Molecular Phylogenetics and Evolution 48 (2008) 408–421

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Diversification of *Lupinus* (Leguminosae) in the western New World: Derived evolution of perennial life history and colonization of montane habitats

Christopher S. Drummond *

Department of Biology, Georgetown University, 406 Reiss Sciences Building, 37th and O Streets, NW, Washington, DC 20057, USA

ARTICLE INFO

Article history:
Received 28 June 2007
Revised 29 February 2008
Accepted 4 March 2008
Available online 14 March 2008

Keywords:
Biogeography
Chloroplast DNA
Elevation
Life history
Lupinus
Leguminosae
Radiation

ABSTRACT

Previous phylogenetic studies of *Lupinus* (Leguminosae) based on nuclear DNA have shown that the western New World taxa form a monophyletic group representing the majority of species in the genus, with evidence for high rates of recent diversification in South America following final uplift of the Andes 2–4 million years ago (Mya). For this study, three regions of rapidly evolving non-coding chloroplast DNA (*trn*L intron, *trn*S–*trn*G, and *trn*T–*trn*L) were examined to estimate the timing and rates of diversification in the western New World, and to infer ancestral states for geographic range, life history, and maximum elevation. The western New World species (5.0–9.3 Mya, 0.6–1.1 spp./My) comprise a basally branching assemblage of annual plants endemic to the lower elevations of western North America, from which two species-rich clades are recently derived: (i) the western North American perennials from the Rocky Mountains, Great Basin, and Pacific Slope (0.7–2.1 Mya, 2.0–5.9 spp./My) and (ii) the predominantly perennial species from the Andes Mountains of South America and highlands of Mexico (0.8–3.4 Mya, 1.4–5.7 spp./My). Bayesian posterior predictive tests for association between life history and maximum elevation demonstrate that perennials are positively correlated with higher elevations. These results are consistent with a series of one or more recent radiations in the western New World, and indicate that rapid diversification of *Lupinus* coincides with the derived evolution of perennial life history, colonization of montane habitats, and range expansion from North America to South America.

1. Introduction

Recent radiations provide compelling evidence for the role of ecological landscapes in speciation, contributing to rapid diversification via a combination of adaptive and/or non-adaptive responses of populations to spatially and temporally heterogeneous environments (Givnish, 1997; Schluter, 2000). Many of the best-known examples of radiations have been observed in isolated habitats that exhibit substantial ecological variability, such as island archipelagos or lake systems, where founder events are followed by exploitation of novel resources, leading to niche diversification, reproductive isolation, and ecological speciation (e.g., Grant, 1986; Schluter and McPhail, 1993; Baldwin, 1997; Jackman et al., 1997; Albertson et al., 1999). In contrast, mainland radiations have been less well-documented.

Molecular estimates of divergence times based on the nuclear DNA (nucDNA) loci ITS1+2 and *LEGCYC1A* (Hughes and Eastwood, 2006) have shown that a derived clade of *Lupinus* in the Andes Mountains of South America underwent a continental-scale mainland radiation beginning ca. 1.5 million years ago (Mya), with rates of recent diversification (2.5–3.7 spp./My) comparable to the highest known rates of speciation in land plants (e.g., Böhle et al., 1996; Baldwin and Sanderson, 1998; Richardson et al., 2001a,b; von Hagen and Kadereit, 2001; Malcomber, 2002; Klak et al., 2003; Verboom et al., 2003; Kay et al., 2005; Good-Avila et al., 2006). Notably, the Andean *Lupinus* exhibit a diversity of phenotypes, ranging from dwarf acaulescent rosettes to large woody shrubs, but are predominantly perennial and typically restricted to elevations >3000 m (Smith, 1938–1951; Hughes and Eastwood, 2006). In the absence of clear evidence for morphological or physiological traits that would promote diversification, the Andean radiation has been hypothesized to result from a combination of geographic subdivision, ecological opportunity, and resource heterogeneity defined by steep altitudinal and climatological gradients created following final uplift of the Andes 2–4 Mya (Hughes and Eastwood, 2006).

ITS1+2 and *LEGCYC1A* data also show that the Andean and Mexican species are derived from a partially unresolved assemblage of western North American taxa (Aïnouche and Bayer, 1999; Aïnouche et al., 2004; Hughes and Eastwood, 2006). The western North American *Lupinus* are both species-rich and morphologically diverse, including a lesser number of annuals found primarily in lowland deserts and grasslands, and a larger number of herbaceous

* Present address: Center for Research on Invasive Species and Small Populations, Department of Fish and Wildlife Resources, University of Idaho, P.O. Box 441141, Moscow, ID 83844-1141, USA. Fax: +1 208 885 9080.
E-mail address: cdrummon@uidaho.edu

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.03.009

and woody perennials extending from coastal to subalpine elevations (Smith, 1944; Barneby, 1989; Riggins and Sholars, 1993; Isely, 1998). The evident diversity of the western North American species indicates that the Andean clade might not represent the only example of a recent radiation in *Lupinus*, and further suggests that perennial life history may have facilitated colonization of novel environments, leading to range expansion and rapid diversification of the genus.

To examine the relationship between geographic distribution, life history, and elevation, a phylogeny of *Lupinus* was inferred from three regions (*trn*L intron, *trn*S–*trn*G, and *trn*T–*trn*L) of non-coding chloroplast DNA (cpDNA) using Bayesian Markov chain Monte Carlo (MCMC) and maximum parsimony (MP) methods. Divergence times and diversification rates were estimated from molecular dates obtained by penalized likelihood based on subsampling of trees from the posterior distribution of Bayesian MCMC analyses. Phylogenetic comparative analyses were conducted using Bayesian stochastic mapping to reconstruct ancestral character states and test for correlations between life history and maximum elevation.

2. Methods

2.1. Study system

Lupinus (Leguminosae) forms a monophyletic genus of approximately 250 species, in addition to numerous subspecific taxa (Smith, 1938–1951; Dunn, 1984; Aïnouche and Bayer, 1999; Roskov et al., 2005; Hughes and Eastwood, 2006). A relatively small number of species are found in the Mediterranean and north Africa (~12 spp.; Gladstones, 1998), eastern South America (~24 spp.; Planchuelo-Ravelo, 1984; Planchuelo and Dunn, 1984; Monteiro and Gibbs, 1986), and eastern North America (~8 spp.; Isely, 1998). However, the majority of species are endemic to the Rocky Mountains and Pacific slope of western North America (~88 spp.; Smith, 1944; Barneby, 1989; Riggins and Sholars, 1993; Isely, 1998), the Andes of western South America (~85 spp.; Smith, 1938–1951; Hughes and Eastwood, 2006), and the highlands of Mexico and Central America (~30 spp.; Smith, 1938–1951; Dunn and Harmon, 1977; C. Hughes, personal communication).

In the most completely sampled phylogenetic study of *Lupinus* to date (Hughes and Eastwood, 2006) combined analyses of the nuclear DNA (nucDNA) loci ITS1+2 and *LEGCYC1A* divided the genus into two principal clades: (i) the Old World species sister to the unifoliolate species from eastern North America and (ii) the eastern South American species and *Subcarnosi* from Texas sister to the western New World species. Other studies based on ITS1+2, *LEGCYC1A*, *LEGCYC1B*, and/or *rbc*L (Käss and Wink, 1997; Aïnouche and Bayer, 1999; Aïnouche et al., 2004; Ree et al., 2004) have supported the monophyly of the western New World species, which exhibit low levels of sequence variation consistent with recent common ancestry.

2.2. Taxon sampling and DNA sequencing

A total of 93 specimens of *Lupinus* were sampled from across the geographic range of the genus (Appendix), including 62 unique species representing each of the major clades identified in previous phylogenetic studies. Because the nomenclature of *Lupinus* remains in flux for many subspecies and varieties, subspecific taxa were treated as single species complexes. Five outgroup taxa were sampled from related genera in the Leguminosae (*Argyrolobium*, *Chamaecytisus*, *Genista*, *Laburnum*, and *Spartium*).

DNA was extracted from fresh, frozen, or silica gel-dried leaf tissue using DNeasy (Qiagen) plant extraction kits. The *trn*S–*trn*G intergenic spacer was amplified by PCR following a protocol modified from Hamilton (1999). The *trn*T–*trn*L intergenic spacer and *trn*L intron were amplified as part of a larger fragment comprising the entire *trn*T–*trn*L–*trn*F region following Taberlet et al. (1991). PCR products were sequenced in both directions using the original PCR primers and internal sequencing primers, obtaining coverage of all regions except the first ~150 bp of the *trn*T–*trn*L spacer closest to the *trn*T exon and the first ~50 bp of the *trn*S–*trn*G spacer closest to the *trn*S exon. Internal primers included B, C, and D from Taberlet et al. (1991) and the following new sequencing primers designed for *Lupinus*: SGint1 GGATTGTGTAAGAATCCACAG (*trn*S–*trn*G); SGint2 GAAATACATTCAACTAAATCTTTG (*trn*S–*trn*G); ABint2 GCTAAGCTACTAAGCTACCG (*trn*T–*trn*L). Chromatograms were edited in Sequencher 4.2 (GeneCodes) and resequenced if necessary. GenBank accession numbers are listed in Appendix.

Sequences were aligned in ClustalX 1.8.2.1 (Thompson et al., 1997) and manually adjusted in MacClade 4.06 (Maddison and Maddison, 2003) according to established principles for identifying microstructural features commonly found in non-coding cpDNA (e.g., Graham et al., 2000; Kelchner, 2000; Simmons and Ochotorena, 2000). Ambiguously aligned regions and exons were excluded from the aligned data matrix (available online at http://www.treebase.org/treebase/). Preliminary phylogenetic analyses were conducted with indels scored as binary characters in SeqState 1.32 (Müller, 2005a) under a modified simple gap-coding scheme (Müller, 2005b). However, indels resulted in over-parameterized Bayesian MCMC models with unstable posterior parameter estimates (data not shown) and did not yield appreciable differences in topological resolution or clade support for either Bayesian MCMC or MP methods. Therefore, analyses of the final data matrix treated indels as missing data.

2.3. Phylogenetic inference

Nucleotide substitution models were examined using three partitioning schemes: (i) concatenated data from all non-coding regions (one partition); (ii) intergenic spacers versus intron (two partitions); and (iii) individual non-coding regions (three partitions). Substitution models for each partition were selected using the Akaike information criterion corrected for small sample size (AIC_c) in MrAIC.pl 1.4.3 (Nylander, 2007b) based on log-likelihood ($\ln L$) scores of maximum-likelihood (ML) topologies inferred in PHYML 2.4.4 (Guindon and Gascuel, 2003). Partitioning schemes were evaluated with Bayes factors (Nylander et al., 2004), which measure the relative support for models using information theoretic criteria (Kass and Raftery, 1995). Bayes factors are calculated as the ratio $B_{ij} = f(X|M_i)/f(X|M_j)$, which can be estimated using $2\ln B_{ij}$ or twice the difference in the harmonic mean of marginal $\ln L$ scores from Bayesian MCMC sampling (Newton and Raftery, 1994). Interpretation of Bayes factors was based on the following guidelines for $2\ln B_{ij}$: <0 support for M_j; 0–2 difference barely worth mentioning; 2–5 positive support for M_i; 5–10 strong support for M_i; >10 very strong support for M_i (Raftery, 1995).

Bayesian MCMC analyses were conducted in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). Analyses included two paired runs, each with one cold chain and three heated chains (temp = 0.05) and a sampling interval of 100 generations. Since cpDNA generally behaves as a single non-recombining, uniparentally inherited organellar genome in angiosperms (Birky, 2001), topologies were linked across partitions under a uniform prior. Because tree lengths and posterior probabilities for clades have been shown to be sensitive to the default branch length prior in MrBayes (Yang and Rannala, 2005; Marshall et al., 2006), the exponential prior for mean branch length ($1/\lambda$) was set to $\lambda = 500$, equivalent to the likelihood estimate of $1/\lambda = 0.002$ from the ML topology obtained under the best-fit model of nucleotide substitution. Rate