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

~~GENOMIC PLANT BARCODING CAN ADDRESS MAJOR CONSERVATION CHALLENGES: POPULATION-LEVEL TRACKING OF A HERBAL MEDICINE'S IN A GLOBAL SUPPLY CHAIN~~

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~~markers~~ 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.

~~Uncontrolled and unsustainable trade in natural resources is an increasingly important threat to global biodiversity. In recent years, molecular identification methods have been proposed as tools to monitor global supply chains, to support regulation and legislative protection of species in trade and enhancing consumer protection by establishing whether a traded product contains the species it is purported to contain. Here, we demonstrate the efficacy of a target-capture genomic DNA barcoding method to establish the identity and geographic origin of samples traded as the red-listed medicinal plant *Anacyclus pyrethrum*.~~

~~DNA-based Samples were identified through by,sequencing, bytarget captureand genome skimmingnumbers of .~Here, we demonstrate the efficacy of a target-capture genomic DNA barcoding method to establish the identity and geographic origin of samples traded as the red-listed medicinal plant *Anacyclus pyrethrum*. This approach is used to unveil product adulteration and substitution in this herbal medicine's national and international supply chains. The captureapproach largely outperforms standard plant DNA barcodes and entire plastid genome sequences.In spite of 'lyout of 110 samples analysed, any of the,ions from other genera and families species from were traded to Indian, Sri Lankan and Moroccan marketsThis aapproach offers the potential to achieve routine species-level DNA-based identification including insights intoinformative of geographic species origins. This represents a major development for biodiversity conservation, including the regulation and monitoring of trade in natural plant products.~~

Keywords: Anacyclus, DNA barcoding,~~, endangered species,~~ genomic barcoding, international trade, supply chain~~, sustainability~~, target capture

7 4

60

8 4

Human exploitation of ~~natural~~ 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 ~~lead to~~create markets in which~~where~~ regulation and authentication ~~is~~ are extremely difficult (Newmaster, Grguric, Shanmughanandhan, Ramalingam, & Ragupathy, 2013; WHO, 2004).

Many traded plant products from wild populations are over-harvested.~~:~~ I~~i~~ncreasing 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 ~~enforce~~ 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 ~~products~~ingredients. ~~In addition, t~~The design, implementation and enforcement of successful conservation actions as well as ~~, and/or~~ assessments of product authenticity and quality ~~and authenticity,~~ 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 ~~and identify~~ traded products ~~to their geographical area of origin~~ are still in their infancy.

DNA barcoding has quickly gained popularity since Hebert et al. (2003) first advocated the use of short and variable DNA sequences, amplified using universal primers, for species identification and discovery of new taxa. DNA barcoding is highly effective for species-level identification in animals using a portion of the mitochondrial marker Cytochrome Oxidase 1 (COI) (P. D. Hebert, Hollingsworth, & Hajibabaei, 2016; P. D. Hebert, Ratnasingham, et al., 2016). In plants, standard DNA barcoding involves one to four plastid DNA regions (*rbcL, matK, trnH-psbA, trnL*), sometimes in combination with internal transcribed spacers of nuclear ribosomal DNA (nrDNA, ITS) (CBOL et al., 2009; Kress, 2017). Although these markers are very informative in many cases, no single marker or combination of markers routinely provide complete species-level resolution, especially in species-rich groups~~(Hollingsworth, 2011)~~, let alone population-level assignment. ~~DNA barcoding is highly effective for species-level identification in animals using a portion of the mitochondrial marker Cytochome Oxidase 1 (COI)(Fazekas et al., 2009). DNA barcoding~~ has ~~quickly gained popularity from 2003 as it advocated the use of short and variable DNA sequences that can be amplified using universal primers for species identification and discovery. In plants, standard DNA barcoding involves~~ using varying combinations of ~~one to four plastid DNA regions (*rbcL, matK, trnH-psbA, trnL*) and/or the internal transcribed spacers of nuclear ribosomal DNA (nrDNA, ITS). Although these markers are very informative in many cases, no single marker or combination of these markers routinely provides complete species-level resolution, especially in species-rich groups(Hollingsworth, 2011), let alone provide population level assignment.~~

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

~~The development of high throughput sequencing (HTS) with new sequencing chemistries and platforms removes the necessity of targeting short and universal barcodes, and offers opportunities to extend the concept of DNA barcoding in plants (Hollingsworth, Li, van der Bank, & Twyford, 2016; Lemmon & Lemmon, 2013). In addition to sequencing standard barcoding loci in a more cost-effective fashion, two major approaches have been proposed for increasing the resolution (and coverage) of plant DNA barcoding.~~ Shallow pass shotgun sequencing (commonly referred as genome skimming) is ~~now frequently~~ used to recover organellar genomes and nuclear ribosomal DNA sequences, increasing the amount of data per sample~~,~~ and leading to some increases in resolution (Manzanilla et al., 2018; Parks, Cronn, & Liston, 2009). Although ~~, and~~ workflows and bioinformatic pipelines are ~~becoming~~ increasingly refined for this approach~~. However, at present~~, current cost constraints mean that most genome skimming barcoding projects only have sufficient sequencing depth to generate ~~recover~~ comparative data ~~from multiple samples~~ for multi-copy regions such as plastid genomes and ribosomal DNA~~,~~; ~~these~~ These regions represent a limited number of independent loci, ultimately constraining resolving power (Soltis & Soltis, 2009; Wood et al., 2009).

Target capture sequencing offers the potential to overcome these deficiencies~~this,~~ by efficiently targeting hundreds of low-copy nuclear markers, providing access to a much greater number of independent data points per unit of sequencing effort (Degnan &

Rosenberg, 2009). Similar to ~~Like~~ genome skimming, target capture is successful in sequencing samples with poor DNA integrity (Brewer et al., 2019; Forrest et al., 2019)~~,~~ and allows sequencing hundreds of samples at the same time (Mamanova et al., 2010). It can also be designed to recover standard DNA barcodes in the same assay (Schmickl et al., 2016). Although target capture has been advocated as a powerful tool for molecular identification of plants (Pillon et al., 2013), its usefulness for determining and tracing the origin of plants in trade remaineds untested~~utility in a barcoding context remains untested to date for plants in trade (Pillon et al., 2013)~~.

We evaluated the power of target-capture DNA barcoding by ~~with an~~ investigati~~o~~ng ~~of~~ the traceability of plant products~~,~~ reputedly derived from ~~focusing on an internationally traded and~~a vulnerable medicinal plant species, *Anacyclus pyrethrum* (L.) Lag.~~,~~ In addition to having a well-established international trade chain ~~widely used in traditional Arabic and Ayurvedic medicine. This exemplar case includes a well-established international trade chain~~ (Rankou, Ouhammou, Taleb, Manzanilla, & Martin, 2015), this species presents ~~, and~~ classic challenges for plant molecular identification such as ~~a~~ recent radiation, frequent hybridization (Humphries, 1979) and ~~a~~ large genome size (Garcia et al., 2013).

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

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

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

~~Anacyclus pyrethrum „is endemic to Morocco, Algeria, and southern Spain (Humphries, 1979; Rosato et al., 2017), and has a long history of use in Islamo-Arabic, European, and Indian Ayurvedic medicine (Adams, Alther, Kessler, Kluge, & Hamburger,~~

2011; De Vos, 2010; Pittle, 2005). considered In the 13th century, Ibn al-Baytār wrote that the plant was "known across the world" and traded from the Maghreb to all other areas (Leclerc, 1877). Its popularity as a medicinal plant stems from the many pharmacological activities of its roots (Manouze et al., 2017). In Morocco, two varieties of A. pyrethrum are distinguished, var. pyrethrum and var. depressus, the first being more potent and up to ten times more expensive than the second (Ouarghidi, Powell, Martin, De Boer, & Abbad, 2012). Today, both varieties are harvested from the wild and used extensively for the treatment of pain and inflammatory disorders across Morocco (Ouarghidi, Martin, Powell, Esser, & Abbad, 2013; Ouarghidi et al., 2012) and Algeria (Benarba, 2016; Ouelbani, Bensari, Mouas, & Khelifi, 2016), as well as the Middle East (Pittle, 2005) and the Indian sub-continent (Tiwari, Poudel, & Uprety, 2004). A. pyrethrum is still traded today from the Maghreb to India (Ved & Goraya, 2007) and Nepal (Tiwari et al., 2004), and is known to be over-harvested and is increasingly difficult to find in local markets in Morocco (Ouarghidi et al., 2013, 2012).

Collectors are proficient in identifying the plant and its two varieties, but material is could be possibly misidentified and adulterated along the chain of commercialisation (de Boer, Ouarghidi, Martin, Abbad, & Kool, 2014; Ouarghidi et al., 2013, 2012).

Here weWe appliedy target-capture genomic barcoding to distinguish *Anacylus* species and geographical races in , to identify traded *Anacylus* root samples in the national and international supply chains., and We compare this novel approach with plastid genome and nrDNA ITS barcodes obtained from genome skimming data.

MATERIALS AND METHODS

Sample collection. We made a reference collection of 72 accessions, consisting of 67 of *Anacyclus* and five of closely related genera, ~~which includes~~including 56 *Anacyclus* accessions and three of closely related genera we collected in Morocco and Spain during field work. We selected a total of 11 samples of *Anacyclus* and two of closely related genera from herbarium voucher specimens of species occurring elsewhere in the Mediterranean. Voucher specimen identifications, collection numbers and locality data are listed in Table S1 and the collection locations are mapped in Figure S1 ~~Forty-two wild populations (n=72 individuals) of *Anacyclus* were sampled in Morocco and Spain to build a reference database. The sampling was complemented with eleven herbarium voucher specimens of species occurring elsewhere in the Mediterranean. The outgroups included two species of *Matricaria*, two species of *Achillea*, and one species of *Otanthus*. The specimen origins and vouchers number are listed in Table S4 and Figure S12~~. We purchased 110 samples, each ~~One hundred and ten trade vouchers~~ consisting of approximately 50g of roots ~~were bought~~ from collectors, herbalists~~herbal shops~~, middle-men, traditional healer, wholesalers, and export companies in Morocco and India (Table S2~~3~~).

Trade information. We ~~To~~ obtained samples of commercialized ~~traded~~ species~~,~~ ~~we~~ by asking~~ed~~ collectors and traders ~~stakeholders for tiguendizt and iguendez~~ in Morocco~~,~~ for tiguendizt and iguendez, two ~~of~~~~he~~ vernacular names used for both ~~*of the two*~~ *A. pyrethrum* varieties (Ouarghidi et al., 2012).~~,~~ We asked traders ~~and~~ in India for akarkara, a common name used there (Ved & Goraya, 2007). When acquiring ~~During~~ the ~~acquisition of~~ samples, we conducted semi-structured interviews about the roots' trade with ~~from collectors and traders in Morocco and India, semi-structured interviews were conducted to enquire about the trade of *Anacyclus*. In total,~~ 39 informants, enquiring ~~were interviewed and asked~~ where the plant material was sourced, to whom it was sold and in what quantities, for what price, and if

some types of roots were of higher quality than others~~there were several qualities of this medicinal plant~~. We ~~The quantity of this product was weighted in herbalist shops and estimated in export companies, estimates were given by informants. Interviews were carried out anonymously and~~ followed the International Society of Ethnobiology Code of Ethics ("ISE Code of Ethics Online," 2018), ensuring free, prior and informed consent, full disclosure and respect for the confidentiality of informants, during all interviews.~~.~~

Extraction and ~~l~~Library preparation. We extracted DNA ~~from~~ of reference and traded vouchers ~~was extracted~~ from approximately 40 mg of dry leaf or root material using the DNeasy Plant Mini Kit (Qiagen). Total DNA (0.2-1.0 µg) was sheared to 500 bp fragments using a Covaris S220 sonicator (Woburn, MA, USA) (Table S1-2). We prepared d~~D~~ual indexed libraries ~~were prepared~~ using the Meyer and Kircher protocol (Meyer & Kircher, 2010) for genome skimming ~~shotgun sequencing~~ and target capture (BioProject PRJNA631886).

Target capture. Using the genome assembly of *A. radiatus* subsp. *radiatus*, w~~W~~e designed 872 low-copy nuclear markers and associated RNA probes by following the Hyb-Seq pipeline (Weitemier et al., 2014)~~[REF] pipeline, with minor adjustments~~ (SI). For target capture enrichment, we prepared twelve equimolar pools ~~were prepared~~ with ten to 24 samples and an average 300 ng of input DNA per pool. The RNA probes were hybridized for 16 hours before target baiting, and 14 PCR cycles were carried out after enrichment following the MyBaits v3 manual. The enriched libraries and genome skimming- libraries ~~shotgun libraries~~ were sequenced on two Illumina HiSeq 3000 lanes (150bp paired-end).

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

Phylogenomics. We aligned tThe recovered matrices (nuclear markers, ITS and plastomes) were aligned with MAFFT v7.471 (Katoh & Standley, 2013), refined with MUSCLE (Edgar, 2004) and filtered with Gblocks v0.91b (Talavera & Castresana, 2007). Phylogenies were inferred using RAxML v8.0.26 (Stamatakis, 2006), with 1000 bootstrap replicates under the GTRGAMMA model. For the low-copy nuclear markers, the species tree was inferred from the individual nuclear markers trees under the multi-species coalescence (MSC) framework

with ASTRAL-III v5.5.9 (Zhang, Sayyari, & Mirarab, 2017). We used t~~T~~he multi-alleles option in ASTRAL-III ~~was used~~ for reconciliation of the independent evolutionary histories of the alleles. The molecular identification of traded roots was assessed from the MSC tree and posterior probabilities (PP) greater than 0.95. For the ITS phylogenetic reconstruction, we used additional Genbank references (Table S3).

We identified t~~Results from morphological identification of the t~~raded roots ~~were identifiedfirst throughg~~based on morphological characters (described in SI)~~)~~, and we used these to triangulate ~~. W(described in (SI)) were~~e ~~combined the morphological identification with results the from~~ molecular identifications. ~~, and s~~Samples were ~~identified~~ categorised according to their position in the supply chain ~~(Figure 2)~~ and geographical origin ~~(Figure 3)~~.

RESULTS

We constructed a reference database of DNA sequences from fresh and herbarium specimens, consisting of 83 individuals of 10 *Anacyclus* species, and 5 individuals representing outgroup species (Figure ~~S1~~S13~~). Four datasets were retrieved from our shotgun and target capture sequencing methods, standard barcode markers (*matK*, *trnH-psbA*, the *trnL* intron and *rbcL*), ITS, along with complete plastid genomes (from shotgun genome skimming) and hundreds of nuclear markers (from target capture) (Figure S2)~~. We ~~then~~ used this reference database to assess the identity and geographic origins of 110 root samples acquired from traded materials by comparing the results from the four~~se~~ different datasets (Figure 1~~, Figures S4~~3~~-S12~~). We show that the target capture approach is the most powerful method to identify plant species in trade and discover their geographic origin.

Data recovery for ~~plastome, standard barcode loci, ITS and~~ genome skimming and target capture data

After quality control filtering, ~~On~~ an average~~,~~ 2.8 million reads (0.42 GB/sample) were obtained per sample from shotgun sequencing ~~per sample after quality control filtering~~ and 2.99 million reads (0.43 GB/sample) per sample for target capture (Figure S~~1~~3, Table S4). ~~The recovery of the nuclear markers from the target capture experiment outperforms shotgun sequencing. On average, 2.3 million reads (0.69 GB/sample) were obtained from shotgun sequencing per sample after quality control filtering and 1.6 million reads (0.48 GB/sample) for target capture.~~ The target capture yielded an average coverage of 303X~~-~~ for the 443 nuclear markers ~~(73 samples, 80%)~~, whereas the unenriched genome skimming yielded an average coverage of 12X for the 443 nuclear data, 20X for the plastome data and 131X for the ribosomal data (Table S5~~1~~). The loci coverage was calculated with *bedtools coverage* (v 2.29.2). ~~On the nuclear dataset, we observe that~~ Samples below an average of 50X coverage in the nuclear dataset~~,~~ show ~~per sample the matrices have~~ a higher missing data rate (>7%) in the matrices. T~~,~~ ~~t~~hese samples were automatically discarded with our pipeline (Figure 2~~3~~, samples in blue). Adulterated samples from other genera~~us~~ have a coverage close to zero (Figure 2~~3~~, samples in orange, yellow and green). To obtain 100x coverage for the nuclear regions ~~U~~using a genome skimming approach, with~~("Sequencing Support – Coverage Calculator," n.d.)~~ an average genome size of the targeted species of 11.72Gb (Garnatje et al., 2011) and a duplication level of the genome skimming libraries of 9% (SI), it would ~~to obtain 100x coverage for the nuclear regions it will~~ require 14 HiSeq 3000/4000 lanes (Illumina, n.d.). From the genome skimming data ~~shotgun data~~ we assembled ITS and the standard barcodes markers, as well as the plastome (Figure S2).~~-~~ Out of the 110 trade samples, we

succeeded in assembling ITS for 102 samples (93%), the standard ~~barcodes~~ barcode regions for 51 to 61 samples (46% to 55%), and plastomes for 49 samples (~~56~~44%) (Figure 1). ~~The average genome size of the targeted species was 11.72Gb (Garnatje et al., 2011), and to obtain similar data to target capture from shotgun sequencing would have required an increased coverage of 25 fold.~~ The resulting aligned matrices for each of the datasets were 633 bp for ITS (including 5.8S), 4408 bp for the standard barcoding ~~markers~~regions, 110,003 bp for the plastome and 289,236 bp for the 443 nuclear markers recovered from the target capture approach (Table S1). The standard barcoding ~~markers~~ regions included the full coding regions of *matK* 1523 bp and *rbcL* 1438 bp, as well as *trnH-psbA* 500 bp and *trnL* 947 bp. The bioinformatics workflow for data analyses is described in Figure S2.

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~~2~~-5~~3~~). 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~~5~~-7~~6~~). The standard barcode regions~~markers~~, *matK*, *rbcL, trnH-psbA* and *trnL* (Figure S8~~7~~-11~~20~~) 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 S1). 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 material 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~~ samples 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~~ in 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~~2~~-5~~3~~, Table S6~~2~~).

We selected 110 individual roots for DNA analysis from the 66 root batches. Of these 9~~8~~ had a morphology consistent with *Anacyclus,* and ~~12~~ 11 which were classed as similar to *Anacyclus* but likely to be non-*Anacyclus* based on their morphology. We recovered partial plastome assemblies ~~shotgun sequence data from nine~~ of these 1~~2~~ non-*Anacylus* roots and using sequence queries against GenBank, ~~the ITS and plastid sequences~~we obtained ~~a genus-level~~ identifications for nine (=*Plantago* spp., *one Primula* spp. and one *Plumeria* spp.~~.~~) (Figure S1~~3~~, Table S2).

Of the 9~~8~~ *Anacyclus* roots, ~~18~~ 10 had no ~~recoverable~~ ~~useful~~identifiable DNA sequences via any of our methods, with ~~89~~80/~~89~~80all of the remaining samples ~~working for~~identified through target capture ~~vs and only 61/80 for through shotgun sequencing)~~ (Figure S1~~3~~4). The ten discarded samples ~~were~~ present~~ing~~ed very fragmented plastome assemblies ~~and high~~with average 89% missing data ~~in average 89%~~ in the nuclear matrices. These samples had a very low DNA integrity.

The plastome sequences enabled identification of seven~~7~~ roots to the species level, with the remainder identified as *Anacyclus* sp. (Figure 1, Table S7). Neither ITS nor any of the standard ~~barcodes~~ barcode regions were able to discriminate any of these samples below the genus level (Figure 1, S4-5~~6~~, S~~-~~8-11~~9~~).

The nuclear markers gave much higher resolution within *Anacyclus* (Table S2, S7). ~~In our investigation of~~For 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 198 samples from ~~8~~ 10 herbalists revealed ~~11~~ 12 *A. ~~nacyclus~~ pyrethrum* var. *pyrethrum* and seven var. *depressus*. Likewise, our ~~seven~~ 11 samples from ~~four~~ six traditional healer sources identified ~~four~~ seven *A. ~~nacyclus~~ pyrethrum* var. *pyrethrum*, and ~~three~~ four var. *depressus*. For material traded in international markets, our ~~16~~ 17 sequenced samples from three export companies in Morocco identified ~~eight~~ nine ~~*Anacyclus*~~ *A. pyrethrum* var. *pyrethrum*, six var. *depressus*, and two *A. homogamous*. Our analysis of ~~28~~ 30 samples from 17 herbalists in India identified three ~~*Anacyclus*~~ *A. pyrethrum* var. *pyrethrum* and ~~25~~ 27 var. *depressus*.

Geographical source

Market samples in Morocco originate ~~both~~ from various Moroccan areas as well as Algeria, and material from all of these populations of origin can be found in Indian market samples (Table S1~~3~~, Figure ~~S1~~23, 6~~-~~). Of the 9890 non-adulterated roots ~~which we~~ identified to species level using target capture, we were able to associate ~~58~~67% to a specific geographic region (Figure 1, ~~23~~2, Table S1~~2~~). Using phylogenetic analysis, root samples of *A. pyrethrum* var. *pyrethrum* clustered with reference material from the High Atlas (Figure 6, case 4), and *A. pyrethrum* var. *depressus* roots clustered with reference material from different regions in Morocco, including the Rif Mountains, the High Atlas and the Middle Atlas (Figure 6, case 2 and 3). Evidence for the international trade from Algeria to Morocco is highlighted by a

distinctive clade that includes traded roots collected from west Algeria (Figure ~~S13~~6, case 1). The geographic origin was only resolvable with target capture data; standard barcoding ~~markers~~regions, ITS and plastome data lacked variation, resolution or both (Figure 1).

DISCUSSION

This study illustrates the potential for target-capture based DNA barcoding to form the next wave of standard plant DNA-barcoding tools and provide ~~the~~ greatly needed species-level resolution. A key rate limiting step for the standard plant barcodes is that they are fundamentally recovering data from just one or two independent loci (plastid DNA and ITS), which often show trans-specific polymorphism and barcode sharing among related species (Hollingsworth, Graham, & Little, 2011). ~~Even using~~The use of complete plastid genome sequences suffers from the same problem, as the data are all physically linked in a single non-recombinant uni-parentally inherited locus. Several recent hybridization events have occurred in the *Anacyclus* genus and entire plastid genomes or plastid barcode markers, and/or ITS provide limited resolution below genus level (Figure 1~~, Figure S4~~). In contrast, our target capture approach using hundreds of nuclear markers yields significantly higher molecular identification success and more accurate resolution to species and even population level (Figure 1).

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

These data provide new insights into trade of *A. pyrethrum* and highlight the extent of adulteration and ~~the~~ scarcity of *A. pyrethrum* var. *pyrethrum* (Figures 2-3, 6). Only a small proportion of the tested samples from herbalists and traditional healers were the potent *A. pyrethrum* var. *pyrethrum,* with the Indian market in particular dominated by var. *depressus*.

In both Morocco and India some individual sellers had entirely or almost entirely adulterated products. ~~All of the~~Most of the non-*Anacyclus* roots we sequenced were identified as *Plantago* spp. ~~-~~by shotgun sequencing, despite being sampled from six different localities including Morocco and India. As *Plantago* roots are similar in appearance to those of *Anacyclus,* it is possible that they are deliberately added as a 'difficult to identify' adulterant which may go unnoticed by non-specialists.

Our analysis of samples from collectors, wholesalers and export companies detected much less adulteration at this point in the supply chain (Figure 4~~2~~, Table ~~S2~~S6). Collection of *A. pyrethrum* var. *pyrethrum* is carried out by professional harvesters employed by export companies who travel across the country and are considered poachers by local communities (Ouarghidi, Powell, Martin, & Abbad, 2017). Local harvesters have increasing difficulty to supply local trade chains (Ouarghidi et al., 2017, 2012), which may finally result in increased adulteration rates in the poorly-governed, national value chains (Figure ~~2~~4), as has also been observed elsewhere (A. Booker et al., 2012) . Our results also identify previously unreported international trade in North Africa prior to export to the Indian sub-continent (Figure ~~2~~3, 6~~3,~~ ~~S11~~). ~~-~~We provide evidence that export companies in Morocco source material not only in this country, but also from neighbouring Algeria (Figure 6). Applying this molecular identification approach enables us to distinguish samples at population level and uncover these hidden international sourcing channels.

Conservation of *A. pyrethrum*

High national and international demand for *A. pyrethrum* likely encourages its overharvesting and adulteration. As the plant is a remedy of the Indian pharmacopoeia, its demand is likely to increase along with that of other Ayurvedic medicines (Kala, Dhyani, & Sajwan, 2006). Although *A. pyrethrum* has been assessed to be vulnerable ~~internationally~~

globally (Rankou et al., 2015) and endangered in Morocco on the IUCN Red List, the plant is not listed in the CITES appendices and its international trade is not regulated. Nonetheless, continued overharvesting is driving wild populations to critical levels of depletion and conservation policies are necessary. Common strategies to conserve overharvested medicinal plants ~~often~~ include collection and trade restrictions as well as cultivation (Schippmann et al., 2002). Cultivation is often proposed as a solution to both conservation issues and sourcing high quality, appropriately identified material (Hamilton, 2004; Schippmann, Leaman, & Cunningham, 2006; Schippmann et al., 2002). ~~However, trade of cultivated plants has not been as successful as expected due to beliefs that plants harvested from the wild are more potent and thus preferable (Alam & Belt, 2009).~~ ~~The~~ eCultivation ~~his kind of demand requirements~~ necessitates engagement with local communities that depend on plant harvest, as well as monitoring of professional trading networks. Both promotion of sustainable harvesting for livelihood security as well as restriction of unsustainable professional trade ~~are~~ may be needed. Only with fine-grained mapping of sourcing areas and supply chains, as our results highlight for *A. pyrethrum*, can meaningful conservation action be achieved. With the implementation of target enrichment, we identified the origin of the harvested populations of medicinal or traded populations and ~~point out~~indicate where conservation efforts ~~sh~~could be ~~firstly~~initially ~~be~~ implemented. We also reveal ~~a~~ previously~~n~~ undocumented harvesting and trade in Algeria (Figure 6, case 1), and show that *Anacyclus* is ~~also~~ harvested at a national scale in Morocco (Figure 6, case 2-4). This study gives scientific evidence~~s~~ to support conservation ~~programs~~initiatives like ~~the XX program from GDF~~Global Diversity Foundation's High Atlas Cultural Landscapes Programme and potentially ~~attract more attention for~~encourage future *Anacyclus* conservation projects ~~for *Anacyclus*~~.

Future prospects for plant DNA barcoding

Key criteria for developing new DNA barcoding approaches for plants include resolving power (~~telling species apart~~differentiating between taxa and discovering their geographical origins), recoverability (enabling use on a wide diversity of tissue sources), and cost and efficiency (enabling scaling over very large sample sets).

In terms of resolving power, the target capture approach used here ~~is a game changer~~offers substantial improvement compared to plant barcodes based on plastid sequences and ITS. The key enabling step is access to multiple nuclear markers, as this reduces sensitivity of the identification to ~~introgression of~~predominance by one or two loci as is the case for barcodes from rDNA or the plastid genome. In this study, we show ~~the~~ high resolution ~~of~~from nuclear markers retrieved by target capture for identification of species in a genus that has ~~both~~ undergone recent hybridization events (Manzanilla, 2018). ~~s~~[65]~~,~~ The successful recovery of these markers from low quality input DNA is also important and ~~combined~~the combination of these observations ~~and is traded through low DNA integrity samples, and~~ makes the case that this method provides a viable solution for future plant barcoding. In contrast, shotgun sequencing requires a substantial sequencing effort to retrieve the same suite of nuclear markers~~, especially~~ for species with large genomes. Thus for ~~with a large genome like~~ *Anacyclus* it would require ~~25~~ 14 HiSeq 3000 lanes to obtain ~~for~~ the same coverage of the loci we have used here. Overall, and regardless of approach, our findings provide empirical evidence to support predictions made in previous review papers (Hollingsworth et al., 2016) that barcoding based on ~~target capture~~ nuclear markers should outperform standard barcoding methods.

The successful recovery of sequences via target capture depends on how closely related the sampled species are to the reference set on which the baits were designed, and/or the level of variation in the loci that form the bait set (McCormack, Tsai, & Faircloth, 2016; Paijmans, Fickel, Courtiol, Hofreiter, & Förster, 2016). In the current study, the baits were designed from *Anacyclus* and related genera. This optimized their specificity to our study group and enabled their successful high-resolution application for assessing trade. A clear challenge for wider use of target capture~~,~~ is applicability over much greater phylogenetic distances. The recently published universal angiosperm bait~~s~~ kit (Buddenhagen et al., 2016; Johnson et al., 2018)~~,~~ designed to recover 353 loci from a wide diversity of flowering plants offer great potential here, as they ~~. The angiosperm universal bait kit (Johnson et al., 2018)~~ can provide species and population-level resolution (Van Andel et al., 2019)[74]. ~~In addition, t~~There is a need for a more general evaluation of when, how and at what scale to most effectively combine taxon-specific bait sets ~~(as used here)~~ with universal bait sets~~,~~ to simultaneously obtain very high resolution and sequencing success over wide phylogenetic distances.

Another important aspect of recoverability is efficacy with degraded DNA. Drying, storage, and transportation affect the quality of plant material in trade and can cause extensive DNA degradation (Anthony Booker et al., 2014). In consequence, traded samples have similar challenges to working with ancient DNA, herbarium samples or archaeological remains. Target capture is particularly well suited to this challenge (McCormack et al., 2016; Paijmans et al., 2016). Although shotgun sequencing can also be very effective on degraded material ~~(Bakker, 2017; Zeng et al., 2018)~~ (Alsos et al., 2020; Bakker, 2017; Zeng et al., 2018) ~~(Citation)~~, our recovery rate in this study was greater for target capture than shotgun sequencing (Figure 1), and we recovered data from 100% of the samples used for establishing

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

~~Although s~~ The target capture methodology presented here, including library construction and sequencing, cost~~ed~~ 70 USD per sample in 2016, a price similar to those presented by Hale et al. (2020). With the optimisations suggested in Hale et al. (2020), prices drop to 22 USD per sample. ~~Tequencing costs continue to fall, the current protocols~~ strategy ~~for target capture (including library construction and sequencing) are still expensive and~~is ~~equate to 70 USD per sample~~ similar to the prices mentioned in Hale et al. (2020) ~~without optimisation. Using the same protocol, with optimisations, the prices per sample could drop to 22 USD~~ (Hale et al., 2020). ~~This is manageable for well-resoureed projects and high value applications and high value species~~Thus, thise~~, but still prohibitively expensive~~ target enrichment approach is quickly becoming affordable for ~~many~~ large-scale biomonitoring projects ~~also removing the necessity to identify specific markers for specific taxonomic groups or less well-resoureed projects~~. With a growing interest and investment in DNA-based identification solutions in medicine and industry (Afshinnekoo et al., 2017; Menegon et al., 2017), ongoing work ~~should~~ is expected to continue to optimize protocols and~~to~~ drive these costs down even ~~more~~further ~~costs down~~, ~~as in~~reaching those of standard barcoding approaches (Hebert et al., 2018).

CONCLUSION

In plants, the frequent sharing of plastid and ribosomal sequences among con-generic species, coupled with the difficulty of routinely accessing multiple nuclear markers, has acted as a constraint on the resolution of DNA barcoding approaches. Current advances in sequencing technology and bioinformatics are removing this constraint and offer the potential 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.

REFERENCES

Adams, M., Alther, W., Kessler, M., Kluge, M., & Hamburger, M. (2011). Malaria in the Renaissance: remedies from European herbals from the 16th and 17th century. *Journal of Ethnopharmacology*, *133*(2), 278–288.

Afshinnekoo, E., Chou, C., Alexander, N., Ahsanuddin, S., Schuetz, A. N., & Mason, C. E. (2017). Precision metagenomics: rapid metagenomic analyses for infectious disease diagnostics and public health surveillance. *Journal of Biomolecular Techniques: JBT*, *28*(1), 40.

Alam, G., & Belt, J. (2009). Developing a medicinal plant value chain: Lessons from an initiative to cultivate Kutki (*Picrorhiza kurrooa*) in Northern India. *KIT Working Papers Series*, (WPS. C5).

Bakker, F. T. (2017). Herbarium genomics: skimming and plastomics from archival specimens. *Webbia*, *72*(1), 35–45.

Benarba, B. (2016). Medicinal plants used by traditional healers from South-West Algeria: An ethnobotanical study. *Journal of Intercultural Ethnopharmacology*, *5*(4), 320.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Booker, A., Johnston, D., & Heinrich, M. (2012). Value chains of herbal medicines-Research needs and key challenges in the context of ethnopharmacology. *Journal of Ethnopharmacology*, *140*(3), 624–633. doi: 10.1016/j.jep.2012.01.039

Booker, Anthony, Frommenwiler, D., Johnston, D., Umealajekwu, C., Reich, E., & Heinrich, M. (2014). Chemical variability along the value chains of turmeric (*Curcuma longa*): a comparison of nuclear magnetic resonance spectroscopy and high performance thin layer chromatography. *Journal of Ethnopharmacology*, *152*(2), 292–301.

Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., … Kortyna, M. (2016). Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. *BioRxiv*, 086298.

de Boer, H. J., Ouarghidi, A., Martin, G., Abbad, A., & Kool, A. (2014). DNA barcoding reveals limited accuracy of identifications based on folk taxonomy. *PLoS ONE*, *9*(1),

584 e84291. doi: 10.1371/journal.pone.0084291
585 De Vos, P. (2010). European materia medica in historical texts: longevity of a tradition and
586 implications for future use. *Journal of Ethnopharmacology*, *132*(1), 28–47.
587 Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and
588 the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340.
589 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
590 throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.
591 Esquinas-Alcazar, J. (2004). International treaty on plant genetic resources for food and
592 agriculture. *Plant Genetic Resources Newsletter*.
593 Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C.,
594 … Husband, B. C. (2009). Are plant species inherently harder to discriminate than
595 animal species using DNA barcoding markers? *Molecular Ecology Resources*, *9*(s1),
596 130–139.
597 Garcia, S., Hidalgo, O., Jakovljević, I., Siljak-Yakovlev, S., Vigo, J., Garnatje, T., & Vallès,
598 J. (2013). New data on genome size in 128 Asteraceae species and subspecies, with
599 first assessments for 40 genera, 3 tribes and 2 subfamilies. *Plant Biosystems - An*
600 *International Journal Dealing with All Aspects of Plant Biology*, *147*(4), 1219–1227.
601 doi: 10.1080/11263504.2013.863811
602 Garnatje, T., Canela, M. A., Garcia, S., Hidalgo, O., Pellicer, J., Sánchez-Jiménez, I., …
603 Vallès, J. (2011). *GSAD: A genome size in the Asteraceae database*.
604 Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes
605 directly from genomic next-generation sequencing reads—a baiting and iterative
606 mapping approach. *Nucleic Acids Research*, *41*(13), e129–e129.
607 Hamilton, A. (2004). Medicinal plants, conservation and livelihoods. *Biodiversity &*
608 *Conservation*, *13*(8), 1477–1517. doi: 10.1023/B:BIOC.0000021333.23413.42
609 Hebert, P. D., Braukmann, T. W., Prosser, S. W., Ratnasingham, S., deWaard, J. R., Ivanova,
610 N. V., … Sones, J. E. (2018). A Sequel to Sanger: amplicon sequencing that scales.
611 *BMC Genomics*, *19*(1), 219.
612 Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proceedings of the*
613 *National Academy of Sciences*, *108*(49), 19451–19452. doi: 10.1073/pnas.1116812108
614 Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant
615 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.

680 Paijmans, J. L., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of
681 enrichment conditions on cross−species capture of fresh and degraded DNA.
682 *Molecular Ecology Resources*, *16*(1), 42–55.
683 Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., …
684 Willerslev, E. (2012). Glacial Survival of Boreal Trees in Northern Scandinavia.
685 *Science*, *335*(6072), 1083–1086. doi: 10.1126/science.1216043
686 Parks, M., Cronn, R., & Liston, A. (2009). Increasing phylogenetic resolution at low
687 taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC*
688 *Biology*, *7*(1), 84.
689 Pillon, Y., Johansen, J., Sakishima, T., Chamala, S., Barbazuk, W. B., Roalson, E. H., …
690 Stacy, E. A. (2013). Potential use of low-copy nuclear genes in DNA barcoding: a
691 comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary*
692 *Biology*, *13*(1), 35.
693 Pittle, K. D. (2005). *Continuity and Change in Islamic Ethnopharmacological Practice: New*
694 *Methods for Cognitive Dialectometry*. Florida State University.
695 Rankou, H., Ouhammou, A., Taleb, M., Manzanilla, V., & Martin, G. (2015). *Anacyclus*
696 *pyrethrum*. *The IUCN Red List of Threatened Species 2015: E.T202924A53798702*,
697 *2015*(2015). Retrieved from http://dx.doi.org/10.2305/IUCN.UK.2015-
698 4.RLTS.T202924A53798702.en
699 Rosato, M., Álvarez, I., Feliner, G. N., & Rosselló, J. A. (2017). High and uneven levels of
700 45S rDNA site-number variation across wild populations of a diploid plant genus
701 (*Anacyclus*, Asteraceae). *PloS One*, *12*(10), e0187131.
702 Särkinen, T., Staats, M., Richardson, J. E., Cowan, R. S., & Bakker, F. T. (2012). How to
703 open the treasure chest? Optimising DNA extraction from herbarium specimens. *PloS*
704 *One*, *7*(8), e43808.
705 Schippmann, U., Leaman, D., & Cunningham, A. (2006). A comparison of cultivation and
706 wild collection of medicinal and aromatic plants under sustainability aspects. *Frontis*,
707 75–95.
708 Schippmann, U., Leaman, D. J., & Cunningham, A. (2002). Impact of cultivation and
709 gathering of medicinal plants on biodiversity: global trends and issues. *Biodiversity*
710 *and the Ecosystem Approach in Agriculture, Forestry and Fisheries*.
711 Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S. C. K., …

712 Suda, J. (2016). Phylogenetic marker development for target enrichment from
713 transcriptome and genome skim data: the pipeline and its application in southern
714 African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources*, *16*(5), 1124–1135. doi:
715 10.1111/1755-0998.12487
716 Soltis, P. S., & Soltis, D. E. (2009). The Role of Hybridization in Plant Speciation. *Annual*
717 *Review of Plant Biology*, *60*(1), 561–588. doi:
718 10.1146/annurev.arplant.043008.092039
719 Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
720 with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.
721 Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev,
722 E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA
723 barcoding. *Nucleic Acids Research*, *35*(3), e14. doi: 10.1093/nar/gkl938
724 Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent
725 and ambiguously aligned blocks from protein sequence alignments. *Systematic*
726 *Biology*, *56*(4), 564–577.
727 Tiwari, N. N., Poudel, R. C., & Uprety, Y. (2004). Study on domestic market of medicinal
728 and aromatic plants (MAPs) in Kathmandu Valley. *Kathmandu: Winrock*
729 *International*.
730 Ved, D., & Goraya, G. (2007). Demand and supply of medicinal plants in India. *NMPB, New*
731 *Delhi & FRLHT, Bangalore, India*, *18*.
732 Veldman, S., Gravendeel, B., Otieno, J. N., Lammers, Y., Duijm, E., Nieman, A., … van
733 Andel, T. R. (2017). High-throughput sequencing of African chikanda cake highlights
734 conservation challenges in orchids. *Biodiversity and Conservation*, *26*(9), 2029–2046.
735 WHO. (2004). *WHO guidelines on safety monitoring of herbal medicines in*
736 *pharmacovigilance systems*.
737 Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., … Taberlet,
738 P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*,
739 *506*(7486), 47–51. doi: 10.1038/nature12921
740 Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L.
741 H. (2009). The frequency of polyploid speciation in vascular plants. *Proc Natl Acad*
742 *Sci U S A*, *106*. doi: 10.1073/pnas.0811575106
743 Wysoker, A., Tibbetts, K., & Fennell, T. (2015). *Picard tools*. Broad Institute.

744 Zeng, C.-X., Hollingsworth, P. M., Yang, J., He, Z.-S., Zhang, Z.-R., Li, D.-Z., & Yang, J.-B.
745 (2018). Genome skimming herbarium specimens for DNA barcoding and
746 phylogenomics. *Plant Methods*, *14*(1), 43.
747 Zhang, C., Sayyari, E., & Mirarab, S. (2017). *ASTRAL-III: Increased Scalability and Impacts*
748 *of Contracting Low Support Branches*. 53–75. Springer.
749

750

DECLARATIONS

Authors' contributions

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

Acknowledgments

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

Reproducibility

For reproducibility purposes, all the scripts used during the data processing are available on the OSF work repository https://osf.io/9bh3p/. New sequencing data have been deposited 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 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-whiskers 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 three 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.