Available online at www.sciencedirect.com
SCIENCE DIRECT®
Organisms, Diversity & Evolution 5 (2005) 147–159
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians

Charles D. Bell*,[1], Michael J. Donoghue

Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 065211, USA

Received 22 April 2004; accepted 26 October 2004

Abstract

Species of Valerianaceae are a common component of the alpine flora throughout the Northern Hemisphere as well as the Andes of South America. Sequence data from three chloroplast markers (*psb*A-*trn*H intron, *trn*K-*mat*K intron, and the *trn*L-F region) along with the internal transcribed spacer region (ITS) of nuclear ribosomal DNA were used to infer relationships within Valerianaceae. Both genomes, as well as a combined data set, provide support for the major clades within the group and do not support a monophyletic *Valeriana*. In addition, these data indicate that *Plectritis* is nested within South American *Valeriana*, as opposed to being sister to *Centhranthus* as previously hypothesized. Valerianaceae appear to have originated in Asia, probably in the Himalayas, and subsequently to have dispersed several times to Europe and to the New World. Our results imply that Valerianaceae colonized South America on multiple occasions from the north. In one of these cases there appears to have been a substantial and rapid radiation, primarily in the high elevation paramo habitat. A variety of methods were used to estimate divergence times to determine when Valerianaceae might have colonized South America. Regardless of the method and fossil constraints applied, our estimates suggest that Valerianaceae colonized South America prior to the formation of the Isthmus of Panama.

Keywords: Andes; Biogeography; Maximum likelihood; Phylogenetic systematics; South America; Valerianaceae

Introduction

Valerianaceae contains ca. 350 species distributed throughout much of the world (except Australia and New Zealand), mostly at high elevations and with many species in alpine zones. These plants are characterized by sympetalous, bilaterally symmetric, or sometimes highly asymmetric flowers; inferior, three-carpelate ovaries, with one fertile carpel at maturity and a single anatropous ovule; achene fruits; and the absence of endosperm. The presence in many species of iridoids of the valepotriate type also characterizes the group (see Backlund and Moritz, 1998).

Valerianaceae exhibit considerable diversity in flower and fruit morphology. Perhaps the most impressive variation in flower morphology concerns the number of stamens, which ranges from four to one (five stamens have also been reported in *Patrinia*; Eriksen, 1989). The trend in the group has been toward reduction in the number of stamens. Donoghue et al. (2003) inferred that

*Corresponding author.
E-mail address: cbell3@xula.edu (C.D. Bell).
URL: http://www.senckenberg.de/odes/05-09.htm.
[1]Present address: Department of Biology, Xavier University of Louisiana, New Orleans, LA 70125, USA.

1439-6092/$ - see front matter © 2005 Gesellschaft für Biologische Systematik. Published by Elsevier GmbH. All rights reserved.
doi:10.1016/j.ode.2004.10.014

there was an initial reduction from the ancestral condition of four to three stamens in the core Valerianaceae, followed by further reduction to two stamens in *Fedia* and, independently, to a single stamen in *Centranthus*. The calyx in Valerianaceae can be persistent and leafy (as in *Nardostachys*), reduced to small teeth (as in *Fedia* and *Valerianella*), featherly and pappus-like (as in *Centranthus* and species of *Valeriana*), or completely lacking (as in species of *Valeriana*). Eriksen (1989) hypothesized that complete reduction of the calyx occurred independently on several occasions within Valerianaceae (e.g., within the Latin American species of *Valeriana*). The degree of reduction of the two sterile locules in the ovary is quite variable, from highly reduced to inflated (as in some *Valerianella* and *Valeriana*). Modifications of the calyx and of the sterile locules are related to dispersal mode.

Valerianaceae traditionally has been subdivided into three tribes (Graebner, 1906): Triplostegieae, Patrinieae, and Valerianeae. Many authors (see Weberling, 1970; Cronquist, 1988; Brummitt, 1992; Backlund, 1996) have recognized 14 genera: a single genus (*Triplostegia*) in the tribe Triplostegieae, two (*Patrinia* and *Nardostachys*) in the tribe Patrinieae, and 11 genera assigned to five subtribes within the Valerianeae. Six genera besides *Valeriana* have been recognized in South America (*Aretiastrum*, *Astrephia*, *Belonanthus*, *Phuodendron*, *Phyllactis*, and *Stangea*). However, recent treatments of the South American taxa have argued for placing these species in *Valeriana*, thus reducing the number of genera within Valerianaceae to eight (Borsini, 1944; Larsen, 1986; Eriksen, 1989).

Since the early 1990s, but especially since 2000, a variety of molecular and morphological phylogenetic analyses have shed light on the relationships of Valerianaceae within the Dipsacales and on relationships among the major lineages within Valerianaceae (Donoghue et al., 1992; Downie and Palmer, 1992; Backlund and Donoghue, 1996; Pyck et al., 1999; Pyck and Smets, 2000; Bell et al., 2001; Donoghue et al., 2001; Pyck et al., 2002; Bell and Donoghue, 2003; Donoghue et al., 2003; Zhang et al., 2003; Bell, 2004a, b; Hidalgo et al., 2004). Regarding relationships within Dipsacales, these studies have concluded that Valerianaceae is the sister group of a clade that includes *Triplostegia* (often treated as a member of Valerianaceae) and the Dipsacaceae. The Valerianaceae/Dipsacaceae clade in turn is most closely related to Morinaceae, together forming an herbaceous group that Donoghue et al. (2001) called the Valerina clade. In turn, Valerina is linked with the Linnaeeae of the former Caprifoliaceae in the Linnina clade of Donoghue et al. (2001).

Regarding relationships within Valerianaceae, it appears (based on all data sets except *mat*K) that the first split separates *Patrinia* from a clade containing all the other species. Within this remaining clade the first split is between *Nardostachys* and a core Valerianaceae clade. These results demonstrate that the traditional Patrinieae (Graebner, 1906), consisting of *Patrinia* and *Nardostachys*, is paraphyletic. Within core Valerianaceae, *Valeriana*, *Centranthus*, and *Plectritis* form a clade that is sister to a clade containing *Fedia* and *Valerianella*. Most recently, Bell (2004a), based on a combined analysis of chloroplast and nuclear DNA sequences, found support for a South American valerian clade.

Overall, the distribution of Valerianaceae matches that of other Dipsacales clades (e.g., *Viburnum*, *Lonicera*). It is primarily found around the Northern Hemisphere, and probably originated in Asia and later moved into Europe and the New World (Bell et al., 2001; Donoghue et al., 2003; Bell, 2004a, b). Although its center of origin may be Asia, the present-day center of Valerianaceae species diversity is in South America, where many different morphological forms—from rosette plants (*V. rigida*) to microphyllous shrubs (*V. microphylla*) to annual vine-like species (*Valeriana chaerophylloides*)—occupy a wide range of habitats. Species of Valerianaceae are abundant and diverse in the Andes, especially in the paramo, scattered along the crests of the highest ranges in the northern Andes or on isolated mountaintops from 3000 to 5000 m.

Eriksen (1989) hypothesized that the diversity of Valerianaceae in South America represents three separate invasions (Fig. 1). She suggested that *Valeriana* twice entered South America through the Isthmus of Panama. One such movement was by the single species *V. chaerophylloides* (sect. *Astrephia* subgen. *Valeriana*), the other gave rise to subgen. *Phyllactis*, with some 160 species. In the case of *Phyllactis* this conclusion was based on the presence of the derived bisporangiate anther condition. Eriksen (1989) envisioned a separate invasion into Venezuela for sect. *Porteria*, using the Caribbean islands as stepping stones.

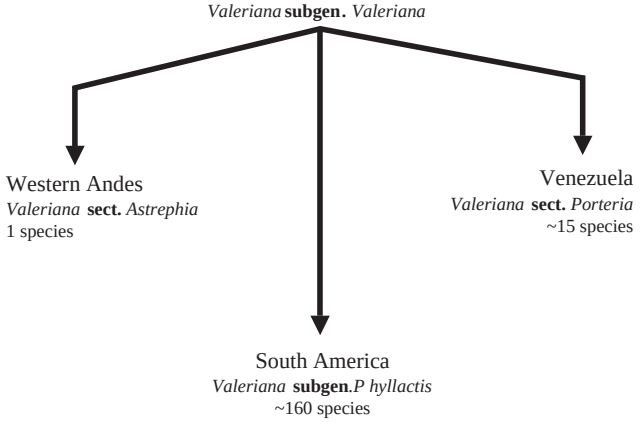

Fig. 1. The assumed pathways of *Valeriana* into the South American continent as proposed by Eriksen (1989).

Knowledge of the history of the Andes places some constraints on the times when Valerianaceae may have entered and radiated into South America. The Andes extend approximately 9000 km, and reach nearly 7000 m in elevation. The southern Andes of Bolivia, Chile, and Argentina are the oldest, with considerable uplift during the early Tertiary period about 50 million years ago. The northern Andes of Colombia and Ecuador are younger, with the major uplift begun in the Miocene, about 20 myr. Paramo vegetation probably originated much more recently, after the final uplift of the Andes, which occurred ca. 5.0–3.5 mya (van der Hammen, 1974; Simpson, 1975; van der Hammen and Cleef, 1986). Nevertheless, the paramo now ranks among the top ecosystems worldwide in floristic diversity and richness of endemic species (Smith and Cleef, 1988; Luteyn, 1999).

The objectives of the present study were: (1) to infer phylogenetic relationships within Valerianaceae based on nuclear and chloroplast DNA sequence data, but with a great increase in the number of sampled taxa, especially from South America; and (2) to further elucidate the biogeographic history of the group, especially to test the Eriksen's (1989) biogeographic hypothesis for the colonization of South America.

Material and methods

Plant material

Sequence data from 90 accessions representing 88 taxa (61 included for the first time in any phylogenetic analysis) were collected from herbarium specimens, silica-preserved and fresh plant material. See Organisms Diversity and Evolution Electronic Supplement 05–09. *Triplostegia* was sequenced and used as an outgroup based on previous work on Dipsacales (Pyck et al., 1999; Pyck and Smets, 2000; Bell et al., 2001; Donoghue et al., 2001; Donoghue et al., 2003).

DNA extraction, PCR, and sequencing

Total DNA was extracted using the CTAB methods of Doyle and Doyle (1990), or with Qiagen DNeasy plant mini extraction kits (Qiagen). Double-stranded copies of all regions were amplified using standard PCR in 25–50 μL volume reactions. All reactions were initially heated at 94 °C for 3 min. Reactions then entailed 35 cycles consisting of 94 °C for 1.5 min, 48–52 °C for 2 min, and 72 °C for 3 min. All amplified PCR products were cleaned prior to sequencing using QIAquick PCR purification kits (Qiagen).

The *trn*K intron and a portion of the 5′ end of the *mat*K protein were amplified with primers *trn*k-11 and *mat*K510R (Young et al., 1999). Two newly designed primers for Dipsacales—*mat*K-seq1 (5′-ATTWGAAT-GATTKCCCGATCTA-3′) along with its reverse complement *mat*K-seq2 (5′-TAGATCGGGMAATCA-TTCWAAT-3′)—were used along with amplification primers for sequencing. All amplification and sequencing of the *trn*L-F IGS region and *trn*L intron were carried out using the universal primers trnlc, trnld, trnle, and trnlf of Taberlet et al. (1991). The ITS region was amplified and sequenced using the primers ITS2, ITS3 (in some cases, ITS3B), ITS4, and ITS5 (White et al., 1990; Urbatsch et al., 2000). The *psb*A-*trn*H intergenic spacer was amplified using forward and reverse primers designed by Sang et al. (1997).

All sequencing was done via dye terminator cycle sequencing using the protocol specified by the manufacturer, and then visualized on either an ABI 377 automated DNA sequencer, an MJ Research BaseStation51 DNA fragment analyzer, or on an ABI 3100 capillary sequencer. Sequence fragments were visualized and edited using the computer package Sequencher (Gene Codes Corporation) to build contig sequences. All sequences were then aligned visually with the help of MacClade version 4.0 (Maddison and Maddison 2000). An aligned matrix is available in TreeBase (www.treebase.org) or on request from the first author.

Phylogenetic analysis

Maximum parsimony searches were conducted using heuristic search methods with tree bisection reconnection (TBR) branch swapping, collapse of zero-length branches, and all characters weighted equally. The analyses were repeated 100 times with the RANDOM ADDITION option. Sets of equally most parsimonious trees were summarized by a strict consensus tree. Bootstrap tests (Felsenstein, 1985) were performed using 300 replicates with heuristic search settings identical to those of the original search. All parsimony and likelihood analyses were performed using the computer software PAUP* vers. 4.0b10 for UNIX (Swofford, 2002). To investigate potential conflict between the partitions of the nuclear and chloroplast data sets, we performed a homogeneity test (i.e. ILD test) as implemented in PAUP*, using 1000 replicate searches.

A series of likelihood ratio tests (LRTs) was performed using the computer software PORN* (Bell, 2001), on a variety of tree topologies, to determine which model of sequence evolution best fit the data. A variety of best fitting models were found, depending on the taxa and data partition being examined. Parameters for each search were simultaneously estimated via maximum likelihood for all data sets. Heuristic search methods were used with TBR branch swapping, and collapse of zero-length branches. The analyses were

repeated 100 times with the RANDOM ADDITION option. Bootstrap tests under maximum likelihood were not performed because they were computationally prohibitive.

Bayesian analyses were carried out using Metropolis-coupled Markov chain Monte Carlo (MCMCMC; MC3) methods as implemented in MrBayes ver. 3.0B (Huelsenbeck and Ronquist, 2001). The chain was run for 40–50 million generations, initiated with a random starting tree, and was sampled every 1000 generations. A total of 40,000–50,000 trees were sampled from the posterior distribution of trees and used to calculate the posterior probabilities of clades. Sample points collected prior to stationarity were eliminated from the sample (i.e. 'burn-in'). The incremental heating scheme for each search used the default settings in MrBayes (i.e. 3 'heated' chains, plus 1 'cold' chain). We used the default settings for the priors on the rate matrix (0–100), branch lengths (0–10), shape parameter (0–10), and the proportion of invariable sites (0–1). A Dirichlet distribution was used for the base frequency parameters, and an uninformative prior was used for the tree topology.

Divergence time estimation

A likelihood ratio test was used to test for departure from clock-like evolution (Felsenstein, 1981) for each of the data sets. In all cases, rate constancy among lineages was rejected. Therefore, we used several recently proposed methods that variously 'relax' the null hypothesis of a uniform rate of molecular evolution across all lineages (Sanderson, 1997; Thorne et al., 1998; Huelsenbeck et al., 2000; Yoder and Yang, 2000; Sanderson, 2002). These analyses used a single candidate tree: the Bayesian maximum likelihood tree from the combined data set (Fig. 1).

Penalized likelihood (PL; Sanderson, 2002) is a semiparametric smoothing method that assumes that there is an autocorrelation in substitution rates and attempts to minimize rate changes between ancestral/descendant branches on a tree (i.e. at the nodes). Branches are allowed to change rates of molecular evolution, but are penalized when rates change from ancestral to descendant branches. A smoothing parameter (λ) can vary from very small, in which case each branch of the phylogeny has a different substitution rate (saturated model), to very large, in which case parameters are essentially clock-like. The crux of the PL method is determining the optimal smoothing level. The program r8s (Sanderson, 2003) implements a data-driven cross-validation procedure that systematically prunes terminals from the tree, then estimates parameters from the submatrix using a given smoothing value. It then tries to predict the data for the pruned taxa using the estimated parameters, and calculates a χ^2 error associated with the difference between the predicted and observed data. The optimal smoothing level is chosen as the one that minimizes the error (Sanderson, 2002). To compute confidence intervals for the divergence times estimated (LF, NPRS, PL), we used the bootstrap resampling method (Baldwin and Sanderson, 1996).

Recent Bayesian methods (Thorne et al., 1998; Kishino et al., 2001) that relax the clock were also used to estimate divergence times, using the computer software MULTIDIVTIME (available from J. Thorne, North Carolina State University). This parametric approach uses a continuous autocorrelation of substitution rates across the phylogeny, and allows for the simultaneous use of several calibrations/time constraints.

Our Bayesian divergence time estimation involved two steps. First, the program ESTBRANCHES estimates branch lengths from the data and a fixed tree topology using the F84 (Felsenstein, 1984; Kishino and Hasegawa, 1989) model of sequence evolution, allowing rates to vary among sites following a discrete gamma (Γ) distribution with four rate categories (Yang, 1994). Parameters for the F84 + Γ model were estimated using the software package PAML (Yang, 2000). Next, the outgroup (*Triplostegia*) was pruned from the tree and MULTIDIVTIME was used to estimate the prior and posterior ages of splits among taxa, their standard deviations, and the 95% credibility intervals via Markov chain Monte Carlo. The Markov chain was for 1,000,000 generations and sampled every 100 generations after an initial burn-in period of 10,000 cycles. To check for convergence of the MCMC, analyses were run at least twice from different starting points. The following priors distributions were used in the analysis of the combined data: 50 my (SD = 5 my) for the expected time between tip and root if there had been no constraints (based on results from Bell and Donoghue, 2005); 0.00010 (SD = 0.00004) substitutions per site per my for the rate of the root node (estimated as the mean rate from the data); 0.02 (SD = 0.02) for the parameter that determines the magnitude of the autocorrelation per my; and 65 my for the largest value of time units between the root and the tips.

For comparative purposes, we also estimated divergence times using nonparametric rate smoothing (NPRS; Sanderson, 1997), and a maximum likelihood method that assumes rate constancy among lineages (LF; Langley and Fitch, 1974). Both of these methods were performed using the computer software r8s (Sanderson, 2003).

Calibration points and time constraints

The fossil record for Valerianaceae is limited. However, the distinctive-winged fruits of *Patrinia* have been documented from the Miocene to Pliocene of Poland

and Russia (Lancucka-Srodoniowa, 1967), as well as from the late Miocene of Japan (Ozaki, 1980). Likewise, *Valeriana* is known on the basis of fossil fruits from the late Miocene and Pliocene of Europe. Another possible time constraint is the presence of lineages in the high Andes. Because the high alpine habitat was not available before ca. 5.0–3.5 mya, it would be parsimonious to assume that the ancestors of paramo clades could not have colonized before that time. These minimum and maximum age constraints were used separately and in combination to estimate the minimum age of the colonization of South America by Valerianaceae. In a previous study interested in divergence times in Dipsacales (Bell and Donoghue, 2005), an age of ca. 60–55 my was estimated for the crown group Valerianaceae. These ages were also used as calibration points, as well as priors distributions in our age estimates.

Results

Phylogenetic analyses

The combined chloroplast data set consisted of 2770 aligned nucleotides (*trn*L = 1059; *mat*K-intron = 1230; *psb*A = 481), of which 1842 were variable and 914 were parsimony informative. The parsimony search resulted in 254 most parsimonious trees 5146 steps in length, with a CI of 0.5375 (0.4455 excluding autapomorphies) and an RI of 0.5991. Maximum likelihood and Bayesian analyses resulted in a single tree topology, with a $-\ln L$ of 33,030.26 under the GTR + Γ model of sequence evolution.

The ITS data set consisted of 758 aligned nucleotide sequences, of which 513 were variable and 359 parsimony informative. The maximum likelihood and Bayesian analyses resulted in a single tree topology, with a $-\ln L$ of 11,803.62 under the GTR + Γ model of sequence evolution. The parsimony analysis resulted in 37,352 trees 2109 steps in length, with a CI of 0.4500 (0.3955 excluding autapomorphies) and an RI of 0.5817.

A parsimony search of the combined data resulted in 320 trees 7428 steps in length, with a CI of 0.5001 (0.4148 excluding autapomorphies) and an RI of 0.5757. The maximum likelihood and Bayesian analyses resulted in a single tree topology, shown in Fig. 2, with a $-\ln L$ of 46,092.91714 under the GTR + Γ model of sequence evolution.

The ITS and the combined data sets were found to be congruent based on the ILD test as performed in PAUP* ($p = 0.09$). In general (see discussion below), results from the ITS, combined cpDNA, and combined data sets were very similar and consistent with previously published results (Bell et al., 2001; Donoghue et al., 2003; Bell, 2004a; Hidalgo et al., 2004). Bootstrap values and posterior probabilities did, however, vary from data set to data set and were generally higher for the combined data.

For the combined data set, as well as the ITS and the cpDNA analyses, *Patrinia* and *Nardostachys* form the basal splits, and *Valerianella* and *Fedia* form a well-supported clade (100% BS, 1.0 PP in the combined analysis) that is sister to a *Valeriana*/*Centranthus*/*Plectritis* clade. The two *Centranthus* species form a clade (100% BS; 1.0 PP) that is sister to a *Valeriana* (except *V. celtica*) plus *Plectritis* clade, and the three *Plectritis* species form a clade (100% BS; 1.0 PP) within *Valeriana*. All three data sets also support the position of *V. celtica* as sister to the *Fedia*/*Valerianella*/*Valeriana*/*Centranthus*/*Plectritis* clade.

The combined data provide support for several small clades (Figs. 2 and 3). At the base of the *Valeriana* + *Plectritis* clade, a clade (Clade III) comprised of exclusively European taxa (*V. supina*, *V. montana*, *V. tripteris*, and *V. pyrenaica*) is sister to the remainder of the clade. A second clade (Clade II) comprises both New World and Old World species of *Valeriana*, but with the New World species (e.g., *V. acutiloba* and *V. sitchensis*) all restricted to North America. In both the Bayesian and the parsimony analyses, the bulk of the South and Central American taxa plus the species of *Plectritis* form a strongly supported clade (Clade I) that is sister to members of *Valeriana* series *Edulis* sensu Meyer (1951; Figs. 2 and 3). Several clades are resolved within Clade I, with varying degrees of support (Fig. 3). All three species of *Plectritis* form a well-supported clade (100% BS; 1.0 PP for all three data sets) that is nested well within *Valeriana*. Hidalgo et al. (2004) also found *Plectritis* to be nested within a clade with South American species of *Valeriana*. However, within the large *Valeriana* + *Plectritis* clade, the chloroplast data provide little support for any major groupings. Within this clade, several groups of High-Andean paramo endemics form clades with moderate to high bootstrap support (Figs. 3 and 4).

Divergence times

The hypothesis of rate constancy among lineages was rejected for all data sets, and we therefore employed methods that variously allow rates to differ among lineages (see above). Our PL (and NPRS) analyses that were calibrated using only the *Patrinia* fossil consistently estimated dates that we believe to be too young based on broader analyses of Dipsacales that incorporated additional fossils (Bell and Donoghue, 2005; authors' unpublished data), especially *Diplodipelta* of the Linnaeeae (Manchester and Donoghue, 1995). This underestimation may result from the default placement of the fossil along the stem of the *Patrinia* clade. If the *Patrinia*

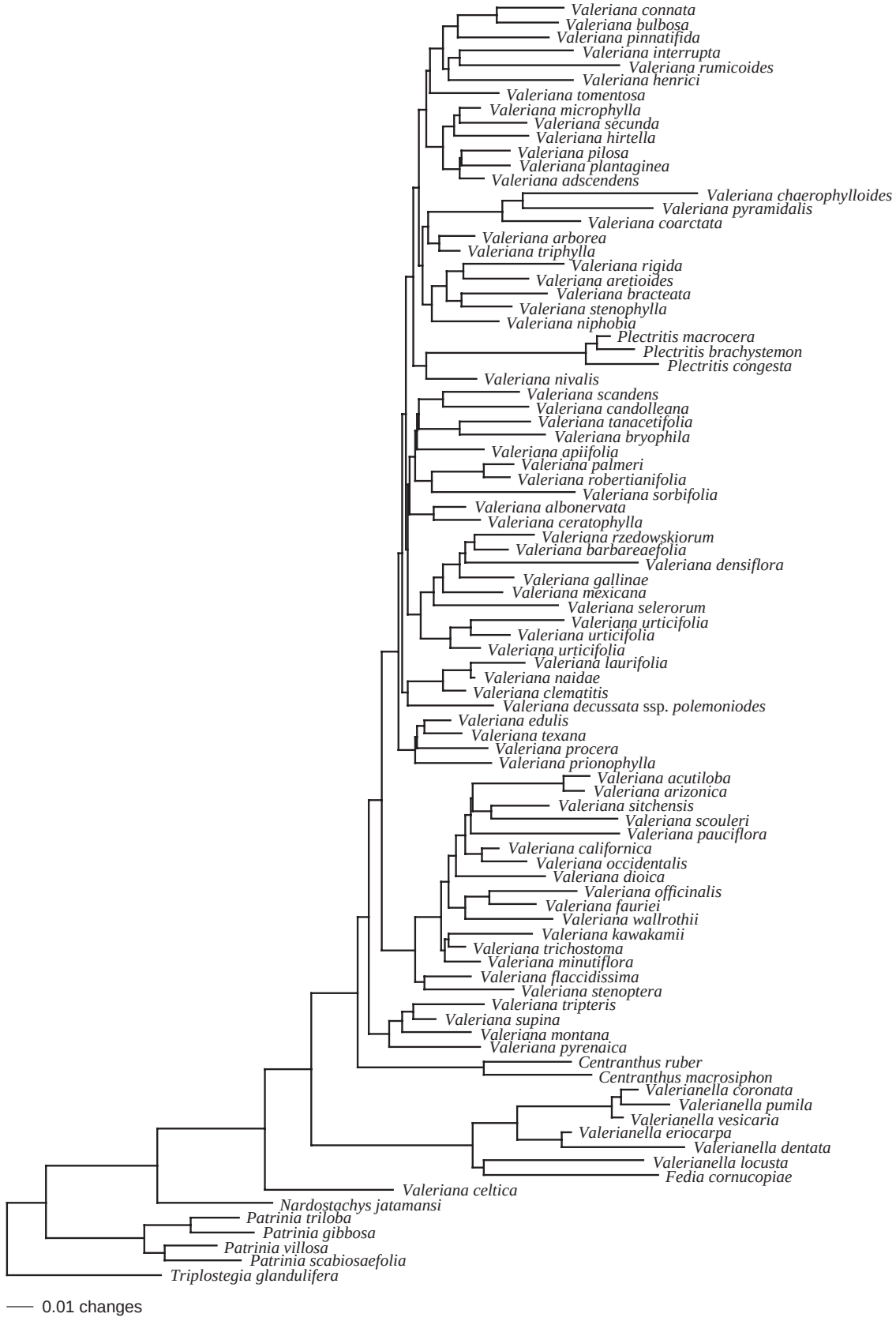

Fig. 2. Maximum likelihood tree topology inferred during Bayesian analysis of the combined nuclear and chloroplast data via MCMC. Branch length proportional to inferred number of substitutions/site under a GTR + Γ model of sequence evolution.

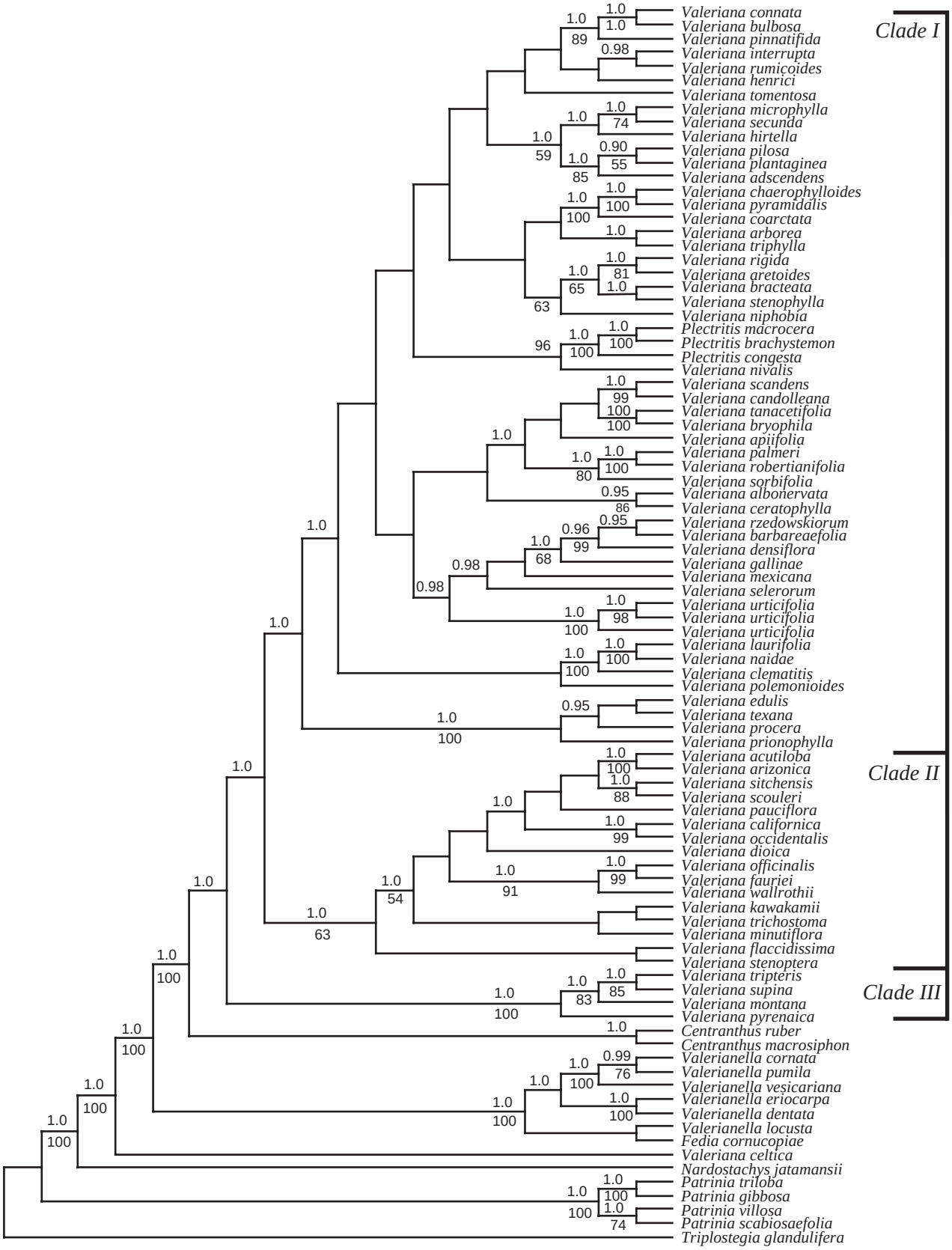

Fig. 3. Maximum likelihood tree topology inferred during Bayesian analysis of the combined nuclear and chloroplast data via MCMC. Numbers above branches correspond to posterior probability values, numbers below branches represent bootstrap proportions from 300 replicate searches under the maximum parsimony criterion.

fossil could be placed with some confidence within the crown clade (based on synapomorphies uniting it with one or more of the extant species), this would push age estimates back in time for all deeper nodes. Likewise, when either of the Andean clades was used to calibrate the trees, estimated dates were too young in comparison with estimates from more inclusive analyses with fossils that could be placed with confidence (Bell and Donoghue, 2005).

When no constraints were enforced, and an age of 55 my (based on age estimates from Bell and Donoghue, 2005) was used as a calibration point for the most recent common ancestor (MRCA) of Valerianaceae, the Bayesian relaxed-clock method gave the youngest ages for the colonization of South America, followed by the LF estimates, then the NPRS estimates, regardless of the data set (Table 1). For the combined data, ages for the MRCA of the Andean-paramo clade were between 9 and 20 my for the Bayesian relaxed-clock analysis. Estimates for this clade were all older than with any of the other methods (i.e. LF, NPRS, and PL). When the maximum age constraint of 5.0 my was enforced on the Andean clade, estimated divergence times for the colonization of South America were more recent for all of the methods and data sets used (Table 1). The NPRS and PL estimates were very similar for each of the three nodes of interest for the ITS data (e.g., node *X*: 25.76 mya for NPRS, 25.72 mya for PL) but not for the chloroplast and combined data sets (see Table 1). Bayesian relaxed-clock ages were significantly younger than either the NPRS or PL estimates for all three data sets when age of the MRCA of the paramo taxa was constrained to 5 my. Regardless of method or data set used, age estimates suggest that the major colonization of South America by Valerianaceae occurred prior to the formation of the Isthmus of Panama (ca. 3.5–3.1 mya; Coates and Obando, 1996).

Discussion

Phylogeny of Valerianaceae and taxonomic implications

The analyses reported here include the largest sample of Valerianaceae to date. These analyses, along with previous studies (Pyck and Smets, 2000; Bell et al., 2001; Pyck et al., 2002; Raymundez et al., 2002, Donoghue et al., 2003; Zhang et al., 2003; Bell, 2004a; Hidalgo et al., 2004), provide support for the following phylogenetic conclusions: (1) *Patrinia* and then *Nardostachys* form the two basal-most divergences within Valerianaceae; (2) species of *Valerianella* and *Fedia* form a well supported clade; (3) *Valeriana* is not monophyletic, with *V. celtica* falling outside of the other sampled species and *Plectritis* being nested within the South American valerians; (4) Valerianaceae originated in Asia, and colonized the New World on a number of occasions; and (5) a large group of exclusively New World valerians form a well supported clade.

These data also support the conclusions of other authors (Borsini, 1944; Eriksen, 1989) that the South America taxa should be treated as *Valeriana*; the recognition of a number of segregate genera would render *Valeriana* paraphyletic (Fig. 4). However, even without the recognition of these additional South American genera species of *Valeriana* do not appear to form a clade: *V. celtica* falls outside of the clade including the remainder of the species, and *Plectritis* is nested well within a clade of South American valerians. This placement of *Plectritis* differs from several other analyses in which it has been linked directly with *Centranthus* (Bell et al., 2001; Donoghue et al., 2003; Bell, 2004a). Unlike the larger chloroplast and ITS data set presented by Bell (2004a), the South American species are not supported as monophyletic; *Plectritis* is strongly supported as falling within this group (Fig. 2). Hidalgo et al. (2004) inferred this same result in their analyses of *mat*K.

The weakly supported nature of these results makes it difficult to conclude much about relationships among many of the South American species. Although some clades are supported by high posterior probability values (i.e. 1.0) from the Bayesian analysis (Fig. 3), bootstrap values are often less the 50% on the same branch. However, these two values represent different statistical measures of support and are not necessarily directly comparable (B. Larget and P.O. Lewis, pers. comm.). These results do, however, draw into question the taxonomic significance of such morphological characters as the lobed anther condition used by Eriksen (1989), as well as other morphological characters such as the presence or absence of a calyx. More confident resolution will require additional data, not only more characters but also more taxa. Until such time, our understanding of character evolution within this morphologically diverse group will remain uncertain.

Biogeography of Valerianaceae

Valerianaceae are distributed throughout much of the Northern Hemisphere, with a current center of species diversity in South America. The Dipsacaceae + *Triplostegia* clade is absent from the New World. The basal lineages of both groups are found in the mountainous regions of Asia, especially in the Himalayas, and DIVA reconstructions imply that the entire Valerina clade originated in eastern Asia and moved from there into

Table 1. Divergence time estimates (in millions of years) in Valerianaceae

	LF	NPRS[a]	NPRS[b]	PL[a]	PL[b]	BRC[a]	BRC[b]
ITS							
Paramo clade	19.62±2.3	26.89±1.9	5.0±0	31.68±2.3	5.0±0	6.02–18.6	3.47–4.98
Node *X*	20.06±2.5	27.63±2.1	25.76±2.0	32.28±2.2	25.72±1.9	6.94–20.3	4.15–10.54
Node *Y*	20.20±3.0	27.97±2.2	26.22±1.6	32.50±1.8	26.18±2.0	7.59–21.72	4.59–11.75
Node *Z*	21.03±2.9	29.68±2.0	28.13±1.5	33.68±1.8	28.11±1.5	8.36–23.36	5.12–13.26
Valerianaceae	55.00±0	55.00±0	55.00±0	55.00±0	55.00±0	49.54–69.71	46.45–65.40
Chloroplast							
Paramo clade	16.63±1.9	21.85±2.2	5.0±0	25.02±2.3	5.0±0	6.22–14.89	3.88–4.99
Node *X*	17.08±2.0	22.88±2.3	18.21±1.7	25.78±2.1	6.33±0.8	6.76–15.97	4.22–6.05
Node *Y*	17.75±1.5	24.28±2.2	20.60±2.0	26.95±1.9	7.13±0.5	7.92–18.02	5.02–8.17
Node *Z*	18.13±1.6	24.89±2.2	21.69±1.8	27.46±2.0	7.74±0.6	8.66–19.41	5.52–9.36
Valerianaceae	55.00±0	55.00±0	55.00±0	55.00±0	55.00±0	51.23–71.20	47.00–65.37
Combined							
Paramo clade	18.91±2.3	20.30±1.9	5.0±0	23.79±2.6	5.0±0	9.46–19.69	4.16–4.99
Node *X*	19.28±2.2	21.53±1.5	21.80±1.8	24.43±2.0	7.61±1.9	9.91–20.51	4.40–5.61
Node *Y*	19.75±1.9	22.86±1.8	23.49±2.1	25.54±2.0	8.80±2.2	11.06–23.72	5.01–7.03
Node *Z*	19.99±1.9	23.40±2.0	24.11±2.0	25.21±2.2	9.26±2.0	11.75–23.72	5.37–7.86
Valerianaceae	55.00±0	55.00±0	55.00±0	55.00±0	55.00±0	45.35–65.82	39.10–59.76

BRC = Bayesian relaxed clock, LF = Langly Fitch, NPRS = nonparametric rate smoothing, PL = penalized likelihood; nodes *X*, *Y*, *Z* represent several possible times at which Valerianaceae might have colonized South America (see also Fig. 4).

[a]Analyses in which no constraints were enforced.

[b]Analyses in which a constraint of 5.0 my was enforced for the age of the most recent common ancestor of the paramo clade.

Europe and, in the case of Valerianaceae, into North America and subsequently into Central and South America (Bell, 2004b).

One remarkable biogeographic pattern concerns the placement of North American *Plectritis* with South American taxa. This North American/South American disjunction has been noted for a number of taxa (Raven and Axelrod, 1974), such as *Linanthus* and *Gilia* of the Polemoniaceae. There are currently five recognized species of *Plectritis*, four that occur in western North America, ranging from Baja California, Mexico, to British Columbia, Canada. One species of *Plectritis*, *P. samolifolia*, has been described from Chile, but we were unable to sample it for this study. In any case, our results support dispersal within this clade from western South America to western North America.

The topology estimated from the combined data in the Bayesian analysis suggests that Valerianaceae potentially made two initial entries into South America: one represented by the clade containing *Valeriana decussata* ssp. *polemoniodes* + *V. laurifolia* (colonizing South America no earlier than divergence at node *Z*; Fig. 4), and a second represented by the remaining South American taxa sampled in this study (at or some time after node *Y*, but prior to node *X*; Fig. 4). However, support for relationships in this area of the tree is not very strong. In any case, the data presented here do not support the three-colonization events hypothesized by Eriksen (1989; Fig. 1). Species of Eriksen's *Valeriana* section *Porteria* (*V. arborea* and *V. triphylla*), as well as *V. chaerophylloides*, appear to be nested within *Valeriana* subgen. *Phyllactis*, albeit weakly supported (Fig. 3).

The time frame of these colonization events does not appear to fit Eriksen's hypothesis. The closure of the Isthmus of Panama took place in the late Pliocene ca. 3.5–3.1 mya (Coates and Obando, 1996). This connection of the two continents clearly served as a major conduit for plant and animal migration in the recent past (Burnham and Graham, 1999). Our data, however, do not support that this was the pathway used by the species of Valerianaceae that first colonized South America. Even when the High-Andean species are constrained to be no older than 5.0 my, the dates inferred for the colonization of South America considerably pre-date the final connection of North and South America (Table 1). This would seem to imply that Valerianaceae arrived in South America, most likely via long-distance dispersal from Central America, at some time during the late Miocene to early Pliocene (based on analyses of the combined data), when the northwestern Andes were not at their current elevations. However, several widespread taxa, such as *V. scandens* and *V. urticifolia*, that are now widespread across North and South America and occur in low- to mid-elevation

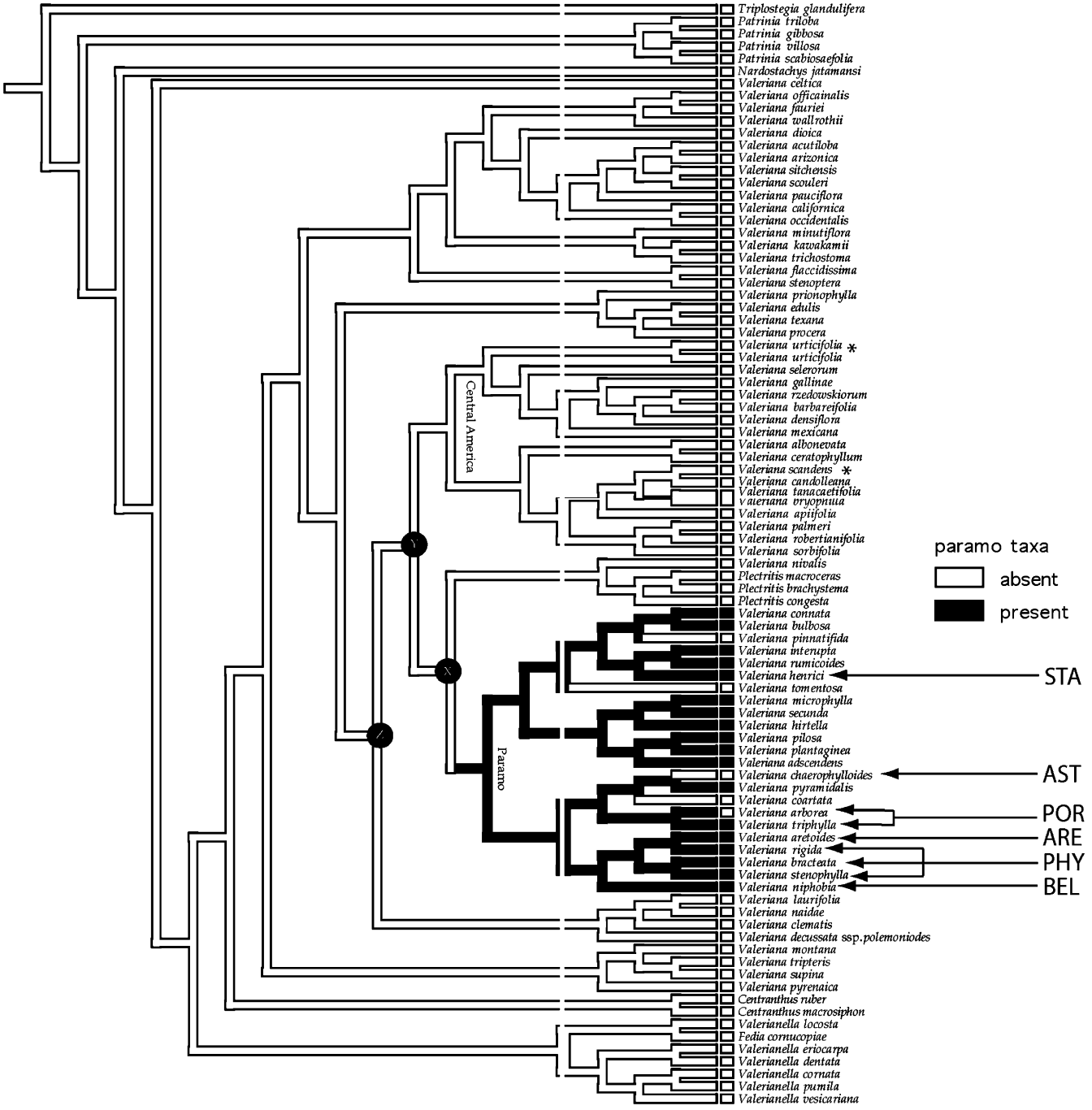

Fig. 4. Maximum likelihood tree topology inferred during Bayesian analysis of the combined nuclear and chloroplast data via MCMC. Maximum parsimony reconstruction of presence (black)/absence (white) from the paramo habitat using DELTRAN optimization. Nodes *X*, *Y*, *Z* represent several possible times at which Valerianaceae might have colonized South America (see also Table 1). * = taxa distributed in North and South America; ARE = *Aretiastrum*; AST = *Astrephia*; BEL = *Belonanthus*; PHY = *Phylactis*; POR = section *Porteria*; STA = *Stangea*.

forest habitats, may well have colonized South America via the isthmus at a more recent time. Age estimates are needed for other high-elevation plant clades that moved from North to South America, to test whether or not such movements corresponded closely with the formation of the land connection.

Valerianaceae in the alpine region

Species of Valerianaceae are a common component of the alpine flora around the Northern Hemisphere as well as in the Andes. As noted above, the basal lineages of Valerianaceae most likely originated and began to

diversify in the mountains of Asia, perhaps especially in the eastern Himalayas. This pattern of diversification is well documented in other groups within Dipsacales (e.g. Morinaceae; Bell and Donoghue, 2003).

Although the vast majority of the species of Valerianaceae in Europe are found in the Mediterranean region (*Valerianella*, *Fedia*, and *Centranthus*), several species of *Valeriana* (Fig. 2, Clade III) do occupy more montane habitats; occurring at high elevations in coniferous sub-alpine forests (*V. montana*) or alpine meadows primarily in the Alps (*V. supina*). Species of Valerianaceae in North America occur primarily in sub-alpine forests, often in moist meadows or along stream beds, but rarely above treeline.

The greatest diversity of alpine Valerianaceae is found in the Andes of South America, especially in the paramo habitat. The species of Valerianaceae that are endemic to the paramo of the northern Andes do form a clade in both the ML and MP analyses (Figs. 2–4), but support values in this area of the phylogeny are weak (Fig. 3). From these data, it cannot be ruled out that the alpine South American species of Valerianaceae do represent a single colonization of this habitat, followed by multiple (at least four) movements out of the paramo. A total of 55 species of Valerianaceae are endemic to the paramo habitat (Luteyn, 1999), of which only 19 (35%) were sampled in this study. If these species are monophyletic and have diversified since the rise of the northern Andes and the origination of this habitat type (ca. 5–3 mya; van der Hammen, 1974; Simpson, 1975; van der Hammen and Cleef, 1986), this would provide an excellent example of a rapid radiation, with something like 0.80–1.34 species/my based on a simple estimator: $[\ln(N)–\ln(N_0)]/T$; where $N =$ standing diversity, $N_0 =$ initial diversity (usually $N_0 = 1$), and $T =$ inferred clade age (Kendall, 1949; Moran, 1951). This method would be an underestimation if, as our phylogenies suggest, species of nonparamo endemics are nested within a clade that colonized the paramo habitat only once (Fig. 4). A similar rapid diversification has been suggested for *Halenia* (Gentianaceae) in the paramo (von Hagen and Kadereit, 2001; Kadereit and von Hagen, 2003). As pointed out by these authors, the colonization of the rising Andes of South America is comparable with the colonization of a newly formed island (e.g., Hawaii), allowing for species to take advantage of many unoccupied niches not already taken by native species. Examination of additional paramo endemics may show this high rate of speciation to be a more general trend.

Acknowledgments

We are grateful to Reto Nyffeler (Zurich Botanic Garden) for inviting CDB to take part in the symposium in Mobile, Alabama 2003, and for his care and patience in handling our manuscript. We thank F. Barrie (Missouri Botanical Garden), M. Dillon (FMNH), B. Eriksen (Göteborg University, Sweden), S. Madriñán (Universidad de Los Andes, Colombia), D. Neill (Missouri Botanical Garden and Herbario Nacional del Ecuador), and J. Schenk (Oregon State University) for their help in obtaining plant material. Thanks also to F. Barrie, B. Eriksen, N. Pyck, and R. Winkworth for valuable discussion, and to two anonymous reviewers for their feedback on an earlier version of this paper. This study was supported by an NSF Biotic Surveys and Inventory Grant to collect plants in the Hengduan Mountain region of China (DEB-9705795), and by a Doctoral Dissertation Improvement Grant (DEB-1014968).

References

Backlund, A.A., 1996. Phylogeny of the Dipsacales. Doctoral Dissertation. Dept. Systematic Botany, Uppsala University, Sweden.

Backlund, A.A., Donoghue, M.J., 1996. Morphology and phylogeny of the order Dipsacales. In: Backlund, A.A. (Ed.), Phylogeny of the Dipsacales, Doctoral Dissertation. Uppsala University, Sweden.

Backlund, A.A., Moritz, T., 1998. Phylogenetic implications of an expanded valepotriate distribution in the Valerianaceae. Biochem. Syst. Ecol. 26, 309–335.

Baldwin, B.G., Sanderson, M.J., 1996. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95, 9402–9406.

Bell, C.D., 2001. PORN*: A Hierarchical Likelihood Ratio Calculator for LINUX. http://www.phylodiversity.net/pornstar, Yale University, New Haven, CT, USA.

Bell, C.D., 2004a. Preliminary phylogeny of Valerianaceae (Dipsacales) inferred from nuclear and chloroplast DNA sequence data. Mol. Phylog. Evol. 31, 340–350.

Bell, C.D., 2004b. Phylogeny and biogeography of Valerina (Dipsacales). Ph.D. Thesis, Yale University, USA.

Bell, C.D., Donoghue, M.J., 2003. Phylogeny and biogeography of Morinaceae (Dipsacales) based on nuclear and chloroplast DNA sequences. Org. Divers. Evol. 3, 227–237.

Bell, C.D., Donoghue, M.J., 2005. Dating the Dipsacales: comparing models, genes, and evolutionary implications. Am. J. Bot. 92, 284–296.

Bell, C.D., Edwards, E.J., Kim, S.-T., Donoghue, M.J., 2001. Dipsacales phylogeny based on chloroplast DNA sequences. Harv. Pap. Bot. 6, 481–499.

Borsini, O.E., 1944. Valerianaceae. In: Genera et Species Plantarum Argentinarum. Descole, Buenos Aires, Argentina, pp. 275–372.

Brummitt, R.K., 1992. Vascular Plant Families and Genera. Royal Botanic Garden, Kew, UK.

Burnham, R.J., Graham, A., 1999. The history of neotropical vegetation: new developments and status. Ann. Missouri Bot. Gard. 86, 546–589.

Coates, A.G., Obando, J.A., 1996. The geologic evolution of the Central American isthmus. In: Evolution and Environments in Tropical America. Chicago Press, Chicago, IL, USA, pp. 21–56.

Cronquist, A., 1988. The Evolution and Classification of Flowering Plants. New York Botanical Gardens, Bronx, NY, USA.

Donoghue, M.J., Olmstead, R.G., Smith, J.F., Palmer, J.D., 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 79, 333–345.

Donoghue, M.J., Eriksson, T., Reeves, P.A., Olmstead, R.G., 2001. Phylogeny and phylogenetic taxonomy of Dipsacales, with special reference to *Sinadoxa* and *Tetradoxa* (Adoxaceae). Harv. Pap. Bot. 6, 459–479.

Donoghue, M.J., Bell, C.D., Winkworth, R.C., 2003. The evolution of reproductive characters in Dipsacales. Int. J. Plant Sci. 164, S453–S464.

Downie, S.R., Palmer, J.D., 1992. Restriction site mapping of the chloroplast DNA inverted repeat: a molecular phylogeny of the Asteridae. Ann. Missouri Bot. Gard. 79, 266–283.

Doyle, J.J., Doyle, J.D., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Eriksen, B., 1989. Note on the generic and infrageneric delimitation in the Valerianaceae. Nord. J. Bot. 9, 179–187.

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

Felsenstein, J., 1984. Distance methods for inferring evolutionary trees: a justification. Evolution 38, 16–24.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Graebner, P., 1906. Die Gattungen der natürlichen Familie der Valerianaceae. Bot. Jahrb. Syst. 37, 464–480.

von Hagen, K.B., Kadereit, J.W., 2001. The phylogeny of *Gentianella* (Gentianaceae) and its colonization of the southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. Org. Divers. Evol. 1, 61–79.

van der Hammen, T., 1974. The Pleistocene changes of vegetation and climate in tropical South America. J. Biogeogr. 1, 3–26.

van der Hammen, T., Cleef, A.M., 1986. High Altitude Tropical Biogeography. Oxford University Press, New York, NY, USA.

Hidalgo, O., Garnatje, T., Susanna, A., Mathez, J., 2004. Phylogeny of Valerianaceae based on *mat*K and ITS markers, with reference to *mat*K individual polymorphisms. Ann. Bot. 93, 283–293.

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

Huelsenbeck, J.P., Larget, B., Swofford, D.L., 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154, 1879–1892.

Kadereit, J.W., von Hagen, K.B., 2003. The evolution of flower morphology in Gentianaceae-Swertiinae and the roles of key innovations and niche width for the diversification of *Gentianella* and *Halenia* in South America. Int. J. Plant Sci. 164, S441–S452.

Kendall, D.G., 1949. Stochastic processes and population growth. J. R. Statist. Soc. B 11, 230–264.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kishino, H., Thorne, J.L., Bruno, W.J., 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18, 352–361.

Lancucka-Srodoniowa, M., 1967. Two new genera: *Hemiptelea* Planch. and *Weigela* Thurnb. in the younger Tertiary of Poland. Acta Palaeobot. 8, 1–17.

Langley, C.H., Fitch, W., 1974. An estimation of the constancy of the rate of molecular evolution. J. Mol. Evol. 3, 161–177.

Larsen, B.B., 1986. A taxonomic revision of *Phyllactis* and *Valeriana* sect *Bracteata* (Valerianaceae). Nord. J. Bot. 6, 427–446.

Luteyn, J.L., 1999. Paramos: a Checklist of Plant Diversity, Geographical Distribution, and Botanical Literature. New York Botanic Garden Press, Bronx, NY, USA.

Maddison, W.P., Maddison, D.R., 2000. MacClade 4: Interactive Analysis of Phylogeny and Character Evolution. Sinauer, Sunderland, MA, USA.

Manchester, S.R., Donoghue, M.J., 1995. Winged fruits of Linnaeeae (Caprifoliaceae) in the Tertiary of western North America: *Diplodipelta* gen. nov. Int. J. Plant Sci. 156, 709–722.

Meyer, F.G., 1951. *Valeriana* in North America and the West Indies (Valerianaceae). Ann. Missouri Bot. Gard. 38, 377–503.

Moran, P.A., 1951. Estimation methods for evolutive processes. J. R. Statist. Soc. B 13, 141–146.

Ozaki, K., 1980. Late Miocene Tatsumitoge flora of Tottori Prefecture, southwest Honshu, Japan. Scientific Report, Yokohama National University, Section 22, pp. 40–42.

Pyck, N., Smets, E., 2000. A search for the position of the seven-son flower (*Heptacodium*, Dipsacales): combining molecular and morphological evidence. Plant Syst. Evol. 225, 185–199.

Pyck, N., Roels, P., Smets, E., 1999. Tribal relationships in Caprifoliaceae: evidence from a cladistic analysis using *ndh*F sequences. Syst. Geogr. Plants 69, 145–159.

Pyck, N., van Lysebetten, A., Stessens, J., Smets, E., 2002. The phylogeny of Patrinieae sensu Graebner (Valerianaceae) revisited: additional evidence from *ndh*F sequence data. Plant Syst. Evol. 233, 29–46.

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

Raymundez, M.B., Mathez, J., Enrech, N.X.de., Duduisson, J.-Y., 2002. Coding of insertion-deletion events of the chloroplast intergene *atp*B-*rbc*L for the phylogeny of the Valerianeae tribe (Valerianaceae). C. R. Biol. 325, 131–139.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.

Sanderson, M.J., 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.

Sang, T., Crawford, D.J., Stuessy, T.F., 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. 84, 1120–1136.

Simpson, B.B., 1975. Pleistocene changes in the flora of the high tropical Andes. Paleobiology 1, 273–294.

Smith, J.M.B., Cleef, A.M., 1988. Composition and origin of the world's tropical alpine flora. J. Biogeogr. 15, 631–645.

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

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

Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15, 1647–1657.

Urbatsch, L.E., Baldwin, B.G., Donoghue, M.J., 2000. Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) and chloroplast DNA restriction site data. Syst. Bot. 25, 539–565.

Weberling, F., 1970. Familie Valerianaceae. In: Hegi, G. (Ed.), Illustrierte Flora von Mitteleuropa. C. Hanser, Munich, Germany, pp. 97–176.

White, T.J., 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.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, CA, USA, pp. 315–322.

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

Yang, Z., 2000. Phylogenetic Analysis by Maximum Likelihood (PAML), 3.0 ed. University College London, UK.

Yoder, A.D., Yang, Z., 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. 17, 1081–1190.

Young, N.D., Steiner, K.E., dePamphilis, C.W., 1999. The evolution of parasitism in Scrophulariaceae/Orobanchaceae: plastid gene sequences refute an evolutionary transition series. Ann. Missouri Bot. Gard. 86, 876–893.

Zhang, W.-H., Chen, Z.-D., Li, J.-H., Chen, Y.-C., Tang, H.-B., 2003. Phylogeny of the Dipsacales s.l. based on chloroplast *trn*L and *ndh*F sequences. Mol. Phylog. Evol. 26, 176–189.