Evolution of rbcL among *Lathyrus* and Kupicha’s classification

S. Marghali, N. Zitouna, M. Gharbi, I. Fadhlaoui and N. Trifi-Farah

Laboratory of Molecular Genetics, Immunology, and Biotechnology, Faculty of Sciences, University of Tunis El Manar, Campus Universitaire, El Manar, Tunis, Tunisia

Corresponding author: N. Trifi-Farah
E-mail: neila.trifi@fst.rnu.tn

Received September 30, 2013
Accepted July 16, 2014
Published October 27, 2014
DOI http://dx.doi.org/10.4238/2014.October.27.14

**ABSTRACT.** Phylogenetic relationships in the *Lathyrus* genus were examined using cpDNA data, particularly data attributed to the “barcode” rbcL gene to construct a possible evolutionary scenario. Plant barcoding can be used to differentiate between species within a genus and to conserve DNA within the same species. We assessed the phylogeny of 29 species of *Lathyrus* using maximum parsimony, maximum likelihood and unweighted pair-group method and arithmetic mean. The classifications did not agree with current morphological and basic *Lathyrus* classification. *Lathyrus belinensis* is a new species that was not described by Kupicha; according to rbcL analysis, the species belongs in the *Lathyrus* genus. Additionally, the genus *Lathyrus* has undergone a rapid population expansion as indicated by neutral selection indices.

**Key words:** Plastid DNA; rbcL evolution; Nucleotide variations; *Lathyrus*; Phylogeny; rbcL
INTRODUCTION

The genus *Lathyrus* (Fabaceae), a member of the tribe Vicieae, includes nearly 160 species of annual and perennial, autogamous, and allogamous herbaceous creeping plants, which are present throughout the temperate regions of the Northern hemisphere and South America. The main centers of diversity include the Mediterranean region, Asia Minor, and North America, as well as temperate South America and East Africa (Simola, 1986; Klamt and Schifino-Wittmann, 2000). Two periods have been proposed to explain the colonization of the Americas by this genus: during the Tertiary period, when Africa and South America were still united (Simola, 1986), or between the Cretaceous and Paleocene periods, through Greenland and Alaska (Kupicha, 1983). In South America, *Lathyrus* species have been found from Colombia to Tierra del Fuego, which likely migrated to the austral part of the continent through the Andean road (Burkart, 1935). All species have $2n = 2x = 14$ chromosomes, with a basic number of $n = x = 7$ (Battistin et al., 1999).

Classification of the *Lathyrus* genus has been carried out in numerous studies. Most studies separate *Lathyrus* into 12 or 13 sections worldwide (Kupicha, 1983; Asmussen and Liston, 1998; ILDIS, 2002; Leht, 2009). Subsequent studies were based on the classification of Kupicha and focused on taxonomic groupings or were regionally limited. All of these classifications were based on morphological characters such as flower color, pod appearance, and size and color of the seeds.

In addition to morphological data, other complex traits have been used to study the infra-structure and evolution of generic species of *Lathyrus* using methods such as enzyme electrophoresis (Chowdhury and Slinkard, 2000; Chtourou-Ghorbel et al., 2001; Ben Brahim et al., 2002), chloroplast DNA restriction site analysis (Asmussen and Liston, 1998), and molecular markers [random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), and inter-simple sequence repeat (ISSR)] (Croft et al., 1999; Chtourou-Ghorbel et al., 2002a,b; Belalad et al., 2006). However, uncertainty still exists regarding the intuitive species relationships among members of the *Lathyrus* genus. Therefore, additional extensive analysis of these species is needed.

Recently, macromolecular characteristics such as nucleotide sequences of DNA or amino acid sequences of proteins have been used to reconstruct phylogenetic relationships of various organisms. The plastid gene ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) has been shown to be useful for addressing phylogenetic relationships at a variety of taxonomic levels and within a number of taxonomic groups. Studies among the genera and species of Cypripedioideae by Albert (1994) and Dendrobiinae by Yukawa and Uehara (1996) indicated that the amount of sequence divergence exhibited by rbcL is sufficient and appropriate for addressing such relationships. Nucleotide sequences of the large subunit of rbcL were used for perform phylogenetic analyses, including in a wide range of angiosperms (Bremer et al., 1995; Olmstead and Reeves, 1995).

The purpose of the present study was to improve the phylogeny of the *Lathyrus* genus by including *Lathyrus* species and related known genera such as *Vicia* and *Pisum* to analyze whether the phylogeny inferred from rbcL nucleotide sequence data is correlated with that of Kupicha (1983). In this study, we examined the following points: 1) whether rbcL is an efficient tool for discriminating between close species, 2) the congruence with Kupicha classification, which was based on morphology, and 3) the evolution on the genus using neutral selection indices.
MATERIAL AND METHODS

The analysis was carried out using 6 species of the *Lathyrus* genus (*L. sativus*, *L. cicera*, *L. ochrus*, *L. sylvestris*, *L. tuberosus*, and *L. aphaca*).

DNA was isolated from young and fresh leaves using the method described by Dellaporta et al. (1983). DNA was quantified by 0.8% agarose gel electrophoresis, which included uncut lambda DNA of known concentration. The DNA was diluted for further analyses.

Laboratory procedures

Polymerase chain reaction (PCR) amplification was carried out according to the following conditions: the 25-µL reaction mixture contained 25 ng genomic DNA, 60 pg primers (1 µL), 100 µM of each dNTP, 2.5 µL 10X *Taq* DNA polymerase buffer, and 2 U *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA). The sequences of the primers used in this study were (F) 5’-ATGTCACCACAAACAGAGACTAAAGC-3’ (Levin et al., 2003) and (R) 5’-GTAAAATCAAGTCCACCAGC-3’ (Kress and Erickson, 2007). PCRs were performed using a Perkin Elmer thermocycler (Waltham, MA, USA) as follows: an initiation step of 4 min at 95°C, followed by 5 cycles for 30 s at 94°C for denaturation, 60 s at 55°C for hybridization, and 60 s at 72°C for elongation. These cycles were followed by 30 additional cycles at 30 s at 94°C for denaturation, 60 s at 54°C for hybridization, and 60 s at 72°C. A final 1-min step at 72°C was included for final elongation.

Products were separated by electrophoresis on a 1.5% agarose gel in 0.5X Tris-Borate EDTA buffer, pH 8.3, stained with ethidium bromide, and visualized under UV light. The size of amplified fragments was estimated by comparison to a 100-bp ladder that had been loaded simultaneously with the amplified DNA. The purified PCR products for the coding rbcL region were sequenced in both strands and the same primers were used for the sequencing step, carried out using the Big Dye Terminator Ready Reaction Kit on an ABI Prism™ 3130 DNA Genetic Analyzer (Applied Biosystems, Hi-Tech Detection Systems, Tunis, Tunisia).

Data analysis

Each *Lathyrus* species was sequenced at least 5 times. Obtained sequences were first evaluated using the BLAST function in the NCBI database to ensure that the sequence was correct. Picos were then verified, and *Taq* errors were corrected to enhance the credibility of the results.

All sequences were submitted and published to the GenBank database NCBI (http://www.ncbi.nlm.nih.gov/), but only 1 sequence was subsequently used to represent the corresponding species.

In the current study, other sequences were retrieved from NCBI and combined with the dataset for the evolutionary study of *Lathyrus* genus. Three *Vicia* and one *Pisum* sequences were included as out-groups. All sequences used in this study and their accession numbers are shown in Table 1.

Multiple alignment was accomplished using ClustalW, implemented using the Mega5 software (Tamura et al., 2011). The non-aligned regions (corresponding to the primer sequences) were excluded. The used primers in this study amplified only the barcoding frame of the rbcL gene. Available and used sequences corresponded, in some cases, to the entire rbcL gene.
The composition and the characteristics of the rbcL nucleotide sequences from the analyzed Lathyrus genus were determined using the Seaview4 software (Gouy et al., 2010). Maximum likelihood, maximum parsimony (MP), and unweighted pair-group method and arithmetic mean (UPGMA) were used to assess phylogeny in the genus. All analyses were performed using the Mega5 software (Tamura et al., 2011) and the bootstrap consensus trees inferred from 1000 replicates were used to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). MP analysis was accompanied by measuring the consistency index (CI) and retention index (RI). In fact, CI and RI both measure the degree of “tree-likeness” in data under a parsimony model for character changes.

The DnaSP software (Rozas, 2009) was used to analyze the sequence contents: polymorphism, conservation, segregating sites, and nucleotide and haplotype diversities. This statistical tool analyzed the evolution of Lathyrus by computing the indices of Tajima’s D, Fu and Li D, Fu, and Li F and determined the mismatch distribution of the pairwise difference.

RESULTS AND DISCUSSION

DNA sequence analysis is a fundamental method for studying the evolution of DNA coding and non-coding regions. In our study, the rbcL-coding frame was sequenced to determine the evolution and phylogeny of Lathyrus species in the Fabaceae family.

rbcL data set

Sequences in the current set were analyzed using the BLAST function on the NCBI database website. BLAST, which was used to verify the sequence identity, provides valuable insight into the efficiency of rbcL short-sequence regions and acts as a barcode (Chase et al., 2005). Complete overlapping sequences for the barcode region of the rbcL gene (between 1-599 bp) were successfully obtained for all accessions studied using the reverse primer designed by Kress and Erickson (2007). These nucleotide sequences were submitted to GenBank and published under the accession numbers listed in Table 1. Total lengths of the newly sequenced regions (from our basic dataset) of L. tuberosus, L. sativus, L. cicera, L. sylvestris, L. aphaca, and L. ochrus were between 551-563 bp (L. ochrus and L. cicera, respectively). This length has been obtained for many plants using modified primers for barcoding analysis. In fact, Kress and Erickson (2007) found a mean size of 530 bp for the rbcL frame.

The nucleotide composition was as follows: 27.6% A, 20.2% C, 22.2% G, and 29.9% T. This composition agrees with results of Sulaiman et al. (2003), who examined 23 species from 16 genera of Mimoseae. A notable feature of the plastid genome is the high AT-content, both in coding regions and in non-coding regions (Howe et al., 2003).

Including the entire dataset (sequenced and retrieved from NCBI), 561 aligned characters, including gaps (488 sites without gaps), were used for phylogenetic analyses. Four hundred and forty-two constant characters (90.6%) and 46 variable characters (9.4%) were identified.

Among the constant characters, 3 long conserved regions were observed. The first was between 47 and 90 bp (5'-AAAACAAAGATACATATCTTGGCAGCATTCCGAGTAA CTCC-3'), the second was located on the 276-316-bp frame (5'-ACCTTTTTGAGAAGGTT TCTGTACTAATACGAGCCACCTCC-3'), and the third frame (5'-AGAGATAAAATTGAAC AGTATGGACGTCC-3') was 434-463 bp.
Of the 46 segregating sites, 29 (63.04%) were parsimony non-informative, while 17 (36.95%) were parsimony informative. A sliding window represents the evolution of the segregating sites among the 561 characters analyzed (Figure 1A). The maximum number of segregating sites was 12. The frame behind the 200 bp was the maximum segregating region, which corresponded to the 379 nucleotide. However, the region upstream and downstream of 300 bp was less parsimonious. These results were consistent with the conserved character of the \textit{rbcL} gene. Among plastid regions, this gene represented the best-characterized frame in species discrimination (CBOL Plant Working Group, 2009).

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Nos.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lathyrus sativus</td>
<td>Kc526196.1</td>
<td>NCBI GenBank</td>
</tr>
<tr>
<td>Lathyrus aphaca</td>
<td>Kc526159.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus cicera</td>
<td>Kc526170.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus ochrus</td>
<td>Kc526177.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus sylvestris</td>
<td>Kc526183.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus tuberosus</td>
<td>Kc526191.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus angulatus</td>
<td>Jn61180.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus davidii</td>
<td>Gq436352.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus linifolius</td>
<td>Jn892522.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus palustris</td>
<td>Jn891868.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus nissolia</td>
<td>Jn892530.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus bauhinii</td>
<td>Jx505475.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus hierosolymitanus</td>
<td>Jx505477.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus inconspicuus</td>
<td>Jx505479.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus pubescens</td>
<td>Jx505483.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus belizensis</td>
<td>Jx505476.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus venosus</td>
<td>Jx505484.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus tingitanus</td>
<td>Hm850109.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus niger</td>
<td>He963532.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus sphaericus</td>
<td>Jn661186.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus pratensis</td>
<td>Jn661186.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus japonicus subsp. maritimus</td>
<td>Hm850107.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus clymenum</td>
<td>Hm850105.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus hirtatus</td>
<td>Hm850106.1</td>
<td></td>
</tr>
<tr>
<td>Lathyrus nervosus</td>
<td>Jx505482.1</td>
<td></td>
</tr>
<tr>
<td>Vicia faba</td>
<td>Jn661200.1</td>
<td></td>
</tr>
<tr>
<td>Vicia sativa</td>
<td>Hm850461.1</td>
<td></td>
</tr>
<tr>
<td>Vicia villosa</td>
<td>Hm850464.1</td>
<td></td>
</tr>
<tr>
<td>Pisum sativum subsp. sativum</td>
<td>Jn661190.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Summary of \textit{Lathyrus} species included in this study and GenBank accession numbers of \textit{rbcL} coding sequence.

Figure 1. Base-pair sliding window of chloroplast DNA coding region for \textit{Lathyrus} genus. Variability of the segregating sites (S) (A) and the nucleotide diversity (Pi) (B) for the \textit{rbcL} gene.
The 29 analyzed sequences were regrouped into 25 haplotypes (Hd = 0.983). This indicates that the *Lathyrus* analyzed species are evolving independently.

Pi values were represented on a sliding window (Figure 1B). Frequencies were balanced between 0.005-0.02. This indicates that the level of polymorphism in this gene is low. The mean nucleotide diversity value was equal to Pi = 0.01449.

The *rbcL* gene is known to be extremely conserved among angiosperms, which is useful for phylogenetic studies of green plants (Zhang et al., 2011; Bafeel et al., 2012). However, lower mutation rates are typically observed among angiosperms in which homologies and similarities between rbcL sequences are greater than 80% (Sulaiman et al., 2003; Zhang et al., 2011).

In the *Lathyrus* genus, some insertions and deletions were observed that have not been reported in other species. This observation confirms the results of Asmussen and Liston (1998) who suggested that the *Lathyrus* genus has an elevated rate of cpDNA evolution compared to other examined genera.

**Phylogeny and evolution**

The sequence data were subjected to three different methods of phylogenetic reconstruction: i) UPGMA, ii) MP, and iii) maximum likelihood (ML).

Using the UPGMA method (Figure 2), the phylogenetic analyses were performed based on the evolutionary distances computed by Tajima and Nei (1984). The phylogenetic relationships established using the UPGMA dendrogram appeared to be independent from Kupicha’s classification. In fact, species belonging to *Lathyrus* and *Orobus* were observed in the UPGMA dendrogram. *Lathyrus belinensis*, a new species, discovered by Maxted and Goyder in 1988 (http://www.lathyrus.info/species/belinensis.html), was not included in Kupicha’s classification. However, this species was morphologically similar to *L. odoratus* (*Lathyrus* section). Moreover, *L. tingitanus* is independent from the other *Lathyrus* species and appears to be similar to *L. clymenum*. This observation is in agreement with the findings of Asmussen and Liston (1998) who examined the ropC and IR regions. The chosen out-groups (*Vicia villosa, V. faba, V. sativa*, and *Pisum sativum*) were not considered to be ancestors of the *Lathyrus* genus.

Using the MP method, the maximum parsimonious tree (length = 73 steps) was obtained (Figure 3). The MP tree was obtained using the close-neighbor-interchange algorithm (Nei and Kumar, 2000) and showed a length of 0.14. MP analyses showed high consistency and retention indices (CI = 0.524 and RI = 0.710). Moreover, composite indices were 0.516 and 0.372 for all sites and parsimony-informative sites, respectively. These values indicate a low rate of homoplasy within the sequence data and demonstrate the utility of the rbcL barcode region for identifying genetic relationships among *Lathyrus* species. According to the generated tree, *Orobus* species (except *L. linifolius*) were clustered together, while the *Lathyrus* species were opposite in the tree. The MP method confirmed the closeness of the 2 sister species *L. sativus* and *L. cicera*. Furthermore, *Vicia* species were clustered with *Clymenum* species, while *Pisum* were similar to the ancestor. The MP tree agreed with the reports of Asmussen and Liston (1998) regarding the closeness of *Orobus, Notolathyrus*, and *Lathyrrostylis* sections.

The ML tree was constructed based on the Hasegawa-Kishino-Yano algorithm (Hasegawa et al., 1985). Otherwise the BIONJ method with an ML distance matrix draws the cladogram. The tree with the highest log likelihood score (lnL = -1190.507) is shown
Figure 2. Unweighted pair-group method and arithmetic mean (UPGMA) dendrogram illustrating the genetic relationships between *Lathyrus* species based on the rbcL barcode region of the chloroplast DNA. Numbers below the nodes indicate bootstrap values.

Figure 3. Strict consensus of the most highly parsimonious trees from the rbcL barcode region data set (length = 73 steps; CI = 0.524; RI = 0.710). Numbers above the nodes indicate bootstrap values.
in Figure 4. This tree was clustered into 2 main groups. The first cluster included the out-groups and *Clymenum* species (*L. ochrus* and *L. clymenum*) with the 3 species of *Lathyrus* section *L. sativus*, *L. cicera*, and *L. tingitanus*. On the other cluster, *L. belinensis* is always close to *Lathyrus* section species; despite its recent discovery, it appears to belong to the *Lathyrus* section if the Kupicha classification is considered.

**Neutral selection**

The null model in population genetics describes a scenario in which natural selection plays no role in species evolution. The test developed by Tajima (1989) was the first aimed specifically at evaluating neutrality in the context of infinite-site models of sequence evolution. Neutral selection indices of Tajima (D), Fu and Li (D and F), and Fu’s (Fs) were also used to characterize the evolution of the *Lathyrus* genus.

As the Tajima’s D index was not 0, the neutral evolution hypothesis under no selective pressure was rejected. In the present study, the value of Tajima’s D, calculated based on the 29 sequences analyzed, deviated from 0 and was negative (-1.789). This value suggests that demographic expansion or selective sweeps occurred.
The results of Tajima’s and Fu and Li’s tests suggested, therefore, an excess of singletons against a low rate of intermediate variants in the entire sample. The mismatch in pairwise nucleotide differences (Figure 5) revealed a unimodal pattern that is characteristic of an expansion of the *Lathyrus* genus as described by Slatkin and Hudson (1991) and Rogers and Harpending (1992). These observations suggest that a demographic scenario is unlikely, as demographic processes are thought to affect all genes across all genomes (Ben Mustapha et al., 2013).

**Figure 5.** Mismatch distribution of the *rbcL* region of *Lathyrus* species based on pairwise nucleotide differences in the *rbcL* gene. The solid and dashed curves indicate the expected and the observed, respectively, distributions under population expansion.

The Fu and Li indices (D and F) were negative but not significant (Fu and Li’s $D^* = -2.35483$; Fu and Li’s $F^* = -2.56018$). A high negative value for Fu’s was found (-16.141), indicating an excess of alleles, which is expected for recent population expansion.

Our results were consistent with balancing selective pressure in the *rbcL* gene and demonstrate that recent population expansion occurred in the legume *Lathyrus* genus.

In summary, we present a robust phylogenetic hypothesis for species in the *Lathyrus* genus determined by analyzing the *rbcL* gene and particularly the barcode frame of this chloroplast gene. Chloroplast DNA loci, which are thought to be uniparentally inherited and non-recombinant, have been used for systematic and phylogenetic analysis of the *Lathyrus* genus. The systematics of this genus remains ambiguous and is often compared with the morphological classification described by Kupicha. The *rbcL* gene, which acts as a barcode region, was efficient for examining the phylogeny of this genus, but our results were not in 100% concordance with those of Kupicha.

**ACKNOWLEDGMENTS**

Research supported by grants from the Ministry of Higher Education and Scientific Research, Tunisia (Project Lab #B02).
REFERENCES


Evolution and phylogeny of *Lathyrus* genus via rbcL frame


