Lei Duan ORCID iD: 0000-0001-6152-5458

Bin-Bin LIU ORCID iD: 0000-0002-0297-7531

Artem Leostrin ORCID iD: 0000-0002-9269-7954

Species delimitation of the liquorice tribe (Leguminosae: Glycyrrhizeae) based on phylogenomic and machine learning analyses

Running title: Species delimitation of liquorice

Lei Duan[1*], Li-Na Han[2], Bin-Bin Liu[3], Artem Leostrin[4], AJ Harris[1], Lin Wang[5], Emine Arslan[6], Kuddisi Ertuğrul[6], Mikhail Knyazev[7], Elena Hantemirova[7], Jun Wen[8*], Hong-Feng Chen[1*]

[1]*Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*

[2]*College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China*

[3]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*

[4]*Komarov Botanical Institute, Russian Academy of Sciences, Professor Popov Street 2, Saint Petersburg 197376, Russia*

[5]*Guangdong Eco-Engineering Polytechnic, Guangzhou 510520, China*

[6]*Department of Biology, Faculty of Science, Selçuk University, Konya 42031, Turkey*

[7]*Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Yekaterinburg, 620144, Russia*

[8]*Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington D.C. 20013-7012, U.S.A.*

[*]Corresponding author

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jse.12902.

E-mail address: duanlei@scbg.ac.cn (L. Duan); *wenj@si.edu* (J. Wen); *h.f.chen@scbg.ac.cn* (H.-F. Chen)

Abstract The liquorice tribe Glycyrrhizeae is a leguminous herbaceous group of plants comprised of the genera *Glycyrrhiza* and *Glycyrrhizopsis*. Some *Glycyrrhiza* taxa contain glycyrrhizin, a pharmacologically significant sweet substance that also has applications in crafting industrial materials. Here, we utilized an expanded taxon sampling of Glycyrrhizeae to reconstruct the phylogenetic relationships in the tribe based on genome skimming data, including whole chloroplast genomes, nuclear ribosomal DNA, and low-copy nuclear DNA. We also launched machine learning analysis (MLA) for one species pair with controversial taxonomic boundary. The integrated results indicated *Glycyrrhizopsis* should be split from *Glycyrrhiza*, while the former genus *Meristotropis* should be treated as part of *Glycyrrhiza*. *Glycyrrhizopsis* includes two species, *G. asymmetrica* and *G. flavescens*, and we recognize 13 species in *Glycyrrhiza*: *G. acanthocarpa*, *G. astragalina*, *G. bucharica*, *G. echinata*, *G. foetida*, *G. glabra*, *G. gontscharovii*, *G. lepidota*, *G. macedonica*, *G. pallidiflora*, *G. squamulosa*, *G. triphylla* and *G. yunnanensis*. We propose a broader *G. glabra* that includes former *G. aspera*, *G. glabra* s.s., *G. inflata* and *G. uralensis*, and represents the glycyrrhizin-contained medicinal group. Our ancestral state inferences show the ancestor of *Glycyrrhiza* lacked glycyrrhizin, presence of glycyrrhizin evolved twice within *Glycyrrhiza* during the last one million years. Our integrative phylogenomics-MLA study not only provides new insights into long-standing taxonomic controversies of Glycyrrhizeae, but also represents a

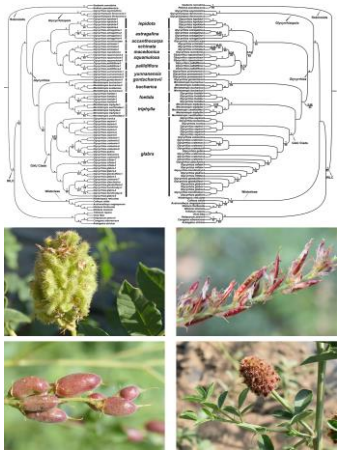
useful approach for future taxonomic studies on other plant taxa.

Graphical abstract

With an expanded taxon sampling of the liquorice tribe, Glycyrrhizeae, we launched phylogenetic analyses based on chloroplast coding sequences (cp CDSs), nuclear ribosomal DNA (nrDNA) and low-copy nuclear (LCN) loci, as well as machine learning analyses (MLAs), to recognize two and 13 species within genera *Glycyrrhizopsis* and *Glycyrrhiza*, respectively.

Our ancestral state inferences show the ancestor of *Glycyrrhiza* lacked glycyrrhizin, and the presence of glycyrrhizin evolved twice within *Glycyrrhiza* during the last one million years.

Our integrative phylogenomics-MLA study not only provides new insights into long-standing taxonomic controversies of Glycyrrhizeae, but also represents a useful approach for future taxonomic studies on other plant taxa.



Key words: character evolution, *Glycyrrhiza*, machine learning analysis, medicinal group, phylogenomics, species delimitation

1. Introduction

The liquorice tribe Glycyrrhizeae is a perennial herbaceous papilionoid group (Fabaceae) that is adapted to mesophytic and xerophytic habitats in temperate Eurasia, North Africa, Australia, and the Americas (Duan et al., 2021a). The long, strong root of liquorice has been widely used as an important traditional medicine in the temperate Old World countries from eastern Asia to the Mediterranean regions due to its efficacy of relieving cough and phlegm (Chinese Pharmacopoeia

Commission, 2015; Dastagir & Rizvi, 2016; Zhou & Jin, 2016; Öztürk et al., 2017; Graebin, 2018; Sharma et al., 2018). Beyond traditional medicine, liquorice has broad economic importance for its utility in modern pharmaceutical products, cosmetics, food additives, tobacco flavoring, and popular confectioneries (Richardson, 2008; Hayashi & Sudo, 2009). Despite the importance and applications of Glycyrrhizeae, there is a long-standing debate with respect to its inter- and infra-generic taxonomy (see below and Table 1), and this affects the medicinal and industrial use of this genus, as well as collection and breeding of liquorice germplasm resources.

Tribe Glycyrrhizeae was proposed by Rydberg (1917) and was resurrected based on genera *Glycyrrhiza* L. and *Glycyrrhizopsis* Boissier (Duan et al., 2021a). Since *Glycyrrhiza* was established by Linnaeus (1753), two satellite genera were erected within it: the trifoliate (rarely 5-foliolate), inflated-fruit species composed of the Central Asian endemic *Meristotropis* Fisher & C.A.Meyer (1843); and the taxa with eglandular, compressed, dehiscent pods representing *Glycyrrhizopsis* (Boissier, 1856; see Fig. 1A-C), which is restricted to southern Anatolia of Turkey and Syria (Çetin, 2015). Historically, *Glycyrrhizopsis* and *Meristotropis* have been controversially treated within or outside of *Glycyrrhiza* among different inter-generic taxonomic revisions. For example, Kruganova (1955) accepted both of *Glycyrrhizopsis* and *Meristotropis*, while Engler and Prantl (1894), Li (1963) and Li & Lu (2015) treated both of them in *Glycyrrhiza*. On the other hand, Boisser (1856) and Meng (2005) merged *Meristotropis* with *Glycyrrhiza*, but

retained the generic status of *Glycyrrhizopsis*, which is consistent with our recent chloroplast phylogenomic results (Duan et al., 2020, 2021a). However, the taxonomic status of Glycyrrhizeae remains unsettled and requires a strengthened inter-generic demarcation based on more robust evidence, e.g., nuclear data-based phylogenetic analyses.

At the specific level, taxonomists delimited *Glycyrrhiza* species based on a series of morphological characters such as leaflet number per leaf, inflorescence/infructescence shape, fruit shape, pericarpic appendage, however, massive morphological transitional phenomenon blurs the specific boundary within *Glycyrrhiza*. The genus has experienced five major revisions (Table 1), a torturous taxonomic history with the number of species varying from 13 (Kruganova, 1955; Meng, 2005) to 36 (Grankina, 2008). Each of the revisions supported one to several controversial species that were rejected by some other workers, e.g., *Glycyrrhiza korshinskyi* Grig., *G. macedonica* Boiss. & Orph. (see Fig. 1F) and *G. yunnanensis* Cheng f. & L.K.Tai ex P.C.Li. Furthermore, a few dubious species were not accepted within any of the five major revisions, e.g., *Glycyrrhiza gontscharovii* Maslenn. and *G. iconica* Hub.-Mor. It is clear that the existing species circumscriptions within *Glycyrrhiza* remain debated.

Not all the liquorice taxa can be used as medicinal plant, and only the glycyrrhizin-contained group has been qualified, i.e., the medicinal group (Chinese Pharmacopoeia Commission, 2015; Li & Lu, 2015; Öztürk et al., 2017). This medicinal group comprises a complex of four generally-accepted species:

Glycyrrhiza aspera Pall., *G. glabra* L., *G. inflata* Batal. and *G. uralensis* Fisch. ex DC., which form a clade in our recent molecular phylogenetic analysis (as the GAU clade in Duan et al., 2020), and many more dubious species erected based on morphology within the group (e.g., Li, 1993; Grankina, 2008). Due to the complex morphology and unclear species boundaries within *Glycyrrhiza*, liquorice resources, especially the medicinal group, are often improperly applied. For instance, sometimes non-medicinal species are sold as medicinal liquorice in market; even a few scientific studies on pharmacology or chemistry used incorrectly identified liquorice materials. An accurate specific delimitation may facilitate the medicinal, industrial and scientific applications of Glycyrrhizeae.

In pharmacology, the most crucial medicinal compound of liquorice is glycyrrhizin (glycyrrhizic acid), a triterpenoid saponin glycoside with hepatoprotective and anti-inflammatory bioactivity, which is also a natural sweetener, 30-50 times sweeter than sucrose (Hayashi & Sudo, 2009; Kao et al., 2014; Li & Lu, 2015; Öztürk et al., 2017; Pandey et al., 2017; Graebin, 2018; Sharma et al., 2018). Additionally, the presence of glycyrrhizin has been regarded as a taxonomically key character to recognize sect. *Euglycyrrhiza* Boiss. (as in Kruganova, 1955), sect. *Glycyrrhiza* ser. *Glabrae* Vass. (as in Li, 1963) or sect. *Glycyrrhiza* P.C.Li (as in Meng, 2005) within *Glycyrrhiza*. Recent studies have shown that the North American species, *G. lepidota* Pursh, also contains glycyrrhizin, albeit at a low concentration, (Hayashi et al., 2005; Li & Lu, 2015), rendering the glycyrrhizin-contained taxa non-monophyletic according to our

prior phylogenetic result (Duan et al., 2020). Thus, the presence of glycyrrhizin may be ancestral trait in *Glycyrrhiza*, or the trait have evolved twice within a relatively recent time. Also, liquiritin is regarded by *Pharmacopoeia of the People's Republic of China* (Chinese Pharmacopoeia Commission, 2015) as a medicinal component of *Glycyrrhiza*. Liquiritin was detected in species beyond the medicinal group (Li & Lu, 2015), but extant researches on liquiritin do not cover all the species in *Glycyrrhiza*, making it difficult to be applied to taxonomic treatment or to ancestral state analyses.

Molecular phylogenetic results are regarded as one of the most reliable sources of evidence to support taxonomic treatments (e.g., Duan et al., 2016, 2019; da Cruz et al., 2018; Wang et al., 2020). Based on molecular phylogeny, at the inter-generic level, *Glycyrrhiza* was resolved as one of the first-diverging taxa within the inverted repeat-lacking clade (IRLC) of Papilionoideae (i.e., papilionoid legumes; Wojciechowski et al., 2000, 2004; Lavin et al., 2005; Duan et al., 2015, 2021b), and several other studies suggested that *Glycyrrhizopsis flavescens* Boiss. was sister to *Glycyrrhiza* (Erayman et al., 2014; Çetin, 2015; Altay et al., 2016). Within *Glycyrrhiza*, two groups were recovered based on the chloroplast *rbcL*: one group included *G. glabra*, *G. uralensis* and *G. inflata* and the other contained *G. echinata* L. and *G. pallidifolia* Maxim. (Hayashi et al., 1998, 2000; Hu & Chang, 2003). Additionally, studies led to the discovery that *G. lepidota* was sister to the rest taxa of *Glycyrrhiza* (Hu et al., 2002; Hayashi et al., 2005; Meng, 2005). However, all of these prior studies suffered from

under-sampling of species (i.e., a maximum of seven species) and, therefore, did not represent strong molecular evidence for an existing or revised taxonomic treatment.

To remedy this, we sampled every generally-accepted species in our prior chloroplast (cp) phylogenomics study, and based on our results, we recognized four main clades within Glycyrrhizeae (Duan et al., 2020), but still lacked sufficient evidence to treat some of the controversial species, e.g., *Glycyrrhiza macedonica*. Furthermore, the uniparental transmission of chloroplast genome (Corriveau & Coleman, 1988; McCauley et al., 2007; Wicke et al., 2011), to some extent, may lead to inaccurate phylogenetic relationship, nuclear data is required to reinforce the analytical reliability. Multiple studies indicated that nuclear ribosomal DNA (nrDNA) is useful to resolve phylogenetic relationships (e.g., Liu et al. 2020; Duan et al., 2021b), but its frequent gene duplication may also cause misleading result. Thus low-copy, even single-copy nuclear genes were considered to be a better data source to construct phylogenetic trees (e.g., Nikolov et al., 2019; White et al., 2019). Recently, Liu et al. (2021) developed a new approach to extract low-copy nuclear (LCN) genes from genome skimming reads, which granted us an opportunity to apply more nuclear data to complete our former cp phylogenomic framework of Glycyrrhizeae in Duan et al. (2020).

In this study, we seek to further clarify specific relationships within controversial species complexes of *Glycyrrhiza* by applying a machine learning analysis (MLA) integrated with phylogenomics. MLAs, especially those that are

image-based, now represent an alternative approach to species delimitation that can avoid possible subjective plant classification (Schuettpelz et al., 2017; Wäldchen & Mäder, 2017; Wang et al., 2017; Kho et al., 2018; Younis et al., 2018; Hussein et al., 2020). Moreover, images of plant specimens are now widely available in high resolution from herbaria around the world, resulting from the individual and the concerted efforts at digitization of collections (Carranza-Rojas et al., 2017), and this facilitates the application of MLAs to taxonomic studies. In general, MLAs comprise the training and application phases. The training phase consists of an analysis of images, which have been unquestionably identified in advance by taxonomists, to generate a set of classifier's parameters that can then be used in the application phase for probabilistically inferring the identities of unidentified specimen images (Mata-Montero & Carranza-Rojas, 2016; Unger et al., 2016; Wäldchen & Mäder, 2017; Wäldchen et al., 2018). However, MLAs are not flawless (Bonnet et al., 2018; Carranza-Rojas et al., 2018a, b; Wäldchen et al., 2018) and should probably not be applied as the sole form evidence in taxonomic treatments at present. While MLAs are capable of measuring and classifying morphological dimensions of images not recognized or recognizable by humans, using them in isolation for taxonomic revision would effectively represent utilization of a morphological species concept. It might be proper to use the specimen-image-based MLA as an additional evidence to the taxonomic circumscription based on the corresponding molecular phylogenetic results and/or

morphological studies. Notably, integrative phylogenetics-MLA approaches to taxonomic delimitation and revision have rarely been reported.

Compared to our prior work (Duan et al., 2020), we presently employed a denser taxon sampling within Glycyrrhizeae, and obtained complete cp genomes, nrDNA and LCN genes from genome skimming data (Straub et al., 2012; Zhang et al., 2015a) to build phylogenetic trees. In addition, we applied the MLA results to test boundary in a species complex that was insufficiently resolved by our phylogenetic trees. We aimed to (1) clarify the generic and specific taxonomy of Glycyrrhizeae and (2) explore whether presence of glycyrrhizin is ancestral in the group. Our results provide a theoretical foundation for proper, efficient, sustainable utilization of the of liquorice resource, and our methods represent a new approach to untangle recalcitrant taxonomic questions by combining phylogenetic trees and image-based MLAs.

2. Materials and methods

2.1 Morphological Study and Taxon Sampling

For the purpose of familiarizing taxonomic revisions and key morphological traits in classification of Glycyrrhizeae, we carried out a thorough literature search on main revisions (Kruganova, 1955; Li, 1963; Meng, 2005; Grankina, 2008; Li & Lu, 2015) and checklists (Lock, 1989; Podlech, 1991; Yakovlev et al., 1996) of this tribe, as well as most floras that record the genera *Glycyrrhiza*, *Glycyrrhizopsis* and *Meristotropis* (e.g., Grigorev & Vasilchenko, 1948; Yeo, 1968; Chamberlain, 1970; Ali, 1977; Scoggan, 1978; Rechinger, 1984; Jeanes, 1996;

Gómer-Sosa, 1999; Bao and Larsen, 2010). Samples and specimens of Glycyrrhizeae were collected in the field from Turkey, Russia, U.S.A. and every province of China where this tribe distributed. Besides, we viewed all the Glycyrrhizeae specimens deposited in the herbaria IBSC, KUN, LE, NY, PE, US, WUK and XJBI, in addition to all the online specimen images from E, G, K, LINN and P, to further reinforce our understanding of the Glycyrrhizeae morphology.

According to the world-wide revisions of *Glycyrrhiza*, we herein provided that a taxon can be recognized as a "widely accepted species" when it was accepted by no less than four main revisions in Table 1, otherwise, the taxon is a "dubious species". In this case, our sampling for molecular phylogenetic analyses consisted of 68 accessions representing all of the 14 widely accepted species and 13 dubious species of Glycyrrhizeae, covering all the species recognized by Kruganova (1955), Li (1963) and Meng (2005), except for the controversial taxa *Glycyrrhiza eurycarpa* P.C.Li and *Glycyrrhizopsis syriaca* Turrill, for which we could not find reliable materials. The dubious species included in this study are: *G. alaschanica* Grankina, *G. asymmetrica* Hub.-Mor., *G. iconica*, *G. glandulifera* Waldst. & Kit., *G. gobica* Grankina, *G. gontscharovii*, *G. korshinskyi*, *G. laxissima* Vassilcz., *G. macedonica*, *G. yunnanensis*, *G. zaissanica* Serg., *Meristotropis kulabensis* Masl. and *M. xanthioides* Vass.

On the other hand, we used all the cp genomes of Glycyrrhizeae published in our previous study (Duan et al., 2020), as well as ten newly sequenced accessions,

including five dubious species (*Glycyrrhiza alaschanica*, *G. gobica*, *G. gontscharovii*, *G. korshinskyi* and *G. zaissanica*), in order to better test the monophyly of each species within this tribe (see Table S1 for details). Outside Glycyrrhizeae, we sampled five species representing four genera within the tribe Wisterieae sensu Compton et al. (2019), and another five species standing for five other genera of the IRLC as in LPWG (2013) and Duan et al. (2021b). Two species representing two genera for the Robinioids (sensu Wojciechowski et al., 2004) were included as outgroups. Most sequences for the study were obtained from field-collected or herbarium specimens (77 accessions, 37 species; see Table S1), except that DNA samples of *Austrocallerya megasperma* (F.Muell.) J.Compton & Schrire, *Wisteria floribunda* (Willd.) DC. and one accession of *Glycyrrhiza astragalina* (#1 in tables and figures) were purchased from the DNA and Tissue Bank of the Royal Botanic Gardens, Kew (https://dnabank.science.kew.org). Raw sequencing reads were deposited in database of NCBI (www.ncbi.nlm.nih.gov; see Table S1).

2.2 DNA Extraction, Genome Assembly, Annotation and Alignment

Total genomic DNA was extracted following a modified CTAB protocol (Doyle & Doyle, 1987). We quantified yield and integrity (size distribution) of genomic DNA extracts by visual assessment on 1% agarose gels, as well as by fluorometric quantification on a Qubit (Invitrogen, Carlsbad, California, USA) using a dsDNA HS kit. Subsequently, all samples were used to build blunt-end DNA libraries via the NEBNext Ultra II DNA library Prep kit for Illumina (New England Bio-labs)

according to the protocol of the manufacturer. The final indexed libraries were then pooled in equimolar ratios and were sequenced in a single lane of an Illumina XTen sequencing system (Illumina Inc.).

We filtered out adaptors and low-quality reads with Trimmomatic v.0.33 (Bolger et al., 2014) from the raw reads. The quality of the remaining reads were checked using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). We then performed *de novo* assembly in SPAdes 3.11 (Bankevich et al. 2012) with k-mer of 75, 85, 95, and 105, and employed a customized python script (Jin et al., 2018) with its default parameters to apply BLAST and a built-in library to connect verified contigs into plastomes in SPAdes. The assembly of the resulting complete cp genomes were annotated using the Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) with *Glycyrrhiza glabra* [GenBank Accession #: NC_024038; Sabir et al. (2014)] as a reference. Most of our samples were members of the IRLC (Wojciechowski et al., 2004), which lack one of the two the inverted repeat (IR) regions (Lavin et al., 1990). Thus, to better align with the cp genomes of Glycyrrhizeae, one of the two IR regions of *Sesbania cannabina* (Retz.) Poir. and *Robinia pseudoacacia* L. were removed.

As per the extraction of nrDNA assembly, we first screened and excluded all the plastid-like reads with abovementioned BLAST approach. Using the remaining reads, *de novo* assemblies were performed with the complete nrDNA sequence (including 5.8S, ITS1, 18S, ITS2 and 26S ribosomal RNA genes) of *Glycyrrhiza uralensis* Fisch. ex DC. (GenBank Accession #: KX530461) with

mapping following Ji et al. (2019). The nuclear ribosomal RNA genes and their boundaries with ITS regions were annotated and defined by comparison with the annotated reference within Geneious Prime 2020 (Kearse et al., 2012). Additionally, based on the HybPiper pipeline (Johnson et al., 2016), we used a newly proposed approach (Liu et al., 2021) to extract low-copy nuclear (LCN) genes from genome skimming data, by mapping a set of published LCN genes (as in Vatanparast et al., 2018) to our sequencing raw reads, and we managed to obtain 496 LCN genes in this way. Given our limited sequencing depth, we found that only five LCN genes, each of which covered over 90% of the sequenced accessions, had acceptable quality (relatively intact gene) for the subsequent phylogenetic analyses (see Table S2 and recovery efficiency in Fig. S1).

2.3 Phylogenetic Analyses

We aligned nrDNA and LCN gene separately using MAFFT v.7 (Katoh & Standley, 2013), applied partitioning to these genes for phylogenetic analyses, and determined the best nucleotide substitution models in PartitionFinder 2 (Lanfear et al., 2016; see Table S2) under the default settings. Whole cp genomes were aligned without partitioning as intergenic spacers can hardly be reliably modeled independently according to many other recent studies (Wei et al., 2017; Wen et al., 2018; Yang et al., 2019). Based on the cp alignment, GTR+G was determined as the best model of nucleotide substitution with PartitionFinder 2.

As for the nuclear sequences, we pretested the performance of the undermentioned Bayesian inference (BI; Rannala & Yang, 1996; Mau et al., 1999)

and maximum likelihood (ML) analyses based on three individual nuclear datasets: nrDNA, concatenated LCN genes, concatenated nrDNA and LCN genes. The results indicated that the BI trees constructed from the three individual datasets have identical topology, but the tree based on concatenated nrDNA and LCN genes had higher support in general than that from nrDNA or concatenated LCN genes (see Trees S1-S2 in supplementary file). The ML trees showed the same case (see Trees S3-S4 in supplementary file). We thus adopted the BI and ML trees based on concatenated nrDNA and LCN genes. In addition, we inferred coalescent-based species tree via ASTRAL-III (Zhang et al., 2018) based on the gene trees of five unlinked LCN genes, and the topological support was generally weak (see Tree S5 in supplementary file).

Based on cp genome and concatenated nuclear data, separate phylogenetic analyses were carried out with BI implemented in the program MrBayes 3.2.5 (Ronquist & Huelsenbeck, 2003; Ronquist et al., 2012) by applying default prior settings. We performed each BI using two independent runs of the Markov chain Monte Carlo (MCMC) for 10 million generations with sampling every 1,000 generations. The first 25% trees were discarded as burn-in and the remaining posterior topologies were summarized as a maximum clade credibility (MCC) tree. Stationarity of the analyses were verified using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer) by ensuring that all ESS values exceeded 200, and convergence was confirmed between independent runs. Apart from BI, ML analyses were also performed with IQ-TREE v.1.6 (Nguyen et al., 2015)

using the following settings: rapid bootstrap analysis with 1,000 replicates followed by a search for best-scoring ML tree starting with a random seed.

2.4 Machine learning analysis

The MLA can be implemented to test the specific status of some dubious taxa, that can hardly be verified by our phylogenetic results (see Fig. 2 below) or previous morphology-based revisions (see Introduction): 1. phylogenetically, a single-accessioned dubious species was sister to a widely accepted species with multiple accessions, indicating that they can be regarded either as two smaller clades, or as one larger clade; 2. taxonomically, workers controversially treated the former as a separate species or a synonym of the latter. According to the criteria, we can thus applied the MLA to the species pair *Glycyrrhiza echinata-G. macedonica* (see Fig. 1D-F), and sought further evidence to clarify their species boundary. This attempt may serve as an example for the future phylogenetics-MLA integrative studies on plant taxonomy.

As input for the MLA, we took 35 and 24 high resolution pictures of herbarium specimens for *G. echinata* and *G. macedonica*, respectively (herbarium codes, specimen barcodings and collection localities see Table S3), including both of their type specimens (Fig. 1D-E). To reduce recognition error in the model training-validating step, we selected specimen images with the uniform reproductive organ of infructescence. The high-resolution specimen photographs were converted into ".png" format and segmented into 100 (10×10) subunits, which were then manually cleaned to remove morphologically uninformative

images, e.g., margin image of the specimen sheet.

Given the trait-based decision-tree models require artificially definition of every morphological trait, which would introduce excessive subjective factor to the consequent MLA, and they consume unaffordable computational resource compared to the widely used pixel-based deep learning models. Therefore, we used a supervised ImageNet approach proposed by Studer et al. (2019) (weight decay = $4e^{-5}$) to pre-train the model, then trained the deep learning models (as in Szegedy et al., 2016) with 90% image subunits of the species pair of *Glycyrrhiza echinata-G. macedonica* (see Fig. 1D-F). The remaining 10% of subunits were applied to validate the model. We performed the training-validating process three times on randomly selected groups of subunits (Samples 1-3). When training, every input image subunit was converted into 448×448 pixel per inch in size, and the training subunits were set to randomly up-down and light-right reverse. The parameters of rotation angle, color saturation, color contrast, brightness, pixel value conversion were randomly assigned values in the range of -pi/2-pi/2, 0.5-1.5, 0.5-1.5, 64.0-255.0. The pixel value conversion was 0-1.0 to -1.0-1.0. We used ca. 200 epochs for each data training, and the RMSProp optimizer was applied (Tieleman & Hinton, 2012) with the following setting: decay = 0.9, momentum = 0.9, epsilon = 1.0, learning rate = 0.001. Four resultant values were calculated to indicate the outcomes:

Accuracy: number of correctly predicted image subunits / number of all image subunits of a certain species pair

Precision: number of the correctly predicted subunits as a certain species / number of all subunits that were predicted as the certain species

Recall rate: number of the correctly predicted subunits as a certain species / number of all subunits that were labeled as the certain species

F1-score: harmonic mean of precision and recall rate

2.5 Ancestral state reconstruction

We used the Bayesian Binary Method (BBM; Ronquist, 2004) implemented in RASP v.4.0 (Yu et al., 2015) to infer the ancestral state of glycyrrhizin in *Glycyrrhiza*. The states of extant species were codified as absence (A) and presence (B) of glycyrrhizin based on a summary of earlier phytochemical researches (Li, 1963; Hayashi et al., 2005; Li & Lu, 2015), as well as according to our taxonomic revision of the medicinal group (see Taxonomic Treatment). As input for the BBM analyses, we used 12,000 post-burnin cp genome trees and the corresponding MCC tree from the abovementioned BI analysis. We trimmed the trees with the ape library in R (Paradis & Schliep, 2018) to include only one accession per species (see Table S1), and each of the four varieties in the enlarged *G. glabra* was represented by one morphologically typical accession (treatment of *G. glabra* see below). We also pruned all other taxa outside of *Glycyrrhiza* as including sparsely sampled related lineages is ill-advised and can lead to erroneous inferences at the root (Ronquist, 1997; Harris et al., 2013). We ran BBM for 100,000 generations, implementing nine hot chains and one cold chain, with a sampling frequency of 100 from the cold chain. We set the maximum

number of allowed states to two, and we applied the F81+G model for changes among character states.

3. Results

Plastomes of all 80 sampled accessions were successfully assembled into complete circular configurations (see Table S1). The sizes of the cp genomes ranged from 122,542 bp to 156,702 bp and the GC contents were between 33.9% and 35.9% (see Table S1 for details). The length of cp genome alignment was 239,967 bp (see Supplementary Alignment 1). We found that the cp genome structure and gene order of Glycyrrhizeae were identical to those reported in previous studies on species in the group (Raveendar et al., 2017; Kang et al., 2018). The alignment of the concatenated nrDNA and LCN genes was 10,548 bp long (see Supplementary Alignment 2), and summary statistics of these nuclear genes are given in Table S2.

3.1 Phylogenetic relationships

We observed that BI and ML approaches in phylogenetic reconstruction yielded the same tree topology for the cp genomes and, separately, for the concatenated nuclear DNA. Thus, we presented the BI tree topologies for these two datasets with both Bayesian posterior probabilities (PP) and ML bootstrap (LBS) values (Fig. 2).

The nuclear results (Fig. 2B) indicated that Glycyrrhizeae is monophyletic (PP = 1, LBS = 100%), forming a well-supported IRLC (PP = 1, LBS = 99%) along with the tribe Wisterieae (PP = 1, LBS = 99%) and five other legume taxa.

Within Glycyrrhizeae, *Glycyrrhizopsis* (PP = 1, LBS = 100%), contained *Glycyrrhizopsis flavescens* (PP = 1, LBS = 100%) and *Glycyrrhiza asymmetrica*. *Glycyrrhiza* (PP = 1, LBS = 100%) was sister to *Glycyrrhizopsis*, and comprised two strongly supported subclades. The first subclade was composed of dubious species of *G. macedonica* and *G. yunnanensis*, as well as six well accepted species: *Glycyrrhiza acanthocarpa* J.M.Black, *G. astragalina* Gillies, *G. echinata*, *G. lepidota*, *G. pallidiflora* Maxim. and *G. squamulosa* Franch., of which five species were robustly supported, except for *G. acanthocarpa* with one single accession. The second subclade contained 17 widely accepted species and dubious species. Accessions of *Meristotropis kulabensis* and *M. xanthioides* were nested within *M. bucharica* (Regel) Kruganova and *M. triphylla* (Fisch. & C.A.Mey.) Fisch. & C.A.Mey., respectively. This subclade also contained a monophyletic group [PP = 1, LBS = 100%; corresponding to the "GAU clade" in Duan et al. (2020), also see Fig. 2A] that consisted of widely accepted species *G. aspera*, *G. glabra*, *G. inflata* and *G. uralensis*, as well as seven dubious species: *G. alaschanica*, *G. glandulifera*, *G. gobica*, *G. iconica*, *G. korshinskyi*, *G. laxissima*, and *G. zaissanica*.

Our cp genome results (Fig. 2A) revealed similar inter-specific relationships with the nuclear tree (Fig. 2B), except that *G. squamulosa* was sister to the *G. echinata-macedonica* clade in the nuclear tree (vs. sister to *G. pallidiflora* in the cp tree). We also observed that species and dubious species (see above) had

unresolved relationships within the GAU clade sensu Duan et al. (2020) in both of our nuclear and cp trees (Fig. 2).

3.2 Automated identification

For the species pair of *Glycyrrhiza echinata-G. macedonica* (Table 2), the machine learning model showed that the prediction accuracy was 91.53% (Sample 1), 97.46% (Sample 2) and 90.68% (Sample 3). Prediction precision, recall rate and F1-score of *G. echinata* images ranged from 88.37% to 100%, from 73.81% to 100%, and from 93.25% to 98.01%, respectively; and those of *G. macedonica* were 87.36%-100%, 76.19%-100% and 86.49%-96.47%, respectively. The results clearly reflected that most images were correctly predicted, and suggest that the model can recognize *G. macedonica* from *G. echinata*. The abovementioned percentual results were calculated based on the confusion matrices for the labels/predictions of images (see Table S4).

3.3 Ancestral state reconstructions

We inferred that the presence of glycyrrhizin evolved twice independently within *Glycyrrhiza*. Our result (Fig. 3) showed that the most recent common ancestor (MRCA) of *Glycyrrhiza*, as well as most of the ancestral nodes of this genus lacked glycyrrhizin (coding: A). Among sampled species, this character only appeared in *G. lepidota* and the enlarged *G. glabra* (coding: B), which are phylogenetically distant from one another.

4. Discussion

4.1 Inter-generic demarcation of *Glycyrrhiza*, *Glycyrrhizopsis* and

Meristotropis*, and species circumscription within *Glycyrrhizopsis

Our phylogenetic results (Fig. 2) indicated a sister relationship between the tribes Glycyrrhizeae and Wisterieae, and these, in turn, formed a clade that is sister to the rest of the IRLC (also see Duan et al., 2020). Within Glycyrrhizeae, *Meristotropis* was formerly established from *Glycyrrhiza* based mainly on its trifoliate leaflets (Fischer & Meyer, 1843; Kruganova, 1955). However, it has since been observed that number of leaflets (per leaf) is not a stable morphological feature for a rigorous taxonomic treatment in Glycyrrhizeae, e.g., *Glycyrrhiza inflata* is occasionally trifoliate (Li and Cui, 1998; Bao & Larsen, 2010), while leaves of *Meristotropis bucharica* and *M. triphylla* rarely have five leaflets (Grigorev & Vasilchenko, 1948; Kruganova, 1955; Meng, 2005). Our recent cp phylogenomic analyses resolved taxa of *Meristotropis* into two non-sister clades, both of which were well nested within *Glycyrrhiza* (Duan et al., 2020), and the results were supported by our present nuclear and cp trees (Fig. 2). We thus follow the view of most taxonomists who have treated *Meristotropis* within *Glycyrrhiza* (Boisser, 1856; Engler & Prantl, 1894; Li, 1963; Meng, 2005; Li & Lu, 2015; Öztürk et al., 2017). Additionally, we agree with Yakovlev et al. (1996), Meng (2005) and Li & Lu (2015) in treating *Meristotropis kulabensis* and *M. xanthioides* as synonyms of *Glycyrrhiza bucharica* (= *Meristotropis bucharica*) and *G. triphylla* (= *M. triphylla*), respectively (see the key in Taxonomic treatment).

Historically, *Glycyrrhizopsis* (Fig. 1A-C) was distinguished from *Glycyrrhiza*

mainly by an aglandular plant with compressed, oblong-rhombic, glabrous, dehiscent legumes (Boisser, 1856; Kruganova, 1955). More recently, *Glycyrrhiza* and *Glycyrrhizopsis* were regarded as independent by Meng (2005) and Meng & Zhu (2010) based on several additional aspects of morphology and palynology. Herein, our nuclear and cp phylogenetic analyses indicated the monophyly of *Glycyrrhizopsis* and its sister relationship with *Glycyrrhiza* (Fig. 2), and our recent spatiotemporal evolutionary study suggested that the ancestors of the two genera diverged nearly 17 million years ago during the Miocene (Duan et al., 2020). Therefore, based on our prior and present results, as well as the morphological difference, we support the generic status of *Glycyrrhizopsis*.

Within *Glycyrrhizopsis*, our sampling comprised two accessions of *Glycyrrhizopsis flavescens* (Fig. 1A-B), which is a widely accepted species included by most taxonomic revisions of Glycyrrhizeae (see Table 1). It is noticeable that this species was sister to *Glycyrrhiza asymmetrica* in our cp and nuclear trees (Fig. 2; also see Duan et al., 2020). The two taxa shared dense rhabdo-raceme with bright yellow flowers, as well as compressed, oblong-rhombic, dehiscent fruits (Huber-Morath, 1965; Chamberlain, 1970). *G. asymmetrica* should, therefore, be transferred in *Glycyrrhizopsis* to ensure the monophyly of the genus *Glycyrrhiza* (see Taxonomic treatment). These two taxa appear to be independent based on their differing base chromosome numbers, which is $x = 8$ for *G. flavescens* and $x = 7$ for *G. asymmetrica* (Çetin, 2015). Additionally, *G. flavescens* and *G. asymmetrica* can be told apart by their

aglandular or glandularly bristled plants, respectively (Huber-Morath, 1965; Chamberlain, 1970; Çetin, 2015). It may be reasonable to treat them as two segregated species as in former revisions (Chamberlain, 1970; Öztürk et al., 2017).

On the other hand, *Glycyrrhizopsis syriaca* (Fig. 1C) was erected from *G. flavescens* by Turrill (1937) on the basis of stenophyllous (oblong to oblanceolate vs. obovate to elliptic), longer (13-28 mm vs. 9-15 mm) leaflets and slimmer (linear vs. lanceolate), longer (ca. 11 mm vs. 4-7 mm) bractlets, which was supported by Meng (2005). There is few specimen or witness record of *G. syriaca* except for the holotype (G. P. Baker s.n., K!), and due to the tense political situation in its distribution from northern Syria, we can barely collect molecular material for the present phylogenetic analyses. However, by examining plenty of *G. flavescens* specimens, we argued that the morphological variation range of *G. syriaca* fits into that of *G. flavescens*. Besides, the distribution of the former well embedded within that of the latter (Meng, 2005). We thus treat *G. syriaca* as a synonym of *G. flavescens*, consistent with Çetin (2015), but its status needs to be further verified with field work and molecular phylogenetic evidence. An additional dubious taxon in *Glycyrrhizopsis* is *G. flavescens* ssp. *antalyensis* Sümbül, Ö.Tufan, O.D.Düşen & R.S.Göktürk (Sümbül et al., 2003), which also merits further taxonomic investigation due to the lack of material for molecular analysis or high resolution images for use in this study.

4.2 Species delimitation in *Glycyrrhiza*

Previously, we built up an infra-generic phylogenetic framework for *Glycyrrhiza* based on cp genomic data (Duan et al., 2020), while as stated above, cp genes have a shortage of uniparental transmission, which may lead to inaccurate phylogenetic relationships. Thus we used a new bioinformatic method (Liu et al., 2021) to selected LCN genes (see Materials and methods), and concatenated them with nrDNA sequences to construct nuclear trees (Fig. 2B), which were applied to complement our cp trees herein (Fig. 2A).

Noteworthily, topology of the nuclear trees was not completely accordant with that of the cp trees, and the discordance laid in the *G. echinata-macedonica-squamulosa-pallidiflora* clade and the GAU clade sensu Duan et al. (2020), i.e., the medicinal group, respectively. Both of the two sets of topological differences were not "hard" incongruence (as described in Mort et al., 2008; Peterson et al., 2015; Spooner et al., 2017), namely the target "incongruent" topologies were only weakly supported, implying the possible invalidation of such incongruence and its insignificant for further discussion. Additionally, for the *G. echinata-macedonica-squamulosa-pallidiflora* clade, the incongruence occurred among deep-branches, which did not affect our monophyly-based species boundary test at a shallower level. We assumed that the weak supports to the two sets of incongruent topologies are probably caused by lacking informative sites in our nuclear data, deeper genomic sequencing is required in order to produce adequate LCN genes in the future. On the other hand, the insignificantly supported, inconsistent topology within the medicinal group may also result from its recent

rapid radiation (see Duan et al., 2020), namely, the genomes of the accessions have not yet sufficiently diverged to resolve clear subcaldes within the group.

The New World species, *Glycyrrhiza lepidota* and *G. astragalina*, comprised a clade in our phylogenetic tree (Fig. 2). these species are endemic to North America (temperate central and western U.S.A. and southwestern Canada; Britton & Brown, 1897; Scoggan, 1978) and South America (around 40°S in Argentina and Chile; Reiche, 1898; Gómer-Sosa, 1999), respectively, while all other *Glycyrrhiza* taxa are disjunctively distributed in the Old World. In addition, *G. lepidota* can be recognized from *G. astragalina* by possessing densely uncinate glandular hairy fruits (vs. glabrous or sparsely glandular fruits of *G. astragalina*). Nearly all of the major, previous revisions of *Glycyrrhiza* accepted these two species (Kruganova, 1955; Li, 1963; Meng, 2005; Grankina, 2008; Li & Lu, 2015; also see Table 1), and this concurs with our present (Fig. 2) and previous phylogenetic results (Duan et al., 2020).

Glycyrrhiza acanthocarpa is the only species of the genus recorded from Australia, where it occurs in the states of Queensland, New South Wales, Victoria, South Australia, and Western Australia (Stanley & Ross, 1983; Weber, 1986; Gardner, 1991; Jeanes, 1996; Western Australian Herbarium, 1998). This species is sister to a clade comprised of five other species (Fig. 2), *G. echinata*, *G. macedonica*, *G. pallidiflora*, *G. squamulosa* and *G. yunnanensis*, which occur in the temperate Old World (Kruganova, 1955; Meng, 2005). The isolated geographic distribution of *G. acanthocarpa* in Australia and distinct phylogenetic

position lends support to this species, which is also accepted in all other major treatments (see Table 1). Similarly, the well-accepted species, *G. squamulosa*, is supported by our results, in which all its accessions form a fully supported clade (Fig. 2). *Glycyrrhiza squamulosa* is morphologically recognizable by having unique globose/reniform, tuberculate pods (Meng, 2005; Bao & Larsen, 2010). We thus agree on previous taxonomic treatments (Table 1) that it merits species status.

Formerly, some taxonomists treated the dubious species *Glycyrrhiza macedonica* as a synonym of the widely accepted species *G. echinata* according to their morphological resemblance (Yeo, 1968; Meng, 2005), and the present phylogenetic analyses resolved a sister relationship between the two taxa (Fig. 2). However, with a single accession of *Glycyrrhiza macedonica*, we cannot verify the species/synonym status of *G. macedonica* solely based on the trees (also see explanation in 2.4 above). The MLA based on high resolution specimen images supports the independence of these species (Table 2), and we took this perspective in our taxonomic treatment, which have preferred by most taxonomists on the ground of inflorescence shapes (oblong for *G. macedonica* vs. subglobose for *G. echinata*) and legume prickles (legume of *G. macedonica* besetting with prickles vs. without prickle in lower part of legume in *G. echinata*) (Kruganova, 1955; Li, 1963; Grigorev & Vasilchenko, 1948; Rechinger, 1984; Li & Lu, 2015).

Glycyrrhiza pallidiflora and *G. yunnanensis* are also morphologically similar, such that Meng (2005) treated the latter as a synonym of the former. However,

both of our analyses of cp genomes (Fig. 2A) and nuclear sequences (Fig. 2B) resolved a distant related relationship between the two well supported species, and they have non-overlapping distributions in northern China-Russian Far East (*G. pallidiflora*; Grigorev & Vasilchenko, 1948; Bao & Larsen, 2010) and southwestern China (*G. yunnanensis*; Yunnan and Sichuan; Bao & Larsen, 2010), respectively. Besides, *G. pallidiflora* and *G. yunnanensis* are distinguishable mainly by shapes of inflorescence and infructescence (prolate vs. subglobose), as well as fruit shape (ovoid vs. prolate) (Grankina, 2008; Bao & Larsen, 2010; also see the key in Taxonomic Treatment). Thus, on the basis of our molecular phylogeny, morphological variations and the geographic distribution, we agree with most of former treatments, which segregated *G. yunnanensis* from *G. pallidiflora* (Li, 1963; Li and Cui, 1998; Grankina, 2008; Bao & Larsen, 2010; Li & Lu, 2015; also see Taxonomic Treatment).

In our phylogenetic trees (Fig. 2), the trifoliate species, *Glycyrrhiza bucharica* (rarely 5-foliolate) and *G. triphylla*, are sister to *G. gontscharovii* [(5)-7-9-foliolate] and *G. foetida* Desf. (9-11-foliolate), respectively. Between these pairs, *Glycyrrhiza triphylla* and *G. foetida* have long been regarded as separate species (Li, 1965; Meng, 2005; Li & Lu, 2015), and they are clearly distinct based on morphology (see the Key in Taxonomic treatment) and phylogeny (Fig. 2). Additionally, they are geographically disjunct in Central Asia (*G. triphylla*; Yakovlev et al., 1996; Meng, 2005) and North Africa and southern Spain (*G. foetida*; Yeo, 1968; Lock, 1989; Mertín, 1999). In contrast, fewer

taxonomists have supported the species status of *G. gontscharovii* (Grigorev & Vasilchenko, 1948; Yakovlev et al., 1996; Öztürk et al., 2017). However, we believe it may be a valid species based on its number of leaflets (per leaf) compared to *Glycyrrhiza bucharica* and its phylogenetic position outside of *G. bucharica* clade. Thus, we recognize all four species, *G. bucharica*, *G. foetida*, *G. gontscharovii* and *G. triphylla*, in this study.

Within the glycyrrhizin-contained medicinal group, which corresponded to the GAU clade in Duan et al. (2020) (also see Fig. 2), we included all four well-accepted medicinal species, i.e., *Glycyrrhiza aspera*, *G. glabra*, *G. inflata* and *G. uralensis*, and seven dubious species in our phylogenetic analyses (see Results). None of the widely-accepted or dubious species were monophyletic in either our cp (Fig. 2A) or nuclear trees (Fig. 2B), except for accessions of *G. inflata*, which formed a clade in the cp tree (Fig. 2A). The chaotic taxonomy of this group may be due to the widely existed morphological intergradations, e.g., legume curvature and density of glandular hairs. The reasons to the phenomena may lie in the rapid diversification of the group (Duan et al., 2020), as well as many hybridization events among morphologically or ecologically distinct populations that yielded fertile progeny (Ashurmetov, 1996; Zhang et al., 1998; Zimnitskaya, 2009; Xie et al., 2014; Chen et al., 2017). Here, we treat the medicinal group (the GAU clade) as a single, morphologically diverse species, *Glycyrrhiza glabra*, which was established by Linnaeus (1753) (see Taxonomic treatment).

Pharmacopoeia of the People's Republic of China (Chinese Pharmacopoeia Commission, 2015) recognized three medicinal species: *Glycyrrhiza glabra* s.s., *G. inflata* and *G. uralensis*. Few study reported significant content differences of glycyrrhizin among the three former species, however, *G. inflata* and *G. uralensis*, in most cases, are distinguishable from *G. glabra* s.s. by their inflated and zigzag fruits, respectively, although there is morphological intergradation phenomenon to some extent. Besides, both of *G. inflata* and *G. uralensis* are commonly used medicinal liquorice taxa and were widely cultivated in Asia (Chinese Pharmacopoeia Commission, 2015; Li & Lu, 2015; Öztürk et al., 2017), it might be prudent to respectively treat these two as two varieties within the broadly defined *G. glabra* (see Taxonomic treatment and Fig. 4) to facilitate sustainable utilization and application of their potential resources. Another previously accepted glycyrrhizin-contained species, *G. aspera* (Li, 1963; Li & Lu, 2015), can be easily recognized from former *G. glabra* s.s., *G. inflata* and *G. uralensis* by being a much shorter plant, by having slender roots and moniliform, and glabrous legumes (Meng, 2005; Bao & Larsen, 2010). Thus, we recognized this former species as a variety as well (Fig. 4). Whereas, due to low yield of its slender, shorter roots, *G. glabra* var. *aspera* is always regarded as unqualified medicinal or industrial materials, even though the variety certainly contains glycyrrhizin (Li, 1963; Meng, 2005; Li & Lu, 2015).

4.3 Trait evolution of glycyrrhizin in *Glycyrrhiza*

The chemical constituent glycyrrhizin is specific to *Glycyrrhiza* based on current

understanding (Li & Lu, 2015; Zhou & Jin, 2016; Öztürk et al., 2017), and it has highly taxonomic, medicinal and industrial value (see Introduction above), yet few worker studied glycyrrhizin in a evolutionary aspect. Our ancestral state inference (Fig. 3) demonstrates that glycyrrhizin is absent in the MRCA of *Glycyrrhiza*, but it appears in two non-sister extant species of the genus: in the broadly defined *G. glabra* (i.e., the medicinal group), and at low concentrations in the North American *G. lepidota* (Hayashi et al., 2005; Li & Lu, 2015). We proposed two alternative hypotheses to explain the result: 1. Our analysis was in line with the evolutionary history of *Glycyrrhiza*, and Glycyrrhizin independently evolved twice from the glycyrrhizin-free *Glycyrrhiza* ancestor; or 2. Glycyrrhizin production may be the ancestral state of liquorice, but extinctions of some key ancestral taxa, especially early diverging species in the genus, may have misled our analysis to produce an inaccurate ancestral trait for the MRCA. In most cases, hypothesis #2 would require many losses of glycyrrhizin production, which is less likely according to the principle of parsimony: an occurrence, phenomenon, or event is the simplest, involving the fewest entities, assumptions, or changes (Sober, 1981; Hine, 2019). In addition, convergent evolution of secondary metabolites is not rare in plants (Pichersky & Lewinsohn, 2011; Zhang et al., 2021), thus, hypothesis #1 seems more plausible.

Glycyrrhizin is of interest for its utility for humans, while in liquorice plants it may serve to promote drought resistance (Nasrollahi et al., 2014; Hosseini et al., 2018, 2020; Xie et al., 2018). However, *G. lepidota* usually grows on moist soil

from riverbanks, open fields, prairies and roadsides (Turner, 1959), albeit in drier regions within the USA and Canada, broadly speaking (e.g., excluding the moist, humid southeast). Thus, the key environmental factor that drove the emergence of glycyrrhizin remains unclear, and production of glycyrrhizin may have facilitated survival of species under dry conditions but have arisen for other reasons. On the other hand, liquorice plant synthesizes glycyrrhizin mainly through the mevalonic acid (MVA) pathway and is regulated by various enzymes (Hayashi et al., 2003; Seki et al., 2008, 2011), and yields many pharmacologically active products of the triterpenoid saponin class, of which glycyrrhizin is a part. Most of the key genes involved in the biosynthesis of glycyrrhizin have been discovered (Li et al., 2017; Mochida et al., 2017), however, further comparative genomic and systems biology studies are still required to elucidate the gene mutations and biochemical changes that trigger the synthesis of glycyrrhizin.

5. Taxonomic treatment

5.1 Key to Glycyrrhizeae

1a. Legume compressed, oblong-rhombic, glabrous, dehiscent; endemic to southern Anatolian plateau.....2

(***Glycyrrhizopsis***)

1b. Legume compressed or inflated, rarely moniliform, ovoid to linear, densely tuberculate, prickly, glandular punctate/hairy, sparsely hairy or glabrous, indehiscent, rarely slightly dehiscent; widely distributed in every continent except for Antarctica.....3

(***Glycyrrhiza***)

2a. Plant glandular, densely bristled; leave 5-9-foliolate; legume simple bristled with thick secretion; 2*n* = 14.....***Glycyrrhizopsis asymmetrica***

2b. Plant aglandular, glabrous, or sparsely hairy; leave 9-19-foliolate; legume glabrous; 2*n* = 16.....***Glycyrrhizopsis flavescens***

3a. Leaf trifoliate, rarely 5-foliolate.....4

3b. Leaf imparipinnate, 5-23-foliolate.....5

4a. Leaflet lanceolate, 20-50 mm; corolla white; legume compressed or slightly inflated, glandular, surface viscid, sometimes sparsely bristly...***Glycyrrhiza bucharica***

4b. Leaflet obovate, 9-15 mm; corolla yellow; legume inflated, densely glandular punctate and glandular hairy.....***Glycyrrhiza triphylla***

5a. Leaflet elliptic to lanceolate; legume globose to ovoid, densely tuberculate or glandular punctate/hairy; seed 2-5; mostly lacking glycyrrhizin.....6

5b. Leaflet ovate to oblong; legume linear to prolate, glandular punctate/hairy, prickly or glabrous, rarely tuberculate; seed 2-11; containing

glycyrrhizin.....15 (***Glycyrrhiza glabra***)

6a. Leaflet oblong to oblong-lanceolate, apex rounded, usually retuse; legume globose to prolate, tuberculate or glandular punctate.....7

6b. Leaflet lanceolate or oblanceolate, apex acuminate; legume globose to ovoid, densely glandular hairy, or prickly.....8

7a. Leaf 7-11-foliolate; legume prolate, 11-22 mm, glabrous or sparsely glandular punctate; endemic to temperate South America.....***Glycyrrhiza astragalina***

7b. Leaf 9-13-foliolate; legume globose or reniform, 5-10 mm, tuberculate; endemic to northern China and Mongolia.....***Glycyrrhiza squamulosa***

8a. Leaflet (5)-7-9-foliolate; legume short bristly, rarely glandular punctate.....***Glycyrrhiza gontscharovii***

8b. Leaflet 7-19-foliolate; legume densely glandular hairy.....9

9a. Raceme loose, non-capitate; legume globose, 5-6 mm; endemic to southern and eastern Australia.....***Glycyrrhiza acanthocarpa***

9b. Raceme dense, capitate; legume ovoid to prolate, 10-20 mm; distributed in temperate Asia, eastern and southern Europe, northern Africa or North America.....10

10a. Legume prolate to ovoid-prolate, densely uncinate glandular hairy; endemic to temperate western North America; containing low concentration of glycyrrhizin.....***Glycyrrhiza lepidota***

10b. Legume ovoid to ovoid-prolate, densely straight glandular hairy, distributed in temperate Asia, eastern and southern Europe or northern Africa; lacking glycyrrhizin.....
...11

11a. Plant fetid; endemic to northern Africa and southern Spain.....***Glycyrrhiza foetida***

11b. Plant non-fetid; distributed in northern and southwestern China, Mongolia, Central Asia, Caucasus or eastern Europe.....12

12a. Standard ovate; distributed in China and Russian Far East.....13

12b. Standard oblong to narrowly elliptic; distributed in Central Asia, Caucasus, western Siberia, West Asia or eastern Europe.....14

13a. Inflorescence and infructescence prolate; legume ovoid, beak acute;

distributed in northern and northeastern China and Russian Far East.....***Glycyrrhiza pallidiflora***

13b. Inflorescence and infructescence subglobose; legume prolate, beak cuspidate; endemic to southwestern China.....***Glycyrrhiza yunnanensis***

14a. Inflorescence capitate, often subglobose; legume densely prickly in upper part, non-prickly in lower part.....***Glycyrrhiza echinata***

14b. Inflorescence non-capitate, somewhat prolate; legume uniformly prickly.....***Glycyrrhiza macedonica***

15a. Plant usually shorter than 30 cm; roots and rhizomes slender; legume moniliform, glabrous.....***G. glabra* var. *aspera***

15b. Plant usually longer than 30 cm; roots and rhizomes strong; legume inflated or compressed, non-moniliform, densely tuberculate, glandular punctate/hairy, sparsely hairy or glabrous.....16

16a. Legume inflated.....***G. glabra* var. *inflata***

16b. Legume compressed.....17

17a. Leaflet margin somewhat undulate; legume zigzag, falcate or ring-like.....***G. glabra* var. *uralensis***

17b. Leaflet margin non-undulate; legume straight or slightly curved.....***G. glabra* var. *glabra***

5.2 Species descriptions of *Glycyrrhizopsis* and new combination of *G. asymmetrica*

Glycyrrhizopsis Boiss., Diagn. Pl. Orient. ser. 2(5): 82. 1856. - Type: *Glycyrrhizopsis flavescens* Boiss.

(1) *Glycyrrhizopsis flavescens* Boiss., Diagn. Pl. Orient. ser. 2(5): 82. 1856. ≡ *Glycyrrhiza flavescens* Boiss., Diagn. Pl. Orient. ser. 1(6): 33. 1846. - Type: TURKEY: Cilicia, Adana, Aucher 994 [holotype: K(barcode 000118443)!] (Fig. 1A-B).

= *Glycyrrhizopsis syriaca* Turrill, Bull. Misc. Inform. Kew 1937(2): 79. 1937 (Fig. 1C).

Description: Perennial herb, 30-85 cm, aglandular, glabrous, or sparsely hairy. Leaf imparipinnate. Leaflets 5-8-paired, elliptic. Racemes many-flowered. Calyx c. 3 mm. Corolla golden-yellow. Legume compressed, oblong-rhombic, glabrous, dark brown, valves contorting on dehiscence, several-seeded. 2*n* = 16.

Note: This species distributes in southern Anatolia of Turkey and northern Syria.

(2) *Glycyrrhizopsis asymmetrica* (Hub.-Mor.) L.Duan, **comb. nov.** ≡ *Glycyrrhiza asymmetrica* Hub.-Mor., Bauhinia ii: 301. 1965. - Type: TURKEY: Antalya District, 15 km east of Antalya, 14 May 1956, A. Huber-Morath 13502 [holotype: G(barcode 00414315)!, isotype: E(barcode 00296537)!].

Description: Perennial herb, 30-70 cm, glandular, densely bristled. Leaf imparipinnate. Leaflets 2-4-paired, widely obovate to orbicular-cuneate. Racemes dense. Calyx 3-5 mm. Corolla yellow. Legume dehiscent, oblong, acuminate, compressed, oblong-rhombic, glandularly bristled, 2-seeded. 2*n* = 14.

Note: This species is endumic to southern Anatolia of Turkey.

5.3 Species descriptions of *Glycyrrhiza* and redelimitation of *G. glabra*

Glycyrrhiza L., Sp. Pl. II: 741. 1753. - Type: *Glycyrrhiza glabra* L.

= *Meristotropis* Fisch. & C.A.Mey., Index Seminum [St.Petersburg (Petropolitanus)] ix: 95. 1843.

(1) *Glycyrrhiza lepidota* Pursh, Fl. Amer. Sept. 2: 480. 1813. - Type: U.S.A.: banks of the Missouri, Pursh s.n. (holotype: K).

Description: Perennial herb, 30-90 cm, erect, branching. Leaf imparipinnate. Leaflets 5-9-paired, lanceolate, or oblong. Peduncles much shorter than the leaves. Racemes dense, many-flowered. Calyx teeth longer than the tube. Corolla yellowish-white. Legume indehiscent, globose to ovoid, densely tuberculate or uncinately glandularly hairy, or prickly, 2-5-seeded.

Note: *G. lepidota* is the only species recorded in North America (temperate central and western U.S.A. and southwestern Canada), which is endemic to this

continent. This species contains low concentration of glycyrrhizin.

(2) ***Glycyrrhiza astragalina*** Gillies, Bot. Misc. 3: 183. 1833. - Type: ARGENTINA: Mendoza, valley of Uspallata, Cuming 812 [holotype: K(barcode 000118162)!].

Description: Perennial herb, 50 cm or more, glabrous or sparsely glandular. Leaf imparipinnate. Leaflets 3-7-paired, elliptic to obovate, rachis with tufts of fleshy hairs on petiolule insertion. Calyx 3.5-5 mm. Racemes initially dense, then lax. Corolla blue or violaceous. Legume indehiscent, compressed, sometimes torulose, mucronate, glandular, 3-4-seeded.

Note: This is the only species recorded in South America (around 40°S in Argentina and Chile), which is restricted in this continent.

(3) ***Glycyrrhiza acanthocarpa*** J.M.Black, Trans. & Proc. Roy. Soc. South Australia XLIII: 351. 1919. ≡ *Indigofera acanthocarpa* Lindl., in Mitch. Three Exped. ii: 17. 1839. - Type: AUSTRALIA: Interior of New Holland, T. L. Mitchell s.n. [holotype: K(barcode 000118166)!].

Description: Subshrub to 100 cm tall, erect, glabrous. Leaves imparipinnate. Leaflets 4-11-paired, linear to elliptic or obovate, both surfaces glandular-punctate: stipules 3-5 mm long. Racemes spike-like, exceeding subtending leaves, many-flowered. Calyx 3-4 mm, teeth about equal to tube. Corolla purple. Legume indehiscent, ovoid or globose, rusty-coloured, covered with hard prickles, 2-5-seeded.

Note: This is the only species recorded in Australia (Queensland, New South

Wales, Victoria, South Australia, and Western Australia), which is endemic to the country.

(4) ***Glycyrrhiza echinata*** L., Sp. Pl. II: 741. 1753. - Type: RUSSIA: desert Tatarize, C. Linnaeus 916.1 (lectotype: LINN!) (Fig. 1D-E).

Description: Perennia herb, 50-100 cm, procumbent or ascending, simple or branching in lower part, glabrous or sparsely hairy and glandular in upper part. Leaf imparipinnate. Leaflets 3-9-paired, elliptic or obovate, glandular at lower side. Racemes dense, subglobose. Calyx 3-4(4.5) mm. Corolla intensively violet-blue. Legume indehiscent, ovoid, elliptic or oblong-elliptic, densely prickly in upper part, non-prickly in lower part.

Note: This species widely ranged from southeastern Europe, through West Asia and Caucasus, to Central Asia and Iran.

(5) ***Glycyrrhiza macedonica*** Boiss. & Orph., Bull. Congr. Bot. St. Petersb.: 135. 1870. - Type: GREECE: Macedonia, T. G. Orphanides 427 [holotype: G, isotype: LE(barcode 00014396)!] (Fig. 1F).

Description: Perennial herb, 100-150 cm, usually branching and glabrous. Leaf imparipinnate. Leaflets 3-6-paired, elliptic or obovate, glandular on both sides. Racemes dense, compact, oblong. Calyx 2.5-3 mm, teeth slightly shorter than tube. Corolla pale violet. Legume indehiscent, ovoid or oblong-ovate, valves more or less uniform at surface, beset with reddish prickles, 2–3-seeded..

Note: *G. macedonica* distributes in eastern Europe, West Asia, Caucasus and Central Asia. Its distribution and morphology are somehow similar to *G. echinata*.

(6) ***Glycyrrhiza squamulosa*** Franch., Fl. David. I: 93. 1884. - Type: MONGOLIA: A. David 2902 [holotype: P(barcode 02297603)!, isotype: PE(barcode 01723786)!, LE(barcode 01024784)!, K(barcode 000118167)!].

Description: Perennial herb, 30–60 cm, densely yellow scaly glandular punctate, glabrous or sparsely pubescent. Leaves imparipinnate. Leaflets 4–6-paired, narrowly elliptic to oblong-obovate, densely scaly glandular punctate. Racemes many-flowered. Calyx campanulate, 2.5–3.5 mm. Corolla white. Legume indehiscent, globose to reniform, tuberculate or glandular punctate, apex mucronate. 2-seeded.

Note: This species distributes in northern China and Mongolia, which is recognizable with its globose or reniform pods.

(7) ***Glycyrrhiza pallidiflora*** Maxim., Prim. Fl. Amur.: 79. 1859. - Type: RUSSIA: Amur, C. J. Maximowicz s.n. [holotype: LE!, isotype: PE(barcode 00022349)!].

Description: Perennial herb, 100–150 cm, striped, densely yellow-brown scaly glandular punctate, nearly glabrous. Leaves imparipinnate. leaflets 4–7-paired, lanceolate or ovate-lanceolate, glabrous. Racemes many-flowered, oblong or globose. Calyx campanulate, 4–5 mm. Corolla purple, or purple-red. Infructescence prolate. Legume indehiscent, ovoid, rigidly spiny, apex abruptly acuminate, 2-seeded.

Note: *G. pallidiflora* mainly grows in North and Northeastern China, sparsely distributes in Russian Far East.

(8) ***Glycyrrhiza yunnanensis*** Cheng f. & L.K.Tai ex P.C.Li, Acta Bot.

Boreal.-Occid. Sin. 4(2): 117. 1984. - Type: CHINA: Yunnan, Lijiang, C. W. Wang 71519 [holotype: PE(barcode 01432591)!].

Description: Perennial herb, 60–100(–120) cm, densely scaly glandular punctate. Leaves imparipinnate. Leaflets 3–7-paired, lanceolate or ovate-lanceolate, densely scaly glandular punctate, sparsely pubescent. Racemes many-flowered. Calyx campanulate, ca. 5 mm. Corolla purple. Infructescence subglobose. Legume dense, indehiscent, prolate, densely spiny, apex cuspidate. Seeds brown.

Note: *G. yunnanensis* is endemic to southwestern China (provinces of Yunnan and Sichuan). This species resembles *G. pallidiflora* in morphology, but they are recognizable from shapes of infructescence and fruit, and their distributional areas are geographically discrete.

(9) *Glycyrrhiza gontscharovii* Maslenn., Trudy Tadzhikistansk. Bazy 8: 617. 1940. ≡ *Meristotropis bucharica* f. *gontscharovii* (Maslenn.) Malzeva, Bot. Mater. Gerb. Bot. Inst. Bot. Acad. Nauk Kazakhsk. S.S.R. 9: 57. 1975. - Type: TAJIKISTAN: Shuro-obod District, J. Linczevski & T. Maslennikova 517 (holotype: LE!, isotype: TASH!).

Description: Perennial herb, 70–100 cm, branching, glabrous, with scattered glands. Leaves imparipinnate. Leaflets 2–4-paired, oblong-ovate or oblong-lanceolate, densely glandular. Peduncles much shorter than leaves. Racemes loose, many-flowered. Calyx campanulate, 3–4 mm. Corolla whitish-yellowish with faint lilac tint. Legume indehiscent, oval, attenuate-

acuminate, short bristly, rarely glandular punctate. Seeds dark gray-green.

Note: This species is restricted in Central Asia.

(10) ***Glycyrrhiza bucharica*** Regel, Trudy Imp. S.-Peterburgsk. Bot. Sada viii: 697. 1884. ≡ *Meristotropis bucharica* (Regel) Kruganova, Trudy Bot. Inst. Akad. Nauk S.S.S.R., Ser. 1, Fl. Sist. Vyssh. Rast. 11: 194. 1955. - Type: TURKMENISTAN: A. Regel 88 (lectotype: LE!).

= *Meristotropis kulabensis* Masl., Trudy Bot. Inst. Akad. Nauk S.S.S.R., 8: 620. 1938.

Description: Perennial herb, 80–100 cm, firm, straight, usually branching, more or less glandular. Leaf trifoliate, rarely 5-foliolate. Leaflet elongate or oblong-lanceolate. Racemes rather poor. Calyx campanulate, 2.5–3 mm. Corolla white. Legume dehiscent (valves white inside), erect, compressed or slightly inflated, glandular, surface viscid, sometimes sparsely bristly with reddish glands, 1–2-seeded.

Note: This species is endemic to Central Asia.

(11) ***Glycyrrhiza foetida*** Desf., Fl. Atlant. 2: 170, t. 199. 1799. - Type: ATLAS MOUNTAINS: M. Desfontaines s.n. [holotype: P(barcode 00288681)!].

Description: Perennial herb, 25-50 cm, fetid. Leaf imparipinnate. Leaflets 4-5-paired, obovate, elliptical or ovate-lanceolate, mucronate. Racemes about equalling the leaves, dense or lax. Corolla pale yellow. Legume indehiscent, fusiform, densely covered with glandular straight bristles and with sessile and short-stalked glands, 2-3-seeded.

Note: Plant of *G. foetida* is malodorous, which are confined in western Mediterranean region (southern Spain and northwestern Africa).

(12) *Glycyrrhiza triphylla* Fisch. & C.A.Mey., Linnaea 10 (Lit.): 91. 1835. ≡ *Meristotropis triphylla* Fisch. & C.A.Mey., Index Seminum [St.Petersburg (Petropolitanus)] ix: 96. 1843. - Type: KAZAKHSTAN: Mangyshlak mountain, s.n. (lectotype: LE!).

= *Meristotropis xanthioides* Vassilcz., Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 11: 120. 1949.

Description: Perennial herb, 20–40 cm, branching from the woody base, beset with scattered glands; Leaf trifoliate, rarely 5-foliolate. Leaflet obovate. Racemes loose, elongate, with discrete flowers. Calyx campanulate, 3 mm. Corolla white-yellow (dry). Legume indehiscent, oval, inflated, sparsely covered with thin prickles beset on rounded glandular punctation, greenish-brownish, sometimes faintly pink, 2–4-seeded.

Note: This species is endemic to Central Asia.

(13) *Glycyrrhiza glabra* L., Sp. Pl. II: 742. 1753. - Type: SWEDEN: Uppsala (cultivated), C. Linnaeus 916.3 [lectotype designed by Ali (1977): LINN!].

= *Glycyrrhiza alaschanica* Grankina, Novosti Sist. Vyssh. Rast. 33: 145. 2001.

= *Glycyrrhiza glandulifera* Waldst. & Kit., Descr. Icon. Pl. Rar. Hung. 1: 20, pl. 21. 1800.

= *Glycyrrhiza gobica* Grankina, Novosti Sist. Vyssh. Rast. 33: 147. 2001.

= *Glycyrrhiza iconica* Hub.-Mor., Bauhinia ii. 302. 1965.

= *Glycyrrhiza korshinskyi* Grig., Izv. Glavn. Bot. Sada S.S.S.R. 29: 94. 1930.

= *Glycyrrhiza laxissima* Vassilcz., Bot. Mater. Gerb. Bot. Inst. Komarova Akad. Nauk S.S.S.R. 11: 120. 1949.

= *Glycyrrhiza zaissanica* Serg., Sist. Zametki Mater. Gerb. Krylova Tomsk. Gosud. Univ. Kuybysheva 1-2: 11. 1933.

Description: Perennial herbs. Roots and rhizomes strong or slender (var. *aspera*). Stem 10-150 cm tall, usually woody at base, with glandular punctation/hair, prickle, or white/brown tomentum. Leaves 2.5-22 cm, (3)-5-17-foliolate; stipules usually caducous, linear, or triangular-lanceolate or ovate-triangular, 4-6 × 2-4 mm, rarely ca. 1 mm in length (var. *inflata*); petiole yellow-brown glandular punctate/hairy and villous; leaflets ovate, oblong, or elliptic, 10-60 × 3-30 mm, abaxially glandular punctate/hair, adaxially glabrous, glabrescent or pilose, base rounded, rarely cuneate (var. *aspera*), margin undulate or not, apex rounded, obtuse, or retuse and with mucro. Racemes many-flowered; rachis densely hairy or glandular punctate, rarely sparsely glandular, prickly, hairy; bracts linear-lanceolate to oblong-lanceolate, 2-6 mm, membranous, grandular or/and pubescent. Calyx campanulate or cylindric, 5-14 mm, glandular punctate and pubescent, 5-toothed; upper two teeth somewhat joined. Corolla purple, white or yellow, 10-16 mm; standard ovate to narrowly elliptic, 6-15 × 4-6.5 mm, base narrowed to clawed, apex retuse or rounded; wings 6-14 mm; keel 7-11 mm. Ovary densely glandular/puberulent or glabrescent. Legume indehiscent, prolate,

ellipsoid to moniliform, straight, falcate, zigzag, or curved in to a ring, compressed or flat, 8-35 mm, somewhat constricted between seeds, densely tuberculate, glandular punctate/hairy, sparsely hairy or glabrous. Seeds 1-11, green to brown, globose or reniform, 2-3 mm in diameter. Fl. May-Aug, fr. Jul-Oct. 2n = 16.

Note: *Glycyrrhiza glabra* is a morphologically diverse species, adapting to a vast range of habitats in the northern temperate zone of the Old World, which is widely used in traditional medicine. Roots and rhizomes of the species contain glycyrrhizin, a natural sweetener that is widely used in modern manufacture of pharmaceuticals, cosmetics, popular confectionaries, cigarettes, etc. Wild resources of *G. glabra* faces exhaustion due to over-collection, and three of the four varieties (except for var. *aspera*) are listed on the Red Book of Chinese Rare and Endangered Plants (http://www.iplant.cn/rep/).

Glycyrrhiza glabra* var. *glabra (Fig. 4A-B)

Note: *G. glabra* var. *glabra* is a morphologically diverse variety. Its roots and rhizomes are strong; leaflet glandular, ovate or elliptic, not undulate; legume prolate, straight or slightly curved, glabrous or with glandular prickles. This variety adapts to a broad range of habitats including Meadows, riparian woodlands and solonetzic steppes, riverbanks, even saline areas, and it widely distributed in northwest China, Mongolia, Siberia, Central Asia, northwestern South Asia, West Asia, Caucasus, eastern and southern Europe and northern Africa. This variety is usually used in traditional medicine and as an industrial raw material.

Glycyrrhiza glabra* var. *aspera (Pall.) L.Duan, **comb. et stat. nov.** ≡ *Glycyrrhiza aspera* Pall., Reise Russ. Reich. 1: 499. 1771. - Type: Missing collection locality information, C. Linnaeus 916.5 [lectotype designed by Grankina (2008): LINN!] (Fig. 4C-D).

Note: Roots and rhizomes of *G. glabra* var. *aspera* are slender; stem short, erect or diffuse; leaflet aglandular, not undulate; legume moniliform, glabrous. This variety grows in dry steppes, semideserts, desert oases, margins of farms, riverbanks or wastelands, distributing in northwest China, Mongolia, Central and West Asia, Caucasus and eastern Europe. This variety contains glycyrrhizin, but its usage is limited by its overall small size and slender root.

Glycyrrhiza glabra* var. *inflata (Batalin) L.Duan, **comb. et stat. nov.** ≡ *Glycyrrhiza inflata* Batalin, Trudy Imp. S.-Peterburgsk. Bot. Sada. Acta Hort. Petrop. St. Petersburg xi: 484. 1891. - Type: CHINA: Xinjiang, Kashgar, Karate, 9 Jul 1889, W. J. Roborowsky s.n. [Lectotype designed by Grankina (2008): LE(barcode 1024780)!] (Fig. 4E-F).

Note: Roots and rhizomes of *G. glabra* var. *inflata* are strong; leaflet glandular, ovate or elliptic, margin somewhat undulate; legume inflated, straight, glabrous or with sparse brown glandular punctation. Habitats of the variety cover riverbanks, dry steppes, semideserts, desert, margins of farms and wastelands, ranging in Central Asia and Gansu, Xinjiang and Inner Mongolia of China. This variety can be used as medicinal plant and manufactured material.

Glycyrrhiza glabra* var. *uralensis (Fisch. ex DC.) L.Duan, **comb. et stat. nov.** ≡

Glycyrrhiza uralensis Fisch. ex DC., Prodr. [A. P. de Candolle] 2: 248. 1825. - Type: RUSSIA: Mt. Ural, Helm. s.n. (holotype: PE!, isotype: LE!) (Fig. 4G-H).

Note: *G. glabra* var. *uralensis* is morphologically diverse, but is recognizable by its undulate leaflets and zigzag, falcate or ring-like (rarely linear) legumes which have dense tubercles, glands, glandular prickles, aglandular prickles or glandular hair. This variety is widely cultivated in Asia, and it is the most common source of liquorice for traditional medicine and industrial materials.

6. Conclusions

Liquorice is a tribe of leguminous herbs with great medicinal and economic importance, previously including the genera *Glycyrrhiza*, *Glycyrrhizopsis* and *Meristotropis*. By integrating phylogenomic evidence, a machine learning approach (MLA), and morphology, as well as based on an extensive sampling, *Glycyrrhizopsis* is treated as a distinct genus while *Meristotropis* is merged into *Glycyrrhiza*. Two and 13 species are recognized within *Glycyrrhizopsis* and *Glycyrrhiza*, respectively. The glycyrrhizin-contained medicinal group of liquorice is treated as an enlarged *Glycyrrhiza glabra*. Furthermore, using character state reconstruction, we inferred that production of glycyrrhizin evolved twice independently within *Glycyrrhiza* during the last one million years, which involved convergent evolution. The ancestral state estimation deepens our understanding towards the evolutionary history of liquorice, but also raises questions on the interaction between *Glycyrrhiza* plants and the environment. We have extracted low-copy nuclear genes from genome skimming data to reconstruct

the phylogenetic trees, further validate the practicability of the new bioinformatic approach. Moreover, we present a novel framework for integration of phylogenomics and MLAs to resolve difficult taxonomic questions.

Acknowledgements

We are grateful to the curators of the following herbaria for providing specimen photos, leaf samples, seeds, or DNA samples: A, CSH, E, G, IBSC, K, KNYA, KUN, LE, LINN, MA, MO, P, PE, TASH, US, WUK and XJBI. We thank Chun Su, Gabriel Johnson, Jing Liu and Qin-Qin Li for assistance in the lab; Robin Everly for literature searches; and Cheng Du, Imalka Kahandawala, Jia-Hui Lu, Jing-Zhi Gong, Kathryn Richardson, Liang Zhao, Pei-Liang Liu, Shi-Min Duan, Virginia Valcárcel, Wei Shi, Zhao-Yang Chang, Zhen-Hai Wu and Zhuo Zhou for sample collecting. This study was supported by grants from Natural Science Foundation of China (Grant no. 32070229, 31600162), Key Research and Development Program of Guangdong Province, China (Grant no. 2020B1111530004), research project of the Komarov Botanical Institute of the Russian Academy of Sciences (Grant no. AAAA-A19-119031290052-1) and the Laboratories of Analytical Biology of the National Museum of Natural History (Smithsonian Institution).

References

Ali SI. 1977. *Glycyrrhiza* L. In: Nasir E, Ali SI eds. *Flora of West Pakistan*. Karachi: University of Karachi. 100: 94-97.

Altay V, Karahan F, Öztürk M, Hakeem KR, Ilhan E, Erayman M. 2016.

Molecular and ecological investigations on the wild populations of *Glycyrrhiza* L. taxa distributed in the East Mediterranean Area of Turkey. *Journal of Plant Research* 129: 1021-1032.

Ashurmetov OA. 1996. Selection of parental pairs for obtaining hybrids in the genera *Glycyrrhiza* L. and *Meristotropis* Fisch. et Mey. *Genetic Resources and Crop Evolution* 43: 167-171.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455-477.

Bao BJ, Larsen K. 2010. *Glycyrrhiza*. In: Wu ZY, Hong DY, Raven PH eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 10: 509-511.

Boissier PE. 1856. *Glycyrrhizopsis*. In: *Diagnoses Plantarum Orientalium Novarum*. Geneva: Typis Ramboz et Schuchardt. II(5): 81-83.

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

Bonnet P, Goëau H, Hang ST, Lasseck M, Šulc M, Malécot V, Jauzein P, Melet JC, You C, Joly A. 2018. Plant identification: experts vs. machines in the era of deep learning. In: Joly A, Vrochidis S, Karatzas K, Karppinen A, Bonnet P eds. *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Cham: Springer. 131-149.

Britton NL, Brown HA. 1897. *Glycyrrhiza* L. In: *An Illustrated Flora of the Northern United States, Canada and the British Possessions*. New York: Charles Scribner's sons. 2: 310.

Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017. Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology* 17: 181.

Carranza-Rojas J, Joly A, Goëau H, Mata-Montero E, Bonnet P. 2018a. Automated identification of herbarium specimens at different taxonomic levels. In: Joly A, Vrochidis S, Karatzas K, Karppinen A, Bonnet P eds. *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Cham: Springer. 151-167.

Carranza-Rojas J, Mata-Montero E, Goeau H. 2018b. Hidden biases in automated image-based plant identification. In: 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI). 1-9.

Çetin Ö. 2015. *Revision of the genus* Glycyrrhiza *L. distributed in Turkey*. Ph.D. Dissertation. Konya: Selçuk University.

Chamberlain DF. 1970. *Glycyrrhiza* L. In: Davis PH ed. *Flora of Turkey and East Aegean Islands*. Edinburgh: Edinburgh University Press. 3: 260-263.

Chen C, Lu J, Li X, Zhou L, Xie L, Li X, Song F. 2017. Inheritance analysis and discovery of chloroplast paternal inheritance in interspecific crossing of *Glycyrrhiza*. *Guihaia* 37: 162-168.

Chinese Pharmacopoeia Commission. 2015. Gancao. In: *Pharmacopoeia of the*

People's Republic of China. Beijing: China Medical Science Press. 1: 86-87.

Compton JA, Schrire BD, Könyves K, Forest F, Malakasi P, Mattapha S, Sirichamorn Y. 2019. The *Callerya* Group redefined and Tribe Wisterieae (Fabaceae) emended based on morphology and data from nuclear and chloroplast DNA sequences. *PhytoKeys* 125: 1-112.

Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75: 1443-1458.

da Cruz DT, Idárraga Á, Banda K, Cogollo Á, van den Berg C, de Queiroz LP, Pennington RT, Lavin M, Cardoso DBOS. 2018. Ancient speciation of the papilionoid legume *Luetzelburgia jacana*, a newly discovered species in an inter-Andean seasonally dry valley of Colombia. *Taxon* 67: 931-943.

Dastagir G, Rizvi MA. 2016. *Glycyrrhiza glabra* L.(Liquorice). *Pakistan Journal of Pharmaceutical Sciences* 29: 1727-1733.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Duan L, Harris AJ, Su C, Zhang ZR, Arslan E, Ertuğrul K, Lôc PK, Hayashi H, Wen J, Chen HF. 2020. Chloroplast phylogenomics reveals the intercontinental biogeographic history of the liquorice genus (Leguminosae: *Glycyrrhiza*). *Frontiers in Plant Science* 11: 793.

Duan L, Harris AJ, Ye W, Deng SW, Song ZQ, Chen HF, Wen J. 2019. Untangling the taxonomy of the *Cladrastis* clade (Leguminosae: Papilionoideae) by

integrating phylogenetics and ecological evidence. *Taxon* 68: 1189-1203.

Duan L, Han LN, Sirichamorn Y, Wen J, Compton JA, Deng SW, Arslan E, Ertuğrul K, Schrire B, Chen HF. 2021a. Proposal to recognise the tribes Adinobotryeae and Glycyrrhizeae (Leguminosae subfamily Papilionoideae) based on chloroplast phylogenomic evidence. *PhytoKeys* 181: 65-77.

Duan L, Li SJ, Su C, Sirichamorn Y, Han LN, Ye W, Lôc PK, Wen J, Compton JA, Schrire B, Nie ZL. 2021b. Phylogenomic framework of the IRLC legumes (Leguminosae subfamily Papilionoideae) and intercontinental biogeography of tribe Wisterieae. *Molecular Phylogenetics and Evolution* 163: 107235.

Duan L, Wen J, Yang X, Liu PL, Arslan E, Ertuğrul K, Chang ZY. 2015. Phylogeny of *Hedysarum* and tribe Hedysareae (Leguminosae: Papilionoideae) inferred from sequence data of ITS, *matK*, *trnL-F* and *psbA-trnH*. *Taxon* 64: 49-64.

Duan L, Yang X, Liu PL, Johnson G, Wen J, Chang ZY. 2016. A molecular phylogeny of Caraganeae (Leguminosae, Papilionoideae) reveals insights into new generic and infrageneric delimitations. *PhytoKeys* 70: 111-137.

Engler A, Prantl K. 1894. *Die natürlichen Pflanzenfamilien*. Leipzig: Wilhelm Engelmann. III: 307-308.

Erayman M, Ilhan E, GÜZEL Y, Eren AH. 2014. Transferability of SSR markers from distantly related legumes to *Glycyrrhiza* species. *Turkish Journal of Agriculture and Forestry* 38: 32-38.

Fischer FE, Meyer CA. 1843. *Index Seminum*. St. Petersburg: Petropolitanus. Ix:

95.

Gardner C. 1991. *Glycyrrhiza* L. In: Harden GJ ed. *Flora of New South Wales*. Kensington: New South Wales University Press. 2: 442.

Gómer-Sosa EV. 1999. *Glycyrrhiza* L. In: Zuloaga FO, Morrone OM eds. *Catálogo de las Plantas Vasculares de la República Argentina*. St. Louis: Missouri Botanical Garden Press. II: 681.

Graebin CS. 2018. The Pharmacological Activities of Glycyrrhizinic Acid ("Glycyrrhizin") and Glycyrrhetinic Acid. In: Mérillon J-M, Ramawat KG eds. *Sweeteners, Reference Series in Phytochemistry*. Cham: Springer. 245-261.

Grankina V. 2008. The system of the genus *Glycyrrhiza* L. (Fabaceae). *Novosti Sistematiki Nizshikh Rastenii* 40: 89-108.

Grigorev YS, Vasilchenko IT. 1948. *Glycyrrhiza*. In: Shishkin BK, Bobrov EG eds. *Flora of the U.S.S.R.* (English version published in 1972). Moscow and Leningrad: Izdatel'stvo Akademii Nauk. 13: 176-184.

Harris AJ, Wen J, Xiang QY. 2013. Inferring the biogeographic origins of inter-continental disjunct endemics using a Bayes-DIVA approach. *Journal of Systematics and Evolution* 51: 117-133.

Hayashi H, Hosono N, Kondo M, Hiraoka N, Ikeshiro Y. 1998. Phylogenetic relationship of *Glycyrrhiza* plants based on *rbcL* sequences. *Biological & Pharmaceutical Bulletin* 21: 782-783.

Hayashi H, Hosono N, Kondo M, Hiraoka N, Ikeshiro Y, Shibano M, Kusano G,

Yamamoto H, Tanaka T, Inoue K. 2000. Phylogenetic relationship of six *Glycyrrhiza* species based on *rbcL* sequences and chemical constituents. *Biological & Pharmaceutical Bulletin* 23: 602-606.

Hayashi H, Huang P, Inoue K. 2003. Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. *Plant and cell physiology* 44: 404-411.

Hayashi H, Miwa E, Inoue K. 2005. Phylogenetic relationship of *Glycyrrhiza lepidota*, American licorice, in genus *Glycyrrhiza* based on *rbcL* sequences and chemical constituents. *Biological & Pharmaceutical Bulletin* 28: 161-164.

Hayashi H, Sudo H. 2009. Economic importance of licorice. *Plant Biotechnology* 26: 101-104.

Hine R. 2019. A Dictionary of Biology (8 ed.). Oxford: Oxford University Press.

Hosseini MS, Samsampour D, Ebrahimi M, Abadía J, Khanahmadi M. 2018. Effect of drought stress on growth parameters, osmolyte contents, antioxidant enzymes and glycyrrhizin synthesis in licorice (*Glycyrrhiza glabra* L.) grown in the field. *Phytochemistry* 156: 124-134.

Hosseini MS, Samsampour D, Ebrahimi M, Abadía J, Sobhani Najafabadi A, Igartua E, Khanahmadi M. 2020. Evaluation of glycyrrhizin contents in licorice (*Glycyrrhiza glabra* L.) under drought and soil salinity conditions using nutrient concentrations and biochemical traits as biomarkers. *Acta Physiologiae Plantarum* 42: 1-7.

Hu JM, Chang SP. 2003. Two new members of the *Callerya* group (Fabaceae) based on phylogenetic analysis of *rbcL* sequences: *Endosamara racemosa* (Roxb.) Geesink and *Callerya vasta* (Kosterm.) Schot. *Taiwania* 28: 118-128.

Hu JM, Lavin M, Wojciechowski MF, Sanderson MJ. 2002. Phylogenetic analysis of nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe-Cyclolobium*, the core Millettieae, and the *Callerya* group. *Systematic Botany* 27: 722-733.

Huber-Morath A. 1965. Novitiae Florae Anatoliceae VII. *Bauhinia* 2: 295-306.

Hussein BR, Malik OA, Ong WH, Slik JWF. 2020. Automated Classification of Tropical Plant Species Data Based on Machine Learning Techniques and Leaf Trait Measurements. In: Alfred R, Lim Y, Haviluddin H, On CK eds. *Computational Science and Technology*. Singapore: Springer. 85-94.

Jeanes JA. 1996. *Glycyrrhiza* L. In: Walsh NG, Entwisle NG eds. *Flora of Victoria*. Melbourne: Inkata Press. 3: 698-699.

Ji Y, Yang L, Chase MW, Liu C, Yang Z, Yang J, Yang J-B, Yi T-S. 2019. Plastome phylogenomics, biogeography, and clade diversification of *Paris* (Melanthiaceae). *BMC Plant Biology* 19: 1-14.

Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv*: 256479.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJ, Wickett NJ. 2016. HybPiper: Extracting coding sequence and introns for

phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Kang SH, Lee JH, Lee HO, Ahn BO, Won SY., Sohn SH, Kim JS. 2018. Complete chloroplast genome and 45S nrDNA sequences of the medicinal plant species *Glycyrrhiza glabra* and *Glycyrrhiza uralensis*. *Genes & Genetic Systems* 93: 83-89.

Kao TC, Wu CH, Yen GC. 2014. Bioactivity and potential health benefits of licorice. *Journal of Agricultural and Food Chemistry* 62: 542-553.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment soft-ware version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Kho SJ, Manickam S, Malek S, Mosleh M, Dhillon SK. 2017. Automated plant identification using artificial neural network and support vector machine. *Frontiers in Life Science* 10: 98-107.

Kruganova EA. 1955. A review of species from the genera *Glycyrrhiza* L. and *Meristotropis* Fish. et C.A.Mey. *Acta Instituti Botanici, Academiae Scientiarum URSS*. 11: 161-197.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott, B. 2016. PartitionFinder

2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772-773.

Lavin M, Doyle JJ, Palmer D. 1990). Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* 44: 390-402.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* 54: 575-594.

Li J, Liu S, Wang J, Li J, Li J, Gao W. 2017. Gene expression of glycyrrhizin acid and accumulation of endogenous signaling molecule in *Glycyrrhiza uralensis* Fisch adventitious roots after *Saccharomyces cerevisiae* and *Meyerozyma guilliermondii* applications. *Biotechnology and applied biochemistry* 64: 700-711.

Li PC. 1963. Study of taxa of genus *Glycyrrhiza* Linn. (Leguminosae) in China. In: *Abstract of 30th Anniversary of Botanical Society of China*. Beijing: Institute of Scientific and Technical Intelligence of China. 90-91.

Li PC, Cui HB. 1998. *Glycyrrhiza*. In: Cui HB ed. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press. 42(2): 167-176.

Li XY. 1993. A study of the System and New taxa of genus *Glycyrrhiza* L. *Bulletin of Botanical Research* 13: 14-43.

Li XY, Lu JH. 2015. *Taxonomy and experimental biology of the genus* Glycyrrhiza

L. Shanghai: Fudan University Press.

Linnaeus C. 1753. *Species Plantarum*. Stockholm: Impensis GC Nauk. 741-742.

Liu BB, Campbell CS, Hong DY, Wen J. 2020. Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. *Molecular Phylogenetics and Evolution* 147: 106784.

Liu BB, Ma ZY, Ren C, Hodel RGJ, Sun M, Liu XQ, Guang-Ning Liu, Hong DY, Zimmer EA, Wen J. 2021. Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: A case study in Vitaceae. *Journal of Systematics and Evolution* 59: 1124-1138.

Lock JM. 1989. *Legumes of Africa: A Check-List*. Richmond: Royal Botanic Gardens, Kew. 261.

LPWG (The Legume Phylogeny Working Group). 2013. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon* 62: 217-248.

Mata-Montero E, Carranza-Rojas J. 2016. Automated plant species identification: Challenges and opportunities. In: Mata FJ, Pont A eds. *ICT for Promoting Human Development and Protecting the Environment-6th IFIP World Information Technology Forum*. Cham: Springer. 26-36.

Mau B, Newton MA, Larget B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55: 1-12.

McCauley DE, Sundby AK, Bailey MF, Welch ME. 2007. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae) evidence from experimental crosses and natural populations. *American Journal of Botany* 94: 1333-1337.

Meng L. 2005. *Systematics of* Glycyrrhiza *L. (Fabaceae) - with a special reference to its relationship to* Glycyrrhizopsis *Boiss.&Bal.* Ph.D. Dissertation. Beijing: Institute of Botany, the Chinese Academy of Sciences.

Meng L, Zhu XY. 2010. Palynological characters of *Glycyrrhiza*, *Glycyrrhizopsis*, and *Meristotropis* (Leguminosae), with special reference to their taxonomic significance. *Journal of Systematics and Evolution* 48: 455-463.

Mertín FG. 1999. *Glycyrrhiza* L. In: Castroviejo S ed. *Flora Iberica*. Madrid: Real Jardín Botánico. VII(I): 350-353.

Mochida K, Sakurai T, Seki H, Yoshida T, Takahagi K, Sawai S, Uchiyama H, Muranaka T, Saito K. 2017. Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *The Plant Journal* 89: 181-94.

Mort ME, Randle CP, Kimball RT, Tadesse M, Crawford DJ. 2008. Phylogeny of Coreopsideae (Asteraceae) inferred from nuclear and plastid DNA sequences. *Taxon* 57: 109-120.

Nasrollahi V, Mirzaie-asl A, Piri K, Nazeri S, Mehrabi R. 2014. The effect of drought stress on the expression of key genes involved in the biosynthesis of

triterpenoid saponins in liquorice (*Glycyrrhiza glabra*). *Phytochemistry* 103: 32-37.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolultion* 32: 268-274.

Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytologist* 222: 1638-1651.

Öztürk M, Altay V, Rehman Hakeem K, Akçiçek E. 2017. *Liquorice: From Botany to Phytochemistry*. Cham: Springer. 5-17.

Pandey S, Verma B, Arya, P. 2017. A review on constituents, pharmacological activities and medicinal uses of *Glycyrrhiza glabra*. *Pharmaceutical Research* 2: 26-31.

Paradis E, Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.

Peterson PM, Romaschenko K, Arrieta YH. 2015. A molecular phylogeny and classification of the Eleusininae with a new genus, *Micrachne* (Poaceae: Chloridoideae: Cynodonteae). *Taxon* 64: 445-467.

Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. *Annual Review of Plant Biology* 62: 549-566.

Podlech D. 1991. *Glycyrrhiza*. In: Lock JM, Simpson K eds. *Legumes of West Asia: A Check-List*. Richmond: Royal Botanic Gardens, Kew. 136-137.

Rannala B, Yang, Z. 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution* 43: 304-311.

Raveendar S, So YS, Lee KJ, Lee DJ, Sung J, Chung JW. 2017. The complete chloroplast genome sequence of *Glycyrrhiza lepidota* (Nutt.) Pursh - An American wild licorice. *Journal of Crop Science and Biotechnology* 20, 295-303.

Rechinger KH. 1984. *Glycyrrhiza* and *Meristotropis*. In: Rechinger KH ed. *Flora Iranica*. Graz: Akademische Druck- und Verlagsanstalt. 157: 164-171.

Reiche C. 1898. *Glycyrrhiza* L. In: *Flora de Chile*. Santiago: Imprenta Cervantes. 2: 113-114.

Richardson T. 2008. *Sweets: A History of Candy*. New York: Bloomsbury Publishing USA. 400.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46: 195-203.

Ronquist F. 2004. Bayesian inference of character evolution. *Trends in Ecology & Evolution* 19, 475-481.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.

Systematic Biology 61: 539-542.

Rydberg PA. 1917. Glycyrrhizeae. In: *Flora of the Rocky Mountains and adjacent plains: Colorado, Utah, Wyoming, Idaho, Montana, Saskatchewan, Alberta, and neighboring parts of Nebraska, South Dakota, North Dakota, and British Columbia*. New York: New York Botanical Garden: 454.

Sabir J, Schwarz E, Ellison N, Zhang J, Baeshen NA, Mutwakil M, Jansen R, Ruhlman T. 2014. Evolutionary and biotechnology implications of plastidgenome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnology Journal* 12: 743-754.

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T. 2008. Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences of the United States of America* 105: 14204-14209.

Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K. 2011. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23: 4112-4123.

Schuettpelz E, Frandsen PB, Dikow RB, Brown A, Orli S, Peters M, Metallo A, Funk VA, Dorr LJ. 2017. Applications of deep convolutional neural networks to digitized natural history collections. *Biodiversity Data Journal* 5: e21139.

Scoggan HJ. 1978. *Glycyrrhiza* L. In: *The Flora of Canada*. Ottawa: National

Museum of Canada. 3: 996.

Sharma V, Katiyar A, Agrawal RC. 2018. *Glycyrrhiza glabra*: chemistry and pharmacological activity. In: Mérillon JM, Ramawat KG eds. *Sweeteners: Pharmacology, biotechnology, and applications*. Cham: Springer International Publishing. 87-100.

Sober E. 1981. The principle of parsimony. *The British Journal for the Philosophy of Science* 32:145-156.

Spooner DM, Ruess H, Iorizzo M, Senalik D, Simon P. 2017. Entire plastid phylogeny of the carrot genus (*Daucus*, Apiaceae): Concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. *American Journal of Botany* 104: 296-312.

Stanley TD, Ross EM. 1983. *Glycyrrhiza* L. In: *Flora of South-Eastern Queensland*. Brisbane: Queensland Department of Primary Industries. I: 302.

Studer L, Alberti M, Pondenkandath V, Goktepe P, Kolonko T, Fischer A, Liwicki M, Ingold R. 2019. A comprehensive study of imagenet pre-training for historical document image analysis. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE. 720-725. arXiv: 1905.09113v1

SüMBüL H, Tufan Ö, DüSEN OD, Göktürk RS. 2003. A new taxon of *Glycyrrhiza* L. (Fabaceae) from southwest Anatolia. *Israel Journal of Plant Sciences* 51: 71-74.

Sun H, Larsen K. 2010. *Eremosparton* Fischer & C.A.Meyer. In: Wu ZY, Raven

PH, Hong DY eds. *Flora of China.* Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 10: 505-506.

Szegedy C, Ioffe S, Vanhoucke V, Alemi A. 2017. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 31(1). arXiv: 1602.07261v2

Tieleman T, Hinton G. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning* 4: 26-31.

Turner BL. 1959. *Glycyrrhiza* L. In: *The legumes of Texas*. Austin and London: University of Texas Press. 201-202.

Turrill WB. 1937. *Glycyrrhizopsis syriaca* Turrill. *Bulletin of Miscellaneous Information, Royal Gardens, Kew* 2: 79.

Unger J, Merhof D, Renner S. 2016. Computer vision applied to herbarium specimens of German trees: testing the future utility of the millions of herbarium specimen images for automated identification. *BMC Evolutionary Biology* 16: 248.

Vatanparast M, Powell A, Doyle JJ, Egan AN. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences* 6: e1036.

Wäldchen J, Mäder P. 2017. Plant species identifcation using computer vision techniques: a systematic literature review. *Archives of Computational Methods in Engineering* 25: 507-543.

Wäldchen J, Rzanny M, Seeland M, Mäder P. 2018. Automated plant species identification-Trends and future directions. *PLoS Computational Biology* 14: e1005993.

Wang YB, Liu BB, Nie ZL, Chen HF, Chen FJ, Figlar RB, Wen J. 2020. Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences via genome skimming. *Journal of Systematics and Evolution* 58: 673-695.

Wang Z, Li H, Zhu Y, Xu T. 2017. Review of plant identification based on image processing. *Archives of Computational Methods in Engineering* 24: 637-654.

Weber JZ. 1986. *Glycyrrhiza* L. In: Jessop JP, Teolken HR eds. *Flora of South Australia, 4th Edn*. Adelaide: South Australian Government Printing Division. II: 603-604.

Wei R, Yan YH, Harris AJ, Kang JS, Shen H, Xiang QP, Zhang XC. 2017. Plastid phylogenomics resolve deep relationships among eupolypod II ferns with rapid radiation and rate heterogeneity. *Genome Biology and Evolution* 9: 1646-1657.

Wen J, Harris AJ, Kalburgi Y, Zhang N, Xu Y, Zheng W, Ickert-Bond SM, Johnson G, Zimmer EA. 2018. Chloroplast phylogenomics of the New World grape species (*Vitis*, Vitaceae). *Journal of Systematics and Evolution* 56: 297-308.

Western Australian Herbarium. 1998. *Glycyrrhiza* L. In: *FloraBase - the Western Australian Flora*. Kensington: Department of Biodiversity, Conservation and Attractions.

White DM, Islam MB, Mason-Gamer RJ. 2019. Phylogenetic inference in section *Archerythroxylum* informs taxonomy, biogeography, and the domestication of coca (*Erythroxylum* species). *American Journal of Botany* 106: 154-165.

Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273-297.

Wojciechowski MF, Lavin M, Sanderson MJ. 2004. A phylogeny of legumes (Leguminosae) based on analyses of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846-1862.

Wojciechowski MF, Sanderson MJ, Steele KP, Liston A. 2000. Molecular phylogeny of the ‘‘temperate herbaceous tribes” of papilionoid legumes: a supertree approach. In: Herendeen PS, Bruneau A eds. *Advances in Legume Systematics, Part 9*. Richmond: Royal Botanic Garden, Kew. 277-298.

Wyman S K, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.

Xie L, Lu J, Li X, Zhang Y, Wei T, Li XY. 2014. The cross compatibility and hybrid seed vigor among three *Glycyrrhiza* species. *Plant Diversity and Resources* 36: 342-348.

Xie W, Hao Z, Zhou X, Jiang X, Xu L, Wu S, Zhao A, Zhang X, Chen B. 2018. Arbuscular mycorrhiza facilitates the accumulation of glycyrrhizin and liquiritin in *Glycyrrhiza uralensis* under drought stress. *Mycorrhiza* 28:

285-300.

Xu LR, Larsen K. 2010. *Alhagi* Gagnebin. In: Wu ZY, Raven PH, Hong DY eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 10: 526-527.

Yakovlev GP, Sytin AK, Roskov YR. 1996. *Legumes of Northern Eurasia: a checklist.* Richmond: Royal Botanic Gardens, Kew, 289-292.

Yang L, Yang Z, Liu C, He Z, Zhang Z, Yang J, Liu H, Yang J, Ji Y. 2019. Chloroplast phylogenomic analysis provides insights into the evolution of the largest eukaryotic genome holder, *Paris japonica* (Melanthiaceae). *BMC Plant Biology* 19: 293.

Yeo PF. 1968. *Glycyrrhiza* L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA eds. *Flora Europaea*. London: Cambridge University Press. 2: 127.

Younis S, Weiland C, Hoehndorf R, Dressler S, Hickler T, Seeger B, Schmidt M. 2018. Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Botany Letters* 165: 377-383.

Yu Y, Harris AJ, Blair C, He X. 2015. RASP (reconstruct ancestral state in phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution* 87: 46-49.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhang Y, Deng T, Sun L, Landis JB, Moore MJ, Wang H, Wang Y, Hao X, Chen J, Li S, Xu M. 2021. Phylogenetic patterns suggest frequent multiple origins of secondary metabolites across the seed-plant 'tree of life'. *National Science Review* 8: nwaa105.

Zhang XL, Li Xy, Wei LJ, Cui JP. 1998. The interspecific hybridization of *Glycyrrhiza* in Xinjiang. *Acta Botanica Boreali-occidentalia Sinica* 18: 132-136.

Zhou CM, Jin GQ. 2016. *Liquorice, 2nd version*. Beijing: China Agriculture Press.

Zimnitskaya SA. 2009. State of the reproductive system of populations of species of the genus *Glycyrrhiza* L.(Fabaceae). *Contemporary Problems of Ecology* 2: 392-395.

Table 1 Species delimitation in main revisions of *Glycyrrhiza*

Kruganova, 1955 (13 spp.)	Li, 1963 (16 spp.)	Meng, 2005 (13 spp.)	Grankina, 2008 (36 spp.)	Li & Lu, 2015 (21 spp.)
G. acanthocarpa	*G. acanthocarpa*	*G. acanthocarpa*	*G. acanthocarpa* *G. alalensis* *G. alaschanica*	*G. acanthocarpa*
G. aspera	*G. aspera*	*G. aspera*	*G. aspera*	*G. aspera*
G. astragalina	*G. astragalina*	*G. astragalina*	*G. astragalina* *G. brachycarpa*	*G. astragalina*
	G. bucharica	*G. bucharica*		*G. bucharica*
G. echinata	*G. echinata*	*G. echinata*	*G. echinata*	*G. echinata*
			G. eglandulosa	*G. eglandulosa* *G. erythrocarpa*
	G. eurycarpa		*G. eurycarpa*	*G. eurycarpa*
*	*G. flavescens*	*		*G. flavescens*
G. foetida	*G. foetida*	*G. foetida*	*G. foetida*	*G. foetida*

			G. foetidissima	
G. glabra		*G. glabra*	*G. glabra*	*G. glabra*
			G. glandulifera	
			G. gobica	
			G. grandiflora	
			G. hirsuta	
			G. hispida	
G. inflata	*G. inflata*	*G. inflata*	*G. inflata*	*G. inflata*
G. korshinskyi	*G. korshinskyi*		*G. korshinskyi*	
			G. laxiflora	
			G. laxissima	
G. lepidota	*G. lepidota*	*G. lepidota*	*G. lepidota*	*G. lepidota*
			G. macrophylla	
			G. michajloviana	
			G. nadezhinae	
G. macedonica	*G. macedonica*			*G. macedonica*
G. pallidiflora **	*G. pallidiflora*	*G. pallidiflora*	*G. pallidiflora*	*G. pallidiflora*
			G. paucifoliolata	*G. paucifoliolata*
				G. prostrata
			G. purpureiflora	
			G. sergievskiana	
			G. shiheziensis	
			G. soongorica	
G. squamulosa ***	*G. squamulosa*	*G. squamulosa*	*G. squamulosa*	*G. squamulosa*
	G. triphylla	*G. triphylla*		*G. triphylla*
G. uralensis		*G. uralensis*	*G. uralensis*	*G. uralensis*
			G. viscida	
	G. yunnanensis		*G. yunnanensis*	*G. yunnanensis*
			G. zaissanica	

Note: **G. flavescens* is supported but as a species in *Glycyrrhizopsis*; ***G. purpureiflora* is treated as *Meristotropis purpureiflora*; ****G. triphylla* is accepted as *Meristotropis triphylla*.

Table 2 Model test results on the *Glycyrrhiza echinata-G. macedonica* species pair using machine learning method

	Sample1		Sample2		Sample3	
Accuracy	91.53%		97.46%		90.68%	
	G. echinata	*G. macedonica*	*G. echinata*	*G. macedonica*	*G. echinata*	*G. macedonica*
Precision	88.37%	100%	98.67%	95.35%	100%	87.36%
Recall rate	100%	76.19%	97.37%	97.62%	73.81%	100%
F1-score	93.83%	86.49%	98.01%	96.47%	93.25%	93.25%

Fig. 1. Inflorescences and type specimens of several widely accepted species/dubious species of *Glycyrrhiza* and *Glycyrrhizopsis*. A: *Glycyrrhizopsis flavescens* (inflorescence); B: *Glycyrrhizopsis flavescens* (holotype, K); C: *Glycyrrhizopsis syriaca* (holotype, K); D: *Glycyrrhiza echinata* (inflorescence); E:

Glycyrrhiza echinata (lectotype, LINN); F: *Glycyrrhiza macedonica* (isotype, LE).

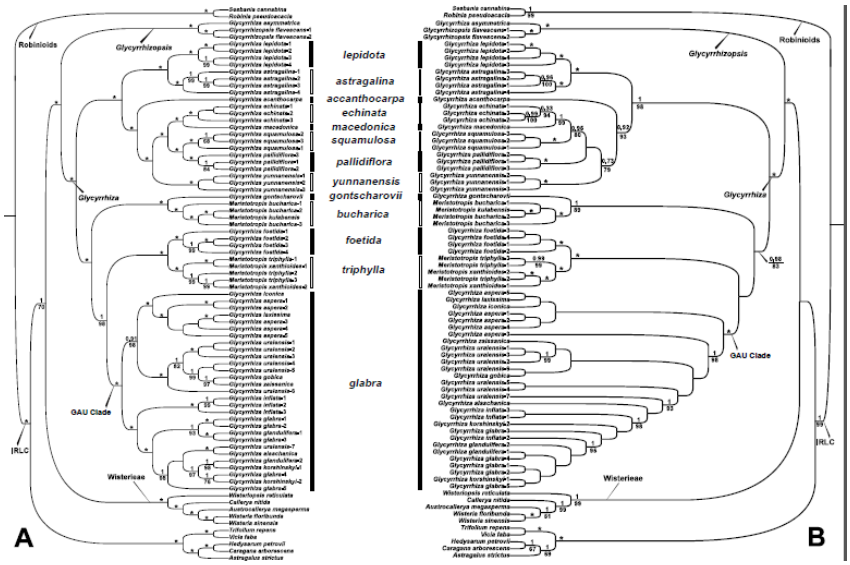


Fig. 2. Maximum clade credibility tree resulting from Bayesian Inference of Glycyrrhizeae based on whole chloroplast genome sequences (**A**) and a concatenation of nuclear ribosomal DNA and five low-copy nuclear loci (**B**). Bayesian posterior probabilities (PP ≥ 0.95) are given above branches, maximum likelihood bootstrap values (LBS ≥ 50%) are below branches. Asterisks indicate both of PP = 1 and LBS = 100%.

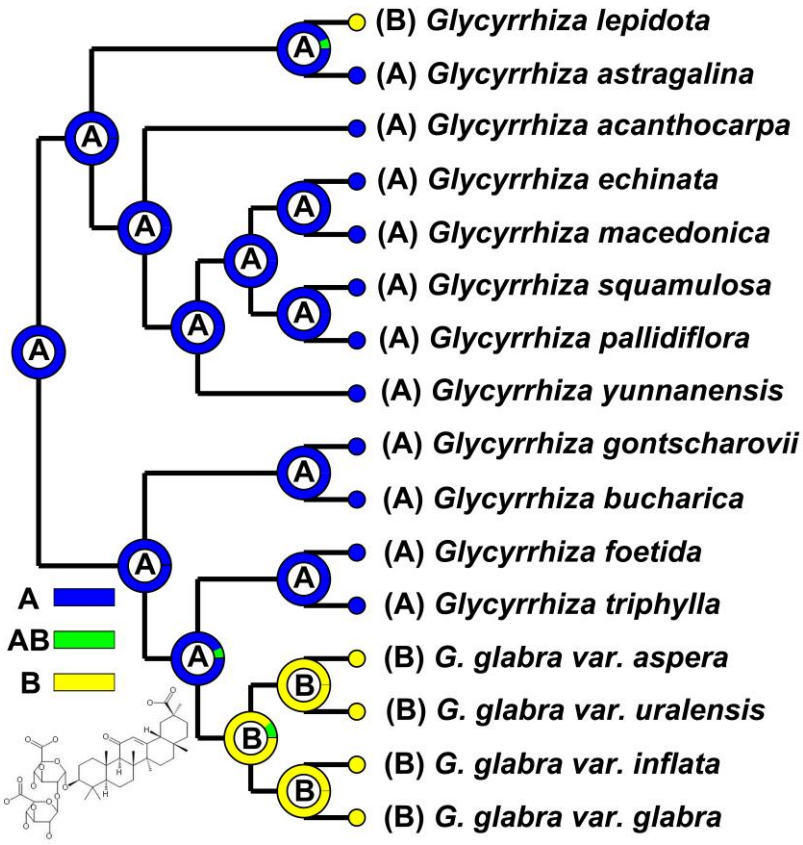


Fig. 3. Ancestral character estimation for the absence (A)/presence (B) of glycyrrhizin in *Glycyrrhiza* under the BBM model implemented in RASP v.4.0 based on the tree obtained from MrBayes. For each species with multiple accessions, the "# 1" accessions were selected for this analysis.

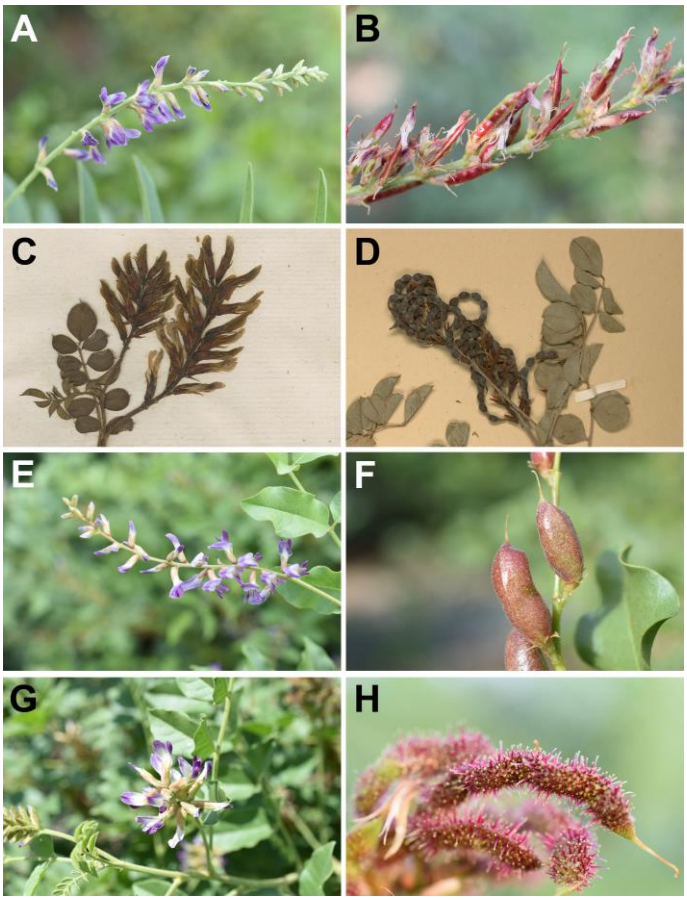


Fig. 4. Flower and Fruit morphologies of four varieties of *Glycyrrhiza glabra*. Flowers (**A**) and fruits (**B**) of *G. glabra* var. *glabra*; flowers [**C**; photographed on holotype specimen: C. Linnaeus 916.5 (LINN!)] and fruits [**D**; photographed on herbarium specimen: C. Kossinsky 119Y (LE!)] of *G. glabra* var. *aspera*; flowers (**E**) and fruits (**F**) of *G. glabra* var. *inflata*; flowers (**G**) and fruits (**H**) of *G. glabra* var. *uralensis*.