DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade

Hugo J. de Boer[1,2,3,*], Abdolbaset Ghorbani[2], Vincent Manzanilla[1], Ancuta-Cristina Raclariu[1], Anna Kreziou[4], Sarawut Ounjai[5], Maslin Osathanunkul[5], Barbara Gravendeel[3]

[1]Natural History Museum, University of Oslo, Oslo, Norway
[2]Department of Organismal Biology, Uppsala University, Uppsala, Sweden
[3]Naturalis Biodiversity Center; Leiden, the Netherlands
[4]Argonafton 30, Thessaloniki, Greece
[5]Department of Biology, Chiang Mai University, Chiang Mai, Thailand

Abstract [191/200]

In eastern Mediterranean countries orchids continue to be collected from the wild for the production of salep, a beverage made of dried orchid tubers. In this study we used nrITS1 and nrITS2 DNA metabarcoding to identify orchid and other plant species present in 55 commercial salep products purchased in Iran, Turkey, Greece, and Germany. Thirty samples yielded a total of 161 plant taxa, and 13 products (43%) contained orchid species and these belonged to 10 terrestrial species with tuberous roots. Another 70% contained the substitute ingredient *Cyamopsis tetraganoloba* (Guar). DNA metabarcoding using the barcoding markers nrITS1 and nrITS2 shows the potential of these markers and approach for identification of species used in salep products. The analysis of interspecific genetic distances between sequences of these markers for the most common salep orchid genera shows that species level identifications can be made with a high level of confidence. Understanding the species diversity and provenance of salep orchid tubers will enable the chain of commercialization of endangered species to be traced back to the harvesters and their natural habitats, and thus allow for targeted efforts to protect or sustainably use wild populations of these orchids.

Keywords: CITES; DNA Metabarcoding; Food safety; Forensic Botany; Orchids; Salep

1. Introduction

Tuberous terrestrial orchids have long been used as medicine and dietary supplements in different parts of the world [1–4]. Orchids are a significant source of nourishment for people in many places, where the starch-rich tubers or pseudobulbs are collected, processed and eaten [5]. In Tanzania, Zambia and Malawi, for example, tubers of terrestrial orchids are used in making a staple food called chikanda [6]. In the eastern Mediterranean, dried tubers of terrestrial orchids are known as salep, which also refers to ground tuber powder and the beverage made from this powder. Salep powder is used in ice-cream production, confectionery and beverages [7,8]. In Greece it is used mainly in a beverage known as *salepi*, which is sold in local markets and is popular as a warming drink during the winter [9].

The orchid tubers for salep originate from wild populations in mainly Turkey, Iran and Greece, and are indiscriminately collected and traded [7,8,10–12]. Harvested tubers are washed in water, boiled in either water or milk, sundried and traded as dried tubers or powder [7,11]. It has been estimated that as much as 30 tons of orchid tubers are harvested annually in Turkey, corresponding to the harvest of 30–120 million individuals [7,13]. In Iran, where orchid tubers are traditionally hardly consumed, an orchid boom is underway in which an estimated 5.5-6.1 million orchids are harvested annually for export to Turkey [11]. In Greece, recent catalysts such as the increasing demand for traditional, organic and alternative foodstuffs have led to a revival of salep consumption, and driven salep prices up to 55–150 euro per kilo [12].

Increasing popularity of salep has raised the demand for salep tubers, and has exacerbated overharvesting of wild orchid populations [12,14]. Whilst alternatives such as cereal starch or synthetic carboxymethyl-cellulose (CMC) are currently common [8,15], the demand for orchid tubers has remained high for those seeking authentic salep [7]. Scarcity of wild orchids in Turkey has forced traders to tap into new sources in adjacent countries [16]. Due to conservation concerns, orchid tuber collection is illegal in Greece, Turkey and Iran, but collection bans are poorly enforced [12,16]. All orchid species are included by the Convention on International Trade of Endangered Species of Fauna and Flora (CITES) on Appendices I or II [17], which means that international trade of products from these species requires specific permits. This large-scale yet poorly visible trade makes it difficult to know which species are targeted and in what quantities. Adequate monitoring would enable identification of priority species for conservation, curbing overexploitation, and targeting high-value species for cultivation. Morphology-based approaches cannot accurately distinguish dried tubers from different genera as tubers from most genera are homogenous in characters [16]. The only genus that can be readily distinguished from other terrestrial tubers is that of *Dactylorhiza* which are palmate in shape. Adulteration with tubers from other terrestrial species has been reported and is a potential health hazard if toxic species are used [11].

The use of DNA barcoding and metabarcoding for the identification of commercialized plant products has evolved with advances in molecular biology and sequencing [18–21], as it can be used to identify and discern taxa at any developmental or processed stage from which DNA

can be extracted [22,23]. DNA metabarcoding is defined as high-throughput multi-taxa identification using the extracellular and/or total DNA extracted from environmental and/or complex DNA sample [24–26]. 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 [27–31]. The combination of these three characteristics has made *trnL* intron P6 loop the marker of choice for ancient DNA and ancient sediment DNA metabarcoding studies [26,27,31–33]. Taberlet et al. [34] 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. The use of nrITS1 and nrITS2 has been limited due to the shorter read length of previous high-throughput sequencing platforms. nrITS is a multicopy nuclear ribosomal marker and concerted evolution make it less suitable for phylogenetic reconstructions [35–37]. However, the markers have been advocated for species-level plant DNA barcoding in taxa-specific studies, as the identification is based on matching query and reference sequences [38,39]. There is a potential to overestimate species richness in diversity studies, at least in the absence of an extensive DNA reference library and associated protocols to account for intragenomic variants [40,41]. Several DNA metabarcoding studies have been published that use nrITS1 or nrITS2 for the identification of fungi [42–45], plants [46,47] or herbal medicines [48–51]. nrITS primers are not truly universal as highlighted by the differences in species composition detected using nrITS1 and nrITS2 on multi-taxa herbal medicines [50,51].

This study takes a novel approach by focusing on nrITS DNA metabarcoding of salep, a complex multi-ingredient food product, made of ground pure or mixed and processed orchid tubers. DNA metabarcoding for species identification has so far not been attempted in orchid trade, and identification of the constituent species in salep has hitherto been impossible. Species level identification of orchid species used in prepared salep would allow us to identify which species are targeted the most, detect the presence of rare, threatened or narrow endemics, and enable us to identify priority species for conservation efforts. The objectives of this study were to: 1) use high-throughput sequencing to determine orchid species composition in salep, 2) evaluate sequencing marker efficacy; 3) investigate species diversity in salep and determine commonly added spices, adulterants and substitutes, 4) study the prevalence of adulterants in salep, and 5) study the prevalence of endangered species in salep.

2. Methods

2.1. Sample collection

Fifty-five processed salep samples were randomly purchased to represent commercially available salep products with the objective to assess the level of adulterated and true orchid based products. The sampling included a representation of producers and vendors, e.g., supermarkets, herbal stores, pharmacies, markets (Table S1). Samples were marketed as loose weight powder (29), processed and packed commercial powders (23), warm salep beverage (2) and salep ice-

cream (1). Products were purchased in Iran (19), Germany (15), Greece (12), and Turkey (9). Four of the products claimed only to contain salep flavor according to the label, whereas the rest was claimed to include genuine salep by the label or vendor.

2.2. Identification of samples by DNA metabarcoding

Total DNA was extracted from all samples in small batches with extraction blanks using the CTAB protocol [52]. Plant DNA was amplified using two plant specific primers pairs for nrITS1 and nrITS2 [53], and in-silico amplification with EcoPCR [54] of GenBank nrITS data was used to determine the suitability of each primer pair in amplifying target orchid species and common expected adulterants. The ITS amplicons were sequenced on an Ion-Torrrent Personal Genome Machine with Ion 316 v2 Chips. FASTQ read files were processed using the HTS-barcode-checker pipeline [55] available as a Galaxy pipeline at the Naturalis Biodiversity Center (http://145.136.240.164:8080/). PRINSEQ [56] was used to inspect read lengths, Phred base qualities and mean quality scores [57]. Reads were selected with a minimal 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 [58] was used to cluster reads into molecular operational taxonomic units (MOTUs) defined by a sequence similarity of >99 % and a minimum number of 2 reads. The consensus sequences of non-singleton MOTUs were queried using BLAST [59] against a local copy of the NCBI/GenBank nucleotide database, with a maximum e-value of 0.05, a minimum hit length of 100 bp and sequence identity of >97 %. Data from samples yielding less than 0.5 ng/µl DNA (Table S1) are analyzed, presented and discussed separately in Table S2 and Figure S1. Detailed methods are available in Text S1 and all raw data as well as all MOTUs and their taxon assignments are deposited in DRYAD doi:10.5061/dryad.5q447.

2.3. Presence, abundance and correlation across samples

Comparison of the species diversity per sample gives an insight into the dominant species in commercial salep products, including those added for texture or flavor of the product. To enable comparison between samples read abundances were normalized using the *standard scale* function in Seaborn 0.7.1 [60] where each column was subtracted by its minimum value and the value divided by its maximum. As a result, the read counts are transformed into a proportion of reads found per species within each sample. The distances between each pair of values in the normalized matrix was subsequently calculated using Euclidean distances, and the hierarchical cluster analysis was done with the UPGMA algorithm (Figure 1). Pearson's correlation was used to test for correlation between the 30 most abundant of taxa across the different salep samples. The matrix of correlation provides an overview of the correlation between the 30 most frequently found species across all the samples (Figure 2).

3. Results

3.1. DNA extraction, PCR amplification and High throughput sequencing

Qubit measurements of the 55 samples gave results for 35 samples with DNA concentrations ranging from 0.5 – 60.4 ng/μl, and for 20 samples concentrations of less than 0.5 ng/μl or not measureable at all. Salep is supposed to be thick and creamy from polysaccharides in the orchid tubers, and all products, even adulterated ones, can therefore be expected to be rich in starches. Nineteen samples yielded a gelatinous DNA extract, and 12 of these did not have measureable DNA concentrations. The results for twenty samples that yielded less than 0.5 ng/μl DNA (Table S1) are presented, analyzed and discussed separately in Table S2 and Figure S1. Results for pilot PCR amplification reactions show a success rate of 64 % (35/55) for nrITS1 and 65 % (36/55) for nrITS2. Thirty samples (55 %) yielded products for both nrITS1 and nrITS2, 5 (9 %) only for nrITS1, 6 (11 %) only for nrITS2, and 14 (25 %) for neither nrITS1 nor nrITS2. For the samples that had no measurable DNA, PCR reactions yielded results respectively for 6 (35 %), 3 (18 %), 4 (24 %), and 4 (24 %) samples. The extraction blanks yielded no amplicons with nrITS1 and nrITS2 primers. PGM chip one, with samples 1-27, had an ion sphere particle (ISP) loading of 88 % and yielded 2,873,882 reads in a final library with a median length of 333 bp. PGM chip two, with samples 28-55, had an ISP loading of 38 % and yielded 1,321,299 reads with a median length of 300 bp. Sequencing success rates were 85.44 % (46/55 samples) for nrITS1 and 87.27 % (48/55 samples) for nrITS2.

3.2. Molecular identification of amplicon MOTUs

For the remaining 35 samples a dataset was obtained comprising 141,285 sequences for nrITS1 and 723,352 sequences for nrITS2. Samples 1, 37, 43, 45, 51 and the extraction blanks yielded no MOTUs for either nrITS1 or nrITS2 and are excluded from the results and discussion. For nrITS1, we found 89 plant taxa (86 at species level and 3 at genus level), and for nrITS2 103 plant taxa (101 at species level and 2 at genus level). Reads and identifications per marker were merged per sample for further analyses, and a total 161 plant taxa (157 at species level and 4 at genus level) were identified (Table S3). Reads per species for nrITS1, nrITS2 and merged per sample). Species detected per salep sample ranged from 1 to 55, with an average of 14.7 species per sample. The following five species were found in over 40 % of the samples: *Cyamopsis tetragonoloba* (L.) Taub., guar bean (70 %), *Triticum dicoccoides* (Körn. ex Asch. & Graebn.) Schweinf., emmer wheat (60 %), *Ipomoea pes-tigridis* L., morning glory (50 %), *Aegilops lorenti* Hochst., Lorent's goatgrass (47 %), *Triticum durum* Desf., durum wheat (47 %), *Secale cereale* L., rye (43 %), and *Triticum aestivum* L., common wheat (43 %). These species were present in resp. 100, 75, 75, 75, 75, 75, and 75 % of the samples that claimed only to include salep flavoring. Plant taxa present in more than 20 % of samples are listed in Table S4.

[Insert] **Figure 1.** Detection of species in salep. Species (y-axis) are colored by relative abundance of normalized read numbers. Species are categorized in gelatinous species (blue) and

non gelatinous species (red), and clustered by Euclidean distances. Salep samples (x-axis) are numbered and grouped by country of purchase.

[Insert] **Figure 2.** Pearson‘s correlation heat map showing correlation of 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 ($r \approx 0$). The histogram in the color key represents the density of the Pearson's correlation coefficients across the matrix.

[Insert] **Figure 3.** Orchid species detected in samples (presence as % in total number of samples).

The detected species can be categorized into species that are rich in starch, and thus suitable as gelatinizing agents for thickening salep, and those that are not. The gelatinous species include orchids, such as *Anacamptis morio* and *Orchis mascula*, cereal crops, such as common wheat, emmer wheat, durum wheat, rye, barley, maize, and legumes, especially guar gum. The non-gelatinous species include a large number of species that are spices and probably added intentionally to flavor the salep, such as ginger, coriander, cinnamon, anise, nigella, mahaleb cherry, poppy and saffron (Figure 1). Figure 2 shows that the widespread use of guar bean flour is strongly correlated with that of morning glory (*I. pes-tigridis* and *I. eriocarpa* R.Br.), maize and mahaleb cherry, but surprisingly not with common cereal substitutes (*S. cereale* and *Triticum* spp.). The use of *Dactylorhiza* species is correlated with the use of the spices ginger and saffron.

Salep orchids were present in 43% (13 out 30) samples, with *Dactylorhiza* being the most common genus present in 18 samples, followed by *Anacamptis* in 5, *Gymnadenia* in 4, and *Orchis* in 1 sample. A total of 12 orchid species were identified: *Anacamptis morio* subsp. *longicornu* (Poir.) H.Kretzschmar, Eccarius & H.Dietr., *Dactylorhiza fuchsii* (Druce) Soó, *Dactylorhiza incarnata* (L.) Soó, *Dactylorhiza majalis* (Rchb.) P.F.Hunt & Summerh., *Dactylorhiza romana* (Sebast.) Soó, *Dactylorhiza saccifera* (Brongn.) Soó, *Dactylorhiza sambucina* (L.) Soó, *Gymnadenia conopsea* (L.) R.Br., *Gymnadenia* x *densiflora* (Wahlenb.) A.Dietr., and *Orchis mascula* (L.) L. (Figure 3).

Turkish samples contained salep orchids in only 1 out of 7 samples, followed by Germany with 3 out of 11, Iran 2 out of 4 and Greece with 7 out of 8. Orchid species were detected in 5 out of 14 powders, 7 out of 19 processed products, 0 out of 1 drinks, and 1 out of 1 ice cream. The sample exclusion threshold of 0.5 ng/µl DNA excluded nearly all samples from Iran (79 %) and these were mostly ground salep powders (46 %). Figure 4 shows the detected presence of orchid species in products from these countries and in samples per category form.

[Insert] **Figure 4.** Detection of orchids in salep, **A.** per country and **B.** per product form.

Discussion

DNA metabarcoding is useful for identifying plant species diversity in a range of products [18,46,48,49,61,62]. Cheng et al. [49] used DNA metabarcoding to analyze nine commercial processed Traditional Chinese Medicines (TCMs) and detected an average of 4.8 species using nrITS2 and 2.8 using *trn*L. Coghlan et al. [18,62] analyzed 15 commercial processed TCMs for presence of both animal and plant ingredients and found over 68 plant families and 8 vertebrate genera in these products. Ivanova et al. [48] used universal nrITS primers to authenticate 15 herbal supplements and found a host of plant and fungi. Raclariu et al. [50,51] used plant-specific nrITS primers to analyse 78 herbal supplements of *Hypericum* and *Veronica* herbal supplements and found large discrepancies between detected species and those listed on the label. Richardson et al. [46], and Hawkins et al. [61] used nrITS2 and *rbc*L, respectively, to analyze DNA from pollen in pollen grains and honey to investigate honey bee foraging preferences. These previous studies have shown that the quality of the extraction substrate influences amplification and sequencing success, and whereas pollen grains and some herbal medicines can have high yields of DNA, more difficult samples such as processed herbal supplements and the food products studied here are harder to work with.

Salep and other processed food and pharmaceutical products have no means of comparing identification methods, in contrast to substrates that can also be used for morphological identification, e.g., pollen clumps [46,63] and pollen in honey [61]. Whereas pharmaceutical products and traditional and complementary alternative medicines will have contents printed on the package [18,48,49,62], salep is often sold as powder in bags or containers on local markets [11,16]. The nature of this study in which salep products with unknown ingredients are studied, makes false negatives harder to detect, and it is difficult to quantify the species diversity that is overlooked by metabarcoding through poor primer fit and amplification bias but some diversity is likely missed [54,64,65]. In this study, we detected a total of 12 orchid taxa in 13 samples, and on average found 14.7 taxa in the 30 samples that passed our quality criteria (Table S3). Most of the identified species are likely ingredients of salep, but some species appear implausible given their distribution or unlikely use. The identification of these plant species 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 species. Tentative candidates for the latter are *Aegilops caudata* L., *Aegilops lorentii* Hochst., *Aegilops speltoides* Tausch, *Anthosachne multiflora* (Banks & Sol. ex Hook.f.) C.Yen & J.L.Yang, *Avena byzantina* K.Koch, *Avena fatua* L., *Boissiera squarrosa* (Sol.) Nevski, *Eleusine coracana* (L.) Gaertn., *Eleusine indica* (L.) Gaertn., *Holcus lanatus* L., *Hordeum vulgare* L., *Lolium temulentum* L., *Poa pratensis* L., *Poa tibetica* Munro ex Stapf, *Secale montanum* var. *anatolicum* Boiss., *Setaria pumila* (Poir.) Roem. & Schult., and *Urtica dioica* L. An additional seven anemophilous species were detected in the separately analyzed low DNA yield samples, *Brachypodium distachyon* (L.) P.Beauv., *Dactyloctenium aegyptium* (L.) Willd., *Echinochloa colona* (L.) Link, *Echinochloa crus-galli* (L.) P.Beauv., *Festuca plebeia* Vickery, *Lolium perenne* L., and *Setaria verticillata* (L.) P.Beauv. confirming previously raised concerns about sensitivity and low template quality.

The presence of eleven identified species is neither likely due their distribution nor wind-dispersed pollen, although some are important in traditional medicine or of horticultural value (Table S5).

The in-silico PCR showed that nrITS1 amplified 914 taxa in 93 families and nrITS2 4001 taxa in 228 families. nrITS2 amplified the main salep orchid species in *Anacamptis*, *Dactylorhiza*, *Himantoglossum*, *Ophrys* and *Orchis*, whereas nrITS1 amplified many of the potential cereal adulterants but no salep orchids. These primer fit issues are also reflected in the number of species found using each marker, with nrITS1 yielding 89 species and nrITS 103, and a total of 58 being identified with only one marker. Other studies have also reported that certain genera and families were not detected with specific markers. Richardson et al. [46] who use the nrITS2 marker did not find amplicons belonging to the genus *Lonicera* and families Lamiaceae and Salicaceae in honey, despite the fact that pollen from these taxa were identified using microscopy. Absence of sequence reads for these species is likely due to poor primer fit caused by sequence divergence in the PCR priming sites [46]. Another problem is limited sequence variation in barcoding markers, and makes certain markers less suitable than others, both in general and for identification of specific families and genera. Hawkins et al. [61] who use *rbc*L and *trn*H-*psb*A as markers to metabarcode pollen find that within Boraginaceae and Euphorbiaceae, species identification remains difficult. Nuclear ribosomal ITS1 and ITS2 are generally variable markers in plants [38], but in our analyses limited variation in *Crocus* (Iridaceae), *Heracleum* (Apiaceae), and *Viola* (Violaceae) impede identification at species level.

Although a limited number of samples was studied, the detection of orchids in 43 % of these is alarming considering that all orchid species are CITES appendix II listed [17] and harvesting is illegal at a national level in the main source countries Greece, Iran and Turkey. Previous studies have shown that illegal collection, local trade and international trade are rampant in these countries, and several authors have raised alarms over the scale and threat of this trade to wild populations of orchids [7,11–13,16]. Ghorbani et al. [11] report that salep tubers are available in many markets in Western Iran, and that wild-collection is on the rise in Iran in recent years to meet Turkish demand for authentic salep. An interesting finding in this study is that it appears that Iranian salep powder is the most highly adulterated salep on the market, with 79 % of samples failing to yield DNA (suggesting the use of synthetic polysaccharides) and only 11 % of samples containing orchids (*Dactylorhiza incarnata*). In Greece on the other hand, where Kreziou et al. [12] report a revival of local foods and traditional medicine, nearly 80 % of salep contained orchids, and from a diversity of genera, *Anacamptis*, *Dactylorhiza*, *Gymnadenia* and *Orchis*. Kasparek & Grimm [7] report on the massive trade of salep from Turkey to the Turkish diaspora in Germany in the 1990s, and although we cannot assess the scale of this trade today, we do see that only 21 % of salep products genuinely contain orchids. Earlier studies have only been able to assess trade in dried tubers, but not to authenticate the bulk of the international trade that is based on salep powder and products. Molecular identification is enabling this and the recent publication of a barcode reference library for

identification of the main salep orchid species is further empowering the use of this approach [72].

Conclusions

Application of high-throughput nrITS1 and nrITS2 DNA metabarcoding to determine the constituents of a product intended to contain illegally harvested and traded terrestrial orchids occurring in the wild in countries around the Mediterranean Sea shows that the method can elucidate species diversity in the products. DNA metabarcoding here provides an insight into a processed product that could previously only be analyzed by analytical chemistry approaches that were unable to verify presence or absence of plant species. The ability of DNA metabarcoding to detect orchid species enables regulatory agencies (e.g., customs, CITES authorities and environmental agencies) to monitor illegal trade and enforce national and international legislation. Implementation of the method has a number caveats due to a lack of universality of methods, markers, analysis, and species delimitation require tailored approaches for different study objectives [25,68]. Quantifying constituents per species on the basis of read numbers can only be approximated in specific cases [33,64,69,71], but species presence and absence scoring can be done with high confidence if the extraction substrates yield enough DNA [68]. The development and further refinement of plant DNA metabarcoding markers, sequencing techniques and analysis pipelines is likely to overcome some of the current challenges involved in this approach. Our data underscore the persistent role of terrestrial orchids in salep, 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. It seems that Greek and Iran salep are most likely to contain real orchid tubers, whereas Turkish salep is more likely to be adulterated. Previous observations have suggested that the market for salep in Turkey has largely depleted local resources and has caused an orchid harvesting boom in neighboring Iran and Greece [7,11,12,16]. We expect that higher quality salep in Turkey contain imported salep tubers, but that common salep found in bazaars and shops is largely adulterated with non-orchid thickeners. This study demonstrates that in addition to the previously documented applicability of DNA metabarcoding to conservation through wildlife forensics [28,30,33,70,71], it can also be in conservation to identify and monitor species affected by illegal plant trade in processed substrates.

Data accessibility. All raw data as well as all MOTUs and their taxon assignments are deposited in DRYAD doi:10.5061/dryad.5q447. Detailed methods and data is available in the Electronic Supplementary Material (ESM).

Author contributions. HdB and BG devised and supervised the project. AG and AK purchased samples. SO, AR and MO performed laboratory work. SO and VM performed data analysis. HdB wrote the manuscript. All authors gave final approval for publication.

Competing interests. The authors have no competing interests.

Funding. The Swedish Science Council supported this research through grant VR-SRL D0664801 to HdB.

Acknowledgments. We thank Audun Schrøder-Nielsen and Jarl Andreas Anmarkrud of the University of Oslo Natural History Museum DNA laboratory for logistical support and Youri Lammers, University of Tromsø for bioinformatics support.

References

1. Bulpitt CJ. 2005 The uses and misuses of orchids in medicine. *QJM* **98**, 625–631. (doi:10.1093/qjmed/hci094)

2. Bulpitt CJ, Li Y, Bulpitt PF, Wang J. 2007 The use of orchids in Chinese medicine. *J. R. Soc. Med.* **100**, 558–563. (doi:10.1258/jrsm.100.12.558)

3. Hossain MM. 2011 Therapeutic orchids: traditional uses and recent advances—An overview. *Fitoterapia* **82**, 102–140.

4. Chinsamy M, Finnie JF, Van Staden J. 2011 The ethnobotany of South African medicinal orchids. *South Afr. J. Bot.* **77**, 2–9. (doi:10.1016/j.sajb.2010.09.015)

5. Arditti J. 1992 *Fundamentals of orchid biology*. New York: Wiley.

6. Veldman S, Otieno JN, Andel T van, Gravendeel B, de Boer HJ. 2014 Efforts urged to tackle thriving illegal orchid trade in Tanzania and Zambia for chikanda production. *TRAFFIC Bull.* **26**, 47–50.

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

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

9. Starin D. 2012 Salepi Extinction, Salepi Survival: How a Change in Ingredients Could Help Safeguard Orchids. *Orchids Bull. Am. Orchid Soc.* **81**, 490–494.

10. Sandal Erzurumlu G, Doran I. 2011 Türkiye'de salep orkideleri ve salep kültürü. *JAgric Fac HRU* **15**, 29–34.

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

12. Kreziou A, de Boer H, Gravendeel B. 2015 Harvesting of salep orchids in north-western Greece continues to threaten natural populations. *Oryx* , 1–4.

13. Sezik E. 2006 Destroying of Ophrys species to obtain Salep in turkey. *J. Eur. Orchid.* **38**, 290.

14. Sezik E. 2002 Destruction and Conservation of Turkish Orchids. In *Biodiversity* (ed B Şener), pp. 391–400. Springer US.

15. Tekinşen KK, Güner A. 2010 Chemical composition and physicochemical properties of tubera salep produced from some Orchidaceae species. *Food Chem.* **121**, 468–471. (doi:10.1016/j.foodchem.2009.12.066)

16. Ghorbani A, Gravendeel B, Zarre S, de Boer HJ. 2014 Illegal wild collection and international trade of CITES-listed terrestrial orchid tubers in Iran. *TRAFFIC Bull.* **26**, 52–58.

17. CITES. 2014 The Convention on International Trade in Endangered Species of Wild Fauna and Flora Appendices.

18. Coghlan M, Haile J, Houston J, Murray D, White N, Moolhuijzen P, Bellgard M, 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. (doi:10.1371/journal.pgen.1002657)

19. Kool A, de Boer HJ, Krüger Å, Rydberg A, Abbad A, Björk L, Martin G. 2012 Molecular identification of commercialized medicinal plants in Southern Morocco. *PLOS ONE* **7**, e39459. (doi:10.1371/journal.pone.0039459)

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

21. De Boer HJ, Ouarghidi A, Martin G, Abbad A, Kool A. 2014 DNA barcoding reveals limited accuracy of identifications based on folk taxonomy. *PLOS ONE* **9**, e84291. (doi:10.1371/journal.pone.0084291)

22. Hebert PDN, Cywinska A, Ball S, de Waard J. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–322.

23. Hajibabaei M, Singer GA, Hebert PD, Hickey DA. 2007 DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **23**, 167–172.

24. Haarsma A-J, Siepel H, Gravendeel B. 2016 Added value of metabarcoding combined with microscopy for evolutionary studies of mammals. *Zool. Scr.* **45**, 37–49.

25. Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, Prins TW, Kok E. 2016 Advances in DNA metabarcoding for food and wildlife forensic species identification. *Anal. Bioanal. Chem.* , 1–16.

26. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012 Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050. (doi:10.1111/j.1365-294X.2012.05470.x)

27. Jørgensen T *et al.* 2012 Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Mol. Ecol.* **21**, 1980–1988. (doi:10.1111/j.1365-294X.2011.05278.x)

28. Hibert F, Taberlet P, Chave J, Scotti-Saintagne C, Sabatier D, Richard-Hansen C. 2013 Unveiling the diet of elusive rainforest herbivores in next generation sequencing era? The tapir as a case study. *PLOS ONE* **8**, e60799.

29. Soininen EM *et al.* 2013 Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content. *Polar Biol.* , 1–8.

30. Valentini A *et al.* 2009 New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* **9**, 51–60. (doi:10.1111/j.1755-0998.2008.02352.x)

31. Boessenkool S, Mcglynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, Nemomissa S, Brochmann C, Popp M. 2014 Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. *Conserv. Biol.* **28**, 446–455. (doi:10.1111/cobi.12195)

32. Parducci L *et al.* 2012 Glacial survival of boreal trees in Northern Scandinavia. *Science* **335**, 1083–1086. (doi:10.1126/science.1216043)

33. Willerslev E *et al.* 2014 Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51. (doi:10.1038/nature12921)

34. Taberlet P *et al.* 2007 Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14–e14.

35. Wendel JF, Schnabel A, Seelanan T. 1995 Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). *Proc. Natl. Acad. Sci.* **92**, 280–284.

36. Alvarez I, Wendel JF. 2003 Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**, 417–434.

37. Koch MA, Dobeš C, Mitchell-Olds T. 2003 Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American Arabis divaricarpa (Brassicaceae). *Mol. Biol. Evol.* **20**, 338–350.

38. Li DZ *et al.* 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.

39. Chen S *et al.* 2010 Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **5**, 1–8. (doi:10.1371/journal.pone.0008613)

40. Flynn JM, Brown EA, Chain FJ, MacIsaac HJ, Cristescu ME. 2015 Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecol. Evol.* **5**, 2252–2266.

41. Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T. 2011 Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ. Microbiol.* **13**, 340–349.

42. Epp LS *et al.* 2012 New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* **21**, 1821–1833. (doi:10.1111/j.1365-294X.2012.05537.x)

43. Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. 2013 ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* **13**, 218–224. (doi:10.1111/1755-0998.12065)

44. Schmidt P-A, Bálint M, Greshake B, Bandow C, Römbke J, Schmitt I. 2013 Illumina metabarcoding of a soil fungal community. *Soil Biol. Biochem.* **65**, 128–132. (doi:10.1016/j.soilbio.2013.05.014)

45. De Beeck MO, Lievens B, Busschaert P, Declerck S, Vangronsveld J, Colpaert JV. 2014 Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *PLOS ONE* **9**, e97629. (doi:10.1371/journal.pone.0097629)

46. Richardson RT, Lin C-H, Sponsler DB, Quijia JO, Goodell K, Johnson RM. 2015 Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Appl. Plant Sci.* **3**, apps.1400066. (doi:10.3732/apps.1400066)

47. Veldman S *et al.* 2017 High-throughput sequencing of African chikanda cake highlights conservation challenges in orchids. *Biodivers. Conserv.* , 1–18.

48. Ivanova NV, Kuzmina ML, Braukmann TW, Borisenko AV, Zakharov EV. 2016 Authentication of Herbal Supplements Using Next-Generation Sequencing. *PloS One* **11**, e0156426.

49. Cheng X, Su X, Chen X, Zhao H, Bo C, Xu J, Bai H, Ning K. 2014 Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for Liuwei Dihuang Wan. *Sci. Rep.* **4**.

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

51. Raclariu AC, Mocan A, Popa MO, Vlase L, Ichim MC, Crisan G, Brysting AK, De Boer HJ. 2017 Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with Veronica chamaedrys. *Front. Pharmacol.* **8**, 378.

52. Doyle JJ, Doyle JL. 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15.

53. Sun Y, Skinner DZ, Liang GH, Hulbert SH. 1994 Phylogenetic analysis of Sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theor. Appl. Genet.* **89**, 26–32.

54. Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P, Pompanon F. 2010 An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434.

55. Lammers Y, Peelen T, Vos RA, 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. (doi:10.1186/1471-2105-15-44)

56. Schmieder R, Edwards R. 2011 Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864.

57. Ewing B, Green P. 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

58. 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.

59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

60. Waskom M *et al.* 2015 *seaborn: v0.7.1*. See https://doi.org/10.5281/zenodo.12710.

61. Hawkins J, de Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B. 2015 Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PloS One* **10**, e0134735.

62. Coghlan ML *et al.* 2015 Combined DNA, toxicological and heavy metal analyses provides an auditing toolkit to improve pharmacovigilance of traditional Chinese medicine (TCM). *Sci. Rep.* **5**.

63. Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, Steffan-Dewenter I, Keller A. 2015 Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecol.* **15**, 20.

64. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. 2011 ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* , gkr732.

65. Pawluczyk M, Weiss J, Links MG, Aranguren ME, Wilkinson MD, Egea-Cortines M. 2015 Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Anal. Bioanal. Chem.* **407**, 1841–1848.

66. Elbrecht V, Leese F. 2015 Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS One* **10**, e0130324.

67. Coissac E, Riaz T, Puillandre N. 2012 Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* **21**, 1834–1847.

68. Ficetola GF *et al.* 2015 Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* **15**, 543–556.

69. Craine JM, Towne EG, Miller M, Fierer N. 2015 Climatic warming and the future of bison as grazers. *Sci. Rep.* **5**, 16738.

70. Soininen EM *et al.* 2009 Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front. Zool.* **6**, 16.

71. Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, Rubenstein DI, Wang W, Pringle RM. 2015 DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc. Natl. Acad. Sci.* **112**, 8019–8024.

72. Ghorbani A, Gravendeel B, Selliah S, Zarre S, de Boer HJ. 2016 DNA barcoding of tuberous Orchidoideae: A resource for identification of orchids used in Salep. *Mol Ecol Resour* (doi:10.1111/1755-0998.12615)

73. Tedersoo L *et al.* 2010 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol.* **188**, 291–301.

Electronic supplementary material (ESM)

Text S1. Detailed methods.

Table S1. Salep samples were randomly purchased to represent commercially available salep products with the objective to assess the level of adulterated and true orchid based products. The sampling includes a representation of producers and vendors, e.g., supermarkets, herbal stores, pharmacies, and markets.

Table S2. Data for products that yielded less than 0.5 ng/µl of DNA. Reads per species for nrITS1, nrITS2 and merged per sample.

Table S3. Data for products that yielded more than 0.5 ng/µl of DNA. Reads per species for nrITS1, nrITS2 and merged per sample.

Table S4. Plant species detected in more than 20 % of samples, excluding orchids (presence as % in total samples in parentheses).

Table S5. Unexpected species detected in samples (presence as % in total samples in parentheses).

Figure S1. Detection of species in salep samples for products that yielded less than 0.5 ng/µl of DNA. Species are colored by relative abundance of normalized read numbers. Species are categorized in gelatinous species (light blue) and non-gelatinous species (blue), and clustered by euclidean distances. Salep samples are numbered and grouped by country of purchase.