New insights into the deep divergences in *Ephedra* (Gnetales) using molecular data

Olle Thureborn

Department of Ecology, Environment and Plant Sciences
Master degree 60 HE credits
Systematic Botany
Biology
Spring term 2014
Supervisor: Catarina Rydin

Stockholms universitet

Abstract

Deep divergences in *Ephedra* have been addressed in several studies in the past and have proven difficult to resolve. One major reason for this is probably the low amount of sequence divergence within *Ephedra*, in combination with distant relationships to the closest living relatives of the genus. I approached these problems using an increased sampling of taxa and information from nine molecular markers, including one (*matK*) that previously has been sequenced only for a few taxa in the genus, and one that has never been used in previous studies of phylogeny in *Ephedra,* the 5'external transcribed spacer (5'ETS). First I tested the hypothesis of a sister relationship between *E. foeminea* and all other *Ephedra* species using a data set including 68 *Ephedra* specimens and outgroups from all other major clades of vascular plants. The results from Bayesian analyses show all species except *E. foeminea* to be monophyletic. The monophyly of *E. foeminea* was not confirmed in Bayesian analyses but was supported in the MP analysis. Subsequently I did further analyses, using a data set that contained 135 specimens of *Ephedra*, and *E. foeminea* as outgroup based on results in the former analysis. The present study lends support for the hypothesis of a sister-relationship between *Ephedra foeminea* and all other species of the genus, and it reveals new insights of the deep divergences of *Ephedra* both in terms of credibility of results (posterior probability values) and resolution. The sequence data from the ETS region is shown to be very useful for the understanding of evolution and relationships within *Ephedra*. A first assessment of the genetic structure and evolution of ETS in *Ephedra* is presented and discussed, as are primer sequences that may be used in future studies of phylogeny and species delimitation.

Keywords: *Ephedra*, external transcribed spacer, Gnetales, phylogeny.

Cover: *Ephedra foeminea* in Omiš, Croatia, July 2013. Photo: Olle Thureborn.

Contents

Introduction 4

Gnetales 4

Ephedra 4

Phylogeny estimation and sequence divergence in *Ephedra* 5

Aim of the present study 5

Material & Methods 6

Field work in Croatia 6

Taxonomic sampling 6

Molecular markers 7

Extraction of DNA 8

DNA amplification 9

Primer design 9

De novo construction of amplification primers for IGS 10

PCR 10

Nested PCR 10

DNA electrophoresis 11

DNA purification 11

Sequencing 11

Sequence assembly and alignment 11

Examining the utility of ETS primers within *Ephedra* 11

Comparison of ETS and ITS 11

Assessments of the ETS structure 12

Phylogenetic Analyses 12

Data partitions 12

Model selection 12

Bayesian analyses 14

Convergence diagnostics 14

Parsimony analysis 14

Results 14

Testing the sister relationship between *Ephedra foeminea* and remaining species of the genus 14

Assessing further relationships within *Ephedra* 15

Single gene analyses; nuclear and chloroplast analyses 15

Assessments of the ETS structure 16

Amplification and sequencing of the ETS 16
Phylogenetic utility of ETS in comparison to ITS 16
Discussion 20
Ephedra foeminea and the root of the *Ephedra* phylogeny 20
Relationships within remaining *Ephedra* 20
Clade A (*E. aphylla-E. major-E. fragilis- E. altissima*) 21
Clade B (*E. fragilis* and *E. altissima*) and clade C (*E. aphylla*, *E. major*, *E. fragilis*) 21
Clade D (*E. alata*) 21
Clade E (*Ephedra milleri*) 22
News within clade F: The core *Ephedra* 22
The polyphyly of *E. major* 22
ETS- a new phylogenetic marker for *Ephedra* 23
Assessments of the ETS structure 23
Conflicting information in ITS and ETS 24
Concluding remarks and outlook 25
Acknowledgments 25
References 26
Appendixes 31
Appendix 1 Collections from field trip in Dalmatia, Croatia 31
Appendix 2 Included Taxa 32

Introduction

Gnetales

The Gnetales are a small group of seed plants, which comprise three extant families: the monogeneric Ephedraceae and Gnetaceae, and the monotypic Welwitschiaceae.

The relationship of the Gnetales to other seed plants has been much debated. Doyle and Donoghue's (1986) "anthophyte hypothesis" based on morphology and fossil evidence resolved the Gnetales as sister to angiosperms, but the hypothesis has not gained support from studies based on molecular data. Molecular studies have mostly placed the Gnetales nested among conifers, and three topologies have been proposed: the Gnetifer hypothesis (Gnetales sister to Pinaceae and cupressophytes (Rydin et al. 2002; Rydin and Korall 2009), the Gnepine hypothesis (Gnetales sister to Pinaceae; e.g., Chaw et al. 2000 and the Gnecup hypothesis (Gnetales sister to cupressophytes; Braukmann et al. 2009; Mathews 2009 and references within). The Gnetales have also been proposed to be the sister to all other extant seed plants (Rydin et al. 2002; Braukmann et al. 2009; Mathews 2009 and references within), and as sister to all other gymnosperms (Schmidt and Schneider-Poetsch 2002).

The three genera of the Gnetales are very different from each other both in terms of morphology and ecology (Price 1996). *Gnetum* comprises c. 30 species, mostly vines with opposite simple leaves, native to tropical rain forests in Africa, Asia, and South and Central America (Simpson 2010). *Welwitschia mirabilis* is an unusual and long lived rosette-like plant endemic to the deserts of Namibia and Angola in southwestern Africa (Simpson 2010). *Ephedra* with its c. 60 species is distributed in arid and semi-arid regions in northern Africa, Eurasia and North and South-America (Kubitzki 1990) and is found at elevations ranging from sea level up to 5000m (e.g., in Himalaya and Andes) (Price 1996).

The distinct differences between the three genera were by Arber and Parkin (1908) seen as an indication of a long evolutionary history and they argued that the extant species are the remains of a former much greater diversity. That suggestion has more recently gained support in the form of fossil findings, for example the extensive record of fossil gnetalean pollen grains from the Early Cretaceous (Crane 1996) as well as Early Cretaceous macrofossils found in different areas of the world (Rydin et al. 2006). However, although different from each other, several features unite the members of the Gnetales, e.g., vessels with porose perforation plates, mostly opposite (decussate) phyllotaxis, single terminal unitegmic ovules, integument with much-elongated beaks (micropylar tube), and ovules surrounded by seed envelope(s) (Rydin and Friis 2010; Simpson 2010), and the first proposals of monophyly of the Gnetales were based on morphology (Arber and Parkin 1908). Later, cladistic analyses by Crane (1985) and Chase et al. (1993) provided evidence for monophyly of the Gnetales; there is strong molecular support for Gnetales being monophyletic with *Ephedra* sister to *Gnetum* and *Welwitschia* (Rydin et al. 2006).

Ephedra

Ephedra is with its approximately 60 species the largest of the three genera, which constitutes the Gnetales. *Ephedra* species are xeromorphic, usually dioecious, often erect or sprawling shrubs, but may also grow as vine-like shrubs or small trees. *Ephedra* species are easily distinguished by their photosynthetic striate stems and much reduced scale-like leaves. Due to the reduced leaves, photosynthesis takes mostly place in the striate stems of annual shoots (Price 1996). The leaves are arranged oppositely or in whorls. The bract of the female cones are arranged in pairs (2-8) or whorls. At seed maturity the bracts of the female cones can be fleshy and colorful or dry and sometimes also winged (Kubitzki 1990; Price 1996), which may be adaptations to bird and wind dispersal, respectively (Price 1996). Of the three genera, *Ephedra* has the smallest ovulate cones with 1-3 ovules per cone (Price 1996). *Ephedra* has been and is still important in the traditional and modern

medicine; some species produce and accumulate high levels of the alkaloid *ephedrine*. Some species are also used as forage plants, grazed by goats, sheep and camels (Freitag and Maier-Stolte 1994) and for bio-control (Al-Qarawi et al. 2011).

While the morphological differences between the genera in Gnetales are substantial, the opposite can be said about the differences between species within *Ephedra*; they are very similar in overall morphology. Studies based on morphological characters (Stapf 1889; Steeves and Barghoorn 1959; Mussayev 1978; Freitag and Maier-Stolte 1994) have resulted in different classifications of the genus. One reason for the difficulties in finding morphological support for phylogenetic relationships within *Ephedra* may be that variation can be pronounced also within species, not only among species (Huang et al. 2005), e.g., number of seeds per ovulate cone and two vs. three leaves per node. It is also obvious that some easily recognizable differences between species, such as fleshy/dry cone bracts, are homoplasious and not clade specific (Ickert-Bond and Wojciechowski 2004).

Ephedra-like plants are known since the Early Cretaceous (Rydin et al. 2006), and despite that the similarities between these fossils and extant *Ephedra* are generally strong, the fossils seem to belong to extinct sister lineages rather than being nested within the now living *Ephedra* clade (Rydin et al. 2010). This is in line with estimates of divergence times of clades based on molecular data (Huang and Price 2003; Ickert-Bond et al. 2009), which indicate that extant *Ephedra* has a recent origin and began to diversify during the Oligocence (c. 30 MA). Although a recent study indicate that the crown group may be somewhat older (Norbäck Ivarsson 2014), *Ephedra* has probably undergone one radiation in the Early Cretaceous, the diversity of which went almost entirely extinct toward the latter part of the Cretaceous, and a second radiation in the Paleogene. This may explain the low molecular and morphological divergence among extant species (Rydin et al. 2010).

Phylogeny estimation and sequence divergence in *Ephedra*

Several attempts have been made to resolve the phylogeny of *Ephedra* using DNA sequence data. Rydin et al. (2004) used 5 regions (18S, 26S, ITS, *rbcL*, *trnL*, *rps4),* Ickert-Bond and Wojciechowski (2004) used *rps4* and ITS1 and Huang et al. (2005) used *rbcL*, *matK*, ITS1. These studies showed (among other things) strong support for geographic groupings. They were incongruent with previous hypotheses of the evolutionary relationships within *Ephedra* based on morphology, for example the traditional classification by Stapf (1889), which was primarily based on characters of the ovulate cone bracts. However, these early molecular studies were all subject of restricted taxon sampling (especially among Old World species). Further, the low sequence divergence in utilized molecular markers has hampered phylogenetic reconstruction (Rydin and Korall 2009), and many nodes were poorly supported statistically. Rydin and Korall (2009) made a further attempt to resolve the *Ephedra* phylogeny with a larger taxon sampling and information from DNA sequences from seven regions. In short, their study resolved *Ephedra foeminea* as sister to all other species of *Ephedra*, but this result was weakly supported. Further, *E. foeminea* was sister to an unresolved trichotomy. Although the results provided more information on the phylogeny of *Ephedra,* several questions remained unanswered or uncertain, for example, relationships among American species and deep divergences in *Ephedra*.

Aim of the present study

Accurate estimates of phylogenies are important bases for other fields of research, such as studies of community ecology, genomics, biogeography and identification and spread of emerging pathogens. Regarding *Ephedra*, a good knowledge about phylogeny is important for the interpretation of morphological traits in an evolutionary perspective, which in turn is critical for the placement of fossils within the ephedroid lineage. So a better understanding of the extant *Ephedra* species is of high importance for further studies concerning the biogeography, ecology, timing of evolutionary events and trait evolution, which may subsequently also lead to new insights of the relationships and evolution of seed plants. The aim of my master project is to further study and resolve the

relationships within *Ephedra*, using an increased taxon sampling from the Mediterranean species complex and new phylogenetic markers. My approach is to first test the hypothesis of a sister relationship between *E. foeminea* and all other *Ephedra* species, and subsequently do further analyses on the whole genus, with particular emphasis on deep divergences and the Mediterranean taxa.

Material & Methods

I used three main approaches in my attempt to resolve the deep divergences in *Ephedra*. First I increased the taxon sampling of the Mediterranean species *E. foeminea*, *E. alata*, *E. fragilis*, *E. aphylla*, *E. foliata*, *E. ciliata* and *E. major* as much as possible, and tried to cover the geographic distribution of each species as well as possible. Second, I utilized a phylogenetic marker not previously used in phylogenetic studies of *Ephedra,* the 5'external transcribed spacer (5'ETS). New ITS and *matK* sequences were also produced for the present study. Finally, I added the new data set to that of Rydin and Korall (2009) for combined analyses.

Field work in Croatia

I conducted a field study in the Dalmatia region of Croatia between 1st and 10th of July 2013. The aim was to collect new *Ephedra* material for the present study (see appendix 1 for voucher information). The field trip was conducted in collaboration with another master student, whose main aim was to collect fertile structures of *Ephedra foeminea*; hence the timing of the trip was of great importance. An appropriate date for collecting fertile *Ephedra* specimens was based upon voucher information from fertile *Ephedra* specimens.

Before the field trip, investigations regarding regulations and permits for collection of plant material within the EU (Croatia became members of the EU 1st July 2013) were made. Since the introduction of UN Convention on Biological Diversity many new regulations have come into place, and of course it is in everyone's interest to comply with those rules. Therefore in March 2013 an inquiry regarding rules and regulations applicable in this specific case was sent to the Swedish Ministry of the Environment. From them we were recommended the Convention on Biological Diversity website (www.cbd.int), from which we located the relevant authorities in charge of our question in respective countries. When the appropriate contacts had been made inquiries were sent. An answer from the Croatian Ministry of Environmental and Nature Protection, Republic Croatia was received early April and an official request for collection of plant material was sent back to them. At the end of April an official permit to collect wild taxa in Croatia and export the same material for scientific/non-commercial research was received. A field trip to Spain was also planned, but unfortunately we did not hear anything from the authorities in Spain and therefore the field trip to Spain was cancelled.

Taxonomic sampling

For this study a total of 135 *Ephedra* specimens representing 45 *Ephedra* species were selected in order to cover a large proportion of the now living diversity of the genus (appendix 2). Of the 135 specimens, 19 specimens are newly extracted for the present study. When selecting ingroup terminals the focus was on the Mediterranean species as defined in Rydin and Korall (2009) and efforts were made to include several individuals of each species to cover their geographical distribution and test monophyly of species. This was not always possible due to either lack of material or existing material being too "old" to yield sufficient amounts of DNA. In order to test the hypothesis of a sister relationship between *E. foeminea* and all other *Ephedra* species, I designed a dataset including 68 *Ephedra* specimens with representatives from all major subgroups as defined in Rydin and Korall (2009), i.e., species from the Mediterranean area, the mainly Asia-clade and the New World-clade. Care was taken to include specimens with DNA sequence data from all nine gene regions. The outgroup selection was the same as in Rydin and Korall (2009), and consisted of taxa

species from the two other genera of the *Gnetales* (*Welwitschia* and *Gnetum*), as well as all other major clades of vascular plants (lycopods, ferns, cycads, *Gingko*, conifers and angiosperms) (see Rydin and Korall 2009).

A subsequent analysis, which addressed major relationships within *Ephedra*, included 135 specimens of *Ephedra*, and used *E. foeminea* as outgroup based on results in the former analysis.

Molecular markers

Because of the low amount of information in molecular markers utilized so far in *Ephedra* studies, I searched the literature for a new, potentially useful phylogenetic marker. The choice fell on the external transcribed spacer (ETS) of the 18S-26S nrDNA repeat. There are in fact two ETS regions present in the intergenic spacer (IGS), the 3'ETS and the 5'ETS (fig. 1). I used the 5'ETS region, which is the most commonly used part in phylogenetic studies. If not stated otherwise, the utilized 5'ETS region will be referred to as ETS throughout this paper. Previous studies (Baldwin and Markos 1998, Linder et al. 2000) of angiosperms have shown that ETS has great potential for intrageneric phylogenetic reconstruction, and could be used together with, or instead of the widely used internal transcribed spacer (ITS). However, there is one major obstacle to sequencing the ETS comparing to the ITS; the absence of a highly conserved region at the 5' end of the spacer suitable for location of universal primers. The 5' end of ETS is flanked by the non-transcribed spacer (NTS), a region evolving at a high rate and thus not a suitable position of universal primers (Baldwin and Markos 1998 and references within). As the ETS had never been used in any study of *Ephedra*, no suitable primers were available in literature. Therefore it would be necessary to amplify the complete intergenic spacer (IGS) and sequence a big part of it for at least one *Ephedra* species, on which production of internal primers (i.e., ETS primers) could be based. Due to the low sequence variation in *Ephedra* in general, it was reasonable to assume that conserved regions would be present, within which primers could be placed.

Further, one of the chloroplast markers utilized in the present study, *matK*, has not been used in previous studies addressing deep divergences in *Ephedra*. A recent study used it for addressing evolutionary and ecological questions on American taxa (Loera et al. 2012), but otherwise, no sequences are available on GenBank. In total, DNA sequence data were retrieved from nine regions for the present study, four nuclear ribosomal DNA regions (the protein coding region 18S and 26S, the non-coding internal transcribed spacer (ITS), and the 5' external transcribed spacer (5'ETS)), and five chloroplast regions (the protein-coding genes *matK, rbcL* and *rps4*, the trn^{SUGA}-$trnfM^{CAU}$ intergenic spacer and the *rpl16* intron). A total of 91 sequences were newly produced for the present study; 50 *matK*, 18 ITS and 23 ETS sequences. These new sequences were obtained from samples extracted in earlier studies (Rydin and Korall 2009; Rydin et al. 2010), or from newly collected, silica dried samples, or from herbarium material. The remaining sequences used in this study were obtained from earlier studies (Ickert-Bond and Wojciechowski 2004; Rydin et al. 2004; Huang et al. 2005; Rydin and Korall 2009; Rydin et al. 2010; Loera et al. 2012). A full list of the specimens used in this study along with their respective voucher and sequence information are given in appendix 2. Sequence sampling was done on specimen-level within *Ephedra*, i.e., all gene regions of a terminal are derived from the same plant. For some terminals in the New World and mainly Asia-clades, and for *Gnetum* and *Welwitschia*, however, the *matK* gene was sampled at the species-level, i.e., for these taxa, the genes of a given terminal are sampled from the same species but not from the same specimen. For outgroup taxa outside of the Gnetales, sampling was done at the generic level (see also appendix 2).

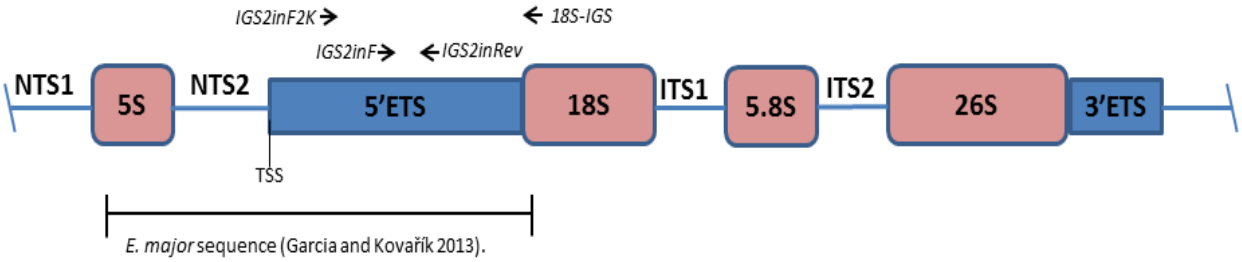

Fig. 1. Schematic representation of a single copy of the nuclear ribosomal DNA modified from (Garcia and Kovařík 2013, Linder et al. 2010, Baldwin and Markos 1998 and Wendland et al 1999). Relative sizes of DNA regions are arbitrary with non-transcribed spacers (NTS1 and NTS2) and internal transcribed spacers (ITS1 and ITS2) represented by lines, the external transcribed spacers (5'ETS and 3'ETS) represented by narrow blue rectangles, and gene coding regions represented by wide red rectangles. The putative transcription start site (TSS) is labeled at the border of the NTS2 and the 5'ETS. The line segment below the schematic drawing represents the 3358 bp long DNA sequence of an *Ephedra major* specimen produced by (Garcia and Kovařík 2013). Orientation and approximate position of primers used for the present study are indicated by arrows. Primers IGS2inF2K, IGS2inF and IGS2inRev were designed for the present study, whereas 18S-IGS was previously designed by Baldwin and Markos (1998). The non-transcribed spacers (NTS1 and NTS2) separate the separately transcribed 5S coding region from the co-transcribed 18S-5.8S-26S block. The gene coding 18S, 5.8S and 26S regions are separated by ITS1 and ITS2.

Extraction of DNA

Total genomic DNA was extracted using a slightly modified version of the CTAB method as described in Doyle and Doyle (1987), and purified with QIAquick® PCR-kit (Qiagen, Solna, Sweden/Hilden, Germany) following the instructions provided by the manufacturer. About 0,02 g dried plant material from each specimen was added to 2 ml eppendorf tubes with 2 steel beads and placed in the tissue lyser for 2 minutes at 5000 rpm for disruption of the material. The crushed plant material was then centrifuged for 2 minutes at 13000 rpm before adding 700 µl of CTAB-mixture to each tube (20 µl of 2-mercaproetanol was added to 10 ml of CTAB holding a temperature of 60°C). Care was taken to ensure that no plant material was stuck to the walls of the tubes. The tubes were then incubated at 60°C (heat helps to disrupt the cells) for 30 minutes. The tubes were turned upside down a couple of times every 10 minutes for mixing purposes. During incubation the content of the cells is released in the solution.

To wash away unwanted material (for example proteins) the contents in the 2 ml tubes were transferred to 1.5 ml eppendorf tubes containing 700 µl of Sevag solution (Chloroform/Isoamyl alcohol 24:1) and put on a shake board for 30 minutes. Next, the tubes were spun at 15 minutes at 10 000 rmp in a centrifuge. This causes the solution to separate into two layers, one bottom layer with unwanted cellular material and chloroform, and the upper layer with DNA solution. The DNA layer was removed and placed in new 1.5 ml eppendorf tubes with much care taken not to bring any of the unwanted material. All but 100 µl of the DNA solution was left in the eppendorf tube to be prepared as doublets. 100 µl of the DNA solution was added to spin-columns, quickspin tubes together with 500 µl of PB and centrifuged at 13 000 rpm for 1 minute. This step causes the DNA to bind to the filter in the spin-column. The used PB was removed from the tube.

Next, PE wash buffer was added to each tube and the tubes were centrifuged at 13 000 rpm for one minute. After removing the used PE the tubes were centrifuged dry for 1 minute at 13 000 rpm. The spin columns were then moved to new clean 1.5 ml eppendorf tubes with 50 µl of EB added to each the spin columns. The DNA the spin columns were then centrifuged at 13 000 for 2 minutes, and then the DNA solution left in each eppendorf tubes was ready for PCR.

In order to precipitate the DNA left in the doublet tubes, each doublet was prepared with isopropanol (3 parts DNA solution 2 parts isopropanol) and spun for 15 minutes at 10 000 rpm. The

solution was then removed, leaving only the precipitated DNA pellet. Next, 750 µl of washing buffer was added to each tube and incubated in room temperature for 20 minutes before the tubes were spun for 15 minutes at 10 000 rpm. The washing buffer was then removed from each tube with only the DNA pellets remaining. The pellets were then left in room temperature to dry before 50-150 µl of EB buffer (10 µM, TrisHcl, pH 8.5) was added to each tube. The tubes were put in the fridge until the pellets had dissolved. When dissolved, the doublets were stored in the freezer until needed.

DNA amplification

The primers used for amplification and sequencing were either newly designed for this study, or synthesized following previous studies. All primers used for PCR amplification and sequencing are presented in table 1.

Table 1. Primers for PCR and sequencing.

Marker	Primer name	Primer sequence (5'-3')	Source
ETS	IGS2inF2K	CGCGACTGTCCTTCGGAGG	This study
ETS	18S-IGS	GAGACAAGCATATGACTACTGGCAGGATCAACCAG	Balwin and Markos 1998
ETS	IGS2inF	GAGGACGCGAGCTTCTCTCG	This study
ETS	IGS2inRev	GATCCGAGATATTTCCCGTGCC	This study
ITS	ITS-18SF	GAACCTTATCGTTTAGAGGAAGG	Rydin et al. 2004
ITS	ITS-26SR	CCGCCAGATTTTCACGCTGGGC	Rydin et al. 2004
ITS	5.8SR	GCGACGTAGGAAAGGAAATAG	Quijada et al. 1997
ITS	820F	CCTACGTCGCTGGGACGTTAAACC	pers. comm. Catarina Rydin
matK	trnK-Ep2	TTCATGAGTCAGGAGAAC	Huang et al. 2005
matK	matK-Ep3R	GTATATACTTCACACGAT	Huang et al. 2005
matK	matKinF	CCCTCTTCGATTCATTCAGAGCTG	This study
matK	matKinRev	GACCATAAGACAATGATTTTTCATG	This study
IGS	26S-IGS	GGATTGTTCACCCACCAATAGGGAACGTGAGCTG	Balwin and Markos 1998
IGS	18S-IGS	GAGACAAGCATATGACTACTGGCAGGATCAACCAG	Balwin and Markos 1998

Primer design

Internal primers for PCR and sequencing of *matK* were based on the consensus of several *Ephedra* sequences. Primers for the ITS region were already available (Rydin et al. 2004, and C.R. pers. comm.) The 5' ETS region had never before been used in phylogenetic studies of *Ephedra* and primers had to be *de novo* designed (see below). When designing the primers care was taken to follow the guidelines from Premier Biosoft (http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html) concerning for example primer melting temperature (T_m), primer annealing temperature (T_a), primer specificity (avoiding homology), primer dimer (primers binding to each other), primer hairpins (primer binding to itself), primer pair T_M match calculation (primers T_M should not differ more than 5°C). Designing of primers, simulation and testing of PCR were performed with the computer program Amplify 3.1, calculation of T_m and T_a were carried out using New England BioLabs TM calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator). Before synthesis, I tested newly designed primers against a suitable target sequence using the software Amplify 3.1 (Engels 2005). Synthesizing of the primer sequences was handled by Eurofins MWG operon, Germany.

De novo construction of amplification primers for IGS

Initial attempts to amplify the complete IGS were conducted using the primers 26S-IGS and 18S-IGS (Baldwin and Markos 1998) and Phusion High-Fidelity DNA polymerase, designed for amplifying long DNA sequences. The PCR protocol for the IGS was as follows: 50 µl reaction volumes including 10 µl 5x Phusion reaction buffer, 4 µl dNTP, 0.5 µl Phusion DNA polymeras (5U/µl), 2.5 µl of each primer (10 µM), 1 µl of DNA template and sterilized H_2O up to 50 µl. The PCR profile for the IGS was: initial denaturation at 98°C for 1 min followed by 35 cycles of 10 s: 98°C (denaturation); 30 s: 71°C (annealing); 1 min: 72°C (extension) and a final extension step of 7 min: 72°C.

The amplification generated bands from two of three tested specimens. Of the four sequencing reactions (2 primers for each specimen) only one (the 26S-IGS of OT10 *E. foeminea*) was successful (the produced sequence was c. 600 bp). Furthermore, the 5'ETS is located upstreams of the 18S gene (fig. 1) and the IGS is in general 4000 bp or more (Jorgensen and Cluster 1988). Therefore, the sequence produced from the 26S-IGS primer did not read long enough to reach into the 5'ETS. Neither could I be sure that the 18S-IGS primer had annealed at the proper site. However, while I was waiting for the first sequencing results, i.e., for the 96-well reaction plate to be filled up with satisfactory amplified PCR-products, and then for the results from Macrogen, a timely paper (Garcia and Kovarik 2013) was released. The paper is on genomic organization of rRNA genes in gymnosperms and the authors had among other things sequenced a large part of the IGS of *Ephedra major* (GenBank JX843794). The IGS in *Ephedra* is divided into IGS1 and IGS2, interrupted by the separately transcribed 5S gene (Garcia and Kovarik 2013). Their sequence read from the 5S gene to the 5' terminus of the 18S gene. As the IGS fragment was long (3358 bp) and contained highly repetitive regions, Garcia and Kovarik (2013) used both primer walking and subcloning strategies when sequencing the IGS. Considering my initial problems described above, and with this in mind, I decided it would be too costly and time-consuming to try to amplify and sequence the complete IGS and I decided to design PCR primers and internal primers for the ETS region using the GenBank accession JX843794 produced by Garcia and Kovarik (2013) as template.

PCR

All polymerase chain reactions (PCR) were carried out in an Eppendorf® Mastercycler® gradient (Bergman & Beving Instrument, Stockholm, Sweden). Unless otherwise stated, the PCR protocols were as follows: 50 µl reaction volumes including 5 µl reaction buffer, 5 µl TMACL, 4 µl dNTP, 0.5 µl Paq DNA polymeras (5U/µl), 0.5 (cpDNA) or 0.7 (nrDNA) µl of each primer (20 µM), 0.5 µl BSA 1%, 1-4 µl of DNA template and sterilized H_2O up to 50 µl.

The PCR profile for the chloroplast *matK* gene was: 97°C followed by 40 cycles of 10 s: 97°C (denaturation); 30 s: 47°C (annealing); 20 s with the addition of 4 s in each consecutive cycle: 72°C (extension) and a final extension step of 7 min: 72°C.

The PCR profile for the ITS was: initial denaturation at 97°C followed by 40 cycles of 10 s: 97°C (denaturation); 30 s: 55°C (annealing); 20 s with the addition of 4 s in each consecutive cycle: 72°C (extension) and a final extension step of 7 min: 72°C.

The PCR profile for the ETS was: initial denaturation at 97°C followed by 40 cycles of 10 s: 97°C (denaturation); 30 s: 58°C (annealing); 20 s with the addition of 4 s in each consecutive cycle: 72°C (extension) and a final extension step of 7 min: 72°C.

Nested PCR

In cases when PCR product yields were low, nested PCR was performed. The primary PCR-product was used as template and two new primers were used, one internal forward primer compatible with the original reverse primer, and one internal reverse primer compatible with the original forward primer. The number of cycles was reduced compared to the first PCR run. This approach was often

successful; however, the risk for contamination is high due to the high concentration of DNA in the template (the primary PCR-product). Extreme care was therefore taken when setting up the reaction. Further, because this is not the first study that addressed infrageneric relationships in *Ephedra*, it is typically easy to detect and assess potential cases of contamination.

DNA electrophoresis

To assess the PCR-products, a DNA ladder (DNA fragments of known size), negative control (control of contamination) and 3.5 µl PCR-product mixed with 1 µl blue dye was loaded into wells on a 1%-agarose gel where an electric field force the DNA to move, a method commonly known as gel electrophoresis. The method uses the fact that DNA has a net negative charge, which makes the DNA molecules migrate towards the anode. The speed, by which the molecules are migrating through the gel, is congruent with their size (smaller molecules migrate slower than larger ones because they experience less hindrance within the gel). In that way the molecules will be separated, which makes it possible to assess whether the PCR was successful or not. The voltage was removed after 30 min and the gel was examined in a UV-light cabinet by comparing the bands from the PCR-products with the bands from the ladder and negative control.

DNA purification

PCR-products were purified by vacuum filtration using Millipore 96-well plates, where each well contains a membrane which excludes small fragments (up to 137 bp) and retain DNA products at the surface. The PCR products were loaded into the plate and mixed with 100 µl distilled H_2O. The plate was put on top of the MultiScreen vacuum manifold and vacuum was turned on for a few minutes (until wells were dry). Next, 100 µl distilled H_2O was added to each well and the plate were moved to a shaking platform for 5 minutes at 300 rpm. Next, the plate was again put on to the manifold and vacuum was turned on for 5 minutes (or until dry). To elute the DNA from the membrane 35 µl distilled H_2O was added to the wells and they were shaken for 10 minutes at 300 rpm. The eluted DNA samples were then moved to new 100 µl tubes for storage in freezer.

Sequencing

Sequencing was prepared by adding 5 µl of the purified PCR-product samples to a MicroAmp® Optical 96-well reaction plate together with 5 µl primer (conc.: 5 µM). Automated sequencing was carried out by Macrogen Sequencing Service, Netherlands.

Sequence assembly and alignment

The raw sequence data obtained from Macrogen was assembled, examined and edited using the Staden package (Staden 1996). The alignment of the new sequences, together with sequences downloaded from GenBank, was conducted manually using the program Se-Al v2.0a11 Carbon. IGS sequences were initially aligned using the online application of MUSCLE (Edgar 2004) available at https://www.ebi.ac.uk/Tools/msa/muscle/, before further assessments were done manually.

Examining the utility of ETS primers within *Ephedra*

To determine whether the *IGS2inF1K* and *IGS2inF2K* primer designed for the ETS region had a general fit within *Ephedra* I attempted to amplify 62 specimens comprising specimens from all major groups as outlined in Rydin and Korall (2009), i.e., the Mediterranean species complex, core *Ephedra*, the New World clade and the mainly Asian clade. The primers were used in combination with the reverse primer *18S-IGS* primer (Baldwin and Markos 1998).

Comparison of ETS and ITS

In order to compare the phylogenetic information in the two spacers (ETS and ITS), I calculated the number of variable characters and informative characters for each spacer using PAUP v. 4.0b10 (Swofford 2003), for identical sets of taxa. Bayesian analyses of both ETS and ITS and the two regions

combined were carried out to examine potential topological conflicts, and as a measure of congruence of the two regions.

Assessments of the ETS structure

As a basis for sequence annotation of 5'ETS, I used the Prediction of PLANT Promoters application (Using RegSite Plant DB, Softberry Inc.) on www.softberry.com (Shamuradov et al. 2003) in order to identify any putative transcription start site (TSS) in the *Ephedra major* sequence (GenBank JX843794), and in the newly produced sequences. The start of the ETS is defined by the transcription start site (TSS) (Poczai and Hyönen 2010). In earlier studies of the organization of the IGS (NTS+ETS) the TSS has been identified by the presence of a so called TATA box (Linder et al. 2000; Grabiele et al. 2012; Galian et al. 2012). A TATA box is a short DNA sequence in promoter regions of eukaryotic organisms, which is located about 25 bp upstreams of the TSS (Campbell et al. 2008). Discovery of this TATA box (which is one of the few conserved regions within the IGS in plants, (Poczai and Hyvönen 2010) in my sequences would indicate that I had sequenced the full ETS. Differences in length of the produced sequences were examined during the alignment procedure.

Phylogenetic Analyses

Data partitions

Because my data consisted of both nuclear and chloroplast sequences I choose to partition the data so that regions that may have evolved under different evolutionary processes were analyzed as separate partitions. For the dataset containing 68 *Ephedra* specimens and outgroup taxa from the remaining vascular plants, I followed the approach in Rydin and Korall (2009) and divided the data into two partitions: chloroplast data and nuclear data. I also analyzed this dataset using additional partitioning schemes, one with four partitions (nuclear introns, nuclear ribosomal DNA, chloroplast introns, and protein-coding chloroplast genes), and one in which all nine gene regions constituted separate partitions. The same partitioning schemes were used for analyses of the dataset including 135 *Ephedra* specimens without outgroups.

Model selection

The best-fitting models for each DNA region and data partition were calculated using the software MrAIC (Nylander 2004). For all datasets, best-fitting models were calculated under corrected Akaike information criterion (AICc; Akaike 1973) and Bayesian information criterion (BIC). The main difference between the two criteria is that BIC penalizes complex models (models with many free parameters) to a higher extent than does the AICc, hence the AICc can lead to overparameterization and BIC can lead to underparameterization (see Burnham and Anderson 2002 for comparison). Model selection is very important in phylogenetic analyses (Huelsenbeck et al. 2004: Posada and Buckley 2004), and may strongly influence the results. How to appropriately choose evolutionary models is however an open question, and I decided to explore the impact of models selection on my results. When AICc and BIC calculated different best-fitting models, I ran two sets of analyses, and compared results employing evolutionary models selected under the BIC criterion as well as the AICc criterion. When choosing evolutionary model for partitions including more than one DNA region (i.e. the partitions in the two and four-partition schemes), and different models were calculated for the included DNA regions, the most parameter-rich model was selected. I further used a third approach, the Reversible Jump MCMC (RJ-MCMC) (Huelsenbeck et al. 2004) procedure for model selection implemented in MrBayes 3.2 (Ronquist et al. 2012). Instead of a priori choosing a model this procedure allows the MCMC chain to sample from all possible time-reversible models in proportion to each model's marginal probability. Hence, this approach accommodates for the uncertainty of selecting the correct evolutionary model (Huelsenbeck et al. 2004; Ronquist et al. 2012) by making it an integral part of the analysis. See table 2 and 3 for a summary of the selected models.

Table. 2. Description of the data set including 100 outgroup taxa and 68 *Ephedra* taxa: number of taxa, total number of characters, number (%) informative and variable sites, and best-fitting evolutionary model under AICc and BIC are shown for each marker as well as the nuclear and chloroplast data sets.

	AICc	BIC	Total taxa	Outgroup taxa	*Ephedra* taxa	Total characters	Variable sites	Informative sites (%) including outgroups	Informative sites (%) within *Ephedra*
18S	GTRIG	SYMIG	113	84	29	1715	675	414 (24.1)	7 (0. 4)
26S	GTRIG	GTRIG	110	80	30	1257	691	475 (37.8)	25 (2)
ETS	GTRG	HKYG	23	NA	23	1868	286	NA	146 (7.8)
ITS	GTRG	GTRG	67	NA	67	1617	299	NA	198 (12.2)
matK	GTRG	GTRG	57	NA	57	1280	55	NA	34 (2.7)
rbcL	GTRIG	GTRIG	129	100	29	1344	710	558 (41.5)	14 (1.0)
rps4	GTRIG	GTRIG	104	69	35	607	424	343 (56.5)	4 (0.7)
rpl16	F81	F81	30	NA	30	700	17	NA	6 (0.9)
$trnS^{UGA}$-$trfM^{CAU}$	GTRG	F81	28	NA	28	837	23	NA	15 (1.8)
Nuclear	GTRIG	GTRIG	168	95	73	6457	1951	1233 (19.1)	376 (5.8)
Chloro-plast	GTRIG	GTRIG	168	100	68	4768	1229	922 (19.3)	73 (1.5)
Total			168	100	68	11225	3180 (28.3)	2189 (19.5)	449 (4.0)

Table. 3. Description of the data set including 135 *Ephedra* taxa: number of taxa, total number of characters, number (%) informative and variable sites, and best-fitting evolutionary model under AICc and BIC are shown for each marker.

	AICc	BIC	No. taxa	Total characters	Variable sites	Informative sites (%)
18S	HKYG	K2P	64	1715	42	11 (0.6)
26S	GTRG	GTRG	61	1257	69	31 (2.5)
ETS	GTRG	HKYG	23	1868	286	146 (7.8)
ITS	GTRIG	GTRIG	134	1617	357	246 (15.2)
matK	GTRG	GTRG	96	1280	89	56 (4.4)
rbcL	GTRI	HKYI	67	1344	41	22 (1.6)
rps4	F81	F81	91	607	11	5 (0.8)
rpl16	F81	F81	61	700	20	8 (1.1)
$trnS^{UGA}$-$trfM^{CAU}$	GTRI	F81G	59	837	32	21 (2.5)
Total			135	11225	947	546 (4.9)

Bayesian analyses

Data sets for each partitioning scheme and the data sets for the ETS vs. ITS comparison were analyzed with a Bayesian approach using the Markov chain Monte Carlo (MCMC) algorithm. All analyses were run using software MrBayes 3.2.2. (Ronquist et al. 2012), on the CIPRES computing cluster (Miller et al. 2010). Metropolis coupled MCMC for all partitioning schemes (e.g., the two, four and nine-partition schemes) included one cold chain and three incrementally heated chains in each of the four runs. Number of generations were set to 20 million and sampling from the MCMC chain occurred every 1000th generation. The single-gene analyses had the same overall settings except that the number of generations was set to 10 million and two runs per analysis were implemented. The default temperature settings for the Markov chain were sometimes lowered when acceptance rates of the Metropolis coupled MCMC sampler was below 20% for any chain (swap frequencies ranging from 20% to 80% are adequate) (McGuire et al. 2007 and references within).

Convergence diagnostics

To evaluate whether the runs had converged, I monitored the standard deviation of split frequencies to be below 0.01, the effective sample size values (EES) to be above 200, and the potential scale reduction factor values (PSRF) (Gelman and Rubin 1992) to be close to 1.0 for all parameters, as recommended in the MrBayes manual (Ronquist et al. 2011). I also used the program Tracer v1.5 (Rambaut et al. 2013), and the online application AWTY (Are We There yet?) (Nylander et al. 2008) to assess performance of the Markov chain and convergence of runs. Tracer was used to examine mixing of parameters and if trace plots of the log likelihood values for the independent runs had reached stationarity. In AWTY, the function cumulative was used to assess when posterior probabilities of clades were stable across the analysis. Substantial variation of those values during the analysis indicates that the chain has not yet converged. The compare command was also used; it plots posterior probabilities of splits for two MCMC runs, and a strong correlation between the runs indicates that the runs have converged (Nylander et al. 2008). After the analysis of convergence, all samples prior to chain convergence were discarded as "burn-in". From the remaining post-burn-in trees a 50% majority rule tree was generated for each analysis.

Parsimony analysis

For the dataset containing 68 *Ephedra* specimens and outgroup taxa a maximum parsimony analysis was undertaken with PAUP v. 4.0b10 (Swofford 2003) using the heuristic search option on equally weighted characters with the tree-bisections-reconnection (TBR) algorithm, 1000 random sequence addition replicates, Multrees option in effect, holding 1 tree at each step and no more than 10 trees saved per replicate.

For estimation of branch support, a bootstrap analysis was performed as follows: 1000 bootstrap replicates, heuristic search with three random sequence addition replicates, each holding 1 tree at each step and no more than 10 trees saved per replicate, Multrees option in effect and TBR branch swapping.

Results

Testing the sister relationship between *Ephedra foeminea* and remaining species of the genus

The result of the two-partition analysis (fig. 2), based on nine gene regions and including a large set of outgroup taxa, resolved the Gnetales as a whole, and the genus *Ephedra* as monophyletic (1.00 posterior probability) and the 13 specimens of *E. foeminea* were separated from the clade comprising all other species of *Ephedra* (0.91). The result from the two-partition analysis is presented as a 50% majority rule consensus tree (fig. 2). Analyses of the four and nine-partition

datasets (not shown) resulted in the same topology, but branch support was generally lower than those obtained from the two partition dataset. The results from MP analysis show *E. foeminea* to be monophyletic with 70% MP bootstrap support (BS). The relationships to the remaining species were collapsed and did thus not contradict the results from the Bayesian analyses.

Assessing further relationships within *Ephedra*

Based on the results described above, I decided to make further analyses of relationships within *Ephedra* using *E. foeminea* as outgroup, and employing a more extensive sampling of specimens. Analyses under different partitioning schemes showed no conflicting topologies. The best resolved and supported trees were obtained from the nine-partition analyses, employing evolutionary models selected under the BIC (fig. 3) and using Reversible Jump MCMC. The phylogenetic estimates under AICc, BIC and Reversible Jump-MCMC were almost identical, except in one case regarding the placement of *E. alata*. The results described below are from the nine-partition analyses employing evolutionary models selected under AICc, BIC and Reversible Jump-MCMC and differing results between the three approaches are reported.

The trees were rooted on *E. foeminea*. The first diverging clade consisted of a species complex of Mediterranean taxa, *E. aphylla*, *E.fragilis*, *E.major* and *E. altisimma* (clade A; 1.00), within which *E. fragilis and E. altissima* (clade B; 1.00) are sisters to the others (clade C; 0.98). Species delimitations within the Mediterranean species complex (clade A) are uncertain.

In the analysis employing evolutionary models selected under the AICc criterion (not shown), the sister group to clade A constitutes an unresolved trichotomy comprising *E. alata* (clade D), *E. milleri* (clade E) and the remaining species (clade F). However, when analyzing the same nine-partition dataset using evolutionary models selected under the BIC criterion, and when using Reversible Jump MCMC, this trichotomy is resolved with *E. alata*, (clade D*:* 1.00) being sister to the remaining *Ephedra* (0.86 and 0.82 respectively), within which *E. milleri* is sister to the rest of the *Ephedra* species (1.00). Clade F (i.e., "core *Ephedra*" of Rydin and Korall 2009) consists of two sister clades. One of the clades consists of *E. foliata* and *E. ciliata* specimens (clade G) (1.00) with the two species being sisters with high support. The other clade comprises two large sister clades, one clade consisting of mainly American and Chinese species (0.99) with the Chinese species (clade H) (1.00) being sister (1.00 PP) to the mainly American species (clade I) (1.00), and the other large sister clade consisting of remaining mainly Asian species (clade J) (1.00). Within clade J, clade K (1.00) is sister to the remaining species. The first divergence of the remaining species is clade L (1.00) comprising *E. pachyclada* and *E. somalensis* being sister to clade M where an undetermined specimen sampled in India (Deoban) is sister (0.94) to the remaining species.

Single gene analyses; nuclear and chloroplast analyses

Results of the single gene analyses of the *matK*, ITS and ETS regions were in general congruent but showed two supported topological conflicts (the trees were rooted on *E. foeminea*). One conflict concerns the placement of *E. foliata* and *E. ciliata*. In the ITS analysis, the ETS analysis, and the combined analyses (fig. 3), the two species are part of the species corresponding to clade F (see above) whereas in the single gene analysis of the *matK* the two species are sisters to the species complex corresponding to clade A (see above). This group (*E. foliata*, *E. ciliata* and clade A) is in turn sister to all other species. The other topological conflict regards the position of *E. alata*. In the ITS analysis it is the first divergence after *E. foeminea,* whereas in the *matK* and ETS analyses it is nested among species corresponding to clade I and J. For the placement of *E. alata* in the combined analyses, see above. For results and discussion of single gene analyses for all other gene regions, see Rydin and Korall (2009). There were no supported topological conflicts within *Ephedra* between nuclear and chloroplast datasets and the combined data set.

Assessments of the ETS structure

The 5'ETS sequences produced for this study contained no putative TATA box, although a putative TATA box was found 1071 bp into the *E. major* sequence (*GenBank* JX843794).

Amplification and sequencing of the ETS

Initial PCRs conducted using the *IGS2inF1K* primer orientated in close proximity to the putative TATA box found in JX843794 did not yield any satisfactory amplification products, whereas PCRs conducted using the *IGS2inF2K* primer orientated c. 1 kb downstream of the putative TATA box was more successful. PCR-products were retrieved from a total of 34 of the tested specimens, of which 31 specimens visualized a single band of similar lengths, and three visualized double bands. Of the successful amplifications, representatives from all major clades were present. The three specimens with double bands were not sent for sequencing. Of the 31 specimens with a single band, 26 were fully or partially sequenced, i.e., not all sequencing reactions per specimen were successful. Due to unsuccessful sequencing reactions and partially poor reads it was not possible to assemble a consensus sequence for all specimens even though some sequencing reactions were of good quality. Specimens with dubious reads were excluded and in the end 23 specimens (representing all major clades) were used for further analysis.

Phylogenetic utility of ETS in comparison to ITS

The lengths of the ETS sequences varied from 1391 bp to 1866 bp, which mainly owes to repeated segments in different numbers. The longest sequence, *E. americana*, differed from the other species by having insertions of in total c. 120 bp that was absent in other sequences.

The number of variable and informative characters were higher in the ETS dataset than in the ITS dataset. The ETS dataset included 1868 bps, of which 15.3% (286) were variable and 7.8% (146) were informative. The ITS dataset included 1617 bps, of which 12.6% (195) were variable and 5.2% (85) were informative.

The results from the Bayesian analyses using evolutionary models selected under AICc and BIC were essentially identical. The results described here are from the Bayesian analyses using models selected under AICc. ETS data resulted in better intraspecific as well as interspecific resolution than did the ITS tree (figs 4A and 4B). There were a few topological conflicts between the two trees. In these analyses, which had a relatively restricted taxon sampling, ITS resolved *E. alata* as sister (0.93) to the remaining species, whereas ETS data placed *E. alata* within a clade comprising *E. americana, E. minuta* and *E. ciliata*. The placement of *E. foliata* also differed between the two trees; in the ETS tree it is sister (1.00) to the *E. major*-clade and the clade comprising *E. americana, E. minuta* and *E. ciliata,* whereas ITS data placed *E. foliata* in a weakly supported clade also comprising *E. americana, E. minuta* and *E. ciliata.*

Within the clade comprising all species of *Ephedra* except *E. foeminea*, the combined ETS+ITS analysis (fig. 4C) resolved *E. aphylla* as sister to remaining species with strong support (0.99). The next divergence is the clade with *E. fragilis* specimens, which are sister to the remaining species with strong support (1.00). *Ephedra alata* is the next diverging species, and it is sister to the remaining species with strong support (1.00).The remaining species are here represented by two sister clades; one, a highly supported clade (1.00) constituting *E. major* specimens, and second, a weakly supported clade consisting of *E. foliata*, *E. minuta*, *E. americana* and *E. ciliata*.

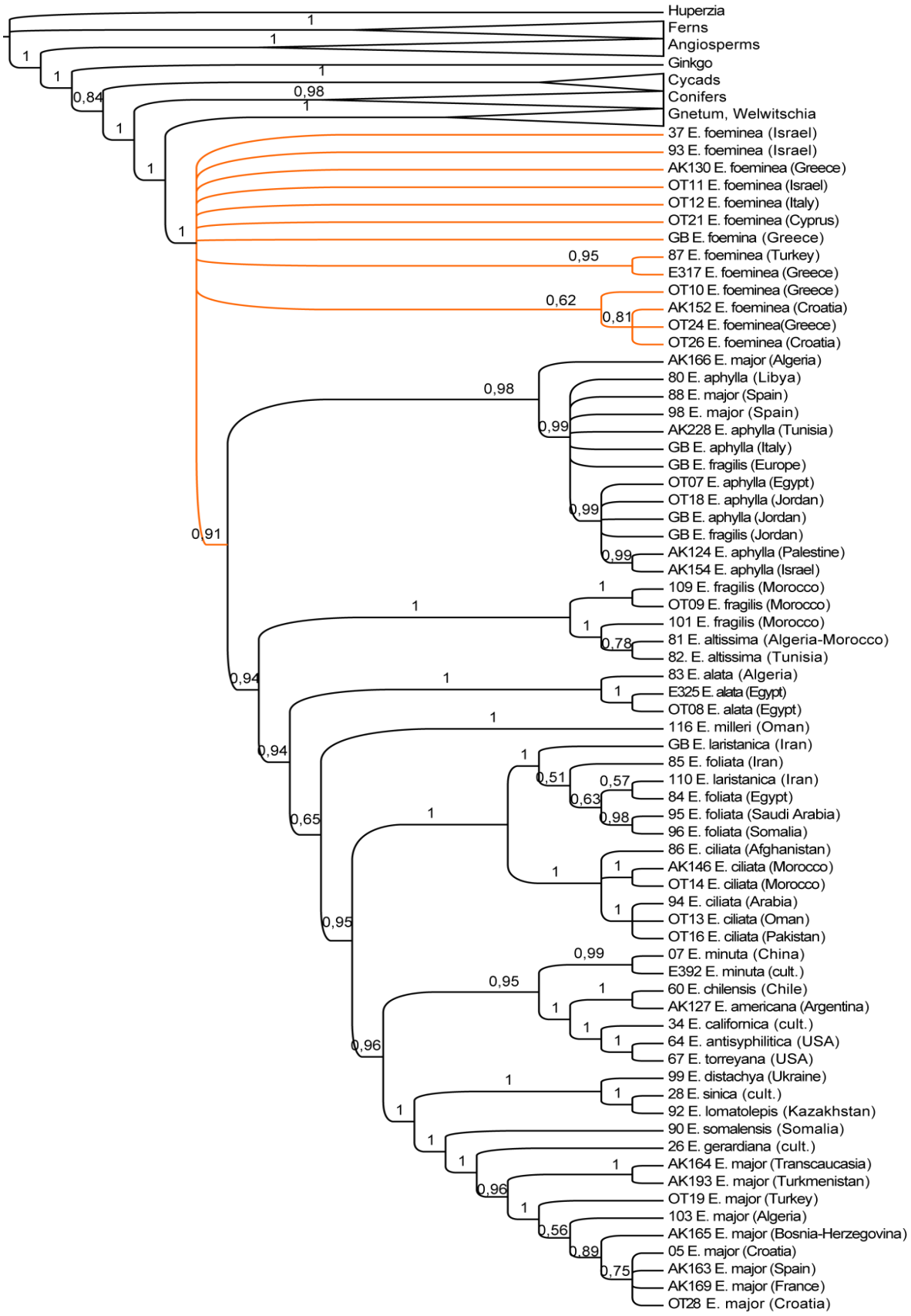

Fig. 2. Phylogenetic estimate of vascular plants and relationships within *Ephedra*: 50% majority rule consensus tree based on 20,000 post-burn in trees from the two-partition dataset of nine phylogenetic markers. Posterior probability values of clades are shown above branches. Colored branches visualize the position of *E. foeminea*. Outgroup clades are pruned to only show the position of the clades, not relationship among taxa within them.

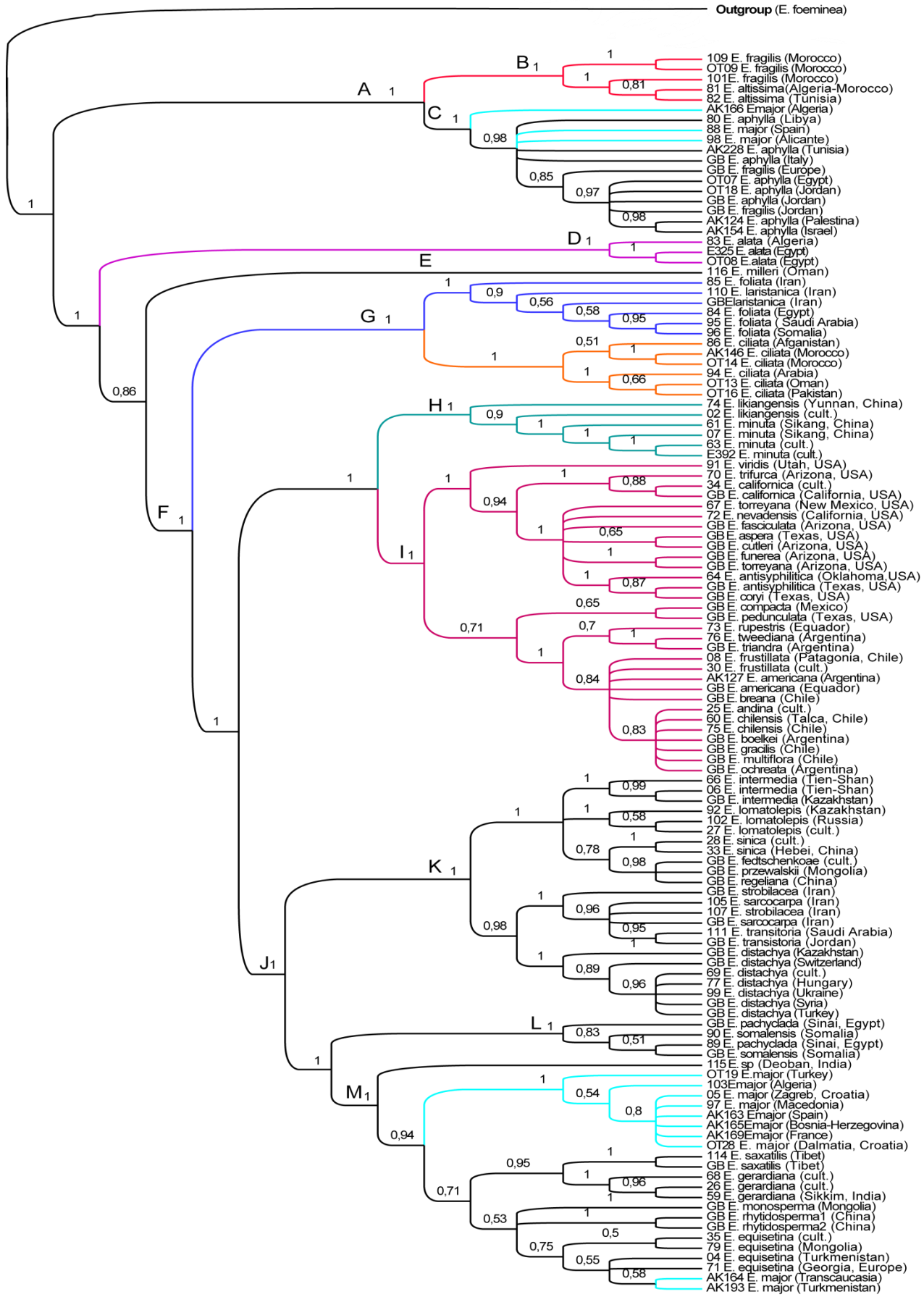

Fig. 3. Phylogenetic estimate of relationships within *Ephedra*: 50% majority rule consensus tree based on 60,000 post-burn in trees from the nine-partition dataset of nine phylogenetic markers using evolutionary models selected under BIC. The tree is rooted on *E. foeminea* . Numbers above branches represents posterior probabilities of clades. Letters above branches represents clades discussed in the text. Colorizations of branches are for easier visualization of clades and sub-clades discussed in the text.

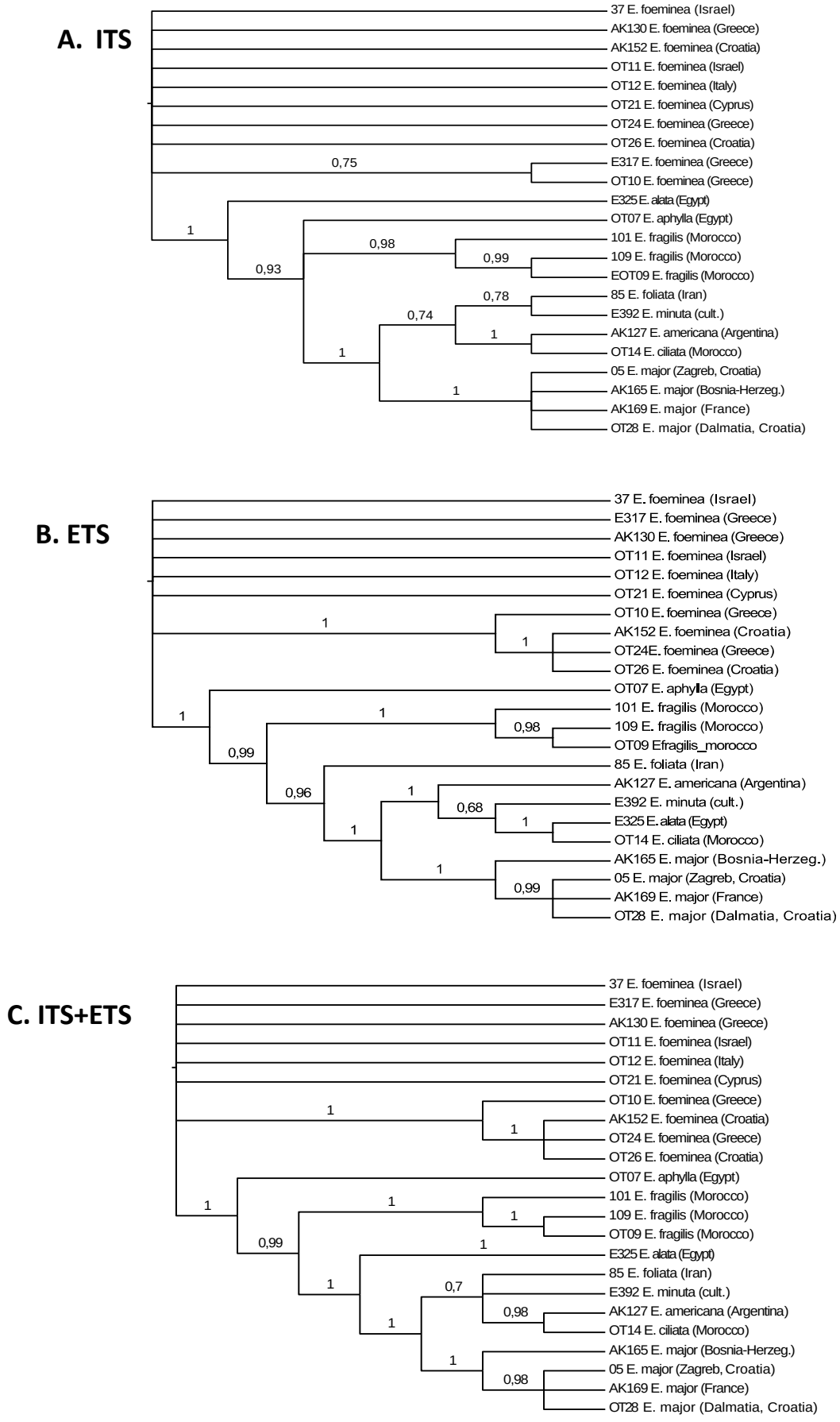

Fig. 4. Phylogenetic estimate based on the ITS, ETS and ITS+ETS combined analyses. For each analysis: 50% majority rule consensus tree based on 15,000 post-burn in trees; evolutionary models selected under AICc. Tree A is based on the ITS data set. Tree B is based on the ETS data set. Tree C is based on a dataset including both ITS and ETS sequences, divided into two partitions. Posterior probability values of clades are given above branches.

Discussion

Ephedra foeminea and the root of the *Ephedra* phylogeny

Deep divergences in *Ephedra* have been addressed in several studies in the past and have proven difficult to resolve. One major reason for this is probably the low amount of sequence divergence within *Ephedra*, in combination with the distant relationships to the closest living relatives of the genus. This has made it difficult to find suitable markers, which contain sufficient information to resolve relationships within the genus and at the same time are alignable to outgroup taxa. The present study lends strong support for the hypothesis of a sister-relationship between *Ephedra foeminea* and all other species of the genus. In the Bayesian analysis of the two-partition dataset including outgroups from all major clades of vascular plants, the *E. foeminea* specimens are separated from the remaining *Ephedra* species with a much higher support compared to the results in Rydin and Korall (2009) (0.91 vs. 0.69). Further, many other nodes in the phylogeny are better supported than in Rydin and Korall (2009) or any previous study. Ickert-Bond et al. (2004) addressed the problem mainly using nrITS data, and mid-point rooting because ITS sequences are not possible to align against those of *Welwitschia* and *Gnetum*. This resulted in a sister relationship between the Iranian taxon *Ephedra laristanica* and remaining *Ephedra*. Rydin et al. (2004) took a different approach and utilized a combination of nrITS data and a set of more slowly evolving gene regions, which resulted in a sister-relationship between a clade of Mediterranean taxa, and a clade comprising all other species of the genus. Neither of these results was well-supported in a statistical sense, and neither attained any obvious support from available morphological information. Therefore; new attempts were made, and results in Rydin and Korall (2009) indicated a possible relationship between the eastern Mediterranean species *E. foeminea* and all other species. *Ephedra laristanica* was instead nested within specimens of the morphologically variable *E. foliata* (Rydin and Korall 2009).

However, the topological results in Rydin and Korall (2009) were weakly supported, and were contradicted not only by results in earlier studies but also by analyses conducted in Rydin and Korall (2009) using parsimony, and by a later study of divergence times of clades using the software BEAST (Ickert-Bond et al. 2009). Even if support values in the present study are not completely satisfactory, my results clearly support the hypothesis of *E. foeminea* being sister to all other *Ephedra* species. However, the 13 *E. foeminea* specimens included here do not form a clade, and as in previous studies, I cannot confirm that *E. foeminea* is a single species. I have not assessed morphological variation within *E. foeminea* but according to previous studies there is no reason to believe that *E. foeminea* should be divided into more than one species based on morphological characters (Freitag and Maier-Stolte 1989), and the results here do not reject the possibility that *E. foeminea* is a single species. On the contrary, two special traits of *E. foeminea,* the presence of pollination drop-producing non-fertile female organs in the male cones (pers. comm. Catarina Rydin), and entomophily (Bolinder 2011) distinguishes *E. foeminea* from all other species of *Ephedra*. Results from my four and nine-partition analysis (not shown) indicate a geographical pattern were specimens from Greece, Croatia and Turkey form a weakly supported sub-clade within *E. foeminea*, which is nested among specimens from Israel, Italy and Cyprus. Future population studies of *E. foeminea* could use these indications as a starting point for analyses of phylogeography and assessments of intraspecific morphological variation.

Relationships within remaining *Ephedra*

Compared to previous studies, the Bayesian analyses in this study resulted in more well-resolved and well-supported estimates of phylogeny in *Ephedra,* and in addition a few noteworthy differences. The results between the alternative partitioning schemes were highly similar and showed no supported conflicts. However, the most well-resolved and well-supported trees were obtained from

nine-partition analyses. Hence the results discussed below are from the analyses of the nine-partition dataset. (See also fig. 2)

Clade A (*E. aphylla-E. major-E. fragilis- E. altissima*)

In Rydin and Korall (2009) the Mediterranean species *E. altissima, E. aphylla*, *E. fragilis*, and *E. major* are members of a weakly supported group, in which *E. fragilis* is polyphyletic and species delimitations are uncertain (Rydin and Korall 2009). With the new data used in this study, the same pattern appears, and with higher support (1.00). Clade A, the Mediterranean clade, is divided into two highly supported sister groups, clades B and C.

Clade B (*E. fragilis* and *E. altissima*) and clade C (*E. aphylla*, *E. major*, *E. fragilis*)

Clade B, which consists of three specimens of *E. fragilis* from Morocco, and two specimens of *E. altissima* from Algeria-Morocco and Tunisia respectively, is highly supported as a monophyletic group and sister to clade C. With the present taxon sampling and considering the apparent polyphyly of *E. fragilis*, it is not possible to assess whether this result reflects the phylogenetic position of the two species *E. fragilis* and *E. altissima*, or if it rather indicates a geographical pattern (perhaps within a single species).

Clade C comprises specimens of *E. aphylla*, *E. major* and *E. fragilis* and indicates uncertain species delimitations (*E. major* will be discussed below). Nevertheless there is a geographical pattern within clade C, where *E. aphylla* specimens from Egypt, Israel, Palestine and Jordan, and one *E. fragilis* from Jordan, form a highly supported sub-group with an "eastern" distribution. This clade is nested among the remaining specimens, which have a more "western" distribution in Algeria, Spain, Tunisia, Italy and Libya. So just as in the case as with *E. foeminea,* my results does not indicate a need to divide *Ephedra* into additional species, but instead phylogeographic patterns within species. It is in fact possible that clades B and C should be interpreted as representing one single species each (clade B = *E. altissima*; clade C = *E. aphylla*), and that the species concept *E. fragilis* is redundant.

Clade D (*E. alata*)

The position of the desert plant *E. alata* was unresolved in Rydin and Korall (2009) where it was represented by one specimen (83) sampled in Algeria. In the present study there is remaining uncertainty regarding the placement of *E. alata,* which is here represented by two additional specimens from Egypt (OT8 and E325). In the analysis using evolutionary models selected under the BIC criterion and the analysis using Reversible Jump MCMC, *E. alata* is placed as the next divergence after clade C, but with moderate support (0.86 and 0.82 respectively). In the nine-partition analysis using evolutionary models selected under the corrected AIC criterion, *E. alata* is found in an unresolved trichotomy consisting of *E. alata*, *E. milleri,* and the remaining species. Furthermore, in single gene analyses, the position of *E. alata* differs and these contradicting results are very difficult to explain. Therefore, examination of morphology, ecology, chromosome number and molecular data for several *E. alata* specimens covering its distribution range would be worthwhile. *Ephedra alata* has a wide distribution range occurring from western North Africa to the Arabian Peninsula. It is an integral part of the Saharo-Sindian region (Freitag and Maier-Stolte 1994) where annual precipitation rarely exceeds 100 mm (Talebi et al. 2014) and is likely to be the species that is most well-adapted for life with very limited supply of water of all species in the Mediterranean area (Freitag and Maier-Stolte 1994). Based on its winged and dry cone bracts at seed maturity, *E. alata* has traditionally (i.e., Stapf 1898) been classified in section *Alatae*. The other species in section *Alatae* that share those features (*E. przewalskii* and *E. strobilaceae*) are in the present study shown to be distantly related to each other and to *E. alata*. Hence dry cone bracts have evolved separately several times, probably as an adaption to extremely arid conditions where wind dispersal would be more favorable than animal dispersal (Rydin and Korall 2009). Another interesting feature is that extracts of *E. alata* have been shown to control mold contamination of soybeans and maize (Al-Qarawi et al. 2011).

Clade E (*Ephedra milleri*)

As in Rydin and Korall (2009), *E. milleri* is in all analyses sister to core *Ephedra* with strong support. Very little is known about this relatively newly described taxon (Freitag and Maier-Stolte 1992). The type material consists of a few herbarium sheets of plants collected in Oman, and initially referred to other taxa. New collects and additional material is clearly needed in order to better understand the morphological distinctiveness and geographic distribution of this species.

News within clade F: The core *Ephedra*

Within core *Ephedra* clade G (*E. foliata* and *E. ciliata*) is sister to the remaining species, with the two species being sister groups with high support, and *E. laristanica* is nested among specimens of *E. foliata*. Nomenclatural considerations of taxa included in clade G have varied in the past, most probably because intraspecific variation is considerable, at least within *E. foliata* (pers. comm. Catarina Rydin). In such cases, the clear results provided by molecular data (Rydin and Korall 2009, and the present study) are helpful, and should be considered in future alpha-taxonomic revisions. A new and rather surprising result regarding the core *Ephedra* is the placement of clade H (1.00), which consists of specimens from two Chinese species, *E. likiangensis* and *E. minuta*. In Rydin and Korall (2009) this clade is sister to all other species in the mainly Asian clade (J). In this study, however, it is sister to the New World species (clade I) with strong support. This result will have considerable implication for future studies of biogeography, dispersal, and character evolution in *Ephedra*. In addition, the specimens of *E. likiangensis* do not form a clade in the present study (but a grade). This result has been seen previously (C. Rydin pers. comm.), but whether or not it indicates uncertain species delimitation among these taxa are difficult to say. Material availability is limited for these two Chinese endemics, and few specimens have been investigated. They differ, however, dramatically in gross morphology; *E. minuta* is tiny dwarf-shrub and *E. likiangenesis* is a large shrub with ascending branchlets.

The polyphyly of *E. major*

As in Rydin and Korall (2009), *E. major* is found in two distantly related and highly supported clades (clade A and M in this study). Three specimens (two from Spain and one from Algeria) are found in clade A whereas the other specimens (from Algeria, Turkey, Croatia, Macedonia, Bosnia-Herzegovina, Spain, France, Transcaucasia and Turkmenistan) are found in clade M. Within clade M, the *E. major* specimens are further divided with the specimens sampled in Algeria, Turkey, Croatia, Macedonia, Bosnia-Herzegovina, Spain and France forming a highly supported group and the more eastern specimens sampled in Transcaucasia and Turkmenistan are found in a poorly supported clade also containing *E. saxatilis*, *E. gerardiana*, *E. monosperma*, *E. rhytidosperma* and *E. equisetina*. Rydin and Korall (2009) examined the *E. major* vouchers found in clade A and argued that their phylogenetic position among species in the "Mediterranean grade" is plausible with respect to morphological characters, but they also notion that most vouchers in question were vegetative and therefore difficult to assess (Rydin and Korall 2009). Rydin et al. (2010) studied the matter further based on data from female reproductive structures, and found clear morphological support for the inclusion of some specimens of *E. major* among Mediterranean taxa (here clade A) and others among Asian taxa (here clade M). And recently, Norbäck Ivarsson (2014) came to the same conclusion based on pollen data. The (neo)type material of *E. major* (Riedl 1993) sampled in Dalmatia (Croatia), most likely corresponds to the specimens found in the Asian clade (pers. comm. Catarina Rydin), and this is in line with the placement of my specimen of *E. major* (EOT28), which was newly collected in Dalmatia for the present study.

In the present study, the position of two central-Asian specimens of *E. major* among other Asian species probably reflects incorrect species determination; these specimens are probably *E. equisetina*, not *E. major*. As so often is the case in *Ephedra*, species determination and species delimitations among these Asian taxa have been debated. Freitag and Maier-Stolte (1994) considered *E. equisetina* and *E. major* synonymous, but my results clearly support (the Asian) *E.*

major as a separate species. The morphological distinctiveness of *E. major*, *E. equisetina* (and *E. gerardiana*) remains unclear, however, but may be of practical importance. Of all species in the genus, *E. major* may be the species that produces and accumulates the highest amounts of ephedrine. Although it is avoided by grazing animals (Freitag and Maier-Stolte 1994), humans may thus have a certain interest in the possibility to detect and collect this particular species in the field without mistaking it for other species.

ETS- a new phylogenetic marker for *Ephedra*

The results from the present study have showed the sequence data from the ETS region to be extremely useful for resolving relationships within *Ephedra,* even more variable and informative than the ITS region, which in previous studies was shown to be the most phylogentically informative marker (Ickert-Bond and Wojciechowski, 2004; Rydin et al. 2004; Huang et al. 2005; Rydin and Korall 2009). For a genus like *Ephedra,* where phylogenetic reconstruction has been hampered by few informative characters in previously utilized molecular markers, the additional information provided by a new nucleotide marker (ETS in this case) is invaluable. I believe additional ETS-data will be useful for resolving the remaining phylogenetic problems in *Ephedra*, e.g., the phylogenetic position of the Central species *E. compacta* and *E. pedunculata*, and perhaps also relationships among closely related taxa in Central Asia and in America, respectively.

Assessments of the ETS structure

The nuclear ribosomal DNA (nrDNA) is arranged in tandem arrays in the plant genome where the regions coding for the large subunit (26S and 5.8S) and the small subunit (18S) are separated from each other by internal transcribed spacers (ITS1 and ITS2) (fig. 1). Spacer regions are referred to non-coding DNA that separate functional gene regions from each other. The co-transcribed 18S-5.8-26S nrDNA gene blocks are separated by an intergenic spacer (IGS). The IGS consists of the non-transcribed spacer (NTS) and two external transcribed spacer (3'ETS and 5'ETS) (Poczai and Hyönen 2010). In *Ephedra* the separately transcribed 5S nrDNA gene divides the IGS in two parts (Garcia and Kovaric 2013).

As the transcription of the 18S-5.8-26S nrDNA gene blocks starts with the 18S gene the ETS, located upstreams of the 18S gene, is labeled 5'ETS and the ETS downstream of the 26S gene is labeled 3'ETS. In phylogenetic studies the 5'ETS is the most commonly used, which is also the case in the present study. The term external derives from the fact that the spacer is located externally with respect to the 18S-5.8-26S nrDNA gene blocks, in contrast to the internal transcribed spacers, which are incorporated inside the gene blocks. The term transcribed explains that the ETS and the ITS are parts of the nrDNA transcriptional unit. Hence the ETSs and ITSs are part of the ribosomal precursor RNA, but are removed during rRNA maturation (van Nues et al. 1995). The start of 5'ETS is defined by the transcription start site (TSS) (Poczai and Hyönen 2010). In earlier studies of the organization of the IGS (NTS+ETS), the location of the TSS has been defined by the presence of a so called TATA box (Linder et al. 2000; Galian et al. 2012; Grabiele et al. 2012). A TATA box is a commonly found DNA sequence in promoter regions of eukaryotic organisms, located about 25 bp upstreams of the transcriptional start point (Campbell et al. 2008). The name TATA box refers to the conserved DNA sequence. In plants, the putative consensus sequence is TATA(R)TA(N)GGG (Galian et al. 2012 and references within). The TATA box is involved in the binding of transcription factors to the DNA, which in turn recruits the transcription enzyme RNA polymerase that forms the transcription initiation complex (Campbell et al. 2008).

As no putative TATA-box was found in the ETS sequences produced for this study, the sequences are most likely incomplete. Attempts were made to utilize the PCR primer *IGS2inF1K*, which is orientated in close proximity to the putative TATA-box, but these attempts were unsuccessful. When analyzing the 3358 bp long IGS sequence of *E. major* (GenBank JX843794) a putative TATA box was found 1071 bp into the sequence, which implies that the ETS is around 2300 bp in *Ephedra major*. However,

when considering the length of ETS within *Ephedra* caution should be taken. First, this is to my knowledge the first attempt to determine the start of the ETS within *Ephedra.* Second, the promoter identifier database used is based on known sequences mostly from angiosperms (a few gymnosperms but no Gnetales). Third, there is increasing evidence suggesting that the promoter sequence is not that well conserved as previously believed (Galian et al. 2012 and references within). Considering the data at hand, it appears however obvious that the ETS sequences of *Ephedra* are quite long compared to that of other plants. For example, the ETS in *Gingko bilboa* is 1641 bp (Galian et al. 2012), *Capsicum pubescence* (Solanaceae) 966 bp long (Grabiele et al. 2012), and *Argyrantheum* (Asteraceae) is 1003 bp. The ETS of *Helianthus* (Asteraceae) ranges from approximately 1600 bp to 2100 bp (Linder et al. 2000), and the variable length is due to different number of subrepeats (Linder et al. 2000). Extensive investigations of the nature and intrageneric variation of the ETS region in *Ephedra* is beyond the scope of the present study but simple eye inspection of my ETS alignment indicates considerable length variation of ETS also in *Ephedra*.

Conflicting information in ITS and ETS

A classical problem in systematics, discussed e.g. in Cunningham (1997) and Li et al. (2006) occurs when different gene regions sequenced from the same specimens results in different topologies. A core question is whether contradicting gene regions should be combined or analyzed separately (Cunningham 1997; Li et al. 2006 and references within). The partly conflicting topologies found here between results from ETS and ITS is an example of this. Despite the conflicts, I chose to analyze them together. I found it reasonable to do this because both spacers occur in the 18S-26S nrDNA transcriptional unit, and because there is evidence that suggests that ITS and ETS may have interdependent roles in rRNA maturation (Good et al. 1997). Therefore coevolution between the ITS and the ETS is likely (Andreasen and Baldwin 2001). Another motivation to combine the datasets was that Cunningham (1997) found that combining two incongruent datasets often generated better resolution than the single datasets alone, something which there probably is numerous empirical examples of (pers. comm. Catarina Rydin), although not always clearly stated in the resulting publications. Moreover, Davolos et al. (2012) considered nodes derived from different datasets to be incongruent if they had ≥0.97 posterior probability. If I would follow their approach, there would be no hard incongruences between the ETS and ITS trees. The reason for the supposed sister species *E. ciliata* and *E. foliata* not grouping together in my ETS tree is most probably due to incomplete sampling of taxa, as well as of characters; only half of the ETS regions of the *E. ciliata* and the *E. foliata* specimens were successfully sequenced.

Concerted evolution

The multicopy nature of then rDNA arrays makes it possible that divergent repeat types of the ETS and ITS are present within the genome. However, the copies are expected to be homogenized by the process of concerted evolution (Poczai and Hyvönen 2010). Concerted evolution is the phenomenon where all copies of genes in a gene family evolve in concert, for example in the nrDNA arrays. This is possible due to two mechanisms: unequal crossing over and gene conversion were the latter is believed to be the main mechanism of concerted evolution in the rDNA arrays. Due to those mechanisms, the homogeneity between the multiple copies of nrDNA arrays in an individual is almost complete (Poczai and Hyvönen 2010 and references within). Nevertheless, directly after events such as hybridization or polyploidization, different repeat types could be present within the genome and concerted evolution might not homogenize the new alleles immediately (Calonje et al. 2009), hence the different alleles can interfere with the sequencing procedure and disrupt the phylogenetic signal. Further, Okuyama et al. (2005) discusses that the ITS region could have a lower rate of concerted evolution than does the ETS region, perhaps due to the presence of the promoter region within IGS (where ETS is included) (Polcano et al. 1998). Conceivably, concerted evolution in the ETS region will homogenize new alleles whereas the slower concerted evolution in the ITS region will take longer to homogenize after a hybridization event (Okuyama et al. 2005). Hence,

incongruences between datasets could ultimately be explained by uneven rates of concerted evolution.

Concluding remarks and outlook

Using an increased sampling of taxa and characters, this study reveals new insights to the divergences in *Ephedra* both in terms of credibility of results (posterior probability values) and resolution. It is somewhat surprising that *E. foeminea* is not resolved as a clade in the Bayesian analyses, in particular as the results from the MP analysis resolves *E. foeminea* as monophyltetic with moderate support (70% BS). Furthermore my study also provides some unexpected results (such as the new position of clade H; *E. minuta* and *E. likiangensis*) that will have considerable implications for future and ongoing studies of evolution and biogeography in the genus. My study is also the first to have developed primers for amplifying the external transcribed spacer (ETS) of the nuclear ribosomal DNA in *Ephedra*. The primers produced here have proven useful for amplifying and sequencing ETS of distantly related species within the genus, although the sequence is not complete and was not successfully amplified for all tested species. Nevertheless, the results have shown that the ETS contains more phylogenetically informative characters than does the internal transcribed spacer (ITS) and it will probably be useful in future studies addressing some difficult and unresolved relationships in *Ephedra*. Most *Ephedra* species are very similar and to determine species can be very difficult; even with the presence of male and/or female cones mistakes are made (Freitag and Stolte 1994). The results regarding the polyphyletic nature of *E. fragilis* and *E. major* and the troublesome species delimitation in clade A (fig. 3) highlight the difficulty of species and lineage delimitation within *Ephedra* and suggests that a taxonomic revision of the genus could be in place. The new insights of the phylogenetic relationships within *Ephedra* in this study in combination with new morphological and ecological information (e.g., Rydin et al 2010; Loera et al. 2012; Norbäck-Ivarsson 2014; Bolinder et al. in progress) could provide a solid and interesting basis for a new classification of *Ephedra*. Further, because of the similarities in overall morphology, and intraspecific variation within *Ephedra,* which often are substantial, it is probably important to utilize molecular data in combination with morphology. The external transcribed spacer ETS emerges as a powerful and promising phylogenetic and taxonomic tool, perhaps even suitable as a “genus-specific barcode” for species identification within *Ephedra*.

Acknowledgments

First and foremost I would like to thank Catarina Rydin for all invaluable help and support during the course of this project. Thank you, Kristina Bolinder and Chen Hou, for great collaboration within the research group, Anbar Khodabandeh and Anna Petri for all your laboratory advices, Niklas Wikström, Kent Kainulainen, Aelys Humphreys and Frida Stångberg for helping out with computer and analytical issues, Sylvain Razafimandimbison for lending of blotters and advice regarding field work, my fellow master student colleagues Eva Larsén, Julia Ferm, Karolina Jerrå, Anna Ginter and Mohannad Yousef for good talk, Lena Norbäck Ivarsson for assistance in field, the whole department for always keeping your doors open, André Aptroot at Pinetum Blijdenstein for access to information regarding their *Ephedra* collection, Moshira Hassan, Prof. Dr. Wafaa Amer and the herbaria BR, CAI, L, S, W and WU for access to material.

References

Akaike H. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60:255-265.

Al-Qarawi A.A., Abd_Allah E.F. and Abeer H. 2011. *Ephedra alata* as biologically-based strategy inhibit aflatoxigenic seedborne mold. African Journal of Microbiology Research 5:2297-2303.

Andreasen K. and Baldwin B.G. 2001. Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): Evidence from 18S–26S rDNA internal and external transcribed spacers. Molecular Biology and Evolution 18:936-944.

Arber E. and Parkin J. 1908. The relationship of the angiosperms to the Gnetales. Annals of Botany 22:489-515.

Baldwin B.G. and Markos S. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). Molecular Phylogenetics and Evolution 10:449-463.

Braukmann T.W., Kuzmina M. and Stefanović S. 2009. Loss of all plastid ndh genes in Gnetales and Conifers: extent and evolutionary significance for the seed plant phylogeny. Current Genetics 55:323-337.

Bolinder K. 2011. Pollination mechanisms in *Ephedra* (Gnetales) in north-eastern Greece. Degree project in biology, Master of Science, Uppsala University.

Burnham K.P. and Anderson D.R. 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach Second Edition. New York, Springer, 488 pp.

Campbell N.A., Reece J.B., Urry L.A., Cain M.L., Wasserman S.A., Minorsky P.V. and Jackson R.B. 2008. Biology 8th Edition. San Francisco, Pearson, 1267 pp.

Calonje M., Martin-Bravo S., Dobes C., Gong W., Jordan-Thaden I., Kiefer C., Kiefer M., Paule J., Schmickl R. and Koch M.A. 2009. Non-coding nuclear DNA markers in phylogenetic reconstruction. Plant Systematics and Evolution 282:257–280.

Chase M. W., Soltis D. E., Olmstead R. G., Morgan D., Les D. H., Mishler B. D., Duvall M. R., Price R. A., Hills H. G. , Qui Y-L., Kron K. A., Rettig J. H., Conti E., Palmer J. D., Manhart J. R., Sytsma K. J., Michaels H. J., Kress W. J., Karol K. G., Clark W. D., Hedrén M., Gaut B. S., Jansen R. K., Kim K-J., Wimpee C. F., Smith J. F., Furnier G. R., Strauss S. H., Xiang Q-Y., Plunkett G. M., Soltis P. S., Swensen S. M., Williams S. E., Gadek P. A., Quinn C. J., Eguiarte L. E., Golenberg E., Learns G. H., Graham S. W., Barrett S. C. H., Dayanandan S., and Albert V. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Annals of the Missouri Botanical Garden 80:528–580.

Chaw S., Parkinson M.C.L., Cheng Y., Vincent T.M. and Palmer J. D. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and the origin of Gnetales from conifers. Proceedings of the National Academy of Sciences 97:4086-4091.

Crane P. R. 1985. Phylogenetic analyses of seed plants and the origin of angiosperms. Annals of the Missouri Botanical Garden 72:716-793.

Crane P. R. 1996. The fossil history of the Gnetales. International Journal of Plant Sciences 157:50-57.

Cunningham C.W. 1997. Can three incongruence tests predict when data should be combined? Molecular Biology and Evolution 14:733-740.

Dávalos L.M., Cirranello A.L., Geisler J.H. and Simmons N.B. 2012. Understanding phylogenetic incongruence: lessons from phyllostomid bats. Biological Reviews 87:991–1024.

Doyle J.A. and Donoghue M.J. 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. Botanical Review 52:321-431.

Doyle J.J. and Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19:11-15.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput Nucleic Acids Research 32:1792-1797

Engels B. 2005. Amplify 3.1 University of Wisconsin. http://engels.genetics.wisc.edu/amplify/

Freitag H. and Maier-Stolte M. 1989. The *Ephedra*-species of P. Forsskål: identity and typification. Taxon 38:545-556.

Freitag H. and Maier-Stolte M. 1992. A new species and a new combination in the genus *Ephedra* from Arabia. Edinburgh Journal of Botany 49:89-93.

Freitag H. and Maier-Stolte M. 1994. Ephedraceae. In Browicz (Ed) Chorology of Trees and Shrubs in South-West Asia and Adjacent regions 10: 5-38. Warszawa, Polish Scientific Publishers.

Galián J.A., Rosato M., and Rosselló J.A. 2012. Early evolutionary co-localization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba*. Heredity 108:640–646.

Garcia S. and Kovařík A. 2013. Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organization. Heredity 111:23-33.

Gelman A. and Rubin D.B. 1992. "Inference from iterative simulation using multiple sequences". Statistical Science 7:457–511.

Good L., Intine R.V.A. and Nazar R.N. 1997. Interdependence in the processing of ribosomal RNA in *Schizosaccharomyces pombe*. Journal of Molecular Biology 273:782-788.

Grabiele M., Debat H.J., Moscone E.A. and Ducasse D.A. 2012. 25S–18S rDNA IGS of *Capsicum*: molecular structure and comparison. Plant Systematics and Evolution 298:313-321.

Huang J. and Price R.A. 2003. Estimation of the age of extant *Ephedra* using chloroplast *rbcL* sequence data. Molecular Biology and Evolution 20:435-440.

Huang J., Giannasi D.E. and Price R.A. 2005. Phylogenetic relationships in *Ephedra* (Ephedraceae) inferred from chloroplast and nuclear DNA sequences. Molecular Phylogenetics and Evolution 35:48-59.

Huelsenbeck J.P., Larget B. and Alfaro M.E. 2004. Bayesian phylogenetic model selection using reversible jump markov chain monte carlo. Molecular Biology and Evolution 21:1123-1133.
Ickert-Bond S.M. and Wojciechowski M.F. 2004. Phylogenetic relationships in *Ephedra* (Gnetales): Evidence from nuclear and chloroplast DNA sequence data. Systematic Botany 29:834-849.

Ickert-Bond S., Rydin C. and Renner S.S. 2009. A fossil-calibrated relaxed clock for *Ephedra* indicates an Oligocene age for the divergence of Asian and New World clades and Miocene dispersal into South America. Journal of Systematics and Evolution 47:444-456.

Jorgensen R.A. and Cluster P.D. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Annals of the Missouri Botanical Garden 75:1238-1247.

Kubitzki K. 1990. Ephedraceae. In: Kubitzki K., Kramer K.U. and Green P.S. (Eds.) The families and genera of vascular plants. Vol.1, Pteridophytes and Gymnosperms, Berlin, Springer-Verlag, 404 pp.

Li Z-M., Li J. and Li X-W 2006. Polyphyly of the genus *Actinodaphne* (Lauraceae) inferred from the analyses of nrDNA ITS and ETS sequences. Acta Phylotaxonomica Sinica 44:272-285.

Linder C.R., Goertzen L.R., Heuvel B.V., Fransisco-Ortega J. and Jansen R.K. 2000. The complete external transcribed spacer of 18S-26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families 14:285-303.

Loera I., Sosa V. and Ickert-Bond S.M. 2012. Diversification in North American arid lands: niche conservatism, divergence and expansion of habitat explain speciation in the genus *Ephedra*. Molecular Phylogenetics and Evolution 65:437-450.

Mathews S. 2009. Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. American Journal of Botany 96:228-236.

McGuire J.A., Witt C.C., Altshuler D.L. and Remsen J.V. Jr. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Systematic Biology 56:837-856.

Miller M.A., Pfeiffer W. and Schwartz T. 2010. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, pp 1-8.

Mussayev I.F. 1978. On geography and phylogeny of some representatives of the genus *Ephedra* L. Botanicheskij Zhurnal 63:523-543.

Norbäck Ivarsson L. 2014. Pollen morphology in *Ephedra* (Gnetales) and implications for understanding fossil ephedroid pollen from the Tibetan Plateau, using a phylogenetic approach. Degree project in biology, Master of Science, Stockholm University.

Nylander J.A.A. 2004. MrAIC.pl. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Nylander J.A.A.., Wilgenbusch J.C., Warren D.L and Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581-583.

Okuyama Y., Fujii N., Wakabayashi M., Kawakita A., Ito M., Watanabe M., Murakami N. and Kato M. 2005. Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Molecular Biology and Evolution 22:285-296.

Poczai P. and Hyvönen J. 2010. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Molecular Biology Reports 37:1897-1912.

Polanco C., González A. I., de la Fuente A. and Dover G.A. 1998. Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. Genetics 149:243–256.

Posada D. and Buckley R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793-808.

Price R.A. 1996. Systematics of the Gnetales: a review of morphological and molecular evidence. International Journal of Plant Sciences 157:S40-S49.

Rambaut A., Suchard M.A., Xie D. and Drummond A.J. 2013. Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer. Accessed 29 October 2013.

Riedl H. 1993. '*Ephedra major*': zur behandlung der Gattung *Ephedra* bei Clusius und Host. Linzer Biologische Beiträge 25:649-655.

Ronquist F., Huelsenbeck J. and Teslenko M. 2011. Draft MrBayes version 3.2 manual: tutorials and model summaries. http://mrbayes.sourceforge.net/mb3.2_manual.pdf.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. and Huelsenbeck J.P. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Rydin C., Källersjö M. and Friis E. M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems and the monophyly of conifers. International Journal of Plant Sciences 170:1031-1043.

Rydin C., Pedersen K.R. and Friis E.M. 2004. On the evolutionary history of *Ephedra*: Cretaceous fossils and extant molecules. Proceedings of the National Academy of Sciences of the United States of America 101:16571-16576.

Rydin C., Pedersen K.R., Crane P. and Friis E. M. 2006. Former diversity of *Ephedra* (Gnetales): evidence from Early Cretaceous seeds from Portugal and North America. Annals of Botany 98:123-140.

Rydin C. and Korall P. 2009. Evolutionary relationships in *Ephedra* (Gnetales), with implications for seed plant phylogeny. International Journal of Plant Sciences 170:1031-1043.

Rydin C., Khodabandeh A. and Endress P.K. 2010. The female reproductive unit of *Ephedra* (Gnetales): comparative morphology and evolutionary perspectives. Botanical Journal of the Linnean Society 163:387-430.

Rydin C. and Friis E.M. 2010. A new Early Cretaceous relative of Gnetales: *Siphonospermum simplex* gen. et sp. nov. from the Yixian Formation of Northeast China. BMC Evolutionary Biology 10:183-188.

Schmidt M. and Schneider-Poetsch H.A. 2002. The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. Journal of Molecular Evolution 54:715-725.

Shahmuradov I.A., Gammerman A.J., Hancock J.M., Bramley P.M. and Solovyev V.V. 2003. PlantProm: a database of plant promoter sequences. Nucleic Acids Research 31:114-117.

Simpson M.G. 2010. Plant Systematics Second Edition. Oxford, Academic Press Inc, 752 pp.

Staden R. 1996. The Staden sequence analysis package. Molecular Biotechnology 5:233-241.

Stapf O. 1889. Die Arten der Gattung *Ephedra*. Denkschriften der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Klasse 56:1-112.

Steeves M.W. and Barghoorn E.S. 1959. The pollen of *Ephedra*. Journal of the Arnold Arboretum Harvard University 40:221-255.

Swofford D.L. 2003. PAUPNew*. Phylogenetic Analysis Using Parsimony (New*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Talebi K.S., Sajedi T. and Pourhashemi M. 2014. Forests of Iran: A Treasure from the Past, a Hope for the Future. Volume 10. Dordrecht, Springer, 152 pp.

van Nues R.W., Venema J., Rientjes, J.M.J., Dirks-Mulder A. and Raué H.A. 1995. Processing of eukaryotic pre-rRNA:the role of transcribed spacers. Biochemistry and Cell Biology 73:789-801.

Wendland J., Pöhlmann R., Dietrich F., Steiner S., Mohr C. and Philippsen P. 1999. Compact organization of rRNA genes in the filamentous fungus *Asbya gossypii*. Current Genetics 35:618-625.

Appendixes

Appendix 1 Collections from field trip in Dalmatia, Croatia

Voucher	Date	Location	Coordinates	Taxon
Lena Norbäck Ivarsson and Olle Thureborn 101	2/7-2013	Dalmatia, close to Slime, along smaller road. On cliffs.	43°24.521'N 16°53.370'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 102	2/7 – 2013	Dalmatia, close to Slime, along smaller road. On cliffs.	43°24.521'N 16°53.370'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 103	3/7 – 2013	Dalmatia, along D8, in the village Pisak. On stone wall.	43°24.135'N 16°51.376'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 104	3/7 – 2013	Dalmatia, along D8, in Ruskamen, next to exit from camping	43°24.605'N 16°44.682'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 105	4/7 – 2013	Dalmatia, Omiš, on the castle ruins.	43°26.695'N 16°41.595'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 106	4/7 – 2013	Dalmatia, Omiš, on the castle ruins.	43°26.699'N 16°41.589'E	*Ephedra foeminea*, female
Lena Norbäck Ivarsson and Olle Thureborn 107	4/7 – 2013	Dalmatia, Omiš, on the castle ruins	43°26.675'N 16°41.583'E	*Ephedra foeminea*, female
Lena Norbäck Ivarsson and Olle Thureborn 108	4/7 – 2013	Dalmatia, Omiš, on the castle ruins	43°26.673'N 16°41.593'E	*Ephedra foeminea*, female
Lena Norbäck Ivarsson and Olle Thureborn 109	5/7 –2013	Dalmatia, Split, south side of Mont Marjan, along road Šetalište vana Mešrovića.	43°30.505'N 16°23.932'E	*Ephedra* sp. Vegetative
Lena Norbäck Ivarsson and Olle Thureborn 110	5/7 –2013	Dalmatia, Split, on Mont Marjan, along Marangunićevo šetalište.	43°30.520'N 16°23.674'E	*Ephedra foeminea*, female
Lena Norbäck Ivarsson and Olle Thureborn 111	5/7 –2013	Dalmatia, Split, on Mont Marjan, along Marangunićevo šetalište.	43°30.527'N 16°23.715'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 112	5/7 –2013	Dalmatia, Split, on Mont Marjan, along Marangunićevo šetalište.	43°30.527'N 16°23.707'E	*Ephedra* sp. Vegetative
Lena Norbäck Ivarsson and Olle Thureborn 113	5/7 –2013	Dalmatia, Split, on Mont Marjan, along Marangunićevo šetalište.	43°30.539'N 16°23.909'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 114	5/7 –2013	Dalmatia, Split, on Mont Marjan, by the chapel "St Jeronim" (S. Girolamo)	43°30.550'N 16°24.077'E	*Ephedra foeminea*, male
Lena Norbäck Ivarsson and Olle Thureborn 115	5/7 –2013	Dalmatia, Split, on Mont Marjan, by the chapel "St Jeronim" (S. Girolamo)	43°30.551'N 16°24.098'E	*Ephedra major*, vegetative
Lena Norbäck Ivarsson and Olle Thureborn 116	7/7 - 2013	Dalmatia, along D8, in Ruskamen, next to exit from camping	43°24.608'N 16°44.656'E	*Ephedra foeminea*, male

Appendix 2 Included Taxa

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
83	E. alata Decne.	Anderberg 480 (S)	Algeria 1980	AY755698	AY755732	AY755774[a]		AY755805	AY755851	FJ958074	FJ958162		Mediterranean
E325	E. alata Decne.	Cult. (S)	Egypt			New*	New*					New*	Mediterranean
OT08	E. alata Decne.	Mr M. Abd Elaeem (CAI)	Egypt 2011			New*						New*	Mediterranean
81	E. altissima Desf.	Samuelsson 6227 (S)	Algeria-Morocco 1936	AY755696	AY755730	AY755772[a]		AY755803	AY755849	FJ958072	FJ958160	New*	North Africa
82	E. altissima Desf.	Bot. Dept. SU 7688 (S)	Tunisia 1972	AY755697	AY755731	AY755773[a]		AY755804	AY755850	FJ958073	FJ958161	New*	North Africa
AK127	E. americana Humb. et Bonpl. ex Willd.	Novara 8219 (S)	Argentina 1988			GU968545	New*					New*	South America
GB	E. americana Humb. et Bonpl. ex Willd.		Ecuador			AY599143[a]			AY591464			[f] (AK127)	South America
25	E. andina Poepp. ex C.A. Mey.	Chase 10140 (K)	(cult.)	AY755670	AY755707	AY755744[a]		AY755782	AY755821	FJ958045	FJ958128		South America
64	E. antisyphilitica Berl. et C.A. Mey.	Hoggard 451 (S)	Oklahoma, USA 2001	AY755682	AY755715	AY755757[a]		AY755789	AY755834	FJ958057	FJ958145	JX217644[df]	North America
GB	E. antisyphilitica Berl. et C.A. Mey.		Texas	L240918		AY599152[a]		AY492031	AY591452			JX217644[df]	North America
OT18	E. aphylla	Rechinger 12991(W)	Jordan 1957			New*						New*	Mediterranean
80	E. aphylla Forssk.	Anderberg 853 (S)	Libya 1982	AY755695	AY755729	AY755771[a]		AY755802	AY755848	FJ958071	FJ958159	New*	Mediterranean
AK124	E. aphylla Forssk.	Kramer 4727 (Z)	Palestine 1981			GU968544[b]						New*	Mediterranean
AK154	E. aphylla Forssk.	Amdursky 402 (S)	Israel 1950			GU968552[b]						New*	Mediterranean
AK228	E. aphylla Forssk.	Bot. Dep. C-7689 (S)	Tunisia 1972			GU968569[b]						New*	Mediterranean
GB	E. aphylla Forssk.		Sicily, Italy			AY599127[a]							Mediterranean

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
GB	E. aphylla Forssk.		Jordan			AY599128[a]							Mediterranean
OT07	E. aphylla Forssk.	Wafaa Amer (CAI)	Egypt 2012			New*	New*					New*	Mediterranean
GB	E. aspera Engelm. ex S. Watson		Texas, USA			AY599146[a]		AF489532	AY591455			AY492010[cf]	North America
GB	E. boelkei F.A. Roig		Argentina			AY599175[a]			AY591473				South America
GB	E. breana Phil.		Chile			AY599177[a]			AY591472				South America
34	E. californica S. Watson	Stedje 68-154 (O)	(cult.)	AY755676	AY755708	AY755750[a]		AY056569	AY755827	FJ958050	FJ958135	AY492011[cf]	North America
GB	E. californica S. Watson		California	U424927		AY599145[a]		AY492033	AY591458			AY492011[cf]	North America
60	E. chilensis C. Presl.	Odum 7780 (E)	Talca, Chile 1980 (cult.)	AY755679	AY755712	AY755754[a]		AY755786	AY755831	FJ958054	FJ958142	AY492012[cf]	South America
75	E. chilensis C. Presl.	Forbes 49.0542 (UC)	Chile 1949 (cult.)	AY755691	AY755725	AY755767[a]		AY755799	AY755844	FJ958067	FJ958155	AY492012[cf]	South America
94	E. ciliata (E.foliata)	Thulin et Al?Gifri 9975 (UPS)	Arabia 1999	FJ957969	FJ957988	FJ958008[a]		FJ958030	FJ958109	FJ958085	FJ958173	New*	Mediterranean–central Asia
OT13	E. ciliata (E.foliata) Fisch. & C.A. Mey.	Ghazanfar 1914 (BR)	Oman 1992			New*						New*	Mediterranean–central Asia
OT14	E. ciliata Fisch. & C.A. Mey.	Staudinger 6714 (W)	Morocco 1999			New*	New*					New*	Mediterranean–central Asia
86	E. ciliata Fisch. & C.A. Mey.	Rechinger (S)	Afghanistan 1962	AY755700	AY755735	AY755776[a]		AY755807	AY755854	FJ958077	FJ958165	New*	Mediterranean–central Asia
AK146	E. ciliata Fisch. & C.A. Mey.	Balls B2487 C-635 (S)	Morocco 1936			GU968548[b]						New*	Mediterranean–central Asia
OT16	E. ciliata Fisch. & C.A. Mey.	Rechinger 28144 (W)	Pakistan 1965			New*						New*	Mediterranean–central Asia
GB	E. compacta Rose		Mexico			AY599163[a]			AY591474			JX217647[df]	Mexico

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
GB	E. coryi Reed		Texas			AY599153[a]			AY591461			JX217649[df]	North America
GB	E. cutleri Peebles		Arizona			AY599156[a]			AY591456			JX217651[df]	North America
69	E. distachya L.	Rydin 69 (S)	(cult.)	AY755686	AY755719	AY755761[a]		AY755793	AY755838	FJ958061	FJ958149	AY492013[cf]	Asia-Europe
77	E. distachya L.	Rácz et al. 35308 (S)	Hungary 1993	AY755693	AY755727	AY755769[a]		FJ958026	AY755846	FJ958069	FJ958157	AY492013[cf]	Asia-Europe
99	E. distachya L.	Pidopliczka s.n. (UPS)	Ukraine 1925	FJ957973	FJ957991	FJ958013[a]		FJ958033	FJ958114	FJ958089		AY492013[cf]	Asia-Europe
GB1	E. distachya L.		Switzerland			AY599136[a]			AY591480			AY492013[cf]	Asia-Europe
GB2	E. distachya L.		Kazakhstan			AY599135[a]		AY492035	AY591481			AY492013[cf]	Asia-Europe
GB3	E. distachya L.		Turkey			AY599134[a]						AY492013[cf]	Asia-Europe
GB4	E. distachya L.		Syria			AY599133[a]			AY591441			AY492013[cf]	Asia-Europe
4	E. equisetina Bunge	Rydin 16 (S)	Turkmenistan 1938 (cult.)	AY755666a	AY755705	AY755740[a]		AY755781	AY755817	FJ958041	FJ958124		Central-East Asia
35	E. equisetina Bunge	Stedje 90-536 (O)	(cult.)		AY755709	AY755751[a]		AY755783	AY755828	FJ958051	FJ958136	AY492014[cf]	Central-East Asia
71	E. equisetina Bunge	Merello et al. 2241 (MO)	Georgia, USA 1999 (cult.)		AY755721	AY755763[a]		AY755795	AY755840	FJ958063	FJ958151		Central-East Asia
79	E. equisetina Bunge	Norlindh 10285 (S)	Mongolia 1972	AY755694	AY755728	AY755770[a]		AY755801	AY755847	FJ958070	FJ958158	AY492014[cf]	Central-East Asia
GB	E. fasciculata A. Nelson		Arizona, USA			AY599180[a]			AY591457				North America
GB	E. fedtschenkoae Pauls.		(cult.)			AY599158[a]			AY591442				Central Asia
OT12	E. foeminea (E. fragilis ssp campylopoda)	Rydberg 6 (S)	Cult. Pinetum. bl. (origin Italy)			New*						New*	Mediterranean
OT21	E. foeminea (E. fragilis ssp campylopoda)	Johannes Walter 7840 (WU)	Cyprus 1996			New*						New*	Mediterranean

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
OT24	E. foeminea	Billiet & Jadin 8250 (BR)	Greece 2006			New*						New*	Mediterranean
OT11	E. foeminea (E.aphylla)	Rydberg 5 (S)	Cult. Pinetum. bl. (origin Israel)			New*	New*					New*	Mediterranean
37	E. foeminea Forssk.	Tait 618 (E)	Israel 1992 (cult.)	AY755677	AY755710	AY755752[a]	New*	AY755784	AY755829	FJ958052	FJ958138	New*	Mediterranean
93	E. foeminea Forssk.	Osvald 26068 (UPS)	Israel 1952	FJ957968	FJ957987	FJ958007[a]		FJ958029		FJ958084	FJ958172	New*	Mediterranean
AK130	E. foeminea Forssk.	Rydin 130 (Z)	Greece cult			GU968546[b]	New*					New*	Mediterranean
AK152	E. foeminea Forssk.	Fries C-7619 (S)	Dalmatia 1938			GU968551[b]	New*					New*	Mediterranean
E317	E. foeminea Forssk.	Bolinder (S)	Greece, Asprovalta			New*	New*					New*	Mediterranean
GB	E. foeminea Forssk.		Greece			AY599131[a]			AY591478				Mediterranean
OT10	E. foeminea Forssk.	Rydin 332 (S)	Greece, Pilos			New*	New*					New*	Mediterranean
OT26	E. foeminea Forssk.	Norbäck Ivarsson & Thureborn 105 (S)	Croatia, Dalmatia			New*						New*	Mediterranean
87	E. foeminea Forssk. (syn. E. campylopoda C.A. Mey.)	Andersen et al. 2098 (S)	Turkey 1971	AY755701	AY755736	AY755777[a]		AY755808	AY755855	FJ958078	FJ958166	New*	Mediterranean
84	E. foliata Boiss.et C.A. Mey.	Täckholm s.n. (S)	Egypt 1929		AY755733		New*		AY755852	FJ958075	FJ958163		Mediterranean–central Asia
85	E. foliata Boiss.et C.A. Mey.	Riedl 3979 (S)	Iran 1960	AY755699	AY755734	AY755775[a]		AY755806	AY755853	FJ958076	FJ958164	New*	Mediterranean–central Asia
95	E. foliata Boiss.et C.A. Mey.	Hedberg et Hedberg 92019A (UPS)	Saudi Arabia 1992	FJ957970	FJ957989	FJ958009[a]		FJ958031	FJ958110	FJ958086	FJ958174	New*	Mediterranean–central Asia
96	E. foliata Boiss.et C.A. Mey.	Thulin 10745 (UPS)	Somalia 2002	FJ957971	FJ957990	FJ958010[a]		FJ958032	FJ958111	FJ958087	FJ958175	New*	Mediterranean–central Asia
101	E. fragilis Desf.	Jonsell 5412 (UPS)	Morocco 1989	FJ957974	FJ957992	FJ958014[a]	New*	FJ958034	FJ958115	FJ958090	FJ958176	New*	Mediterranean

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
109	E. fragilis Desf.	Denk 179?01 (S)	Morocco 2005	FJ957979	FJ957997	FJ958019[a]	New*	FJ958037	FJ958119	FJ958095	FJ958181	New*	Mediterranean
GB1	E. fragilis Desf.		Jordan			AY599129[a]		AF489533	AY591440				Mediterranean
GB2	E. fragilis Desf.		Europe			AY599130[a]			AY591477				Mediterranean
OT09	E. fragilis Desf.	Denk & Gruber 178-02 (S)	Morocco 2005			New*	New*					New*	Mediterranean
8	E. frustillata Miers	Rydin 08 (S)	Patagonia, Chile 1994 (cult.)	AY755669a	AY755706	AY755743[a]		AY056570	AY755820	FJ958044	FJ958127		South America
30	E. frustillata Miers	Chase 10218 (K)	(cult.)	AY755674a	AY056490	AY755748[a]		AY056564	AY755825	FJ958048	FJ958131		South America
GB	E. funerea Coville et Morton		Arizona, USA			AY599168[a]			AY591454			JX217652[df]	North America
26	E. gerardiana Wall. ex Florin	Chase 10141 (K)	(cult.)	AY755671a	AY056486	AY755745[a]		AY056560	AY755822	FJ958046	FJ958129	EF053133[ef]	Central Asia
59	E. gerardiana Wall. ex Florin	MacPherson s.n. (E)	Sikkim, India 1983 (cult.)	AY755678	AY755711	AY755753[a]		AY755785	AY755830	FJ958053	FJ958141	EF053133[ef]	Central Asia
68	E. gerardiana Wall. ex Florin	Stedje 74-460 (O)	(cult.)	AY755685	AY755718	AY755760[a]		AY755792	AY755837	FJ958060	FJ958148	EF053133[ef]	Central Asia
GB	E. gracilis Phil.		Chile			AY599150[a]			AY591465				South America
6	E. intermedia Schrenk et C.A. Mey.	Rydin 64a (S)	Tien?Shan 1971 (cult.)	AY755667a	AY056492	AY755741[a]		AY056566	AY755818	FJ958042	FJ958125	AY492016[cf]	Central-East Asia
66	E. intermedia Schrenk et C.A. Mey.	Rydin 66 (S)	Tien?Shan 1971 (cult.)	AY755683	AY755716	AY755758[a]		AY755790	AY755835	FJ958058	FJ958146	AY492016[cf]	Central-East Asia
GB	E. intermedia Schrenk et C.A. Mey.		Kazakhstan			AY599179[a]						AY492016[cf]	Central-East Asia
110	E. laristanica Assadi	Davis et Bokhari D.56211B (E)	Iran 1974	FJ957980	FJ957998	FJ958020[a]			FJ958120	FJ958096	FJ958182		Iran
GB	E. laristanica Assadi		Iran			AY599126[a]			AY591437				Iran

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
2	E. likiangensis Florin	Rydin 03-926 (S)	(cult.)	AY755665a	AY056485	AY755739 [a]		AY755780	AY755816	FJ958040	FJ958123		China
74	E. likiangensis Florin	Forbes 94.0389 (UC)	Yunnan, China 1994 (cult.)	AY755690	AY755724	AY755766 [a]		AY755798	AY755843	FJ958066	FJ958154		China
92	E. lomatolepis Schrenk et C.A. Mey	Baitulin et al. s.n. (UPS)	Kazakhstan 1997	FJ957967	FJ957986	FJ958006 [a]		FJ958028	FJ958108	FJ958083	FJ958171	AY492017 [cf]	Central-East Asia
102	E. lomatolepis Schrenk et C.A. Mey	Titow 488 (S)	Russia 1914	FJ957975	FJ957993	FJ958015 [a]			FJ958116	FJ958091	FJ958177	AY492017 [cf]	Central-East Asia
27	E. lomatolepis? Schrenk et C.A. Mey	Chase 10142 (K)	(cult.)	AY755672	AY056487	AY755746 [a]		AY056561	AY755823			AY492017 [cf]	Central-East Asia
5	E. major Host	Rydin 37 (S)	Yugoslavia 1955 (cult.)	FJ957965a	FJ957984	FJ958003 [a]	New*	FJ958025	FJ958107			New*	Mediterranean–central Asia
88	E. major Host	Debreczy et al. 46206 (S)	Andalusia, Spain 1995	AY755702	AY755737	AY755778 [a]		AY755809	AY755856	FJ958079	FJ958167	New*	Mediterranean–central Asia
97	E. major Host	E. Behr 1936 (UPS)	Macedonia 1936	FJ957972		FJ958011 [a]			FJ958112			New*	Mediterranean-central Asia
98	E. major Host	Julin 1979 (UPS)	Alicante, Spain 1979			FJ958012 [a]			FJ958113	FJ958088			Mediterranean–central Asia
103	E. major Host	Uggla s.n. (S)	Algeria 1936	FJ957976	FJ957994	FJ958016 [a]		FJ958035	FJ958117	FJ958092	FJ958178	New*	Mediterranean–central Asia
AK163	E. major Host	Montserrat 319171 (Z)	Spain 1971			GU968554 [b]						New*	Mediterranean–central Asia
AK164	E. major Host	Grossheim (Z)	Transcaucasia 1924			GU968555 [b]						New*	Mediterranean–central Asia
AK165	E. major Host	Baenitz (Z)	Herzegovina 1898			GU968556 [b]	New*					New*	Mediterranean–central Asia
AK166	E. major Host	Hofmann 013-1971 (Z)	Algeria 1971			GU968557 [b]						New*	Mediterranean–central Asia
AK169	E. major Host	Zogg & Gassner 8388 (Z)	France			GU968559 [b]	New*					New*	Mediterranean–central Asia
AK193	E. major Host	Kurbanov 630 (MO)	Turkmenistan 2001			GU968565 [b]						New*	Mediterranean–central Asia

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
OT19	E. major Host	Aldasoro et al 2570 (BR)	Turkey 2001			New*						New*	Mediterranean–central Asia
OT28	E. major Host	Norbäck Ivarsson & Thureborn 115 (S)	Croatia, Dalmatia			New*	New*					New*	Mediterranean–central Asia
116	E. milleri Freitag et Maier-Stolte	Miller 7667 (E)	Oman 1985	FJ957983	FJ958002	FJ958024[a]			FJ958121	FJ958100	FJ958186		Oman
7	E. minuta Florin	Rydin 2 (S)	Sikang, China 1934 (cult.)	AY755668	AY056493	AY755742[a]		AY056567	AY755819	FJ958043	FJ958126	AY492019[cf]	China
61	E. minuta Florin	Rydin 61 (S)	Sikang, China 1934 (cult.)	AY755680	AY755713	AY755755[a]		AY755787	AY755832	FJ958055	FJ958143	AY492020[cf]	China
63	E. minuta Florin	Rydin 63 (S)	(cult.)	AY755681	AY755714	AY755756[a]		AY755788	AY755833	FJ958056	FJ958144	AY492021[cf]	China
E392	E. minuta Florin	Cult. (S)				New*	New*					New*	China
GB	E. monosperma J.G. Gmel. ex C.A. Mey.		Mongolia			AY599139[a]			AY591443				Central-East Asia
GB	E. multiflora Phil. ex Stapf		Chile			AY599173[a]			AY591471				South America
72	E. nevadensis S. Watson	Forbes 66.1033 (UC)	California, USA 1966 (cult.)	AY755688	AY755722	AY755764[a]		AY755796	AY755841	FJ958064	FJ958152	AY492020[cf]	North America
GB	E. ochreata Miers		Argentina			AY599176[a]			AY591463				South America
89	E. pachyclada Boiss.	Danin S-2455 (S)	Sinai, Egypt 1974	AY755703	AY755738	AY755779[a]		AY755810	AY755857	FJ958080	FJ958168		West Asia
GB	E. pachyclada Boiss.		Sinai, Egypt			AY599138[a]							West Asia
GB	E. pedunculata Engelm. ex S. Watson		Texas, USA			AY599144[a]		AY492043	AY591460			AY492021[cf]	Mexico, Texas
GB	E. przewalskii (Stapf) Andrz.		Mongolia			AY394072[a]		AY492044	AY591484			AY492022[cf]	Central-East Asia
GB	E. regeliana Florin		China			AY599160[a]			AY591449				Central-East Asia

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
GB1	E. rhytidosperma Pachom.		China	DQ212959		DQ212958[a]		DQ212957				DQ212960[ef]	China
GB2	E. rhytidosperma Pachom.		China	DQ028781		DQ028782[a]		DQ028779				DQ212960[ef]	China
73	E. rupestris Benth.	Ornduff 9675 (UC)	Ecuador 1987 (cult.)	AY755689	AY755723	AY755765[a]		AY755797	AY755842	FJ958065	FJ958153		South America
105	E. sarcocarpa Aitch. et Hemsl.	Aellen et Estafandri 2786 (S)	Iran 1963	FJ957977	FJ957995	FJ958017[a]			FJ958118	FJ958093	FJ958179	AY492023[cf]	Central Asia
GB	E. sarcocarpa Aitch. et Hemsl.		Iran			AY599137[a]		AY492045				AY492023[cf]	Central Asia
114	E. saxatilis Florin	Hedin C-218 (S)	Tibet 1909	FJ957981	FJ958000	FJ958022[a]		FJ958038		FJ958098	FJ958184		Central Asia
GB	E. saxatilis Florin		Tibet	U424937		AY599140[a]			AY591445				Central Asia
28	E. sinica Stapf	Chase 10143 (K)	(cult.)	AY755673	AY056488	AY755747[a]		AY056562	AY755824	FJ958047	FJ958130	AY492024[cf]	East Asia
33	E. sinica Stapf	Schönenberger s.n. (S)	Hebei, China 2000	AY755675	AY056491	AY755749[a]		AY056565	AY755826	FJ958049	FJ958134	AY492024[cf]	East Asia
90	E. somalensis Freitag et Maier-Stolte	Thulin 10925A (UPS)	Somalia 2002	FJ957966	FJ957985	FJ958004[a]		FJ958027		FJ958081	FJ958169	New*	Horn of Africa
GB	E. somalensis Freitag et Maier-Stolte		Somalia			AY599141[a]			AY591444				Horn of Africa
115	E. sp. L.	Achyutananda Borna s.n. (S)	Jaunsar, Deoban, India 1923	FJ957982	FJ958001	FJ958023[a]		FJ958039		FJ958099	FJ958185		
107	E. strobilacea Bunge	Aellen et Esfandari 2703 (S)	Iran 1960	FJ957978	FJ957996	FJ958018[a]		FJ958036		FJ958094	FJ958180		Central Asia
GB	E. strobilacea Bunge		Iran			AY599162[a]			AY591448				Central Asia
67	E. torreyana S. Watson	Eriksson 04?487 (S)	New Mexico, USA 2001	AY755684	AY755717	AY755759[a]		AY755791	AY755836	FJ958059	FJ958147	New*	North America

ID	Taxon	Vouchers	Collection sites	18S[a]	26S[a]	nrITS	ETS	rbcL[a]	rps4[a]	rpL16[a]	trnS-trnfM[a]	matK	Distribution
GB	E. torreyana S. Watson		Arizona, USA	U424147		AY599166[a]		AY492047	AY591453			[f] (67)	North America
111	E. transitoria Riedl	Collenette 9095 B (E)	Saudi Arabia 1994		FJ957999	FJ958021[a]				FJ958097	FJ958183		West Asia
GB	E. transitoria Riedl		Jordan			AY599172[a]			AY591450				West Asia
GB	E. triandra Tul.		Argentina			AY599165[a]			AY591468				South America
70	E. trifurca Torr.	Miller et Harder 8187 (MO)	Arizona 1994 (cult.)	AY755687	AY755720	AY755762[a]		AY755794	AY755839	FJ958062	FJ958150	AY492026[cf]	North America
76	E. tweediana Fisch. C.A. Mey.	Forbes 66.0742 (UC)	Argentina (cult.)	AY755692	AY755726	AY755768[a]		AY755800	AY755845	FJ958068	FJ958156	AY492027[cf]	South America
91	E. viridis Coville	Holmgren et al. 1826 (UPS)	Utah, USA 1965			FJ958005[a]				FJ958082	FJ958170	AY492028[cf]	North America

* New sequences produced for this study.

[a] Sequences used in previous study, Rydin and Korall (2009).

[b] Sequences used in previous study, Rydin et al. (2010).

[c] Sequences used in previous study, Huang et al. (2005).

[d] Sequences used in previous study, Loera et al. (2012)

[e] Sequences from Wang et al. (2005) directly submitted on GenBank.

[f] Indicates that that gene is sampled on species level, i.e. the gene is sampled from the same species but not from the same specimen. When applicable, numbers within brackets indicates which specimen the sequence is sampled from.