Y-L, Jiang P, Zhang C, Pan Y-Z, Liu B, Xiao H-X, Li L-F. The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). BMC Plant Biol. 2015;15(1):297.
13. Komatsu K, Zhu S, Fushimi H, Qui TK, Cai S, Kadota S. Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of *Panax vietnamensis* and five related species. Planta Med. 2001;67(05):461–5.
14. Yap KY-L, Chan SY, Lim CS. Infrared-based protocol for the identification and categorization of ginseng and its products. Food Res Int. 2007;40(5):643–52.
15. Li Y-M, Sun S-Q, Zhou Q, Qin Z, Tao J-X, Wang J, Fang X. Identification of American ginseng from different regions using FT-IR and two-dimensional correlation IR spectroscopy. Vib Spectrosc. 2004;36(2):227–32.
16. Mihalov JJ, Marderosian AD, Pierce JC. DNA identification of commercial ginseng samples. J Agric Food Chem. 2000;48(8):3744–52.
17. Liu D, Li Y-G, Xu H, Sun S-Q, Wang Z-T. Differentiation of the root of cultivated ginseng, mountain cultivated ginseng and mountain wild ginseng using FT-IR and two-dimensional correlation IR spectroscopy. J Mol Struct. 2008;883:228–35.
18. Zhu S, Fushimi H, Cai S, Komatsu K. Species identification from ginseng drugs by multiplex amplification refractory mutation system (MARMS). Planta Med. 2004;70(02):189–92.
19. Park M-J, Kim MK, In J-G, Yang D-C. Molecular identification of Korean ginseng by amplification refractory mutation system-PCR. Food Res Int. 2006;39(5):568–74.
20. Qin J, Leung FC, Fung Y, Zhu D, Lin B. Rapid authentication of ginseng species using microchip electrophoresis with laser-induced fluorescence detection. Anal Bioanal Chem. 2005;381(4):812–9.
21. Kim J, Jo BH, Lee KL, Yoon E, Ryu GH, Chung KW. Identification of new microsatellite markers in *Panax ginseng*. Mol Cells. 2007;24(1):60.
22. Ratnasingham S, Hebert PD. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One. 2013;8(7): e66213.
23. Hao X, Jiang R, Chen T. Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. Bioinformatics. 2011;27(5):611–8.
24. Boyer F, et al. Obitools: a unix–inspired software package for DNA metabarcoding. Molecular Ecology Resources. 2016;16(1):176–82.
25. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 2013;29(22):2869–76.
26. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.
27. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol. 2004;2(10):e312.
28. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol. 2012;21(8):1864–77.
29. Brown SD, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour. 2012;12(3):562–5.
30. Meier R, Zhang G, Ali F. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Syst Biol. 2008;57(5):809–13.
31. Ross HA, Murugan S, Sibon Li WL. Testing the reliability of genetic methods of species identification via simulation. Syst Biol. 2008;57(2):216–30.
32. Kapli P, et al. Multi-rate poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics. 2017;33(11):1630–638.
33. Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol. 2013;62(5):707–24.
34. Yang Z, Rannala B. Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Mol Ecol. 2017;26(11):3028–36.
35. Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. Proc Natl Acad Sci. 2017;114(7):1607–12.
36. Wen J. Evolution of the *Aralia–Panax* complex (Araliaceae) as inferred from nuclear ribosomal ITS sequences. Edinb J Bot. 2001;58(02):243–57.
37. Plunkett GM, Wen J, Lowry Ii PP. Infrafamilial classifications and characters in Araliaceae: insights from the phylogenetic analysis of nuclear (ITS) and plastid (trnL-trnF) sequence data. Pl Syst Evol. 2004;245(1):1–39.
38. Wen J, Zimmer EA. Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. Mol Phylogenet Evol. 1996;6(2):167–77.
39. Hara H. On the Asiatic species of the genus Panax. J Jpn Bot. 1970.
40. Zhou J, Huang W, Wu M, Yang C, Feng K, Wu Z. Triterpenoids from *Panax* Linn. and their relationship with taxonomy and geographical distribution. Acta Phytotaxonomica Sin. 1975;13(2):29–45.
41. Choi H-K, Wen J. A phylogenetic analysis of *Panax* (Araliaceae): integrating cpDNA restriction site and nuclear rDNA ITS sequence data. Pl Syst Evol. 2000;224(1):109–20.
42. Kumar Sharma S, Krishan Pandit M. A new species of *Panax* L. (Araliaceae) from Sikkim Himalaya, India. Syst Bot. 2009;34(2):434–8.
43. Duy NV, Trieu LN, Chinh ND, Tran VT. A new variety of *Panax* (Araliaceae) from Lam Vien Plateau, Vietnam and its molecular evidence. Phytotaxa. 2016;277(1):12.
44. Yi T, Lowry PP, Plunkett GM. Chromosomal evolution in Araliaceae and close relatives. Taxon. 2004;53(4):987–1005.
45. Choi HI, Kim NH, Lee J, Choi BS, Do Kim K, Park JY. Evolutionary relationship of *Panax ginseng* and P. quinquefolius inferred from sequencing and

comparative analysis of expressed sequence tags. Genet Resour Crop Evol. 2013;60:1377–87.
46. Wen J, Plunkett GM, Mitchell AD, Wagstaff SJ. The evolution of Araliaceae: a phylogenetic analysis based on ITS sequences of nuclear ribosomal DNA. Syst Bot. 2001;26(1):144–67.
47. Kim NH, Choi HI, Kim KH, Jang W, Yang TJ. Evidence of genome duplication revealed by sequence analysis of multi-loci expressed sequence tag-simple sequence repeat bands in Panax ginseng Meyer. J Ginseng Res. 2014;38:130–5.
48. Jiang P, Shi F-X, Li Y-L, Liu B, Li L-F. Development of highly transferable microsatellites for *Panax ginseng* (Araliaceae) using whole-genome data. Appl Plant Sci. 2016;4(11):1600075.
49. Ma K-H, Dixit A, Kim Y-C, Lee D-Y, Kim T-S, Cho E-G, Park Y-J. Development and characterization of new microsatellite markers for ginseng (*Panax ginseng* CA Meyer). Conserv Genet. 2007;8(6):1507–9.
50. Van Dan N, Ramchiary N, Choi SR, Uhm TS, Yang T-J, Ahn I-O, Lim YP. Development and characterization of new microsatellite markers in *Panax ginseng* (CA Meyer) from BAC end sequences. Conserv Genet. 2010;11(3):1223–5.
51. Choi H-I, Kim N-H, Kim J-H, Choi B-S, Ahn I-O, Lee J-S, Yang T-J. Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. J Ginseng Res. 2011;35(4):399–412.
52. Liu H, Xia T, Zuo Y-J, Chen Z-J, Zhou S-L. Development and characterization of microsatellite markers for *Panax notoginseng* (Araliaceae), a Chinese traditional herb. Am J Bot. 2011;98(8):e218–20.
53. Joly S, et al. Genetic structure of the American ginseng (*Panax quinquefolius* L.) in Eastern Canada using reduced-representation high-throughput sequencing. Botany. 2016;95(4):429–34.
54. Cruse-Sanders JM, Hamrick J. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L.(Araliaceae). Am J Bot. 2004;91(4):540–8.
55. Bai D, Brandle J, Reeleder R. Genetic diversity in North American ginseng (*Panax quinquefolius* L.) grown in Ontario detected by RAPD analysis. Genome. 1997;40(1):111–5.
56. Schluter C, Punja ZK. Genetic diversity among natural and cultivated field populations and seed lots of American ginseng (*Panax quinquefolius* L.) in Canada. Int J Plant Sci. 2002;163(3):427–39.
57. Cruse-Sanders J, Hamrick J. Spatial and genetic structure within populations of wild American ginseng (*Panax quinquefolius* L., Araliaceae). J Hered. 2004; 95(4):309–21.
58. Grubbs HJ, Case MA. Allozyme variation in American ginseng (*Panax quinquefolius* L.): variation, breeding system, and implications for current conservation practice. Conserv Genet. 2004;5(1):13–23.
59. Boehm C, Harrison H, Jung G, Nienhuis J. Organization of American and Asian ginseng germplasm using randomly amplified polymorphic DNA (RAPD) markers. J Am Soc Hortic Sci. 1999;124(3):252–6.
60. Li S, Li J, Yang X-L, Cheng Z, Zhang W-J. Genetic diversity and differentiation of cultivated ginseng (*Panax ginseng* CA Meyer) populations in North-East China revealed by inter-simple sequence repeat (ISSR) markers. Genet Resour Crop Evol. 2011;58(6):815–24.
61. Reunova GD, Koren OG, Muzarok TI, Zhuravlev YN. Microsatellite analysis of *Panax ginseng* natural populations in Russia. Chin Med. 2014;5(04):231.
62. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotech. 2012;30(5):434–9.
63. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11(5):759–69.
64. Coissac E, et al. From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol. 2016;25(7):1423–28.
65. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J. 2011;9(3):328–33.
66. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. Genomics of chloroplasts and mitochondria: Dordrecht: Springer; 2012. p. 103–26.
67. Givnish TJ, Ames M, McNeal JR, McKain MR, Steele PR, Graham SW, Pires JC, Stevenson DW, Zomlefer WB, Briggs BG. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales1. Ann Mo Bot Gard. 2010;97(4):584–616.
68. Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, Boore JL, Jansen RK. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol Biol. 2006;6(1):77.
69. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009;7(1):84.
70. Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD. Complete plastome sequences of *Equisetum arvense* and Isoetes flaccida: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol Biol. 2010;10(1):321.
71. Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14(1):151.
72. Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston A. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl Plant Sci. 2014;2(9):1400042.
73. Schmickl R, et al. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). Mol Ecol Resour. 2016;16(5):1124–35.
74. Mandel JR, Dikow RB, Funk VA. Using phylogenomics to resolve mega-families: an example from Compositae. J Syst Evol. 2015;53(5):391–402.
75. Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, Kozik A, Michelmore RW, Rieseberg LH, Burke JM. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. Appl Plant Sci. 2014;2(2):1300085.
76. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PLoS One. 2012;7(8):e43808.
77. Steele PR, Hertweck KL, Mayfield D, McKain MR, Leebens-Mack J, Pires JC. Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. Am J Bot. 2012;99(2):330–48.
78. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012;99(2):349–64.
79. Twyford AD, Ness RW. Strategies for complete plastid genome sequencing. Mol Ecol Resour. 2017;17(5):858–68.
80. Obae GS. Nuclear DNA, content and genome size of American ginseng. J Med Plant Res. 2012;6.
81. Pan YZ, Zhang YC, Gong X, Li FS. Estimation of genome size of four *Panax* species by flow cytometry. Plant Diversity Res. 2014;36
82. Du FK, Lang T, Lu S, Wang Y, Li J, Yin K. An improved method for chloroplast genome sequencing in non-model forest tree species. Tree Genet Genomes. 2015;11(6):114.
83. Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci. 2010;107(19):8689–94.
84. Yigit E, Hernandez DI, Trujillo JT, Dimalanta E, Bailey CD. Genome and metagenome sequencing: using the human methyl-binding domain to partition genomic DNA derived from plant tissues. Appl Plant Sci. 2014; 2(11):1400064.
85. Kim K, Nguyen VB, Dong J, Wang Y, Park JY, Lee S-C, Yang T-J. Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 Panax-related species. Sci Rep. 2017;7:4917.
86. Bock DG, Kane NC, Ebert DP, Rieseberg LH. Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol. 2014;201(3):1021–30.
87. Li R, Ma P-F, Wen J, Yi T-S. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PLoS One. 2013;8(10):e78568.
88. Yao X, Liu Y-Y, Tan Y-H, Song Y, Corlett RT. The complete chloroplast genome sequence of Helwingia himalaica (Helwingiaceae, Aquifoliales) and a chloroplast phylogenomic analysis of the Campanulidae. PeerJ. 2016;4:e2734.
89. Wang L, Du X-J, Li X-F. The complete chloroplast genome sequence of the evergreen plant Dendropanax dentiger (Araliaceae). Mitochondrial DNA Part A. 2016;27(6):4193–4.
90. Kim K-J, Lee H-L. Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11(4):247–61.
91. Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z. Comparative chloroplast genomes of Camellia species. PLoS One. 2013;8(8):e73053.
92. Chen Q, Feng X, Li M, Yang B, Gao C, Zhang L, Tian J. The complete chloroplast genome sequence of Fatsia japonica (Apiales: Araliaceae) and the phylogenetic analysis. Mitochondrial DNA Part A. 2016;27(4): 3050–1.

93. Kim K, Lee S-C, Lee J, Lee HO, Joh HJ, Kim N-H, Park H-S, Yang T-J. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within Panax ginseng species. PLoS One. 2015;10(6):e0117159.
94. Zhao Y, Yin J, Guo H, Zhang Y, Xiao W, Sun C, Wu J, Qu X, Yu J, Wang X, et al. The complete chloroplast genome provides insight into the evolution and polymorphism of Panax ginseng. Front Plant Sci. 2015;5:696.
95. Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genet. 2014;15(1):138.
96. Nguyen B, Kim K, Kim Y-C, Lee S-C, Shin JE, Lee J, Kim N-H, Jang W, Choi H-I, Yang T-J. The complete chloroplast genome sequence of *Panax vietnamensis* Ha et Grushv (Araliaceae). Mitochondrial DNA Part A. 2017;28(1):85–6.
97. Zong X, Song J, Lv J, Wang S. The complete chloroplast genome sequence of Schefflera octophylla. Mitochondrial DNA Part A. 2016;27(6):4685–6.
98. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–120.
99. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic acids research. 2013;41(13):e129–e129.
100. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 2007;52(5):267–74.
101. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
102. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
103. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
104. Lanfear R, Calcott B, Ho SY, Guindon S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 2012;29(6):1695–701.
105. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90.
106. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
107. Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 2008;24(4):581–3.
108. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–3.
109. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genomics. 2016;2(4):e000056.
110. Quinlan AR: BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014;11–12.
111. Valcárcel V, Fiz-Palacios O, Wen J. The origin of the early differentiation of Ivies (Hedera L.) and the radiation of the Asian Palmate group (Araliaceae). Mol Phylogenet Evol. 2014;70:492–503.
112. Lee C, Wen J. Phylogeny of *Panax* using chloroplast trnC–trnD intergenic region and the utility of trnC–trnD in interspecific studies of plants. Mol Phylogenet Evol. 2004;31(3):894–903.
113. Álvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol. 2003;29(3):417–34.
114. Bailey C. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol Phylogenet Evol. 2003;29
115. Fehrer J, Gemeinholzer B, Chrtek J, Bräutigam S. Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella hawkweeds* (Hieracium, Cichorieae, Asteraceae). Mol Phylogenet Evol. 2007;42(2):347–61.
116. Soltis DE, Kuzoff RK. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution. 1995;49(4): 727–42.
117. Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J. Phylogenomic analyses of species relationships in the genus Sabal (Arecaceae) using targeted sequence capture. Biol J Linn Soc. 2016;117(1):106–20.
118. Manzanilla V, Bruneau A. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. Mol Phylogenet Evol. 2012;65(1):149–62.
119. Novikova PY, et al. Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat Genet. 2016;48(9):1077.
120. Popp M, Oxelman B. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol Phylogenet Evol. 2001;20(3):474–81.
121. Wendel JF, Doyle JJ. Phylogenetic incongruence: window into genome history and molecular evolution. Molecular systematics of plants II. Boston: Springer; 1998. p. 265–96.
122. Stephens JD, Rogers WL, Mason CM, Donovan LA, Malmberg RL. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. Am J Bot. 2015;102(6):910–20.
123. Raclariu AC, Mocan A, Popa MO, Vlase L, Ichim MC, Crisan G, Brysting AK, de Boer H. Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with Veronica chamaedrys. Front Pharmacol. 2017;8:378.
124. Raclariu AC, Paltinean R, Vlase L, Labarre A, Manzanilla V, Ichim MC, Crisan G, Brysting AK, de Boer H. Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. Sci Rep. 2017;7(1):1291.
125. Veldman S, et al. High-throughput sequencing of African chikanda cake highlights conservation challenges in orchids. Biodivers Conserv. 2017;26(9): 2029–46.

SUPPORTING MATERIAL

Table S1. Araliaceae species used for this study and their accession numbers.

Table S2. Partition finder scheme.

Table S3. selected markers and their primer sequences.

Figure S1. Fragment analyzer DNA report of P. bipinnatifidus, P. sp. (puxailaileng), P. stipuleanatus, P. vietnamensis samples, for the genomic DNA (gDNA), for the non-methylated and methylated fractions.

Figure S2. Annotated plastid genome for P. binnatifidus

Figure S3. Annotated plastid genome for P. sp. (puxailaileng).

Figure S4. Annotated plastid genome for P. vietnamensis.

Figure S5. Annotated plastid genome for P. stipuleanatus.

Figure S6. Results of the mPTP species delimitation analysis on the full plastid genome matrix. The red lines illustrate the branches representing speciation and the brown lines the branches representing coalescence processes. The numbers on the branches represent the Bayesian posterior probabilities for the delimited species.

Figure S7. ML phylogeny for marker trnC-rps16. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation.

Figure S8. ML phylogeny for marker trnE-trnM. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation.

Figure S9. ML phylogeny of marker trnS-trnG. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation.

Figure S10. ML phylogeny of the marker psbM-trnD. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation.

Figure S11. ML phylogeny for the concatenated matrix with the four markers, trnC-rps16, trnS-trnG, trnE-trnM and psbM-trnD. The bootstrap values are represented in italic on the branches. The red branches represent supported species delimitation.

Table S1

NCBI Reference Sequence	Genus	Species	Cultivar	Reference article
KT153023	*Aralia*	*elata*		(Kim et al. 2017)
NC_022810	*Aralia*	*undulata*		(Bock et al. 2014)
KC456163	*Aralia*	*undulata*		(Shi et al. 2015) (Li et al. 2013)
NC_022811	*Brassaiopsis*	*hainla*		(Li et al. 2013)
KC456164	*Brassaiopsis*	*hainla*		(Shi et al. 2015) (Li et al. 2013)
KP271241	*Dendropanax*	*dentiger*		(Shi et al. 2015) (Yao et al. 2016)
NC_026546	*Dendropanax*	*dentiger*		(Wang et al. 2016)
KR136270	*Dendropanax*	*morbifer*		(Kim et al. 2017)
NC_027607	*Dendropanax*	*morbifer*		(Kim et al. 2016)
KT153020	*Eleutherococcus*	*gracilistylus*		NA
JN637765	*Eleutherococcus*	*senticosus*		(Shi et al. 2015) (Li et al. 2013)
NC_016430	*Eleutherococcus*	*senticosus*		(Kim et al. 2017) (Yang et al. 2013)
NC_027685	*Fatsia*	*japonica*		(Chen et al. 2016)
KR021045	*Fatsia*	*japonica*		(Chen et al. 2016)
HM596070	*Hydrocotyle*	*verticillata*		(Shi et al. 2015) (Yao et al. 2016)
KC456167	*Kalopanax*	*septemlobus*		(Shi et al. 2015) (Yao et al. 2016)
NC_022814	*Kalopanax*	*septemlobus*		(Li et al. 2013)
NC_022812	*Metapanax*	*delavayi*		(Li et al. 2013)
KC456165	*Metapanax*	*delavayi*		(Shi et al. 2015) (Li et al. 2013)
MF377620	*Panax*	*bipinnatifidus*		This article.
KM067386	*Panax*	*ginseng*	Cheongsun	(Kim et al. 2015)
KM088019	*Panax*	*ginseng*	Chunpoong	(Kim et al. 2017) (Kim et al. 2015)
KC686331	*Panax*	*ginseng*	damaya	(Kim et al. 2015) (Zhao et al. 2015)
KC686332	*Panax*	*ginseng*	Ermaya	(Shi et al. 2015) (Kim et al. 2015) (Zhao et al. 2015)
KC686333	*Panax*	*ginseng*	gaolishen	(Kim et al. 2015) (Zhao et al. 2015)
KM067387	*Panax*	*ginseng*	Gopoong	(Kim et al. 2015)
KM067388	*Panax*	*ginseng*	Gumpoong	(Kim et al. 2015)
KM067394	*Panax*	*ginseng*	Hwangsook	(Kim et al. 2015)
KM067389	*Panax*	*ginseng*	Jakyung	(Kim et al. 2015)
KM067393	*Panax*	*ginseng*	Sunhyang	(Kim et al. 2015)

NCBI Reference Sequence	Genus	Species	Cultivar	Reference article
KM067390	*Panax*	*ginseng*	Sunone	(Kim et al. 2015)
KM067391	*Panax*	*ginseng*	Sunpoong	(Kim et al. 2015)
KM067392	*Panax*	*ginseng*	Sunun	(Kim et al. 2015)
KM088020	*Panax*	*ginseng*	Yunpoong	(Kim et al. 2015)
AY582139	*Panax*	*ginseng*		(Li et al. 2013) (Yao et al. 2016)
NC_006290	*Panax*	*ginseng*		(Kim et al. 2015) (Zhao et al. 2015)
KF431956	*Panax*	*ginseng*		(Shi et al. 2015) (Zhao et al. 2015)
KP036469	*Panax*	*japonicus*		(Kim et al. 2017)
NC_028703	*Panax*	*japonicus*		(Kim et al. 2017)
KJ566590	*Panax*	*notoginseng*		(Shi et al. 2015) (Dong et al. 2014)
NC_026447	*Panax*	*notoginseng*		(Dong et al. 2014)
KP036468	*Panax*	*notoginseng*		(Kim et al. 2017)
KR021381	*Panax*	*notoginseng*		NA
KT001509	*Panax*	*notoginseng*		NA
KT028714	*Panax*	*quinquefolius*		NA
KM088018	*Panax*	*quinquefolius*		(Kim et al. 2017) (Kim et al. 2015)
NC_027456	*Panax*	*quinquefolius*		(Kim et al. 2015)
NC_030598	*Panax*	*stipuleanatus*		NA
KX247147	*Panax*	*stipuleanatus*		NA
MF377622	*Panax*	*stipuleanatus*		This article.
MF377621	*Panax*	*vietnamensis*		This article.
MF377623	*Panax*	*vietnamensis*		This article.
KU059178	*Panax*	*vietnamensis*		(Nguyen et al. 2017)
KP036470	*Panax*	*vietnamensis*		(Kim et al. 2017)
KP036471	*Panax*	*vietnamensis*		(Kim et al. 2017)
NC_028704	*Panax*	*vietnamensis*		(Kim et al. 2017)
KC456166	*Schefflera*	*delavayi*		(Shi et al. 2015) (Li et al. 2013)
NC_022813	*Schefflera*	*delavayi*		(Li et al. 2013)
NC_029764	*Schefflera*	*heptaphylla*		(Zong et al. 2016)
KT748629	*Schefflera*	*heptaphylla*		(Zong et al. 2016)

Bock D.G., Kane N.C., Ebert D.P., Rieseberg L.H. 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol. 201:1021–1030.

Chen Q., Feng X., Li M., Yang B., Gao C., Zhang L., Tian J. 2016. The complete chloroplast genome sequence of *Fatsia japonica* (Apiales: Araliaceae) and the phylogenetic analysis. Mitochondrial DNA Part A. 27:3050–3051.

Dong W., Liu H., Xu C., Zuo Y., Chen Z., Zhou S. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genet. 15:138.

Kim K., Lee S.-C., Lee J., Lee H.O., Joh H.J., Kim N.-H., Park H.-S., Yang T.-J. 2015. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. PLOS ONE. 10:e0117159.

Kim K., Lee S.-C., Yang T.-J. 2016. The complete chloroplast genome sequence of *Dendropanax morbifera* (Léveillé). Mitochondrial DNA Part A. 27:2923–2924.

Kim K., Nguyen V.B., Dong J., Wang Y., Park J.Y., Lee S.-C., Yang T.-J. 2017. Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. Sci. Rep. 7:4917.

Li R., Ma P.-F., Wen J., Yi T.-S. 2013. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PLOS ONE. 8:e78568.

Nguyen B., Kim K., Kim Y.-C., Lee S.-C., Shin J.E., Lee J., Kim N.-H., Jang W., Choi H.-I., Yang T.-J. 2017. The complete chloroplast genome sequence of *Panax vietnamensis* Ha et Grushv (Araliaceae). Mitochondrial DNA Part A. 28:85–86.

Shi F.-X., Li M.-R., Li Y.-L., Jiang P., Zhang C., Pan Y.-Z., Liu B., Xiao H.-X., Li L.-F. 2015. The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). BMC Plant Biol. 15:297.

Wang L., Du X.-J., Li X.-F. 2016. The complete chloroplast genome sequence of the evergreen plant *Dendropanax dentiger* (Araliaceae). Mitochondrial DNA Part A. 27:4193–4194.

Yang J.-B., Yang S.-X., Li H.-T., Yang J., Li D.-Z. 2013. Comparative chloroplast genomes of *Camellia* species. PLoS One. 8:e73053.

Yao X., Liu Y.-Y., Tan Y.-H., Song Y., Corlett R.T. 2016. The complete chloroplast genome sequence of *Helwingia himalaica* (Helwingiaceae, Aquifoliales) and a chloroplast phylogenomic analysis of the Campanulidae. PeerJ. 4:e2734.

Zhao Y., Yin J., Guo H., Zhang Y., Xiao W., Sun C., Wu J., Qu X., Yu J., Wang X., Xiao J. 2015. The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. Front. Plant Sci. 5.

Zong X., Song J., Lv J., Wang S. 2016. The complete chloroplast genome sequence of *Schefflera octophylla*. Mitochondrial DNA Part A. 27:4685–4686.

Table S2

Subset	Best Model	# sites	Partition names
1	GTR+I+G+X	662	*intron_1, intron_77*
2	GTR+G+X	1316	*psbA, psbE*
3	GTR+G+X	1757	*intron_16, intron_69, intron_3, intron_51*
4	GTR+G+X	2219	*intron_75, intron_31, intron_38, matK*
5	GTR+I+G+X	1856	*intron_4*
6	GTR+I+G+X	5510	*rps16, rps19, atpF, intron_29, ndhA, intron_73*
7	GTR+G+X	2414	*intron_72, intron_115, intron_5*
8	GTR+X	2663	*trnI-GAU, trnI-GAU, trnV-GAC, trnS-UGA, rrn5, trnC-GCA, intron_130, intron_128, trnC-GCA*
9	GTR+X	1239	*intron_30, intron_54, intron_63, intron_6*
10	GTR+G+X	2214	*psbZ, psbD, rps12, intron_97, psbK, petG, petL, petN, psbI*
11	GTR+G+X	1860	*intron_80, intron_110, intron_7, intron_48, intron_123, rps19, intron_120*
12	GTR+G+X	2729	*intron_26, intron_111, rps19, intron_8, intron_41*
13	GTR+I+G+X	2419	*atpB, psaC, intron_101, trnS-GCU, intron_131*
14	GTR+G+X	1725	*intron_84, intron_74, intron_9, intron_119*
15	GTR+I+G+X	3465	*trnG-UCC, intron_19, intron_27, intron_85, intron_24*
16	GTR+X	380	*intron_10, psbM*
17	GTR	988	*trnS-GGA, trnN-GUU, rrn4_5, trnL-CAA, trnC-GCA, trnT-GGU, trnI-CAU, trnI-CAU, trnR-UCU, psbN, psbF*
18	GTR+I+G+X	1649	*rpl16, intron_11, intron_90*
19	GTR+G+X	4693	*intron_44, atpA, rps4, petA, rps2*
20	GTR+G+X	2067	*intron_13, intron_83, intron_20, intron_57, intron_21*
21	GTR+G+X	5095	*ndhC, atpH, psaA, psaB*
22	GTR+G+X	1239	*intron_14*

23	GTR+I+G+X	2752	*rpl33, atpI, ndhI, ndhJ, psbJ, rpl32, ndhG*
24	GTR+G+X	2245	*intron_43, intron_22, intron_15, intron_70*
25	GTR+I+G+X	6513	*intron_76, trnT-UGU, rpoC2, trnV-UAC, ndhH, infA*
26	GTR+G+X	1256	*psbT, atpE, rpl22, intron_17*
27	GTR+I+G+X	2934	*rpoC1*
28	GTR+G+X	3218	*rpoB*
29	GTR+X	308	*intron_64, psbM, rps19*
30	GTR+G+X	3113	*intron_23, intron_53, intron_61, intron_56*
31	GTR+I+G+X	4948	*rrn23, rrn16, rrn4_5_2, trnL-CAA, trnD-GUC, trnS-GGA, trnW-CCA, trnE-UUC, trnW-CCA, trnW-CCA*
32	GTR+I+G+X	1550	*rrn5, trnY-GUA, intron_102, trnG-UCC, trnV-GAC, trnA-UGC, trnL-UAG, intron_104, trnM-CAU*
33	GTR+I+G+X	1519	*petB, intron_25*
34	GTR+I+G+X	3125	*intron_124, psbC, psbB*
35	GTR+G+X	2110	*intron_52, intron_122, intron_32, intron_37, intron_112*
36	GTR+I+G+X	1801	*intron_34, ndhD, intron_33*
37	GTR+G+X	2147	*intron_99, rps14*
38	GTR+G+X	2555	*intron_78, intron_66, intron_50, intron_36*
39	GTR+I+G+X	3993	*psaI, rps3, rpl20, ycf3, cemA, rpl36*
40	GTR+X	500	*intron_39, intron_141*
41	GTR+I+G+X	2249	*petD, trnL-UAA, ndhE, rps19*
42	GTR+X	996	*intron_42, intron_65, intron_55, intron_47*
43	GTR+G+X	1840	*intron_46, intron_62, intron_114, intron_117*
44	GTR+X	1975	*rps18, trnT-GGU, psbL, psbT, intron_135, intron_96*
45	GTR+I+G+X	1466	*rbcL*
46	GTR+G+X	2104	*accD, ycf4*
47	GTR+G+X	158	*petL*
48	GTR+X	919	*rps11, intron_58, rpl14*

49	GTR+G+X	344	*intron_68*
50	GTR+I+G+X	3130	*rpoA, ccsA, rps15, intron_71*
51	GTR+I+G+X	2106	*clpP*
52	GTR+G+X	910	*rps8, rps19, psbH*
53	GTR+I+G+X	288	*intron_87, intron_79*
54	GTR+G+X	113	*intron_81*
55	GTR+X	351	*intron_82, intron_118*
56	GTR+G+X	87	*rps19*
57	GTR+I+G+X	207	*intron_89*
58	GTR+X	3370	*intron_105, intron_132, ycf15, rpl2, rpl23, intron_138, intron_129, rps7*
59	GTR+I+G+X	2031	*ycf1, intron_139, intron_92*
60	GTR+X	2601	*ndhB, rpl23, intron_93*
61	GTR+I+G+X	13807	*ycf2, intron_106, intron_108, ycf2, intron_126*
62	GTR+G+X	3152	*intron_133, intron_103, ycf15*
63	GTR+I+G+X	319	*intron_95*
64	GTR+X	3009	*ndhB, rps12*
65	GTR+X	2725	*intron_100, intron_127, intron_134, rpl2, rps7*
66	GTR+I+G+X	4390	*rrn23, rrn16, trnS-GGA*
67	GTR+I+G+X	1254	*intron_107, intron_125*
68	GTR+I+G+X	2371	*ndhF*
69	GTR+G+X	970	*intron_113*
70	GTR+I+G+X	119	*intron_116*
71	GTR+I+G+X	6186	*ycf1*
72	GTR+I+G+X	887	*trnA-UGC*
73	GTR+I+G+X	319	*intron_136*

Figure S1

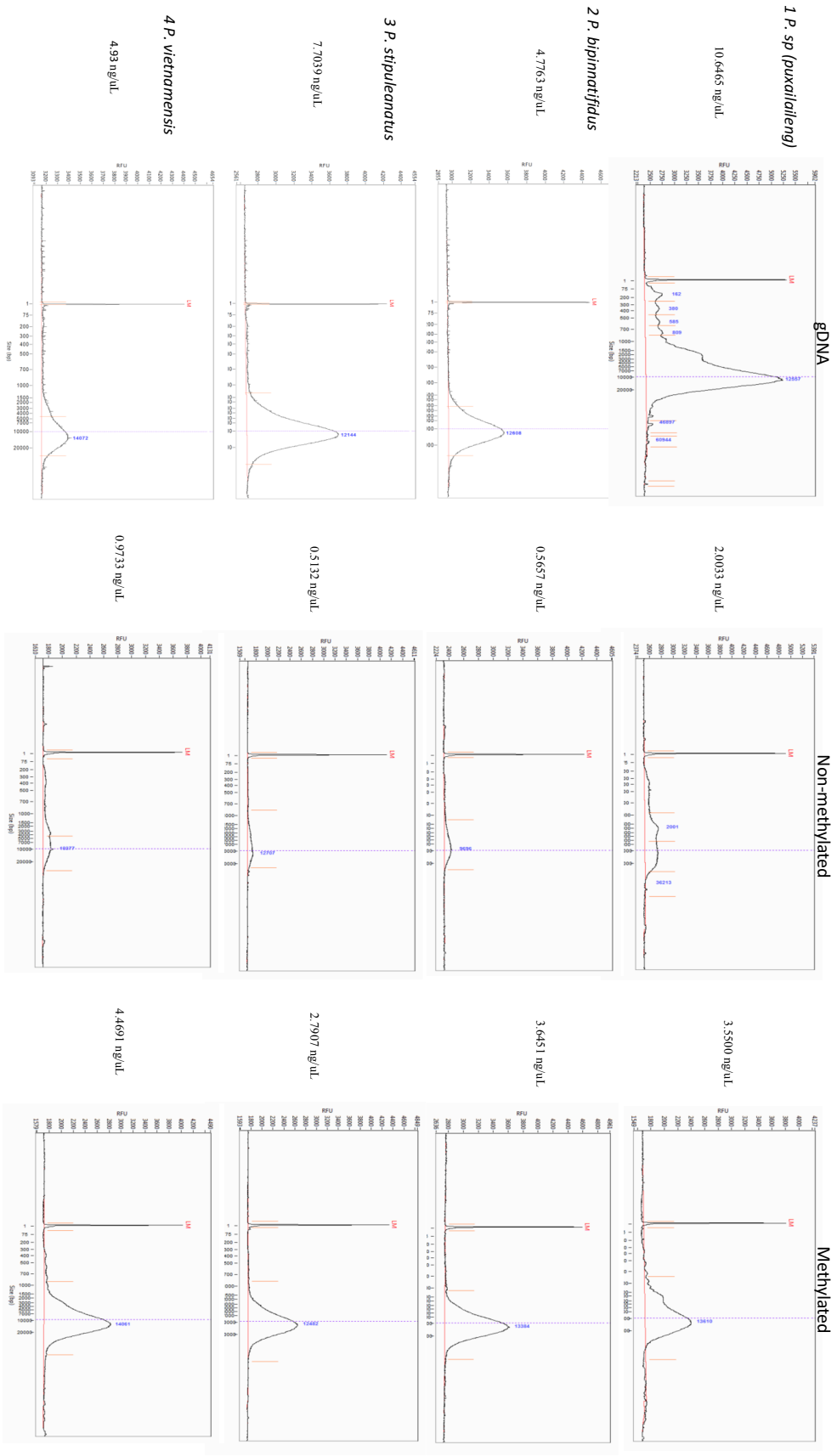

1 P. sp (puxailaileng)
2 P. bipinnatifidus
3 P. stipuleanatus
4 P. vietnamensis
gDNA
Non-methylated
Methylated
10.6465 ng/uL
4.7763 ng/uL
7.7039 ng/uL
4.93 ng/uL
2.0033 ng/uL
0.5657 ng/uL
0.5132 ng/uL
0.9733 ng/uL
3.5500 ng/uL
3.6451 ng/uL
2.7907 ng/uL
4.4691 ng/uL

Figure S2

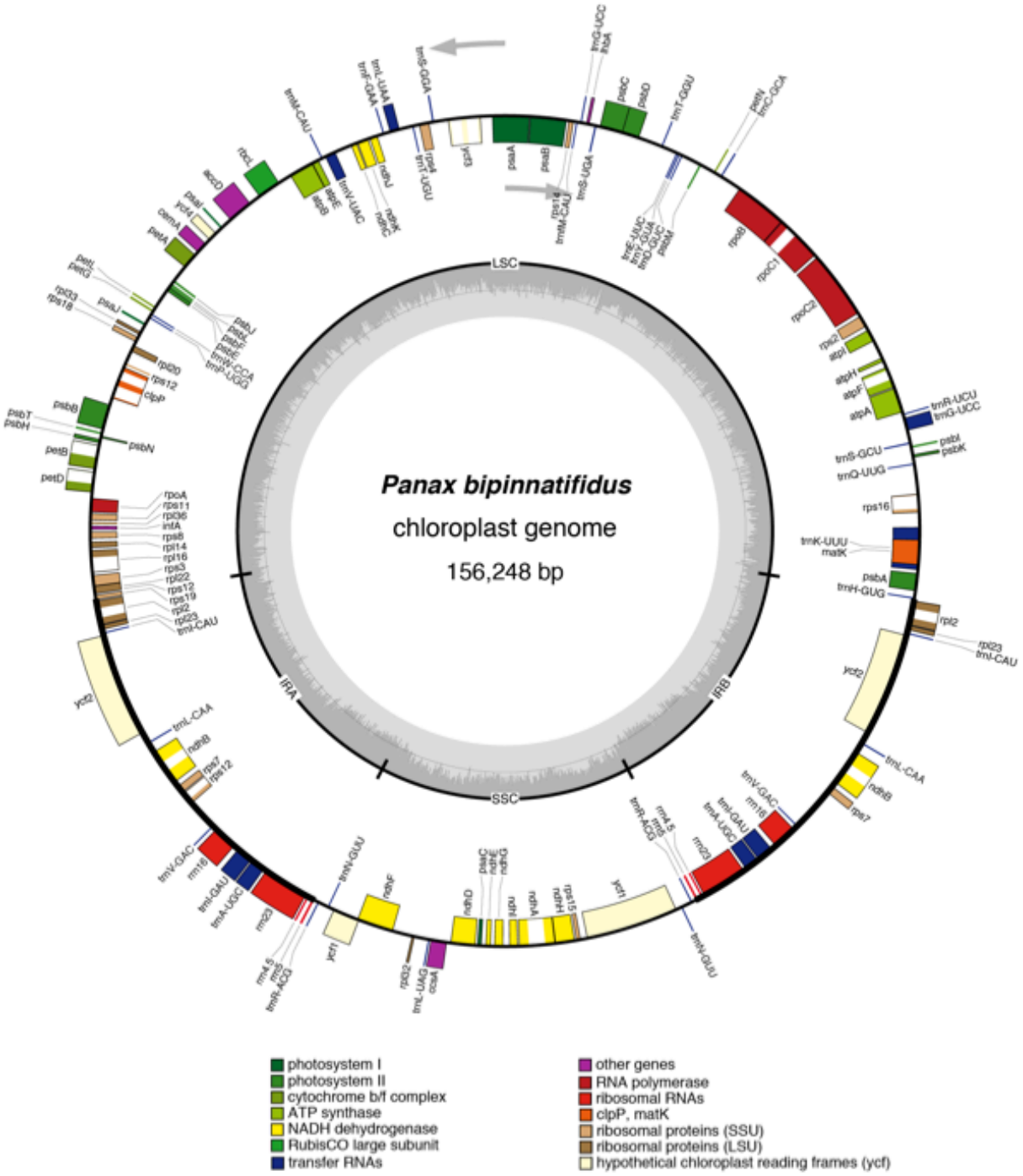

Panax bipinnatifidus
chloroplast genome
156,248 bp
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
transfer RNAs
other genes
RNA polymerase
ribosomal RNAs
clpP, matK
ribosomal proteins (SSU)
ribosomal proteins (LSU)
hypothetical chloroplast reading frames (ycf)

Figure S3

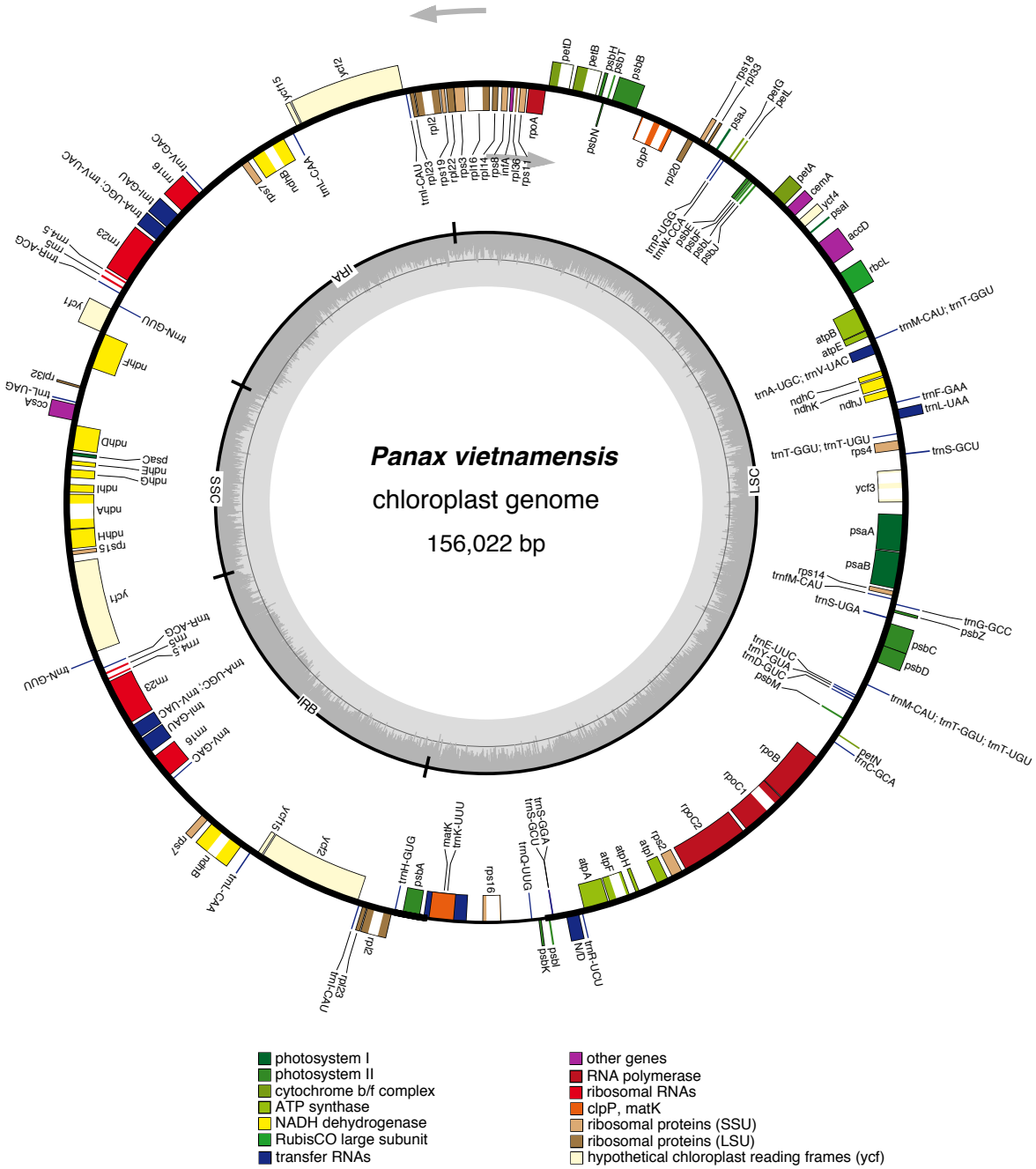

Panax vietnamensis
chloroplast genome
156,022 bp
IRA
IRB
LSC
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
transfer RNAs
other genes
RNA polymerase
ribosomal RNAs
clpP, matK
ribosomal proteins (SSU)
ribosomal proteins (LSU)
hypothetical chloroplast reading frames (ycf)

Figure S4

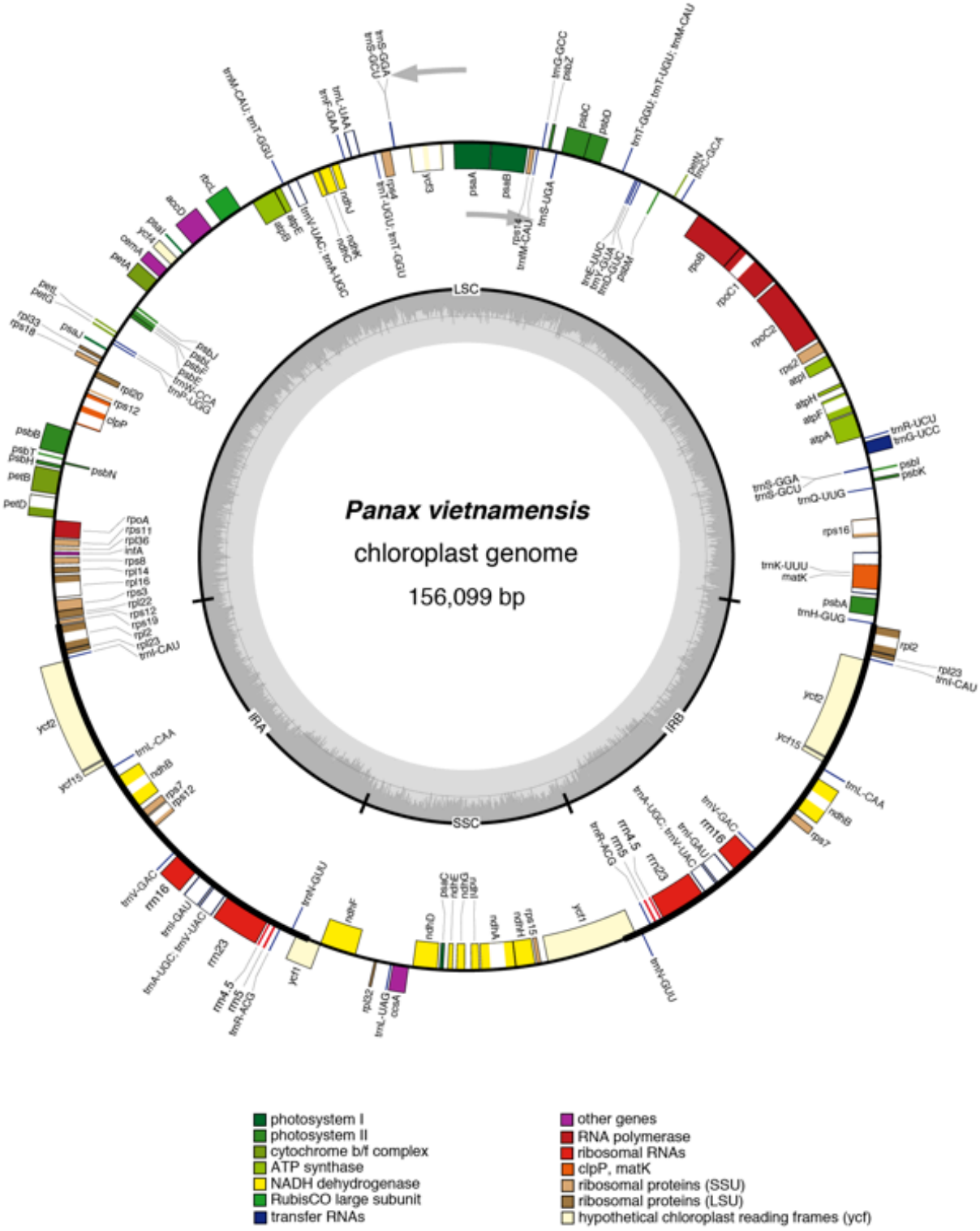

Panax vietnamensis
chloroplast genome
156,099 bp
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
transfer RNAs
other genes
RNA polymerase
ribosomal RNAs
clpP, matK
ribosomal proteins (SSU)
ribosomal proteins (LSU)
hypothetical chloroplast reading frames (ycf)

Figure S5

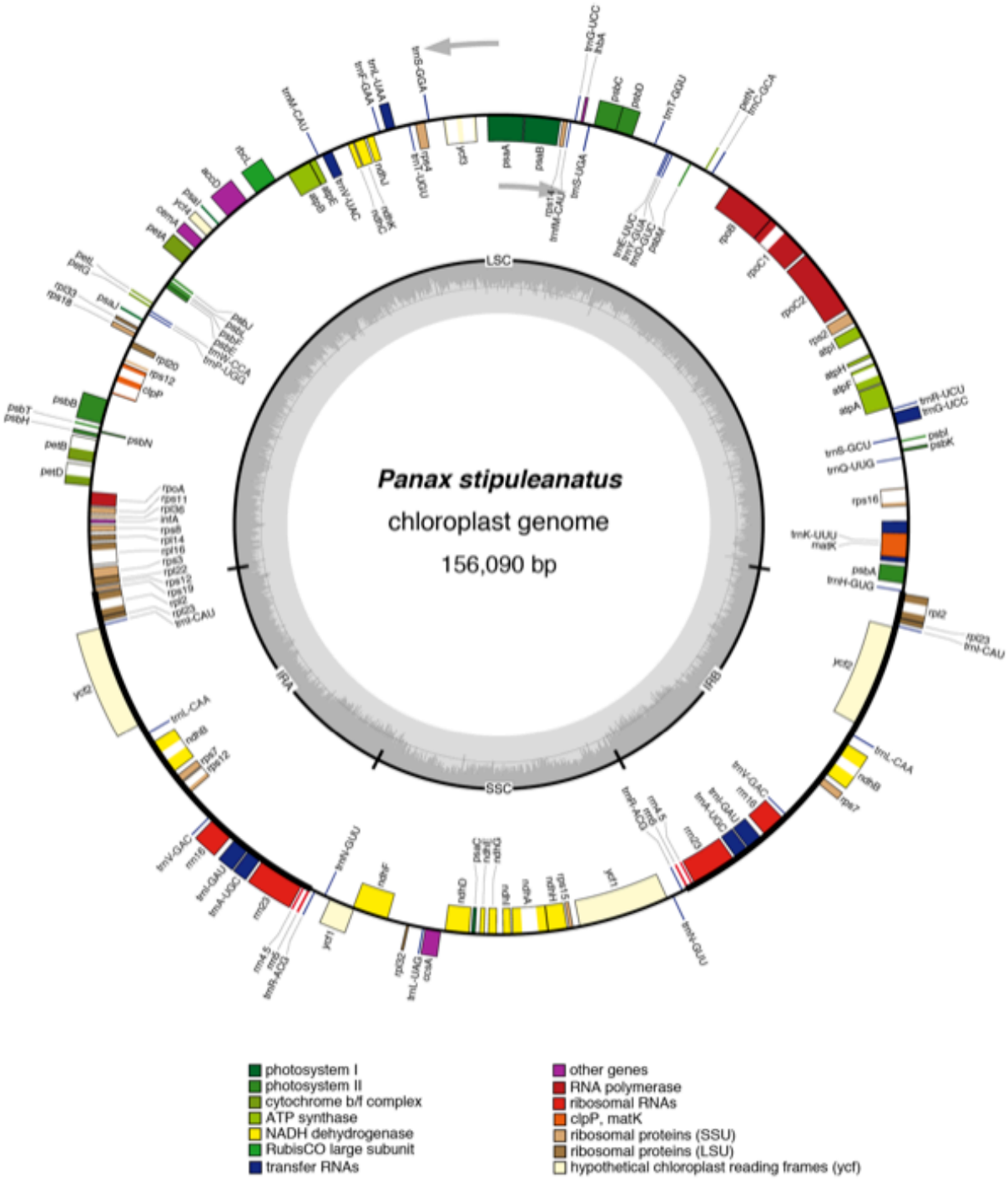

Panax stipuleanatus
chloroplast genome
156,090 bp
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
transfer RNAs
other genes
RNA polymerase
ribosomal RNAs
clpP, matK
ribosomal proteins (SSU)
ribosomal proteins (LSU)
hypothetical chloroplast reading frames (ycf)

Figure S6

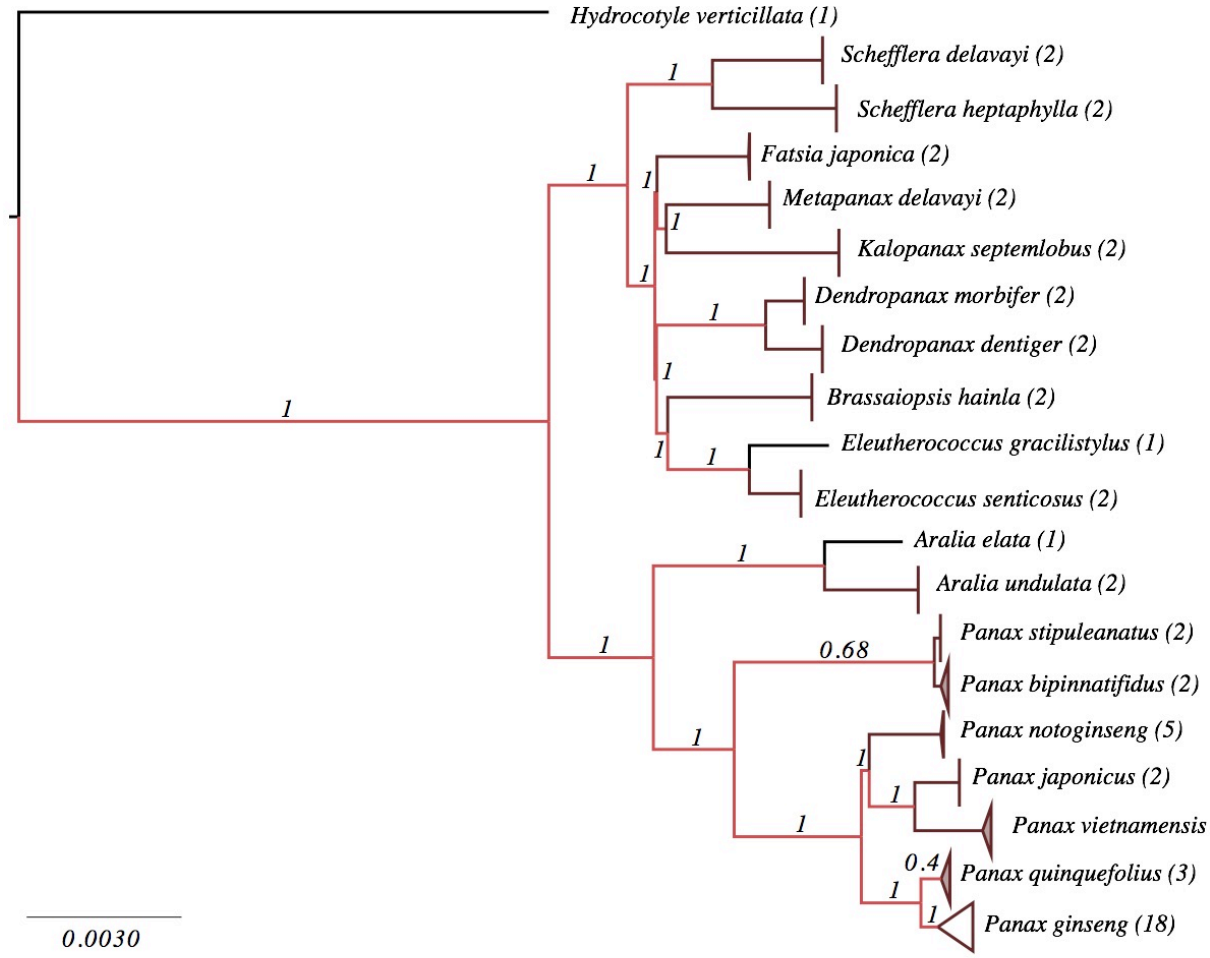

Hydrocotyle verticillata (1)
Schefflera delavayi (2)
Schefflera heptaphylla (2)
Fatsia japonica (2)
Metapanax delavayi (2)
Kalopanax septemlobus (2)
Dendropanax morbifer (2)
Dendropanax dentiger (2)
Brassaiopsis hainla (2)
Eleutherococcus gracilistylus (1)
Eleutherococcus senticosus (2)
Aralia elata (1)
Aralia undulata (2)
Panax stipuleanatus (2)
Panax bipinnatifidus (2)
Panax notoginseng (5)
Panax japonicus (2)
Panax vietnamensis
Panax quinquefolius (3)
Panax ginseng (18)
0.68
0.4
0.0030

Figure S7

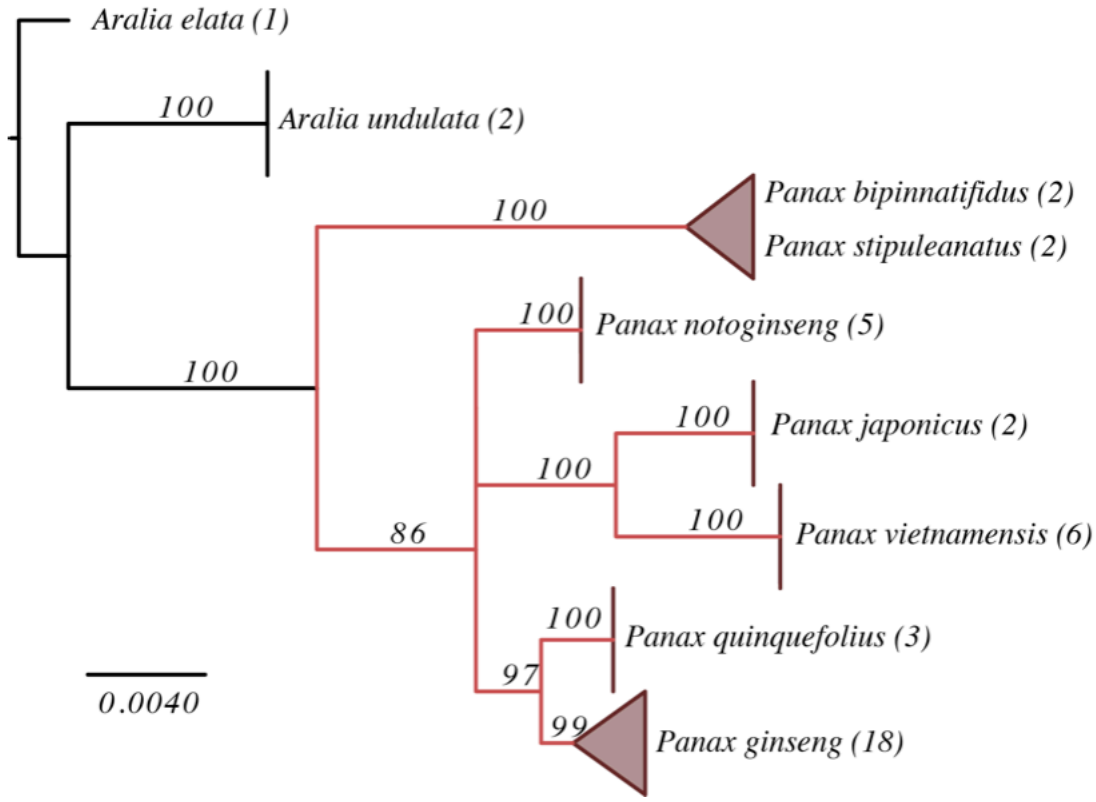


Figure S8

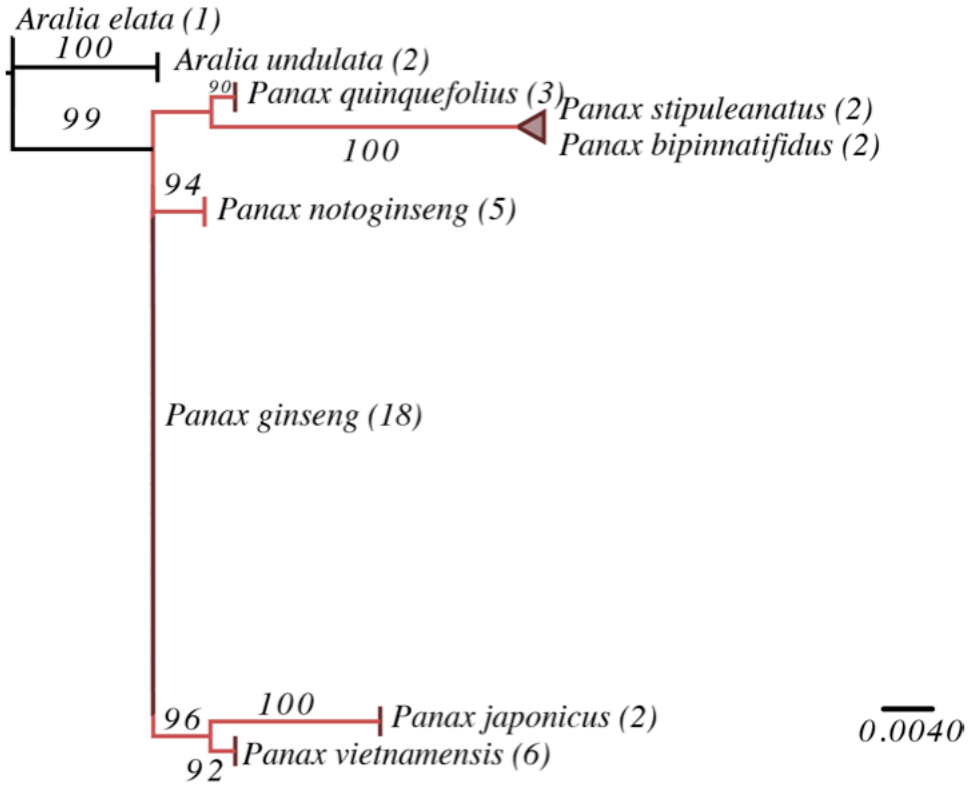


Figure S9

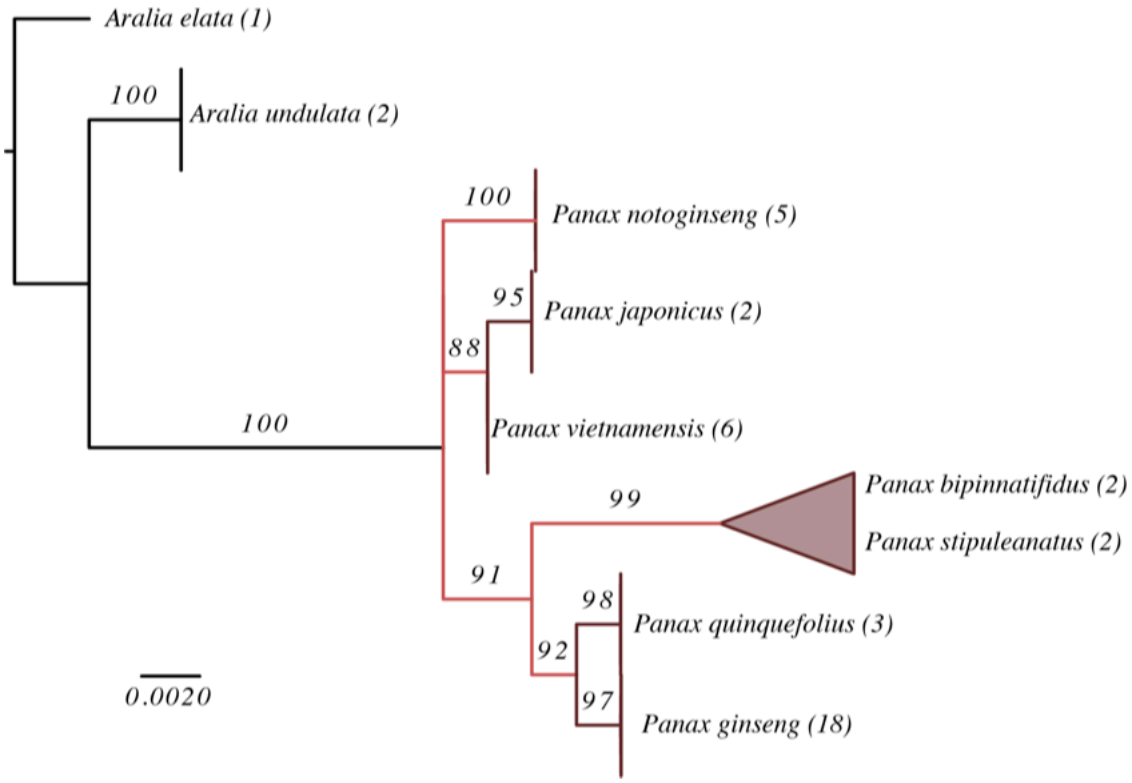
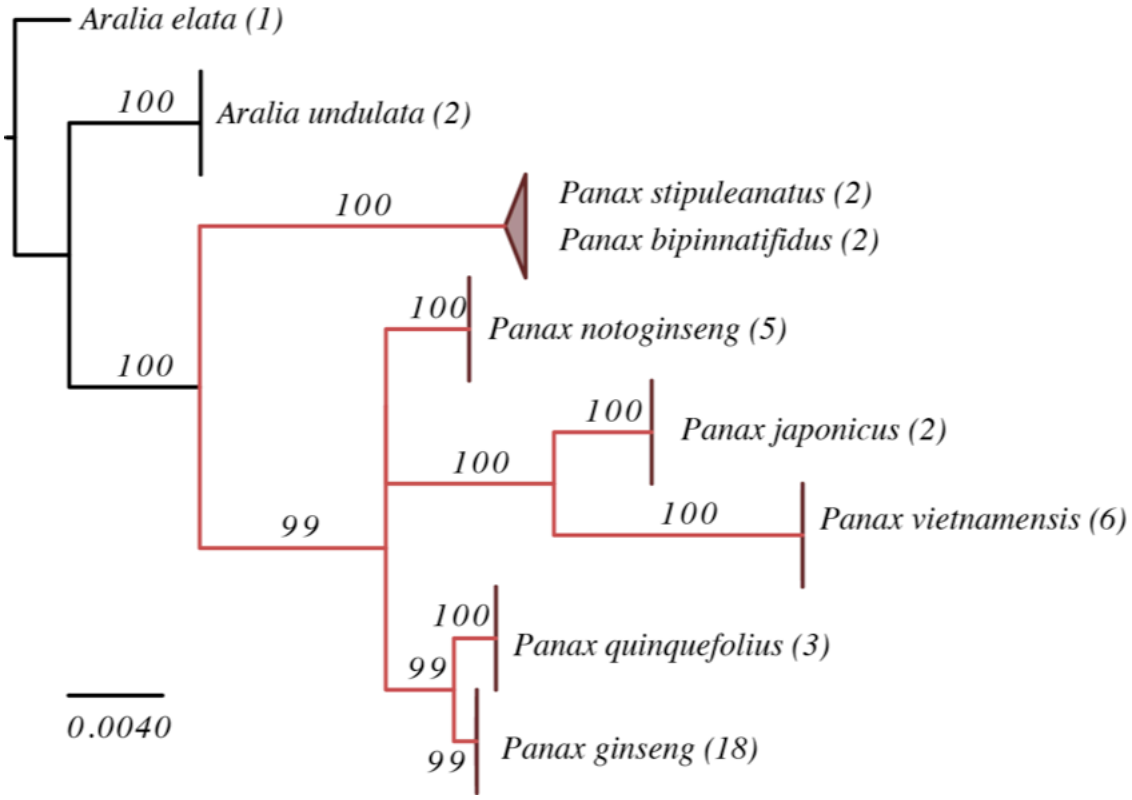


Figure S10



IV

V