

DNA barcoding of tuberous Orchidoideae: A resource for identification of orchids used in Salep

Abdolbaset Ghorbani^{1,2}, Barbara Gravendeel^{3,4}, Sugirthini Selliah⁵, Shahin Zarré⁶, Hugo de Boer^{1,3,5,*}

¹ Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236, Sweden

² Traditional Medicine & Materia Medica Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³ Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

⁴ University of Applied Sciences Leiden, Leiden, The Netherlands

⁵ The Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318 Oslo, Norway

⁶ Department of Plant Sciences, University of Tehran, Iran

* Corresponding author: Hugo de Boer, Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318 Oslo, Norway. Email: hugo.deboer@nhm.uio.no

Word count: 4286; Tables 3, Figures 3.

Keywords: CITES; Molecular identification; Overharvesting; Orchid conservation; Plant DNA barcoding; Wildlife Trade.

21 **Abstract**

22 Tubers of terrestrial orchids are harvested and traded from the eastern Mediterranean to the Caspian Sea for the
23 traditional product Salep. Over-exploitation of wild populations and increased middle-class prosperity have escalated
24 prices for Salep, causing overharvesting, depletion of native populations and providing an incentive to expand
25 harvesting to untapped areas in Iran. Limited morphological distinctiveness among traded Salep tubers renders species
26 identification impossible, making it difficult to establish which species are targeted and affected the most. In this study,
27 a reference database of 490 nrITS, *trnL*-F spacer and *matK* sequences of 133 taxa was used to identify 150 individual
28 tubers from 31 batches purchased in 12 cities in Iran to assess species diversity in commerce. The sequence reference
29 database consisted of 211 nrITS, 158 *trnL*-F, and 121 *matK* sequences, including 238 new sequences from collections
30 made for this study. The markers enabled unambiguous species identification with tree-based methods for nrITS in 67%
31 of the tested tubers, 58% for *trnL*-F and 59% for *matK*. Species in the genera *Orchis* (34%), *Anacamptis* (27%) and
32 *Dactylorhiza* (19%) were the most common in Salep. Our study shows that all tuberous orchid species in this area are
33 threatened by this trade, and further stresses the urgency of controlling illegal harvesting and cross-border trade of Salep
34 tubers.

35

36 **Introduction**

37

38 Tuberous terrestrial orchids have long been used as medicine and dietary supplements in different parts of the world
39 (Bulpitt 2005; Bulpitt *et al.* 2007; Hossain 2011; Chinsamy *et al.* 2011). In the eastern Mediterranean, Asia Minor and
40 the Middle East, tubers of different orchid species are collected indiscriminately from the wild and are traded for
41 production of Salep tuber powder (Kasperek & Grimm 1999; Ece Tamer *et al.* 2006; Sandal Erzurumlu & Doran 2011;
42 Ghorbani *et al.* 2014a; Kreziou *et al.* 2015). Harvested tubers are washed in water, boiled in either water or milk, sun-
43 dried and traded as dried tubers (Kasperek & Grimm 1999). The tubers are ground into a powder and used in preparing
44 a hot beverage known as Salep or Salepi and also in ice cream production (Sezik 2002a; Ece Tamer *et al.* 2006; Starin
45 2012). Salep drink was once common in Europe (Landerer 1850), but is now consumed mainly in Turkey and Greece
46 (Bulpitt 2005; Ece Tamer *et al.* 2006; Starin 2012). It is estimated that as much as 30 tons of orchid tubers are harvested
47 annually in Turkey, which requires the destruction of 30-120 million orchid plants (Kasperek & Grimm 1999; Sezik
48 2006). Increasing popularity of Salep has increased the demand for Salep tubers, which in turn has led to further
49 overharvesting of wild orchid populations (Sezik 2002b; Kreziou *et al.* 2015). Scarcity of wild orchids in Turkey has
50 forced traders to tap into new sources in adjacent countries (Ghorbani *et al.* 2014b). In Iran, where orchid tubers are
51 traditionally hardly consumed, an orchid boom is underway in which an estimated 5.5-6.1 million orchids are harvested
52 annually for export to Turkey (Ghorbani *et al.* 2014a). Conservation concerns have made orchid tuber collection illegal
53 in Greece, Turkey and Iran, but collection bans are poorly enforced (Ghorbani *et al.* 2014b; Kreziou *et al.* 2015). All
54 orchid species are included by the Convention on International Trade of Endangered Species of Fauna and Flora
55 (CITES) on Appendices I or II (CITES 2014), which means that international trade of these species and derived
56 products is regulated. Most of the Salep tuber trade from Iran to Turkey takes place without CITES permits, and tubers
57 are often mislabeled as low-value nuts or other products to circumvent taxes and permit requirements (Kasperek &
58 Grimm 1999; Ghorbani *et al.* 2014b; Kreziou *et al.* 2015). This large-scale, yet poorly visible trade makes it difficult to
59 ascertain which species are targeted and in what quantities. Morphology-based approaches for identification are
60 insufficient and cannot even accurately distinguish dried tubers from different genera. Other methods for salep
61 identification, such as GCMS, HPLC, gravimetric, absorbance and rheological analyses, all indicate that identification
62 to species level is not possible using only chemical analyses (Dogan *et al.* 2007; Tekinşen & Güner 2010; Babbar &
63 Singh 2016). Adequate monitoring would enable identification of priority species for conservation measures such as
64 curbing overexploitation, and targeting high-value species for cultivation.

65 DNA barcoding provides an accurate and reliable alternative to morphology-based identification of biological
66 material (Hebert *et al.* 2003). As a method it can be used to identify and discern species at any developmental or
67 processing stage from which DNA can be extracted (Hebert *et al.* 2003; Hajibabaei *et al.* 2007), and even from the
68 minute amounts such as those found in dung (Hibert *et al.* 2013), pollen (Richardson *et al.* 2015), degraded herbarium
69 vouchers (Särkinen *et al.* 2012), permafrost preserved subfossils (van Geel *et al.* 2008), and ancient sediment cores
70 (Williams *et al.* 2000; Posadzki *et al.* 2012). Plant DNA barcoding has been applied in many fields, for example
71 molecular systematics (Liu *et al.* 2011; van Velzen *et al.* 2012), biodiversity inventories (Aubriot *et al.* 2013; Thompson
72 & Newmaster 2014), wildlife forensics (Deguilloux *et al.* 2002; Ogden *et al.* 2009), bio-piracy control (Parveen *et al.*
73 2012), and authentication of herbal products (Kool *et al.* 2012; Coghlan *et al.* 2012; Newmaster *et al.* 2013; de Boer *et al.*
74 2014; Vassou *et al.* 2015).

75 Several genetic regions have been proposed as standard barcodes for land plants, the ideal barcode being both
76 easily amplifiable and efficiently retrievable from any of the 300,000+ species of plants (Kress *et al.* 2005; Fazekas *et al.*
77 2008). Most studies now employ a tiered multilocus approach, which is based on the use of a common, easily
78 amplified and aligned region such as *rbcL*, *rpoC1*, *trnL* or *trnL-F* spacer that can act as a scaffold on which to place
79 data from a more variable noncoding region such as *matK*, *trnH-psbA*, nrITS, or nrITS2. Most species (approximately
80 75-85%) can be identified using such an approach, and the subsequent addition of surrogate regions can increase
81 barcoding success to over 90% in some floras (Ebihara *et al.* 2010; Burgess *et al.* 2011; de Vere *et al.* 2012; Kuzmina *et al.*
82 2012; Liu *et al.* 2015). In Orchidaceae, several plastid and nuclear molecular markers including *rbcL*, *psaB*, *psbC*-
83 *trnS*, *rpl16*, *matK*, *ycf1*, *trnH-psbA*, *trnH-trnK*, *trnL-F* and nrITS have been applied for phylogenetic analysis (Cameron
84 2004; Xiang *et al.* 2011; Parveen *et al.* 2012; Inda *et al.* 2012; Kim *et al.* 2014). These studies suggest that a multi-locus
85 combination of coding and non-coding regions with different evolutionary rates is necessary for effective identification
86 of species in Orchidaceae.

87 This study tests the hypothesis that molecular identification using DNA barcoding can be used for
88 identification of orchid species comprising boiled and dried tuber samples traded in the main export market hubs in
89 Iran. We address the following research questions: 1) Can DNA be extracted, amplified and sequenced from boiled and
90 dried Salep tubers? 2) What marker or markers are optimal for the identification of Salep tubers traded in the markets of
91 Iran? 3) What genera and species are most common among the tubers included in our sampling? 4) Can the most
92 common traded species be used to predict the main source areas of orchid tubers exported to Turkey? The aim was to
93 test and establish a DNA barcoding protocol to identify dried orchid tubers from markets and to show the potential of
94 this technique to curb illegal trade of CITES listed orchid tubers.

95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124

Methods

Collection of reference and market material

Flora Iranica vol. 126 (Renz 1978), Flora of Iran vol. 57 (Shahsavari 2008) and Orchids of Europe, North Africa and the Middle East (Delforge 2006) were used to estimate that a total of 47 orchid species occur in Iran, including 32 species with tuberous roots that could potentially be targeted for Salep collection. During fieldwork in 2013-2014, a total of 127 herbarium vouchers representing 30 species and subspecies of orchids were collected from natural populations in different parts of Iran (Suppl. 1). Vouchers were identified (Renz 1978; Delforge 2006; Shahsavari 2008) and deposited at the herbarium of Tehran University (TUH). Sequences generated from these vouchers (Suppl. 1) as well as selected vouchered sequences from NCBI GenBank were used to construct a DNA barcode reference library (Suppl. 2).

Markets in 12 cities and towns in Iran (Tehran, Kermanshah, Sanandaj, Tabriz, Urmia, Mahabad, Shahindezh, Kashan, Ardabil, Aq-Emam, Marave-Tappe and Kalaleh) were visited and 31 batch samples of unidentified Salep tubers containing 15-50 tubers each were purchased. Figure 1 shows the distribution of orchids in Iran at genus level based on indexed vouchers from TUH and W, plus the location of the 12 main Salep markets. Per sample, tubers were subsequently categorized based on shape and size, and a total of 150 random tubers were selected as query tubers for DNA barcoding. Salep tubers in trade are hard to identify, although palmate *Dactylorhiza* tubers differ from those of other tuberous genera (Figure 2).

DNA extraction, amplification and sequencing

For reference samples, total genomic DNA was extracted from silica-gel dried leaf material using a modified CTAB protocol (Doyle & Doyle 1987). The query tubers were ground into powder using liquid nitrogen, and subsequently DNA was extracted using a STE-CTAB protocol (Shepherd & McLay 2011). The STE-CTAB protocol was necessary to reduce gel formation due to the high glucomannan content of tubers. A gelatinous layer, which was formed after adding CTAB buffer, caused difficulties in extraction procedures and low DNA yields. Extracted DNA was purified using a GE Illustra GFXTM PCR DNA and Gel Band Purification kit following the manufacturer's protocol (GE Healthcare, Buckinghamshire, UK).

Three barcode regions, nrITS (ITS1-5.8S-ITS2), *trnL*-F spacer and *matK* were amplified by a standard polymerase chain reaction (PCR). The nrITS (ITS1-5.8S-ITS2) region was amplified using the following primers:

17SE_F (5'-ATGGTCCGGTGAAGTGTTTC-3'), 26SE_R (5'-CCCGGTTTCGCTCGCCGTTAC-3'), 5.8I-1_R (5'-GTTGCCGAGAGTCGT-3') and 5.8I-2_F (5'-GCCTGGGCGTCACGC-3') (Sun *et al.* 1994). The *trnL*-F spacer was amplified using the following primers: C_F (5'-CGAAATCGGTAGACGCTACG-3'), C2_F (5'-GGATAGGTGCAGAGACTCAAT-3') and F_R (5'-ATTTGAACTGGTGACACGAG-3') (Taberlet *et al.* 1991; Bellstedt *et al.* 2001). *MatK* was amplified using the following four primers: 19_F (5'-CGTTCTGACCATATTGCACTATG-3') and 881R (5'-TMTTCATCAGAATAAGAGT-3') (Gravendeel *et al.* 2001); F2_F (5'-CTAATACCCCATCCCATCCAT-3') (Steele & Vilgalys 1994) and R1_R (5'-CATTTTTTCATTGCACACGRC-3') (Kocyan *et al.* 2004). PCR amplification was performed in a 50 µl reaction volume containing 5 µl reaction buffer IV (10x), 5 µl MgCl₂ (25mM), 1 µl dNTP (10 µM), 0.25 µl Taq-polymerase (5 U/µl), 0.5 µl BSA, 1 µl of each primer (10 mM) and 1 to 4 µl of template DNA. The PCR protocols of 95°C 3 min., (95°C 20 s., 55°C 1 min., 72°C 2 min.) x 35, 72°C 10 min., 8°C ∞ for nrITS, 95°C 3 min., (95°C 15 s., 55°C 50 s., 72°C 4 min.) x 35, 72°C 8 min., 8°C ∞ for *trnL*-F spacer and 95°C 3 min., (95°C 34 s., 59°C 45 s., 72°C 1 min.) x 35, 72°C 7 min., 8°C ∞ for *matK* were applied. Sanger sequencing was performed by MacroGen Europe Inc. (Amsterdam, the Netherlands) on an ABI3730XL automated sequencer (Applied Biosystems). Primers used for PCR amplification were also used for sequencing reactions.

Reference database preparation

The reference database was compiled from a total of 490 source sequences of 133 taxa, including both voucher specimens collected from the field including 85 nrITS sequences (19 species), 90 *trnL*-F (26 species), 63 *matK* (20 species) and publicly available DNA sequences from NCBI GenBank including 126 nrITS sequences (102 species), 68 *trnL*-F (56 species) and 58 *matK* (55 species) (Table 1). All sequences were downloaded from the listed tuberous genera in the tribe Orchideae (Orchidaceae), including synonymous genera and/or species: *Anacamptis* Rich., *Cephalanthera* Rich., *Chamorchis* Rich., *Dactylorhiza* Neck. ex Nevski (including *Coeloglossum* Hartm.), *Gennaria* Parl., *Gymnadenia* R.Br., *Himantoglossum* W.D.J.Koch (incl. *Barlia* Parl. and *Comperia* K.Koch), *Neottia* Rchb.f., *Neottia* Guett. (incl. *Listera* R.Br.), *Neottianthe* Schltr., *Ophrys* L., *Orchis* L. (incl. *Aceras* R.Br.), *Serapias* L., *Limodorum* Boehm., *Platanthera* Rich., and *Stenandria* Schltr. Representative accessions were included for non-tuberous genera and tuberous species occurring close to the study area: *Corallorhiza trifida* Châtel, *Epipactis helleborine* (L.) Crantz, *Goodyera repens* (L.) R.Br., *Habenaria macroceratitis* Willd., *Herminium monorchis* (L.) R.Br., *Pecteilis gigantea* (Sm.) Raf., *Peristylus densus* (Lindl.) Santapau & Kapadia, *Pseudorchis albida* (L.) Á.Löve & D.Löve, *Satyrion bicorne* (L.) Thunb., *Spiranthes aestivalis* (Poir.) Rich., *Spiranthes spiralis* (L.) Chevall. and *Zeuxine*

155 *strateumatica* (L.) Schltr (Suppl. 2). Where there were more than two accessions per marker per species, only two
156 accessions were selected, giving priority to those accessions with associated vouchers plus optimal read length and
157 quality. Representative accessions of *Brownleea parviflora* Harv. ex Lindl., *Disa uniflora* P.J.Bergius and *Disperis*
158 *lindleyana* Rchb.f. were selected as outgroups based on Inda et al. (2012).

159

160 *Data analysis*

161 Contigs were assembled and edited in SeqTrace (Stucky 2012). All sequences including reference sequences and query
162 tuber sequences were aligned using MUSCLE (Edgar 2004) as implemented in Aliview v. 1.15 aligner (Larsson 2014).
163 Final manual inspections were performed and adjustments were done if necessary. Sequences generated for this study
164 were submitted to NCBI GenBank (Suppl. 1, Suppl. 3).

165 Bayesian inference (BI) and maximum likelihood (ML) analysis were performed for each marker separately
166 and on concatenated datasets, using RAxML-HPC v.8 (Stamatakis 2014) and MrBayes v.3.2.2 (Ronquist *et al.* 2012) on
167 CIPRES Science Gateway v.3.3 (Miller *et al.* 2010) and the high performance computing facility available at University
168 of Oslo, Lifeportal (<https://lifeportal.uio.no/root>). Gaps were treated as missing data.

169 For Bayesian analyses, the model GTR + G was selected for all datasets. Two independent runs with sixteen
170 MCMC chains were simultaneously performed for 20 million rearrangements initiated with a random starting tree, and
171 sampling one tree every 1000 generations, except for *matK*. For *matK*, we performed eight MCMC chains and a total of
172 10 million generations using the default heating temperature. Convergence of runs with default parameters was assessed
173 on preliminary analyses. Where convergence did not occur, the heating parameter was adjusted to reach a convergence.
174 Convergence of runs was assessed using Tracer v. 1.6 (Rambaut *et al.* 2014). Twenty-five percent of trees were
175 discarded as burn-in, and the remaining trees were used to generate a consensus tree with Bayesian posterior
176 probabilities (PP) values. Only PP values over 0.95 were considered and included for each marker and concatenated
177 topologies. The number of trees retained for each analysis is presented in Table 2.

178 For maximum likelihood analyses with RAxML, the model GTR + G was selected for all datasets, and a rapid
179 bootstrap analysis with 1000 trees was conducted. Single marker trees were compared for incongruence prior to
180 concatenation. Datasets were concatenated using Geneious v. 6.1.8 (Kearse *et al.* 2012). Multiple GenBank reference
181 sequences for a single species were merged in order to obtain one consensus species sequence (cf. Suppl. 2). The unlink
182 option was used to estimate the parameters for each partition.

183 The BI and ML phylogenetic trees were used to identify the query tubers (Suppl. 4-11). The tubers were
184 considered successfully identified to species level when they were monophyletically clustered with related individuals

185 of the same species. When tubers were clustered with individuals of different species of the same genus, only a genus
186 level identification was assigned (Suppl. 12).

187 Sequence similarity search using Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990) is often
188 used in DNA barcoding (Little & Stevenson 2007; Sass *et al.* 2007; Kool *et al.* 2012; de Boer *et al.* 2014). BLAST+
189 (Camacho *et al.* 2009) features implemented in NCBI BLAST were used to query unknown tuber sequences against the
190 compiled reference database. All top hits less than 15 points lower than the max score were considered for
191 identification: if the retained top hits (max score -15 points) included only a single species then a species level
192 identification was estimated; if the retained top hits (max score -15 points) included multiple species in the same genus
193 then a genus level identification was estimated; if the retained top hits (max score -15 points) included multiple species
194 in different genera then a family level identification was estimated (Suppl. 12).

195 Final consensus identifications were made based on the results from all markers and methods, BLAST, ML,
196 and BI (Suppl. 12). Species level identification was assigned if all markers with species level identifications yielded the
197 same species identification. Genus level identification was assigned if identifications resulted in multiple species of the
198 same genus.

199

200

201 **Results and discussion**

202

203 *Amplification and sequencing success*

204 Sequencing success rates were different for reference samples and market tuber samples. For the reference leaf samples
205 (L), sequencing success was 67% (85 samples) for nrITS, 71% (90) for the *trnL*-F spacer and 47% (63) for *matK* (Table
206 2). Out of the 127 samples, all three markers could be sequenced for 34 samples, solely nrITS for 29, solely *trnL*-F
207 spacer for 14 and solely *matK* for 7. For tuber samples (T), sequencing success was 69% for nrITS (104 samples), 63%
208 for the *trnL*-F spacer (94) and 19% for *matK* (28) (Table 2). Out of 150 tuber samples, all three markers could be
209 sequenced for 8 samples, solely nrITS for 53, solely *trnL*-F spacer for 29, and none for *matK* only. In general, low
210 sequencing success might be due to degraded DNA as a result of boiling and drying the tubers during processing.
211 Sequencing success for nrITS might be affected by fungal contamination during the drying process and orchid
212 mycorrhizal associations producing a mix of plant and fungal nrITS sequences. *MatK* had the lowest amplification
213 success, and it has been shown that this locus cannot be amplified with 'universal' orchid primers due to the presence of

214 alternative translation initiation codons in orchids (Barthet *et al.* 2015), and therefore requires 'case by case'
215 optimization for each genus.

216

217

218 *Species identifications*

219 The similarity-based approach using BLAST using nrITS marker data identified 59 out of 104 tuber samples (57%) to
220 genus level and 45 (43%) to species level. Using *trnL*-F spacer, 61 out of 94 tuber samples (65%) were identified to
221 genus level and 33 (35%) to species level. Using *matK*, 11 out of 28 tuber samples (39%) were identified to genus and
222 17 (61%) to species level. The consensus of the BLAST identification of the three markers resulted in genus level
223 identification in 93 samples (62%) and species level in 57 samples (38%) (Table 3; Suppl. 12).

224 The tree-based approach using RAxML maximum likelihood using nrITS marker data identified 34 out of 104
225 tubers (33%) to genus level and 70 (67%) to species level (Suppl. 4, Suppl. 12). Using *trnL*-F spacer, 39 out of 94
226 samples (42%) were identified to genus level and 55 (58%) to species level (Suppl. 5, 12). Using *matK*, 12 out of 28
227 tuber samples (43%) were identified to genus and 16 (57%) to species level (Suppl. 6, 12). Concatenated data identified
228 87 samples (58%) to genus level and 63 (42%) to species level (Suppl. 7, 12). The ML consensus identification of the
229 three markers identified 60 samples (40%) to genus level and 90 samples (60%) to species level (Table 3; Suppl. 12).

230 The tree-based approach using MrBayes Bayesian inference using nrITS marker data identified 33 out of 104
231 tubers (32%) to genus level and 71 (68%) to species level (Suppl. 8, 12). Using *trnL*-F spacer, 39 out of 94 samples
232 (42%) were identified to genus level and 55 (58%) to species level (Suppl. 9, 12). Using *matK*, 9 out of 28 tuber
233 samples (32%) were identified to genus and 19 (68%) to species level (Suppl. 10, 12). Concatenated data identified 48
234 samples (32%) to genus level and 102 (68%) to species level (Suppl. 11, 12). The BI consensus identification of the
235 three markers identified 53 samples (35%) to genus level 97 samples (65%) to species level (Table 3; Suppl. 12).

236 The final identification that combines consensus identification results of ML, BI and BLAST approaches
237 produced an identification of 49 tubers (32.7%) to the genus level and 101 (67.3%) to the species level (Suppl. 12).

238

239 *Species composition of Salep*

240 Similarity-based identifications using BLAST showed that *Orchis* (51 samples), *Anacamptis* (40 samples), *Dactylorhiza*
241 (29 samples), *Ophrys* (18 samples) and *Himantoglossum* (11 samples) and *Steveniella* (1 sample) were constituents of
242 the studied Salep samples from Iran (Suppl. 12). *Orchis simia* Lam. and *O. mascula* (L.) L. were the main *Orchis*
243 species in Salep. *Anacamptis pyramidalis* (L.) Rich., *A. coriophora* (L.) R.M.Bateman, Pridgeon & M.W.Chase and *A.*

244 *palustris* (Jacq.) R.M.Bateman, Pridgeon & M.W.Chase were the main *Anacamptis* species. *Dactylorhiza umbrosa*
245 (Kar. & Kir.) Nevski was the only identified *Dactylorhiza* species. However, 24 out of 29 *Dactylorhiza* samples were
246 identified only to genus level. It is known that *Dactylorhiza* has a dynamic system of hybridization and allopolyploidy
247 formation (Hedrén *et al.* 2001, 2008). These allopolyploids show no clear genetic differentiations despite phenotypic
248 differences (Balao *et al.* 2015) and it is therefore difficult to identify these samples to species level using the applied
249 markers. Similarly, *Ophrys* was found to be one of the constituents of Salep but discerning the species used as Salep
250 with the BLAST similarity search was not possible. Species delimitation in closely related taxa of the genus *Ophrys* has
251 been challenging because of continuous introgression and absence of complete lineage sorting (Devey *et al.* 2008).

252 Tree-based identifications using ML and BI showed similar results: *Orchis* (51 samples for ML and 52 for BI),
253 *Anacamptis* (40 ML; 39 BI), *Dactylorhiza* (29 ML; 29 BI), *Ophrys* (18 ML; 18 BI), *Himantoglossum* (11 ML; 11 BI)
254 and *Steveniella* (1 ML; 1 BI) were the constituents of Salep (Suppl. 12). *Anacamptis* species in Salep samples are *A.*
255 *palustris*, *A. morio* (L.) R.M.Bateman, Pridgeon & M.W.Chase, *A. pyramidalis* and *A. coriophora*. *Orchis* species
256 contributing to Salep are *O. mascula*, *O. militaris* L. and *O. simia*. It was not possible to identify *Ophrys* and
257 *Dactylorhiza* samples to species level using the applied markers.

258 Figure 3 shows the species composition of studied Salep tubers based on final consensus identifications
259 including all markers and methods (Suppl. 12). The phylogenetic relationships among genera is based on Inda *et al.*
260 (2012). Based on final identification results the genera *Orchis* (51 samples), *Anacamptis* (40 samples), *Dactylorhiza* (29
261 samples), *Ophrys* (18 samples), *Himantoglossum* (11 samples) and *Steveniella* (1 sample) are the main the constituents
262 of studied Salep samples. All tuberous orchid species are used for Salep with a preference for species in the genera
263 *Orchis*, *Anacamptis* and *Dactylorhiza*.

264

265 *Generic composition of tubers per geographic origin*

266 The analyzed tubers can be geographically categorized into three zones of origin: a western zone (Ardabil, Eastern and
267 Western Azarbaijan, Kurdistan and Kermanshah provinces), a northern zone (Golestan) and a central zone (Tehran and
268 Esfahan). Sixty-five tubers originate from the western zone, and these include 26 tubers (38%) of *Anacamptis*, 22 tubers
269 (32%) of *Dactylorhiza* and 11 tubers (16%) of *Himantoglossum*. The generic composition of the 66 tubers from the
270 northern zone is different, and these include 42 tubers (64%) of *Orchis* and 15 tubers (23%) of *Ophrys*. The 15 tubers
271 from the central zone are mainly *Anacamptis* (8 samples, 53%) and *Dactylorhiza* (5 samples, 33%). Although
272 distribution and abundance of orchids in Iran is poorly documented, the results show that *Dactylorhiza* tubers, that trade
273 at a lower value in the market, are harvested in the western and central zones, whereas high-value *Orchis* tubers are

most commonly collected in the northern zone. Kasperek and Grimm (1999) report the presence of Iranian Salep in eastern Turkey in the 1990s, and Ghorbani *et al.* (2014a; b) writes that orchid tuber collection in western Iran has a longer history than in the north and east of Iran, where a recent boom is escalating harvesting and trade. The results could indicate that the resources for superior quality Salep tubers from *Orchis* species have been depleted in the western zone, and that Salep collection is now targeting the more inferior quality *Dactylorhiza* tubers. In the northern zone *Orchis* tubers are still readily available, but as natural populations dwindle collectors will target other genera.

280

281

282 **Conclusions**

This study has produced a resource of 238 reference sequences and 226 tuber sequences that can be used for identification of Orchidaceae species in the poorly documented Salep trade in Turkey, Greece and Albania. It also shows that genomic DNA of sufficient quality can be extracted and sequenced from highly processed Salep tubers. However, extraction of DNA is accompanied with some difficulties as a result of gel formation due to the high glucomannan content in the tubers. Post-harvest storage time of the tubers and boiling time during processing may also affect the quality of extracted DNA. Among the applied markers, nrITS and *trnL*-F spacer were easier to amplify and sequence than *matK* and these markers also show a higher discriminatory power for most of the genera. However, *Dactylorhiza* and *Ophrys*, that are known for allopolyploidy and hybridization, are challenging for barcoding using the applied markers, and a high-throughput sequencing gene capture approach would probably yield the right read depth for phasing of alleles and accurate species identification (Weitemier *et al.* 2014; Schmickl *et al.* 2015). The results also show that the genera most affected by Salep harvesting are *Orchis*, *Anacamptis*, *Dactylorhiza* and *Ophrys*. Geographic clustering of Salep tubers show clear differences in generic composition per zone with significant implications for harvesting pressure and resource depletion. *Dactylorhiza* and *Anacamptis* are more abundant as Salep tubers from the western zone, whereas *Orchis* and *Ophrys* are more abundant as Salep tubers from the northern zone. *Himantoglossum* was only present in Salep from the western zone. The results expose the overharvested species in each region that should be targeted for tailored conservation activities, and confirms the finding by Ghorbani *et al.* (2014a) that overharvesting of superior value *Orchis* tubers in western parts has led Salep middlemen and traders to tap into new areas in northern parts of the country. Conservation measures should be implemented in western, central and northern Iran to protect wild orchid populations from immediate threats due to unsustainable over-exploitation and to prevent their disappearance before many of them have even been studied properly.

303

304

305 **Acknowledgments**

306 Vincent Manzanilla is acknowledged for the graphic design and creation of Figure 3. Caroline Nilsson is acknowledged
307 for the graphic design and creation of Figure 1. Financial support from the Swedish Research Link program of the
308 Swedish Science Council is gratefully acknowledged. The Carl Tryggers Foundation for Scientific Research provided a
309 postdoctoral research stipend to AG (through HdB). Further support for fieldwork was gratefully received from the
310 Anne S. Chatham fellowship of the Garden Club of America (AG), Sven och Dagmar Saléns stiftelse (HdB, AG) and
311 Helge Ax:son Johnsons Stiftelse (HdB).

312

313

314 **Author contributions**

315 AG, HdB and BG devised the project. AG carried out the vast majority of the fieldwork, assisted by SZ, HdB and BG.
316 AG, SS and HdB analyzed the data. AG, SS and HdB wrote the first draft of the manuscript. All authors have read and
317 approve the final manuscript.

318

319

320 **References**

- 321 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular*
322 *Biology*, **215**, 403–410.
- 323 Aubriot X, Lowry PP, Cruaud C, Couloux A, Haevermans T (2013) DNA barcoding in a biodiversity hot spot: potential
324 value for the identification of Malagasy *Euphorbia* L. listed in CITES Appendices I and II. *Molecular Ecology*
325 *Resources*, **13**, 57–65.
- 326 Babbar, SB, Singh, DK (2016) Protocols for In Vitro Mass Multiplication and Analysis of Medicinally Important
327 Phenolics of a Salep Orchid, *Satyrium nepalense* D.Don (“Salam Mishri”). In: *Protocols for In Vitro Cultures*
328 *and Secondary Metabolite Analysis of Aromatic and Medicinal Plants, Second Edition* (ed Mohan Jain S), pp
329 1-11, Springer Protocols, Springer US.
- 330 Balao F, Tannhäuser M, Lorenzo MT, Hedrén M, Paun O (2015) Genetic differentiation and admixture between sibling
331 allopolyploids in the *Dactylorhiza majalis* complex. *Heredity*.

Barthet MM, Moukarzel K, Smith KN, Patel J, Hilu KW (2015). Alternative translation initiation codons for the plastid maturase MatK: unraveling the pseudogene misconception in the Orchidaceae. *BMC Evolutionary Biology*, **15**, 210.

Bellstedt DU, Linder HP, Harley EH (2001) Phylogenetic relationships in *Disa* based on non-coding *trnL-trnF* chloroplast sequences: evidence of numerous repeat regions. *American Journal of Botany*, **88**, 2088–2100.

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.

Bulpitt CJ (2005) The uses and misuses of orchids in medicine. *QJM*, **98**, 625–631.

Bulpitt CJ, Li Y, Bulpitt PF, Wang J (2007) The use of orchids in Chinese medicine. *Journal of the Royal Society of Medicine*, **100**, 558–563.

Burgess KS, Fazekas AJ, Kesanakurti PR *et al.* (2011) Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. *Methods in Ecology and Evolution*, **2**, 333–340.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cameron KM (2004) Utility of plastid *psaB* gene sequences for investigating intrafamilial relationships within Orchidaceae. *Molecular Phylogenetics and Evolution*, **31**, 1157–1180.

Chinsamy M, Finnie JF, Van Staden J (2011) The ethnobotany of South African medicinal orchids. *South African Journal of Botany*, **77**, 2–9.

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

Coghlan ML, Haile J, Houston J *et al.* (2012) Deep Sequencing of Plant and Animal DNA Contained within Traditional Chinese Medicines Reveals Legality Issues and Health Safety Concerns. *PLoS Genetics*, **8**.

Deguilloux M-F, Pemonge M-H, Petit RJ (2002) Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proceedings of the Royal Society B*, **269**, 1039–46.

Delforge P (2006) *Orchids of Europe, North Africa and the Middle East*. A&C Black, London, UK.

Devey DS, Bateman RM, Fay MF, Hawkins JA (2008) Friends or relatives? Phylogenetics and species delimitation in the controversial European orchid genus *Ophrys*. *Annals of Botany*, **101**, 385–402.

Dogan M, Kayacier A, Ic E (2007). Rheological characteristics of some food hydrocolloids processed with gamma irradiation. *Food Hydrocolloids*, **21**, 392–396.

Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

362 Ebihara A, Nitta JH, Ito M (2010) Molecular species identification with rich floristic sampling: DNA barcoding the
363 pteridophyte flora of Japan. *PLOS ONE*, **5**, e15136.

364 Ece Tamer C, Karaman B, Utku Copur O (2006) A traditional Turkish beverage: Salep. *Food Reviews International*,
365 **22**, 43–50.

366 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids*
367 *Research*, **32**, 1792–1797.

368 Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome
369 discriminate plant species equally well. *PLOS ONE*, **3**, e2802.

370 Van Geel B, Aptroot A, Baittinger C *et al.* (2008) The ecological implications of a Yakutian mammoth's last meal.
371 *Quaternary Research*, **69**, 361–376.

372 Ghorbani A, Gravendeel B, Naghibi F, de Boer HJ (2014a) Wild orchid tuber collection in Iran: a wake-up call for
373 conservation. *Biodiversity and Conservation*, **23**, 2749–2760.

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

376 Gravendeel B, Chase MW, de Vogel EF *et al.* (2001) Molecular phylogeny of *Coelogyne* (Epidendroideae;
377 Orchidaceae) based on plastid RFLPS, matK, and nuclear ribosomal ITS sequences: evidence for polyphyly.
378 *American Journal of Botany*, **88**, 1915–1927.

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

381 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings.*
382 *Biological Sciences / The Royal Society*, **270**, 313–321.

383 Hedrén M, Fay MF, Chase MW (2001) Amplified fragment length polymorphisms (AFLP) reveal details of polyploid
384 evolution in *Dactylorhiza* (Orchidaceae). *American Journal of Botany*, **88**, 1868–1880.

385 Hedrén M, Nordström S, Ståhlberg D (2008) Polyploid evolution and plastid DNA variation in the *Dactylorhiza*
386 *incarnata/maculata* complex (Orchidaceae) in Scandinavia. *Molecular Ecology*, **17**, 5075–5091.

387 Hibert F, Taberlet P, Chave J *et al.* (2013) Unveiling the Diet of Elusive Rainforest Herbivores in Next Generation
388 Sequencing Era? The Tapir as a Case Study. *PLoS ONE*, **8**, e60799.

389 Hossain MM (2011) Therapeutic orchids: traditional uses and recent advances-An overview. *Fitoterapia*, **82**, 102–140.

Inda LA, Pimentel M, Chase MW (2012) Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Annals of Botany*, **110**, 71–90.

Kasperek M, Grimm U (1999) European trade in Turkish Salep with special reference to Germany. *Economic Botany*, **53**, 396–406.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, **28**, 1647–1649.

Kim HM, Oh S-H, Bhandari GS, Kim C-S, Park C-W (2014) DNA barcoding of Orchidaceae in Korea. *Molecular Ecology Resources*, **14**, 499–507.

Kocyan A, Qiu Y-L, Endress PK, Conti E (2004) A phylogenetic analysis of Apostasioideae (Orchidaceae) based on ITS, trnL-F and matK sequences. *Plant Systematics and Evolution*, **247**.

Kool A, de Boer HJ, Krüger A *et al.* (2012) Molecular identification of commercialized medicinal plants in southern Morocco. *PloS One*, **7**, e39459.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences*, **102**, 8369–8374.

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

Kuzmina ML, Johnson KL, Barron HR, Hebert PD (2012) Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. *BMC Ecology*, **12**, 25.

Landerer X (1850) Beiträge zur Pharmakognosie. *Archiv der Pharmazie*, **112**, 177–180.

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276–3278.

Little DP, Stevenson DW (2007) A comparison of algorithms for the identification of specimens using DNA barcodes : examples from gymnosperms. *Cladistics*, **23**, 1–21.

Liu JIE, Moeller M, Gao LM, Zhang D., Li DZ (2011) DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources*, **11**, 89–100.

Liu J, Yan H-F, Newmaster SG *et al.* (2015) The use of DNA barcoding as a tool for the conservation biogeography of subtropical forests in China. *Diversity and Distributions*, **21**, 188–199.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Gateway Computing Environments Workshop (GCE), 2010*, pp. 1–8.

420 Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S (2013) DNA barcoding detects
 421 contamination and substitution in North American herbal products. *BMC Medicine*, **11**, 222.

422 Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics—bridging the gap between conservation genetics
 423 and law enforcement. *Endangered Species Research*, **9**, 179–195.

424 Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian
 425 Paphiopedilum species. *Molecular Ecology Resources*, **12**, 82–90.

426 Posadzki P, Watson L, Ernst E (2012) Contamination and adulteration of herbal medicinal products (HMPs): an
 427 overview of systematic reviews. *European Journal of Clinical Pharmacology*, **69**, 295–307.

428 Rambaut A, Suchard M, Xie D, Drummond J (2014) Tracer v1.6.

429 Renz J (1978) *Flora Iranica. Part 126: Orchidaceae*. Naturhistorisches Museum, Vienna, Austria.

430 Richardson RT, Lin C-H, Sponsler DB *et al.* (2015) Application of ITS2 Metabarcoding to Determine the Provenance
 431 of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences*, **3**, 1400066.

432 Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and
 433 model choice across a large model space. *Systematic Biology*, **61**, 539–542.

434 Sandal Erzurumlu G, Doran I (2011) Türkiye’de salep orkideleri ve salep kültürü. *J.Agric. Fac. HR.U.*, **15**, 29–34.

435 Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT (2012) How to Open the Treasure Chest? Optimising
 436 DNA Extraction from Herbarium Specimens. *PLoS ONE*, **7**, e43808.

437 Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA barcoding in the cycadales: testing the potential of proposed
 438 barcoding markers for species identification of cycads. *PLOS ONE*, **2**, e1154.

439 Schmickl R, Liston A, Zeisek V *et al.* (2015) Phylogenetic marker development for target enrichment from
 440 transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae).
 441 *Molecular ecology resources [Early View]*.

442 Sezik E (2002a) Turkish orchids and Salep. *Acta Pharmaceutica Turcica*, **44**, 151–157.

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

445 Sezik E (2006) Destroying of Ophrys species to obtain Salep in turkey. *Journal Europäischer Orchideen*, **38**, 290.

446 Shahsavari A (2008) *Flora of Iran. Part 57: Orchidaceae*. Research Institute of Forests and Rangelands, Tehran, IR
 447 Iran.

448 Shepherd LD, McLay TGB (2011) Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant
 449 tissue. *Journal of Plant Research*, **124**, 311–314.

450 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
451 *Bioinformatics*, **btu033**.

452 Starin D (2012) Salepi Extinction, Salepi Survival: How a Change in Ingredients Could Help Safeguard Orchids.
453 *Orchids: the Bulletin of the American Orchid Society*, **81**, 490–494.

454 Steele KP, Vilgalys R (1994) Phylogenetic Analyses of Polemoniaceae Using Nucleotide Sequences of the Plastid Gene
455 *matK*. *Systematic Botany*, **19**, 126–142.

456 Stucky BJ (2012) SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. *Journal of*
457 *Biomolecular Techniques*, **23**.

458 Sun Y, Skinner DZ, Liang GH, Hulbert SH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal
459 transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, **89**, 26–32.

460 Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of
461 chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

462 Tekinşen KK, Güner A (2010) Chemical composition and physicochemical properties of tubera salep produced from
463 some Orchidaceae species. *Food Chemistry*, **121**, 468–471.

464 Thompson KA, Newmaster SG (2014) Molecular taxonomic tools provide more accurate estimates of species richness
465 at less cost than traditional morphology-based taxonomic practices in a vegetation survey. *Biodiversity and*
466 *Conservation*, **23**, 1411–1424.

467 Vassou SL, Kusuma G, Parani M (2015) DNA barcoding for species identification from dried and powdered plant parts:
468 A case study with authentication of the raw drug market samples of *Sida cordifolia*. *Gene*, **559**, 86–93.

469 Van Velzen R, Weitschek E, Felici G, Bakker FT (2012) DNA barcoding of recently diverged species: relative
470 performance of matching methods. *PloS One*, **7**, e30490.

471 De Vere N, Rich TC, Ford CR *et al.* (2012) DNA barcoding the native flowering plants and conifers of Wales. *PLoS*
472 *One*, **7**, e37945.

473 Weitemier K, Straub SC, Cronn RC *et al.* (2014) Hyb-Seq: Combining target enrichment and genome skimming for
474 plant phylogenomics. *Applications in Plant Sciences*, **2**, 1400042.

475 Williams VL, Balkwill K, Witkowski ETF (2000) Unraveling the commercial market for medicinal plants and plant
476 parts on the witwatersrand, South Africa. *Economic Botany*, **54**, 310–327.

477 Xiang X-G, Hu H, Wang W, Jin X-H (2011) DNA barcoding of the recently evolved genus *Holcoglossum*
478 (Orchidaceae: Aeridinae): a test of DNA barcode candidates. *Molecular Ecology Resources*, **11**, 1012–1021.

479

480 **Data Accessibility**

481 The concatenated sequence matrix of all reference and tubers species, as well as resulting BI and ML phylogenetic trees
482 (Suppl. 4-11) are deposited in Dryad, <http://dx.doi.org/10.5061/dryad.qb36g>.

483

484 **Tables and Figures**

485 **Figure 1.** Distribution of Salep genera in Iran and location of main markets. Shaded areas show the three zones of
486 origin: western, central and northern zones.

487 **Figure 2.** Tuber samples of different morphology purchased from the markets. A. Samples of *Orchis/Anacamptis* type
488 tubers. B. Samples of *Dactylorhiza* type tubers.

489 **Figure 3.** Identifications of screened Iranian Salep tubers. A. Shaded genera occur in Iran. Phylogeny adapted from
490 Inda et al. 2012. B. Proportion of identified genera. C. Filled circles represent the number of tubers identified to a
491 particular species.

492

493 **Table 1.** Species and samples per genus in sequence reference library.

494 **Table 2.** Sequence matrix and Bayesian analysis data.

495 **Table 3.** Molecular identification of Salep tuber to species and genus level.

496

497 **Supplemental Data**

498 **Supplement 1.** Reference sequences derived from vouchers collected for this study.

499 **Supplement 2.** Reference sequences derived from external NCBI GenBank accessions.

500 **Supplement 3.** GenBank accession numbers of the Salep tubers.

501 **Supplement 4.** RAxML maximum likelihood phylogenetic tree for nrITS.

502 **Supplement 5.** RAxML maximum likelihood phylogenetic tree for *trnL-F* spacer.

503 **Supplement 6.** RAxML maximum likelihood phylogenetic tree for *matK*.

504 **Supplement 7.** RAxML maximum likelihood phylogenetic tree for the concatenated matrix.

505 **Supplement 8.** MrBayes bayesian phylogenetic tree for nrITS.

506 **Supplement 9.** MrBayes bayesian phylogenetic tree for *trnL-F* spacer.

507 **Supplement 10.** MrBayes bayesian phylogenetic tree for *matK*.

508 **Supplement 11.** MrBayes bayesian phylogenetic tree for the concatenated matrix.

509 **Supplement 12.** Molecular identifications of tubers based on similarity- and tree-based approaches.

510

511 **Table 1.** Species and samples per genus in sequence reference library.

	Identification Reference Resource			
	GenBank		Field collections	
Genus	# Samples	# Species	# Samples	# Species
<i>Anacamptis</i>	6	5	17	3
<i>Brownleea</i>	1	1	NA	NA
<i>Cephalanthera</i>	3	3	10	5
<i>Chamorchis</i>	1	1	NA	NA
<i>Corallorhiza</i>	1	1	NA	NA
<i>Dactylorhiza</i>	20	18	17	3
<i>Disa</i>	1	1	NA	NA
<i>Disperis</i>	1	1	NA	NA
<i>Epipactis</i>	1	1	7	2
<i>Gennaria</i>	1	1	NA	NA
<i>Goodyera</i>	1	1	NA	NA
<i>Gymnadenia</i>	8	7	NA	NA
<i>Habenaria</i>	1	1	NA	NA
<i>Herminium</i>	1	1	NA	NA
<i>Himantoglossum</i>	10	10	7	2
<i>Limodorum</i>	NA	NA	2	1
<i>Neotinea</i>	5	4	NA	NA
<i>Neottia</i>	3	3	3	1
<i>Neottianthe</i>	3	2	NA	NA
<i>Ophrys</i>	41	23	35	4
<i>Orchis</i>	8	7	21	4
<i>Pecteilis</i>	1	1	NA	NA
<i>Peristylus</i>	1	1	NA	NA
<i>Platanthera</i>	2	2	5	2
<i>Pseudorchis</i>	1	1	NA	NA
<i>Satyrium</i>	1	1	NA	NA
<i>Serapias</i>	8	7	NA	NA
<i>Spiranthes</i>	2	2	NA	NA
<i>Steeniella</i>	1	1	3	1
<i>Zeuxine</i>	1	1	NA	NA

513 **Table 2.** Sequence matrix and Bayesian analysis data.

Markers	No. of sequences			
	Reference (R)	Leaf (L)	Tuber (T)	Total
nrITS	126	85	104	315
<i>trnL</i> -F	68	90	94	252
<i>matK</i>	58	63	28	149
Concatenated	138	135	150	423

514

Markers	Alignment matrix		Bayesian analysis
	Seq length incl. gaps (bp)	Min/max length without gaps (bp)	No. trees retained
nrITS	822	209/722	30 002
<i>trnL</i> -F	1663	287/1032	30 002
<i>matK</i>	1173	365/1105	15 002
Concatenated	3658	209/2677	30 002

515

516 **Table 3.** Molecular identification of Salep tuber to species and genus level.
517

	Samples for which sequences were obtained									
	ITS		<i>trnL</i> -F		<i>matK</i>		Concatenated		Consensus	
Sequenced samples	104		94		28		150		150	
	Similarity (BLAST) identification									
Species	45	43%	33	35%	17	61%	-	-	57	38%
Genus	59	57%	61	65%	11	39%	-	-	93	62%
	Maximum likelihood (RAxML) identification									
Species	70	67%	55	59%	16	57%	63	42%	90	60%
Genus	34	33%	39	41%	12	43%	87	58%	60	40%
	Bayesian inference (MrBayes) identification									
Species	71	68%	55	59%	19	68%	102	68%	97	65%
Genus	33	32%	39	41%	9	32%	48	32%	53	35%

518