AoB PLANTS http://aobplants.oxfordjournals.org/

Open access – Research article

Biogeography and divergence times of genus *Macroptilium* (Leguminosae)

Shirley M. Espert[1,2] and Alicia D. Burghardt[1*]

[1] Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Received: 24 June 2010; **Returned for revision:** 23 August 2010; **Accepted:** 14 October 2010; **Published:** 19 October 2010

Citation details: Espert SM, Burghardt AD. 2010. Biogeography and divergence times of genus *Macroptilium* (Leguminosae). *AoB PLANTS* **2010**: plq018, doi:10.1093/aobpla/plq018

Abstract

Background and aims *Macroptilium* is a herbaceous legume genus with 18 currently accepted species, seven of them with economic importance due to their use as forage, green fertilizer and in medicine. The genus is strictly American, with an unknown biogeographic history. The aim of this study was to infer a biogeographic pattern of *Macroptilium* and to estimate its divergence times, using sequences from the nuclear ribosomal DNA internal transcribed spacers.

Methodology To study the historical biogeography of *Macroptilium*, two approaches were used: area optimization on a previously obtained phylogeny and a dispersal–vicariance analysis. Divergence times were calculated by Bayesian methods.

Principal results The analyses revealed that *Macroptilium* has its origin in the middle Pliocene, with an estimated age that ranges from 2.9 to 4 million years. The biogeographic analyses placed its origin in South America, specifically on the Chaquean sub-region, where most of the cladogenetic events of the genus took place.

Conclusions *Macroptilium* constitutes a further example of the geographic pattern displayed by numerous Neotropical taxa that moved north from South America to dominate the Central American lowlands after the land connection across the Isthmus of Panama was established.

Introduction

Macroptilium (Benth.) Urban is a herbaceous legume genus with 18 currently accepted species, seven of them used as forage, green fertilizer and in medicine (Barbosa Fevereiro 1986). It was first described by Bentham (1837) as a section of the common bean genus *Phaseolus*, but several years later Urban (1928) considered that the taxa deserved genus ranking, due to differences with remaining *Phaseolus* species. In this new genus, members of two formerly *Phaseolus* sections, Macroptilium and Microcochle, were included.

Based on morphological, biochemical and molecular data, recent studies support the monophyly of *Macroptilium* (Espert *et al.* 2007). Some aspects of the floral morphology, such as the upper teeth of the calyx, the shape of the style and stigma, and the size of the wing petals (bigger than the standard), are the main contributors

[2] Present address: Department of Microbiology and Immunology, University of Miami Miller School of Medicine, Miami, FL 33136, USA.
* Corresponding author's e-mail address: alibu@bg.fcen.uba.ar

AoB PLANTS Vol. 2010, plq018, doi:10.1093/aobpla/plq018, available online at www.aobplants.oxfordjournals.org

to the differentiation of the genus from the related species of Phaseolinae. Two monophyletic groups are recovered within *Macroptilium*, with several synapomorphies and good support values (Espert *et al.* 2007). Each of them is composed of species assigned by Lackey (1983) to the sections Macroptilium and Microcochle. Both these sections are characterized by attributes of the inflorescence, calyx, stigma and pollen.

The 18 species of *Macroptilium* are distributed between the southern part of the USA and middle Argentina. Mexico, with nine taxa, is the major centre of variability of the genus in North America, while Brazil and Paraguay, with 12 taxa, are the main diversity centres in South America. The genus is one of the eight genera of Phaseolinae that grew exclusively on the American continent. The current distribution observed in *Macroptilium* has never been explained. Little is known about the biogeographic history of other Phaseolinae species, except for the large American genus *Phaseolus* (Delgado-Salinas *et al.* 2006) and the African *Wajira* (Thulin *et al.* 2004).

In this work, we conducted a biogeographic analysis and estimated divergence times of the species of Macroptilium, using sequences from the nuclear ribosomal DNA internal transcribed spacers (ITSs). Internal transcribed spacer regions showed a high substitution rate, which makes them very useful for the study of molecular evolution between related species. However, some organisms may harbour ITS paralogues with different evolutionary histories, which could lead to a misinterpretation of the results. Mistaken labelling of paralogues as orthologues can be problematic for divergence time studies because gene duplications often long pre-date speciation events that separate the taxa of interest, leading to a downward bias in age estimates (Sanderson *et al.* 2004).

To infer a biogeographic pattern of genus *Macroptilium*, we used two approaches: area optimization on a previously obtained phylogeny and a dispersal–vicariance analysis. The expected divergence between homologous sequences is determined by the time since common ancestry and the rate at which differences have been accumulating (Sanderson *et al.* 2004). Many methods are available that estimate dating times. Since none of these methods is clearly superior to any other, the selection of one model over another is problematic. Once the molecular clock for the data is rejected, it is necessary to choose between a handful of methods that address rate heterogeneity (see Rutschmann 2006 for a detailed review). Since our input data is only a cladogram, and we want to estimate branch lengths in order to propose divergence times, we have to choose between a group of methods that only require a topology for the analysis. The Bayesian dating method implemented in Multidivtime (Thorne *et al.* 1998; Kishino *et al.* 2001; Thorne and Kishino 2002) uses a fully probabilistic and high parametric model to describe the change in evolutionary rate over time and uses the Markov chain Monte Carlo (MCMC) procedure to derive the posterior distribution of rates and times. The Bayesian method represents a promising alternative for estimating rates and divergence times in the absence of rate constancy (Magallon 2004). This is due to the fact that, in simulation studies, this method leads to age estimates that agree with independent fossil data. The problem with this technique is that it relies on the topology provided by the user (Rutschmann 2006).

The major aim of our work was to provide new evidence to advance understanding of the biogeography of genus *Macroptilium*. In addition, divergence times of *Macroptilium* species were estimated.

Materials and methods

Taxon sampling and topology

The molecular sampling included the sequence data from Espert *et al.* (2007). Fifteen of the 18 species of the genus *Macroptilium* were analysed, along with five related species belonging to subtribe Phaseolinae, which formed the outgroup. Sequences of the ITS regions of nuclear DNA (ITS-1 and ITS-2) from the studied species were retrieved from GenBank [see Additional information]. The possible existence of paralogous sequences for the ITS-1 and ITS-2 regions were studied in Espert *et al.* (2007), by analysing the length of the sequences, the G + C content, the presence and length of conserved domains, secondary structures and free energy, as suggested by Mayol and Rossello (2001). Given the results, none of the sequences is paralogous, and they appear to be functional regions of the nuclear genome.

Sequence alignments were made with the DIALIGN program (Morgenstern *et al.* 1998), using a threshold value of 10. Data matrices used in this study are deposited in TreeBase (PIN number 4749; S.M.E., submitter).

Three species, *M. pedatum*, *M. martii* and *M. monophyllum*, were excluded from the molecular dating analysis because the material was not available. However, they were included in the biogeographic study.

The topology used in this work is one of the eight most parsimonious trees obtained by Espert *et al.* (2007), in their combined analysis of morphological, biochemical and molecular data of the species mentioned above (Fig. 1). Combined and individual dataset analyses gave similar topologies.

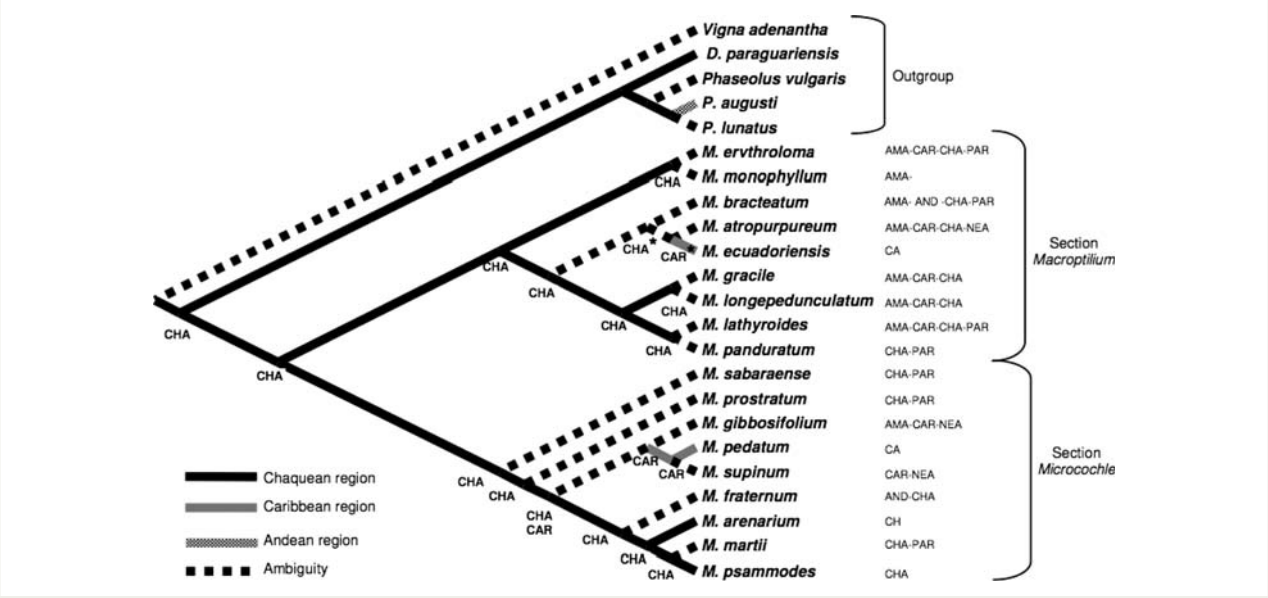

Fig. 1 One of the eight most parsimonious trees obtained by Espert *et al.* (2007). Areas where species are distributed are listed next to their names. Results of the optimization are represented with different colours of the tree branches. DIVA areas' assignation is shown on each node. An asterisk indicates ambiguous reconstructions of nodes. AMA, Amazonian sub-region; AND, Andean region; CAR, Caribbean sub-region; CHA, Chaquean sub-region; NEA, Nearctic region; PAR, Paranaense sub-region.

Biogeographic analysis

The areas of *Macroptilium* species and their phylogeny were used to reconstruct their distribution history. Areas were assigned according to the American regional division devised by Morrone (2001). Species distribution was gathered from field trips, literature and herbarium specimens. All the locations were marked on a map, and then the biogeographic areas proposed by Morrone (2001) were circumscribed to the same map. The intersection of the locations with the areas resulted in the final assignation, and is shown in Fig. 1. The cladogram of Espert *et al.* (2007) was used as the input tree. Two methods were used; in the first place, areas were treated as an unordered multistate character and optimized onto the topology, using Winclada (Nixon 2002). Secondly, the program Dispersal Vicariance Analysis (DIVA; Ronquist 1996) was used. Dispersal Vicariance Analysis is based on the dispersal–vicariance approach (Ronquist 1997), which consists of the optimization of a three-dimensional cost matrix derived from a simple biogeographic model. It favours vicariance events, by minimizing the number of dispersals and extinctions.

With DIVA, it is possible to reconstruct the distribution history of individual groups in the absence of a general hypothesis of area relationships. The model used in DIVA does not assume anything about the shape or the existence of general area relationships.

Molecular dating

The molecular clock behaviour for the sequences was tested with the likelihood ratio test (LRT), which compares the likelihood of a topology with and without constant rate assumption. The program Modeltest 3.7 (Posada and Crandall 1998) was used to produce an appropriate nucleotide substitution model for our data. The Tajima and Nei with Gamma distribution (TrN + G) model was chosen, and then the likelihood values assuming molecular clock (L_0, null hypothesis) and with variable lengths for each branch (L_1, alternative hypothesis) were estimated using PAUP* 4.0b10 (Swofford 2002). The significance of the LRT statistic was approximated using the chi distribution.

Because the LRT was highly significant, we used one of the approaches that have been proposed for use under a non-molecular clock scenario: the Bayesian method (Thorne and Kishino 2002). The ages of speciation events were estimated using the Bayesian relaxed molecular clock as implemented in Multidivtime, following the step-by-step manual compiled by Rutschmann (2005). First, we estimated the model parameters with BASEML of the PAML package (Yang 1997). Secondly, the branch lengths for the chosen cladogram and a variance–covariance matrix were calculated using Estbranches, a component of the Multidivtime package, under the F84 + Γ substitution model (as suggested in

the Estbranches manual). Finally, we ran a MCMC for estimating the mean posterior divergence times on nodes with associated standard deviations from the variance–covariance matrix produced by Estbranches. The Markov chain was sampled 10 000 times every 100 cycles after a burn-in stage of 100 000 cycles. The tree was calibrated using the ages reported in Thulin *et al.* (2004): the basal node of the chosen cladogram (Espert *et al.* 2007) has an estimated time of 4.5–6.2 My (million years). These ages were calculated by setting the root of the Phaseolinae clade to 10.7 My, a value obtained after a large-scale analysis using 12 minimum age constraints derived from the fossil record of the Leguminosae (Lavin *et al.* 2005). The analysis was run twice in order to compare the results and establish whether convergence had occurred.

Results

Biogeographic analysis

Phylogenetic relationships of *Macroptilium* species based on three different sources of data (morphology, seed protein patterns and DNA sequence), along with the optimization of the geographic regions, are shown in Fig. 1.

Biogeographic patterns were also reconstructed with DIVA. The areas of the ancestral nodes are presented in Fig. 1. Only one vicariant event and several dispersals are required to explain the actual distribution of the 18 species of *Macroptilium*. Two of the 17 nodes were ambiguously reconstructed, they are all indicated in Fig. 1.

Molecular dating

Figure 2 shows the phylogenetic relationships of *Macroptilium* species, where the different branch length represents the variation in the substitution rate. Both dating runs using the Bayesian approach gave similar results; hence, the Markov chain reached its stationary distribution. The estimates after calibrating the tree are given in Table 1. *Macroptilium* crown clade has an average estimated age of 3.466 My. The ancestor of section Microcochle has an earlier origin compared with that of section Macroptilium, since its estimated age is 3.127 My, 400 000 years older than the other section.

Discussion

Recent studies on subtribe Phaseolinae established that the New World genera, including *Macroptilium*, form a monophyletic clade with an *estimated* age of 8 My (Thulin *et al.* 2004). Our analyses revealed that Macroptilium has its origin in the middle Pliocene, with an estimated age that ranges from 4 to 2.9 My. Unfortunately, no fossil records of *Macroptilium* are available; therefore, the age and biogeographic inference are not supported by this type of data. The genus has a broad distribution along the American continent, but the biogeographic analyses placed its origin on South America, specifically on the Chaquean sub-region. This area comprises the north and centre of Argentina, southern Bolivia, the west and centre of Paraguay, and the centre and north-eastern Brazil (Morrone 2001). Geological evidence corroborates this observation, since the older species of *Macroptilium* that arose prior to the establishment of the Isthmus of Panama, ~3 My ago (Gentry 1982*a*), are strictly South American or have a wide distribution. On the other hand, North American species have a late origin, subsequent to the land connection between Central and South America. Even though an alternative hypothesis where the genera might have originated in North America with a subsequently southern migration cannot be ruled out, most angiosperms, and specifically legumes, have migrated from South America (Lavin and Luckow 1993).

The present distribution of the genus requires one vicariant event and several dispersals after the establishment of the common ancestor to all species. Most of the cladogenetic events of the genus took place on the Chaquean sub-region. The first of these events resulted on the ancestors of each of the two sections of the genus, Macroptilium and Microcochle.

Almost all the cladogenetic events on section Macroptilium occurred on the Chaquean sub-region, ~1.6 and 3 My ago. Species of this section have a wider distribution than those of section Microcochle, probably due to their habit: they are erect or climber plants, in contrast to the prostrate habit of their sister group. In addition, differences in the floral morphology of the taxa of the two sections (Espert *et al.* 2007) could also affect the pattern of distribution of the plants, by having different types of pollinators. Despite the fact that the greatest diversification on the section occurred on the Chaquean sub-region, almost all extant species also grow on the Amazonian sub-region, and might have originated there as well. In this area, changes in precipitation were pronounced, resulting in the reduction of the extent to scattered pockets, and providing optimal conditions for speciation (Gentry 1982*b*). Within section Macroptilium, an ambiguous reconstruction of the biogeographic pattern is found on the clade composed of *M. bracteatum*, *M. atropurpureum* and *M. ecuadoriensis* (Fig. 1). The distribution of the common ancestor of the group is ambiguously located on the Chaquean sub-region or in a broader area formed by this territory and

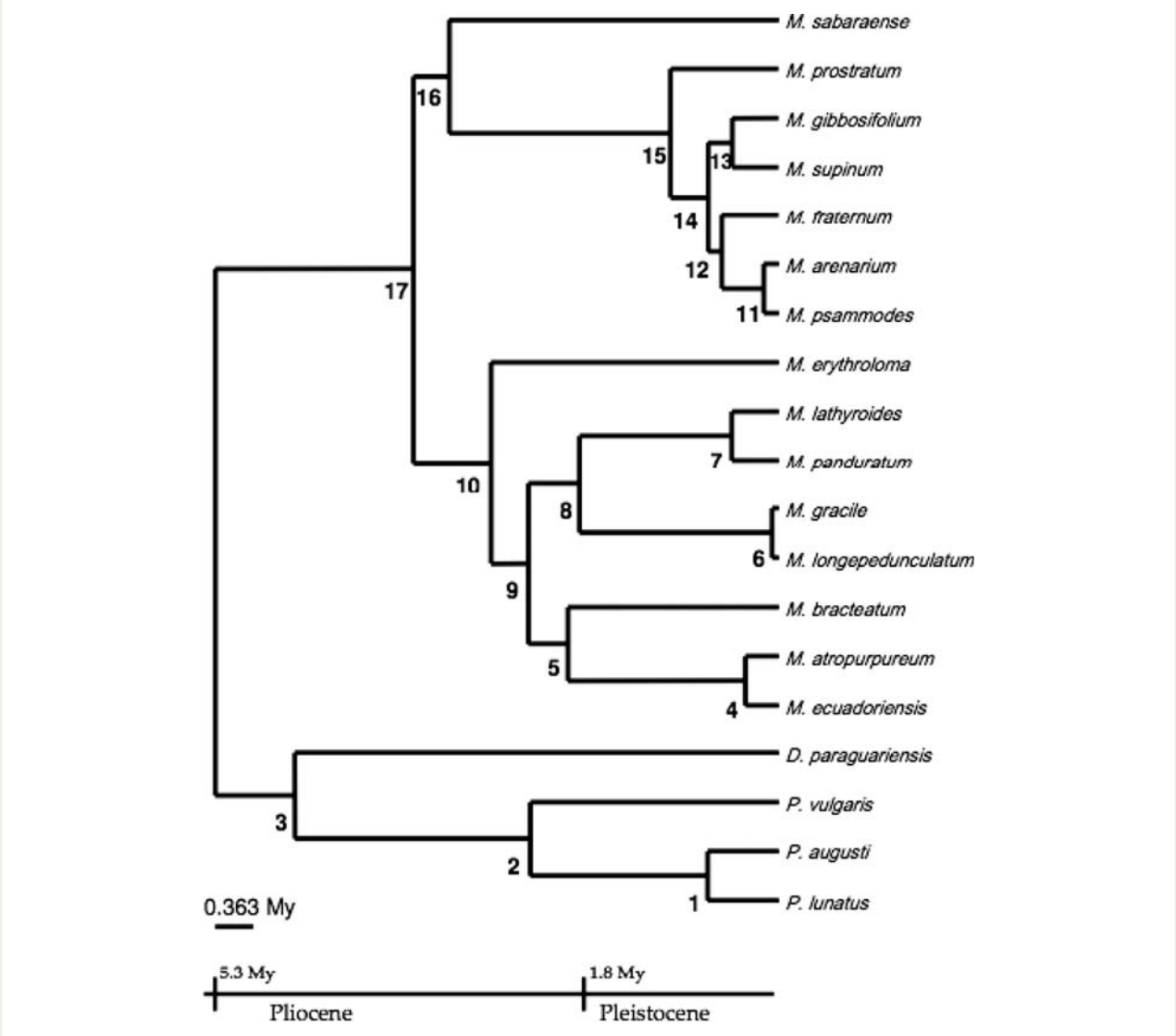

Fig. 2 Phylogram of *Macroptilium* species obtained with Multidivtime. Numbering of nodes is shown below branches.

the Caribbean one. The occurrence of the species on this latter area could be explained by the dispersal of some individuals from the Chaquean sub-region through the Andean region; but the most likely explanation would be the dispersal over the Amazonian area due to the existence of *M. bracteatum* and *M. atropurpureum* on this sub-region at the present time. *Macroptilium ecuadoriensis* has a restricted distribution to the Caribbean sub-region, and has an estimated age of 300 000 years. Despite recent studies having raised doubts about whether to consider this taxon as a species or as a variety of *M. atropurpureum* (Espert *et al.* 2007), we decided to treat it as a differentiated entity, until further analysis corroborates its status. However, the evidence presented here does not support the existence of *M. ecuadoriensis* as a species; rather, it is possibly evolving towards its establishment, but has not reached that status yet.

The only vicariant event inferred by the biogeographic analysis took place within section Microcochle, resulting in one ancestor on each sub-region by allopatric speciation. Long-distance dispersal took place from Chaco to the Caribbean sub-region. Since these two areas have no connection, the dispersal of individuals could have happened through the Amazonian sub-region, or through the Andean region. On both areas extant species of this section grow: *M. gibbosifolium* is located on the Amazonian area as well as on the Nearctic

Table 1 Age estimates of the interior nodes of the phylogeny of *Macroptilium*.

Node	Mean age (My)	Standard deviation (My)	95 % confidence range (My)
1	0.569/0.783	0.229/0.3151	0.236–1.122/0.325–1.5457
2	1.983/2.7313	0.477/0.6566	1.206–3.05/1.6609–4.2011
3	3.864/5.3208	0.577/0.7939	3.09–5.22/4.2569–7.1898
4	0.265/0.3647	0.200/0.2759	0.015–0.764/0.0211–1.0516
5	1.684/2.3191	0.335/0.4613	1.07–2.383/1.4736–3.2817
6	0.052/0.0713	0.051/0.0698	0.002–0.187/0.0021–0.2573
7	0.374/0.5154	0.178/0.2454	0.111–0.811/0.1525–1.1162
8	1.591/2.1905	0.336/0.4629	0.991–2.306/1.3644–3.1758
9	1.999/2.7540	0.332/0.4576	1.379–2.666/1.8994–3.6720
10	2.299/3.1654	0.325/0.4477	1.691–2.941/2.3292–4.0498
11	0.119/0.1649	0.102/0.1404	0.005–0.394/0.0075–0.5429
12	0.456/0.6274	0.208/0.2867	0.151–0.954/0.2077–1.3139
13	0.372/0.5121	0.197/0.2712	0.079–0.848/0.1100–1.1676
14	0.563/0.776	0.232/0.3188	0.226–1.134/0.3108–1.5609
15	0.864/1.1897	0.295/0.4061	0.408–1.538/0.5612–2.1186
16	2.631/3.6228	0.292/0.4020	2.072–3.168/2.8537–4.3629
17	2.916/4.0151	0.256/0.3531	2.47–3.332/3.4015–4.5885

Nodes are numbered according to Fig. 2. A minimum age of 4.5 My and a maximum of 6.2 My were used to calibrate the tree; these results are indicated on the left side and the right side, respectively.

and Caribbean sub-region, while *M. fraternum* grows on the Andean region and Chaquean sub-region. Because geographic information was gathered from the literature and herbarium specimens, probably leading to an underestimation of the actual record, we presume that other taxa never sampled may arise in one or both areas as well. In this genus, this problem is even worse due to the small size of the plants, which makes the detection of the individuals, and therefore sampling, difficult. Diversification on this section was fairly rapid, mostly during about a 1 My window, on the late Pleistocene.

Given the present distribution of the genus in small populations, it could be inferred that their ancestors displayed a similar structure pattern. If the Chaquean subprovince were a big open habitat originated by the uplift of the Andes in the first place, and the appearance of the Serra do Mar later (Iriondo 1999), many populations could have dispersed and established in new areas, genetic drift occurring then and resulting in several differentiated groups of individuals. These new populations occasionally migrated towards other areas (e.g. Amazonic and Caribbean sub-regions) where some other species, like *M. supinum* and *M. gibbosifolium*, arose also by stochastic phenomena. Biogeographic studies, based on other angiosperm families, conclude that most of the South American taxa moved north to completely dominate the Central American lowlands after the isthmian connection closed. Most of these invasions have been so recent that even at the specific level there has been little differentiation, and perhaps the northward migration is still taking place (Gentry 1982*a*). Many Neotropical families show a clear northward decrease in the number and the diversity of their species (Gentry 1982*b*). We therefore conclude that genus *Macroptilium* provides another example of this pattern of geographical distribution.

Conclusions and forward look

With an estimated age of 4–2.9 My, genus *Macroptilium* is shown to have originated in the Pliocene, prior to the establishment of the Isthmus of Panama. We propose a south to north migration of its species after the Isthmus was established. The addition of gene sequences, especially from the chloroplast, will help to corroborate the hypothesis proposed in this work, and to give more insights into the biogeographic history of this legume genus.

Additional information

The following additional information is available in the online version of this article –

List of taxa used in this study, along with the GenBank accession numbers for ITS sequences and its references.

Sources of funding

This work was funded by Universidad de Buenos Aires, Grants AX33 (2000–2001) and EX061 (2001–2003); and by Consejo Nacional de Investigaciones Cientificas y Tecnicas, Argentina, Grant PIP02521 (2004–2005).

Contributions by the authors

All the authors contributed to this paper to a similar extent.

Acknowledgement

We thank two anonymous reviewers for helpful comments on the manuscript.

Conflict of interest statement

None declared.

References

Barbosa Fevereiro VP. 1986. *Macroptilium* (Bentham) Urban do Brasil (Leguminosae – Faboideae – Phaseoleae – Phaseolinae). *Arquivos do Jardim Botanico do Rio de Janeiro* **28:** 109–180.

Bentham G. 1837. *Commentationes de leguminosarum generibus*. Vienna: Sollingeri, 72–78.

Delgado-Salinas A, **Bibler R**, **Lavin M. 2006.** Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Systematic Botany* **31:** 779–791.

Espert SM, **Drewes S**, **Burghardt A. 2007.** Phylogeny of *Macroptilium* (Leguminosae): morphological, biochemical and molecular evidence. *Cladistics* **23:** 119–129.

Gentry AH. 1982*a*. Neotropical floristic diversity: phytogeographical connections between central and south America, pleistocen climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden* **69:** 557–593.

Gentry AH. 1982*b*. Phytogeographic patterns as evidence for a Chocó refuge. In Prance G. ed. *Biological diversification in the tropics*. New York: Plenum Press, 112–136.

Iriondo M. 1999. The neogene of the Llanos-Chaco-Pampa depression. *Episodes* **22:** 226–231.

Kishino H, **Thorne JL**, **Bruno WJ. 2001.** Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* **18:** 352–361.

Lackey JA. 1983. A review of generic concepts in American Phaseolinae (Fabaceae, Faboideae). *Iselya* **2:** 21–64.

Lavin M, **Luckow M. 1993.** Origins and relationships of tropical North America in the context of the Boreotropics hypothesis. *American Journal of Botany* **80:** 1–14.

Lavin M, **Herendeen PS**, **Wojciechowski M. 2005.** Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* **54:** 575–594.

Magallon S. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. *International Journal of Plant Sciences* **165:** S7–S21.

Mayol M, **Rosello J. 2001.** Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution* **19:** 167–176.

Morgenstern B, **Frech K**, **Dress A**, **Werner T. 1998.** DIALIGN: finding local similarities by multiple sequence alignment. *Bioinformatics* **14:** 290–294.

Morrone JJ. 2001. *Biogeografía de América Latina y el Caribe*. Zaragoza: M&T-Manuales y Tesis SEA.

Nixon KC. 2002. *WinClada*. Ithaca, NY.

Posada D, **Crandall KA. 1998.** MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Ronquist F. 1996. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se).

Ronquist F. 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46:** 195–203.

Rutschmann F. 2005. *Bayesian molecular dating using PAML/multidivtime. A step-by-step manual*. Zurich, Switzerland: University of Zurich.

Rutschmann F. 2006. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. *Diversity and Distributions* **12:** 35–48.

Sanderson MJ, **Thorne JL**, **Wikstrom N**, **Bremer K. 2004.** Molecular evidence on plant divergence times. *American Journal of Botany* **91:** 1656–1665.

Swofford DL. 2002. *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Sunderland: Sinauer Associates.

Thorne JL, **Kishino H. 2002.** Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51:** 689–702.

Thorne JL, **Kishino H**, **Painter I. 1998.** Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15:** 1647–1657.

Thulin M, **Lavin M**, **Pasquet R**, **Delgado-Salinas A. 2004.** Phylogeny and biogeography of *Wajira* (Leguminosae): a monophyletic segregate of *Vigna* centered in the horn of Africa region. *Systematic Botany* **29:** 903–920.

Urban I. 1928. Plantae cubenses novae vel rariores. L. Elkman lectae. IV. *Symbolae Antillanae* **9:** 433–543.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13:** 555–556.