Anales del Jardín Botánico de Madrid 80 (1): e133
https://doi.org/10.3989/ajbm.2630
ISSN-L: 0211-1322

Supermatrix analyses reveal the importance of outgroup, gene and taxon sampling in *Onosma* (Boraginaceae) phylogenetics

Deniz AYGÖREN ULUER[1*]

[1]Department of Plant and Animal Production, Ahi Evran University, Kirşehir, Turkey

[*]Correspondence: d.aygoren@ahievran.edu.tr
[1]https://orcid.org/0000-0002-2095-3816

Abstract. Tribe Lithospermeae (Boraginaceae) consists of ca. 26 genera and 470 species, in which *Onosma* constitutes approximately one third of the species (~150). Although the tribe is strongly supported as monophyletic, both generic and species boundaries remain ambiguous. Among them, not only the phylogenetic position of Eastern Asian *Onosma* species, but also the taxonomic limits of the genus remain unclear. Whether Eastern Asian *Onosma* is monophyletic, or the genus should be widened to include *Maharanga*, and maybe *Cystostemon*, are still open questions. For these reasons, I performed 16 phylogenetic analyses with different taxon coverages, alignments, gene regions and outgroups, with up to 746 taxa of tribe Lithospermeae and with five DNA regions, using data from GenBank. The results, with the widest taxon coverage to date, show that while genus *Onosma* is not monophyletic in any of the analyses, the phylogenetic relationships among *Onosma* s.s., Eastern Asian *Onosma*, *Maharanga* and *Cystostemon* differ among analyses. However, the approximately unbiased (AU) test showed that the topology (((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) *Onosma* s.s.) is overwhelmingly supported. Therefore, the current study highlights the importance of taxon, gene and outgroup sampling in *Onosma* phylogenetics.

Keywords. *Cystostemon*, Lithospermeae, *Maharanga,* *Onosma*, outgroup, phylogeny, taxon/gene sampling.

Resumen. La tribu Lithospermeae (Boraginaceae) consta de ca. 26 géneros y 470 especies, en las que *Onosma* constituye aproximadamente un tercio de las especies (~150). Si bien la tribu ha sido fuertemente apoyada como monofilética, los límites taxonómicos a nivel de género y especie son todavía ambiguos. Entre ellos, no solo la posición filogenética de las especies de *Onosma* de Asia oriental, sino también la delimitación del género son confusas. Si *Onosma* de Asia oriental es monofilético o si el género debe ampliarse para incluir a *Maharanga* y tal vez *Cystostemon* son todavía cuestiones por resolver. Por estas razones, realicé 16 análisis filogenéticos con diferentes coberturas de taxones, alineamientos, regiones de genes y grupos externos, con hasta 746 taxones de la tribu Lithospermeae y con cinco regiones de ADN, usando datos de GenBank. Los resultados, con la cobertura taxonómica más amplia hasta la fecha, demuestran que si bien el género *Onosma* no resultó monofilético en ninguno de los análisis, las relaciones filogenéticas entre *Onosma* s.s., *Onosma* de Asia oriental, *Maharanga* y *Cystostemon* difieren entre análisis. Sin embargo, el test AU ("approximately unbiased") mostró que la topología (((*Onosma*-Asia Oriental+Maharanga) *Cystostemon*) *Onosma* s.s.) es ampliamente compatible. Por lo tanto, este trabajo destaca la importancia del muestreo de taxones, genes y grupos externos en la filogenética de *Onosma*.

Palabras clave. *Cystostemon,* Lithospermeae, filogenia, grupo externo, *Maharanga,* muestreo de taxones/genes, *Onosma*.

How to cite this article: Aygören Uluer D. 2023. Supermatrix analyses reveal the importance of outgroup, gene and taxon sampling in *Onosma* (Boraginaceae) phylogenetics. *Anales del Jardín Botánico de Madrid* 80: e133. https://doi.org/10.3989/ajbm.2630

Title in Spanish: Análisis de supermatrices demuestran la importancia del muestreo de grupos externos, genes y taxones en la filogenética de *Onosma* (Boraginaceae).

Associate editor: Javier Fuertes-Aguilar. Received: 17 January 2022; accepted: 23 February 2023; published online: 14 June 2023.

INTRODUCTION

The family Boraginaceae Juss. (Boraginales) is subdivided in five subfamilies, including subfamily Boraginoideae Arn. with two tribes, namely Boragineae and Lithospermeae (Chacón & al. 2019). Tribe Lithospermeae with ca. 26 genera and 470 species (Chacón & al. 2019) is characterized by several morphological traits: herbaceous habit, sympetalous corolla with generally basal and faucal scales, fruit with four nutlets with basal attachments (Cohen 2014; Weigend & al. 2016). While Lithospermeae has been strongly supported as monophyletic (Thomas & al. 2008; Cohen & Davis 2009; Cohen 2014; Selvi & al. 2017), within the tribe both generic and species boundaries are ambiguous, and the phylogenetic relationships differ among studies (Weigend & al. 2009; Cohen 2011; Cohen 2014; Chacón & al. 2019).

Onosma L. is the largest genus of Lithospermeae with ~150 species (Chacón & al. 2017). It is distributed in north-western Africa, Europe and Asia, but the centre of diversity is in the dry, rocky and sunny habitats of Turkey and

Iran (Mehrabian & al. 2011; Nasrollahi & al. 2019). While *Onosma* has been considered as a taxonomically difficult group; only a few morphological and molecular studies have focussed on this genus (e.g., Cecchi & al. 2011; Nasrollahi & al. 2019). Similarly, to date, its sampling in the previous studies was limited [Table 1]. For instance, while Nasrollahi & al. (2019) sampled 122 *Onosma* samples (87 species), they included only five outgroups in their analyses. Similarly, while Chacón & al. (2019) included 257 Lithospermeae samples (180 species and 13 subspecies), their sampling of *Onosma* remained insufficient with only 64 newly sequenced samples (50 species and 4 subspecies). Furthermore, some recently described taxa (e.g., *Onosma atila-ocakii* O.Koyuncu & Yaylaci; Koyuncu & al. 2013) have never been included in a phylogenetic study, and their phylogenetic position remain unknown. In addition, it was reported that genus *Onosma* is not monophyletic, with the Eastern Asian genus *Maharanga* recovered as closely related to Eastern Asian *Onosma*, namely, *O. rostellata* Lehm., *O. paniculata* Bureau & Franch., *O. hookeri* C.B. Clarke, *O. waltonii* Duthie, *O. sinicum* and *O. pyramidalis* Hook.f. (Cecchi & al. 2011; Chacón & al. 2019; Nasrollahi & al. 2019), so not only the phylogenetic position of the Eastern Asian species, but also the taxonomic limits of the genus *Onosma* (i.e., whether Eastern Asian *Onosma*+*Maharanga* constitutes a different genus) remain unclear.

For these reasons, molecular analyses of *Onosma* with a wide taxon coverage and more genetic regions are needed. However, such a worldwide study would be both time and money consuming. Therefore, before attempting such an expensive job, a supermatrix analysis which covers more taxa and all available published sequences would be the first step to overcome these problems. With this aim, in the current study, I performed 16 phylogenetic analyses with

Table 1. Comparison of taxon sampling of the present study and the previous studies on genus *Onosma* and/or tribe Lithospermeae. The highest taxon coverage is indicated in bold. EA: Eastern Asian.

	Cohen (2011)	**Cohen (2014)**	**Chacón & al. (2019)**	**Nasrollahi & al. (2019)**	**Current study ITS**	**Current study Total evidence**
Total taxa		60	258	122	**746**	**350**
EA *Onosma*			5	8	**21**	**12**
Aegonychon			2		**3**	1
Alkanna		3	28	1	**37**	30
Ancistrocarya			1			
Arnebia	1	2	9		**39**	16
Buglossoides	4	4	4		**50**	20
Cerinthe	1	2	6		**27**	9
Cystostemon		1	3		**4**	**4**
Echium	1	10	55	2	**118**	58
Echiostachys		1	3		3	2
Glandora	2	3	7		**10**	6
Halacsya	1	1	1		**5**	1
Huyhnia		1	2		1	3
Lithodora	2	2	4		**25**	16
Lithospermum	37	9	44		**91**	60
Lobostemon		2	3		**7**	5
Macrotomia		1			1	1
Maharanga	1	1	2	1	2	2
Mairetis	1	1	2		**4**	3
Megacaryon			2		**3**	2
Moltkia	1	5	6		**16**	9
Moltkiopsis		1	1		**7**	4
Neatostema	1	1	2		**4**	3
Onosma	1	6	64	117	**278**	88
Paramoltkia	1	1	1		**5**	2
Podonosma	1	1	2	1	**3**	2
Pontechium		1	2		**3**	**3**
Stenosolenium			1		1	1

Table 2. Total number of taxa, alignment length, total variable characters, and the number of parsimony informative sites for each region sampled in the present study.

Region	Number of taxa	Alignment length (bp)	Total variable characters	Parsimony informative characters (%)
ITS (1)	746	814	611	510 (62.7)
ITS (2)	555	810	555	489 (60.4
ITS (3)	323	802	525	447 (55.)
Plastid	339	2,753	1,073	662 (24)
Total evidence (1)	350	3,555	1,598	1,109 (31.2)
Total evidence (2)	345	3,555	1,579	1,101 (31)
Total evidence with distant OG (*Vahlia*)	347	3,611	1,750	1,321 (36.6)
Total evidence with closely related OG (*Anchusa*)	347	3,629	1,585	1,138 (31.4)

different taxon coverage, gene regions and outgroup(s), with up to 746 taxa and four plastid and one nuclear DNA regions, namely, *rps*16, *trn*S-G spacer, *trn*L gene, *trn*L-F spacer and nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS). I also performed approximately unbiased (AU) (Shimodaira & Hasegawa 1999) test to evaluate the p-values for the phylogenetic relationships among *Onosma*, *Maharanga* and *Cystostemon*.

MATERIAL AND METHODS

Taxon sampling, alignment and phylogenetic analyses

The present study includes up to 746 taxa from 24 genera of tribe Lithospermeae (Table 1). Five loci, namely, nuclear ITS, *trn*L gene, *trn*L-F spacer, *trn*S-G spacer and *rps*16 downloaded from GenBank (Appendices 1 and 2). The number of taxa, alignment length, number of variable characters, and the number of parsimony informative sites for each region are provided in Table 2.

First, for the ITS region, three data matrices with different taxon sampling were created: dataset 1) ITS matrix with missing data (i.e., either ITS1 or ITS2 regions, or both available for each taxon) with 746 taxa; dataset 2) ITS data matrix without missing data with 555 taxa (i.e., both the ITS1 and ITS2 regions are available for each taxon); dataset 3) ITS data matrix without missing data with 323 taxa (i.e., for this dataset, the difficult taxa to align, possibly due to misidentifications and/or ITS region problems, were removed). Second, for the plastid data only one data matrix was created: dataset 4) *trn*L gene, *trn*L-F spacer, *rps*16 and *trn*S-G spacer were included for 339 taxa. Third, for the total evidence analyses (i.e., all five concatenated loci), four matrices were created: dataset 5) total evidence matrix with 350 taxa; dataset 6) total evidence matrix with 345 taxa [five taxa with only one sequence for each region, namely *Halacsya* Dörfl., *Macrotomia* DC. ex Meisn., *Stenosolenium* Turcz., *Lithospermum tschimganicum* B.Fedtsch (= *Ulugbekia*/*Arnebia tschimaganica*) and *Aegonychon* Gray were excluded]. With the addition of two different outgroup taxa (i.e., out of tribe Lithospermeae), namely *Vahlia* Thunb. and *Anchusa* L., two more datasets were created: dataset 7) total evidence matrix with distant outgroups, namely two *Vahlia* taxa, and 345 ingroup taxa; and finally dataset 8) total evidence matrix with closely related outgroups, namely two *Anchusa* taxa, and 345 ingroup taxa (Table 2). Other than these last two datasets (7 and 8), no outgroup(s) out of tribe Lithospermeae were used. In terms of taxon coverage for each dataset, see Table 3. These strategies were used to detect the effect of outgroups (e.g., datasets 5, 6, 7 and 8), different alignments (e.g., dataset 1, 6 and 3), genes (e.g., datasets 3, 4 and 6) and missing data (e.g., datasets 1 and 2) on the phylogenetic relationships within *Onosma* (Smith 1994; Qiu & al. 2001). Sequences were aligned by Geneious Pro 4.8.4 (Kearse & al. 2012) and manually edited for 16 different Maximum likelihood (ML) and Bayesian Inference (BI) analyses (Table 2). ML analyses were conducted with RAxML version 8.2.12 (Stamatakis & al. 2014) on the CIPRES Science Gateway (Miller & al. 2011). The GTRGAMMA model was selected as the best-fit model by using the Akaike information criterion (AIC) in the software jModelTest2.1.10 (Guindon & Gascuel 2013; Darriba & al. 2012), and the "Let RAxML halt bootstrapping automatically" options selected. A rapid bootstrap analysis/search for best-scoring ML tree was performed. BI analyses were implemented using MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003) on the CIPRES Science Gateway (Miller & al. 2011), for 10 million generations, sampling every 100 generations, with a random starting tree. The first 25% of trees were discarded as "Burn-in" and the remaining trees were used to build the consensus tree. Tracer v1.5 (Rambaut & Drummond 2007 was used to inspect the MCMC output and to determine convergence of the two chains. iTOL (Interactive Tree of Life) (Letunic & Bork 2016) was used to create the tree image.

Alternative topology testing

The approximately unbiased (AU) (Shimodaira & Hasegawa 1999) test was used to evaluate the p-values for the possible (((Eastern Asian *Onosma*+*Maharanga*) *Onosma* s.s.) *Cystostemon*) and (((Eastern Asian *Onosma*+*Maha*-

Table 3. Results of the phylogenetic analyses with different genetic regions and different taxon sampling. Both BS (%) and PP (0–1) support values are given for each clade. The difference between the taxon sampling of different analyses is explained in Materials and Methods. Taxon numbers are indicated within parentheses. Empty cells indicate only one sequence was sampled. C-OG, close outgroup; D-OG, distant outgroup; X, the taxon was not recovered as monophyletic. Tree topology: T1, (((Eastern Asian *Onosma+Maharanga*) *Cystostemon*) distantly related to *Onosma* s.s); T2, (((Eastern Asian *Onosma+Maharanga*) *Onosma* s.s) *Cystostemon*)); T3, (((Eastern Asian *Onosma+Maharanga*) *Cystostemon*) *Onosma* s.s).

Clade name	ITS (323)	ITS (555)	ITS (746)	Plastid (339)	Total evidence (350)	Total evidence (345)	Total evidence, D-OG (347)	Total evidence, C-OG (347)
Tree topology	X	T1	T1	X	T2	T3	T3	T3
Onosma s.s.	100; 1.00	100; 1.00	100; 1.00	X	100; 1.00	100; 1.00	93; 1.00	100; 1.00
E Asian *Onosma+ Maharanga*	100; 1.00	100; 1.00	100; 1.00	80; 1.00	98; 1.00	97; 1.00	94; 0.96	96; 1.00
Maharanga	98; 1.00	98; 1.00	98; 1.00	98; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Asian *Onosma*	91; 1.00	87; 0.98	90; 0.99	X	X	X	X	X
Alkanna	84; 1.00	80; 0.95	80; 0.94	87/ 0.80	99; 1.00	99; 1.00	99; 1.00	99; 1.00
Arnebia (excluding *A. hispidissima*)	X	X	X	X	X	87; 1.00	92; 1.00	87; 0.99
Buglossoides	X	X	X	X	X	97; 1.00	87; 0.96	96; 1.00
Cerinthe	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Cystostemon	92; 1.00	92; 0.98	92 0.97	98; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Echiostachys	98; 1.00	95; 1.00	95; 0.99	62; 0.80	99; 1.00	98; 1.00	97; 1.00	98; 1.00
Echium	99; 1.00	91; 0.99	91; 0.99	X	92; 0.99	89; 0.99	92; 1.00	90; 0.97
Halacsya		100; 1.00	100; 1.00					
Huynhia				85; 0.99	89; 1.00	89; 1.00	87; 0.99	86; 0.95
Lithodora I	99; 1.00	98; 1.00	99; 1.00	X	95; 1.00	100; 1.00	100; 1.00	100; 1.00
Lithodora II (including *Glandora*)	100; 1.00	79; 0.96	75; 0.95	X	100; 1.00	96; 1.00	93; 1.00	94; 1.00
Lithospermum	97; 1.00	93; 0.99	98; 1.00	X	100; 1.00	100; 1.00	99; 1.00	100; 1.00
Lobostemon (excluding *L. trigonus*)	100; 1.00	99; 1.00	98; 1.00	X	97; 1.00	97; 1.00	85; 1.00	97; 1.00
Mairetis	100; 1.00	100; 1.00	100; 1.00	98; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Megacaryon	100; 1.00	100; 1.00	100; 1.00	92; 0.86	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Moltkia	100; 1.00	100; 1.00	100; 1.00	63; 0.72	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Moltkiopsis	100; 1.00	100; 1.00	100; 1.00		100; 1.00	100; 1.00	100; 1.00	100; 1.00
Neatostema	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Paramoltkia	100; 1.00	100; 1.00	100; 1.00	X	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Podonosma	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00
Pontechium	100; 1.00	100; 1.00	100; 1.00	96; 1.00	100; 1.00	100; 1.00	100; 1.00	100; 1.00

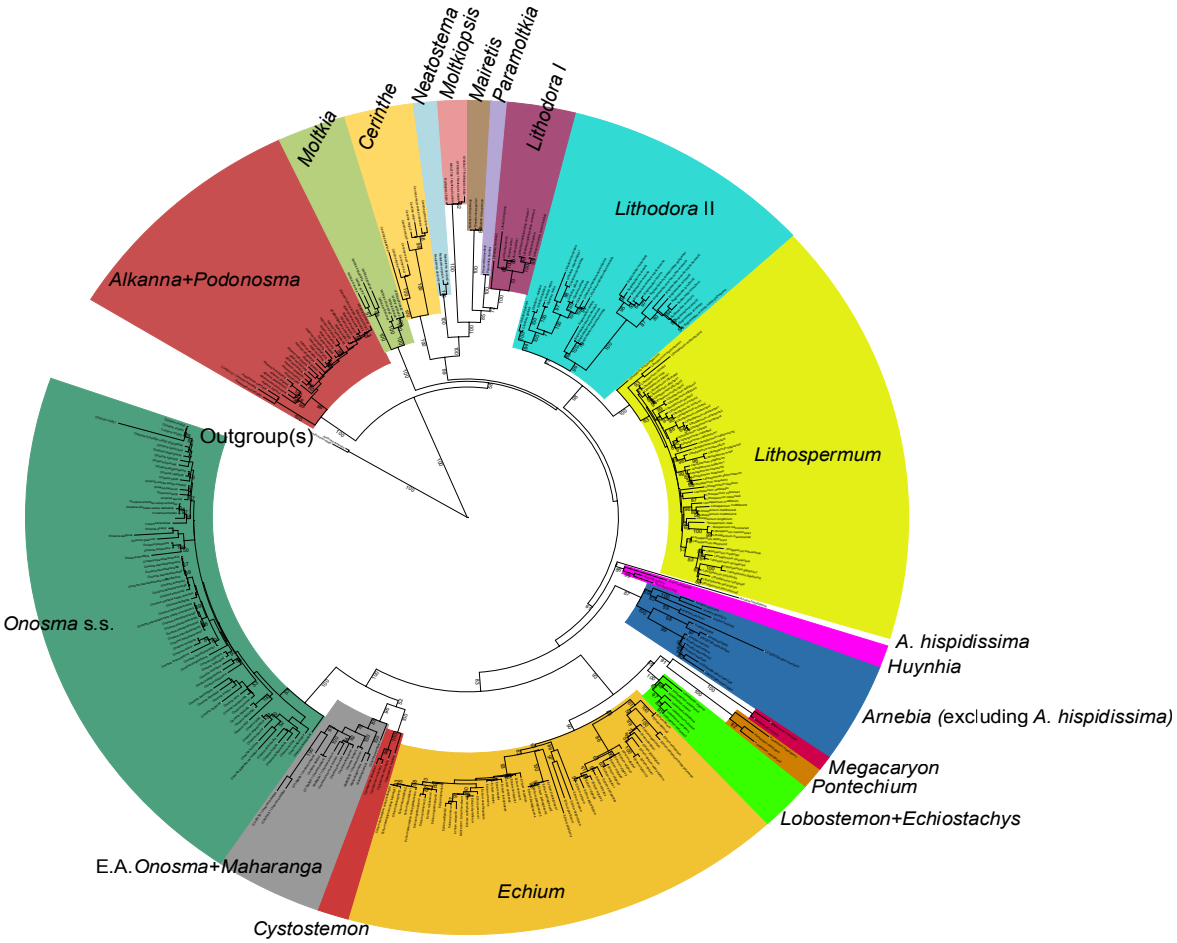

Fig. 1. Summary of the phylogenetic tree from ML analysis of *rps*16+*trn*S-G+*trn*L+*trn*L-F+ITS dataset showing the phylogenetic relationships within tribe Lithospermeae. Bootstrap support values and posterior probabilities are indicated above branches. Outgroups are indicated, Lithospermeae genera are colour coded.

ranga) *Cystostemon*) *Onosma* s.s.) topologies resulted from the phylogenetic analyses by W-IQ-TREE (http://iq-tree.cibiv.univie.ac.at/, Trifinopoulos & al. 2016), and by using the '-au' option and 10,000 bootstrap replicates with the "total evidence dataset with 350 taxa". Note that due to the possible problems with the ITS region (explained below), the AU test was performed just for the two possible topologies, indicated in Figures 2b and 2c.

RESULTS

The genus *Onosma* and its subsections, namely, *Onosma*, *Haplotricha* and *Heterotricha* Riedl (1967) did not resulte monophyletic in any of the analyses (Table 3, Fig. 1) (Appendix 3–10). In contrast, *Onosma* s.s. was monophyletic (93-100% BS, 0.99-1.00 PP) in all analyses, except the plastid matrix analysis, where the clade was not well-resolved

The topologies describing the phylogenetic relationships among Eastern Asian genera *Onosma*, *Maharanga*, *Onosma* s.s and *Cystostemon* differed depending on the analysis (Table 3, Fig. 2). First, *Onosma* s.s. was not sister to the Eastern Asian *Onosma+Maharanga* clade in neither the ITS with 555 taxa nor the ITS with 746 taxa analyses (Fig. 2a) although note that the relationships among *Onosma* s.s., Eastern Asian *Onosma*, *Maharanga* and *Cystostemon* were not resolved in the ITS with 333 taxa and plastid data analyses. Second, the "total evidence tree with 350 taxa" analysis yielded a (((Eastern Asian *Onosma+Maharanga*) *Onosma* s.s.) *Cystostemon*) topology (Fig. 2b). However, a possible (((Eastern Asian *Onosma+Maharanga*) *Cystostemon*) *Onosma* s.s.) relationship was recovered for the first time here, and apart than the distant and closely related outgroup analyses, this topology was also recovered from the "total evidence tree with 345 taxa" analysis (Fig. 2c).

While the Eastern Asian *Onosma+Maharanga* clade was strongly supported as monophyletic (94-100% BS, 1.00 PP) in all analyses, only in the ITS with 555 and 746 taxa analyses the Eastern Asian *Onosma* emerged mono-

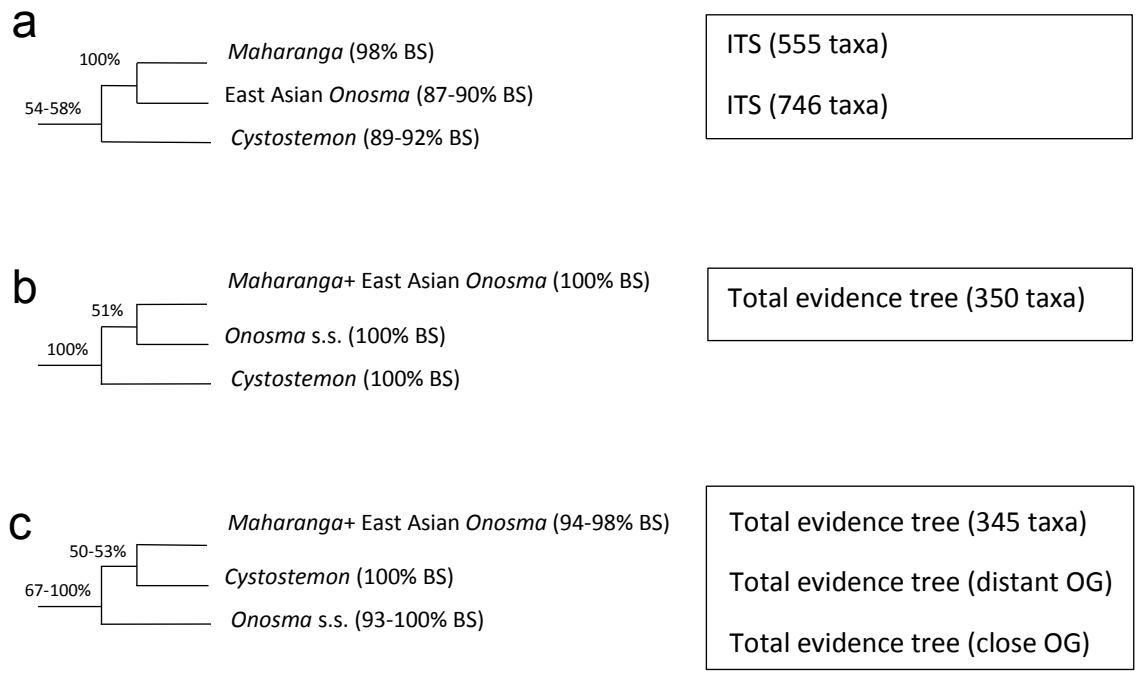

Fig. 2. Phylogenetic relationships within the *Onosma* s.s.+Eastern Asian *Onosma* +*Maharanga*+ *Cystostemon* clade with different gene and different taxon sampling analyses. OG: outgroup, total evidence tree: *rps*16+*trn*S-G+*trn*L+*trn*L-F+ITS; **a,** ((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) topology (in which *Onosma* s.s. was not sister to this clade) recovered from two analyses, namely, ITS with 55 taxa and ITS with 746 taxa; **b,** (((Eastern Asian *Onosma*+*Maharanga*) *Onosma* s.s.) *Cystostemon*) topology recovered from the total evidence analysis with 350 taxa; **c,** (((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) *Onosma* s.s.) topology recovered from three analyses, namely, the total evidence analyses with 345 taxa, distant OG analysis and closely related OG analysis.

phyletic (87-89% BS, 0.96-1.00 PP), in the other analyses the *Maharanga* samples were embedded in the Eastern Asian *Onosma*. Similarly, *Cystostemon* (89–100% BS, 0.93–1.00 PP) and *Maharanga* (98–100% BS, 1.00 PP) were also monophyletic in all analyses. In terms of the differences between nuclear (ITS) and plastid (*trn*L+*trn*L--F+*trn*S-G+*rps*16) datasets, while *Onosma* s.s. (100% BS, 1.00 PP), Eastern Asian *Onosma* (87–91% BS, 0.92–0.98 PP) was monophyletic in all ITS analyses, the clade was not monophyletic in the plastid data analysis (Table 3). Moreover, while the Eastern Asian *Onosma*+*Maharanga* clade received a 100% BS and 1.00 PP in all ITS analyses, in the plastid data analysis this clade was supported with 80% BS and 1.00 PP (Table 3).

The AU test showed that the (((Eastern Asian *Onosma*+*Maharanga*) *Onosma* s.s.) *Cystostemon*) was rejected (p-value = -0.00133), but the (((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) *Onosma* s.s.) topology was overwhelmingly supported (p-value = 0.999) (Table 4). In terms of outgroups (tribe Lithospermeae) *Arnebia* Forssk. (excluding *A. hispidissima* (Lehm.) A.DC.) and *Buglossoides* Moench. were monophyletic only in the total evidence analyses with 345 taxa (total evidence analyses with distant and closely related outgroup analyses) (Table 3). One *Aegonychon* sample was embedded within the *Buglossoides* samples, one *Macrotomia* sample also rendered *Arnebia* non-monophyletic in the remaining trees (i.e., three ITS analyses, and total evidence analysis). The remaining genera of Lithospermeae were only non-monophyletic in the plastid data analyses (unresolved) (Table 3).

DISCUSSION

To date, the current study is the first to encompass not only the widest taxon coverage in the tribe Lithospermeae, but also the Eastern Asian *Onosma*+*Maharanga*+ *Cystostemon* clade. The results clearly show that the phylogenetic relationships within the clade are clearly dependent on changes in taxon, gene and outgroup sampling (Table 3, Fig. 2). For instance, the "total evidence with 350 taxa" analysis yielded a topology, (((Eastern Asian *Onosma*+*Maharanga*) *Onosma* s.s.) *Cystostemon*), similar to the molecular dating analysis of total evidence tree of Chacón & al. (2019). However, this topology is rejected by the AU test (p-value = -0.00133) (Table 4, Fig. 2). On the other hand, the results of AU test overwhelmingly supported the (((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) *Onosma* s.s.) topology (p-value = 0.999) which was recovered from the distant and closely related outgroup analyses, and the "total evidence tree with 345 taxa" analysis (Table 4, Fig. 2), a topology never reported before.

Apart from the molecular data here presented, except for the corolla morphology, *Cystostemon* shares great

Table 4. Topology test results of two phylogenetic relationships (topologies) of *Onosma*, *Maharanga* and *Cystostemon* (p-value < 0.05 indicates statistical rejection). AU: approximately unbiased.

Topology	logL	AU test p-values	Significantly worse
(((Eastern Asian *Onosma*+*Maharanga*) *Onosma* s.s.) *Cystostemon*)	-39593.0028	-0.00133	YES
(((Eastern Asian *Onosma*+*Maharanga*) *Cystostemon*) *Onosma* s.s.)	-39400.53477	0.999	NO

morphological similarities with *Onosma* and *Maharanga* (Cohen 2014). However, whether genus *Onosma* is not monophyletic, or the genus should be extended to include *Maharanga* and maybe *Cystostemon,* or the "*O. rostellata* and Sino-Indian *Onosma* +*Maharanga*" constitutes a different lineage are questions that remain unanswered (Cohen 2014; Chacón & al. 2019; Nasrollahi & al. 2019). The main obstacle to make a taxonomic decision is that the taxon sampling of these clades is still very limited in each of the above-mentioned studies and the present one. Any further phylogenetic analyses for taxonomic rearrangement should include more Eastern Asian *Onosma, Maharanga* and *Cystostemon* samples and perhaps more genetic regions. For instance, the genus *Maharanga* comprises around 10 species and *Cystostemon* includes 16; yet sequences from only two and four species, respectively, are currently available in GenBank. Similarly, for the Eastern Asian *Onosma*, except *O. hookerii* (seven individuals), only a few (one to three) sequences have been published. These limitations highlight the importance of taxon sampling in phylogenetic studies as have been indicated before (Zwickl & al. 2002; Heath & al. 2008).

Another unresolved question addressed here is whether the Eastern Asian *Onosma* species are monophyletic or not. On one hand, ITS analysis of Nasrollahi & al. (2019) and plastid/ITS analyses of Chacón & al. (2019) revealed a non-monophyletic Eastern Asian *Onosma* (*Maharanga* samples embedded in the clade); on the other hand, the ITS analysis of Cecchi & al. (2011) and molecular dating analysis of Chacón & al. (2019) with ITS+ *trn*L-F+*rps*16+*trn*S-*trn*G data yielded a monophyletic Eastern Asian *Onosma* clade. The results are equivocal in the present study (Table 3, Fig. 2). Eastern Asian *Onosma* may indeed include *Maharanga*, because the two genera share great morphological similarities other than corolla and anther morphology (Zhu & al. 1995); however, these results may be a consequence of sampling gaps, or the known problems associated to the ITS region, such as, concerted evolution, paralogy, gene duplication and incomplete lineage sorting (Álvarez & Wendel 2003). Although these issues have never been reported for Boraginaceae (Cecchi & al. 2011); it is still possible that ITS data problems may be responsible for these peculiar phylogenetic relationships. Similarly, as in the previous studies in tribe Lithospermeae (e.g., Cecchi & Selvi 2009; Weigend & al. 2009; Cecchi & al. 2014; Coppi & al. 2015), the results of this study have possibly indicated that the relatively low number of parsimony-informative characters of the plastid data matrix (24%) caused the phylogenetic uncertainties, and therefore, limited-plastid data may not be informative enough to solve the genus *Onosma* and tribe Lithospermeae phylogenetic relationships.This was already pointed out byNasrollahi & al. (2019) who suggested to use "fast evolving genes" for future phylogenetic studies of *Onosma*. Unfortunately the results of the present study hint that even these types of genes may not be enough to solve the phylogenetic inconsistency, due to the reason of the effect of taxon sampling on the group phylogeny and possible complex evolutionary histories of both clades (Nasrollahi & al. 2019). Besides, morphological characters may also not be helpful to answer the phylogenetic questions within the genus (i.e., similar morphologies of *Onosma*, *Maharanga* and *Cystostemon)*, even when they are supported by molecular data. Thus, both the data and taxon sampling (i.e., both the ingroup and outgroup) hold great importance in *Onosma* phylogenetic analyses.

While Nasrollahi & al. (2019) reported a monophyletic position for *Onosma hookeri* and to date only one *O. sinicum* Diels sample was included in the previous studies (e.g., Chacón & al. 2019), the analyses of the current study showed that neither *O. sinicum* nor *O. hookeri* are monophyletic (Fig. 1). Indeed, misidentifications are common in Boraginales (Dr. Aslı Doğru Koca, pers. comm.) and the amount of wrong identifications in voucher specimens is reported to be very high in public molecular repositories (Nilsson & al. 2006; Wu & al. 2021). In the current study, I did not filter GenBank sequences by their reliable IDs or deposited voucher numbers. Indeed, this could be one of the reasons for the non-monophyly of these taxa. For example, a quick survey has shown that among the 746 ITS sequences included in the current study, only 489 of them (~66%) have their voucher numbers included in GenBank (results not shown). Although a voucher specimen will not guarantee a correct identification it provides scientific reliability and possibility of verification in the future (Wu & al. 2019). Therefore, future studies should be mindful about possible wrong identifications while adding GenBank sequences in their datasets.

The monophyly of most outgroup genera was well supported (Table 3). However, future studies aiming to eva-

luate the phylogenetic relationships within tribe Lithospermeae should be aware of including a more thorough sampling, particularly from genera *Buglossoides*, *Arnebia*, *Halacsya*, *Macrotomia, Stenosolenium* and *Aegonychon.*

The causes for phylogenetic incongruency reported for many angiosperm clades such as, Fabales (Aygoren Uluer & al. 2020a), Saxifragales (Jian & al. 2008), tribe Lithospermeae (Cohen 2011) and tribe Mirbelieae (Barret & al. 2021), can be explained by several non-exclusive reasons, such as, rapid radiation of the clade (e.g., Cohen 2011), inadequate data (Zeng & al. 2017), long branch attraction (LBA) (e.g., Qiu & al. 2001), conflicting gene trees (e.g., Yang & al. 2013), effect of outgroup sampling (e.g., Smith 1994; Aygoren Uluer & al. 2020b), reticulate events (e.g., hybridization, introgression and horizontal gene transfer, incomplete lineage sorting) (e.g., Solís-Lemus & Ané 2016) and/or stochastic noise.Further studies with a comprehensive ingroup and in outgroup sampling may answer possible causes for the phylogenetic problems among *Onosma*, *Cystostemon* and *Maharanga.* Therefore, future studies should increase the amount of data, if not the use of whole genome sequencing, the use of wide-genome RAD-seq data. Certainly, an increased taxon sampling of Turkish *Onosma* species, Eastern Asian *Onosma* species, *Cystostemon* and *Maharanga* should be a priority in future molecular studies.

REFERENCES

Álvarez I. & Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

Aygoren Uluer D., Hawkins J.A. & Forest F. 2020a. Interfamilial relationships in order Fabales: new insights from the nuclear regions *sqd* 1 and 26S rDNA. *Plant Systematics and Evolution* 306: 1–14.

Aygoren Uluer D., Forest F. & Hawkins J.A. 2020b. Supermatrix analyses and molecular clock rooting of Fabales: Exploring the effects of outgroup choice and long branch attraction on topology. *Botany* 98: 231–247.

Barrett R.L., Clugston J.A., Cook L.G., Crisp M.D., Jobson P.C., Lepschi B.J., Renner M.A. & Weston P.H. 2021. Understanding Diversity and Systematics in Australian Fabaceae Tribe Mirbelieae. *Diversity* 13: 391.

Cecchi L. & Selvi F. 2009. Phylogenetic relationships of the monotypic genera *Halacsya* and *Paramoltkia* and the origins of serpentine adaptation in circum-mediterranean Lithospermeae (Boraginaceae): insights from ITS and *mat*K DNA sequences. *Taxon* 58: 700–714.

Cecchi L., Coppi A. & Selvi F. 2011. Evolutionary dynamics of serpentine adaptation in *Onosma* (Boraginaceae) as revealed by ITS sequence data. *Plant Systematics and Evolution* 297: 185–199.

Cecchi L., Coppi A., Hilger H.H. & Selvi F. 2014. Non-monophyly of *Buglossoides* (Boraginaceae: Lithospermeae): Phylogenetic and morphological evidence for the expansion of *Glandora* and reappraisal of *Aegonychon*. *Taxon* 63: 1065–1078.

Chacón J., Luebert F. & Weigend M. 2017. Biogeographic events are not correlated with diaspore dispersal modes in Boraginaceae. *Frontiers in Ecology and Evolution* 5: 26.

Chacón J., Luebert F., Selvi F., Cecchi L. & Weigend M. 2019. Phylogeny and historical biogeography of Lithospermeae (Boraginaceae): Disentangling the possible causes of Miocene diversifications. *Molecular Phylogenetics and Evolution* 14: 106626.

Cohen J.I. & Davis J.I. 2009. Nomenclatural changes in *Lithospermum* (Boraginaceae) and related taxa following a reassessment of phylogenetic relationships. *Brittonia* 61: 101–111.

Cohen J.I. 2011. A phylogenetic analysis of morphological and molecular characters of *Lithospermum* L. (Boraginaceae) and related taxa: evolutionary relationships and character evolution. *Cladistics* 27: 559–580.

Cohen J.I. 2014. A phylogenetic analysis of morphological and molecular characters of Boraginaceae: evolutionary relationships, taxonomy, and patterns of character evolution. *Cladistics* 30: 139–169.

Coppi A., Cecchi L., Nocentini D. & Selvi F. 2015. *Arnebia purpurea*: a new member of formerly monotypic genus *Huynhia* (Boraginaceae-Lithospermeae). *Phytotaxa* 204: 123–136.

Darriba D., Taboada G.L., Doallo R. & Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Guindon S. & Gascuel O. 2003 A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52: 696–704.

Heath T.A., Hedtke S.M. & Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Jian S., Soltis P.S., Gitzendanner M.A., Moore M.J., Li R., Hendry T.A., Qiu Y.L., Dhingra A., Bell C.D. & Soltis D.E. 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57: 38–57.

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

Koyuncu O., Yaylacı Ö.K., Özgişi K., Sezer O. & Öztürk D. 2013. A new *Onosma* (Boraginaceae) species from central Anatolia, Turkey. *Plant Systematics and Evolution* 299: 1839–1847.

Letunic I. & Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44: W242–W245.

Mehrabian A.R., Sheidai M., Noormohammadi Z., Asrei Y. & Mozafarian V. 2011. Inter-simple sequence repeats (ISSR) and morphological diversity in *Onosma* L. (Boraginaceae) species in Iran. *African Journal of Biotechnology* 10: 10831–10838.

Miller M.A., Pfeiffer W. & Schwartz T. 2011. The CIPRES science gateway: a community resource for phylogenetic analyses. *In*: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, Louisiana.

Nasrollahi F., Kazempour-Osaloo S., Saadati N., Mozaffarian V. & Zare-Maivan H. 2019. Molecular phylogeny and divergence times of *Onosma* (Boraginaceae ss) based on nrDNA ITS and plastid *rpl32-trnL* (UAG) and *trnH–psbA* sequences. *Nordic Journal of Botany* 37: e02060.

Nilsson R.H., Ryberg M., Kristiansson E., Abarenkov K., Larsson K.H. & Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PloS ONE* 1: e59.

Qiu Y.L., Lee J., Whitlock B.A., Bernasconi-Quadroni F. & Dombrovska O. 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Molecular Biology and Evolution* 18: 1745–1753.

Rambaut A. & Drummond A.J. 2007. Tracer v1. 4. Website:https://beast.community/tracer [accessed: 23 Mar. 2022].

Ronquist F. & Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Selvi F., Cecchi L., Hilger H.H. & Coppi A. 2017. A reappraisal of the genus *Megacaryon* (Boraginaceae, Lithospermeae) based on molecular, morphological, and karyological evidence. *Systematics and Biodiversity* 15: 552–563.

Shimodaira H. & Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114.

Smith A.B. 1994. Rooting molecular trees: problems and strategies. *Biological Journal of the Linnean Society* 51: 279–292.

Solís-Lemus C. & Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* 12: e1005896.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313

Thomas D.C., Weigend M. & Hilger H.H. 2008. Phylogeny and systematics of *Lithodora* (Boraginaceae—Lithospermeae) and its affinities to the monotypic genera *Mairetis*, *Halacsya* and *Paramoltkia* based on ITS1 and *trn*LUAA-sequence data and morphology. *Taxon* 57: 79–97.

Trifinopoulos J., Nguyen L.T., von Haeseler A. & Minh B.Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 44: W232–W235.

Weigend M., Gottschling M., Selvi F. & Hilger H.H. 2009. Marbleseeds are gromwells–Systematics and evolution of *Lithospermum* and allies (Boraginaceae tribe Lithospermeae) based on molecular and morphological data. *Molecular Phylogenetics and Evolution* 52: 755–768.

Wu H.Y., Chan K.T., But G.W.C. & Shaw P.C. 2021. Assessing the reliability of medicinal *Dendrobium* sequences in GenBank for botanical species identification. *Scientific Reports* 11: 1–9.

Yang H.M., Zhang Y.X., Yang J.B. & Li D.Z. 2013. The monophyly of *Chimonocalamus* and conflicting gene trees in Arundinarieae (Poaceae: Bambusoideae) inferred from four plastid and two nuclear markers. *Molecular Phylogenetics and Evolution* 68: 340–356.

Zeng L., Zhang N., Zhang Q., Endress P.K., Huang J. & Ma H. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist* 214: 1338–1354.

Zhu G.L., Harald R. & Rudolf K. 1995. Boraginaceae. *In*: Wu Z. Y. & Raven P. H. (eds), *Flora of China* 16: 329–427. Science Press, Beijing, & Missouri Botanical Garden Press, St. Louis.

Zwickl D.J. & Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.

Appendix 1. Alignment of ITS (746 taxa). **Appendix 2.** Alignment of total evidence analysis with 350 taxa. **Appendix 3.** ITS (323 taxa) RAxML result. **Appendix 4.** ITS (555 taxa) RAxML result. **Appendix 5.** ITS (746 taxa) RAxML result. **Appendix 6.** Plastid (339 taxa) RAxML result. **Appendix 7.** Total evidence (350 taxa) RAxML result. **Appendix 8.** Total evidence (345 taxa) RAxML result. **Appendix 9.** Total evidence with distant outgroup (347 taxa) RAxML result. **Appendix 10.** Total evidence with closely related outgroup (347 taxa)RAxML result. All appendices files are available in the Mendeley data repository at https://data.mendeley.com/datasets/f9ghrnzghx/1