USING TARGET CAPTURE TO ADDRESS CONSERVATION CHALLENGES: POPULATION-LEVEL TRACKING OF A GLOBALLY-TRADED HERBAL MEDICINE.

Vincent Manzanilla*[1], Irene Teixidor-Toneu[1], Gary J. Martin[2], Peter M. Hollingsworth[3], Hugo J. de Boer[1], Anneleen Kool[1]

[1] Natural History Museum, University of Oslo, Sars gate 1, 0562 Oslo, Norway

[2] Global Diversity Foundation, 37 St. Margarets Street, Canterbury, Kent CT1 2TU, England

[3] Royal Botanic Garden Edinburgh, Inverleith Row, Edinburgh EH3 5LR, Scotland

* Corresponding author: Vincent Manzanilla Email: vincent.manzanilla@gmail.com

ABSTRACT

The promotion of responsible and sustainable trade in biological resources is widely proposed as one solution to mitigate currently high levels of global biodiversity loss. Various molecular identification methods have been proposed as appropriate tools for monitoring global supply chains of commercialized animals and plants. We demonstrate the efficacy of target capture genomic barcoding in identifying and establishing the geographic origin of samples traded as *Anacyclus pyrethrum*, a medicinal plant assessed as globally vulnerable in the IUCN Red List of Threatened Species. Samples collected from national and international supply chains were identified through target capture sequencing of 443 low-copy nuclear makers and compared to results derived from genome skimming of plastome, standard plastid barcoding regions and ITS. Both target capture and genome skimming provided approximately 3.4 million reads per sample, but target capture largely outperformed standard plant DNA barcodes and entire plastid genome sequences. Despite the difficulty of distinguishing among closely related species and infraspecific taxa of *Anacyclus* using conventional taxonomic methods, we succeeded in identifying 89 of 110 analysed samples to subspecies level without ambiguity through target capture. Of the remaining samples, we determined that eleven contained plant material from other genera and families and ten were unidentifiable regardless of the method used. Furthermore, we were able to discern the geographical origin of *Anacyclus* samples collected in Moroccan, Indian and Sri Lankan markets, differentiating between plant materials originally harvested from diverse populations in Algeria and Morocco. With a recent drop in the cost of analysing samples, target capture offers the potential to routinely identify commercialized plant species and determine their geographic origin. It promises to play an important role in monitoring and regulation of plant species in trade, supporting biodiversity

conservation efforts, and in ensuring that plant products are unadulterated, contributing to consumer protection.

Keywords: Anacyclus, DNA barcoding, genomic barcoding, international trade, supply chain, target capture

Human exploitation of biological resources is a major challenge for biodiversity conservation and sustainable development. Global trade and consumer demand for natural products provide increasing threats to species (Lenzen et al., 2012), and at the same time create markets in which regulation and authentication are extremely difficult (Newmaster, Grguric, Shanmughanandhan, Ramalingam, & Ragupathy, 2013; WHO, 2004).

Many traded plant products from wild populations are over-harvested. Increasing scarcity results in higher prices, and incentivises adulteration, substitution and poaching (Hamilton, 2004; Schippmann, Leaman, & Cunningham, 2002; Veldman et al., 2017). In recent decades, multilateral environmental agreements including the Convention on Biological Diversity (CBD) and the Convention on International Trade in Endangered Species (CITES) have addressed the trade of threatened species. In parallel, the World Health Organisation has developed guidelines for safety monitoring of herbal medicines in pharmacovigilance systems (WHO, 2004) and the Food and Agriculture Organisation, through the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), has set up a multilateral system to promote sustainable agriculture (Esquinas-Alcazar, 2004). However, implementing these regulations and guidelines is hampered by difficulties identifying plant products in trade. Multiple, complex and interacting supply chains can co-exist for a single plant product (A. Booker, Johnston, & Heinrich, 2012). Traded plant products are often not identifiable to species by their morphology or chemistry, as they may be dried, powdered, processed, or commercialised in mixtures with other ingredients. The design, implementation and enforcement of successful conservation actions as well as assessments of product authenticity and quality often require the identification of the geographic origin of species in trade. This is difficult as the development of efficient methods to identify and trace traded products are still in their infancy.

63 DNA barcoding has quickly gained popularity since Hebert et al. (2003) first
64 advocated the use of short and variable DNA sequences, amplified using universal primers,
65 for species identification and discovery of new taxa. DNA barcoding is highly effective for
66 species-level identification in animals using a portion of the mitochondrial marker
67 Cytochrome Oxidase 1 (COI) (P. D. Hebert, Hollingsworth, & Hajibabaei, 2016; P. D.
68 Hebert, Ratnasingham, et al., 2016). In plants, standard DNA barcoding involves one to four
69 plastid DNA regions (*rbcL, matK, trnH-psbA, trnL*), sometimes in combination with internal
70 transcribed spacers of nuclear ribosomal DNA (nrDNA, ITS) (CBOL et al., 2009; Kress,
71 2017). Although these markers are very informative in many cases, no single marker or
72 combination of markers routinely provide complete species-level resolution, especially in
73 species-rich groups, let alone population-level assignment.

74 The development of high throughput sequencing (HTS) with new reagents and
75 platforms expands the application of DNA barcoding in plants in a cost-effective fashion, in
76 part because it removes the need of targeting short universal barcodes (Hollingsworth, Li, van
77 der Bank, & Twyford, 2016; Lemmon & Lemmon, 2013). Two major approaches, shallow
78 pass shotgun sequencing and target capture sequencing, have been proposed for increasing the
79 coverage and resolution of plant DNA barcoding.

80 Shallow pass shotgun sequencing (commonly referred as genome skimming) is used to
81 recover organellar genomes and nuclear ribosomal DNA sequences, increasing the amount of
82 data per sample and leading to some increases in resolution (Manzanilla et al., 2018; Parks,
83 Cronn, & Liston, 2009). Although workflows and bioinformatic pipelines are increasingly
84 refined for this approach, current cost constraints mean that most genome skimming barcoding
85 projects only have sufficient sequencing depth to generate comparative data for multi-copy
86 regions such as plastid genomes and ribosomal DNA. These regions represent a limited

87 number of independent loci, ultimately constraining resolving power (Soltis & Soltis, 2009;
88 Wood et al., 2009).

89 Target capture sequencing offers the potential to overcome these deficiencies by
90 efficiently targeting hundreds of low-copy nuclear markers, providing access to a much
91 greater number of independent data points per unit of sequencing effort (Degnan &
92 Rosenberg, 2009). Similar to genome skimming, target capture is successful in sequencing
93 samples with poor DNA integrity (Brewer et al., 2019; Forrest et al., 2019) and allows
94 sequencing hundreds of samples at the same time (Mamanova et al., 2010). It can also be
95 designed to recover standard DNA barcodes in the same assay (Schmickl et al., 2016).
96 Although target capture has been advocated as a powerful tool for molecular identification of
97 plants (Pillon et al., 2013), its usefulness for determining and tracing the origin of plants in trade
98 remained untested.

99 We evaluated the power of target-capture DNA barcoding by investigating the
100 traceability of plant products reputedly derived from a vulnerable medicinal plant species,
101 *Anacyclus pyrethrum* (L.) Lag. In addition to having a well-established international trade chain
102 (Rankou, Ouhammou, Taleb, Manzanilla, & Martin, 2015), this species presents classic
103 challenges for plant molecular identification such as recent radiation, frequent hybridization
104 (Humphries, 1979) and large genome size (Garcia et al., 2013).

105 The genus *Anacyclus* (Asteraceae) comprises 12 species of annual and perennial
106 weedy herbs with partly overlapping geographic ranges around the Mediterranean basin
107 (Humphries, 1979; Rosato, Álvarez, Feliner, & Rosselló, 2017). Some species are abundant
108 and have wide geographical ranges (for example, *A. clavatus* (Desf.) Pers. and *A. radiatus*
109 Loisel.), whereas others are rare and have restricted ranges (for example, *A. maroccanus*
110 (Ball) Ball and *A. pyrethrum* (L.) Lag.).

111 *A. pyrethrum* has a long history of use in traditional Arabic and Islamic, Ayurvedic and
112 European medicine (Adams, Alther, Kessler, Kluge, & Hamburger, 2011; De Vos, 2010; Pittle, 2005).
113 In the 13th century, Ibn al-Baytār wrote that the plant was known across the world and traded from the
114 Maghreb to all other areas (Leclerc, 1877). Its popularity as a medicinal plant stems from the many
115 attested and putative pharmacological activities of its roots (Manouze et al., 2017). Traded from the
116 Maghreb to India (Ved & Goraya, 2007) and Nepal (Tiwari et al., 2004), it is known to be over-
117 harvested and is increasingly difficult to find in local markets in Morocco (Ouarghidi et al., 2013,
118 2012).

119 There are two accepted varieties, *A. pyrethrum* var. *pyrethrum* and var. *depressus*, both
120 endemic to Algeria, Morocco, and southern Spain (Humphries, 1979; Rosato et al., 2017). In Morocco,
121 *A. pyrethrum* var. *pyrethrum* is considered more potent and is up to ten times more expensive than var.
122 *depressus* (Ouarghidi, Powell, Martin, de Boer, & Abbad, 2012). Both varieties are harvested from the
123 wild and used extensively for the treatment of pain and inflammatory disorders across Morocco
124 (Ouarghidi, Martin, Powell, Esser, & Abbad, 2013; Ouarghidi et al., 2012) and Algeria (Benarba,
125 2016; Ouelbani, Bensari, Mouas, & Khelifi, 2016), as well as the Middle East (Pittle, 2005) and the
126 Indian sub-continent (Tiwari, Poudel, & Uprety, 2004). Although collectors are proficient in
127 differentiating the two varieties, material traded as *A. pyrethrum* is adulterated and misidentified along
128 the chain of commercialisation (de Boer, Ouarghidi, Martin, Abbad, & Kool, 2014; Ouarghidi et al.,
129 2013, 2012).

130 We applied target-capture genomic barcoding to distinguish *Anacylus* species and
131 geographical races in root samples in the national and international supply chains. We compare
132 this novel approach with plastid genome and nrDNA ITS barcodes obtained from genome
133 skimming data.

134 **MATERIALS AND METHODS**

135 **Sample collection.** We made a reference collection of 72 accessions, consisting of 67 of *Anacyclus*
136 and five of closely related genera, which includes 56 *Anacyclus* accessions and three of closely related
137 genera we collected in Morocco and Spain during field work. We selected a total of 11 samples of
138 *Anacyclus* and two of closely related genera from herbarium voucher specimens of species occurring
139 elsewhere in the Mediterranean. Voucher specimen identifications, collection numbers and locality
140 data are listed in Table S1 and the collection locations are mapped in Figure S1. We purchased 110
141 samples, each consisting of approximately 50g of roots from collectors, herbalists, middle-men,
142 traditional healer, wholesalers, and export companies in Morocco and India (Table S2).

143 **Trade information.** We obtained samples of commercialized species by asking collectors and traders
144 in Morocco, for tiguendizt and iguendez, two of vernacular names used for both *A. pyrethrum*
145 varieties (Ouarghidi et al., 2012). We asked traders in India for akarkara, a common name used there
146 (Ved & Goraya, 2007). When acquiring the samples, we conducted semi-structured interviews about
147 the roots' trade with 39 informants, enquiring where the plant material was sourced, to whom it was
148 sold and in what quantities, for what price, and if some types of roots were of higher quality than
149 others. We followed the International Society of Ethnobiology Code of Ethics ("ISE Code of Ethics
150 Online," 2018), ensuring free, prior and informed consent, full disclosure and respect for the
151 confidentiality of informants, during all interviews.

152 **Extraction and library preparation.** We extracted DNA from reference and traded vouchers
153 from approximately 40 mg of dry leaf or root material using the DNeasy Plant Mini Kit
154 (Qiagen). Total DNA (0.2-1.0 µg) was sheared to 500 bp fragments using a Covaris S220
155 sonicator (Woburn, MA, USA) (Table S1-2). We prepared dual indexed libraries using the
156 Meyer and Kircher protocol (Meyer & Kircher, 2010) for genome skimming and target
157 capture (BioProject PRJNA631886).

158 **Target capture.** Using the genome assembly of *A. radiatus* subsp. *radiatus*, we designed 872
159 low-copy nuclear markers and associated RNA probes by following the Hyb-Seq pipeline

160 (Weitemier et al., 2014) (SI). For target capture enrichment, we prepared twelve equimolar
161 pools with ten to 24 samples and an average 300 ng of input DNA per pool. The RNA probes
162 were hybridized for 16 hours before target baiting, and 14 PCR cycles were carried out after
163 enrichment following the MyBaits v3 manual. The enriched libraries and genome skimming
164 libraries were sequenced on two Illumina HiSeq 3000 lanes (150bp paired-end).

165 **Data Processing.** We retrieved four datasets from the genome skimming and target capture
166 sequencing methods: (1) standard barcode markers (*matK*, *trnH-psbA*, the *trnL* intron and
167 *rbcL*), (2) ITS, (3) complete plastid genomes (from shotgun genome skimming), and (4)
168 hundreds of nuclear markers (from target capture) (Figure S2). The sequencing runs were
169 trimmed and quality filtered using Trimmomatic (Bolger, Lohse, & Usadel, 2014). Low-copy
170 nuclear markers and their alleles were retrieved for each sample. First, the reads were mapped
171 against the selected low-copy nuclear markers (SI) using BWA v0.7.5a-r40 (Heng Li &
172 Durbin, 2009). Duplicate reads were removed using Picard v2.10.4 (Wysoker, Tibbetts, &
173 Fennell, 2015). Alleles were phased for each marker and individual using SAMtools v1.3.1
174 (H. Li et al., 2009). The last step of the pipeline combined the retrieved alleles into single
175 gene matrices. We recovered plastome and ITS sequences by pooling shotgun and target
176 enrichment sequencing data. Plastid genomes were built using MITOBim v1.8 (Hahn,
177 Bachmann, & Chevreux, 2013). ITS sequences were recovered using BWA by mapping the
178 reads to the reference ITS of *Anacyclus pyrethrum* (KY397478) for *Anacyclus* species and
179 traded samples, to the reference ITS of *Achillea pyrenaica* Sibth. ex Godr. (AY603247) for
180 *Otanthus* and *Achillea*, and to the reference ITS of *Matricaria aurea* (Loefl.) Sch.Bip.
181 (KT954177) for *Matricaria* samples. During the mapping and the assembly steps, we retained
182 sequences only with a minimum coverage of 10X.

183 **Phylogenomics.** We aligned the recovered matrices (nuclear markers, ITS and plastomes) with
184 MAFFT v7.471 (Katoh & Standley, 2013), refined with MUSCLE (Edgar, 2004) and filtered
185 with Gblocks v0.91b (Talavera & Castresana, 2007). Phylogenies were inferred using
186 RAxML v8.0.26 (Stamatakis, 2006), with 1000 bootstrap replicates under the GTRGAMMA
187 model. For the low-copy nuclear markers, the species tree was inferred from the individual
188 nuclear markers trees under the multi-species coalescence (MSC) framework with ASTRAL-
189 III v5.5.9 (Zhang, Sayyari, & Mirarab, 2017). We used the multi-alleles option in ASTRAL-III
190 for reconciliation of the independent evolutionary histories of the alleles. The molecular
191 identification of traded roots was assessed from the MSC tree and posterior probabilities (PP)
192 greater than 0.95. For the ITS phylogenetic reconstruction, we used additional Genbank
193 references (Table S3).

194 We identified traded roots based on morphological characters (described in (SI)), and we used
195 these to triangulate molecular identifications. Samples were categorised according to their
196 position in the supply chain and geographical origin.

197 RESULTS

198 We constructed a reference database of DNA sequences from fresh and herbarium specimens,
199 consisting of 83 individuals of 10 *Anacyclus* species, and 5 individuals representing outgroup
200 species (Figure S1). We used this reference database to assess the identity and geographic
201 origins of 110 root samples acquired from traded materials by comparing the results from the
202 four different datasets (Figure 1). We show that the target capture approach is the most
203 powerful method to identify plant species in trade and their geographic origin.

204 Data recovery for genome skimming and target capture data

205 After quality control filtering, an average 2.8 million reads (0.42 GB/sample) were obtained
206 per sample from shotgun sequencing and 2.99 million reads (0.43 GB/sample) per sample for
207 target capture (Figure S3, Table S4). The target capture yielded an average coverage of 303X
208 for the 443 nuclear markers, whereas the unenriched genome skimming yielded an average
209 coverage of 12X for the 443 nuclear data, 20X for the plastome data and 131X for the
210 ribosomal data (Table S5). The loci coverage was calculated with *bedtools coverage* (v
211 2.29.2). Samples below an average of 50X coverage in the nuclear dataset, show a higher
212 missing data rate (>7%) in the matrices. These samples were automatically discarded with our
213 pipeline (Figure 2, samples in blue). Adulterated samples from other genera have a coverage
214 close to zero (Figure 2, samples in orange, yellow and green). To obtain 100x coverage for the
215 nuclear regions using a genome skimming approach, with an average genome size of the
216 targeted species of 11.72Gb (Garnatje et al., 2011) and a duplication level of the genome
217 skimming libraries of 9% (SI), it would require 14 HiSeq 3000/4000 lanes (Illumina, n.d.).
218 From the genome skimming data we assembled ITS and the standard barcodes markers, as
219 well as the plastome (Figure S2). Out of the 110 trade samples, we succeeded in assembling
220 ITS for 102 samples (93%), the standard barcode regions for 51 to 61 samples (46% to 55%),
221 and plastomes for 49 samples (44%) (Figure 1). The resulting aligned matrices for each of the
222 datasets were 633 bp for ITS (including 5.8S), 4408 bp for the standard barcoding regions,
223 110,003 bp for the plastome and 289,236 bp for the 443 nuclear markers recovered from the
224 target capture approach (Table S1). The standard barcoding regions included the full coding
225 regions of *matK* 1523 bp and *rbcL* 1438 bp, as well as *trnH-psbA* 500 bp and *trnL* 947 bp.
226 The bioinformatics workflow for data analyses is described in Figure S2.

227 **Comparative levels of species discrimination using different approaches**

The ITS, plastome, and standard barcodes phylogenies highlight the complex evolutionary history of *Anacyclus*. The ITS phylogeny lacks resolution in general (Figure S4-5). The outgroups *Tanacetum*, *Matricaria*, *Achillea*, *Othanthus* and *Tripleurospermum* have well-supported bootstrap values, but within the genus *Anacyclus,* only *A. atlanticus* Litard. & Maire, *A. maroccanus* and *A. radiatus* are highly supported. The plastome phylogeny shows very good support at genus level for the *Anacyclus* node, and at species level for the outgroups. The lack of variation in the plastid genome within the genus *Anacyclus* results in little phylogenetic support with no species-specific clusters recovered (Figure S6-7). The standard barcode regions, *matK*, *rbcL, trnH-psbA* and *trnL* (Figure S8-11) displayed low levels of resolution at the species level, even using the full coding regions of *matK* and *rbcL* (e.g. rather than the 800-900 bp of *matK* and 654 bp of *rbcL* typically recovered using standard barcoding primers;(Alsos et al., 2020; Hollingsworth, Graham, & Little, 2011)).

The 443 nuclear markers recovered by target capture, led to a well-resolved phylogeny and high levels of species discrimination: all the genera in the Matricariinae tribe and all interspecific relationships are well-supported, with most nodes showing posterior probabilities (PP) of 1 (Figure 3 and S12. Within *Anacyclus,* all species, sub-species and varieties are well supported. PP are lower for *A. monanthos* (PP = 0.75). The complex of hybrid species composed of *A. clavatus*, *A. homogamos*, and *A. valentinus* is polyphyletic and shows signs of hybridization and incomplete lineage sorting. Intraspecific nodes have PP varying between 0.27 and 1, mostly depending on species population structure.

Assessment of *Anacyclus* trade

Interviews with 39 harvesters, middlemen, retailers, and wholesalers in various Moroccan cities, indicate that the national and international trade of *Anacyclus pyrethrum* follow two separate supply chains (SI). Retailer herbalists in Moroccan cities are supplied by middlemen who acquire the plant from local harvesters from rural communities. These retailers typically hold between a few hundred grams to one kilogram of the plant material in their shops. In contrast, wholesalers who export the plant internationally, hire professional harvesters who travel across the geographical range of the species to collect plant material. Harvested roots are brought directly from the wild to the export companies in Rabat, Casablanca and Tangier, from where they enter the international market, including supply of material to India. According to informants from export companies, between 3-10 tons of the plant product can be stocked at a time.

Our examination of material in trade involved screening a total of 66 bags each containing an average of 25g of dry roots. Initial morphological examination of these samples identified obvious non-*Anacyclus* adulterants in 39/66 batches. The adulterants were present with a proportion from 3% to 100% with an average of 42%. The non-*Anacyclus* adulterants were found at high frequency in collections from traditional healers and herbalists, less so from collectors, wholesalers and export companies (Figures 4-5, Table S6).

We selected 110 individual roots for DNA analysis from the 66 root batches. Of these 99 had a morphology consistent with *Anacyclus,* and 11 which were classed as similar to *Anacyclus* but likely to be non-*Anacyclus* based on their morphology. We recovered partial plastome assemblies of these 11 non-*Anacylus* roots and using sequence queries against GenBank, we obtained identifications for nine *Plantago* spp., *one Primula* spp. and one *Plumeria* spp. (Figure S13, Table S2).

Of the 99 *Anacyclus* roots, 10 had no identifiable DNA sequences via any of our methods, with 89/89 of the remaining samples identified through target capture (Figure S13). The ten discarded samples presented very fragmented plastome assemblies with average 89% missing data in the nuclear matrices. These samples had a very low DNA integrity.

The plastome sequences enabled identification of seven roots to the species level, with the remainder identified as *Anacyclus* sp. (Figure 1, Table S7). Neither ITS nor any of the standard barcode regions were able to discriminate any of these samples below the genus level (Figure 1, S4-5, S8-11).

The nuclear markers gave much higher resolution within *Anacyclus* (Table S2, S7). In our investigation of plant material traded within Morocco, our analyses of six individual root samples from four rural community collectors identified three *Anacyclus* var. *pyrethrum* and three var. *depressus*. Our analysis of five samples from three wholesaler 'middle-men' in Morocco identified two *Anacyclus* var. *pyrethrum* and three var. *depressus*. Our sequences from 19 samples from 10 herbalists revealed 12 *A. pyrethrum* var. *pyrethrum* and seven var. *depressus*. Likewise, our 11 samples from six traditional healer sources identified seven *A. pyrethrum* var. *pyrethrum*, and four var. *depressus*. For material traded in international markets, our 17 sequenced samples from three export companies in Morocco identified nine *A. pyrethrum* var. *pyrethrum*, six var. *depressus*, and two *A. homogamous*. Our analysis of 30 samples from 17 herbalists in India identified three *A. pyrethrum* var. *pyrethrum* and 27 var. *depressus*.

Geographical source

293 Market samples in Morocco originate both from various Moroccan areas as well as Algeria,
294 and material from all of these populations of origin can be found in Indian market samples
295 (Table S1, Figure 3, 6). Of the 99 non-adulterated roots which we identified to species level
296 using target capture, we were able to associate 67% to a specific geographic region (Figure 1,
297 3, Table S1). Using phylogenetic analysis, root samples of *A. pyrethrum* var. *pyrethrum*
298 clustered with reference material from the High Atlas (Figure 6, case 4), and *A. pyrethrum*
299 var. *depressus* roots clustered with reference material from different regions in Morocco,
300 including the Rif Mountains, the High Atlas and the Middle Atlas (Figure 6, case 2 and 3).
301 Evidence for the international trade from Algeria to Morocco is highlighted by a distinctive
302 clade that includes traded roots collected from west Algeria (Figure 6, case 1). The
303 geographic origin was only resolvable with target capture data; standard barcoding regions,
304 ITS and plastome data lacked variation, resolution or both (Figure 1).

305 **DISCUSSION**

306 This study illustrates the potential for target-capture based DNA barcoding to form the next
307 wave of standard plant DNA-barcoding tools and provide the greatly needed species-level
308 resolution. A key rate limiting step for the standard plant barcodes is that they are
309 fundamentally recovering data from just one or two independent loci (plastid DNA and ITS),
310 which often show trans-specific polymorphism and barcode sharing among related species
311 (Hollingsworth, Graham, & Little, 2011). Even using complete plastid genome sequences
312 suffers from the same problem, as the data are all physically linked in a single non-
313 recombinant uni-parentally inherited locus. Several recent hybridization events have occurred
314 in the *Anacyclus* genus and entire plastid genomes or plastid barcode markers, and/or ITS
315 provide limited resolution below genus level (Figure 1). In contrast, our target capture

316 approach using hundreds of nuclear markers yields significantly higher molecular
317 identification success and more accurate resolution to species and even population level
318 (Figure 1).

319 *Species identification, species in trade, and geographic origins of* Anacyclus

320 These data provide new insights into trade of *A. pyrethrum* and highlight the extent of
321 adulteration and the scarcity of *A. pyrethrum* var. *pyrethrum* (Figures 2-3, 6). Only a small
322 proportion of the tested samples from herbalists and traditional healers were the potent *A.*
323 *pyrethrum* var. *pyrethrum,* with the Indian market in particular dominated by var. *depressus.*
324 In both Morocco and India some individual sellers had entirely or almost entirely adulterated
325 products. Most of the non-*Anacyclus* roots we sequenced were identified as *Plantago* spp. by
326 shotgun sequencing, despite being sampled from six different localities including Morocco
327 and India. As *Plantago* roots are similar in appearance to *Anacyclus,* it is possible that they
328 are deliberately added as a 'difficult to identify' adulterant which may go unnoticed by non-
329 specialists.

330 Our analysis of samples from collectors, wholesalers and export companies detected
331 much less adulteration at this point in the supply chain (Figure 4, Table S6). Collection of *A.*
332 *pyrethrum* var. *pyrethrum* is carried out by professional harvesters employed by export
333 companies who travel across the country and are considered poachers by local communities
334 (Ouarghidi, Powell, Martin, & Abbad, 2017). Local harvesters have increasing difficulty to
335 supply local trade chains (Ouarghidi et al., 2017, 2012), which may finally result in increased
336 adulteration rates in the poorly-governed, national value chains (Figure 4), as has also been
337 observed elsewhere (A. Booker et al., 2012) . Our results also identify previously unreported
338 international trade in North Africa prior to export to the Indian sub-continent (Figure 3, 6).

339 We provide evidence that export companies in Morocco source material not only in this
340 country, but also from neighbouring Algeria (Figure 6). Applying this molecular identification
341 approach enables us to distinguish samples at population level and uncover these hidden
342 international sourcing channels.

343 **Conservation of *A. pyrethrum***

344 High national and international demand for *A. pyrethrum* likely encourages its
345 overharvesting and adulteration. As the plant is a remedy of the Indian pharmacopoeia, its
346 demand is likely to increase along with that of other Ayurvedic medicines (Kala, Dhyani, &
347 Sajwan, 2006). Although *A. pyrethrum* has been assessed to be vulnerable internationally
348 (Rankou et al., 2015) and endangered in Morocco on the IUCN Red List, the plant is not
349 listed in the CITES appendices and its international trade is not regulated. Nonetheless,
350 continued overharvesting is driving wild populations to critical levels and conservation
351 policies are necessary. Common strategies to conserve overharvested medicinal plants often
352 include collection and trade restrictions as well as cultivation (Schippmann et al., 2002).
353 Cultivation is often proposed as a solution to both conservation issues and sourcing high
354 quality, appropriately identified material (Hamilton, 2004; Schippmann, Leaman, &
355 Cunningham, 2006; Schippmann et al., 2002). The cultivation necessitate engagement with
356 local communities that depend on plant harvest, as well as monitoring of professional trading
357 networks. Both promotion of sustainable harvesting for livelihood security as well as
358 restriction of unsustainable professional trade are needed. Only with fine-grained mapping of
359 sourcing areas and supply chains, as our results highlight for *A. pyrethrum*, can meaningful
360 conservation action be achieved. With the implementation of target enrichment, we identified
361 the origin of the harvested populations of medicinal or traded populations and point out where
362 conservation efforts should be firstly implemented. We also reveal a previously

363 undocumented harvesting and trade in Algeria (Figure 6, case 1), and show that *Anacyclus* is
364 harvested at a national scale in Morocco (Figure 6, case 2-4). This study gives scientific
365 evidence to support conservation programs like GDF's High Atlas Cultural Landscapes
366 Programme and potentially attract more attention for future conservation projects for
367 *Anacyclus*.

368

369 *Future prospects for plant DNA barcoding*

370 Key criteria for developing new DNA barcoding approaches include resolving power (telling
371 species apart), recoverability (enabling use on a wide diversity of tissue sources), and cost and
372 efficiency (enabling scaling over very large sample sets).

373 In terms of resolving power, the target capture approach used here offers substantial
374 improvement compared to plant barcodes based on plastid sequences and ITS. The key
375 enabling step is access to multiple nuclear markers, as this reduces sensitivity of the
376 identification to introgression of one or two loci as is the case for barcodes from rDNA or the
377 plastid genome. In this study, we show high resolution from nuclear markers retrieved by
378 target capture for identification of species in a genus that has undergone recent hybridization
379 events (Manzanilla, 2018). The successful recovery of these markers from low quality input
380 DNA is also important and combined these observations make the case that this method
381 provides a viable solution for future plant barcoding. In contrast, shotgun sequencing requires
382 a substantial sequencing effort to retrieve the same suite of nuclear markers for species with
383 large genomes. Thus for *Anacyclus* it would require 14 HiSeq 3000 lanes to obtain the same
384 coverage of the loci we have used here. Overall, and regardless of approach, our findings
385 provide empirical evidence to support predictions made in previous review papers

386 (Hollingsworth et al., 2016) that barcoding based on nuclear markers should outperform
387 standard barcoding methods.

388 The successful recovery of sequences via target capture depends on how closely
389 related the sampled species are to the reference set on which the baits were designed, and/or
390 the level of variation in the loci that form the bait set (McCormack, Tsai, & Faircloth, 2016;
391 Paijmans, Fickel, Courtiol, Hofreiter, & Förster, 2016). In the current study, the baits were
392 designed from *Anacyclus* and related genera. This optimized their specificity to our study
393 group and enabled their successful high-resolution application for assessing trade. A clear
394 challenge for wider use of target capture, is applicability over much greater phylogenetic
395 distances. The recently published universal angiosperm baits (Buddenhagen et al., 2016;
396 Johnson et al., 2018), designed to recover 353 loci from a wide diversity of flowering plants
397 offer great potential here. The angiosperm universal bait kit (Johnson et al., 2018) can provide
398 species and population-level resolution (Van Andel et al., 2019). In addition, there is a need
399 for a more general evaluation of when, how and at what scale to most effectively combine
400 taxon-specific bait sets (as used here) with universal bait sets, to simultaneously obtain very
401 high resolution and sequencing success over wide phylogenetic distances.

402 Another important aspect of recoverability is efficacy with degraded DNA. Drying,
403 storage, and transportation affect the quality of plant material in trade and can cause extensive
404 DNA degradation (Anthony Booker et al., 2014). In consequence, traded samples have similar
405 challenges to working with ancient DNA, herbarium samples or archaeological remains.
406 Target capture is particularly well suited to this challenge (McCormack et al., 2016; Paijmans
407 et al., 2016). Although shotgun sequencing can also be very effective on degraded material
408 (Alsos et al., 2020; Bakker, 2017; Zeng et al., 2018), our recovery rate in this study was
409 greater for target capture than shotgun sequencing (Figure 1), and we recovered data from

410 100% of the samples used for establishing the reference library and over 70% of the samples
411 in trade. This 70% success rate for hundreds of nuclear markers providing high resolution
412 from suboptimal tissue of traded samples is noteworthy. The other mainstream approach for
413 highly degraded DNA is a portion of the chloroplast *trnL* (UAA) intron, specifically the P6
414 loop (10–143 bp)(Taberlet et al., 2007). This has been highly successful in recovering
415 sequence data from degraded samples (Parducci et al., 2012; Willerslev et al., 2014).
416 However, this short region of the plastid genome has a low variation at the species level and
417 does not typically discriminate among con-generic species (Taberlet et al., 2007).

418 The target capture methodology presented here, including library construction and
419 sequencing, cost 70 USD per sample in 2016, a price similar to those presented by Hale et al.
420 (2020). With the optimisations suggested in Hale et al. (2020), prices drop to 22 USD per
421 sample. Thus, this target enrichment approach is quickly becoming affordable for large-scale
422 biomonitoring projects. With a growing interest and investment in DNA-based identification
423 solutions in medicine and industry (Afshinnekoo et al., 2017; Menegon et al., 2017), ongoing
424 work is expected to continue to optimize protocols and drive these costs down even further,
425 reaching those of standard barcoding approaches (Hebert et al., 2018).

426 **CONCLUSION**

427 In plants, the frequent sharing of plastid and ribosomal sequences among con-generic
428 species, coupled with the difficulty of routinely accessing multiple nuclear markers, has acted
429 as a constraint on the resolution of DNA barcoding approaches. Current advances in
430 sequencing technology and bioinformatics are removing this constraint and offer the potential
431 for a new wave of high-resolution identification tools for plants. These approaches, such as

target capture have the capability to distinguish species and populations, providing insights into diversity and ecology, as well as the multitude of societal applications which require information on the identification and provenance of biological materials.

436 **REFERENCES**

437 Adams, M., Alther, W., Kessler, M., Kluge, M., & Hamburger, M. (2011). Malaria in the Renaissance:
438 remedies from European herbals from the 16th and 17th century. *Journal of*
439 *Ethnopharmacology*, *133*(2), 278–288.
440 Afshinnekoo, E., Chou, C., Alexander, N., Ahsanuddin, S., Schuetz, A. N., & Mason, C. E. (2017).
441 Precision metagenomics: rapid metagenomic analyses for infectious disease diagnostics and
442 public health surveillance. *Journal of Biomolecular Techniques: JBT*, *28*(1), 40.
443 Alam, G., & Belt, J. (2009). Developing a medicinal plant value chain: Lessons from an initiative to
444 cultivate Kutki (*Picrorhiza kurrooa*) in Northern India. *KIT Working Papers Series*, (WPS.
445 C5).
446 Bakker, F. T. (2017). Herbarium genomics: skimming and plastomics from archival specimens.
447 *Webbia*, *72*(1), 35–45.
448 Benarba, B. (2016). Medicinal plants used by traditional healers from South-West Algeria: An
449 ethnobotanical study. *Journal of Intercultural Ethnopharmacology*, *5*(4), 320.
450 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
451 sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170
452 Booker, A., Johnston, D., & Heinrich, M. (2012). Value chains of herbal medicines-Research needs
453 and key challenges in the context of ethnopharmacology. *Journal of Ethnopharmacology*,
454 *140*(3), 624–633. doi: 10.1016/j.jep.2012.01.039
455 Booker, Anthony, Frommenwiler, D., Johnston, D., Umealajekwu, C., Reich, E., & Heinrich, M.
456 (2014). Chemical variability along the value chains of turmeric (*Curcuma longa*): a
457 comparison of nuclear magnetic resonance spectroscopy and high performance thin layer
458 chromatography. *Journal of Ethnopharmacology*, *152*(2), 292–301.
459 Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., … Kortyna,
460 M. (2016). Anchored phylogenomics of angiosperms I: Assessing the robustness of
461 phylogenetic estimates. *BioRxiv*, 086298.
462 de Boer, H. J., Ouarghidi, A., Martin, G., Abbad, A., & Kool, A. (2014). DNA barcoding reveals
463 limited accuracy of identifications based on folk taxonomy. *PLoS ONE*, *9*(1), e84291. doi:
464 10.1371/journal.pone.0084291
465 De Vos, P. (2010). European materia medica in historical texts: longevity of a tradition and
466 implications for future use. *Journal of Ethnopharmacology*, *132*(1), 28–47.
467 Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the
468 multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340.
469 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.

Nucleic Acids Research, *32*(5), 1792–1797.

Esquinas-Alcazar, J. (2004). International treaty on plant genetic resources for food and agriculture. *Plant Genetic Resources Newsletter*.

Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C., … Husband, B. C. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, *9*(s1), 130–139.

Garcia, S., Hidalgo, O., Jakovljević, I., Siljak-Yakovlev, S., Vigo, J., Garnatje, T., & Vallès, J. (2013). New data on genome size in 128 Asteraceae species and subspecies, with first assessments for 40 genera, 3 tribes and 2 subfamilies. *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology*, *147*(4), 1219–1227. doi: 10.1080/11263504.2013.863811

Garnatje, T., Canela, M. A., Garcia, S., Hidalgo, O., Pellicer, J., Sánchez-Jiménez, I., … Vallès, J. (2011). *GSAD: A genome size in the Asteraceae database*.

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129–e129.

Hamilton, A. (2004). Medicinal plants, conservation and livelihoods. *Biodiversity & Conservation*, *13*(8), 1477–1517. doi: 10.1023/B:BIOC.0000021333.23413.42

Hebert, P. D., Braukmann, T. W., Prosser, S. W., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., … Sones, J. E. (2018). A Sequel to Sanger: amplicon sequencing that scales. *BMC Genomics*, *19*(1), 219.

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *108*(49), 19451–19452. doi: 10.1073/pnas.1116812108

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLoS ONE*, *6*(5), e19254. doi: 10.1371/journal.pone.0019254

Hollingsworth, P. M., Li, D.-Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: from barcodes to genomes. *Phil. Trans. R. Soc. B*, *371*(1702), 20150338.

Humphries, C. J. (1979). A revision of the genus *Anacyclus* L. (Compositae: Anthemideae). *Bulletin of the British Museum (Natural History), Historical Series*, (7), 83–142.

ISE Code of Ethics Online. (2018, March). Retrieved March 3, 2018, from International Society of Ethnobiology website: http://www.ethnobiology.net/what-we-do/core-programs/ise-ethics-program/code-of-ethics/code-in-english/

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., … Kim, J. T. (2018). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*, *68*(4), 594–606.

Kala, C. P., Dhyani, P. P., & Sajwan, B. S. (2006). Developing the medicinal plants sector in northern

India: challenges and opportunities. *Journal of Ethnobiology and Ethnomedicine*, *2*(1), 32.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Leclerc, L. (1877). *Traité des simples* (Vol. 1). Imprimerie nationale.

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 99–121.

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., & Geschke, A. (2012). International trade drives biodiversity threats in developing nations. *Nature*, *486*(7401), 109.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N. (2009). Genome project data processing S. The sequence alignment/Map format and SAMtools. *Bioinformatics*, *25*. doi: 10.1093/bioinformatics/btp352

Li, Heng, & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., … Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111.

Manouze, H., Bouchatta, O., Gadhi, A. C., Bennis, M., Sokar, Z., & Ba-M'hamed, S. (2017). Anti-inflammatory, antinociceptive, and antioxidant activities of methanol and aqueous extracts of Anacyclus pyrethrum roots. *Frontiers in Pharmacology*, *8*, 598.

Manzanilla, V., Kool, A., Nguyen Nhat, L., Nong Van, H., Le Thi Thu, H., & de Boer, H. J. (2018). Phylogenomics and barcoding of Panax: toward the identification of ginseng species. *BMC Evolutionary Biology*, *18*(1), 44. doi: 10.1186/s12862-018-1160-y

McCormack, J. E., Tsai, W. L., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, *16*(5), 1189–1203.

Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., … Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS One*, *12*(10).

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448. doi: 10.1101/pdb.prot5448

Newmaster, S., Grguric, M., Shanmughanandhan, D., Ramalingam, S., & Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine*, *11*(1), 222. (doi:10.1186/1741-7015-11-222).

Ouarghidi, A., Martin, G., Powell, B., Esser, G., & Abbad, A. (2013). Botanical identification of medicinal roots collected and traded in Morocco and comparison to the existing literature.

Journal of Ethnobiology and Ethnomedicine, *9*(1), 59. (doi:10.1186/1746-4269-9-59).

Ouarghidi, A., Powell, B., Martin, G. J., & Abbad, A. (2017). Traditional Sustainable Harvesting Knowledge and Distribution of a Vulnerable Wild Medicinal Root (*A. pyrethrum* var. *pyrethrum*) in Ait M'hamed Valley, Morocco. *Economic Botany*, *71*(1), 83–95.

Ouarghidi, A., Powell, B., Martin, G. J., De Boer, H., & Abbad, A. (2012). Species substitution in medicinal roots and possible implications for toxicity of herbal remedies in Morocco. *Economic Botany*, *66*(4), 370–382.

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

Paijmans, J. L., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*(1), 42–55.

Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., … Willerslev, E. (2012). Glacial Survival of Boreal Trees in Northern Scandinavia. *Science*, *335*(6072), 1083–1086. doi: 10.1126/science.1216043

Parks, M., Cronn, R., & Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, *7*(1), 84.

Pillon, Y., Johansen, J., Sakishima, T., Chamala, S., Barbazuk, W. B., Roalson, E. H., … Stacy, E. A. (2013). Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary Biology*, *13*(1), 35.

Pittle, K. D. (2005). *Continuity and Change in Islamic Ethnopharmacological Practice: New Methods for Cognitive Dialectometry*. Florida State University.

Rankou, H., Ouhammou, A., Taleb, M., Manzanilla, V., & Martin, G. (2015). *Anacyclus pyrethrum*. *The IUCN Red List of Threatened Species 2015: E.T202924A53798702*, *2015*(2015). Retrieved from http://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T202924A53798702.en

Rosato, M., Álvarez, I., Feliner, G. N., & Rosselló, J. A. (2017). High and uneven levels of 45S rDNA site-number variation across wild populations of a diploid plant genus (*Anacyclus*, Asteraceae). *PloS One*, *12*(10), e0187131.

Särkinen, T., Staats, M., Richardson, J. E., Cowan, R. S., & Bakker, F. T. (2012). How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PloS One*, *7*(8), e43808.

Schippmann, U., Leaman, D., & Cunningham, A. (2006). A comparison of cultivation and wild collection of medicinal and aromatic plants under sustainability aspects. *Frontis*, 75–95.

Schippmann, U., Leaman, D. J., & Cunningham, A. (2002). Impact of cultivation and gathering of

medicinal plants on biodiversity: global trends and issues. *Biodiversity and the Ecosystem Approach in Agriculture, Forestry and Fisheries*.

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S. C. K., … Suda, J. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources*, *16*(5), 1124–1135. doi: 10.1111/1755-0998.12487

Soltis, P. S., & Soltis, D. E. (2009). The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology*, *60*(1), 561–588. doi: 10.1146/annurev.arplant.043008.092039

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14. doi: 10.1093/nar/gkl938

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564–577.

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

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

Veldman, S., Gravendeel, B., Otieno, J. N., Lammers, Y., Duijm, E., Nieman, A., … van Andel, T. R. (2017). High-throughput sequencing of African chikanda cake highlights conservation challenges in orchids. *Biodiversity and Conservation*, *26*(9), 2029–2046.

WHO. (2004). *WHO guidelines on safety monitoring of herbal medicines in pharmacovigilance systems*.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486), 47–51. doi: 10.1038/nature12921

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A*, *106*. doi: 10.1073/pnas.0811575106

Wysoker, A., Tibbetts, K., & Fennell, T. (2015). *Picard tools*. Broad Institute.

Zeng, C.-X., Hollingsworth, P. M., Yang, J., He, Z.-S., Zhang, Z.-R., Li, D.-Z., & Yang, J.-B. (2018). Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods*, *14*(1), 43.

Zhang, C., Sayyari, E., & Mirarab, S. (2017). *ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches*. 53–75. Springer.

614 **DECLARATIONS**

615 **Authors' contributions**

616 The project was coordinated by AK, GM, HdB and VM. VM did the design of the study and
617 performed data analysis. AK, HdB, ITT, PH and VM wrote the manuscript. All authors
618 provided useful contributions to data analysis and interpretation of the results. All authors
619 have read and approved the final version of the manuscript.

620 **Acknowledgments**

621 Maxime Borry for the DNA extraction; Julie Hawkins for hosting V.M. in Reading during the
622 phylogenomics analyses; the Global Diversity Foundation team in Morocco for their logistic
623 support during the field work. The PET group for their useful discussions and
624 recommendations during the project. This work was performed on the Abel Cluster, owned by
625 the University of Oslo and the Norwegian metacenter for High Performance Computing
626 (NOTUR), and operated by the Department for Research Computing at USIT, the University
627 of Oslo IT-department, http://www.hpc.uio.no/.

628 **Reproducibility**

629 For reproducibility purposes, all the scripts used during the data processing are available on
630 the OSF work repository https://osf.io/9bh3p/. New sequencing data have been deposited
631 under a single NCBI BioProject accession PRJNA631886.

Funding

This project was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under the Grant agreement no. 606895 to the FP7-MCA-ITN MedPlant, "Phylogenetic Exploration of Medicinal Plant Diversity".

Figure 1: Sequencing recovery and identification success for the traded samples for each dataset. The figure shows the percentage of samples for which useful marker sequences were successfully retrieved for molecular identification for standard barcodes, ITS, plastomes and nuclear markers. No data recovered (NDR) is used to indicate samples for which no sequence data was recovered or where no identification at the genus level or below could be made. For the samples that produced useable data, the proportion of samples that resulted in identification at the genus, species and population levels is given.

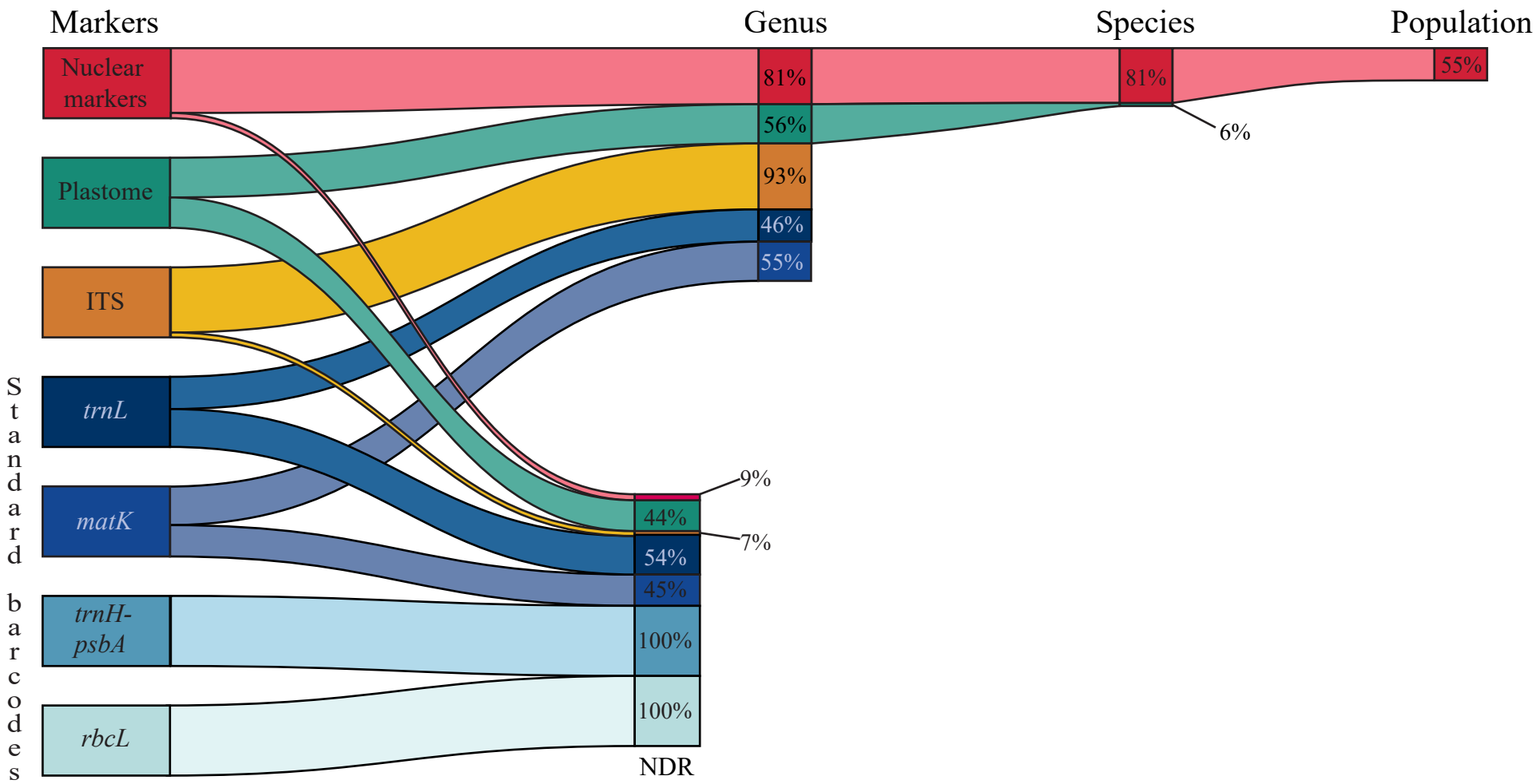
Figure 2: Box-and-whisker plots showing retrieved coverage of 443 targeted markers for each sample. Identified and non-identified samples are colour coded.

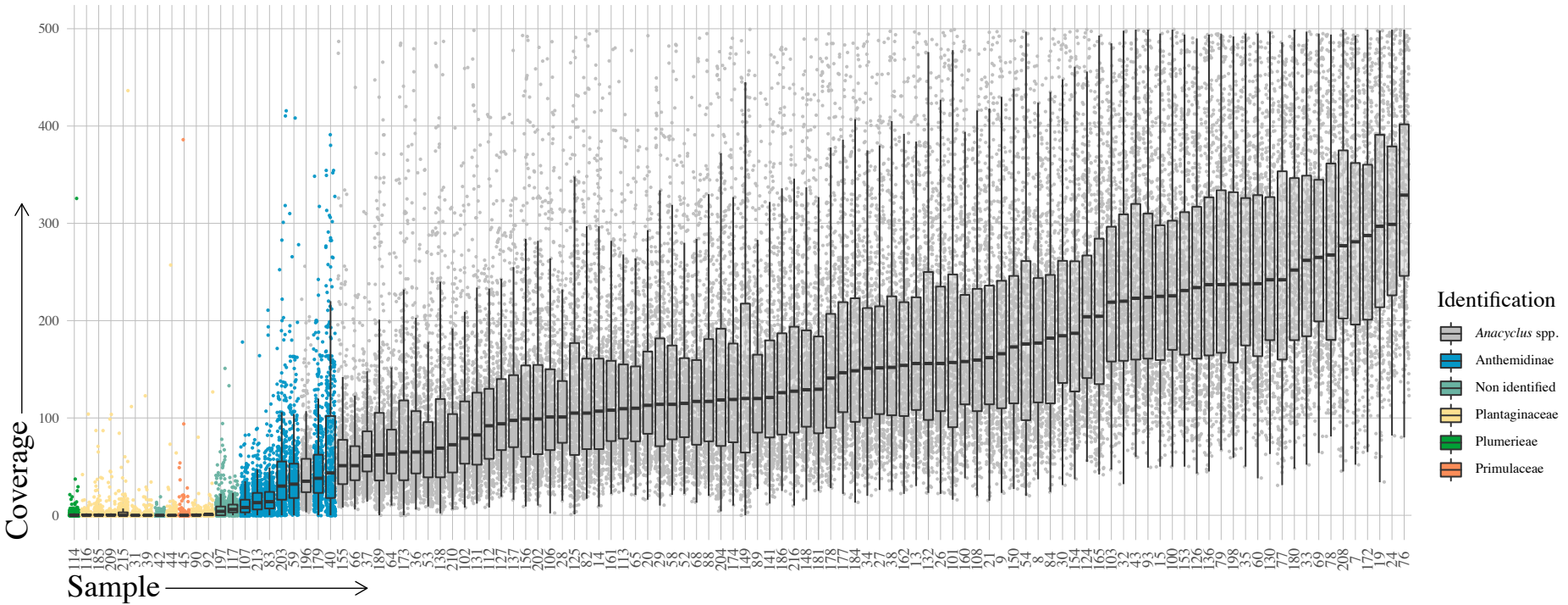
Figure 3: Multispecies coalescent phylogenetic tree of the nuclear loci dataset. The reference dataset includes taxon labels and herbarium accession numbers and the traded samples are numbered. Supported clades with an associated geographic origin are indicated by coloured bars.

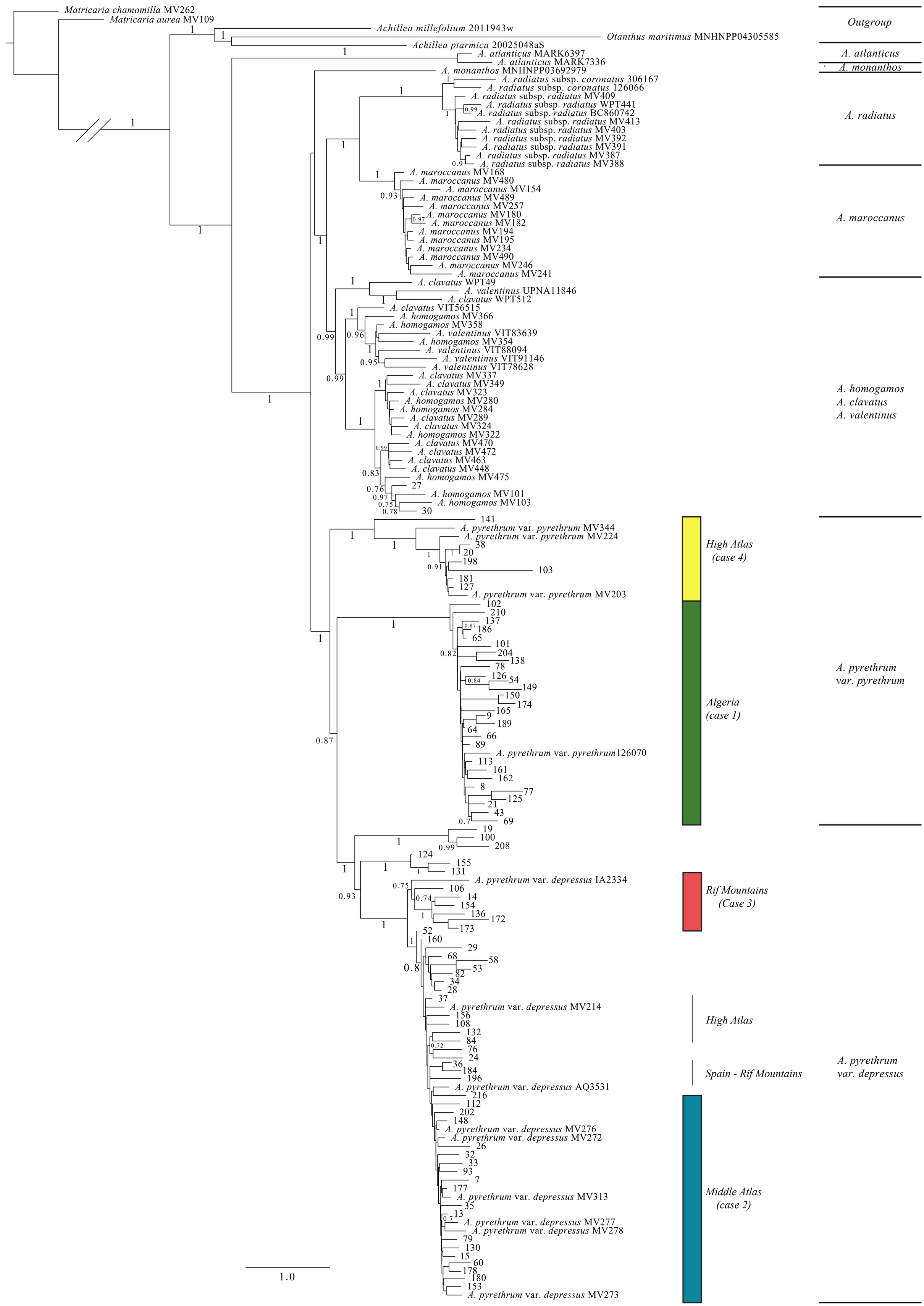
Figure 4: National and international supply chains of *A. pyrethrum*. Pie charts represent the proportion of *A. pyrethrum* (light and dark blue represent var. *depressus* and var. *pyrethrum* respectively) and adulterated samples (orange and brown for *A. homogamos* and other adulterants) by each stakeholder. We were unable to obtain samples from wholesalers/middlemen in India or professional collectors in Morocco (indicated by square boxes).

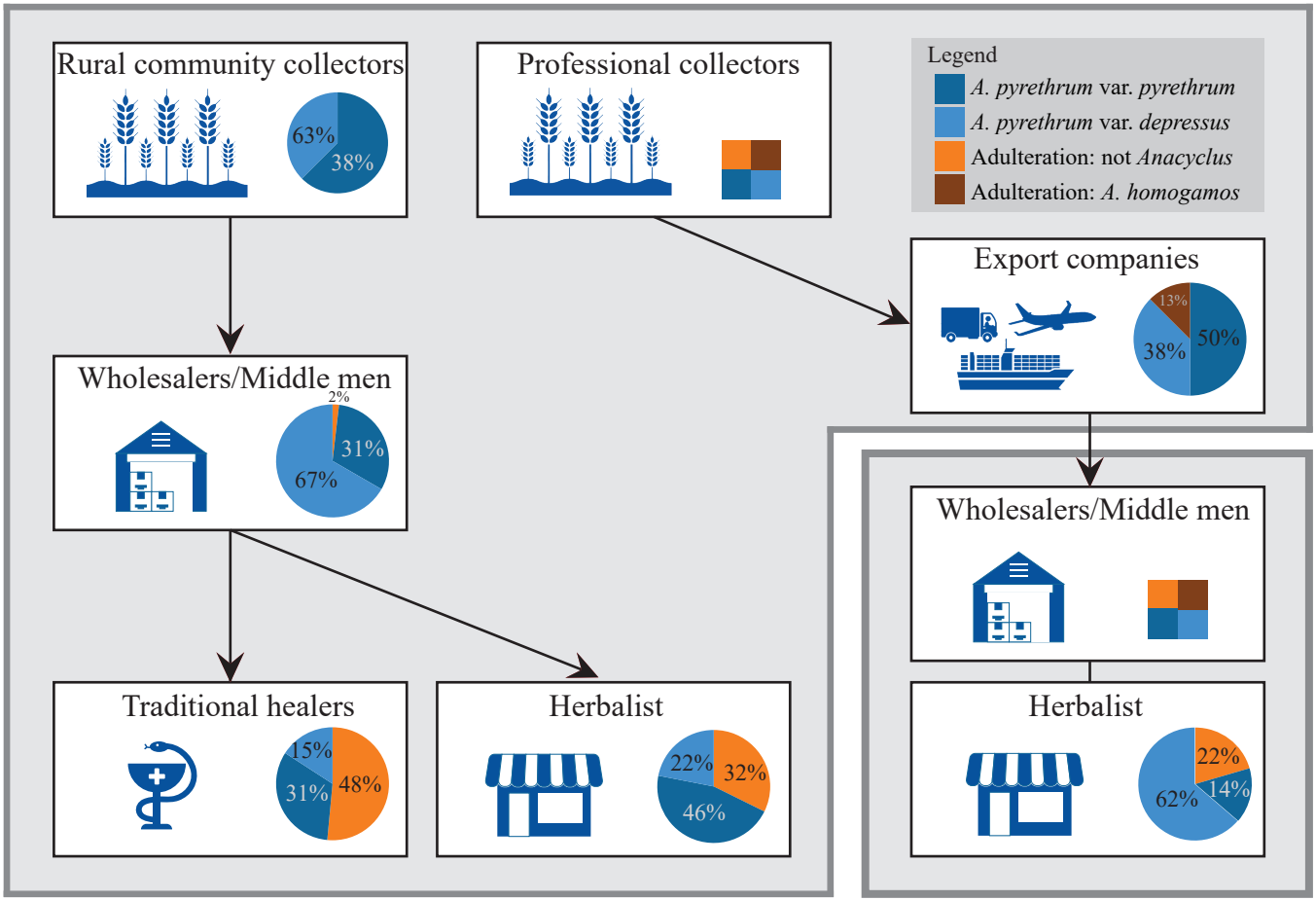
Figure 5: Species identification of market samples. Sample locations are shown with coloured circles according to the type of stakeholder. A pie-chart with the proportions of adulteration and identified species is represented for each location in (a) Morocco (native range) and (b) India (exported material).

Figure 6: Map of the origin of the traded samples. The figure relates four supported clades mentioned in the Figure 3. The stars indicate reference samples and the dots the traded samples. The number in the circle or the start corresponds to the number of individuals.

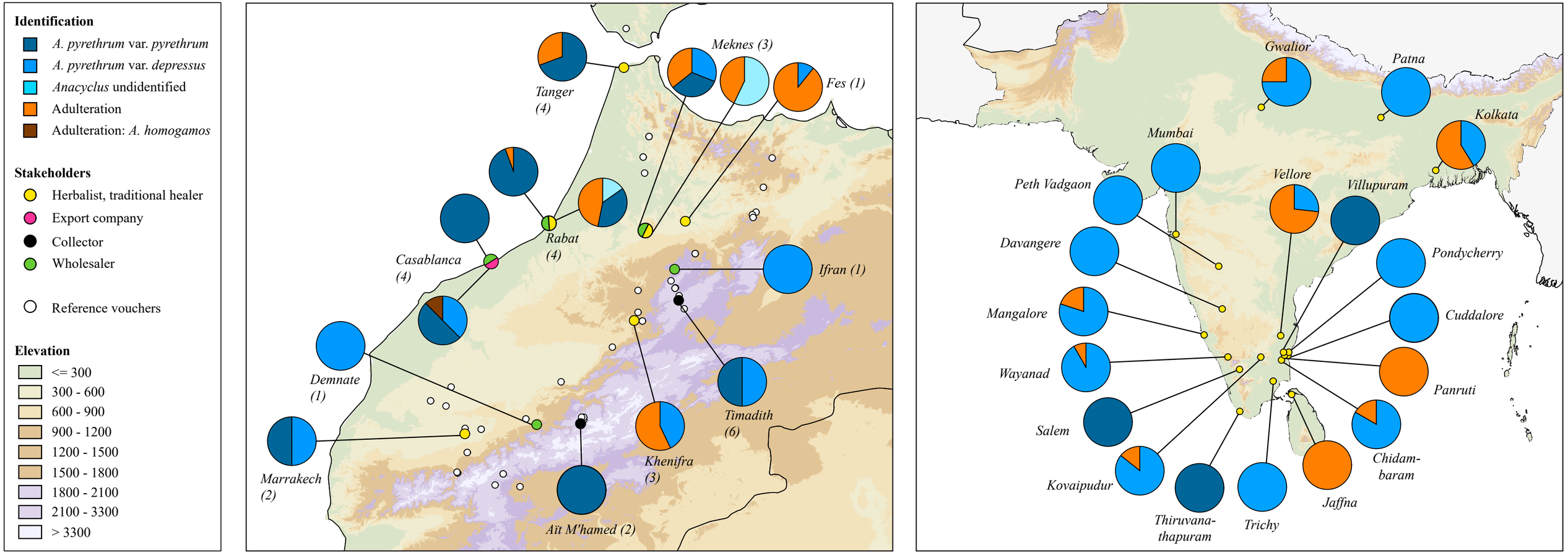

Markers
Genus
Species
Population
Nuclear markers
Plastome
ITS
trnL
matK
trnH-psbA
rbcL
Standard barcodes
81%
56%
93%
46%
55%
44%
54%
45%
100%
100%
NDR
9%
7%
81%
6%
55%


Coverage
500
400
300
200
100
0
Sample
Identification
Anacyclus spp.
Anthemidinae
Non identified
Plantaginaceae
Plumerieae
Primulaceae

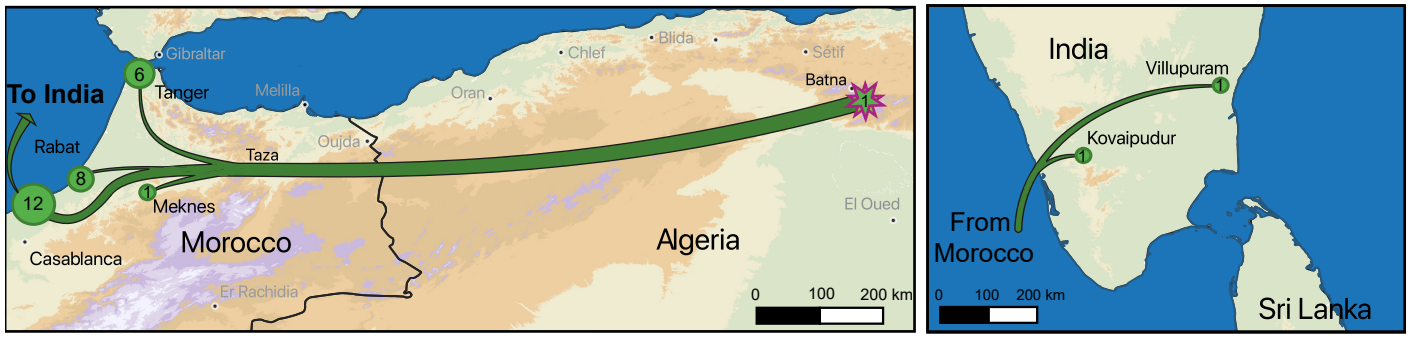

Matricaria chamomilla MV262
Matricaria aurea MV109
Achillea millefolium 2011943w
Otanthus maritimus MNHNPP04305585
Achillea ptarmica 20025048aS
A. atlanticus MARK6397
A. atlanticus MARK7336
A. monanthos MNHNPP03692979
A. radiatus subsp. coronatus 306167
A. radiatus subsp. coronatus 126066
A. radiatus subsp. radiatus MV409
A. radiatus subsp. radiatus WPT441
A. radiatus subsp. radiatus BC860742
A. radiatus subsp. radiatus MV413
A. radiatus subsp. radiatus MV403
A. radiatus subsp. radiatus MV392
A. radiatus subsp. radiatus MV391
A. radiatus subsp. radiatus MV387
A. radiatus subsp. radiatus MV388
A. maroccanus MV168
A. maroccanus MV480
A. maroccanus MV154
A. maroccanus MV489
A. maroccanus MV257
A. maroccanus MV180
A. maroccanus MV182
A. maroccanus MV194
A. maroccanus MV195
A. maroccanus MV234
A. maroccanus MV490
A. maroccanus MV246
A. maroccanus MV241
A. clavatus WPT49
A. valentinus UPNA11846
A. clavatus WPT512
A. clavatus VIT56515
A. homogamos MV366
A. homogamos MV358
A. valentinus VIT83639
A. homogamos MV354
A. valentinus VIT88094
A. valentinus VIT91146
A. valentinus VIT78628
A. clavatus MV337
A. clavatus MV349
A. clavatus MV323
A. homogamos MV280
A. homogamos MV284
A. clavatus MV289
A. clavatus MV324
A. homogamos MV322
A. clavatus MV470
A. clavatus MV472
A. clavatus MV463
A. clavatus MV448
A. homogamos MV475
27
A. homogamos MV101
A. homogamos MV103
30
141
A. pyrethrum var. pyrethrum MV344
A. pyrethrum var. pyrethrum MV224
38
20
198
103
181
127
A. pyrethrum var. pyrethrum MV203
102
210
137
186
65
101
204
138
78
126
54
149
150
174
165
9
189
64
66
89
A. pyrethrum var. pyrethrum126070
113
161
162
8
77
125
21
43
69
19
100
208
124
155
131
A. pyrethrum var. depressus IA2334
106
14
154
136
172
173
52
160
29
68
58
53
82
34
28
37
A. pyrethrum var. depressus MV214
156
108
132
84
76
24
36
184
196
A. pyrethrum var. depressus AQ3531
216
112
202
148
A. pyrethrum var. depressus MV276
A. pyrethrum var. depressus MV272
26
32
33
93
7
177
A. pyrethrum var. depressus MV313
35
13
A. pyrethrum var. depressus MV277
A. pyrethrum var. depressus MV278
79
130
15
60
178
180
153
A. pyrethrum var. depressus MV273
1.0
Outgroup
A. atlanticus
A. monanthos
A. radiatus
A. maroccanus
A. homogamos
A. clavatus
A. valentinus
A. pyrethrum var. pyrethrum
A. pyrethrum var. depressus
High Atlas (case 4)
Algeria (case 1)
Rif Mountains (Case 3)
High Atlas
Spain - Rif Mountains
Middle Atlas (case 2)



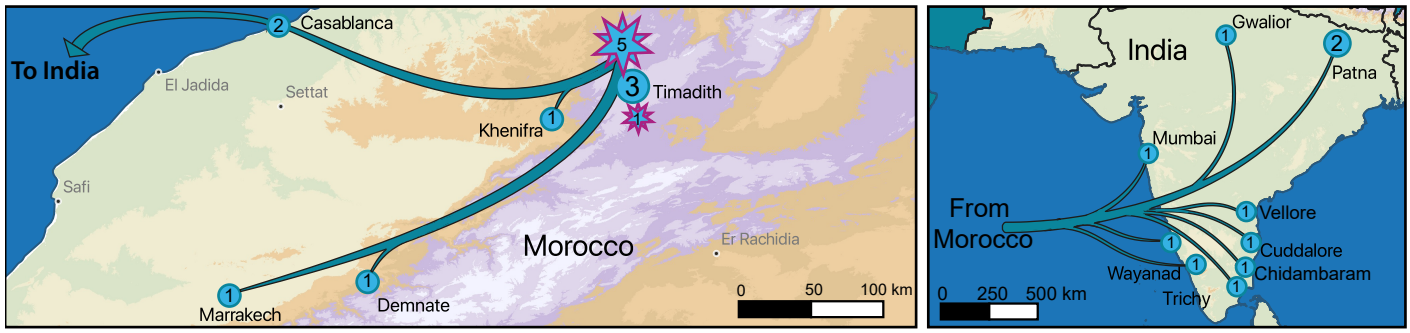
Moroccan national market **International market**


Identification
A. pyrethrum var. pyrethrum
A. pyrethrum var. depressus
Anacyclus undidentified
Adulteration
Adulteration: A. homogamos
Stakeholders
Herbalist, traditional healer
Export company
Collector
Wholesaler
Reference vouchers
Elevation
<= 300
300 - 600
600 - 900
900 - 1200
1200 - 1500
1500 - 1800
1800 - 2100
2100 - 3300
> 3300
Tanger (4)
Meknes (3)
Fes (1)
Rabat (4)
Casablanca (4)
Ifran (1)
Demnate (1)
Timadith (6)
Khenifra (3)
Marrakech (2)
Aït M'hamed (2)
Gwalior
Patna
Kolkata
Mumbai
Peth Vadgaon
Vellore
Villupuram
Davangere
Pondycherry
Mangalore
Cuddalore
Wayanad
Panruti
Salem
Chidam-baram
Kovaipudur
Jaffna
Thiruvana-thapuram
Trichy

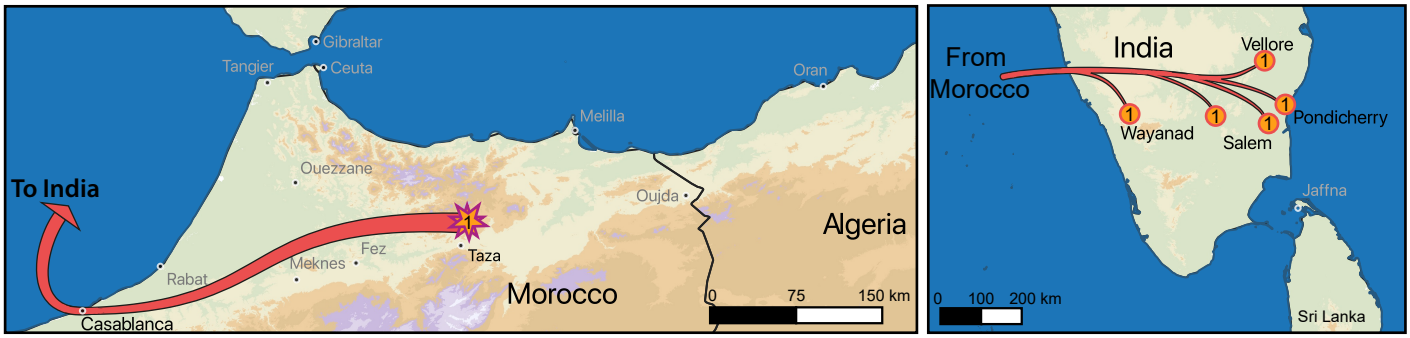
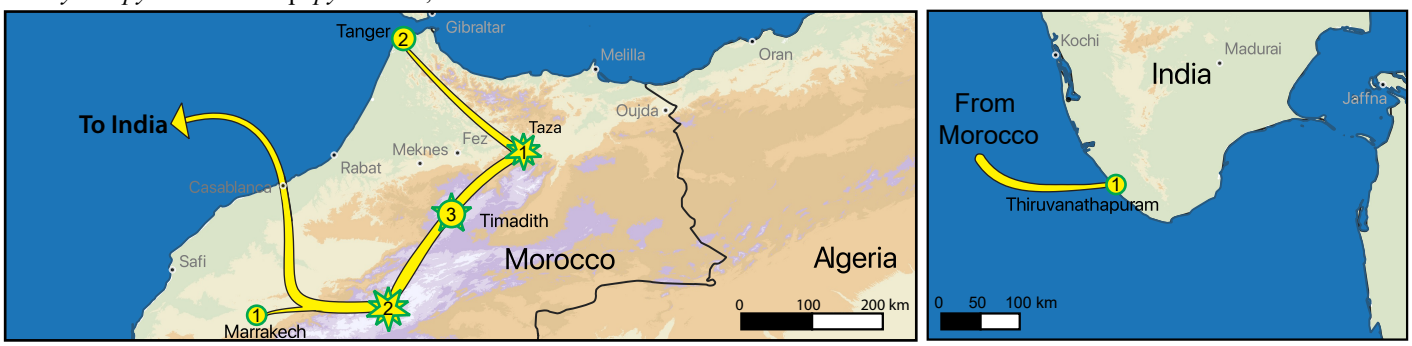
Anacyclus pyrethrum subsp. *pyrethrum*, case 1



Anacyclus pyrethrum subsp. *depressus*, case 2



Anacyclus pyrethrum subsp. *depressus*, case 3



Anacyclus pyrethrum subsp. *pyrethrum*, case 4