ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2004) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Phylogenetic relationships and biogeography of the desert plant genus *Fagonia* (Zygophyllaceae), inferred by parsimony and Bayesian model averaging

B.-A. Beier [a,*], J.A.A. Nylander [b], M.W. Chase [c], M. Thulin [a]

[a] *Department of Systematic Botany, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden*
[b] *Department of Systematic Zoology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden*
[c] *Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK*

Received 14 October 2003; revised 5 May 2004

Abstract

Phylogenetic relationships within *Fagonia* were inferred from analyses of plastid *trnL* intron and nuclear ribosomal ITS DNA sequences. Sampling of the genus was nearly complete, including 32 of 34 species. Phylogenetic analysis was carried out using parsimony, and Bayesian model averaging. The latter method allows model-based inference while accounting for model-selection uncertainty, and is here used for the first time in phylogenetic analyses. All species of *Fagonia* in the Old World, except *F. cretica*, form a weakly supported clade, and all *Fagonia* species of the New World, except *F. scoparia*, are well supported as sister to the Old World clade. *Fagonia scoparia*, from Mexico, and *F. cretica*, from Northern Africa, are well supported as sisters to all other *Fagonia* species. Vicariance–dispersal analysis, using DIVA, indicated that the occurrences of *Fagonia* in South America and southern Africa are due to dispersals, and also, that the ancestor of *Fagonia* had a distribution compatible with the boreotropics hypothesis.

Keywords: Bayesian model averaging; Biogeography; Boreotropics; Arid regions; *Fagonia*; Parsimony; Phylogeny; Zygophyllaceae

1. Introduction

The genus *Fagonia* is a member of Zygophyllaceae, which is included in the eurosid I clade (APG II, 2003). In a phylogenetic analysis of Zygophylloideae, based on molecular and morphological data, *Fagonia* was shown to have a well-supported position as sister to the genus *Melocarpum*, endemic to the Horn of Africa region (Beier et al., 2003). *Fagonia* consists of shrubs with free spinescent or pointed stipules, pink or purple petals, and an obconical, more or less pubescent, loculicidal capsule, often with persisting sepals. Most species of *Fagonia* have three-foliolate leaves, but there are also several species that are one-foliolate. The free stipules and the pubescent, obconical capsules are particularly important in the circumscription of *Fagonia*.

The delimitation of species in *Fagonia* is known for being notoriously difficult. This is caused by the great variation in most morphological characters in many of the species. The first complete modern treatment has been finished only recently (Beier, in press). Following this revision, *Fagonia* is a genus of 34 species confined to warm and arid areas of all continents except Australia, with concentrations of species in the Horn of Africa region and Baja California.

1.1. Geographical distribution

Fagonia has a remarkably disjunct distribution covering arid areas of the New and Old World (Fig. 1). It is found in Mexico, southwestern USA, Chile, and Peru

* Corresponding author.
E-mail address: Bjorn-Axel.Beier@ebc.uu.se (B.-A. Beier).

1055-7903/$ - see front matter © 2004 Published by Elsevier Inc.
doi:10.1016/j.ympev.2004.05.010

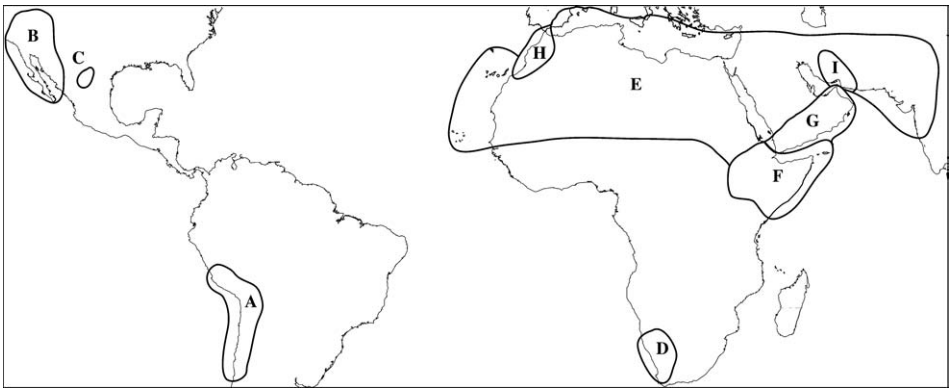

Fig. 1. Distribution map of *Fagonia* with the nine areas of endemism: A, South America; B, North America; C, Northeastern Mexico; D, Southern Africa; E, Saharo-Sind including parts of Macaronesia; F, Horn of Africa region; G, Yemen and Oman; H, Western Morocco; and I, Southern Iran.

in the New World, whereas in the Old World it is known from southern Africa, parts of Macaronesia, North Africa south to the Sahel regional transition zone, and the southernmost parts of Europe, including many of the Mediterranean islands. *Fagonia* is also known from the Horn of Africa region including Kenya, western Asia including Lebanon, Turkey, and the Arabian Peninsula east to Afghanistan and western India. Species of *Fagonia* are restricted to distinct regions in the Old or New World. A single species in South America is endemic to Chile and Peru, and the two species of southern Africa are confined to Botswana, Namibia, and South Africa. Areas rich in endemic species of *Fagonia* are Baja California with six species, and the Horn of Africa region (including southern Yemen, and Oman) with eight species. Neither of the two species in Macaronesia is endemic there, and the genus has not been reported from Madagascar.

The disjunct distribution of *Fagonia* has been explained by several different hypotheses. Engler (1896, 1915) proposed that the New World populations of *Fagonia* had been established by seeds of *F. cretica* introduced by cargo ships from the Iberian Peninsula and that *Fagonia* was a truly Old World genus. However, Engler (1931) later considered the distribution as prehistoric and not a result of anthropogenic dispersal. Johnston (1940) explained the distribution of *Fagonia* as remnants of a widespread early Tertiary desert flora. Axelrod (1950) initially considered the distribution as a result of a break-up of the range of a common pantropical Tertiary ancestor and subsequent speciation. Later, Axelrod (1970) considered the distribution as being remnant of the dry flora that inhabited Gondwanaland prior to its break-up during the Cretaceous, whereas Stebbins and Day (1967) considered that the distribution was a result of a pre-Tertiary migration from the Old to the New World via the Bering land bridge. Porter (1974) proposed a combination of seafloor spreading and long-distance dispersal from east to west during the early Tertiary when the continents of the Old and New Worlds were much closer than at present.

1.2. Systematics

Phylogenetic relationships in *Fagonia* have been the subject of only a few studies. Ozenda and Quézel (1957) grouped the North African *Fagonia* species into four "natural groups": (1) the *F. kahirina–cretica–flamandii* group, (2) the *F. arabica–bruguieri* group, (3) the *F. glutinosa–latifolia* group, and (4) the *F. microphylla*-group. El Hadidi (1966), based on Ozenda and Quézel (1957), instead divided the North African *Fagonia* species into three groups. The relationships of the groups were only commented on briefly. El Hadidi (1973) later modified the "natural groups" and divided them into species "with tri- or unifoliolate leaves" or "with simple leaves." The group with "tri- or unifoliolate leaves" was in turn divided into three "complexes": (1) the "*F. arabica*-complex," (2) the "*F. bruguieri*-complex," and (3) the "*F. indica*-complex." The two last complexes contained many taxa now no longer recognized (Beier, in press). Subsequently, El Hadidi (1974) made changes in the division of the complexes and divided them into: (1) the "*F. isotricha*-complex" (*F. isotricha* is now a synonym of *F. latifolia*), (2) the "*F. glutinosa*-complex," (3) the "*F. sinaica*-complex" (*F. sinaica* is now a synonym of *F. scabra*), (4) the "*F. thebaica*-complex" (*F. thebaica* is now a synonym of *F. arabica*). The species of the New World, *F. californica* and *F. chilensis* (as *F. subaphylla*), were hypothesized to be associated with the *F. sinaica*-complex. El Hadidi (1974) also recognized that the three groups of *Fagonia* species with short stipules and trifoliolate leaves were more closely related to each other than to the group of species with long stipules and one- to trifoliolate leaves.

To summarize, the "natural groups" presented by Ozenda and Quézel (1957) and El Hadidi (1966, 1973, 1974) were based on vague criteria and included only the Old World species. Also, the taxonomy used at spe-

cies level differed substantially from the taxonomy presented by Beier (in press).

1.3. Bayesian model averaging

Maximum parsimony (MP) has been a successfully used method for phylogeny reconstruction for several decades, and it has been an important part of the renaissance of systematics. Nevertheless, plant systematics is currently in a period of re-evaluation of the methodology employed to estimate phylogeny (Archibald et al., 2003). Limitations of the phylogenetic methods are the driving forces behind the development of new methods. MP has some well-known limitations, which will result in possible inconsistency. For example, long-branch attraction, that is long branches and unequal branch lengths predispose it towards inconsistency (Felsenstein, 2004). One of the recently introduced approaches for coming to terms with limitations experienced with MP is Bayesian inference (BI). It differs from MP in being a model-based method including prior beliefs. It possesses advantages over other model-based methods, such as maximum likelihood (ML), in terms of ability to use complex models of evolution and computational efficiency (Huelsenbeck et al., 2001; Nylander et al., 2004). However, negative attributes that have to be considered when using a model-based inference method are, for example, the possible effects of using an inappropriate model in the analysis. All inference under both BI and ML is conditional on a (single) pre-specified model of evolution—a character substitution model. A common observation is that results may vary with different substitution models, and when results differ, it may be difficult to tell which result to trust. More importantly, it has been shown that the failure to use an appropriate model can lead to bias and incorrect conclusions (Gaut and Lewis, 1995; Sullivan and Swofford, 1997). To come to terms with this problem, statistical methods for estimating an appropriate model have been applied in phylogenetics. The most commonly used approaches have been the use of likelihood-ratio testing (LRT; Felsenstein, 1981; Posada and Crandall, 1998), or the use of model-selection criteria such as the Akaike information criterion (AIC, Akaike, 1973) and the Bayesian information criterion (BIC, Schwartz, 1978). However popular these methods have become, they are not without problems. For example, only nested models can be compared using LRT, and the correct use of AIC is dependent on finding the ML estimate under each model used (Burnham and Anderson, 2002), which is computationally demanding and therefore seldom applied. Furthermore, Buckley et al. (2002) have pointed out that there can be substantial selection uncertainty involved when choosing a model for phylogenetic inference. Model selection is the process of using the data to select one model from a set of models. Model selection uncertainty indicates that there is an alternative model, or models, that fit the data nearly as well as the one selected. Importantly, if there is no model that is superior to the others in the set of candidate models, it might be hazardous to base all inference on a single model, especially if results differ among models. Furthermore, ignoring model-selection uncertainty often leads to inferior inferences and overstatements of accuracy (Burnham and Anderson, 2002).

Although systematists have described how to measure model-selection uncertainty, few have taken it into account in the actual inference of phylogeny. One way to account for model-selection uncertainty in phylogenetic inference is to use model averaging. Model averaging is the process of estimating some parameter for each model in a set and then averaging the estimates according to how likely each model is (Wasserman, 2000). Bayesian inference and especially the technique of Markov chain Monte Carlo (MCMC) allow us to do this in a sophisticated manner (Carlin and Chib, 1995; Green, 1995). However, some applications of Bayesian model averaging (BMA) are difficult to implement. Here, we demonstrate an approximate application of BMA, which is easy to implement using existing software and in which phylogenetic inference is made conditional on a set of models, each weighted according to its posterior probability (PP).

1.4. Aims of study

The main aims of this study are to: (1) present a hypothesis of the phylogenetic relationships within *Fagonia*, (2) gain a better understanding of the historical biogeography of *Fagonia*, and (3) assess the method of Bayesian model averaging as a method of phylogenetic analyses (Nylander, 2004).

2. Materials and methods

2.1. Data preparation

Plastid *trnL* intron sequences and nuclear ribosomal ITS sequences, including the 5.8S gene, for 44 specimens were used in our study (Table 1). All species of *Fagonia* and *Melocarpum* as well as *Zygophyllum xanthoxylum* and *Roepera billardieri* analyzed in the study by Beier et al. (2003) were included here. The choice of outgroup is based on Beier et al. (2003). In their study of the subfamily Zygophylloideae, they found *Melocarpum* to be well supported as sister to *Fagonia*. Representatives of two more genera of Zygophylloideae, *Roepera* and *Zygophyllum*, are also included in the outgroup of this study. The sampling of *Fagonia* is nearly complete, and only two species, *F. densispina* (newly described) and *F. californica* (morphologically similar and presum-

Table 1
Species used in this analysis, together with information on voucher, geographical origin of plant material, and database accession numbers

Species	Voucher information	ITS Database Accession No.	*trnL* Database Accession No.
Fagonia acerosa Boiss.	Davis 56261 (E) Iran	AY641617	AY641579
F. arabica L.	Leonard 4887 (S) Libya	AY641618	AY641580
F. bruguieri DC. (a)	Thulin et al. 9986 (UPS) United Arab Emirates	AY641619	AY641582
F. bruguieri DC. (b)	Hedge et al. 7663 (E) Afghanistan	AY641620	AY641581
F. charoides Chiov.	Thulin et al. 10587 (UPS) Somalia	AY641621	AY641583
F. chilensis Hook. & Arn.	Penailillo s.n. (UTALCA) Chile	AY641622	AY641584
F. cretica L. (a)	Beier 125 (UPS) Canary Islands	AY641623	AY641585
F. cretica L. (b)	Kilian 2937 (B) Cape Verde Islands	AY641624	AY641586
F. densa I.M.Johst.	Rebman 3171 (SD) Mexico	AY641625	AY641587
F. glutinosa Delile	Davis 49654 (K) Libya	AY641627	AY641588
F. gypsophila Beier & Thulin	Thulin et al. 9473 (UPS) Somalia	AY641626	AY641589
F. hadramautica Beier & Thulin	Thulin et al. 9808 (UPS) Yemen	AY641628	AY641590
F. harpago Emb. & Maire	Podlech 40630 (RSA) Morocco	AY641629	AY641591
F. indica Burm.f. (a)	Thulin et al. 9835 (UPS) Yemen	AY641630	AY300768
F. indica Burm.f. (b)	Thulin et al. 10024 (UPS) United Arab Emirates	AY641631	AY641592
F. indica Burm.f. (c)	Baxter et al. 4202 (E) Jordan	AY641632	AY641593
F. laevis Standl. (a)	Beier 97 (UPS) USA	AY641633	AY641594
F. laevis Standl. (b)	Beier 95 (UPS) USA	AY641634	AY641595
F. lahovarii Volkens & Schweinf.	Thulin et al. 9522 (UPS) Yemen	AY641635	AY641596
F. latifolia Delile	Scholz 174 (B) Chad	AY641626	AY641597
F. latistipulata Beier & Thulin	Thulin et al. 10833 (UPS) Somalia	AY641636	AY641598
F. longispina Batt.	Podlech 53369 (M) Morocco	AY641637	AY641599
F. luntii Baker	Thulin et al. 9881 (UPS) Yemen	AY641638	AY300769
F. mahrana Beier	Thulin et al. 9682 (UPS) Yemen	AY641639	AY641600
F. minutistipula Engl.	Giess and Müller 13952 (K) Namibia	AY641641	AY300770
F. mollis Delile	Townsend 86/12 (K) Jordania	AY641643	AY641600
F. olivieri DC.	Samuelsson 4357 (S) Syria	AY641646	AY641602
F. orientalis J.Presl & C.Presl	Collenette 7516 (E) Saudi Arabia	AY641648	AY641603
F. pachyacantha Rydb. (a)	Beier 93 (UPS) USA	AY641649	AY641604
F. pachyacantha Rydb. (b)	Beier 103 (UPS) Mexico	AY641651	AY300771
F. palmeri Vasey & Rose	Hastings 75 (SD) Mexico	AY641653	AY641605
F. paulayana Wagner & Vierh. (a)	Thulin et al. 9515 (UPS) Yemen	AY641654	AY641606
F. paulayana Wagner & Vierh. (b)	Thulin et al. 9507 (UPS) Yemen	AY641652	AY641607
F. paulayana Wagner & Vierh. (c)	Ghafoor and Goodman 4413 (E) Pakistan	AY641650	AY641608
F. rangei Engl.	Leistner 3388 (K) South Africa	AY641647	AY641609
F. scabra Forssk.	Davis 49662 (E) Libya	AY641645	AY300767
F. scoparia Brandegee	Johnston 9461(SD) Mexico	AY641644	AY300772
F. subinermis Boiss.	Grey-Wilson and Hewer 285 (W) Iran	AY641642	AY641610
F. villosa P.D.Porter	K. 5915 (RSA) Mexico	AY641640	AY641611
F. zilloides Humbert	Davis 49047 (E) Morocco	AY641655	AY641612
Melocarpum hildebrandtii (Engl.) Beier & Thulin	Thulin et al. 9012 (UPS) Somalia	AY641615	AJ387971
M. robecchii (Engl.) Beier & Thulin	Thulin et al. 9537 (UPS) Yemen	AY641616	AY300773
Roepera billardieri (DC.) G.Don	R. 417 (Adelaide B.G.) Australia	AY641613	AJ387969
Zygophyllum xanthoxylum (Bunge) Engl.	Chase 1700 (K) China	AY641614	AJ387975

ably closely related to *F. laevis*), are missing from the analysis. The following widely distributed and variable species are sampled with more than one specimen in our analysis: *F. bruguieri*, *F. cretica*, *F. indica*, *F. laevis*, *F. pachyacantha*, and *F. paulayana*. These species were sampled with several different specimens to represent most of their morphological variation as well as the geographical distribution. DNA was extracted from 33 specimens of *Fagonia* using a slightly modified version of the 2XCTAB method of Doyle and Doyle (1987). The DNA from silica-dried material was precipitated with ethanol, and the DNA from herbarium material was precipitated in isopropanol, following the recommendations of Fay et al. (1998). DNA was purified using the QIAquick purification kit (Qiagen) following the manufacturer's protocol.

The *trnL* intron region was amplified and sequenced using the forward primer c and reverse primer d (Taberlet et al., 1991). Larger fragments including the *trnL-F* spacer (e and f primers of Taberlet et al., 1991), could be amplified, but these proved to be impossible to sequence for most taxa due to several long homopolymer regions (A or T) causing *Taq* to make many mistakes and drop bases, leading to electropherograms that were difficult to edit with confidence (Beier et al., 2003). For amplification and sequencing of the ITS region, the following primers

were used in different combinations: ITS4 (White et al., 1990), P17 and P26S-82R (Popp and Oxelman, 2001), P16 and P25 (Oxelman and Lidén, 1995), and 175E and 265E (Sun et al., 1994). PCRs were made in final volumes of 50 µl, containing 1–4 µl DNA, 5 µl of 10× Taq-buffer (500 mM KCl, 100 mM Tris–HCl, pH 9.0), 5 µl $MgCl_2$, 2×1 µl of 12.5 pmol primer, 2–4 µl dNTPs (10 mM), 0.25 U *Taq* polymerase, 0.5 µl of 20 mg/ml BSA, and H_2O up to 50 µl. The PCR protocol for *trnL* was 2 min at 94 °C, then 28 cycles at 94 °C for 1 min, 50 °C for 30 s, 72 °C for 1 min, and ending with one cycle of 72 °C for 6 min. Amplification of ITS was achieved by denaturation for 60 or 90 s at 97 °, then 25 or 34 cycles with 1 min or 20 s at 97 °C and 1 min at 50 or 55 °C and 3 min or 90 s at 72 °C, and finally 6 or 10 min at 72 °C.

PCR products were purified using the QIAquick purification kit (Qiagen) or the MultiScreen PCR-plates (Millipore), according to the manufacturer's protocols. Cycle sequencing was done using the Dye Terminator Cycle Sequencing kit version 1.0 (Applied Biosystem) or with the DYEnamic ET terminator Cycle Sequencing Kit (Amersham Biosciences). Manufacturer's protocols were followed with the exception of the dye concentration, which was reduced to 1 µl in a reaction volume of 10 µl buffer for the Dye Terminator Cycle Sequencing kit. Reactions were run on an ABI 377 automated sequencer or a MegaBACE 1000 DNA Analysis system.

Multiple sequence alignment was performed manually following the guidelines of Kelcher (2000). Underlying patterns, usually repeating units, were taken into account when aligning. The aligned matrix is available upon request from B-AB and MWC.

2.2. Parsimony analysis

The MP analyses were completed using the parsimony options of PAUP* 4.0b10 (Swofford, 2002). Characters were given equal weight and treated as unordered (Fitch parsimony; Fitch, 1971). The ITS and the *trnL* data sets were analyzed separately and in combination. A heuristic search with 10,000 replicates of random taxon additions was run with TBR branch swapping, holding five trees at each step during stepwise addition, and saving only one tree per replicate (MULTREES off) to reduce time spent swapping on large islands of trees. The combination of the ITS and *trnL* data sets was considered justified since no incongruence, as defined below, was observed between the two data sets. We assessed the congruence of the separate data sets manually by checking the bootstrap percentage (BP) of the clades. The trees were considered incongruent if they displayed high bootstrap support, above 80%, for any of the incongruent clades (Seelanan et al., 1997; Wiens, 1998).

Clade support was calculated with 10,000 bootstrap replicates in PAUP* 4.0b10 (Swofford, 2002) using TBR branch swapping with simple addition and MULTREES off. In this study we arbitrarily considered clades with bootstrap percentage from 50 up to 70% as weakly supported, from 71 up to 80% as moderately supported, and above 80% as well supported. Branch lengths used DELTRAN optimization.

2.3. Bayesian model averaging

Phylogenetic inference under approximate BMA was accomplished as follows: first, a set of nucleotide substitution models was put forward as candidate models; the HKY model (Hasegawa et al., 1985), and the GTR model (Lanave et al., 1984; Rodríguez et al., 1990; Tavaré, 1986), each combined with two different ways to model rate variation, the gamma distribution (Γ) (Yang, 1994), or the gamma distribution combined with the assumption that a proportion of sites were invariable (Γ + I) (Gu et al., 1995). These models were chosen on the basis that they have been found to sufficiently describe empirical data (Hasegawa et al., 1985; Yang, 1994), but also from a practical point of view; they are all incorporated in the software used to estimate posterior distributions (see below). The different substitution models were applied to the two gene fragments (partitions) individually giving a total of 6 by 6 = 36 joint models. Then, a neighbor-joining tree for the combined data set was calculated using Jukes-Cantor distances in PAUP. Each data set was then evaluated separately on this topology using the maximum likelihood criterion (estimating all parameters), and the log likelihoods for the different partitions were summed to give the overall likelihood for the 36 models. Posterior probabilities for the models were estimated using Akaike weights (Burnham and Anderson, 2002) based on the Bayesian information criterion (BIC) (Schwartz, 1978). In situations where prior information is small relative to the information provided by the data, the model with the highest posterior probability is the one that minimizes

$$\mathrm{BIC} = -2\log L + K\log n,$$

where L is the maximum of the likelihood under a specific model, K is the number of parameters, and n is the sample size. Used as a model-selection criterion, the model with the lowest BIC should be chosen.

Let R be the number of candidate models in the set $i = 1,2,\ldots,R$. For model i, we can get a relative distance, $\Delta\mathrm{BIC}_i = \mathrm{BIC}_i - \mathrm{BIC}_{\min}$, to the best model in the set, $\mathrm{BIC}_{\min}$. By normalizing the differences to sum to 1, we derive Akaike weights, which can be interpreted as the posterior probability that a model is the "quasi-true model" (the model that is the smallest dimension representation of a "true," data generating, unknown model) in a set of models, given the data, and prior probabilities on models (Burnham and Anderson, 2002). Thus, probability of model i, given data D, and prior model probabilities p_i, is

$$P(m_i|D) = \frac{\exp(-\frac{1}{2}\Delta\mathrm{BIC}_i)p_i}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta\mathrm{BIC}_r)p_r}.$$

The prior probability is typically set to $1/R$ (equal or flat prior). After posterior model probabilities were calculated, MCMC analyses were run under the models with the highest posterior probability using the program MrBayes v.3b4 (Ronquist and Huelsenbeck, 2003), and using the default priors of the program for the parameters. The two gene partitions were allowed to have their own set of parameters, using a rate multiplier (Nylander et al., 2004; Yang, 1996). Four parallel chains (Metropolis coupling; Huelsenbeck and Ronquist, 2001), with a heating parameter set to 0.2, were run for 3 million generations sampling every 100th generation. Two separate runs, starting from randomly chosen trees, were made to ensure that the chains had converged on the same target distribution. Trees sampled during the burn-in phase were discarded.

If a parameter θ is common over all models in the set, instead of using an estimate from a single model we can use the weighted average over all models. Thus, for model i in set R, weighted as w_i, we get an average estimate of θ by,

$$\hat{\hat{\theta}} = \sum_{i=1}^{R} w_i\hat{\theta}_i.$$

In our case, the tree topology and the branch lengths are the parameters θ in common to all models, and the weights w_i are the posterior model probabilities $P(m_i|\ D)$ or Akaike weights. To accomplish a weighted average, we sampled trees from each MCMC run under each model in proportion to the posterior probability of the models, and pooled them into one tree file. Finally, a majority-rule consensus tree was calculated from the pooled sample using the *sumt* command in MrBayes, to yield the final Bayesian estimate of phylogeny.

2.4. Biogeographic analysis

The historical biogeography of *Fagonia* was explored by dispersal–vicariance analysis using DIVA v. 1.1 (Ronquist, 1996). DIVA reconstructs ancestral distributions by optimizing areas on a given cladogram based on the vicariance model, while at the same time allowing dispersals and extinctions to occur (Ronquist, 1997). The method is event-based, and assumptions are not made about the existence of a general biogeographic pattern. Duplications and vicariance receive a cost of 0 in DIVA, whereas dispersals and extinctions cost 1 per area added or deleted, respectively. This is because "dispersals and extinctions are unpredictable events that can wipe out the traces of phylogenetically constrained processes like vicariance and duplication" (Sanmartín, 2003), and unless extinctions and dispersals are assigned a cost spurious events may be introduced in optimal reconstructions (Sanmartín and Ronquist, 2002).

The distributions of *Fagonia* species were assigned to nine areas of endemism. An area of endemism is here defined as a geographic region with at least two species that exhibit distributional congruence, following Harold and Moor (1994), except in two cases of isolated occurrences of single endemics. The areas are as follows: (A) South America, consisting of Chile and Peru, counting one species, (B) North America consisting of Baja California and adjacent parts of Mexico, and southwestern USA, with six species, (C) northeastern Mexico consisting of the region of Coahuila, with one species, (D) southern Africa, consisting of Botswana, Namibia, and South Africa, with two species, (E) the Saharo-Sindian region (modified after White and Léonard, 1991) from Mauritania in the west to India in the east, south to circa 15°N in the Sahel regional transition zone and north to southern Italy and Spain, and Macaronesia except the Azores, with 12 species, (F) the Horn of Africa region consisting of Djibouti, eastern Eritrea, eastern Ethiopia, northeastern Kenya, Socotra, and Somalia, with eight species, (G) southern Arabian Peninsula consisting of continental Yemen and Oman, with seven species, (H) western Morocco, with five species, and (I) southern Iran, with fives species as well. All areas except E correspond to geographically small areas well known for their local endemism (Friis et al., in press; Goldblatt, 1978; Medail and Quezel, 1999; Msanda et al., 2002; Thorne, 2000; Thulin, 1994; Wiggins, 1980), whereas the Saharo-Sindian region (E) is well known for its rather depauperate, species-poor flora and low numbers of endemic plants (Raven and Axelrod, 1974). We first analyzed our biogeographic data without constraints, but also explored the assumption that *Fagonia* arose in a small area, a "center of origin" by constraining the number of ancestral areas to 2, 3, and 4.

In a DIVA analysis a single tree has to be selected, because DIVA can only handle fully bifurcate trees. Here we have chosen to use an all compatible consensus tree based on the combined data set of *trnL* and ITS from the BMA analyses. This fully bifurcate tree has the advantage of showing the support, as PP, for each clade, including those with very low support. However, before choosing this strategy we inspected the shortest trees resulting from the parsimony analysis of the combined data set, concluding that differences between the shortest trees are small and confined to some of the uppermost clades and not the deeper nodes. The outgroup was reduced to *Melocarpum hildebrandtii* and *M. robecchii*, excluding *Roepera billardieri* and *Zygophyllum xanthoxylum* because they do not overlap in their distributions with any of the species of *Fagonia* or *Melocarpum*. Also, the distributions of *R. billardieri* and *Z. xanthoxylum* represent only two of several areas of endemism found in the genera *Roepera* and *Zygophyl-*

lum. All species with more than one terminal were pruned, leaving only one terminal for each species.

3. Results

3.1. Parsimony analysis

The aligned *trnL* matrix was 493 base pairs (bp) long, of which 339 positions were constant and 47 (9.5%) potentially parsimony-informative. The analysis produced 2881 minimum length trees (MLT), of 371 steps with consistency index (CI) 0.82 (excluding uninformative characters) and retention index (RI) 0.86. The MLT first found is shown in Fig. 2, with branch lengths (DELTRAN optimization) indicated above branches and BP in bold text. There is little resolution within *Fagonia*. The genus is weakly supported (BP 65; clade A), and *F. scoparia* is sister to the rest of *Fagonia* (BP 68; clade B). The other *Fagonia* species of the New World form a clade (BP 70; clade C), and *F. acerosa*, endemic to Iran, forms a clade with *F. olivieri*, known from Lebanon, Syria, Iran, Iraq, and Jordan (BP 61; clade P).

The aligned ITS matrix was 742 bp long, of which 406 positions were constant and 189 (25.5%) potentially parsimony-informative. The analysis produced 2843 MLTs, 985 steps long with a CI of 0.71 (excluding uninformative characters) and a RI of 0.74. The MLT first found is shown in Fig. 3; branch lengths (DELTRAN optimization) are indicated above branches and BP in bold text. *Fagonia* forms a well-supported clade (BP 100; clade A) and *F. scoparia*, endemic to Mexico, is weakly supported as sister to the rest of *Fagonia* (BP 65; clade B). Within clade B there is a well-supported clade with all the other New World species (BP 98; clade C). A well-supported clade is formed by *F. acerosa*, endemic to Iran, and the widespread North African-Asian *F. bruguieri* (BP 99; clade D). Another well-supported clade comprises *F. lahovarii*, endemic to the Horn of Africa region including Yemen, *F. latistipulata*, endemic to Somalia, and *F. mollis*, restricted to Egypt, Israel/Palestine, Jordan, and northern Saudi Arabia (BP 99; clade E). A clade consisting of *F. harpago*, endemic to Morocco, and *F. longispina*, endemic to Morocco and Algeria, is also well supported (BP 93; clade F). The two species from southern Africa, *F. minutistipula* and *F. rangei*, form a weakly supported clade (BP 65; clade G). All samples of *F. indica* and *F. paulayana*, which are widespread North African-Asian species, and *F. subinermis*, endemic to Iran, form a moderately supported clade (BP 76; clade H), within which there is little internal structure.

Analysis of the combined matrix produced 2835 trees of 1365 steps with CI of 0.72 (excluding uninformative characters) and RI of 0.75. The strict consensus tree is shown with BP indicated above the branches in Fig. 4. The topology of the consensus tree of the combined data set shows the same eight clades, A–H, as in the analysis of ITS only. However, the BP is higher in three of the clades (B, C, and H) and nearly the same in two (E and G). Also, the resolution within clade B is better in the combined tree than in the tree based on ITS only. *Fagonia cretica*, with a distribution from the Cape Verde Islands to southern Europe and North Africa east to Egypt, is weakly supported as sister to the rest of clade B (BP 64; clade I). Clade C, with all the New World species except *F. scoparia*, is weakly supported as sister to a clade with all the Old World species except *F. cretica* (BP 62; clade J). North American *F. laevis* is weakly supported (BP 63; clade K) as sister to the rest of the New World species in clade C, nesting the South American *F. chilensis* within the North American species. Also, *F. charoides*, endemic to Somalia, is weakly supported (BP 61; clade L) as sister to *F. harpago* and *F. longispina* (clade F).

3.2. Bayesian model averaging

The results of the likelihood evaluation and calculations of BIC and Akaike weights are summarized in Table 2. Even though all models were considered equally plausible a priori (each having a prior probability of 1/36), four models received practically all the posterior probability after calculating the BIC and Akaike weights. These four models were all parameter-rich, having six different substitution rates for each of the data partitions (the GTR model), combined with parameters for rate variation (and/or a proportion of invariable sites, I). The model receiving the highest posterior probability (0.428) was the GTR + Γ/GTR + Γ. The second most probable model (GTR + Γ/GTR + Γ + I) had a posterior probability of 0.287, and with the third and fourth having 0.171 (GTR + Γ + I/GTR + Γ), and 0.115 (GTR + Γ + I/GTR + Γ + I), respectively. Thus, the best model according to BIC was not much more probable than the second best (or third, or fourth). It can be noted that the model with the highest number of parameters, the GTR + Γ + I/GTR + Γ + I model, received a posterior probability of only 0.115. Furthermore, the two other models having the same number of parameters as the most probable model received a posterior probability of less than $1/10^6$. Neither of these models included parameters for rate variation for both data partitions, thus illustrating the importance to do so for a good fit of the model to this combined data set.

The tree resulting from the BMA analysis of the combined data set revealed the same clades (clade A–L) as the MP analysis (Fig. 5). The clades A–F, H, and J all have posterior probabilities (PP) of 1.0, and clades, I, K, and L have a PP between 0.85 and 0.94. Clade G has a PP of 0.66. The BMA analysis unveiled a clade consisting of clades E, H, and *F. gypsophila* as well as *F. mahrana* (PP 0.82; clade M), and a clade including

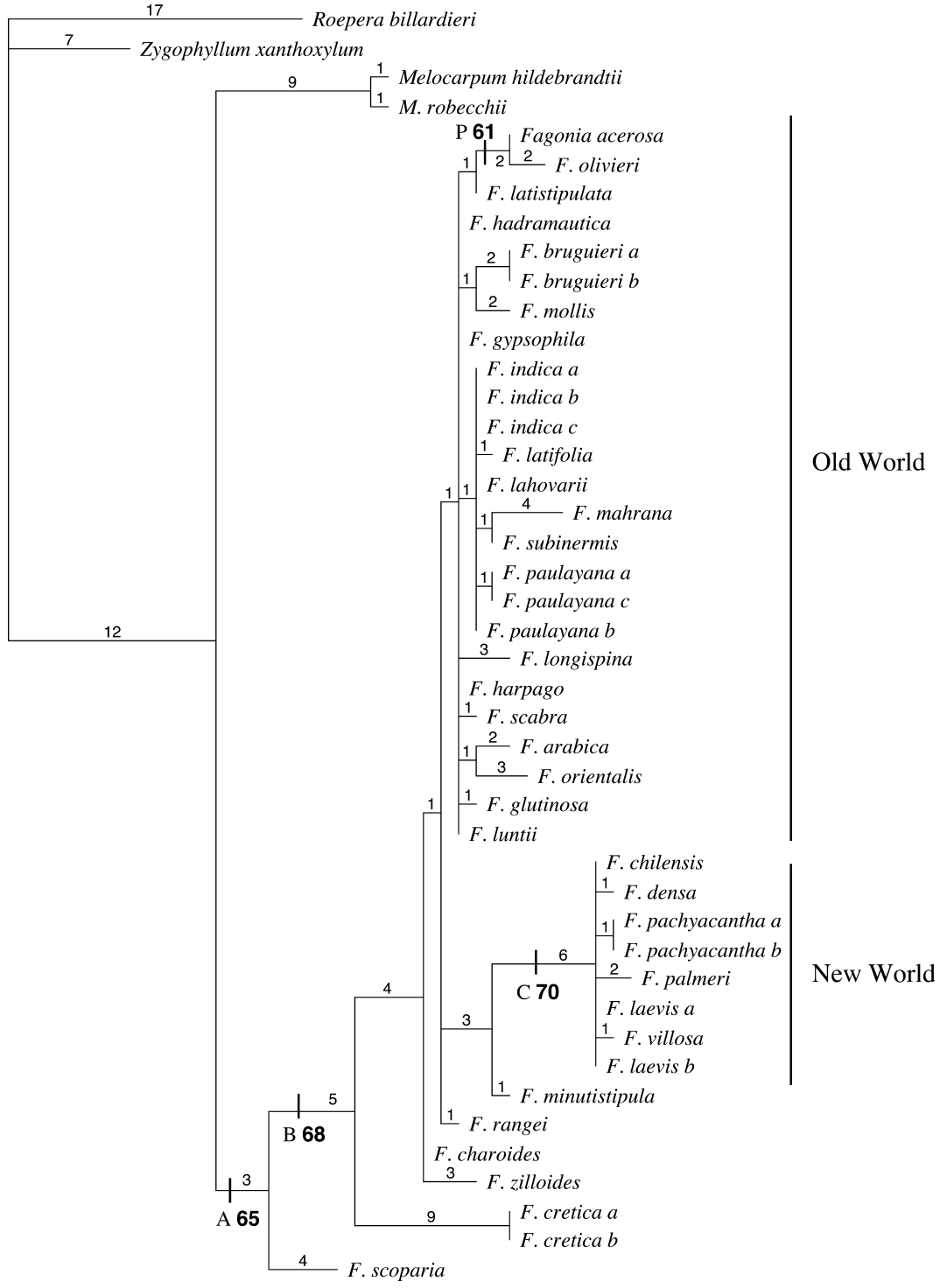

Fig. 2. The first minimum length tree based on *trnL* only. The analysis produced 2881 minimum length trees, of 371 steps with consistency index (CI) 0.82 (excluding uninformative characters) and retention index (RI) 0.86. The branch lengths (DELTRAN optimization) are indicated above the branches and bootstrap percentages in bold text.

F. glutinosa, *F. latifolia*, *F. orientalis*, and *F. zilloides* from the Saharo-Sind region, *F. hadramautica* from Yemen, and clade G, including F. *minutistipula* and *F. rangei* from southern Africa (PP 0.94; clade N). A clade with all Old World species except *F. cretica* and *F. luntii* was also found (PP 0.68; clade O).

For comparative reasons, we also performed a MCMC analysis using the single model indicated as best fit according to the BIC (GTR + Γ/GTR + Γ). Only minor differences between the results from the averaged and the single selected model could be seen in the topology estimates (results not shown). In fact, the only

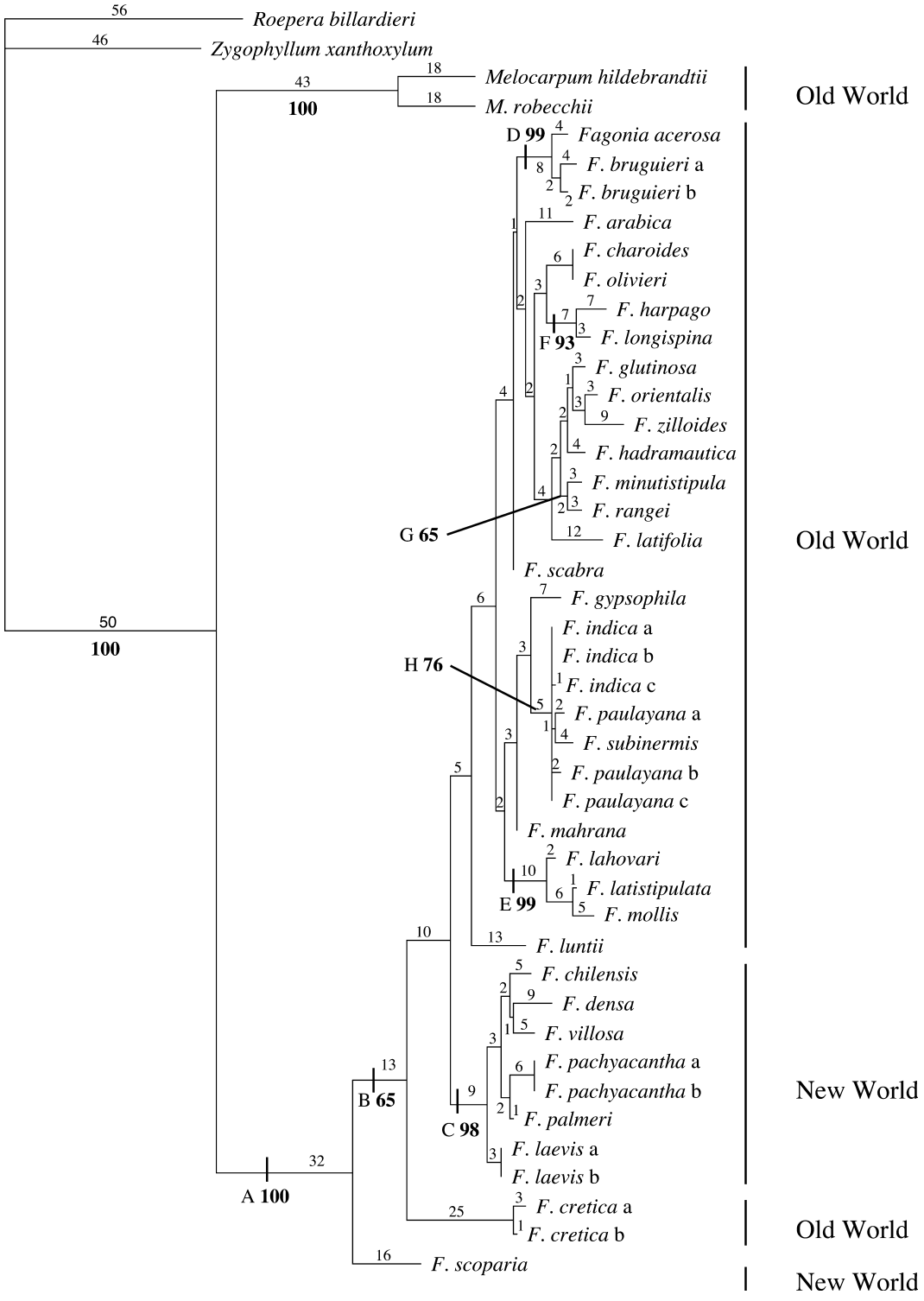

Fig. 3. The first minimum length tree based on ITS only. The analysis produced 2843 minimum length tree, 985 steps long with a CI of 0.71 (excluding uninformative characters) and a RI of 0.74. Branch lengths (DELTRAN optimization) indicated above branches, and bootstrap percentages below in bold.

differences observed were slight changes in PP for some of the clades with low PP's. This was probably due to the fact that the models receiving the highest posterior probability were reasonably similar, at least in the distributions of the topology parameter.

3.3. Biogeographic analysis

The result of the DIVA analysis is shown in Fig. 6, with the different optimal historical area reconstructions indicated at the nodes. The optimal reconstruction re-

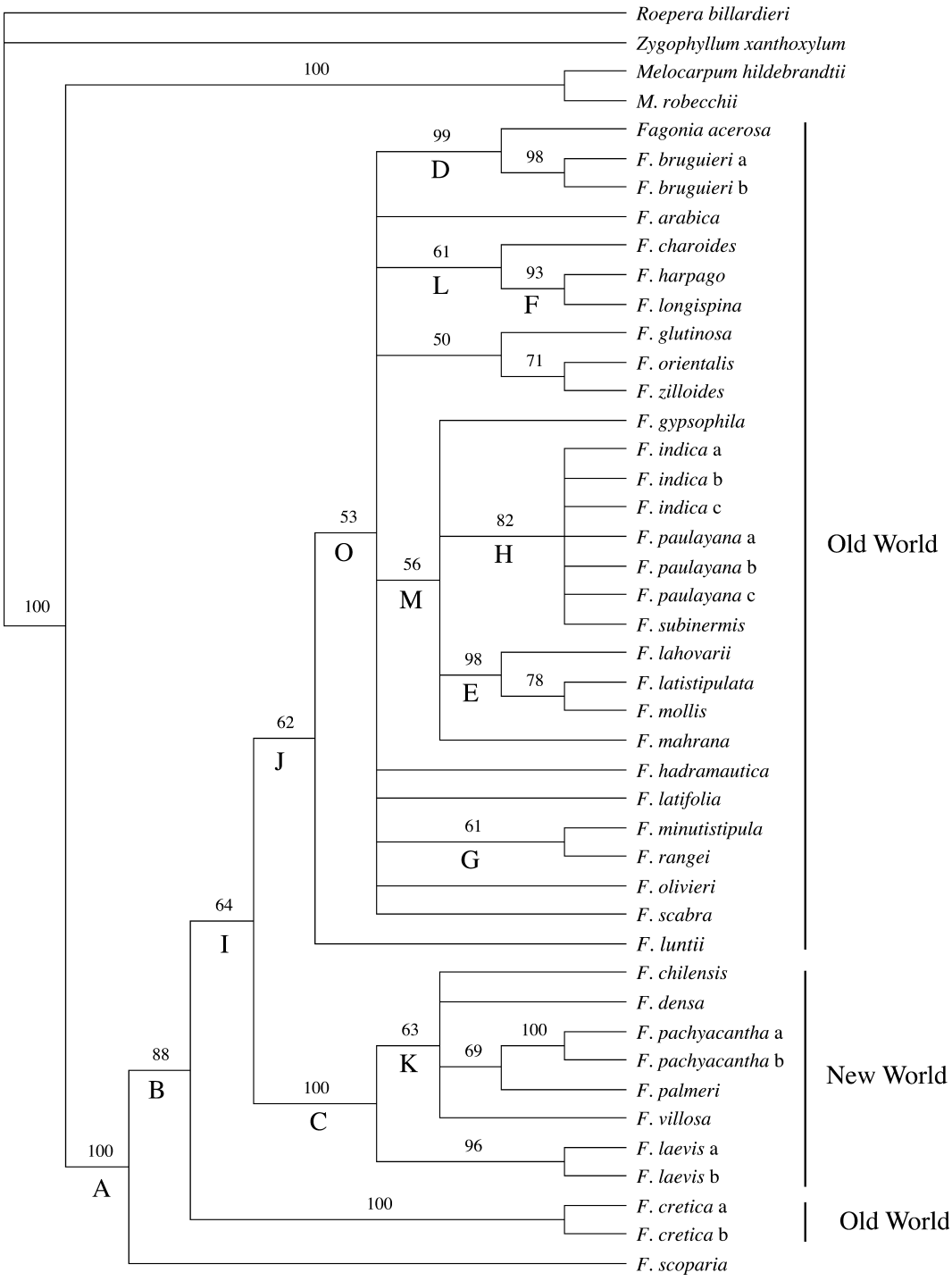

Fig. 4. Strict consensus tree based on a combined data set of *trnL* and ITS. The analysis produced 2835 trees, of 1365 steps with CI of 0.72 (excluding uninformative characters) and RI of 0.75. Bootstrap percentages are indicated above the branches.

quired 28 dispersals, and the historical distribution of the ancestor of *Fagonia* is given two alternative ancestral distributions in the analysis: North America (B), northeastern Mexico (C), and Saharo-Sind (E) with the bordering area of western Morocco (H), or the same area, but including the Horn of Africa (F). The ancestor to all species of *Fagonia* except *F. scoparia* is also given two alternative ancestral distributions: North America (B), Saharo-Sind (E) with the bordering area of Morocco (H), or the same area plus the Horn of Africa (F). Vicariance is inferred as the event that separated the ancestor of all *Fagonia* species, except *F. cretica* and *F. scoparia*, to the areas of North America (B) and the Horn of Africa region (F). Considering that vicariance took place between two areas not adjoined today, North America (B) and Horn of Africa (F), extinction is inferred at this

Table 2
Result of the likelihood evaluation of 36 different substitution models for the trnL and the ITS data

Model (*trn* L/ITS)	Number of parameters	Log likelihood	BIC	Posterior model probability[a]
HKY/HKY	8	−5607.32	11239.37	<0.001
HKY/HKY + Γ	9	−5453.93	10935.69	<0.001
HKY/HKY + Γ + I	10	−5452.18	10935.27	<0.001
HKY/GTR	12	−5564.13	11165.36	<0.001
HKY/GTR + Γ	13	−5416.95	10874.09	<0.001
HKY/GTR + Γ + I	14	−5415.80	10874.89	<0.001
HKY + Γ/HKY	9	−5580.57	11188.97	<0.001
HKY + Γ/HKY + Γ	10	−5427.19	10885.29	<0.001
HKY + Γ/HKY + Γ + I	11	−5425.43	10884.87	<0.001
HKY + Γ/GTR	13	−5537.39	11114.96	<0.001
HKY + Γ/GTR + Γ	14	−5390.21	10823.70	<0.001
HKY + Γ/GTR + Γ + I	15	−5389.26	10824.50	<0.001
HKY + Γ + I/HKY	10	−5579.78	11190.48	<0.001
HKY + Γ + I/ HKY + Γ	11	−5426.40	10886.80	<0.001
HKY + Γ + I/ HKY + Γ + I	12	−5424.64	10886.37	<0.001
HKY + Γ + I/GTR	14	−5566.59	11116.47	<0.001
HKY + Γ + I/GTR + Γ	15	−5389.41	10825.20	<0.001
HKY + Γ + I/GTR + Γ + I	16	−5388.27	10826.00	<0.001
GTR/HKY	12	−5589.46	11216.02	<0.001
GTR/HKY + Γ	13	−5436.08	10912.34	<0.001
GTR/HKY + Γ + I	14	−5434.32	10911.92	<0.001
GTR/GTR	16	−5546.27	11142.01	<0.001
GTR/GTR + Γ	17	−5399.09	10850.74	<0.001
GTR/GTR + Γ + I	18	−5397.95	10851.54	<0.001
GTR + Γ/HKY	13	−5563.15	11166.50	<0.001
GTR + Γ/HKY + Γ	14	−5409.77	10862.82	<0.001
GTR + Γ/HKY + Γ + I	15	−5408.01	10862.40	<0.001
GTR + Γ/GTR	17	−5519.97	11092.49	<0.001
GTR + Γ/GTR + Γ	18	−5372.79	10801.23	0.428
GTR + Γ/GTR + Γ + I	19	−5371.64	10802.03	0.287
GTR + Γ + I/HKY	14	−5562.53	11168.33	<0.001
GTR + Γ + I/HKY + Γ	15	−5409.14	10864.66	<0.001
GTR + Γ + I/HKY + Γ + I	16	−5407.38	10864.23	<0.001
GTR + Γ + I/GTR	18	−5519.34	11094.33	<0.001
GTR + Γ + I/GTR + Γ	19	−5372.16	10803.06	0.171
GTR + Γ + I/GTR + Γ + I	20	−5371.01	10803.86	0.115

[a] The posterior model probability is equal to the Akaike weight, w_i, based on the Bayesian information criterion (see text for details).

node. All the New World species of *Fagonia* are indicated to be a result of successive duplications (i.e., speciations within the area), with the exception of *F. chilensis* in South America, the distribution of which is conceived as a result of dispersal of the ancestor.

The ancestor of all species of the Old World, except *F. cretica*, was according to our analysis, endemic to the Horn of Africa (F), from where dispersal to Saharo-Sind (E), or western Morocco (H), and southern Arabia (G) took place. The ancestor of all species of the Old World, except *F. cretica* and *F. luntii*, had a wide distribution covering the Horn of Africa (F) and Saharo-Sind (E), or the Horn of Africa (F) and western Morocco (H).

The ancestor of clade M, which includes: *F. gypsophila*, *F. indica*, *F. lahovarii*, *F. latistipulata*, *F. mahrana*, *F. mollis*, *F. paulayana*, and *F. subinermis*, was endemic to the Horn of Africa (F), from where it dispersed into southern Arabia (G) and southern Iran (I), as well as to Saharo-Sind (E).

The ancestor to the species included in the clade sister to clade M, is here indicated as endemic to Saharo-Sind (E) or western Morocco (H). Subsequently, speciation within the area occurred, after which dispersal to southern Iran (I), southern Arabia (G), southern Africa (D) and to the Horn of Africa (F) took place. Some of the terminal nodes seem to be results of recent dispersals, and narrow endemics such as *F. hadramautica* are indicated to have evolved after relatively recent dispersals.

Constraining the number of ancestral areas in the DIVA analyses to 2, 3, and 4, respectively, resulted in more than 29 dispersals being required in each of the constrained runs. Also, constraining the ancestral areas to 3–4 resulted in a dramatic increase in alternative putative ancestral areas being proposed in the deep nodes. Each of the four deepest nodes was given between 9 and 20 different ancestral areas if a constraint of 3–4 areas was enforced.

The weak support in some parts of the *Fagonia* tree, as well as the unresolved clades in the strict consensus

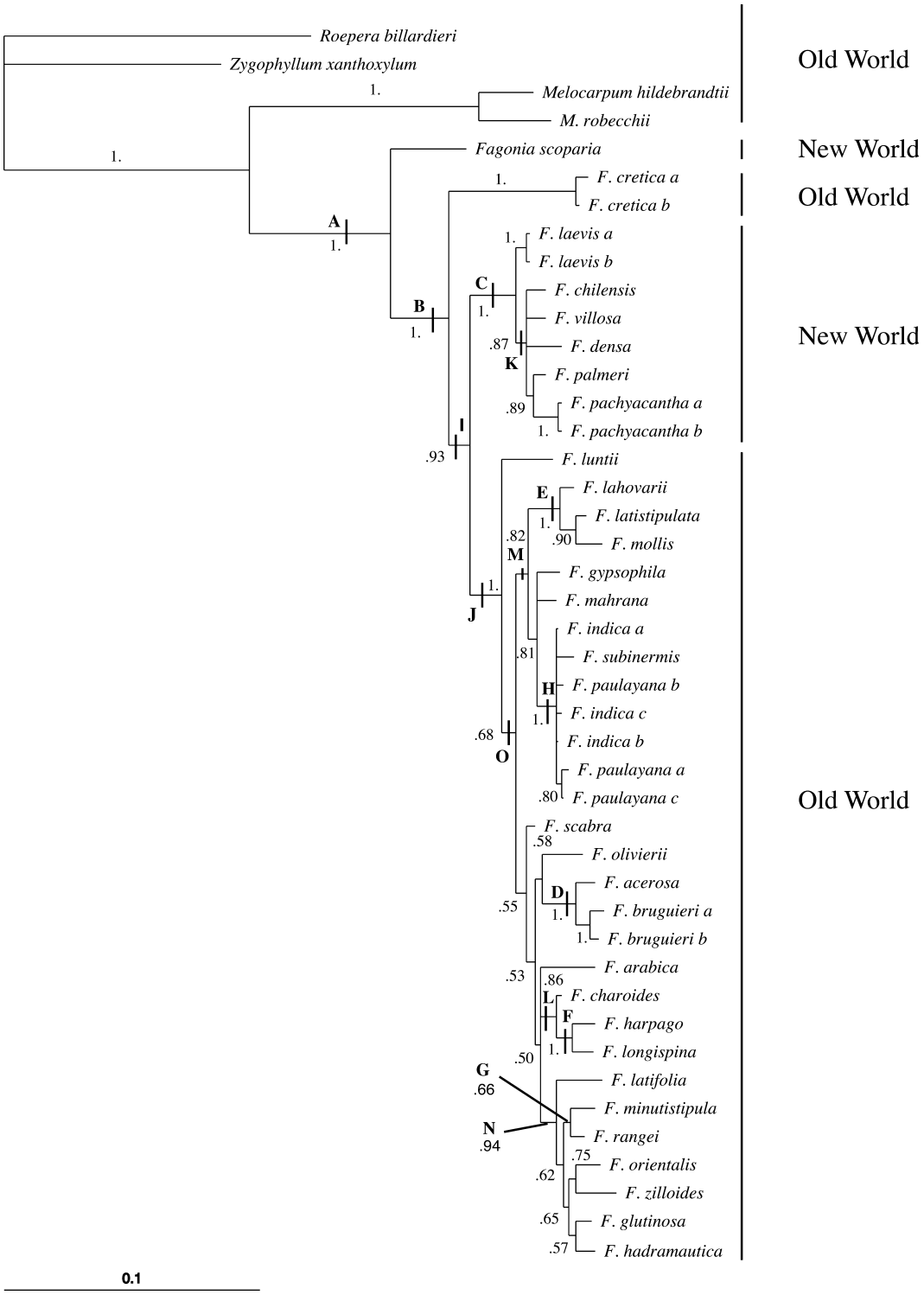

Fig. 5. Fifty percent majority-rule concensus tree produced by Bayesian model averaging. Posterior probabilities equal or above 0.5 are indicated at the branches.

tree of the MP analysis (Fig. 4), and in that inferred by BMA (Fig. 5), make conclusions about the historical biogeography for several of the clades unattainable. Nevertheless, the dispersal–vicariance analysis shows that the occurrence of *Fagonia* in South America and southern Africa is most likely a result of dispersal.

4. Discussion

4.1. Phylogenetic relationships

In this study we have only used a DNA data set to infer phylogenetic relationships in *Fagonia*. As was shown

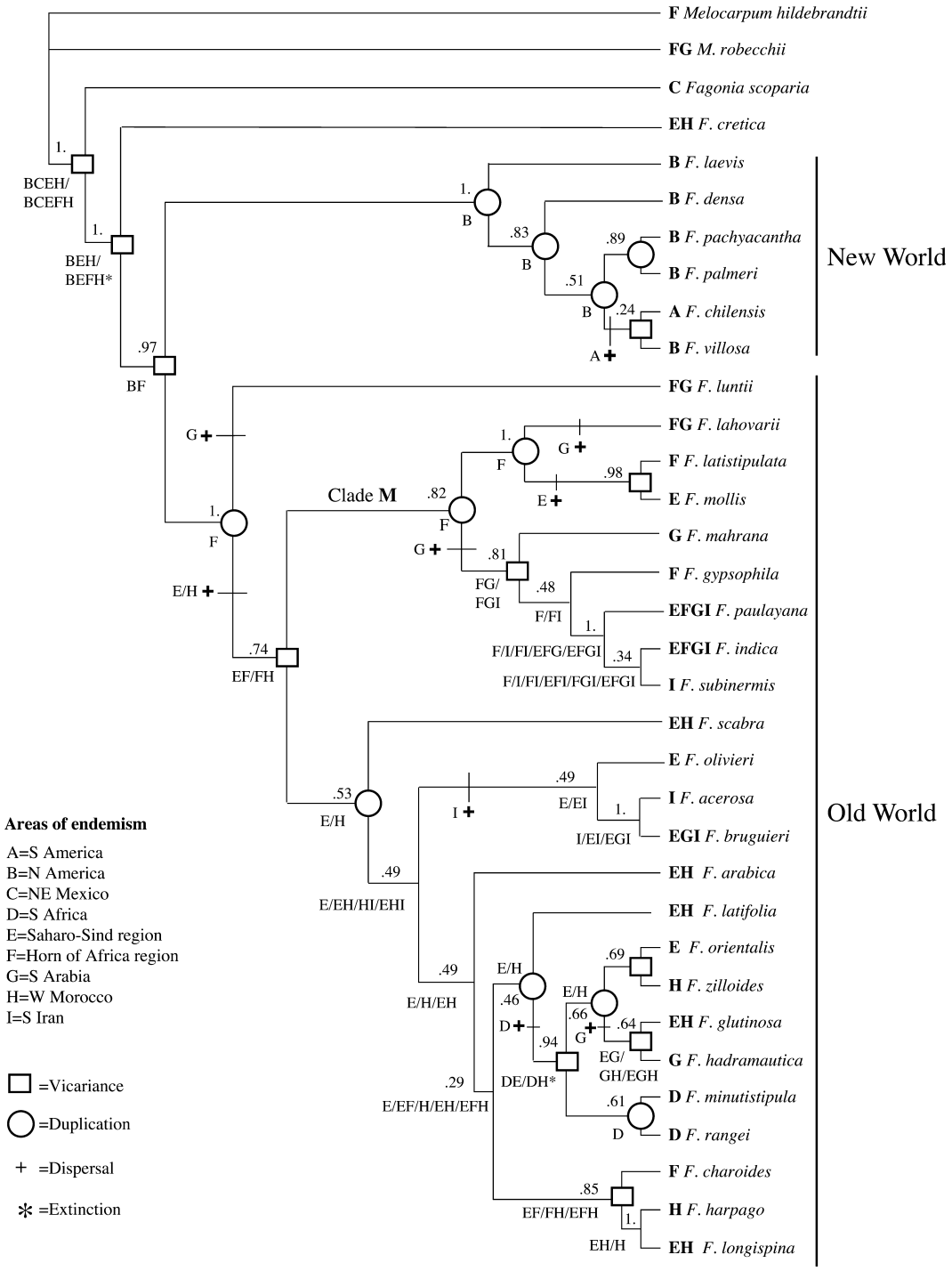

Fig. 6. The inferred historical distribution of *Fagonia*, using DIVA. The optimal reconstruction required 28 dispersals. Equally optimal distributions are separated by slash (/). Posterior probabilities are indicated at the branches.

by Beier et al. (2003), morphological data supported no internal resolution for the five species of *Fagonia* included, although the genus *Fagonia* was well supported by morphology (free stipules, pubescent and obconical fruits, and outer testa of seeds mucilaginous without internal structures). The apparent extensive homoplasy in the morphological characters is in our opinion reason for a molecular-only approach when examining phylogenetic relationships within *Fagonia*. In the analysis based on a combined data set in Beier et al. (2003), *F. scoparia* was shown to be sister to the rest of *Fagonia*, and a clade with the New World species, except *F. scoparia*, was weakly supported, and a clade with all Old World species was moderately supported. These results are confirmed here as well supported. However, this study differs in that *F. cretica* is found to be sister

to all species of *Fagonia* except *F. scoparia*. This difference originates from a misidentification of *F. scabra* in Beier et al. (2003), which emphasizes the importance of correct voucher information making it possible to trace a sequenced specimen.

The well-supported clades (e.g., clade C, D, E, and F in Figs. 4 and 5) in our study are not corroborated by any obvious morphological characters, which does not surprise us considering the plasticity of most morphological characters in *Fagonia* (Beier, in press). However, it should be noted that *F. indica* and *F. paulayana* are morphologically distinct, but not possible to distinguish on the basis of our sequence data. *Fagonia paulayana* is separated from the sympatric *F. indica* mainly by the deciduous (not persistent) sepals.

The schemes of relationships and classifications presented by Ozenda and Quézel (1957) and El Hadidi (1966) are not supported by our results. For example, *F. cretica* and *F. orientalis* (as "*F. flamandii*") were placed as sister taxa in their classification, as were *F. arabica* and *F. zilloides*. These relationships are far from supported by our results (Figs. 3–5). *Fagonia cretica* is sister to all *Fagonia* species except *F. scoparia*, and *F. zilloides* is sister to *F. orientalis*. Later, El Hadidi (1973) modified the classification and grouped the Old World *Fagonia* species into those with "tri- or unifoliolate leaves" including, e.g., *F. acerosa*, *F. bruguieri*, and *F. indica*, and those "with simple leaves" including *F. harpago*, *F. luntii*, and *F. ovalifolia* (now a synonym of *F. indica*). Species with simple leaves are scattered throughout the tree, and the group is clearly not supported by our results. Also, grouping species of *Fagonia* with the help of number of leaflets is according to our results not possible if monophyletic groups are aimed at. El Hadidi (1974) has also proposed that *Fagonia* species with short stipules and trifoliolate leaves were more closely related to each other than to species with long stipules and uni- to trifoliolate leaves. Different combinations of all these characters are found within several species of *Fagonia* (Beier, in press), and no pattern was observed in our analyses supporting the delimitation proposed by El Hadidi (1974).

4.2. Bayesian model averaging

It has been asserted that selecting a single model and using it for inference ignores model uncertainty, resulting in inferior inferences and overstatements of accuracy (Burnham and Anderson, 2002). The Bayesian approach allows us to incorporate the uncertainty by using model averaging. The Bayesian approach also allows us to take uncertainty in particular parameters into account in inference. Thus, the method possesses a number of advantages compared to other model based methods for phylogenetic inference (Huelsenbeck et al., 2002). However, recent computer simulations have shown that with certain types of model misspecifications, Bayesian inference can result in excessively high posterior probabilities for clades (Erixon et al., 2003; Suzuki et al., 2002). The reason for this is not yet clear (Erixon et al., 2003; Suzuki et al., 2002).

In this paper we have used two different methods for phylogenetic inference. An advantage of this approach is that a reference point is given by the traditionally used method of parsimony. Within the context of model averaging, one is tempted to extend the reasoning for the use of different phylogenetic inference methods for the same data. If we use two methods, e.g., MP and BMA (as in this paper), is there a way to combine the results for a final inference of a parameter (phylogenetic tree)? The statistical literature describes methods of comparing parametric and non-parametric methods within a Bayesian framework (e.g., Berger and Guglielmi, 1999), but methods for doing this in phylogenetics are still to be proposed.

Model averaging is a natural extension of the Bayesian framework of phylogenetic analysis. However, it should be remembered that all inference is conditional on both the data and the set of a priori models considered. This puts a focus on the choice of models to be contained in the set. Even though a model might receive a high posterior probability among a set of models, that particular model might still be highly inaccurate (cf. Bollback, 2002). We anticipate more work in the direction of model comparison, and the evaluation of model accuracy.

4.3. Historical biogeography

Fagonia is not unique in having a disjunct distribution in the Old and New World arid areas. There are several other genera among flowering plants, e.g., *Helianthemum* (Cistaceae), *Hoffmannseggia* (Fabaceae), *Prosopis* (Fabaceae), and *Thamnosma* (Rutaceae), with distributions including disjunctions between arid regions of South Africa, South America, North America, as well as North Africa and Asia (Mabberley, 1997). However, *Fagonia* is one of few genera that occur in all five of these regions.

Johnston (1940) and Axelrod (1950) considered the disjunct distribution of *Fagonia* to be a remnant of a widespread early Tertiary flora, which is in agreement with the boreotropics hypothesis (Davis et al., 2002; Lavin and Luckow, 1993; Lavin et al., 2000; Sanmartín et al., 2001; Tiffney, 1985; Wolfe, 1975). This hypothesis asserts that a land bridge existed in the North Atlantic between North America, Europe, and Africa during the Tertiary, coinciding with Eocene thermal maxima (Tiffney, 1985). The land bridge was in existence until at least the Eocene or early Oligocene, (30 mya), prior to the opening of the Red Sea, and during a period when South America was an isolated land mass (Axelrod and

Raven, 1978; Raven and Axelrod, 1974; Sanmartín et al., 2001; Tiffney, 1985; Wolfe, 1975).

The two successive vicariance events leading to the split of *Fagonia* (Fig. 6) into New and Old World lineages may be explained within the framework of the boreotropics hypothesis, whereas the occurrences of *Fagonia* in southern Africa and South America are here understood to be results of dispersals.

A history of dispersal from North America to South America, as indicated for the ancestor of *F. chilensis*, is shared by at least 106 species-pairs or identical species (Carlquist, 1983; Morrell et al., 2000; Thorne, 2000). All these plant species were most likely dispersed by migrating birds (Carlquist, 1983; Cruden, 1966). Also, during the Pleistocene, the climate provided considerably larger areas of similar ecology in California and Chile, making dispersal events even more likely (Carlquist, 1983).

The deserts and semi-deserts of Africa are the result of a development of a broad belt of drier climate that started to spread over areas previously covered by savanna or woodland in the early Miocene (23.8–20.5 mya). The widespread deserts of today are phenomena of even later periods, and the current desert vegetation of North Africa dates from the late Pliocene (3.6–1.8 mya) (Axelrod and Raven, 1978; Wickens, 1984). The exchange of taxa between the deserts of northern and southern Africa was made possible by an arid corridor between southwestern Africa and the Horn of Africa during repeatedly drier and colder times during the Pleistocene (1.8–0.01 mya) and probably also earlier (Goldblatt, 1978; van Zinderen Bakker, 1978). The distribution of *Fagonia* in southern Africa could thus be explained by an expanding distribution, i.e., dispersal, of the ancestor through this arid corridor and subsequent isolation, which would be in agreement with our vicariance analysis. However, for *Fagonia* with its sticky seeds, long distance dispersal by annually migrating birds along the eastern or western coast of Africa cannot be ruled out as a possibility for the establishment of the ancestor of the *Fagonia* species in southern Africa. Today at least 49 species of birds are known to migrate regularly between southern Africa and Eurasia (Sinclair et al., 1993; P. Alström, pers. comm.).

A disjunction between the arid zones of northern and southern Africa is known in many genera and species (Liston et al., 1989; Thulin, 1994; de Winter, 1971). Such disjunctions within species have generally been explained as a result of arid corridors during dry periods of the Pleistocene, whereas the disjunct genera may reflect more ancient histories of dispersals and vicariance.

The ancestor of all the Old World *Fagonia* species, except *F. cretica*, was according to our analysis restricted to the Horn of Africa region, from where dispersal took place to southern Arabia and the region of Saharo-Sind or western Morocco (H). However, there are indications of later vicariance between the Horn of Africa and southern Arabia in the case of *F. mahrana*. Also, in the biogeographic history of, e.g., *Zygocarpum* (Fabaceae), the occurrences in the Horn of Africa and southern Arabia have been interpreted as a result of vicariance (Thulin and Lavin, 2001).

The distributions of *F. cretica* on the Canary and Cape Verde archipelagos and *F. latifolia* on the Cape Verde Islands are indicated to be results of independent dispersal events. This agrees with the hypothesis that the neighboring African flora, and especially that of Morocco, has been the principal source for the colonization of the Canary and Cape Verde Islands (Bramwell, 1976; Medail and Quezel, 1999).

Assigning a time for the vicariance events that separated *Fagonia* of the New and the Old World, by estimating the nucleotide substitution rate and using the fossil record for calibration, is not made here. No fossils, to our knowledge, of *Fagonia* or *Melocarpum* are available. However, an extended study of the historical biogeography of Zygophyllaceae may make an estimation of the nucleotide substitution rate possible, since fossils of *Guiacum* (Zygophyllaceae) are available from the Rocky Mountains (45 mya) in the New World and the Rhine valley (20–23.8 mya) in the Old World (Menzel, 1914; Leopold and MacGinitie, 1972). This also indicates that other groups of Zygophyllaceae may have had a boreotropical history.

There is a difference in diversification between the amphi-Atlantic lineages of *Fagonia*, with 26 species in the Old World and only eight species in the New World. In the Old World, *Fagonia* is found over a much larger area than in the New World, where six out of eight species are restricted to Baja California, California, and three adjacent states. This may be part of the explanation of the difference in species diversity. However, the well-documented rapid climatic changes and limited areas appropriate as refugia may have caused higher extinction rates in the New World than in the Old World, but the history of the thermophilic boreotropical taxa in North America is not well known (Sanmartín et al., 2001; Thorne, 2000; Tiffney, 1985).

5. Conclusions

The hypothesis that the Tertiary North Atlantic land bridge had a significant influence on the development of modern continental biotas (Davis et al., 2002; Lavin et al., 2000; Sanmartín et al., 2001), including that of Africa and North America, is compatible with our study. The proposal of Stebbins and Day (1967), that the distribution of *Fagonia* is the result of a pre-Tertiary migration from the Old to the New World via the Beringian land bridge is in our view much less likely. This is because the climate was less favorable in the Bering area for thermophilic taxa, and also, the phylogenetic

relationships are more in line with a boreotropic history for the group.

Long-distance dispersal across the Atlantic, from Africa to South America as suggested by Raven and Axelrod (1974) and Porter (1974), is not supported by our results. However, this does not rule out the possible dispersals between the islands that made up the Tertiary North Atlantic land bridge. Also, Engler's (1896, 1915) idea about anthropogenic dispersal from the Old to the New World is highly unlikely since the species of the New World differ considerably from those of the Old World with respect to molecular and morphological characters.

Acknowledgments

We thank the directors and curators of the herbaria at B, E, K, M, RAS, S, SD, UPS, and W for allowing us to take samples for DNA extraction and A.G. Marticorera for sending material of *F. chilensis*. We also thank D. Devey, N. Heidari, and J. Joseph for excellent assistance with sequencing. I. Sanmartín, G. Thor and two anonymous reviewers are thanked for valuable comments on the manuscript. Mikael Thollesson and the Linnaeus Centre of Bioinformatics, Uppsala provided computational facilities. This work has been supported by grants to B.-A. Beier from Helge Ax:son Johnssons stiftelse, Regnells stiftelse, S. J. Enanders fond, Stiftelsen Hiertas Minne, Tullbergs fond, Uddenberg-Nordingska Stiftelsen, and the EC-funded TMR Programme through The Natural History Museum, London.

References

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

APG II, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. Bot. J. Linn. Soc. 141, 399–436.

Archibald, J.K., Mort, M.E., Crawford, D.J., 2003. Bayesian inference of phylogeny: a non-technical primer. Taxon 52, 187–191.

Axelrod, D.I., 1950. Evolution of desert vegetation in western North America. Publ. Carnegie Inst. Washington 590, 215–306.

Axelrod, D.I., 1970. Mesozoic paleogeography and early angiosperm evolution. Bot. Rev. 36, 277–319.

Axelrod, D.I., Raven, P.H., 1978. Late Cretaceous and tertiary vegetation history of Africa. In: Werger, M.J.A. (Ed.), Biogeography and Ecology of Southern Africa. Junk, Hague, pp. 79–119.

Beier, B.-A., Chase, M.W., Thulin, M., 2003. Phylogeny and taxonomy of the subfamily Zygophylloideae (Zygophyllaceae) based on molecular and morphological data. Plant Syst. Evol. 240, 11–40.

Beier, B.-A., in press. A revision of *Fagonia* (Zygophyllaceae). Syst. Biodiv.

Berger, J.O., Guglielmi, A., 1999. Bayesian testing of a parametric model versus nonparametric alternatives. J. Am. Stat. Assoc. 96, 174–184.

Bollback, J.P., 2002. Bayesian model adequacy and choice in phylogenetics. Mol. Biol. Evol. 19, 1171–1180.

Bramwell, D., 1976. The endemic flora of the Canary Islands; distribution, relationships and phytogeography. In: Kunkel, G. (Ed.), Biogeography and Ecology in the Canary Islands. Junk, Hague, pp. 207–240.

Buckley, T.R., Arensburger, P., Simon, C., Chambers, G.K., 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. Syst. Biol. 51, 4–18.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference, A Practical Information-theoretic Approach, second ed. Springer, New York.

Carlin, B.P., Chib, S., 1995. Bayesian model choice via Markov chain Monte Carlo methods. J. R. Statist. Soc. B 57, 473–484.

Carlquist, S., 1983. Intercontinental dispersal. Sonderbd. Naturwiss. Ver. Hamburg 7, 37–47.

Cruden, R.W., 1966. Birds as agents of long-distance dispersal for disjunct plant groups of the temperate western hemisphere. Evolution 20, 517–532.

Davis, C.C., Bell, C.D., Mathews, S.M., Donoghue, M.J., 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malphigiaceae. Proc. Natl. Acad. Sci. USA 99, 6833–6837.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of leaf tissue. Phytochem. Bull. Bot. Soc. Am. 19, 11–15.

Engler, A., 1896. Über die geographische Verbreitung der Zygophyllaceen im Verhältnis zu ihrer systematischen Gliederung. Abh. Königl. Akad. Wiss. Berlin 2, 1–36.

Engler, A., 1915. In: Engler, A., Drude, O. (Eds.), Vegetation der Erde 9, 3(1), p. 731.

Engler, A., 1931. Zygophyllaceae. In: Engler, A., Prantl, K. (Eds.), Die natürlichen Planzenfamilien, second ed., vol. 19. Engelmann, Leipzig, pp. 144–184.

El Hadidi, M.N., 1966. The genus *Fagonia* in Egypt. Candollea 21, 13–54.

El Hadidi, M.N., 1973. Revision of *Fagonia* species with tri- to unifoliolate and simple leaves. Österr. Bot. Z. 121, 269–278.

El Hadidi, M.N., 1974. Weitere Beobachtungen an der Gattung *Fagonia*. Mitt. Bot. Staatssamml. München 11, 387.

Erixon, P., Svennbland, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Fay, M.F., Bayer, C., Alverson, W.S., de Bruijn, A.Y., Chase, M.W., 1998. Plastid *rbc*L sequence data indicate a close affinity between *Diegodendron* and *Bixa*. Taxon 47, 43–50.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fitch, W.M., 1971. Towards defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406–416.

Friis, I., Thulin, M., Adersen, H., Bürger, A.-M., in press. Patterns of plant diversity and endemism in the Horn of Africa. Biol. Skr.

Gaut, B.S., Lewis, P.O., 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol. Biol. Evol. 12, 152–162.

Goldblatt, P., 1978. An analysis of the flora of southern Africa: its characteristics, relationships, and origins. Ann. Missouri Bot. Gard. 62, 369–436.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Gu, X., Fu, Y.-X., Li, W.-H., 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. 12, 546–557.

Harold, A.S., Moor, R.D., 1994. Areas of endemism: definition and recognition criteria. Syst. Biol. 43, 261–266.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21, 160–174.

Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Biometrics 17, 754–755.

Johnston, I.M., 1940. The floristic significance of shrubs common to North and South America. J. Arnold Arbor. 21, 224–233.

Kelcher, S., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Missouri Bot. Gard. 87, 482–498.

Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

Lavin, M., Luckow, M., 1993. Origins and relationships of tropical North America in the context of the boreotropics hypothesis. Am. J. Bot. 80, 1–14.

Lavin, M., Thulin, M., Labat, J.-N., Pennington, R.T., 2000. Africa, the odd man out: molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. Syst. Bot. 25, 449–467.

Leopold, F., MacGinitie, H.D., 1972. Development and affinities of tertiary floras in the rocky Mountains. In: Graham, A. (Ed.), Floristics and Paleofloristics of Asia and Eastern North America. Elsevier, Amsterdam, pp. 170–200.

Liston, A., Rieseberg, L.H., Elias, T.S., 1989. Genetic similarity is high between intercontinental disjunct species of *Senecio* (Asteraceae). Am. J. Bot. 76, 383–388.

Mabberley, D.J., 1997. The Plant Book, second ed. Cambridge University Books, Cambridge.

Medail, F., Quezel, P., 1999. The phytogeographical significance of SW Morocco compared to the Canary Islands. Plant Ecol. 140, 221–244.

Menzel, P., 1914. Beitrag zur Flora der Niederrheinischen Braunkohlenformation. Jahrb. Königl. Preuss. Geolog. Landesanst. Berlin 34, 1–36.

Morrell, P.L., Porter, J.M., Friar, E.A., 2000. Intercontinental dispersal: the origin of the widespread South American plant species *Gilia laciniata* (Polemoniaceae) from a rare California and Oregon coastal endemic. Pl. Syst. Evol. 224, 13–32.

Msanda, F., El-Aboudi, A., Peltier, J.-P., 2002. Originalitéde la flore et de la vegetation de l' Anti-Atlas sud-occidental (Maroc). Feddes Repert. 113, 603–615.

Nylander, J.A.A., 2004. Bayesian Phylogenetics and the Evolution of Gall Wasps. Dissertation, Acta Univ. Upsal. 937.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.-L., 2004. Bayesian phylogenetic inference of combined data. Syst. Biol. 53, 47–67.

Oxelman, B., Lidén, M., 1995. Generic boundaries in the tribe Sileneae (Caryophyllaceae) as inferred from nuclear rDNA sequences. Taxon 44, 525–542.

Ozenda, P., Quézel, P., 1957. Les Zygophyllacées de l' Afrique du Nord et du Sahara. Trav. Inst. Rech. Sah. 14, 43–44.

Popp, M., Oxelman, B., 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol. Phylogen. Evol. 20, 474–481.

Porter, D.M., 1974. Disjunct distributions in the New World Zygophyllaceae. Taxon 23, 339–346.

Posada, D., Crandall, K.A., 1998. Modeltest, testing the models of substitution. Bioinformatics 14, 817–818.

Raven, P.H., Axelrod, D.I., 1974. Angiosperm biogeography and past continental movements. Ann. Missouri Bot. Gard. 61, 539–673.

Rodríguez, J., Oliver, L., Marín, A., Medina, R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Ronquist, F., 1996. DIVA v. 1.1 Computer program for MacOS and Win32. Available from: <http://morphbank.ebc.uu.se/DIVA>.

Ronquist, F., 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46, 195–203.

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

Sanmartín, I., 2003. Dispersal versus vicariance in the Mediterranean: historical biogeography of the palearctic Pachydeminae (Coleoptera, Scarabaeoidea). J. Biogeogr. 30, 1883–1997.

Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73, 345–390.

Sanmartín, I., Ronquist, F., 2002. New solutions to old problems: widespread taxa, redundant distributions, and missing areas in event-based biogeography. Animal Biodiv. Conserv. 25, 75–93.

Schwartz, G., 1978. Estimating the dimensions of a model. Ann. Stat. 6, 461–464.

Seelanan, T., Schnabel, A., Wendel, J.F., 1997. Congruence and consensus in the cotton tribe (Malvaceae). Syst. Bot. 22, 259–290.

Sinclair, I., Hockey, P., Tarboton, W., 1993. Birds of Southern Africa. Struik, Cape Town.

Stebbins, G.L., Day, A., 1967. Cytogenetic evidence for long continued stability in the genus *Plantago*. Evolution 21, 409–428.

Sullivan, J., Swofford, D.L., 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. 4, 77–86.

Sun, Y., Skinner, D.Z., Lang, G.H., Hulbert, S.H., 1994. Phylogenetic analysis of sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor. Appl. Gen. 89, 26–32.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Nat. Acad. Sci. USA 99, 16138–16143.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.

Taberlet, P., Gielly, G., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lec. Math. Life Sci. 17, 57–86.

Thorne, R.F., 2000. Phytogeography of North America north of Mexico. In: Flora of North America Editorial Committe (Eds.) North American Flora, 1, p. 250.

Thulin, M., 1994. Aspects of disjunct distributions and endemism in the arid parts of the Horn of Africa. In: Seyani, J.H., Chikuni, A.C. (Eds.), Proc. XIIIth meeting AETFAT 2, pp. 1105–1119.

Thulin, M., Lavin, M., 2001. Phylogeny of the *Ormocarpum* group (Fabaceae): a new genus *Zygocarpum* from the Horn of Africa region. Syst. Bot. 26, 299–317.

Tiffney, B., 1985. The Eocene North Atlantic bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. J. Arnold Arbor. 66, 243–273.

Wasserman, L., 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44, 92–107.

White, T., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), PCR protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.

White, F., Léonard, J., 1991. Phytogeographical links between Africa and Southwest Asia. In: Engel, T., Frey, W., Kürschner, H. (Eds.),

Contributiones Selectae ad Floram et Vegetationem Orientis. J. Cramer, Berlin, pp. 229–246.

Wickens, G.E., 1984. Flora. In: Cloudsley-Thompson, J.L. (Ed.), Key Environments: Sahara Desert. Pergamon, Oxford, pp. 67–75.

Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.

Wiggins, I.L., 1980. Flora of Baja California. Stanford University Press, Stanford.

de Winter, B., 1971. Floristic relationships between the northern and southern arid areas in Africa. Mitt. Bot. Staatssamml. München 10, 424–437.

Wolfe, J.A., 1975. Some aspects of plant geography of the Northern Hemisphere during the Late Cretaceous and Tertiary. Ann. Missouri Bot. Gard. 62, 264–279.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39, 306–314.

Yang, Z., 1996. Maximum likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.

van Zinderen Bakker, E.M., 1978. Quarternary vegetation changes in southern Africa. In: Werger, M.J.A. (Ed.), Biogeography and Ecology of Southern Africa. Junk, Hague, pp. 79–119.