Acta Bot. Croat. 71 (1), 87–94, 2012

CODEN: ABCRA 25
ISSN 0365–0588
eISSN 1847-8476

The correct phylogenetic position of *Lotus conimbricensis* Brot. (Leguminosae, Loteae) based on nuclear ribosomal ITS sequences

MIGUEL A. FARIA[1], D. JAMES HARRIS[2,3,*], TATIANA VISNEVSCHI-NECRASOV[2], MANUEL TAVARES DE SOUSA[4], EUGÉNIA NUNES[2]

[1] REQUIMTE, Laboratório de Bromatologia e Hidrologia, Faculdade de Farmácia da Universidade do Porto, Rua de Aníbal Cunha, 164, 4099-030 Porto, Portugal

[2] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.

[3] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal.

[4] Estação Nacional de Melhoramento de Plantas, Estrada de Gil Vaz, 7350-951 Elvas, Portugal

Abstract – The nrDNA ITS sequence determined in *Lotus conimbricensis* in a previous phylogenetic study was unusual, in that it was almost identical to those retrieved from the morphologically distinct species *L. subbiflorus*. In the present study we sequenced new specimens of both species to reassess the phylogenetic position of *L. conimbricensis*. We conclude that the ITS sequence of *L. conimbricensis* used in the earlier analyses was most likely erroneous, and in fact *L. conimbricensis* is not closely related to *L. subbiflorus*. Critical reexamination of previously published data indicates that several other similar errors may exist for other *Lotus* species, and these should be checked before taxonomic conclusions are made.

Key words: *Lotus conimbricensis*, *Lotus subbiflorus*, ITS, phylogeny, taxonomy, nrDNA

Introduction

Lotus is the largest genus within the tribe *Loteae*, with approximately 130 species. Historically there has been little agreement in the taxonomic literature regarding the generic limits of *Lotus* and its infrageneric subdivision (DEGTJAREVA et al. 2006). However, this has changed considerably with the advent of phylogenetic studies based on nrITS sequences. These have clearly shown that the New World species of *Lotus* are not closely related to the

* Corresponding author, e-mail: james@mail.icav.up.pt

Old World species (ALLAN and PORTER 2000), and in particular DEGTJAREVA et al. (2006) revised sectional classifications proposed by KRAMINA and SOKOLOFF (2003) and SOKOLOFF (1999a, b). Some sections appeared as non-monophyletic, including the section *Lotus,* which was resolved as paraphyletic since *Lotus conimbricensis* Brot. *(Lotus* sect. *Erythrolotus* Brand) had an ITS sequence type identical to those found in *Lotus subbiflorus* Lag. (*Lotus* sect. *Lotus*).

Most *Lotus* species are distinct from each other in their ITS sequences, so it is surprising that two morphologically dissimilar species such as *L. subbiflorus* and *L. conimbricensis* should be identical in this DNA region. Errors in GenBank are well known (HARRIS 2003), and so unusual results that may have important taxonomic implications deserve careful investigation. Furthermore, the ITS region is known to display intra-individual variation in some taxonomic groups, that can confound phylogenetic studies (HARRIS and CRANDALL, 2000). DEGTJAREVA et al. (2006) combined new data with previously published sequences and noted that in the case of *Lotus creticus* they obtained a very different sequence and phylogenetic placement when compared to the sequence by ALLAN et al. (2003, 2004).

The aim of this study was to sequence several individuals of both *L. subbiflorus* and *L. conimbricensis* to resolve their phylogenetic relationship, and to look for some other possible examples of discordance in previous studies by examination of sequence data available for different *Lotus* species.

Materials and methods

Specimens were obtained from the Estação Nacional de Melhoramento de Plantas (Elvas, Portugal). DNA was isolated from leaf tissue using standard methodologies (SAMBROOK et al. 1989). The entire ITS1 and 1TS2 region was amplified using universal primers (WHITE et al. 1990). Amplifications were performed in a Biometra T3 thermalcycler (Biometra, Goettingen, Germany) in 20 μL reactions consisting of approximately 10 ng DNA template, 1 μM of each primer, 200 μM of each dNTP, 0.5 U EcoTAQ DNA polymerase (Ecogen, Barcelona, Spain), 2 μL of 10X PCR buffer and 1.5 mM $MgCl_2$, using the following amplification protocol: initial denaturation at 95 °C for 2 min followed by 30 cycles of 95 °C for 30s, 53 °C for 30s and 72 °C for 1 min. A final extension step at 72 °C for 7 minutes was performed.

PCR products were purified using the JetQuick (Genomed, Löhne, Germany) micro spin kit based on a surface modified silica membrane and sequenced using the same primers on an ABI 3730 DNA sequencer (Applied Biosystems, Foster City, USA) using the kit BigDyeTerminator v3.1 from the same supplier.

Six specimens of *L. conimbricensis*, and two *L. subbiflorus* were sequenced. 101 sequences of *Lotus* were taken from GenBank, as well as the three closest outgroups, following DEGTJAREVA et al. (2006) – *Cytisopsis pseudocytisus*, *Hammatolobium lotoides* and *Anthyllis tetraphylla*. Sequences were aligned using ClustalW with default parameters (THOMPSON et al. 1997). Phylogenetic analysis was run using MrBayes v3.1 (HUELSENBECK and RONQUIST 2001), using the GTR+I+G model, as selected by Modeltest (POSADA and CRANDALL 1998). Parameters were estimated as part of the analysis, with four Markov chains. The analysis was run for 10^6 generations, saving one tree every 100 generations.

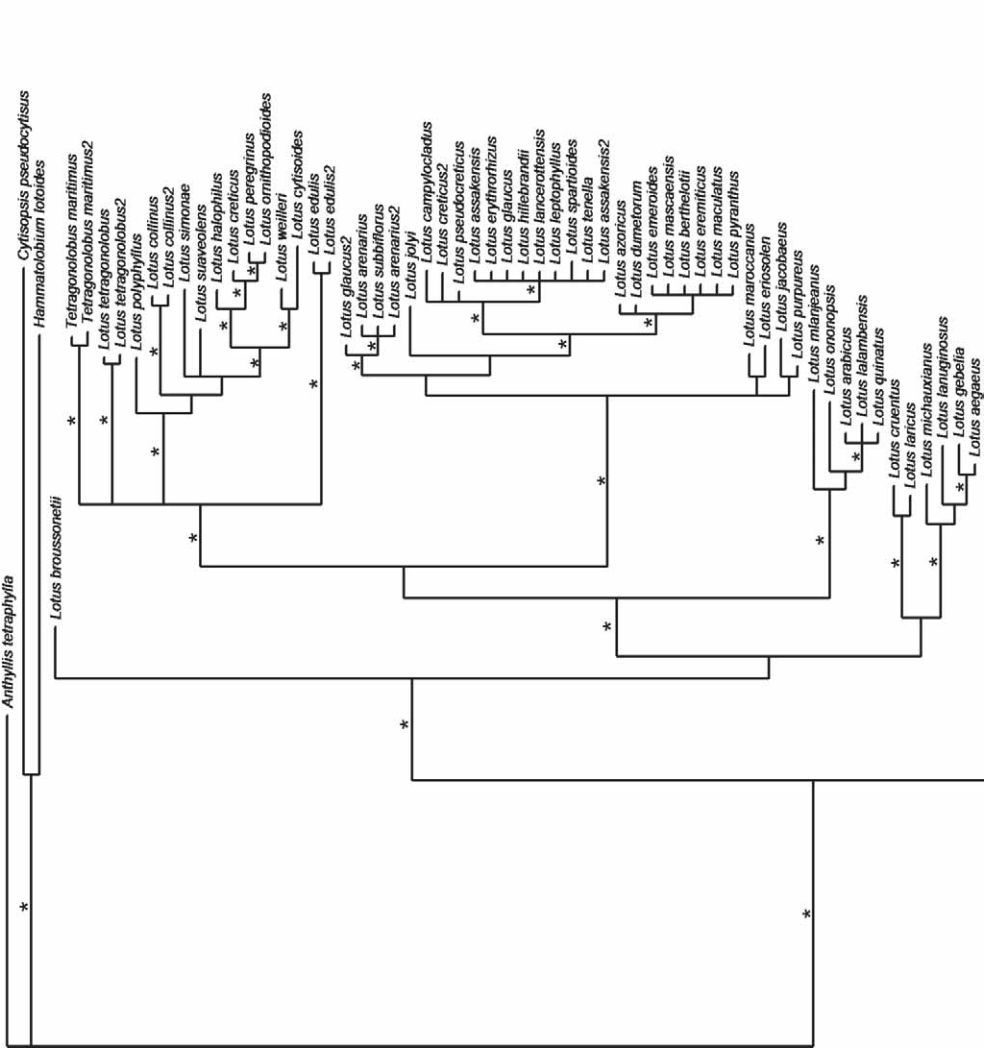
The log-likelihood values of the sample point were plotted against the generation time and trees obtained prior to reaching stationary (25%) were discarded. Remaining trees were combined in a 50% majority consensus tree . Two independent runs were made to check for convergence.

Results

In total 106 taxa were analysed, including the three outgroup taxa. Aligned sequences were 646 nucleotides long (383 constant sites, 208 informative sites). As expected the two species sequenced as part of this study showed limited intraspecific variation – both samples of *L. subbiflorus* were identical, while of the six samples of *L. conimbricensis* five were identical and one differed by a single C-G mutation. Therefore in the analysis only one sequence per species was included. All new sequence data have been deposited in GenBank (accession numbers JQ655098 to JQ655105). The phylogram estimated from the Bayesian analysis is shown in figure 1. The ITS sequences of *L. subbiflorus* generated here are identical to several other *L. subbiflorus* sequences from GenBank. One *L. subbiflorus*, however, has a very different sequence (AF450160), being identical to that from two specimens of *Lotus arenarius* (AF450193 and AF218528). The ITS sequence of *L. conimbricensis* previously published (AF450186) is almost identical to *L. subbiflorus* as sequenced in this and other studies. However, sequences from our six samples of L. *conimbricensis* are very different from the previously reported sequence, and appeared in a different part of the tree (Fig. 1). The closest relative of *L. conimbricensis* would be a sample of *Lotus halophilus* (AF450208), which differs by just two or three mutations. However, another sample of *Lotus halophilus* (DQ160283) is extremely distinct and part of another clade.

Discussion

The use of nrITS sequences has revolutionized *Lotus* systematics. Nevertheless, some conclusions based on poor data or uncritical data treatment can be premature or even incorrect. An example is the single specimen of *Lotus conimbricensis*, previously sequenced and found to be almost identical to *Lotus subbiflorus*. Later studies included the same sequence from GenBank and reached the same conclusions (e.g. DEGTJAREVA et al. 2006). However, it is now clear that several published sequences in GenBank are anomalous, with extremely divergent sequence types recovered by different authors from the same species. There may be various explanations for it. One is that in some groups intraindividual variation is very common (HARRIS and CRANDALL 2000). In these cases divergent copies obtained from the same individual can appear in different phylogenetic positions. This seems unlikely in the case of *Lotus* however, since when this happens many heterozygous positions are usually observed, unless the initial PCR products are cloned. Also, when multiple individuals are sequenced, variation would be expected. Although we sequenced six *L. conimbricensis,* only a single nucleotide difference was found. Another explanation may be that considerable variation exists within species. For instance, DEGTJAREVA et al. (2006) identified an ITS sequence in *L. creticus*, which was very different from those reported by ALLAN et al. (2003), and suggested that further studies are needed to assess intraspecific variation within this species. Again, however, we think that this explanation is unlikely in most cases


Anthyllis tetraphylla
Cytisopsis pseudocytisus
Hammatolobium lotoides
Lotus broussonetii
Tetragonolobus maritimus
Tetragonolobus maritimus2
Lotus tetragonolobus
Lotus tetragonolobus2
Lotus polyphyllus
Lotus collinus
Lotus collinus2
Lotus simonae
Lotus suaveolens
Lotus halophilus
Lotus creticus
Lotus peregrinus
Lotus ornithopodioides
Lotus weilleri
Lotus cytisoides
Lotus edulis
Lotus edulis2
Lotus glaucus2
Lotus arenarius
Lotus subbiflorus
Lotus arenarius2
Lotus jolyi
Lotus campylocladus
Lotus creticus2
Lotus pseudocreticus
Lotus assakensis
Lotus erythrorhizus
Lotus glaucus
Lotus hillebrandii
Lotus lancerottensis
Lotus leptophyllus
Lotus spartioides
Lotus tenella
Lotus assakensis2
Lotus azoricus
Lotus dumetorum
Lotus emeroides
Lotus mascaensis
Lotus berthelotii
Lotus eremiticus
Lotus maculatus
Lotus pyranthus
Lotus maroccanus
Lotus eriosolen
Lotus jacobaeus
Lotus purpureus
Lotus maritimus
Lotus onobrychis
Lotus arabicus
Lotus lalambensis
Lotus quinatus
Lotus cruentus
Lotus ianicus
Lotus michauxianus
Lotus ianuginosus
Lotus gebelia
Lotus aegaeus

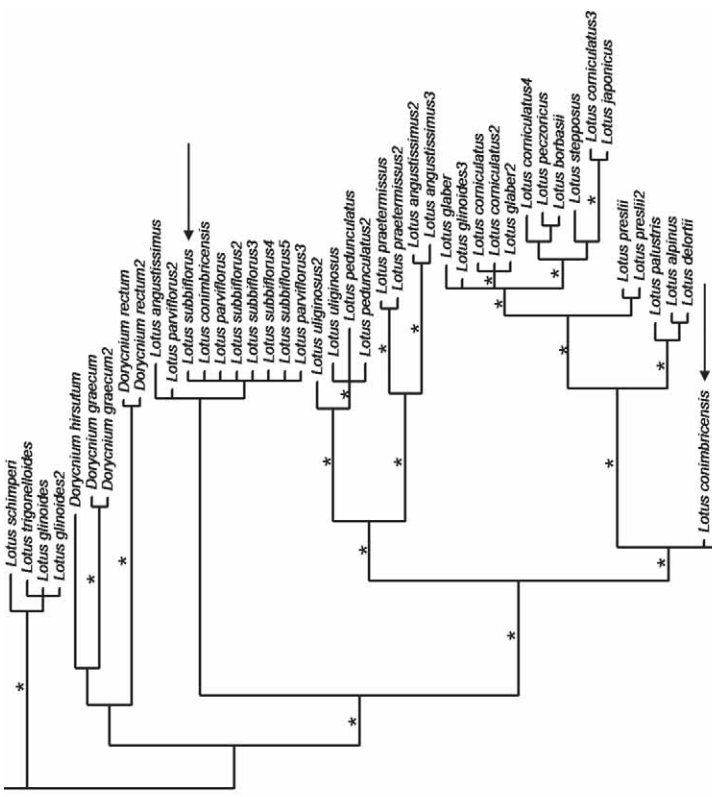


Fig 1. Phylogram of relationships of available *Lotus* species estimated using a Bayesian approach. * Indicate nodes with >95% Bayesian Posterior Probability support. Arrows indicate newly sequenced haplotypes.

of *Lotus*. As in this study, when multiple individuals are sequenced, limited variation is generally found. This can also be seen in figure 1. When haplotypes are different, intraspecific variation can be expected to be much greater than variation found between sister species. However, in the cases of *L. creticus, L. subbiflorus* and *L. conimbricensis*, where different haplotypes were found, they were placed in completely different lineages within *Lotus*. More importantly, the unexpected haplotypes are identical, or nearly so, to those of other species. Thus one *L. subbiflorus* is identical to *L. arenarius*, and one *L. conimbricensis* is almost identical to the more typical *L. subbiflorus*. In these cases it seems that errors have been made either in identifying the sample sequenced, or during the PCR and sequencing procedures. The example of *Lotus creticus* has recently been reassessed using additional samples (SANDRAL et al. 2010), confirming that the sample *L. creticus* 2 reflects its actual relationships. Unfortunately, this seems to be quite common in *Lotus*, as it is clearly illustrated here for *L. conimbricensis*. Various published sequences appear unusual in being very different from the other of the same species, but similar or identical to sequences of a different species. For example, *Lotus corniculatus* and *L. subbiflorus* show a similar pattern to this. Another example is the highly divergent sequences of *Lotus glinoides*. Two samples (DQ166220 and DQ166282, from DAGTJAREVA et al. 2006) cluster with *Lotus schimperi*, while another (AF450189, from ALLAN et al. 2003) clusters unexpectedly as part of a clade including *L. corniculatus*. This latter sequence could be another possible error.

Regarding *L. conimbricensis* our results unambiguously indicate that it is not closely related to *L. subbiflorus* as previously reported. In addition, DEGTJAREVA et al. (2008) also sequenced another individual of *L. conimbricensis*, and found it was not closely related to *L. subbiflorus*. Our newly sequenced accessions of *L. subbiflorus* are identical to some of those previously published (DEGTJAREVA et al. 2006), so there can be no doubt regarding the identification of this species. Rather our six new sequences of *L. conimbricensis* are quite distinct from other species of *Lotus*, except for one sample of *Lotus halophilus*. Unfortunately, this is another critical species, with two specimens sequenced being very different from each other. In the analysis of ALLAN et al. (2003) the relationships of these taxa are unresolved in the strict consensus trees. However, given the highly divergent and unrelated sequences, we suspect that this *L. halophilus* sample is another error.

Although many phylogenetic relationships established in earlier studies of *Lotus* are maintained in our analysis, we recommend extreme caution in making taxonomic changes based on single specimens sequenced for this marker. Not only are there general articles regarding possible errors in GenBank, but various studies have found similar problems in other plant families. For example, KRISTIANSEN et al. (2005) recently highlighted how errors in GenBank were responsible for the incorrect phylogenetic placement of the genus *Oxychloe* (Juncaceae) in an analysis based on rbcL sequences. By sequencing multiple individuals from the same species, the probabilities of errors are greatly reduced. At the same time true levels of intraspecific variation can be assessed in different species and sections. As has been stressed by others (e.g. HODKINSON et al. 2007), DNA databanks require integration with herbaria and seed banks so that cross-referencing can maximize the utilization and value of the DNA collections. Finally, users of the published sequences of *Lotus* in particular should be aware that many apparent misidentifications or errors exist, and should be especially wary of basing taxonomic decisions on single specimens.

Acknowledgements

This work was partially funded by the FCT (Fundação para a Ciência e Tecnologia) project POCTI/AGR/55696/2004. M. A. Faria gratefully acknowledges the FCT „Ciência e Inovação 2010", for the attribution of grant BPD/20725/2004. Thanks to the two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

References

ALLAN, G. J., PORTER, J. M., 2000: Tribal delimitation and phylogenetic relationships of Loteae and Coronilleae (Faboideae: Fabaceae) with special reference to *Lotus*: evidence from nuclear ribosomal ITS sequences. American Journal of Botany 87, 1871–1881.

ALLAN, G. J., FRANCISCO-ORTEGA, J., SANTOS-GUERRA, A., BOERNER, E., ZIMMERF, E. A., 2004: Molecular phylogenetic evidence for the geographic origin and classification of Canary Island *Lotus* (Fabaceae: Loteae). Molecular Phylogenetics and Evolution 32, 123–138.

ALLAN, G. J., ZIMMER, E. A., WAGNER, SOKOLOFF, W. L., 2003: Molecular Phylogenetic analysis of tribe Loteae (Leguminosae): Implications for classification and biogeography. In: Klitgaard, B. B., Bruneao, A. (eds.), Advances in legume systematics, 371–393. Royal Botanic Garden, Kew.

DEGTJAREVA, G. V., KRAMINA, T. E., SOKOLOFF, D. D., SAMIGULLIN, T. H., VALIEJO-ROMAN, C. M., ANTONOV, A. S., 2006: Phylogeny of the genus *Lotus* (Leguminosae, Loteae): evidence from nrITS sequences and morphology. Canadian Journal of Botany 84, 813–830.

DEGTJAREVA, G. V., KRAMINA, T. E., SOKOLOFF, D. D., SAMIGULLIN, T. H., SANDRAL, G., VALIEJO-ROMAN, C. M., 2008: New data on nrITS phylogeny of *Lotus* (Leguminosae, Loteae). Wulfenia 15, 35–49.

HARRIS, D. J., CRANDALL, K. A., 2000: Intragenomic variation within ITS1 and ITS2 of crayfish (Decapoda, Cambaridae): implications for phylogenetic and microsatellite studies. Molecular Biology and Evolution 17, 284–291.

HARRIS, D. J., 2003: Can you bank on GenBank? Trends in Ecology and Evolution 18, 317–319.

HODKINSON, T. R., WALDREN, S., PARNELL, J. A. N., KELLEHER, C. T., SALAMIN, K., SALAMIN, N., 2007: DNA banking for plant breeding, biotechnology and biodiversity evaluation. Journal of Plant Research 120, 17–19.

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

KRAMINA, T. E., SOKOLOFF, D. D., 2003: On *Lotus* sect. Erythrolotus and related taxa (Leguminosae). Bulletin of the Moscow Society of Naturalists Biological Series 108, 59–62.

KRISTIANSEN, K. A.; CILIEBORG, M., DRÁBKOVÁ, L., JØRGENSEN, T., PETERSEN G., SEBERG O., 2005: DNA Taxonomy – the Riddle of Oxychloë (Juncaceae). Systematic Botany 30, 284–286.

POSADA, D., CRANDALL, K. A., 1998: Modeltest: Testing the model of DNA substitution. Bioinformatics 14, 817–818.

SANDRAL, G., DEGTJAREVA, G. V., KRAMINA, T. E., SOKOLOFF, D. D., SAMIGULLIN, T. H., HUGHES, S., VALIEJO-ROMAN, C. M., 2010: Are *Lotus creticus* and *Lotus cytisoides* (Leguminosae) closely related species? Evidence from nuclear ribosomal ITS sequence data. Genetic Resources and Crop Evolution 57, 501–514.

SAMBROOK, J., FRITSCH E.F., MANIATIS,T., 1989: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.

SOKOLOFF, D. D., 1999a: *Ottleya*, a new genus of Papilionaceae – Loteae from North America. Feddes Repertorium 110, 89–97.

SOKOLOFF, D. D., 1999b: What is *Tetragonolobus wiedmannii* Boiss. (Fabaceae)? Novitales Systematicae Plantarum Vascularium 31, 139–142.

THOMPSON, J. D., HIGGINS, D. G., GIBSON, T. J., 1994: Clustal W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position--specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

WHITE, T. J., BRUNS, T. D., LEE, S. B., TAYLOR, J. W., 1990: Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics. In: Innis, N., Gelfand, D., Sninsky, J., White, T. (eds.), PCR – Protocols and applications – a laboratory manual, 315–322. Academic Press, New York.