

RESEARCH ARTICLE

Hedysarum L. (Fabaceae: Hedysareae) Is Not Monophyletic – Evidence from Phylogenetic Analyses Based on Five Nuclear and Five Plastid Sequences

Pei-Liang Liu¹, Jun Wen^{2*}, Lei Duan³, Emine Arslan⁴, Kuddisi Ertuğrul⁴, Zhao-Yang Chang^{1*}

1 College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, China, **2** Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington D.C., United States of America, **3** Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, China, **4** Department of Biology, Faculty of Science, Selçuk University, Konya, Turkey

* czybbg@nwsuaf.edu.cn (ZYC); wenj@si.edu (JW)



OPEN ACCESS

Citation: Liu P-L, Wen J, Duan L, Arslan E, Ertuğrul K, Chang Z-Y (2017) *Hedysarum* L. (Fabaceae: Hedysareae) Is Not Monophyletic – Evidence from Phylogenetic Analyses Based on Five Nuclear and Five Plastid Sequences. PLoS ONE 12(1): e0170596. doi:10.1371/journal.pone.0170596

Editor: Shilin Chen, Chinese Academy of Medical Sciences and Peking Union Medical College, CHINA

Received: July 29, 2016

Accepted: January 6, 2017

Published: January 25, 2017

Copyright: © 2017 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All DNA sequences are available from the NCBI GenBank. Accession numbers are listed in Supporting Information S1 Appendix.

Funding: This work was supported by the National Natural Science Foundation of China (www.nsf.gov.cn, grant numbers 30270106, 30870155, to ZYC) and China Scholarship Council (www.csc.edu.cn, grant number 201406300108, to PLL). The funders had no role in study design, data collection

Abstract

The legume family (Fabaceae) exhibits a high level of species diversity and evolutionary success worldwide. Previous phylogenetic studies of the genus *Hedysarum* L. (Fabaceae: Hedysareae) showed that the nuclear and the plastid topologies might be incongruent, and the systematic position of the *Hedysarum* sect. *Stracheya* clade was uncertain. In this study, phylogenetic relationships of *Hedysarum* were investigated based on the nuclear ITS, ETS, *PGDH*, *SQD1*, *TRPT* and the plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG*, *petN-psbM* sequences. Both nuclear and plastid data support two major lineages in *Hedysarum*: the *Hedysarum s.s.* clade and the *Sartoria* clade. In the nuclear tree, *Hedysarum* is biphyletic with the *Hedysarum s.s.* clade sister to the *Corethroedendron + Eversmannia + Greuteria + Onobrychis* clade (the CEGO clade), whereas the *Sartoria* clade is sister to the genus *Taverniera* DC. In the plastid tree, *Hedysarum* is monophyletic and sister to *Taverniera*. The incongruent position of the *Hedysarum s.s.* clade between the nuclear and plastid trees may be best explained by a chloroplast capture hypothesis via introgression. The *Hedysarum* sect. *Stracheya* clade is resolved as sister to the *H. sect. Hedysarum* clade in both nuclear and plastid trees, and our analyses support merging *Stracheya* into *Hedysarum*. Based on our new evidence from multiple sequences, *Hedysarum* is not monophyletic, and its generic delimitation needs to be reconsidered.

Introduction

The legume family (Fabaceae) is the third largest flowering plant family with about 19,500 species in about 751 genera. The family exhibits a high level of species diversity and evolutionary success in various ecosystems worldwide [1]. The genus *Hedysarum* L. (Fabaceae: Hedysareae)

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

consists of about 160 species of perennial herbs to rarely shrublets. It mainly distributes in temperate Eurasia, with a few species in North Africa and North America. Species of *Hedysarum* adapt to diverse habitats in temperate forests, steppes, alpine regions and the Tibetan plateau. This genus is generally characterized by brown membranous connate stipules, imparipinnate leaves, non-persistent corollas at fruit maturity, backward turning standard petal, right- to obtuse-angle-shaped keels and most prominently, lomented legumes with 2-several seeds [2, 3]. Some species of *Hedysarum* are good fodder plants, such as *H. dahuricum* Turcz. ex B. Fedtsch. and *H. petrovii* Yakovlev in arid regions, and *H. neglectum* Ledeb. in alpine regions [4, 5].

The generic delimitation of *Hedysarum* has been highly problematic. Established by Linnaeus [6], *Hedysarum* consisted of 54 species at his time [7]. However, *Hedysarum* sensu Linnaeus was unnatural with 51 species subsequently transferred to 16 other genera such as *Alhagi* Gagnebin, *Desmodium* Desv., *Onobrychis* Mill. and *Sulla* Medik. [2, 7]. With more species being discovered, Fedtschenko [8] divided *Hedysarum* into seven sections mainly based on their habits and morphology of stems and loment. Choi & Ohashi [2] revised Fedtschenko's classification of *Hedysarum* based on their comprehensive morphological studies. They segregated the genera *Corethrodedron* Fischer & Basiner and *Sulla* from *Hedysarum*, and transferred the monotypic genus *Stracheya* Benth. [9] into *Hedysarum* as *H.* sect. *Stracheya* (Benth.) B.H.Choi & H.Ohashi. They also merged *H.* sect. *Multicaulia* (Boiss.) B.Fedtsch., *H.* sect. *Crinifera* (Boiss.) B.Fedtsch. and *H.* sect. *Subacaulia* (Boiss.) B.Fedtsch. into a broadly defined *H.* sect. *Multicaulia* s.l. [2].

Recent phylogenetic studies of Hedysareae [10, 11] did not support the monophyly of *Hedysarum* as circumscribed by Fedtschenko [8] nor by Choi & Ohashi [2]. Separations of the genera *Corethrodedron* and *Sulla* from *Hedysarum* were supported by molecular data [10, 11]. A new genus, *Greuteria* Amirahmadi & Kaz. Osaloo., was split from *Hedysarum* based on phylogenetic reconstruction [10, 11]. The monotypic *Sartoria* Boiss. & Heldr. was nested within *Hedysarum* sect. *Multicaulia* s.l. Therefore *Sartoria hedysaroides* Boiss. & Heldr. was transferred to *Hedysarum* as *H. anatolicum* Amirahmadi & Kaz. Osaloo [10]. After these treatments, *Hedysarum* was suggested to be monophyletic in the plastid trees [10, 11]. However, *Hedysarum* was not monophyletic in the nuclear ITS trees, and its relationship with other genera in Hedysareae was uncertain due to low support values of the ITS trees [10, 11].

With an extensive taxon sampling scheme and more molecular data, Duan et al. [11] recognized three main clades of *Hedysarum*: the *H.* sect. *Hedysarum* clade; the re-defined *H.* sect. *Multicaulia* s.l. clade excluding *H. kumaonense* Benth. ex Baker and *H. lehmannianum* Bunge, but including *Sartoria hedysaroides*; and the re-delimited *H.* sect. *Stracheya* clade consisting of *H. tibeticum* (Benth.) B.H.Choi & H.Ohashi, *H. kumaonense* and *H. lehmannianum*. The *H.* sect. *Stracheya* clade was sister to the *H.* sect. *Multicaulia* s.l. clade with very low support [11]. However, this relationship was not supported by Amirahmadi et al. [10]. The systematic position of the *H.* sect. *Stracheya* clade within *Hedysarum* was thus uncertain.

Previous phylogenetic hypotheses on *Hedysarum* and Hedysareae largely relied on plastid data since the nuclear ITS sequence showed limited resolution concerning the deep relationships [10, 11]. Sequencing more plastid genes may increase phylogenetic precision, however, accurate inference of the phylogenetic history of this group requires nuclear data [12, 13]. Nuclear genes are considered as an important complement or alternative to plastid ones due to their biparental inheritance, and are less vulnerable to hybridizations and introgressions than organelle genes [1, 14, 15].

To our knowledge, only the ribosomal internal transcribed spacer (ITS) sequence from the nuclear genome has been used to infer the phylogeny of *Hedysarum*. The ribosomal external transcribed spacer (ETS) has also been widely used in phylogenetic studies [13], and has an

evolutionary rate at least as fast as the ITS sequence [16]. Furthermore, several studies have developed other nuclear sequences for phylogenetic reconstruction in various taxonomic groups of Fabaceae [12, 17, 18]. The *SQD1* (UDP sulfoquinovose synthase gene) belongs to the low copy conserved ortholog set (COS) genes [19]. It has been used in a phylogenetic study of the caesalpinoid legumes [18]. The nuclear coding *PGDH* (putative phosphogluconate dehydrogenase gene) and *TRPT* (putative triosephosphate translocator gene) sequences are exon-derived, putative orthologs, single copy genes [12], and were used in a phylogenetic study of the Hologalegina legumes [17].

In the present study, we employed the nuclear ITS, ETS, *PGDH*, *SQD1* and *TRPT* sequences and the plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG*, and *petN-psbM* sequences to test (1) the monophyly of *Hedysarum*, and (2) the systematic position of the *Hedysarum* sect. *Stracheya* clade.

Materials and Methods

Ethics statement

The plant materials used in this study did not involve protected or endangered species. No specific permits were required for the collection of samples. Voucher information of all the samples was given in [S1 Appendix](#).

Taxon sampling

A total of 58 accessions were included in this study ([S1 Appendix](#)), representing all genera in the tribe Hedysareae and all sections and major infra-sectional clades in *Hedysarum* as recognized by Duan et al. [11]. *Hedysarum boveanum* Bunge ex Basiner, *H. denticulatum* Regel, *H. minjanense* Rech.f., *H. poncinsii* Franch., and *H. syriacum* Boiss. were placed in sect. *Multicaulia* s.l. [2] based on morphological characters [20–22], and they were sampled for the first time to test their phylogenetic positions. *Alhagi sparsifolia* Shap. ex Keller & Shap. was selected as the outgroup based on previous results [10,11]. Some of the ITS, *psbA-trnH*, and *trnL-trnF* sequences used in this study were published by Duan et al. [11] and Amirahmadi et al. [10], while all other sequences were generated by the present study ([S1 Appendix](#)). Voucher information of DNA sequences was listed in [S1 Appendix](#).

DNA extraction, PCR and sequencing

Total genomic DNAs were extracted from silica-gel dried leaf material or herbarium specimen using either the Plant DNA Extraction Kit AGP965/960 (AutoGen, Holliston, Massachusetts, USA) or the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA).

Primers used for amplification and sequencing were “ITS5a” and “ITS4” for ITS [23], “ETS-Hedy” (CCYTGWGCYRTTGTGCCTTGG, designed in this study) and “18S-IGS” [16] for ETS, forward and reverse primers for *SQD1* [19], forward and reverse primers for *PGDH* and *TRPT* [12], “psbAF” and “trnHR” for *psbA-trnH* intergenic region [24], “trnC” and “petN 1R” for *trnC-petN* intergenic region [25], “trnS” [26] and “5’tRN_{G2S}” [27] for *trnS-trnG* intergenic region, “ycf6F” and “psbMR” [27] for the *petN-psbM* intergenic region (the *ycf6* gene has been renamed as *petN*, see [28, 29]), “c” and “f” for *trnL* intron plus *trnL-trnF* intergenic region [30].

Polymerase chain reaction (PCR) was performed in a 25 μ l volume with the following components: 10 \times reaction buffer, 200 μ mol \cdot L⁻¹ of each dNTP, 10 μ g BSA, 0.4 μ mol \cdot L⁻¹ of each primer, 2.5mmol \cdot L⁻¹ MgCl₂, 1U of BIOLASE DNA Polymerase (Bioline USA Inc., Taunton, Massachusetts), 1–7.5 μ l template DNA. The amplification conditions were 3 min at 95°C,

followed by 36–40 cycles of 1 min at 94°C, 1 min at 50–60°C, and 1–1.5 min at 72°C, then a final extension at 72°C for 7–10 min. The PCR products were purified using ExoSAP-IT (USB Corporation, Cleveland, Ohio, USA) or the polyethylene glycol (PEG) precipitation procedure [31]. Cycle sequencing reactions were conducted in both directions using the amplification primers and the BigDye 3.1 reagents. After being cleaned up by the Sephadex columns, the sequencing products were run on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, California, USA). Sequences were assembled using the program Geneious v.8.1.2 [32] (<http://www.geneious.com/>). All sequences have been deposited in GenBank and the accession numbers were listed in [S1 Appendix](#).

Phylogenetic analysis

Multiple sequence alignments were conducted in Geneious v.8.1.2 using MUSCLE [33] with default settings, followed by manual adjustments. For the non-coding sequences, ambiguously aligned regions were removed from the matrix prior to phylogenetic analysis. Insertions and deletions (indels) in the data matrices were coded as binary characters using the program SeqState [34] according to the “simple coding” method [35]. The binary characters were combined with the DNA data as the additional partition of the matrix. SequenceMatrix [36] was employed to assemble combined datasets.

Single locus analysis

Phylogenetic reconstructions were first conducted by using single locus datasets. The best-fit nucleotide substitution model for each of the 10 individual sequences was determined using the Bayesian Information Criterion (BIC) in jModelTest v.2.1.7 [37]. For the ITS dataset, boundaries of the 5.8S, ITS1 and ITS2 regions were determined by comparing with the published 5.8S sequence of *Vicia faba* L. [38], and models for the ITS1, 5.8S and ITS2 regions were determined separately (see [Table 1](#)).

Bayesian inferences (BI) were conducted in MrBayes v.3.2.5 [39, 40]. When models (such as TIM3ef, TrNef, TPM1, TVM) determined by BIC were not directly available in MrBayes, we transformed a GTR model in MrBayes by fixing the six nucleotide substitution rates and four nucleotide state frequencies to the values as calculated by jModelTest for each dataset. The model applied to all the coded binary partitions was a default Standard Discrete Model in MrBayes [41]. For the ITS dataset, the partitions were done for ITS1, 5.8S and ITS2 separately.

Table 1. Characteristics of individual datasets: alignment length, the number and percentage of constant, variable and potentially parsimony-informative (Pi) sites, the number of coded indel(s), and the best-fit nucleotide substitution model determined by BIC.

Dataset	Length	Constant (%)	Variable (%)	Pi (%)	Indel	Model
ETS	346	131 (37.9%)	215 (62.1%)	149 (43.1%)	45	HKY+G
ITS1	285	126 (44.2%)	159 (55.8%)	112 (39.3%)	55	TIM3ef+G
5.8S	165	152 (92.1%)	13 (7.9%)	6 (3.6%)	2	TPM3+I
ITS2	229	108 (47.2%)	121 (52.8%)	86 (37.6%)	22	TrNef+G
PGDH	405	304 (75.1%)	101 (24.9%)	67 (16.5%)	0	K80+I
SQD1	273	212 (77.7%)	61 (22.3%)	41 (15.0%)	0	TPM1+G
TRPT	330	215 (65.2%)	115 (34.8%)	65 (19.7%)	7	HKY+G
psbA-trnH	403	300 (74.4%)	103 (25.6%)	58 (14.4%)	31	TPM1uf+I+G
trnC-petN	1221	882 (72.2%)	339 (27.8%)	213 (17.4%)	110	TVM+G
trnL-trnF	1068	881 (82.5%)	187 (17.5%)	110 (10.3%)	60	TVM+G
trnS-trnG	692	485 (70.1%)	207 (29.9%)	142 (20.5%)	69	TPM1uf+G
petN-psbM	1316	999 (75.9%)	317 (24.1%)	198 (15.0%)	104	TVM+G

doi:10.1371/journal.pone.0170596.t001

In the Bayesian inference, two independent analyses starting from different random trees with three heated and one cold chain were run for 10,000,000 generations, and trees were sampled every 1,000 generations (10,000 trees sampled in total). The first 2,500 trees (25%) were discarded as burn-in, and the remaining trees were used to construct a 50% majority-rule consensus tree and posterior probabilities (PP). Tree visualization was achieved in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Phylogenetic reconstructions were also carried out with maximum parsimony (MP) and maximum likelihood (ML) analyses using PAUP* 4.0b10 [42] and RAxML v.8.2 [43], respectively. The MP bootstrap analysis was performed with the following settings: heuristic search, TBR branch-swapping, 1,000 bootstrap replicates, random addition sequence with 10 replicates, a maximum of 1,000 trees saved per round. The ML rapid bootstrap analysis was conducted with a random seed, 1,000 alternative runs, and the same partition scheme as in the Bayesian analysis. Model parameters were estimated and optimized for each partition of the dataset by RAxML with the GTRCAT commands. Bootstrap support values from the MP (PBS) and ML (LBS) analyses were labeled on the corresponding branches of the BI trees.

Concatenated data analysis

We conducted incongruence length difference (ILD) test [44] to assess whether different datasets could be concatenated for phylogenetic reconstruction. The ILD tests were performed in PAUP* 4.0b10 [42] by using only the informative characters [45] with heuristic search, 1,000 replicates, simple addition sequence (with a maximum of 1,000 trees saved per replicate), and tree-bisection-reconnection (TBR) branch-swapping algorithm. When the ILD test found a *p* value greater than 0.01, datasets were concatenated for phylogenetic analysis [46].

BI, MP and ML analyses were conducted for the concatenated datasets by using the same methods as in the single locus analyses. For BI and ML, the best partitioning scheme and substitution model were determined by using PartitionFinder v.1.1.1 [47] with the following settings: linked branch length, BIC metric for model and partitioning selection, and the greedy algorithm. Data blocks in PartitionFinder were set as the followings: the ETS, ITS1, 5.8S and ITS2 datasets, the first, second and third codon position of each of the *PGDH*, *SQD1* and *TRPT* datasets, and the *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG* and *petN-psbM* datasets.

We performed the approximately unbiased (AU) tests [48] using CONSEL v.0.2 [49] to test statistical support of incongruence between nuclear and plastid topologies. Topological constraints were generated from the ML trees using MEGA v.6.0 [50]. Per site log likelihoods for each topology were estimated in RAxML v.8.2 [43] under the GTR + G model. The dataset partition scheme was the same as in the ML analysis.

Coalescent analysis

The Bayesian coalescent-based analyses for multi-locus data were performed using *BEAST [51] as implemented in the BEAST v.2.4.3 package [52]. Two sets of analyses were conducted: one with all five nuclear sequences, another with all five plastid sequences. The substitution model for each sequence was that used in the single locus analysis, and model parameters were unlinked among partitions. The lognormal relaxed clock model, the linear with constant root population function and the Yule speciation model were used. Four independent runs of 100,000,000 and 500,000,000 generations were performed for the nuclear data and the plastid data, respectively. Samples were stored every 5,000 generations. Tracer v.1.6 (<http://beast.bio.ed.ac.uk/Tracer>) was used to check the effective sample sizes (ESSs) of the parameters and the convergence of different runs. After removing the 10% burn-in of each run, the log and tree files from different runs were combined by using LogCombiner [52]. In the combined results,

the ESSs of the sampled parameters all exceeded 200. The maximum clade credibility tree was annotated using TreeAnnotator [52].

Bayesian concordance analysis

We carried out the Bayesian Concordance Analysis (BCA) [53] for the five nuclear loci as well as the five plastid loci using the BUCKy v.1.1.4 package [54]. Gene trees of each locus were obtained from the single locus BI (.t files generated by MrBayes), and were summarized by using the program mbsum [54] with a 25% burn-in. The output files of mbsum were then used to construct the primary concordance tree and concordance factor (CF) in the program bucky [54]. The *a priori* level of discordance among loci (α) was set to 0.01, 1 and 100 in alternative runs to assess the effect of α on the BCA results. Each run was conducted with 10,000,000 updates, four replicates, one cold and three hot chains of the Metropolis coupled Markov chain Monte Carlo (MCMCMC). In MCMCMC, a swap was proposed once every 100 updates, and the α multiplier was set to 10. Different runs with α set to 0.01, 1 and 100 all recovered the same primary concordance tree topology and CFs, hence the BCA results obtained with the default $\alpha = 1$ were reported. The primary concordance tree was visualized in FigTree, and branch labels were annotated as the sample-wide posterior mean CFs and their 95% credibility intervals.

Results

Nuclear data

Characteristics of the individual nuclear datasets were summarized in Table 1. Treating indels as missing data or binary characters did not change the topologies, but the analyses that included indels as characters increased support values in general. The results presented here were based on the datasets with indels treated as binary characters. Trees generated from each individual datasets supported the monophyly of each of the genera *Ebenus* L., *Greuteria*, *Sulla*, and *Taverniera* DC., and several small lineages within *Hedysarum* and *Onobrychis*. *Hedysarum denticulatum* and *H. minjanense* were nested with *H. sect. Stracheya* as redefined by Duan et al. [11] in all five nuclear trees. Deep relationships within Hedysareae were largely unsolved by single sequences. However, *Hedysarum* was not supported to be monophyletic in any of the five single-sequence trees.

In order to search trees with better resolution, we concatenated individual datasets in our analyses. To explore whether two or more nuclear datasets could be concatenated, we conducted 10 pairwise ILD tests for the five nuclear datasets. The *p* values were shown in Table 2. Congruence was detected between any pair in the ETS, ITS, *PGDH* and *TRPT* datasets (all with $p > 0.01$). We first concatenated the non-coding ribosomal ETS and ITS datasets (pairwise ILD: $p = 0.13$), as well as the coding *PGDH* and *TRPT* datasets (pairwise ILD: $p = 0.061$) for two separate analyses.

In the concatenated ETS + ITS tree (S1 Fig), *Corethrodedron*, *Ebenus*, *Greuteria*, *Onobrychis*, *Sulla* and *Taverniera* were each supported to be monophyletic. *Corethrodedron*,

Table 2. P values of pairwise ILD tests for the nuclear ETS, ITS, PGDH, TRPT and SQD1 datasets.

	ITS	PGDH	TRPT	SQD1
ETS	0.130	0.036	0.025	0.001
ITS		0.331	0.170	0.002
PGDH			0.061	0.001
TRPT				0.003

doi:10.1371/journal.pone.0170596.t002

Eversmannia Bunge, *Greuteria* and *Onobrychis* together formed a well supported clade (the CEGO clade), with Bayesian posterior probability (PP) = 1, maximum parsimony bootstrap percent (PBS) = 100%, and maximum likelihood bootstrap percent (LBS) = 100%, but relationships among these four genera were poorly resolved. The *Hedysarum* sect. *Hedysarum* clade and the *H.* sect. *Stracheya* clade were sisters (the *Hedysarum* s.s. clade, PP = 1, PBS = 70%, LBS = 94%). *Sartoria hedysaroides* plus some members of the *Hedysarum* sect. *Multicaulia* s.l. formed a well supported clade (the *Sartoria* clade, PP = 1, PBS = 100%, LBS = 100%). The genus *Hedysarum* was not monophyletic, because the *Hedysarum* s.s. clade grouped with the CEGO clade, while the *Sartoria* clade formed a clade with *Taverniera*. However, these relationships had low support values in the ITS + ETS tree (not supported by the BI tree, poorly supported by the MP and ML trees).

In the concatenated *PGDH* + *TRPT* tree (S2 Fig), the *Hedysarum* sect. *Hedysarum* clade and the *H.* sect. *Stracheya* clade were supported to be sister to each other and constituted the *Hedysarum* s.s. clade (PP = 0.98, PBS = 54%, LBS = 80%). Additionally, the *Hedysarum* s.s. clade grouped with *Corethrodedron*, *Eversmannia*, *Greuteria*, and *Onobrychis* (PP = 0.96, PBS = 34%, LBS = 35%), but relationships within this clade were unclear. The *Sartoria* clade (PP = 1, PBS = 71%, LBS = 76%) and *Taverniera* were supported to be sisters with low support (PP = 0.8, PBS = 52%, LBS = 64%).

In both the ETS + ITS tree (S1 Fig) and the *PGDH* + *TRPT* tree (S2 Fig), *Hedysarum* was not monophyletic because the *Hedysarum* s.s. clade grouped with *Corethrodedron*, *Eversmannia*, *Greuteria* and *Onobrychis*, while the *Sartoria* clade was sister to the genus *Taverniera*. Although support values varied in these two trees, *Hedysarum* was shown to be biphyetic. Pairwise ILD test between the ETS + ITS dataset and the *PGDH* + *TRPT* dataset found $p = 0.703$, thus we concatenated these four sequences in our further analyses.

In the concatenated ETS + ITS + *PGDH* + *TRPT* tree (S3 Fig), *Ebenus* diverged first, followed by *Sulla* (PP = 0.71, PBS = 84%, LBS = 40%). *Corethrodedron*, *Ebenus*, *Greuteria*, *Sulla*, and *Taverniera* were each highly supported to be monophyletic (all with PP = 1, PBS = 100%, LBS = 100%). *Onobrychis* was weakly supported to be monophyletic (PP = 0.68, PBS = 60%, not supported by the ML tree) with two strongly supported subclades (both with PP = 1, PBS = 100%, LBS = 100%). The CEGO clade was strongly supported (PP = 1, PBS = 100%, LBS = 100%). *Hedysarum* was biphyetic with the *Hedysarum* s.s. clade sister to the CEGO clade (PP = 1, PBS = 86%, LBS = 94%), and the *Sartoria* clade forming a clade with *Taverniera* (PP = 0.97, PBS = 69%, LBS = 92%). The CEGO + *Hedysarum* s.s. clade was then sister to the *Sartoria* + *Taverniera* clade (PP = 1, PBS = 82%, LBS = 90%).

The *SQD1* tree (S4 Fig) showed limited resolution as well as low support values concerning the deep relationship in Hedysareae. The *SQD1* dataset has the smallest number (41) of parsimony-informative sites compared with the other four nuclear sequences (see Table 1). *Hedysarum* was, nevertheless, not monophyletic in the *SQD1* tree. Members of the *Sartoria* clade grouped with *Taverniera*, *Hedysarum longigynophorum* C.C.Ni, *H. dentatoalarum* K.T.Fu and *H. chinense* (B.Fedtsch.) Hand.-Mazz., but the support values were low (PP = 0.85, PBS = 52%, LBS = 53%). *Sulla*, *Ebenus*, *Eversmannia*, *Corethrodedron*, *Greuteria*, *Onobrychis* and the rest of *Hedysarum* formed another clade (PP = 1, PBS = 62%, LBS = 68%). We could see that the *SQD1* tree (S4 Fig) was not contrary to the ETS + ITS + *PGDH* + *TRPT* tree (S3 Fig) with regard to the non-monophyly of *Hedysarum*. Hence we tentatively concatenated the *SQD1* dataset with the other four nuclear datasets.

In the concatenated ETS + ITS + *PGDH* + *TRPT* + *SQD1* tree (Fig 1), relationships of the CEGO, *Hedysarum* s.s., *Sartoria* clades and *Taverniera* remained the same as in the concatenated ETS + ITS + *PGDH* + *TRPT* tree (S3 Fig). With the *SQD1* dataset added, the support values were even higher (Fig 1): the *Hedysarum* s.s. clade was sister to the CEGO clade

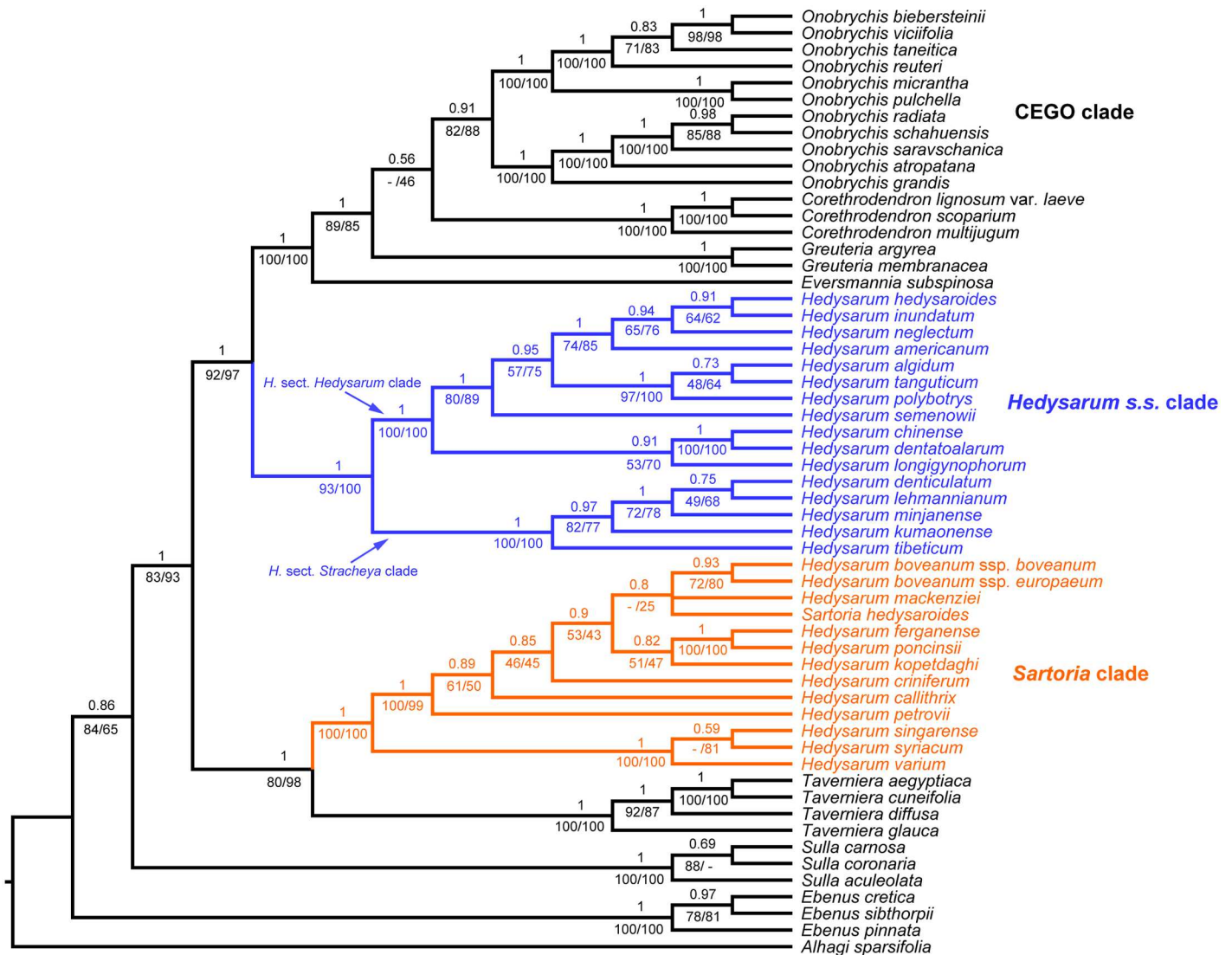


Fig 1. Bayesian tree based on the concatenated nuclear ITS, ETS, PGDH, TRPT and SQD1 sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. Dashes indicate branches that are not found in the maximum parsimony tree or the maximum likelihood tree.

doi:10.1371/journal.pone.0170596.g001

with PP = 1, PBS = 92%, LBS = 97% (compared with PP = 1, PBS = 86%, LBS = 94% in S3 Fig), and the *Sartoria* clade grouped with *Taverniera* with PP = 1, PBS = 80%, LBS = 98% (compared with PP = 0.97, PBS = 69%, LBS = 92% in S3 Fig).

The coalescent tree based on all five nuclear sequences (S5 Fig) revealed that the *Hedysarum* s.s clade was sister to the CEGO clade (PP = 0.95), and the *Sartoria* clade grouped with *Taverniera* (PP = 0.96). The CEGO, *Hedysarum* s.s., the *Sartoria* clade and *Taverniera* were each supported to be monophyletic (all with PP = 1).

The primary concordance tree based on all five nuclear sequences (S6 Fig) had similar topology compared with the concatenated data tree (Fig 1) and the coalescent tree (S5 Fig). It supported the sister relationship between the *Hedysarum* s.s and the CEGO clades (mean CF = 0.26), and that between the *Sartoria* clade and *Taverniera* (mean CF = 0.3).

Table 3. P values of pairwise ILD tests for the plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG* and *petN-psbM* datasets.

	<i>trnC-petN</i>	<i>trnL-trnF</i>	<i>trnS-trnG</i>	<i>petN-psbM</i>
<i>psbA-trnH</i>	0.751	0.266	0.423	0.315
<i>trnC-petN</i>		0.143	0.192	0.894
<i>trnL-trnF</i>			0.260	0.657
<i>trnS-trnG</i>				0.250

doi:10.1371/journal.pone.0170596.t003

Plastid data

Characteristics of plastid datasets were presented in Table 1. Trees based on individual plastid datasets had limited resolution. The 10 pairwise ILD tests for the five plastid datasets all found $p > 0.01$ (Table 3), therefore we concatenated these five plastid datasets in our analyses. Because there was incongruence between the plastid and nuclear trees, the plastid and nuclear datasets were not concatenated in our analysis.

In the concatenated plastid tree based on *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG* and *petN-psbM* sequences (Fig 2), *Sulla* diverged first, followed by *Ebenus* with high support (PP = 1, PBS = 100%, LBS = 100%). *Corethrodedron*, *Ebenus*, *Greuteria*, *Hedysarum*, *Onobrychis*, *Sulla*, and *Taverniera* were each supported to be monophyletic. The CEGO clade was also highly supported (PP = 1, PBS = 100%, LBS = 100%), which was sister to the *Hedysarum* + *Taverniera* clade (PP = 1, PBS = 98%, LBS = 100%). *Hedysarum* was monophyletic (PP = 1, PBS = 92%, LBS = 100%), and was sister to *Taverniera* (PP = 0.97, PBS = 88%, LBS = 97%). Within *Hedysarum*, the *Sartoria* clade (PP = 1, PBS = 100%, LBS = 100%) and the *Hedysarum* s.s. clade (PP = 1, PBS = 88%, LBS = 100%) were each well supported. *H. denticulatum* and *H. minjanense* were nested within *H. sect. Stracheya* as redefined by Duan et al. [11]. The *Hedysarum* sect. *Hedysarum* clade was sister to the *H. sect. Stracheya* clade (PP = 1, PBS = 88%, LBS = 100%).

The coalescent tree based on all five plastid sequences (S7 Fig) showed that *Hedysarum* was monophyletic (PP = 1), and was sister to *Taverniera* (PP = 0.83). The *Hedysarum* + *Taverniera* clade was then sister to the CEGO clade (PP = 1).

The primary concordance tree based on all five plastid sequences (S8 Fig) also had similar topology compared with the concatenated plastid data tree (Fig 2) and the coalescent tree (S7 Fig). *Hedysarum* was supported to be monophyletic (mean CF = 0.67) and was sister to *Taverniera* (mean CF = 0.35). The *Hedysarum* + *Taverniera* clade was then sister to the CEGO clade (mean CF = 0.56).

The AU tests of the concatenated nuclear ETS + ITS + *PGDH* + *TRPT* dataset against the concatenated five plastid sequence dataset were performed. The nuclear dataset rejected the plastid topology ($p < 0.01$), and the plastid dataset rejected the nuclear topology ($p < 0.01$) as well, suggesting incongruence between the nuclear and plastid datasets. The concatenated five nuclear sequence dataset (ETS + ITS + *PGDH* + *TRPT* + *SQD1*) also rejected the plastid topology ($p < 0.01$), and vice versa.

Discussion

Non-monophyly of *Hedysarum*

Our results revealed incongruence between the nuclear and the plastid trees in *Hedysarum* (Fig 3). In the nuclear trees (Figs 1 and 3), *Hedysarum* is biphyetic; the *Sartoria* clade is sister to *Taverniera*; and the *Hedysarum* s.s. clade is sister to the CEGO clade. On the other hand, *Hedysarum* is monophyletic and sister to *Taverniera* in the plastid tree (Figs 2 and 3). The

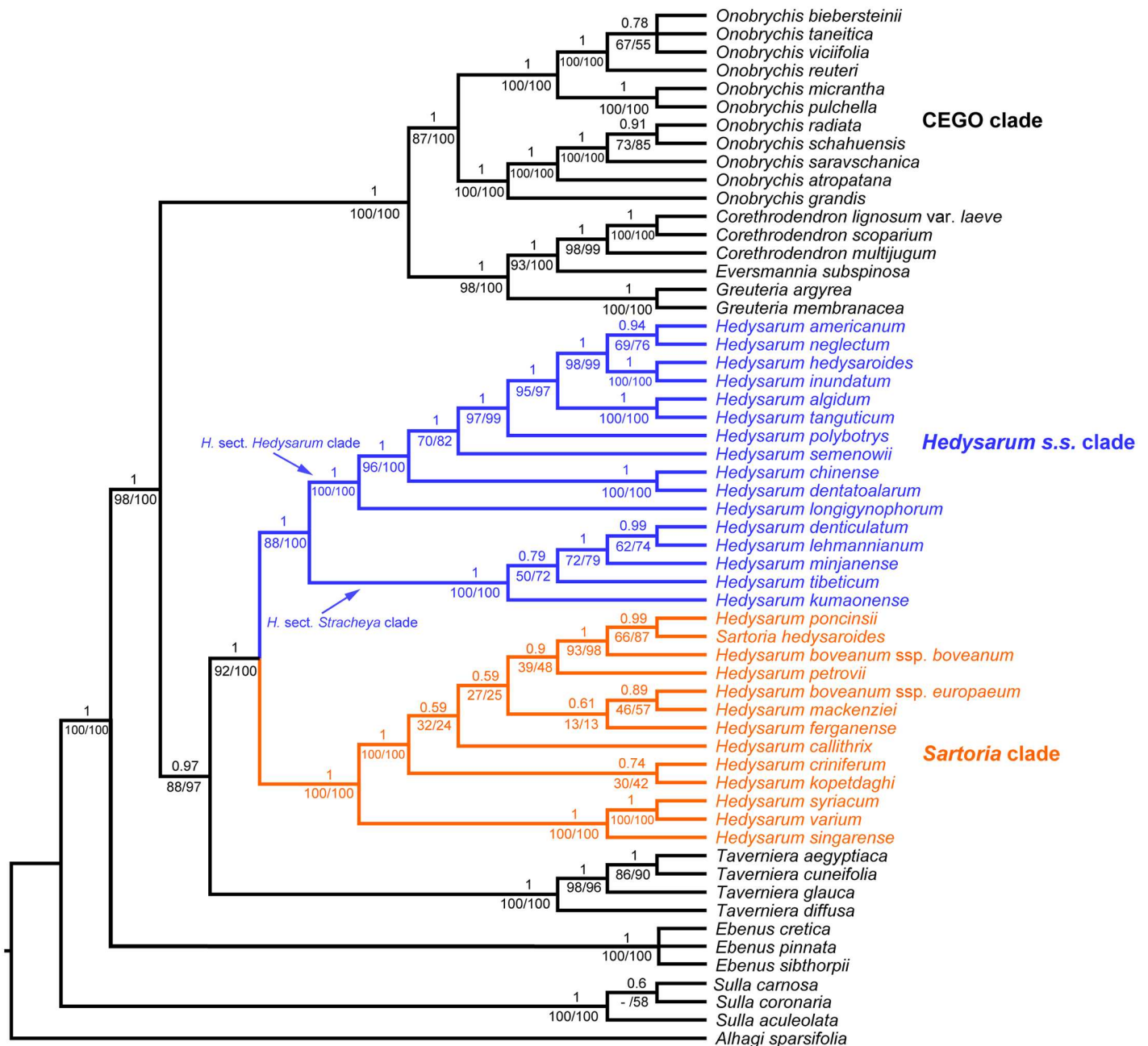


Fig 2. Bayesian tree based on the concatenated plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG*, and *petN-psbM* sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. A dash indicates a branch that is not found in the maximum parsimony tree.

doi:10.1371/journal.pone.0170596.g002

topological incongruence concerning *Hedysarum* is robustly supported (Figs 1 and 2). The difference between the nuclear and the plastid topologies lies on the position of the *Hedysarum* s. s. clade, with the *Hedysarum* s.s. clade sister to the CEGO clade in the nuclear tree, in contrast with its sister relationship with the *Sartoria* clade in the plastid tree (Fig 3).

Our trees based on concatenated nuclear and plastid data are well supported concerning the position of the *Hedysarum* s.s. clade (Figs 1 and 2). However, there is concern that data

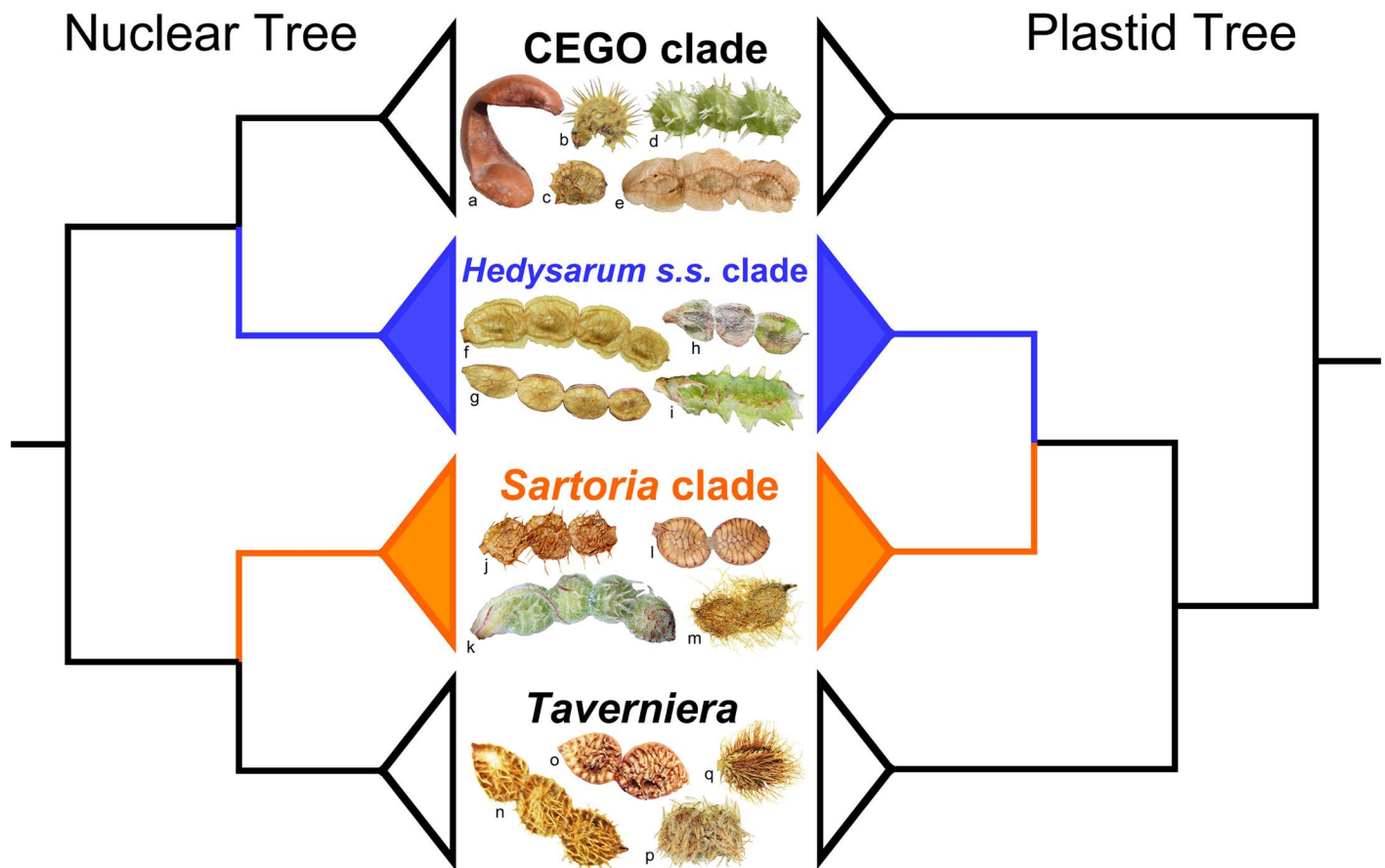


Fig 3. Summary of the incongruent position of the *Hedysarum s.s.* clade in the nuclear (left) and the plastid (right) trees, with selected legume species from each clade (middle). The legumes: a, *Eversmannia subspinosa* B.Fedtsch.; b, *Onobrychis atropatana* Boiss.; c, *O. viciifolia* Scop.; d, *Corethroedron multijugum* (Maxim.) B.H.Choi & H.Ohashi; e, *Greuteria membranacea* (Coss. & Balansa) Amirahm. & Kaz.Osaloo; f, *Hedysarum semenowii* Regel & Herder; g, *H. americanum* (Michx.) Britton; h, *H. kumaonense* Benth. ex Baker; i, *H. tibeticum* (Benth.) B.H.Choi & H.Ohashi; j, *H. gremiale* Rollins; k, *H. petrovii* Yakovlev; l, *H. boveanum* Bunge ex Basiner; m, *H. callithrix* Bunge ex Boiss.; n, *Taverniera diffusa* (Cambess.) Thulin; o, *T. glauca* Edgew.; p, *T. lappacea* DC.; q, *T. longisetosa* Thulin; photoed by PLL and LD.

doi:10.1371/journal.pone.0170596.g003

concatenation may lead to overconfidence in incorrect species tree [55]. Thus we also used the coalescent method and BCA to infer the phylogenetic relationships using multi-locus data. Our coalescent trees (S5 and S7 Figs) agree with the concatenated trees (Figs 1 and 2), and they are also well supported. In BCA, the concordance factor (CF) of a certain clade is the proportion of the genes or the genome for which that given clade is true, and should not be confused with a clade support value [53, 56]. In our BCA results (S6 and S8 Figs), CFs of many clades are below 1, which is likely due to the limited number of phylogenetic informative sites in each individual locus especially in the *PGDH*, *SQD1*, *TRPT* and *psbA-trnH* sequences (Table 1). The single locus trees show limited resolution with many polytomies, such as the *SQD1* tree (S4 Fig). The nuclear sequences have fewer informative sites than the plastid ones (Table 1), and the CFs from the nuclear BCA are lower than those from the plastid BCA (S6 and S8 Figs). However, in the nuclear as well as in the plastid BCA results, the primary concordance tree is built from the clades with highest CFs and shows the relationships shared by a large proportion of the sampled genes [56]. It agrees with the concatenated data tree and the multi-locus coalescent tree. Closer inspection of the nuclear BCA result shows that in the list of clades which are not in the nuclear primary concordance tree but with estimated CF > 0.05, there is no such

clade of a monophyletic *Hedysarum* as in the plastid phylogeny. Also, in the list of clades which are not in the plastid primary concordance tree but with estimated CF > 0.05, there is no contradictory clade of a biphyetic *Hedysarum* as in the nuclear phylogeny. In our analyses, various methods of phylogenetic reconstructions, i.e., BI, MP and ML based on concatenated data, multi-locus coalescent, and BCA, all uncover the same incongruent pattern between the nuclear and the plastid trees (Fig 3).

Various mechanisms have been proposed to explain topological incongruence between different gene trees, such as hybridization/introgression, lineage sorting, allopolyploidy, horizontal gene transfer, and many other factors [14, 57–59].

Insufficient data is probably responsible for the low resolution of the *SQD1* tree, however it can be ruled out from the candidate mechanisms that explain the incongruent position of the *Hedysarum s.s.* clade because both the nuclear and plastid topologies are robustly supported (Figs 1 and 2, S3 Fig). Gene choice and rate of molecular evolution [58] can also be ruled out for the same reason. Sequencing error probably affects phylogenetic accuracy where there are a small number of informative sites [58]. In our trees, sequencing error at most only influences the closely related terminals where the support is low, but should not affect well supported deep relationships. Our sampling covers all genera of the tribe Hedysareae and all sections and major infra-sectional clades in *Hedysarum* as recognized by Duan et al. [11]. Furthermore the different sequences were generated from the same DNA extraction (except for the *PGDH* sequence of *Sulla coronaria*, see S1 Appendix) for each accession. Insufficient taxon sampling and sample misidentification are thus not likely to be the causes of the deep incongruence.

Incomplete homogenization of tandem repeats and the possibly uniparental inheritance of the nuclear ribosomal gene region, and orthology/paralogy conflation of the nuclear low or single copy genes, may affect the accuracy of phylogenetic reconstruction [13, 58, 60, 61]. These issues may lead to incongruence between different nuclear gene trees. We cannot fully rule out these issues in our nuclear data. However, the results based on ribosomal gene regions (S1 Fig) and low or single copy genes (S2 and S4 Figs) all support a biphyetic *Hedysarum*. Furthermore, the monophyly of *Hedysarum* is not detected in any of the nuclear trees. Therefore, the incongruence between the nuclear and the plastid trees cannot be simply explained by such mechanisms.

Allopolyploidy can be excluded from being considered as a mechanism to explain the incongruent position of the *Hedysarum s.s.* clade. We checked the chromosome number counts for Hedysareae in the *Index to Plant Chromosome Numbers* (IPCN, <http://www.tropicos.org/Project/IPCN>). Most species in *Hedysarum* as well as those of other genera in Hedysareae have been reported to have chromosome numbers $2n = 16$ or $2n = 14$, suggesting that they are diploid because their basic chromosome numbers are indicated as $x = 8$ or $x = 7$ [2, 62]. To summarize the available data from IPCN, deep lineages of Hedysareae are predominantly diploid, e.g., $2n = 2x = 16$ for *Alhagi*, *Sulla* (species name still under *Hedysarum* in IPCN; with one report of $2n = 18$), *Taverniera*, *Corethroedendron* (species name still under *Hedysarum* in IPCN), *Eversmannia*, the *Hedysarum* sect. *Stracheya* clade (*H. minjanense* and *H. denticulatum*), and $2n = 2x = 14$ for *Ebenus* (with one report of $2n = 18$). Most species in the *Sartoria* clade, the *H.* sect. *Hedysarum* clade and *Onobrychis* are diploid with chromosome numbers $2n = 14$ or 16 . Polyploidy in *Hedysarum* and *Onobrychis* occurs mostly as multiple cytotypes within one species (e.g., $2n = 14, 28$ for *H. hedysaroides* Schinz & Thell., $2n = 16, 32$ for *H. mackenziei* Richardson), with some being exclusively polyploid such as $2n = 28$ for *H. inundatum* Turcz. and *O. biebersteinii* Sirj. [63].

Occasionally horizontal gene transfer between species through nonsexual means may explain gene tree incongruence [58]. But the variational pattern in our study would require multiple horizontal transfers in the nuclear genes (*ETS*, *ITS*, *PGDH*, *TRPT*, *SQD1*) in a

particular ancestor of *Hedysarum*, which simultaneously leads to discordant gene trees with the plastid tree. This hypothesis is viewed difficult, as many unlinked genes are not likely to be horizontally transferred in parallel [64]. Stegemann et al. [65] showed that an entire chloroplast genome could be horizontally transferred between sexually incompatible species through grafting. Their results were based on artificial grafting and selection process. It is not known whether chloroplast genome could be horizontally transferred through natural grafting. Additionally, unlike in woody plants, species of Hedysareae are not likely to form natural grafting because most species are herbaceous perennials.

Lineage sorting of ancestral polymorphisms can cause incongruent gene trees. The phylogenetic consequence of lineage sorting is usually difficult to distinguish from that of the other processes, such as hybridization/introgression [58, 66]. In general, lineage sorting is expected to be a possible cause of incongruence at lower taxonomic ranks often in population and/or species-level studies [58]. But the incongruence in our results involves deep relationships among different genera. Lineage sorting may contribute to incongruence at higher taxonomic ranks when polymorphisms are maintained by balancing selection through many speciation events [58]. Concerning the position of the *Hedysarum s.s.* clade, however, our tree based on coding genes (*PGDH*, *TRPT*) agrees with the tree based on non-coding sequences (ETS, ITS), which are thought to be neutral. Our tree based on the plastid non-coding data (Fig 2) also agrees with the tree based on the plastid coding *matK* gene in a previous study [10]. Hence lineage sorting may not be the best explanation for the incongruence in our results.

Our phylogenetic pattern may be best explained by a chloroplast capture hypothesis via introgression. In most angiosperm taxa, the plastid genome is uniparentally inherited, whereas the nuclear genome is biparentally inherited [13, 67]. After many generations of backcrossing through introgression, the introgressant inherited the plastid genome from one parent, and nearly all nuclear genes from the other parent. A significant feature of chloroplast capture is that plastid genome introgression is always not accompanied by nuclear introgression [14, 58, 65]. Our results are consistent with this scenario, i.e., multiple nuclear datasets are not supporting the monophyletic *Hedysarum* as suggested by the plastid tree. Chloroplast capture is probably one of the most common causes of phylogenetic incongruence [58], and has been reported in a wide range of taxa at various taxonomic levels [14, 15, 68–75].

Our phylogenetic analyses of the nuclear sequences show that *Hedysarum* is not monophyletic. Its generic delimitation needs to be reconsidered. The *Sartoria* clade and the *Hedysarum s.s.* clade can be distinguished from each other by several morphological characters. Most species in the *Sartoria* clade have grayish-green leaves, obscure lateral veins in the leaflets, biconvex and non-winged loments which always possess prickles, bristles, and ribs, whereas species in the *Hedysarum s.s.* clade always have bright-green leaves, visible lateral veins in the leaflets, compressed and more or less winged loments which are always unarmed [2, 3, 5]. Loment morphology of the *Sartoria* clade is very similar to that of *Taverniera* by its prickles, bristles, and ribs [2, 76], corresponding to the sister relationship between the *Sartoria* clade and *Taverniera* in the nuclear tree (Fig 3). The compressed, unarmed, more or less winged loment is also found in *Greuteria* and *Eversmannia* in the CEGO clade (Fig 3) [10]. Mironov [77] found that species of *H. sect. Gamotion* (members of the *Hedysarum s.s.* clade) differed from species of *H. sect. Multicaulia* (members of the *Sartoria* clade) in several pericarp anatomical characters that were thought to be of systematic significance.

Additionally, the *Hedysarum s.s.* clade and the *Sartoria* clade also show eco-geographical differentiation. In general, members of the *Hedysarum s.s.* clade adapt to mesic and/or psychric habitats in temperate montane forests, alpine and arctic regions of Eurasia and North America, whereas species of the *Sartoria* clade are distributed in xeric habitats in arid/semi-arid areas

and steppes of central and western Asia, the Mediterranean region and western North America [11, 78].

Systematic position of the *Hedysarum* sect. *Stracheya* clade

Stracheya was originally established as a monotypic genus [9], which was then treated as a section of *Hedysarum* based on comprehensive morphological analyses [2, 3]. *Hedysarum tibeticum* was the sole member of *H. sect. Stracheya* when Choi and Ohashi first constructed this section in *Hedysarum* [2]. However, previous phylogenetic analyses showed different systematic positions of the *H. sect. Stracheya* clade in Hedysareae with low support values [10, 11]. With more sequence data, our nuclear and plastid trees resolved congruent sister relationship between the *H. sect. Hedysarum* clade and a clade of five species including *H. tibeticum*, *H. kumaonense*, *H. lehmannianum*, *H. denticulatum*, and *H. minjanense* (Figs 1 and 2, S3 Fig). Our analysis supports merging *Stracheya* with *Hedysarum* and expanding *H. sect. Stracheya* to include four additional species. These four species were previously treated as members of *H. sect. Subacaulia* because of their strongly reduced stems [5, 20]. However, they did not form a clade with other members of *H. sect. Subacaulia* such as *H. petrovii*, *H. ferganense* Korsh., and *H. poncinsii* (Figs 1 and 2, S3 Fig). Some species of *H. sect. Hedysarum*, for example, *H. tanguticum* B.Fedtsch. and *H. pseudoastragalus* Ulbr., also have reduced stems [5]. Thus, the strongly reduced stem is most likely a homoplasious convergence in *Hedysarum*. The dwarf habit of the *H. sect. Stracheya* lineage might be an adaptation to the cold and arid habitats in the pan-Himalayan region and the adjacent eastern part of central Asia [79, 80].

Supporting Information

S1 Appendix. Taxon names, geographical locality, voucher information and herbarium codes, and GenBank accession numbers for the sequences used in this study. Herbarium codes follow the Index Herbariorum (<http://sweetgum.nybg.org/ih/>). For each taxon, GenBank accession numbers of the ten sequences are given in the sequence of ETS, ITS, *PGDH*, *SQD1*, *TRPT*, *trnL-trnF*, *psbA-trnH*, *trnS-trnG*, *trnC-petN*, and *petN-psbM*. New sequences generated in this study are indicated by an asterisk (*). Missing sequences are indicated by a dash (-).

(XLS)

S1 Fig. Maximum parsimony tree based on the concatenated nuclear ITS and ETS sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. Dashes indicate branches that are not found in the Bayesian tree or the maximum likelihood tree.

(TIF)

S2 Fig. Bayesian tree based on the concatenated nuclear *PGDH* and *TRPT* sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. Dashes indicate branches that are not found in the maximum parsimony tree or the maximum likelihood tree.

(TIF)

S3 Fig. Bayesian tree based on the concatenated nuclear ITS, ETS, *PGDH* and *TRPT* sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. Dashes indicate branches that are not found in the maximum parsimony or the

maximum likelihood trees.
(TIF)

S4 Fig. Bayesian tree based on nuclear *SQD1* sequences. The Bayesian posterior probabilities are above the branches, and the maximum parsimony (left) and maximum likelihood (right) bootstrap support values are below the branches. Dashes indicate branches that are not found in the maximum parsimony or the maximum likelihood trees.
(TIF)

S5 Fig. Coalescent tree based on nuclear *ETS*, *ITS*, *PGDH*, *SQD1* and *TRPT* sequences. The posterior probabilities are above the branches.
(TIF)

S6 Fig. Primary concordance tree based on nuclear *ETS*, *ITS*, *PGDH*, *SQD1* and *TRPT* sequences. The sample-wide posterior mean concordance factors are above the branches, and their 95% credibility intervals are below the branches.
(TIF)

S7 Fig. Coalescent tree based on plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG* and *petN-psbM* sequences. The posterior probabilities are above the branches.
(TIF)

S8 Fig. Primary concordance tree based on plastid *psbA-trnH*, *trnC-petN*, *trnL-trnF*, *trnS-trnG* and *petN-psbM* sequences. The sample-wide posterior mean concordance factors are above the branches, and their 95% credibility intervals are below the branches.
(TIF)

Acknowledgments

We thank curators of the United States National Herbarium (US) and the Field Museum (F) for permission to sample their specimens. Laboratory work was conducted in the Laboratories of Analytical Biology of the National Museum of Natural History, the Smithsonian Institution. We are grateful to Xue Yang, Ning Zhang, Gabriel Johnson, Matthew Kweskin and Jeffrey Hunt for their assistance in the experiments and data analyses, and Ashley Egan for helpful discussions in the early phase of the study.

Author Contributions

Conceptualization: ZYC.

Data curation: PLL LD.

Formal analysis: PLL EA KE.

Funding acquisition: ZYC PLL.

Investigation: PLL LD.

Methodology: JW.

Project administration: ZYC JW.

Resources: JW ZYC.

Visualization: PLL.

Writing – original draft: PLL.

Writing – review & editing: JW LD ZYC EA KE.

References

1. LPWG (The Legume Phylogeny Working Group). Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. *Taxon*. 2013; 62: 217–248.
2. Choi B-H, Ohashi H. Generic criteria and an infrageneric system for *Hedysarum* and related genera (Papilionoideae-Leguminosae). *Taxon*. 2003; 52: 567–576.
3. Xu L-R, Choi B-H. *Hedysarum* L. In: Wu Z-Y, Raven PH, Hong D-Y, editors. *Flora of China*. Beijing: Science Press, St. Louis: Missouri Botanical Garden Press; 2010. Volume 10: 514–525.
4. Sultanov B. *Hedysarum neglectum* Ldb. and *Hedysarum austrosibiricum* B.Fedtsch. in *Flora of Kirghizia*. In: Nikitina EV, Sudnicyna IG, Aidarov RA, Sultanov RM, Arbaeva ZS, Sultanov B et al., editors. *Flora of Kirghiz SSR Supplement*. Frunze: Ilim Publishing House; 1970. Volume 2: 31–34.
5. Xu L-R. *Hedysarum* L. In: Cui H-B, editor. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press; 1998. Volume 42(2): 176–221.
6. Linnaeus C. *Species plantarum*. 1st ed. Stockholm: Laurentii Salvii; 1753.
7. Jarvis C. *Order out of Chaos: Linnaean Plant Names and their Types*. 1st ed. London: The Linnean Society of London & the Natural History Museum, London; 2007.
8. Fedtschenko B. *Generis Hedysari revisio*. *Trudy Imp S-Peterburgsk Bot Sada*. 1902; 19: 183–349.
9. Bentham G. On three new genera connected with the Indian Flora. *Hooker's J Bot Kew Gard Misc*. 1853; 5: 304–309.
10. Amirahmadi A, Osaloo SK, Moein F, Kaveh A, Maassoumi AA. Molecular systematics of the tribe Hedy-sareae (Fabaceae) based on nrDNA ITS and plastid *trnL-F* and *matK* sequences. *Plant Syst Evol*. 2014; 300: 729–747.
11. Duan L, Wen J, Yang X, Liu P-L, Arslan E, Ertuğrul K, et al. Phylogeny of *Hedysarum* and tribe Hedysar- eae (Leguminosae: Papilionoideae) inferred from sequence data of ITS, *matK*, *trnL-F* and *psbA-trnH*. *Taxon*. 2015; 64: 49–64.
12. Choi H-K, Luckow MA, Doyle J, Cook DR. Development of nuclear gene-derived molecular markers linked to legume genetic maps. *Mol Genet Genomics*. 2006; 276: 56–70. doi: [10.1007/s00438-006-0118-8](https://doi.org/10.1007/s00438-006-0118-8) PMID: [16642337](https://pubmed.ncbi.nlm.nih.gov/16642337/)
13. Zimmer EA, Wen J. Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol Phy- logen Evol*. 2012; 65: 774–785.
14. Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol Trends Plants*. 1991; 5: 65–84.
15. Soltis DE, Kuzoff RK. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution*. 1995; 49: 727–742.
16. Baldwin BG, Markos S. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycaenia* (Compositae). *Mol Phylogen Evol*. 1998; 10: 449–463.
17. Farruggia FT, Howard AJ. Examination of five nuclear markers for phylogenetic study of Hologalegina (Leguminosae). *Brittonia*. 2011; 63: 489–499.
18. Babineau M, Gagnon E, Bruneau A. Phylogenetic utility of 19 low copy nuclear genes in closely related genera and species of caesalpinoid legumes. *S Afr J Bot*. 2013; 89: 94–105.
19. Li M, Wunder J, Bissoli G, Scarponi E, Gazzani S, Barbaro E, et al. Development of COS genes as uni- versally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics*. 2008; 24: 727–745.
20. Fedtschenko B. *Hedysarum* L. In: Komarov VL, editor. *Flora of USSR*. Moskva-Leningrad: Izdatelstvo Akademii Nauk SSSR; 1948. Volume 13: 259–319.
21. Hedge IC. *Hedysarum* L. In: Davis PH, editor. *Flora of Turkey and the East Aegean Islands*. Edinburgh: Edinburgh University Press; 1970. Volume 3: 549–560.
22. Valdés B. *Hedysarum* L. In: Talavera S, Aedo C, Castroviejo S, Herrero A, Romero Zarco C, Salgueiro FJ, et al., editors. *Flora Iberica*. Madrid: Real Jardín Botánico, CSIC; 2000. Volume 7(2): 943–955.
23. Stanford AM, Harden R, Parks CR. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *matK* and ITS sequence data. *Am J Bot*. 2000; 87: 872–882. PMID: [10860918](https://pubmed.ncbi.nlm.nih.gov/10860918/)
24. Sang T, Crawford DJ, Stuessy TF. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am J Bot*. 1997; 84: 1120–1136. PMID: [21708667](https://pubmed.ncbi.nlm.nih.gov/21708667/)
25. Lee C, Wen J. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. *Mol Phylogen Evol*. 2004; 31: 894–903.

26. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot.* 2007; 94: 275–288. doi: [10.3732/ajb.94.3.275](https://doi.org/10.3732/ajb.94.3.275) PMID: [21636401](https://pubmed.ncbi.nlm.nih.gov/21636401/)
27. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot.* 2005; 92: 142–166. doi: [10.3732/ajb.92.1.142](https://doi.org/10.3732/ajb.92.1.142) PMID: [21652394](https://pubmed.ncbi.nlm.nih.gov/21652394/)
28. Hager M, Biehler K, Illerhaus J, Ruf S, Bock R. Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome *b₆f* complex. *EMBO J.* 1999; 18: 5834–5842. doi: [10.1093/emboj/18.21.5834](https://doi.org/10.1093/emboj/18.21.5834) PMID: [10545095](https://pubmed.ncbi.nlm.nih.gov/10545095/)
29. Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, et al. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J Syst Evol.* 2015; 53: 458–468.
30. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol.* 1991; 17: 1105–1109. PMID: [1932684](https://pubmed.ncbi.nlm.nih.gov/1932684/)
31. Hiraishi A, Kamagata Y, Nakamura K. Polymerase chain reaction amplification and restriction fragment length polymorphism analysis of 16S rRNA genes from methanogens. *J Ferment Bioeng.* 1995; 79: 523–529.
32. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012; 28: 1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/)
33. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
34. Müller K. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Appl Bioinformatics.* 2005; 4: 65–69. PMID: [16000015](https://pubmed.ncbi.nlm.nih.gov/16000015/)
35. Simmons MP, Ochoterena H. Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol.* 2000; 49: 369–381. PMID: [12118412](https://pubmed.ncbi.nlm.nih.gov/12118412/)
36. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene dataset with character set and codon information. *Cladistics.* 2011; 27: 171–180.
37. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9: 772.
38. Yokota Y, Kawata T, Iida Y, Kato A, Tanifuji S. Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. *J Mol Evol.* 1989; 29: 294–301. PMID: [2514271](https://pubmed.ncbi.nlm.nih.gov/2514271/)
39. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003; 19: 1572–1574. PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/)
40. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012; 61: 539–542. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)
41. Ronquist F, Huelsenbeck J, Teslenko M. MrBayes version 3.2 manual: tutorials and model summaries. 2011. http://mrbayes.sourceforge.net/mb3.2_manual.pdf
42. Swofford DL. PAUP*: phylogenetic analysis using parsimony (*and other methods) Version 4.0b10. Sunderland, Massachusetts: Sinauer Associates; 2002.
43. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30: 1312–1313. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) PMID: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/)
44. Farris JS, Källersjö M, Kluge AG, Bult C. Testing significance of incongruence. *Cladistics.* 1995; 10: 315–319.
45. Lee MSY. Uninformative characters and apparent conflict between molecules and morphology. *Mol Biol Evol.* 2001; 18: 676–680. PMID: [11264420](https://pubmed.ncbi.nlm.nih.gov/11264420/)
46. Cunningham CW. Can three incongruence tests predict when data should be combined? *Mol Biol Evol.* 1997; 14: 733–740. PMID: [9214746](https://pubmed.ncbi.nlm.nih.gov/9214746/)
47. Lanfear R, Calcott B, Ho SYW, Guindon S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 2012; 29: 1695–1701. doi: [10.1093/molbev/mss020](https://doi.org/10.1093/molbev/mss020) PMID: [22319168](https://pubmed.ncbi.nlm.nih.gov/22319168/)
48. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 2002; 51: 492–508. doi: [10.1080/10635150290069913](https://doi.org/10.1080/10635150290069913) PMID: [12079646](https://pubmed.ncbi.nlm.nih.gov/12079646/)
49. Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics.* 2001; 17: 1246–1247. PMID: [11751242](https://pubmed.ncbi.nlm.nih.gov/11751242/)

50. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
51. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol.* 2010; 27: 570–580. doi: [10.1093/molbev/msp274](https://doi.org/10.1093/molbev/msp274) PMID: [19906793](https://pubmed.ncbi.nlm.nih.gov/19906793/)
52. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014; 10: e1003537. doi: [10.1371/journal.pcbi.1003537](https://doi.org/10.1371/journal.pcbi.1003537) PMID: [24722319](https://pubmed.ncbi.nlm.nih.gov/24722319/)
53. Ané C, Larget B, Baum DA, Smith SD, Rokas A. Bayesian estimation of concordance among gene trees. *Mol Biol Evol.* 2007; 24: 412–426. doi: [10.1093/molbev/msl170](https://doi.org/10.1093/molbev/msl170) PMID: [17095535](https://pubmed.ncbi.nlm.nih.gov/17095535/)
54. Larget BR, Kotha SK, Dewey CN, Ané C. BUCKY: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics.* 2010; 26: 2910–2911. doi: [10.1093/bioinformatics/btq539](https://doi.org/10.1093/bioinformatics/btq539) PMID: [20861028](https://pubmed.ncbi.nlm.nih.gov/20861028/)
55. Edwards SV. Is a new and general theory of molecular systematics emerging? *Evolution.* 2009; 63: 1–19. doi: [10.1111/j.1558-5646.2008.00549.x](https://doi.org/10.1111/j.1558-5646.2008.00549.x) PMID: [19146594](https://pubmed.ncbi.nlm.nih.gov/19146594/)
56. Baum DA. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon.* 2007; 56: 417–426.
57. Doyle JJ. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst Bot.* 1992; 17: 144–163.
58. Wendel JF, Doyle JJ. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ, editors. *Molecular systematics of plants II: DNA sequencing.* Norwell: Kluwer Academic Publishers; 1998. pp. 265–296.
59. Zou X-H, Ge S. Conflicting gene trees and phylogenomics. *J Syst Evol.* 2008; 46: 795–807.
60. Álvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogen Evol.* 2003; 29: 417–434.
61. Blattner FR. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. *Mol Phylogen Evol.* 2004; 33: 289–299.
62. Goldblatt P. Cytology and the phylogeny of Leguminosae. In: Polhill RM, Raven PH, editors. *Advances in Legume Systematics, part 2.* London: Royal Botanic Gardens, Kew; 1981. pp. 427–463.
63. Doyle JJ. Polyploidy in Legumes. In: Soltis PS, Soltis DE, editors. *Polyploidy and genome evolution.* Berlin Heidelberg: Springer-Verlag; 2012. pp. 147–180.
64. Sang T, Zhong Y. Testing hybridization hypotheses based on incongruent gene trees. *Syst Biol.* 2000; 49: 422–434. PMID: [12116420](https://pubmed.ncbi.nlm.nih.gov/12116420/)
65. Stegemann S, Keuthe M, Greiner S, Bock R. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci USA.* 2012; 109: 2434–2438. doi: [10.1073/pnas.1114076109](https://doi.org/10.1073/pnas.1114076109) PMID: [22308367](https://pubmed.ncbi.nlm.nih.gov/22308367/)
66. Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ. The reticulate history of *Medicago* (Fabaceae). *Syst Biol.* 2008; 57: 466–482. doi: [10.1080/10635150802172168](https://doi.org/10.1080/10635150802172168) PMID: [18570039](https://pubmed.ncbi.nlm.nih.gov/18570039/)
67. Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011; 76: 273–297. doi: [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4) PMID: [21424877](https://pubmed.ncbi.nlm.nih.gov/21424877/)
68. Wolfe AD, Elisens WJ. Evidence of chloroplast capture and pollen-mediated gene flow in *Penstemon* sect. *Peltanthera* (Scrophulariaceae). *Syst Bot.* 1995; 20: 395–412.
69. Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, et al. Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Mol Biol Evol.* 2005; 22: 285–296. doi: [10.1093/molbev/msi016](https://doi.org/10.1093/molbev/msi016) PMID: [15483320](https://pubmed.ncbi.nlm.nih.gov/15483320/)
70. Fehrer J, Gemeinholzer B, Chrtek J, Bräutigam S. Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Mol Phylogen Evol.* 2007; 42: 347–361.
71. Acosta MC, Premoli AC. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Mol Phylogen Evol.* 2010; 54: 235–242.
72. Xu B, Wu N, Gao X-F, Zhang L-B. Analysis of DNA sequences of six chloroplast and nuclear genes suggests incongruence, introgression, and incomplete lineage sorting in the evolution of *Lespedeza* (Fabaceae). *Mol Phylogen Evol.* 2012; 62: 346–358.
73. Deng T, Nie Z-L, Drew BT, Volis S, Kim C, Xiang C-L, et al. Dose the Arcto-Tertiary biogeographic hypothesis explain the disjunct distribution of northern hemisphere herbaceous plants? The case of *Meehania* (Lamiaceae). *PLoS ONE.* 2015; 10: e0117171. doi: [10.1371/journal.pone.0117171](https://doi.org/10.1371/journal.pone.0117171) PMID: [25658699](https://pubmed.ncbi.nlm.nih.gov/25658699/)

74. Yi T-S, Jin G-H, Wen J. Chloroplast capture and intra- and inter-continental biogeographic diversification in the Asian—New World disjunct plant genus *Osmorhiza* (Apiaceae). *Mol Phylogen Evol.* 2015; 85: 10–21.
75. Duan L, Yang X, Liu P, Johnson G, Wen J, Chang Z. A molecular phylogeny of Caraganeae (Leguminosae, Papilionoideae) reveals insights into new generic and infrageneric delimitations. *PhytoKeys.* 2016; 70: 111–137.
76. Thulin M. Revision of *Taverniera* (Leguminosae-Papilionoideae). *Symb Bot Upsal.* 1985; 25: 43–95.
77. Mironov YM. Pericarp anatomy of East European species of the genus *Hedysarum* L. (Papilionaceae): sections *Gamotion* and *Multicaulia*. *Bull Mosc Soc Nat Ser Biol.* 2000; 105: 50–53.
78. Xu L-R. The ecological differentiation of the *Hedysarum* L. and geographical distribution in China. *Acta Bot Bor-Occ Sinica.* 1985; 4: 275–285.
79. Wen J, Zhang J-Q, Nie Z-L, Zhong Y, Sun H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front Genet.* 2014; 5: 4. doi: [10.3389/fgene.2014.00004](https://doi.org/10.3389/fgene.2014.00004) PMID: [24575120](https://pubmed.ncbi.nlm.nih.gov/24575120/)
80. Meng H-H, Gao X-Y, Huang J-F, Zhang M-L. Plant phylogeography in arid Northwest China: retrospectives and perspectives. *J Syst Evol.* 2015; 53: 33–46.