MOLECULAR PHYLOGENY OF Lallemantia L. (LAMIAEAE): INCONGRUENCE BETWEEN PHYLOGENETIC TREES AND THE OCCURRENCE OF HGT

Masoud SHEIDAI*, Fahimeh KOOHDAR, Zeinab Moradiyan POODE

Faculty of Life Sciences & Biotechnology, Shahid Beheshti University, Tehran, Iran.


The genus Lallemantia Fisch. & C. A. Mey. (Family Lamiaceae) is a small genus with only five species that are herbaceous annual or biennial plants with food and medicinal value. This genus is of Caucasian origin and contains 5 species in Iran. The aims of the present study were: 1- to examine the occurrence of phylogenetic conflict between nuclear (ribosome ITS), nuclear repetitive sequences (ISSRs) and plastid (rps16 intron, cp) sequences in the genus Lallemantia (Lamiaceae), 2- to investigate the occurrence of inter-specific hybridization in this genus, and 3- to compare the time of divergence of the species from the basal line by ITS and cp-DNA molecular data. This is the first analysis on these evolutionary aspects of the genus Lallemantia. We provided the first molecular evidence for the occurrence of inter-specific hybridization in the genus Lallemantia and illustrated that phylogenetic signals in cp-DNA and ITS sequences differ significantly.

Keyword: Lallemanti, hybridization, cpDNA, ITS

INTRODUCTION

Lamiaceae (Mint family), is the sixth largest angiosperm family, and contains more than 7000 species that are distributed all over the world (Li et al., 2016).

Despite extensive progress that has been made in molecular phylogeny studies in this family (Nakano and Sasaki, 2011), its phylogenetic backbone has never been well resolved. Its molecular phylogeny is complicated due to conflict between molecular data and potential inter-specific hybridization as well as horizontal gene transfer (Oliveira et al., 2007; Salma et al.,

Corresponding author: Masoud Sheidai, Faculty of Life Sciences & Biotechnology, Shahid Beheshti University, Tehran, Iran. e-mail: msheidai@yahoo.com
Phylogenetic incongruence between “gene trees” and “species trees” has been widely acknowledged. Conflicts may emerge from several processes including paralogy, hybridization, and incomplete lineage sorting (ZHANG et al., 2015).

The phenomenon of polyploidy and hybridization usually results in novel genetic combinations, leading to complex, reticulate evolution and incongruence among gene trees, which in turn may show different phylogenetic histories than the inherent species tree (ROY et al., 2015).

Congruence of gene trees (or sub trees) is often considered the most desirable outcome of phylogenetic analysis, because such a result indicates that all sequences in the clade are orthologs (homologs derived from the same ancestral sequence without a history of gene duplication or lateral transfer), and that discrete monophyletic clades can be unambiguously identified, perhaps supporting novel or previously described taxa. In contrast, gene trees that are incongruent are often considered problematic because the precise resolution of speciation events seems to be obscured. Thus, it would also be very useful to identify significant incongruencies in gene trees because these represent non-canonical evolutionary processes (e.g., MADDISON and KNOWLES, 2006; LIU et al., 2008).

Lamiaceae family is no exception to this phenomenon. For example, BOGLER et al. (2004) carried out phylogenetic analyses of nucleotide sequences of the internal transcribed spacers and 5.8S subunit of nuclear ribosomal DNA and the trnL gene and trnL-trnF spacer of the chloroplast genome for 33 of the 72 genera in the Mentheae tribe and reported conflict between the results of different molecular data. Maximum parsimony analysis of the combined data set showed that Bystropogon is the sister genus to the Old World taxa Acinos, Ziziphora, and Clinopodium. However, separate analysis of the ITS and trnL/F data sets did not agree as to the sister group to Bystropogon, and the cpDNA phylogeny strongly supported a relationship of Bystropogon with a clade of New World mint taxa. Due to the apparent conflict between the chloroplast and nuclear characters observed in the phylogenies, they could not draw a certain conclusion on the true biogeographic relationship of Bystropogon. There are also many methods of constructing phylogenetic trees (e.g. Distance, Parsimony or Likelihood), which can produce different trees. Given this situation, it is desirable to compare phylogenetic trees from a set of sequences constructed by different methods and/or to compare phylogenetic trees from different sets of homologs (PUIGBO et al., 2007).

Once the final tree is constructed, there are several metrics for calculating the correctness of this tree, Ti, when compared to the actual tree, T0.

One distance metrics which is widely used to determine phylogenetic trees difference is the Robinson-Foulds distance, (ROBINSON and FOULDS, 1981). The Robinson-Foulds distance between tree T1 and T2 can be calculated as:

\[
RF(T_1, T_2) = |C(T_1) - C(T_2)| + |C(T_2) - C(T_1)|
\]

Essentially, this could be interpreted as the number of false positive and false negative bipartitions in T2 when compared to T1. For n taxa, the maximum value of the RF distance is 2n-6.

One advantage of applying this method is that it is very fast. Pair wise RF distances can be computed for t trees on n taxa in O(n log2 n) time. Another advantage to this metric is that, as
the number of taxa grows large, the average normalized mean distance (the average distance from one tree to another random tree, divided by the total possible score) approaches 1, and the variance approaches 0 (CLEMENT, 2011).

The genus Lallemantia Fisch. & C. A. Mey. (Family Lamiaceae) is a small genus with five species. They are herbaceous annual or biennial plants that have served as food and medicine. For example, Lallemantia iberica Fisch. & C. A. Mey. is cultivated in Iran and southern parts of the former USSR as an oil-seed plant (RIVERA-NUNEZ and OBRONE-GASTRO, 1992) and L. royleana seeds displayed significant anti-bacterial activity and can be a good remedy for skin disease and gastro-intestinal problems caused by human pathogenic bacterial strains (MAHMOOD et al., 2013).

Lallemantia species are distributed in Afghanistan, China, India, Kazakhstan, Kyrgyzstan, Pakistan, Iran, Russia, Tajikistan, Turkmenistan, Uzbekistan, SW Asia and Europe. This genus is a Caucasian originated genus and contains 5 species in Iran (RECHINGER, 1982).

The aims of the present study were: 1- to examine the occurrence of phylogenetic conflict between nuclear (ribosome ITS), nuclear repetitive sequences (ISSRs) and plastid (rps16 intron) sequences in the genus Lallemantia (Lamiaceae), 2- to investigate the occurrence of inter-specific hybridization in this the genus, and 3- to compare time of divergence of the species by ITS and cp-DNA molecular data. This is the first report on these evolutionary aspects of the genus Lallemantia.

MATERIALS AND METHODS

Plant materials

Extensive field work and collections were undertaken during 2013–2015, and plant specimens were randomly collected from 5 geographical populations. We identified five species, L. allemantia royleana, L. canescens, L. baldschuanica, L. iberica and L. peltata. Five leaves per species were randomly selected and used for molecular phylogeny and ISSR study. The voucher specimens are deposited in herbarium of Shahid Beheshti University (HSBU).

Table1. The list of Lallemantia species studied, their localities and voucher numbers

<table>
<thead>
<tr>
<th>Province</th>
<th>Voucher number</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Lallemantia peltata</td>
<td>HSBU 201209</td>
<td>Alborz</td>
</tr>
<tr>
<td>2 Lallemantia conecens</td>
<td>HSBU 201200</td>
<td>Qazvin</td>
</tr>
<tr>
<td>3 Lallemantia baldschuanica</td>
<td>HSBU 201208</td>
<td>Khorasan</td>
</tr>
<tr>
<td>4 Lallemantia iberica</td>
<td>HSBU 201210</td>
<td>Zanjan</td>
</tr>
<tr>
<td>5 Lallemantia royleana</td>
<td>HSBU 201212</td>
<td>Khorasan</td>
</tr>
</tbody>
</table>

CP-DNA

The intron in the gene for ribosomal protein L16 (rpl16) was amplified and sequenced with universal primers following the methodology of SHAW et al. (2005) and TIMMER et al. (2007). The rpl16 forward primer was 5’- GATAAGGTACTTATTAGTGTCGTTT -3’ and, the reverse primer was 5’- TCCTTACCATTTAGTGTCGTTT -3’. Each 20 µl of PCR tube contained 10 µl of 2x PCR buffer, 0.5 mM of each primer, 200 mM of each dNTP, 1 Unit of Taq DNA polymerase (Bioron, Germany), and 1 µl of template genomic DNA at 20 ng µl⁻¹. The
amplification reaction was performed in Techne thermocycler (Germany) with the following program: 2 min initial denaturation step 94°C, 1 min at 94°C; 1 min at 54°C and 1 min at 72°C. The reaction was completed by final extension step of 6 min at 72°C. PCR products were visualized on 2.5% agarose gel through GelRed™ Nucleic Acid Gel Staining. Fragment sizes were estimated using a 100 bp size ladder (Thermo- Fisher Scientific, Waltham, MA USA).

**ITS nuclear gene**

The primers used for ITS region are provided in Table 2.

<table>
<thead>
<tr>
<th>Number</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ITS5 5′- GGA AGT AAA AGTCGT AAC AAG G- 3′</td>
</tr>
<tr>
<td>2</td>
<td>ITS4 5′- TCC GCT TATTGA TAT GC- 3’</td>
</tr>
</tbody>
</table>

Each 20 µl of PCR tube contained 10 µl of 2x PCR buffer, 0.5 mM of each primer, 200 mM of each dNTP, 1 Unit of Taq DNA polymerase (Bioron, Germany), and 1 µl of template genomic DNA at 20 ng µl⁻¹. The amplification reaction was performed in Techne thermocycler (Germany) with the following program: 2 min initial denaturation step 94°C, 1 min at 94°C; 1 min at 53°C and 1min at 72°C. The reaction was completed by final extension step of 6 min at 72°C. PCR products were visualized on 2.5% agarose gel through GelRed™ Nucleic Acid Gel Staining. Fragment sizes were estimated using a 100 bp size ladder (Thermo- Fisher Scientific, Waltham, MA USA).

**Phylogenetic analyses**

The intron in the gene for ribosomal protein L16 (rpl16) and ITS- sequences were aligned and used to study the species relationship by performing using different methods such as Neighbor Joining (NJ), UPGMA, and maximum parsimony (ML) as performed in MEGA 5 software (TAMURA et al., 2011).

The molecular clock test was performed as implemented in MEGA 5. The test was done by comparing the ML value for the given topology with and without the molecular clock. Constraints under the TAMURA and NEI (1993). Before estimating time of divergence, we used MEGA 5 to test the molecular clock and to find the best substitution model for the given sequences. The equal evolutionary rate of the studied sequences was rejected at a 5% significance level and therefore we used the relaxed molecular clock model in further analyses (MINAEIFAR et al., 2016). Moreover, HKY was the best substitution model identified by model test as implemented in MEGA 5 (TAMURA et al., 2011).

**Test for incongruence between phylogenetic trees**

We used 3 different methods for testing difference between phylogenetic trees: 1- The Robinson-Foulds distance, (ROBINSON and FOULDS, 1981), as implemented in PHYLIP ver. 3.1 (1998).
2- The statistical significance of a match between two incongruent phylogenetic trees, reported as P-values (THEOBALD, 2004). This method provides a table and the java script calculator that provide values for the statistical significance of a match between two incongruent phylogenetic trees, reported as P-values. These P-values give the probability that two bifurcating rooted trees, with a given number (or less) of mismatching branches, would match by chance.

The number of incongruent branches is determined relative to the maximum agreement sub tree (MAST) between two trees. A MAST is the "core" sub-tree that is common between two trees. The number of incongruent branches is equal to the minimum number of branches that must be pruned from one of the real trees to get the MAST. P is the ratio of the maximum number of possible incongruent trees over the total number of possible trees.

3- The ILD test. (= Partition Homogeneity Test (PHT) as implemented in PAUP. For this, first we combined the data sets by using Sequence Matrix 1.8 ver. 100., and then the combined data were exported to PAUP ver. 4 (Swofford, 2002) and tested by PHT.

**Time of species divergence**

BEAST v1.6.1 (DRUMMOND et al., 2010a, b) was used for the Bayesian MCMC inferred analyses of the nucleotide sequence data (DRUMMOND and Rambaut, 2007). BEAUti (Bayesian Evolutionary Analysis Utility version) v1.6.1 (DRUMMOND et al., 2010a, b) was utilized to generate initial xml files for BEAST. A Yule process of speciation (‘a pure birth’ process) was used as a tree prior for all the tree model analyses. The Yule tree prior is widely recognized as giving the best fit model for trees describing the relationships between different species. The parameter can be regarded as explaining the net speciation rate (nee, 2006). For the MCMC posterior analyses, the length of chain was 1000000. After 100 tree burn-in processing, 10000 trees were used for the analyses. The BEAUti xml file (DRUMMOND et al., 2010a, 2010b) was run in the BEAST v1.6.1 program (DRUMMOND et al., 2010a, 2010b) and the maximum clade credibility (MCC) chain generations were repeated five times for each molecular clock model with independent runs to ensure suitable convergence and adequate mixing. The MCC tree was generated under the relaxed clock model (HKY substitution). Because no fossils are available for the studied species, we used a rate of evolution of the plastid sequence (u = 1.0 X 10^{-9} s^{-1} year^{-1}) (Zurawska et al., 1984).

Tracer v1.5 software (Rambaut and Drummond, 2007) was used for the output of the model parameters to examine the sampling and convergence results obtained from BEAST. Tree Annotator v1.6.1 software (Rambaut and Drummond, 2007) was used to annotate the phylogenetic results generated by BEAST as a form of single ‘target’ tree. Summary statistics of posterior probabilities of the nodes are shown on the target trees: the 95% highest posterior density (HPD) limits of the node heights, rates and the posterior estimates. For the annotated BEAST MCC tree output analyses, the FigTree v1.3.1 (Rambaut, 2009) program was also used. The posterior probability was set to 0.5. This is equivalent to the bootstrapping value in PAUP (Phylogenetic Analysis Using Parsimony analysis) analyses (Wan-Pyo Hong and Jury, 2011).

**Biogeography**

The distribution range of Lallemantia species studied was divided into 5 areas (provinces). These areas are: A (Alborz), B (Qazvin), C (Razavi Khorasan), and D (Zanjan). We used S-DIVA (Statistical Dispersal-Vicariance Analysis) and BBM (Bayesian Binary Method) analyses implemented in RASP to reconstruct the possible ancestral ranges on the phylogenetic
trees (YU et al., 2010). In these methods, the frequencies of an ancestral range at a node in ancestral reconstructions are averaged over all trees (YU et al., 2010). To account for uncertainties in phylogeny, we used the tree obtained from BEAST analysis. The possible ancestral ranges at each node on a selected tree were obtained.

RESULTS

Phylogenetic trees and species relationship

The ISSR tree (Figure 1) of the studied species revealed close affinity between L. iberica and L. canescens; L. peltata and L. royleana are also close to each other. Similarly, L. baldchuanica is close to L. canescens and L. iberica.

![Figure 1. ISSR NJ tree of the Lallemantia species.](image)

Cp-DNA phylogenetic trees

After cp-DNA multiple sequence alignment by MUSCLE and curing the sequences, 385 sequences remained for phylogenetic tree construction. Out of these 91 sequences was parsimony informative.

Different methods of ML and Bayesian phylogenetic trees as well as networking of Lallemantia species based on C-DNA sequences (Figure 2) produced similar results. In all of them, close affinity was observed between L. iberica and L. canescens; followed by L. baldchuanica. Similarly, L. peltata and L. royleana are very close to each other but are far from the other three species.
**ITS phylogenetic trees**

After ITS multiple sequence alignment by MUSCLE and curing the sequences, 226 sequences remained for phylogenetic tree construction. Out of these 124 sequences were parsimony informative.

Phylogenetic trees of ML and Bayesian methods and networking (Figure 3) of *Lallemantia* species based on ITS sequences produced similar results. In all of them, *L. iberica* was placed far from the other species, followed by *L. royleana*. Three species of *L. peltata*, *L. baldschuanica* and *L. canescens* were also grouped together. However, in ML and Bayesian trees, the first two species were closer to each other, while in network, *L. peltata* and *L. canescens* were placed close to each other.

Therefore, the phylogenetic relationship of analyzed species differed to a great degree in cp-DNA based phylogenetic trees versus ITS-based phylogenetic trees.
Incongruence test between cp-DNA and ITS gene

We found evidence for significant phylogenetic incongruence between plastid and nuclear data in the genus. The Robinson-Foulds distance was 4, while the statistical significance test of Theobald (2004) as well as the partition-homogeneity test (PHT) of PAUP produced significant difference between the two data sets ($P = 0.02$). Therefore, ITS and Cp-DNA sequences produced different evolutionary signals in *Lallemantia*.

Horizontal Gene Transfer (HGT) test

HGT test performed by T-REX between ISSR and ITS trees, as well as ITS and Cp-DNA trees, revealed two events of horizontal gene transfer, first between *L. baldschuanaica* and *L. peltata*, and second one between *L. canescens* and *L. peltata*.

Time of divergence in Lallemantia

BEAST chronogram of Cp-DNA suggests the probable time of divergence of *Lallemantia* is 2-27 mya, while according to ITS sequences it is 7-14 mya. This date incongruence between these two genes and sequences makes uncertain any suggestive date of divergence in the genus.
DISCUSSION

Lamiaceae is the sixth largest angiosperm family, and contains more than 7000 species distributed all over the world. This family has many phylogenetically unresolved genera and therefore many species are of not determined relationship LI et al. (2016). Its molecular phylogeny is complicated due to the conflict between molecular data and potential inter-specific hybridization as well as horizontal gene transfer (OLIVEIRA et al., 2007; SALMAKI et al., 2013; LI et al., 2016).

The present study revealed that Cp-DNA phylogenetic tree is in accordance with ISSR tree of the species relationship, but ITS tree differed from them. Hence, it is the Cp-DNA that is of phylogenetic application to resolve the species relationship in Lallemantia, and that combined Cp-DNA and ITS nuclear sequences should not be used for molecular phylogenetic studies in this genus. This may suggest cp-DNA as a suitable molecular marker for studying Lallemantia evolutionary history and phylogeny.

HGT analysis suggested that the occurrence of inter-specific hybridization is one of the potential causes of incongruence observed. Therefore, it seems that inter-specific hybridization is occurring in the genus Lallemantia and this phenomenon may be one of the possible causes of incongruence between ITS and Cp-DNA phylogenetic trees observed.

Polyplody and hybridization can result in novel genetic combinations and the occurrence of complex, reticulate evolution and therefore, could lead to incongruence among gene trees, as well as to picturing different phylogenetic histories from the inherent species tree (ROY et al., 2015).

Other reasons for Phylogenetic incongruence between “gene trees” and “species trees” include paralogy, hybridization, and incomplete lineage sorting (ZHANG et al., 2015).

In several molecular studies carried out within Lamiaceae, we encountered reports on incongruence between phylogenetic relationships produced by nuclear and plastid genome. However, the results of these investigations differ in showing which genus is concordant with taxonomy of the studied plant group. For example, OLIVEIRA et al. (2007) carried out molecular phylogenetic study of the genus Dicerandra, an endemic mint of the southeastern United States that comprises nine species. They used data from both nuclear and plastid genomes and found incongruence in the nuclear and plastid trees in their placement of two perennial taxa, D. cornulissima and D. immaculata var. savannarum, perhaps due to ancient hybridization or to lineage sorting. Similarly, SALMAKI et al. (2013) investigated molecular phylogeny of tribe Stachydeae (Lamiaceae) and reported that within Eurystachys, monophyly is supported by both nuclear and plastid data but majority of recognized taxa appeared to be para- or polyphyletic. The taxon compositions of most sub-clades were congruent between the plastid and nuclear tree topologies, whereas their relative phylogenetic placements are often not. They concluded that plastid–nuclear incongruence was due to the occurrence of hybridization within Stachydeae.

ROY et al. (2015) also studied molecular phylogeny the Hawaiian endemic mints of Stachydeae and considered allopolyploidy, hybridization and incomplete lineage sorting as the potential causes of speciation in this group. They utilized five independently inherited, low-copy nuclear loci, and performed both individual gene tree analysis as well as multi-locus coalescence-based tree reconstructions. The results showed incongruence between individual gene trees, grouping the Hawaiian mints with both temperate North American and Meso-South American Stachys clades, while multi-locus coalescence tree was concurrent with previous nrDNA results placing them within the temperate North American Stachys clade.
Molecular investigation performed by Gobert et al. (2006) in a hybridization planed experiment in *M. × piperita* by utilizing three data partitions of nuclear DNA (ITS), chloroplast DNA (non-coding regions *trnL* intron, intergenic spacers *trnL-trnF*, and *psbA-trnH*), and AFLP and ISSR, markers, produced incongruities between ITS, chloroplast DNA, and AFLP-ISSR phylogenetic trees. Only DNA fingerprinting data (AFLP-ISSR) were congruent with morphological classification.

They concluded that the event of chloroplast capture occurred in *M. × piperita*. Moreover, direct sequencing of ITS failed to provide evidence of the existence of the two parental copy types for *M. × piperita*, a sterile hybrid that has had no opportunity for concerted evolution of ITS copies.

Since, AFLP-ISSR data clustered *M. × piperita* with the parent that had the largest genome; they concluded that differential introgression of different genome regions occurred in mint hybrids.

In constructing molecular phylogenetic trees different sets of putatively orthologous genes often yield strongly supported but incompatible tree topologies (Beiko and Hamilton, 2006). Incongruence in tree topologies can be explained by such processes as horizontal gene transfer events, hidden paralogy (Crevey et al., 2004) and model misspecification.

There are also many methods of constructing phylogenetic trees (e.g. Distance, Parsimony or Likelihood), which can produce different trees. Given this situation, it is desirable to compare phylogenetic trees from a set of sequences constructed by different methods and/or to compare phylogenetic trees from different sets of homologs (Puigbó et al., 2007).

In the field of phylogenetics, much research has been focused on the amount of data that accurately create a species tree. Too little data produce unreliable tree, while too much data may be problematic due to incongruence between genes. By using individual genes to create an individual gene trees, and then combining them intelligently to form species tree, the result will not only be a stable species tree, but will also produce individual, accurate gene trees as a byproduct. These individual gene trees can be analyzed and compared to the species tree, looking for statistically significant differences that could allow the researcher to infer different evolutionary pressures (Clement, 2011).

Significant phylogenetic incongruence between plastid and nuclear data may affect taxonomy and other phylogeny related aspects like time of divergence and reconstruction of ancestral areas (Zhang et al., 2015).

In conclusion, we provided the first molecular evidence for the occurrence of inter-specific hybridization in the genus *Lallemantia* and illustrated that phylogenetic signals in cp-DNA and ITS sequences differ significantly.

Received, December 31st, 2018
Accepted October 18th, 2018

REFERENCES


MOLEKULARNA FILOGENIJA Lallemantia L. (LAMIACEAE): NEPODUDARNOST FILOGENETSKIH STABALA I POJAVE HGT

Masoud SHEIDAI*, Fahimeh KOOH DAR, Zeinab Moradiyan POODE

Fakultet za prirodne nauke i biotehnologiju, Shahid Beheshti Univerzitet, Teheran, Iran

Izvod

Rod Lallemantia Fisch. & C. A. Mey. (Familija Lamiaceae) je mali rod sa samo pet vrsta koje su jednogodišnje i dvogodišnje zeljaste biljke od medicinskog i hranljivog značaja. Rod je kavkaskog porekla i u Iranu sadrži pet vrsta. Ciljevi rada su bili sledeći: 1- ispitati pojavu filogenetskog “sukoba” između jedarne (ribozomalne) sekvence, jedarnih repetitivnih sekvenci i sekvenci plastida (rps16 intron, cp) u rodu Lallemantia (Lamiaceae), 2- ispitati pojavu inter-specifične hibridizacije unutar ovog roda, i 3-uporediti vreme divergentnosti vrsta od osnovne linije na osnovu molekularnih podataka. Ovo je prva analiza evolutivnog aspekta roda Lallemantia. Mi smo prvi dali molekularni prikaz pojave inter-specifične hibridizacije u rodu Lallemantia I prikazali da se filogenetski signali cp-DNK i ITS sekvenci značajno razlikuju.

Primljeno 31. XII. 2018.
Odobreno 18. X. 2018.