DNA sequence variation in the fruit tree *Inga inicuil* (Leguminosae: Mimosoideae) from contrasting altitude distinguishes two different species

Ma. Teresa P. Pulido-Salas[1], Dolores González[2]*, Arturo Pérez-Vázquez[3] and Francisco Lorea-Hernández[2]

Preliminary version

Botanical Sciences
94 (4): 1-14, 2016

DOI: 10.17129/botsci.638

Abstract

Background: *Inga inicuil* is a leguminous tree in Veracruz, Mexico whose fruits are locally commercialized. One taxonomic hypothesis suggests that *I. inicuil* is a species with a wide altitudinal range, while another segregates *I. paterno* from *I. inicuil* by altitude and morphology. The aim of this study was to explore variation in non-coding regions from the chloroplast and nuclear genomes within individuals of *I. inicuil* from contrasting altitude.
Hypothesis: Phylogenetic analyses with sequence data would separate exemplars of *I. inicuil* in two different species and as the case with other metabolites, the sugar content in the sarcotesta would vary with altitude.
Studied species: *Inga inicuil*
Study site: The mountainous region of Coatepec and the coastal region of Tolome and San Pancho in Veracruz, Mexico.
Methods: Representative individuals of *I. inicuil* were selected from localities with contrasting altitude. Maximum Likelihood and Bayesian Phylogenetic Inference approaches were used for phylogenetic analyses with sequence data from the ITS1-5.8S-ITS2 and *trn*L-F. Sugar content in ripe fruit was measured using a refractometer.
Results: Sugar content was not related to altitude or with sequence variation. However, sequences from the *trn*L-F revealed an insertion/deletion of approximately 309 nucleotides present only in trees growing at the coastal plain. Phylogenetic analyses with sequence data support the hypothesis that *I. inicuil* and *I. paterno* are two different species.
Conclusions: The remarkable differences of sequence data from the *trn*L-F within the species make necessary further research increasing taxon sampling along with its distribution range for clarifying its taxonomy.
Key words: edible fruit, indel, *Inga paterno*, molecular phylogeny, sequence variation.

La variación en las secuencias de ADN en el árbol frutal *Inga inicuil* (Leguminosae: Mimosoideae) de altitud contrastante distingue dos especies diferentes

Resumen

Antecedentes: *Inga inicuil* es una leguminosa arbórea en Veracruz, México cuyo fruto se comercializa localmente. Una hipótesis taxonómica sugiere que *I. inicuil* es una especie con amplio rango altitudinal, mientras que otra segrega *I. paterno* de *I. inicuil* por su altitud y morfología. El objetivo de este estudio fue explorar la variación en regiones no codificadoras del cloroplasto y núcleo en individuos de *I. inicuil* de altitud contrastante.
Hipótesis: Análisis filogenéticos con secuencias génicas separarán ejemplares de *I. inicuil* en dos especies diferentes, como en otros metabolitos el contenido de azúcar en la sarcotesta variará con la altitud.
Especie en estudio: *Inga inicuil*
Sitio de estudio: La región montañosa de Coatepec y la región costera de Tolome y San Pancho en Veracruz, México.
Métodos: Se seleccionaron individuos de *I. inicuil* de localidades con altitud contrastante. Se utilizó Máxima Verosimilitud e Inferencia Filogenética Bayesiana para los análisis filogenéticos con las secuencias del ITS1-5.8S-ITS2 y del *trn*L-F. El contenido de azúcar en la fruta madura se midió con un refractómetro.
Resultados: El contenido de azúcar no está relacionado con altitud o variación de las secuencias. Sin embargo, las secuencias del *trn*L-F revelaron una inserción/deleción de aproximadamente 309 nucleótidos solo en los árboles que crecen en la planicie costera. Los análisis filogenéticos apoyan la hipótesis de que *I. inicuil* e *I. paterno* son dos especies diferentes.
Conclusiones: Las notables diferencias en los datos de secuencias del *trn*L-F hacen necesario investigación adicional aumentando el muestreo a lo largo del rango de distribución de esta especie para aclarar su taxonomía.
Palabras clave: árbol frutal, filogenia molecular, indel, *Inga paterno*, variación de secuencias.

[1] Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán, Mérida, Yuc., México
[2] Red de Biodiversidad y Sistemática, Instituto de Ecología, A.C., Xalapa Ver., México
[3] Colegio de Postgraduados. Campus Veracruz, Tepetates, Mpio. Manlio Fabio Altamirano, Ver., México
* Corresponding author: dolores.gonzalez@inecol.mx

Inga inicuil Schltdl. & Cham. ex G. Don (Ingeae, Mimosoideae, Leguminosae) is a tree native to Mesoamerica where it is used as a shade tree in coffee plantations (Pennington & Sarukhán 1998). In the mountainous region of Coatepec, Veracruz, Mexico this tree and fruit are commonly known as "jinicuil". The seeds are appreciated for their sweetness fleshy sarcotesta and the pods are sold as fresh fruit. Jinicuil is also the common name used at the coastal plains of Veracruz, where people there eat the seeds cooked and salted as a snack. The cotyledons contain protein (Geilfus 1994, Bressani 2010) and germination can reach 100 % (Pennington 1997), which gives the species high potential as a food crop. Additionally, pods are sold in local markets in Coatepec. In spite of its cultural importance, there is not any agricultural management of the tree and the rate of local loss of trees in this area has been estimated at 2.4 % annually (Pulido-Salas 2009, 2013).

One challenge in achieving agricultural sustainability is the development of agroecosystems with a predominance of native species (Altieri & Nicholls 2000). To accomplish this, it is necessary to learn more about these species (Vázquez-Yanes *et al*. 1999) as well as characterize them in order to design specific management programs (Martínez-Castillo *et al*. 2004). The diversity and genetic structure of populations constitute a basic starting point for developing management protocols for native species or for initiating genetic conservation programs. If biodiversity is not preserved, heterozygosity may be reduced and introgression can occur with the concomitant deterioration of local genetic populations and of the species itself (Baverstock & Moritz 1996). Knowing the genetic diversity of native species allows for the development of long-term programs to manage, monitor, and preserve populations (Karp *et al*. 1997). To achieve this, molecular approaches allow the identification of polymorphisms in different regions of the genome with various mutation rates (Parker *et al*. 1998). There are several molecular techniques for exploring genetic diversity. Both Amplified Fragment Length Polymorphisms (AFLPs) and Random Amplified Polymorphic DNA (RAPDs) can detect genetic variation across the entire genome. Others include Restriction Fragment Length Polymorphisms (RFLPs), Simple Sequence Repeats (SSR, microsatellites) or genic sequences (González 1998), which can detect variation in more specific regions of the genome. For example, in an AFLP study with 56 samples of 30 varieties of *Macadamia* spp., it was possible to quantify the degree of polymorphism in the varieties and their progenitors, and to propose a possible differentiated agricultural management (Robledo 2003). In another study, using RFLP (Restriction Fragment Length Polymorphism) markers in fragmented populations of *Sorbus torminalis*, a species appreciated for its fruit and wood, significant genetic variation was found that could be attributed to recent changes in silvicultural practices in Central Europe (Angelone *et al*. 2007). Also, a study with SSR markers for the characterization and analysis of genetic diversity in cultivars of *Corylus avellana* found high genetic diversity and polymorphism that was suitable for distinguishing different hazelnut lineages (Bassil *et al*. 2013). Furthermore, research with DNA sequence of non-coding regions has also provided insight into genetic diversity in a variety of organisms because these segments of DNA are less functionally constrained and are, therefore, more variable (González & Vovides 2002). For example, a study with the *trn*H-*trn*K non-coding region of chloroplast DNA sequence data in apricot germplasm revealed significant genetic diversity of nucleotide sequences in both haplotypic and nucleotide diversity tests. The results provided some clues for the origin of apricot species and useful information for the management of apricot genetic resources (Batnini *et al*. 2014).

There are few studies on genetic diversity for species of *Inga*, particularly for *I. inicuil* and *I. paterno* Harms. Those that exist focus mainly on the preservation of species in tropical forests, such as *I. edulis* (*e.g*. Hollingsworth *et al*. 2005, Dawson *et al*. 2008, Rollo *et al*. 2016), *I. thibaudiana* (Schierenbeck *et al*. 1997), and *I. vera* (Cruz-Neto *et al*. 2014). Other studies of *Inga* have proposed hybridization between species that grow together in coffee plantations. Pennington (1997) recorded seven species that were auto-incompatible and inter-sterile and observed that the best fruit of *I. inicuil* (*i.e*. longer pods and sweeter sarcotesta) was on trees separated by over one kilometer. The possibility of hybridization, however, has not been proven genetically.

On the other hand, chemical studies of *Inga* species have discovered a relationship between altitude and the amount of some chemical compounds. For example, it has been found that the amount of pipeolitic acid (a non-proteinic amino acid) in the leaves of several species varies according to altitude. Although *I. inicuil* was not included in that study, it was shown that species

growing at a similar altitude have similar patterns of pipeolitic acid, which is interpreted as a defense strategy against predators or pathogens such as ants and fungi (Morton *et al*. 1991, Kite 1997). Also, Koptur (1985) reported varying concentrations of phenolic compounds in leaves of *I. densiflora* and *I. punctata* that are also correlated with altitude. Nonetheless, it is unknown if there is a connection between altitude and sugar content in the fruit of *I. inicuil*.

Inga has great morphological variation (Pennington 1997). It is considered a genus in the process of diversification (Richardson *et al*. 2001). It includes around 300 species within the tropical Americas, ranging in altitude from sea level to 3,000 m. The genus has had a complex taxonomic history and unstable nomenclature (*e.g*. Sousa 1993, 2001, 2009, Pennington 1997, Brown *et al*. 2008). Currently, two taxonomic hypotheses exist for *I. inicuil*. The first states that what is commonly known as "jinicuil" belongs to two species described as *I. inicuil* and *I. paterno*, with morphological differences and separated by altitude (Sousa 1993, 2001, Ricker *et al*. 2013). *Inga paterno* was described having stipules persistent, from 9 to 22 mm long, and pedicel thin, up to 5 mm long growing between 0-800 m a.s.l.; while *I. inicuil* has stipules soon deciduous, from 5 to 9 mm long, and pedicel short and robust, of 1.5 mm long and grows above 850 m a.s.l. The second hypothesis proposes that jinicuil is a single species (*I. inicuil*) with a wide altitudinal distribution, spanning from the mountainous cloud forest to the regions with evergreen and sub-deciduous forests near sea level and with wide ranges of morphological variation (Pennington 1997). Under this scenario the name *I. paterno* falls into synonymy with the older name *I. inicuil* (Pennington 1997, Pennington & Sarukhán 1998, Groom 2012). In a phylogenetic analysis with nuclear and chloroplast DNA, the genus *Inga* as a whole was shown to be monophyletic. However, very little variation was found in the sequences generating unresolved topologies, which led to the conclusion that *Inga* represents a recently diversified genus (Richardson *et al*. 2001). In a molecular phylogenetic analysis aimed at elucidating the evolution of defenses to herbivory in the genus, several species of *Inga* were analyzed (Kursar *et al*. 2009), and the results confirmed the lack of variation within species of *Inga*. None of these studies included *I. inicuil* or *I. paterno*. The aim of this study was to explore variation in sequences of non-coding regions from the chloroplast and nuclear genomes within individuals of *I. inicuil* from contrasting altitude. We hypothesized that phylogenetic analyses with sequence data would separate exemplars of *I. inicuil* in two different species and as the case with other metabolites, the sugar content in the sarcotesta would vary with altitude.

Materials and methods

Material studied. From 57 located trees, 22 adult individuals were selected (DBH > 15 cm). Trees were not physically close to each other to avoid possible kinship, and were chosen from contrasting altitude. From these, 19 were from the municipality of Coatepec, Veracruz, located between 850 and 1,530 m a.s.l. on the edge of the central mountainous region of Mexico and were used as shade trees in coffee plantations, while three were from the town of Tolome, municipality of Paso de Ovejas, Veracruz in the coastal region at 50 m a.s.l. Fertile branches/samples from the chosen trees were collected for herbarium specimens and some were also sent to a taxonomist specialist on *Inga* for an accurate identification. Representative collections were deposited at the herbaria XAL, Instituto de Ecología, A. C. in Xalapa, Veracruz and MEXU, Instituto de Biología, Universidad Nacional Autónoma de México in Mexico City.

Sugar content in fruit. Given that the sweetness of the pulp covering the seeds is one of the characteristics that increases the commercial potential of *I. inicuil*, and with the hypothesis that, as in the case with other metabolites, there could be a connection between altitude and sugar content, we took a sample of 16 pods per tree from 19 trees from the mountain region and three from the coastal region, four for each quarter of an imaginary square on the crown (Martínez-Moreno *et al*. 2006). The sugar content was measured (°Brix average/tree) in ripe fruit of the juice coming out from the sarcotesta using a refractometer (ATAGO ATC-1; range 0-32 °Brix) during the summer. The four highest values per tree were used to calculate the average. Different seeds from the same pod were measured separately. The Pearson correlation coefficient was measured in Excel.

DNA extraction and amplification. For DNA extraction, 25 representative trees were selected from localities with contrasting altitude (Table 1). Twelve were from the mountainous region (municipality of Coatepec) and 13 from the coastal region (three from Tolome and ten from San Pancho, in the municipalities of Paso de Ovejas and La Antigua, respectively). Genomic DNA was extracted from young recently collected leaves, as suggested by Doyle & Doyle (1987). Extraction and purification of DNA was conducted with the DNeasy plant mini kit (Qiagen, Valencia, CA, USA), following the instructions of the manufacturer. To verify the quantity and quality of the DNA, an aliquot was taken from the extraction and loaded in a 1.2 % agarose gel dyed with ethidium bromide. A molecular-weight marker of known concentration (25 ng/μl) was included in the gel. Each sample was amplified with specific primers for two molecular markers, one nuclear and the other from chloroplast. The nuclear marker corresponded to the ITS1-5.8S-ITS2 region (ITS), and the chloroplast marker to a section of the *trn*L-F intergenic spacer. Primers used for amplifying the ITS were ITS1 (5'- TCCGTAGGTGAACCTGCGG -3') and ITS-4 (5'-TCCTCCGCTTATTGATATGC-3'; White *et al*. 1990). Primers for the *trn*L-F were "e" and "f" (5'-GGTTCAAGTCCCTCTATCCC-3' and 5'-ATTTGAACTGGTGACACGAG-3' respectively; Taberlet *et al*. 1991). Reactions were performed in a 25 μl mixture containing 10-20 ng of DNA, 5 μl of PCR buffer, 200 μM of each of the four deoxynucleoside triphosphates, 5 pmol of each primer, 2.5 mM $MgCl_2$, 2.5 U of Go Taq flexi DNA polymerase (Promega, Madison, WI, USA), and distilled water to volume. The amplifications were performed on a thermocycler (Mastercycler, Eppendorf, Hamburg, Germany). The amplification program included an initial denaturation at 94 °C for 5 min, followed by 35 cycles with denaturation at 94 °C for 1 min, annealing at 51-59 °C for 1 min, and extension at 72°C for 2 min, and a final extension for 7 min at 72 °C.

Table 1. Municipality, altitude and GenBank accession numbers for specimens used in phylogenetic analyses. *indicates specimens with an insertion/deletion (indel) of about 309 bp in the *trn*L-F region.

Collection code	Municipality (Veracruz, Mexico)	Altitude (m a.s.l.)	GenBank accession number	
			ITS	*trn*L-F
MS III	Coatepec	850	KR186220	KR186245
JMcv I	Coatepec	1,050	KR186221	KR186246
JMcv II	Coatepec	1,050	KR186222	KR186247
JMcv IV	Coatepec	1,050	KR186223	KR186248
JMor IV	Coatepec	1,060	KR186224	KR186249
MA III	Coatepec	1,150	KR186225	KR186250
CG I	Coatepec	1,150	KR186226	KR186251
CG III	Coatepec	1,150	KR186227	KR186252
TS III	Coatepec	1,150	KR186228	KR186253
RR I	Coatepec	1,260	KR186229	KR186254
JB I	Coatepec	1,528	KR186230	KR186255
JB II	Coatepec	1,350	KR186231	KR186256
RoTo	Paso de Ovejas	50	KR186232	*KR186257
CaTo	Paso de Ovejas	50	KR186233	*KR186258
3To	Paso de Ovejas	50	KR186234	*KR186259
M5	San Pancho	20	KR186235	*KR186260
M5-A	San Pancho	20	KR186236	*KR186261
M6	San Pancho	20	KR186237	*KR186262
M6-A	San Pancho	20	KR186238	*KR186263
M7	San Pancho	20	KR186239	*KR186264
M7-A	San Pancho	20	KR186240	*KR186265
M8	San Pancho	20	KR186241	*KR186266
M8-A	San Pancho	20	KR186242	*KR186267
M9	San Pancho	20	KR186243	*KR186268
M9-A	San Pancho	20	KR186244	*KR186269

DNA sequencing and analyses. Amplified DNA was purified before sequencing with the Wizard SV gel and PCR clean-up system kit as described by the manufacturer (Promega, Madison, WI, USA), and sequenced using ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) according to manufacturer's instructions. Cycle sequence products were cleaned with an isopropanol precipitation and electrophoresed using an ABI 310 genetic analyzer (Applied Biosystems, Foster City, CA, USA). The resulting sequences were edited with BioEdit software version 7.1.3.0 (Hall 1999) and aligned using ClustalW with default parameters for gap opening and extension penalties (Thompson *et al*. 1994). Sequence variation was obtained using the Arlequin software v. 3.1 (Excoffier *et al*. 2005). Phylogenetic analyses of sequence data were performed separately and combined with Maximum Likelihood (ML) using GARLI v. 0.951 (Zwickl 2006) and Bayesian Phylogenetic Inference (BPI) using MrBayes 3.1.2 (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003, Altekar *et al*. 2004). BPI analyses were executed with model parameters GTR + Γ (nst = 6; rates = gamma), and default values for priors. ML analyses were performed specifying in the script that model parameters were computed at the same time during the searches. Analyses consisted of 10 replicates to ensure that results were consistent and reproducible. Branch support for ML was determined simultaneously by performing 100 non-parametric bootstrap iterations in each of the 10 replicates. BPI analyses comprised two independent 1-million generation runs, with four chains (one cold and three hot) each, until an average standard deviation of split frequencies of 0.01 or less was reached. We sampled trees every 100th generation and discarded initial samples applying a "burnin" value of 25 % before calculating the majority consensus tree and posterior probabilities (PP) for clades.

Four matrices were prepared for analyses. One included only *trn*L-F sequences from 69 terminals. They corresponded to twenty-five sequences generated in this study (12 from the mountainous region and 13 from the coastal plains), and 38 from GenBank. From the GenBank sequences, 28 correspond to identified *Inga* species (*I. alba*, *I. edulis* and *I. thibaudiana* with two, four and two representatives respectively), five to non-identified *Inga* spp., and six to representatives from the tribe Ingeae, which were used as the outgroup (*Zapoteca sousae*, *Pithecellobium diversifolium*, *Abarema piresii*, and three for *Zygia* spp.). Two matrices contained only ITS sequences. One comprised 464 terminals, 25 from this study and 434 from GenBank. From these, 45 correspond to identified *Inga* species (some species having up to 27 representatives), 52 to non-identified *Inga* spp., and five representatives from the outgroup. Other matrix had only 110 terminals. This included 25 from this study, two of each identified *Inga* species (10 species with only one) and the representatives of the tribe Ingeae. A concatenated matrix was also constructed. This included the 25 terminals from this study, two species of *Inga* with same voucher (*I. edulis* and *I. thibaudiana*) that have sequences in GenBank for both loci, and three representatives of the tribe Ingeae. All cladograms were edited with Adobe Illustrator v 13.0.2. Alignment and resulting trees from multi-locus gene analyses are deposited in TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S17506).

Results

Sugar content and sequence analyses. The analysis of sugar content in the fruits revealed unexpected variation. Values obtained do not show any discernible pattern (Figure 1) and are not homogeneous among seeds even in the same pod. For example, data obtained from seeds in one of the pods tested varied from 13 to 18 °Brix. There was also no relationship of sugar content and altitude (Figure 2; r^2 = -0.05) or a relationship of sugar content with DNA sequence variation.

The results for sequence data confirmed that sequences from both loci had little variation. Nevertheless, DNA sequences from ITS and *trn*L-F in exemplars from the coastal and mountainous regions varied according to altitude. The *trn*L-F region did not have any nucleotide substitution among the 25 obtained sequences, but there was an insertion/deletion (indel) of about 309 base pairs (bp) exclusively in the 13 trees from the coastal region (marked with an asterisk in Table 1). The sizes of the amplicons for the *trn*L-F with the primers e-f were 495 bp for the 12 exemplars from the mountain region and 186 bp for the 13 trees from the coastal region (Figure 3). This indel is not present in any of the 33 species of *Inga* that have sequences for this locus

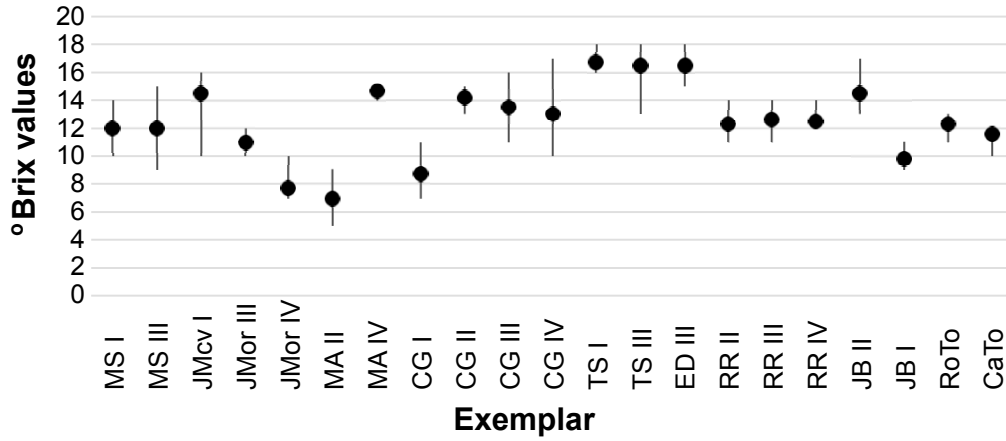

Figure 1. Values of sugar content in each tree and its ranges of variability.

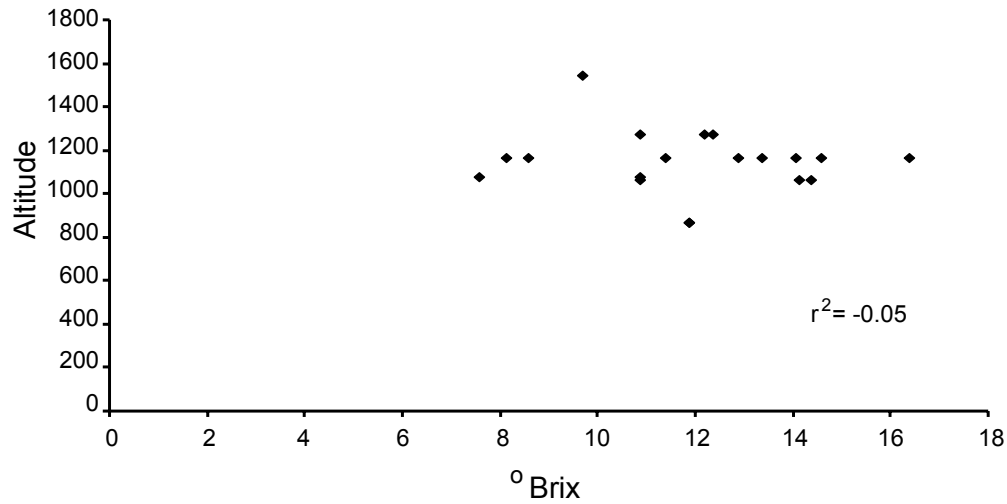

Figure 2. Relationship between altitude and sugar content in the fleshy seedcoat of *Inga inicuil*.

in GenBank. In contrast, sequences from the ITS have only nucleotide substitutions among the 25 sequences obtained in this study. The size of the amplicon for the ITS with the primers ITS1-ITS4 was 657 bp (Table 2). As with the *trn*L-F locus, there was distinctive variation in nucleotide substitutions for the 13 exemplars from the coastal plains (Figure 4) among 437 exemplars of *Inga* that have ITS sequences in GenBank.

Phylogenetic analyses. ML and BPI analyses with the *trn*L-F locus and 69 terminals generated unresolved cladograms (Figure 5). Low DNA sequence variation was a problem for the analy-

Table 2. Summary of sequence variation in the fruit tree *Inga inicuil*. Most of the polymorphic sites are due to the indel in exemplars from the coastal region.

Gene partition	Tree Provenance and (number)	Amplicon size (bp)	After alignment	No. of polymorphic sites	Ts	Tv	Indels of one nucleotide	Longer indels (bp)
*trn*L-F	Mountainous region (12)	495	495	316	6	1	1	1 (309)
	Coastal region (13)	186						
ITS	Mountainous region (12)	657	657	14	12	2	0	0
	Coastal region (13)							

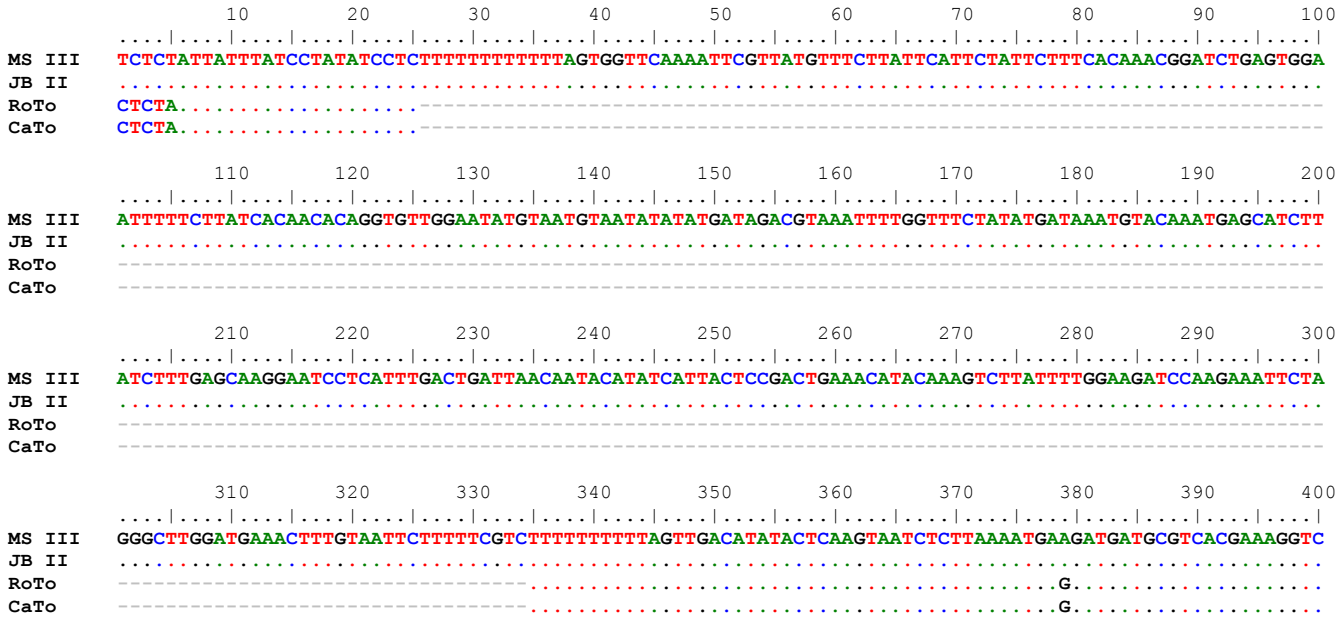

Figure 3. Sequence alignment of a section of the *trn*L-F gene in four exemplars of *Inga* showing the large indel in exemplars of the coastal plains (Roto and CaTo). MS III and JB II were collected from the mountainous region.

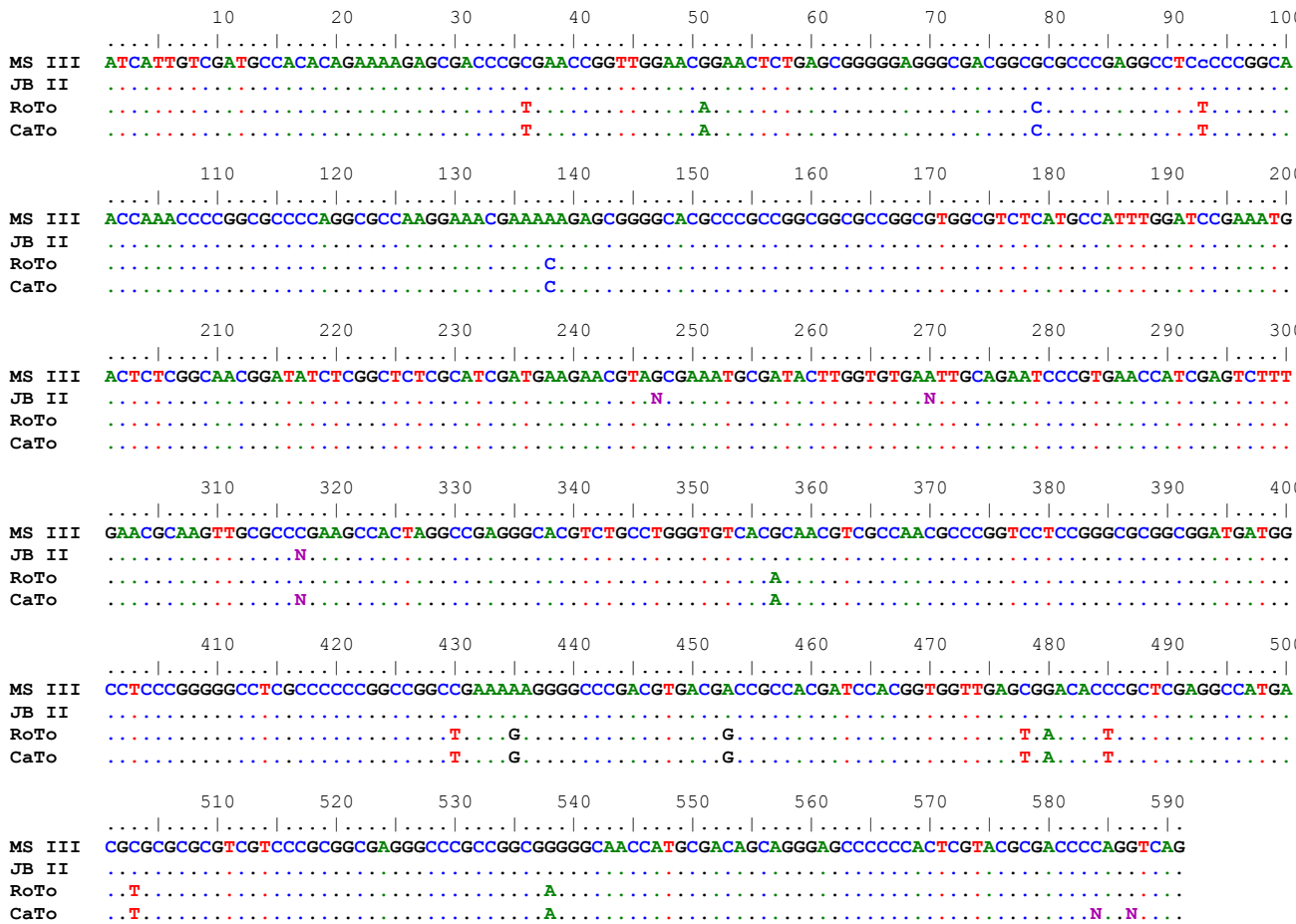

Figure 4. Sequence alignment of a section of the ITS gene in four exemplars of *Inga* showing differences in nucleotide substitutions between exemplars from the coastal plains (Roto and CaTo), and from the mountainous region (MS III and JB II).

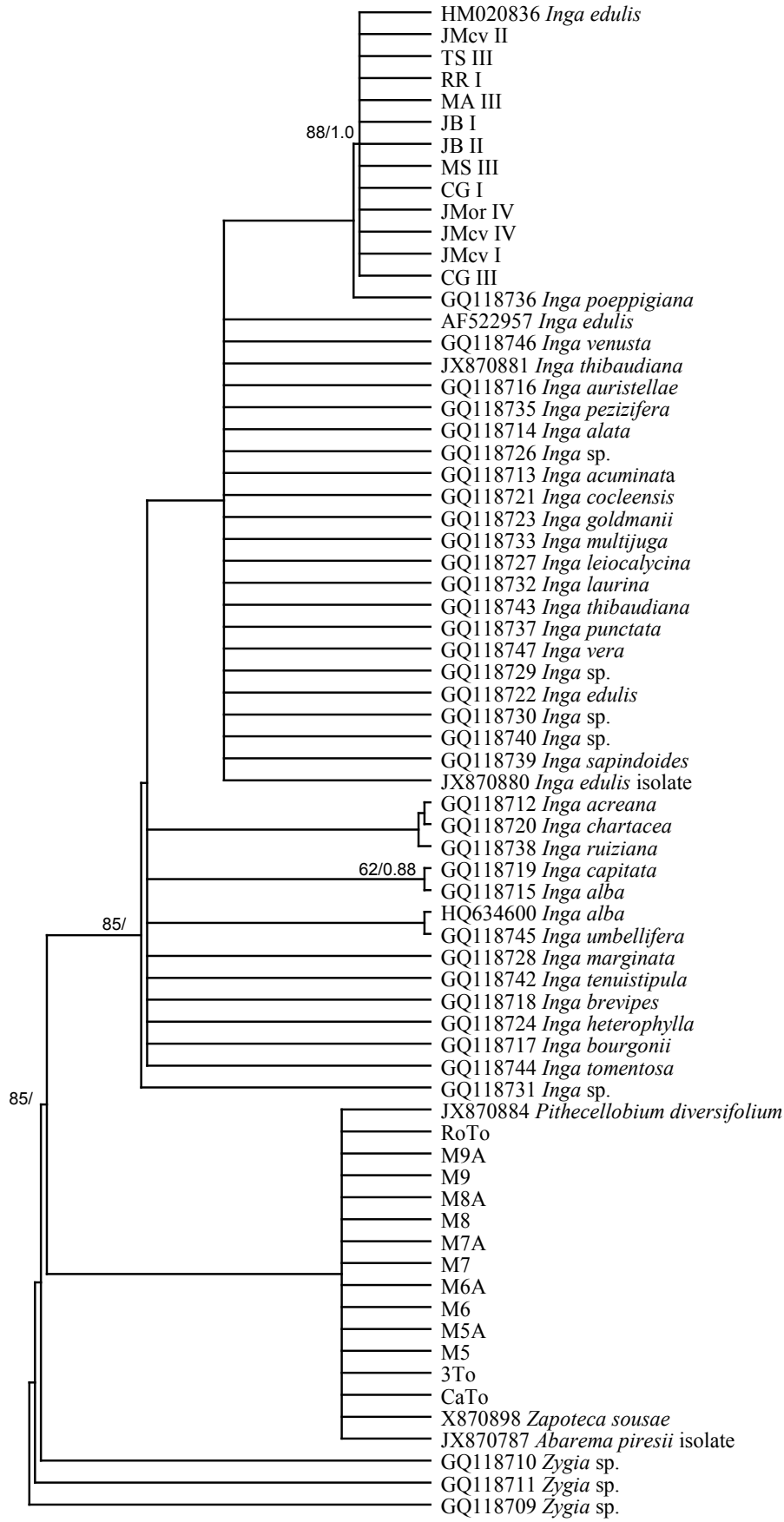

Figure 5. Phylogenetic tree for the *trn*L-F spacer showing the topological placement of *Inga*'s exemplars from the mountainous and coastal regions relative to other *Inga* spp. Analysis was performed with Maximum Likelihood. Values of support are indicated above branches (ML bootstrap/PP).

Figure 6. Phylogenetic tree for the ITS region showing the topological placement of *Inga*'s exemplars from the mountainous and coastal regions relative to other *Inga* spp. Analysis was performed with Maximum Likelihood. Values of support are indicated above branches (ML bootstrap/PP).

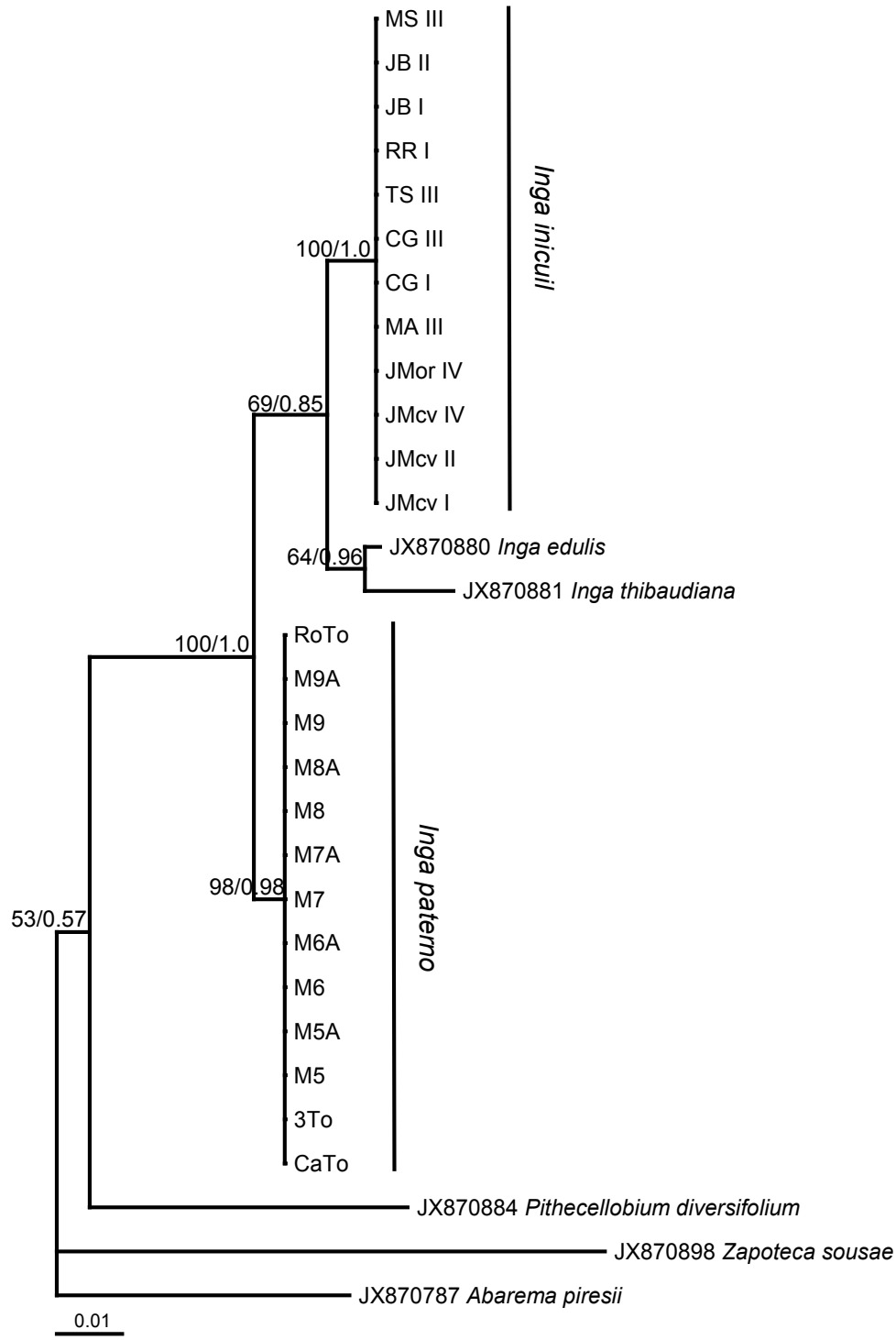

Figure 7. Hypothesis of inferred relationships of *Inga inicuil* and *I. paterno*, based on Maximum Likelihood of the concatenated data set of two loci (ITS and *trn*L-F). Values of support are indicated above branches (ML bootstrap/PP).

ses due to few characters. The best tree from the ML analysis had a log likelihood (-lnL) score of -1333.8458 and revealed a relationship between the 13 exemplars from the coastal plains and *Pithecellobium diversifolium*, *Zapoteca sousae*, and *Abarema piresii*. However, this result is very likely artificial since it is the result of lack of characters in the sequences with the 309 bp indel. In contrast, ML analysis with the matrix of 464 sequences for the locus ITS generated a tree with several groups in terminal branches, but without resolution in basal branches (not shown), supporting the proposed recent divergence of the species within this genus. In the second ML analysis with only 110 ITS sequences (-lnL = 2755.6728), *Inga auristellae* was the

Table 3. Comparative features of *Inga inicuil* and *I. paterno*

Collection code	Collector number (Pulido-Salas)	Locality, Ver., Mexico	Identified taxon	Distinctive morphological characters (Sousa 1993)
RoTo	1016	Tolome	*I. paterno*	Altitude: 0-800 m above sea level
CaTo	1015	Tolome	*I. paterno*	Stipules: persistent, from 9 to 22 mm.
				Pedicel: thin, up to 5 mm long.
TS III	1031	Coatepec	*I. inicuil*	Altitude: 1400-2400 m above sea level
RR I	1034	Coatepec	*I. inicuil*	Stipules: soon deciduous, from 5 to 9 mm.
JB I	1037	Coatepec	*I. inicuil*	Pedicel: short and robust, of 1.5 mm long

sister group to the 13 exemplars of the coastal plains and *I. suaveolens* to the 12 exemplars of the mountainous region (Figure 6). Our results indicate that the sequences generated in this study do not correspond to any of *Inga* species deposited in GenBank. Analyses of ML and BPI of combined sequence data showed a clear distinction between the exemplars of the coastal plains and the mountainous region. Both clades had high support values. Clade containing exemplars of the coastal plains had a bootstrap value of 98 %, and a PP of 0.98 while clade containing exemplars from the mountainous region had 100 and 1.0 respectively (Figure 7). ML analysis recovered a tree with a log likelihood score of -3106.1068. Consequently, molecular data and morphologically distinctive characters noted by the taxonomist of the genus on herbarium specimens (Table 3) show clearly that both sources of information distinguish between exemplars from the mountainous region and the coastal region, supporting the hypothesis that they are two different species.

Discussion

Variation in the sugar content of the sarcotesta did not support one of our hypotheses given that there was no relationship between altitude and sugar content, as has been observed with phenolic compounds and pipeolitic acid. The possible causes for the heterogeneous distribution of sugar could due to intrinsic factors in the production of fruits in *Inga* similar to those observed in other species, such as auto-incompatibility, limited resources in the production of fruits, selective abortion, and issues related to the ontogenetic development of the fruit (Koptur 1983, 1984, 1985). Also, the possible spatial overlap of several species of *Inga* in coffee plantations may favor cross-pollination as has been suggested by Richardson *et al*. (2001). It can be inferred that the heterogeneity in the sugar content in the fruit shows the lack of genetic enhancement of this fruit-tree, which opens the possibility of planning genetic improvement programs for more productive use of the seeds and fruits.

Molecular variation and altitude. The variation found in the *trn*L-F and ITS sequences reflects two clearly defined lineages according to the altitudinal origin of the samples. DNA sequences were homogeneous in all 12 trees sampled from the mountainous region (municipality of Coatepec, Ver.); similarly, in the 13 trees from the coastal plains (Tolome and San Pancho, municipality of Paso de Ovejas and La Antigua Ver.). However, there were important dissimilarities at the nucleotide level between both groups. The conserved indel in trees from the coastal region appears to be a synapomorphic character delimiting natural lineages as has been observed in other organisms (*e.g*. Calviño & Downie 2007, Chiari *et al*. 2009, Soltis *et al*. 1998). Indels from the *trn*L-F region have resulted good phylogenetic markers for some taxa (*e.g*. Richardson *et al*. 2000, Holt *et al*. 2004, Ghamkhar *et al*. 2007, Drábková *et al*. 2004), as well as homoplasious for others (*e.g*. Kellermann & Udovicic 2007). The *trn*L-F region is often of different length, which can make sequence alignment and the determination of homologous bases a matter of concern in phylogenetic analyses (González *et al*. 2006). However, considering our results we conclude that although not devoid of homoplasy, indels can be useful markers of shared history at lower taxonomic levels as in the case of *Inga inicuil* and *I. paterno*. Differences found in this

work support previous observations based on morphological characters (Sousa 1993) suggesting that *I. inicuil* and *I. paterno* are distinct species occupying habitats at different altitudes: *Inga inicuil* corresponds to specimens collected from Coatepec, Ver., (range 850 to 1,530 m a.s.l.), while trees sampled in Tolome and San Pancho (20–50 m a.s.l.) correspond to *I. paterno* (Sousa 1993). Due to the clear differences found between the trees from the mountainous versus the coastal regions, it would be useful to continue the study with a detailed ethnobotanical exploration of the uses for seeds and fruits of these species by people in localities at different altitudes. Also, the remarkable differences of sequence data from the *trn*L-F region within the species make further research necessary, increasing taxon sampling along with its distribution range. The study of DNA variation in what popularly is called "jinicuil" in the physiographically diverse state of Veracruz, Mexico, provides useful data that help to clarify the taxonomy of this group and which in turn, considering the uses and potential food of these species, provide a firmer basis for designing sustainable management of these native fruit-trees.

Acknowledgments

Thanks to the Consejo Nacional de Ciencia y Tecnología (CONACyT-México) for the scholarships 197852 to MTPP-S, and the Colegio de Postgraduados (colpos.mx) through the research group Agronegocios, Agroecoturismo y Arquitectura del Paisaje. We thank to Griselda Benítez from INECOL for help in field work; Mario Sousa Sánchez at the herbarium MEXU (Universidad Nacional Autónoma de México in Mexico City) for identifying the *Inga* specimens. We also acknowledge to David Gernandt and Tom Wendt for the edition of the English language, and to Bertha Pérez and Elsa Utrera for some technical work for obtaining sequences. We also thank the use of equipment financed with project 103158 CONACyT-México to DG.

Literature cited

Altekar G, Dwarkadas S, Huelsenbeck JP. Ronquist F. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**: 407-415.

Altieri M, Nicholls C. 2000. *Agroecología teoría y práctica para una agricultura sustentable*. Serie Textos Básicos para la Formación Ambiental, Programa de las Naciones Unidas para el Medio Ambiente, México D.F.

Angelone S, Hilfiker K, Holderegger R, Bergamini A, Hoebee SE. 2007. Regional population dynamics define the local genetic structure in *Sorbus torminalis*. *Molecular Ecology* **16**: 1291-1301.

Bassil N, Boccacci P, Botta R, Postman J, Mehlenbacher S. 2013. Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. *Genetic Resources and Crop Evolution* **60**: 543-568.

Batnini M-A, Bourguiba H, Trifi-Farah N, Krichen L. 2014. Chloroplast DNA sequence data provides new insights into genetic diversity and phylogenetic relationships of Tunisian apricot germplasm. *Scientia Horticulturae* **178**: 241-247.

Baverstock PR, Moritz C. 1996. Project design. *In*: Hillis DM, Moritz C Mable BK, eds. *Molecular Systematics*. Sunderland: Sinauer, 17-47

Bressani R. 2010. Valoración química nutricional de la harina de semillas de diferentes especies de *Inga* (*I. jinicuil I. laurina/paterno I. vera*). *Informe final Proyecto FONACYT-SENACYT-CONCYT. Guatemala*.

Brown GK, Murphy DJ, Miller JT, Ladiges PY. 2008. *Acacia s.s.* and its relationship among tropical legumes, Tribe Ingeae (Leguminosae: Mimosoideae). *Systematic Botany* **33**: 739-751.

Calviño CI, Downie SR. 2007. Circunscription and phylogeny of Apiaceae subfamily Saniculoideae based on chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* **44**: 175-191.

Chiari Y, van der Meijden A, Madsen O, Vences M, Meyer A. 2009. Base composition, selection, and phylogenetic significance of indels in the recombination activating gene-1 in vertebrates. *Frontiers in Zoology* **6**: 32.

Cruz-Neto O, Aguiar AV, Twyford AD, Neaves LE, Pennington RT, Lopes AV. 2014. Genetic and ecological outcomes of *Inga vera* subsp. *affinis* (Leguminosae) tree plantations in a fragmented tropical landscape. *PLoS ONE* **9**:e99903.

Dawson IK, Hollingsworth PM, Doyle JJ, Kresovich S, Weber JC, Sotelo-Montes C, Pennington TD, Pennington RT. 2008. Origin and genetic conservation of tropical trees in agroforestry systems: a case study from the Peruvian Amazon. *Conservation Genetics* **9**: 361–372

Drábková L, Kirschner J, Vlček Č, Pačes V. 2004. *Trn*L-*trn*F intergenic spacer and *trn*L intron define major clades within *Luzula* and *Juncus* (Juncaceae): importance of structural mutations. *Journal of Molecular Evolution* **59**: 1-10.

Doyle JJ, Doyle JL 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Excoffier L, Laval G, Schneider S. 2005. Arlequin Version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**: 47-50.

Geilfus F. 1994. *El árbol al servicio del agricultor*. Manual de agroforesteria para el desarrollo rural. CATIE Enda-Caribe, Turrialba.

Ghamkhar K, Marchant AD, Wilson KL, Bruhl JJ. 2007. Phylogeny of Abildgaardieae (Cyperaceae) inferred from ITS and *trn*L-F data. *Aliso: A Journal of Systematic and Evolutionary Botany* **23**: 149-164.

González D. 1998. Marcadores moleculares para los estudios comparativos de la variación en ecología y sistemática. *Revista Mexicana de Micología* **14**: 1-21.

González D, Vovides AP. 2002. Low intralineage divergence in *Ceratozamia* (Zamiaceae) detected with nuclear ribosomal DNA ITS and chloroplast DNA *trn*L-F non-coding region. *Systematic Botany* **27**: 654-661.

González D, Cubeta MA, Vilgalys R. 2006. Phylogenetic utility of indels within ribosomal DNA and β-tubulin sequences from fungi in the *Rhizoctonia solani* species complex. *Molecular Phylogenetics and Evolution* **40**: 459-470

Groom A. 2012. *Inga jinicuil*. The IUCN Red List of Threatened Species 2012: e.T19893012A19999543. http://www.iucnredlist.org/details/19893012/0 (accessed April 8, 2015).

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT *Nucleic Acids Symposium Series* **41**: 95-98.

Holt SDS, Horová L, Bureš P. 2004. Indel patterns of the plastid DNA *trn*L-*trn*F region within the genus *Poa* (Poaceae). *Journal of Plant Research* **117**: 393-407.

Hollingsworth PM, Dawson IK, Goodall-Copestake WP, Richardson JE, Weber JC, Sotelo-Montes C, Pennington RT. 2005. Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. *Molecular Ecology* **14**: 497-501

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. 1997. *Molecular tools in plant genetic resources conservation a guide to the technologies PLGRI* International Plant Genetic Resources Institute, Rome.

Kellermann J, Udovicic F. 2007. Large indels obscure phylogeny in analysis of chloroplast DNA (*trn*L-F) sequence data: Pomaderreae (Rhamnaceae) revisited. *Telopea* **12**: 1-22.

Kite G. 1997. Non-protein amino acid chemistry. *In*: Pennington TD, ed. *The genus Inga: Botany*. Kew: Royal Botanic Gardens, 35-70.

Koptur S. 1983. Flowering phenology and floral biology of *Inga* (Fabaceae Mimosoideae). *Systematic Botany* **8**: 354-368.

Koptur S. 1984. Outcrossing and pollinator limitation of fruit set: breeding systems of neotropical *Inga* trees (Fabaceae Mimosoideae). *Evolution* **38**: 1130-1143.

Koptur S. 1985. Alternative defenses against hervibores in *Inga* (Fabaceae Mimosoideae) over an elevation gradient. *Ecology* **66**: 1639-1650.

Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG, Murakami ET, Drake C, McGregor R, Coley PD. 2009. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 18073-18078.

Martínez-Castillo J, Zizumbo-Villarreal D, Perales-Rivera H, Colunga-GarcíaMarín P. 2004. Intraspecific diversity and morpho-phenological variation in *Phaseolus lunatus* L. from the Yucatan Peninsula Mexico. *Economic Botany* **58**: 354-380.

Martínez-Moreno E, Corona-Torres T, Avitia-García E, Castillo-González AM, Terrazas-Salgado T, Colinas-León MT. 2006. Caracterización morfométrica de frutos y semillas de nanche (*Byrsonima crassifolia* (L.) H.B.K.). *Revista Chapingo Serie Horticultura* **12**: 11-17.

Morton CT, Zektzer AS, Rife JP, Romeo JT. 1991. Trans-4-methoxypipecolin acid, an amino acid from *Inga paterno*. *Phytochemistry* **30**: 2397-2399.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA. 1998. What molecules can tell us about populations: Choosing and using a molecular marker. *Ecology* **79**: 361-382.

Pennington RT, Sarukán J. 1998. *Árboles tropicales de México. Manual para la identificación de las principales especies*. Ciudad de México: Universidad Nacional Autónoma de México / Fondo de Cultura Económico.

Received:
XXXXX

Accepted:
XXXXX

Pennington TD. 1997. *The genus Inga*: Botany. Kew: Royal Botanic Gardens.

Pulido-Salas MT. 2009. *Inga jinicuil ("jinicuil") en Coatepec, Ver., México, como árbol frutal nativo: potencial económico y factores que afectan su permanencia*. Tesis de Doctorado, Colegio de Postgraduados — Campus Veracruz, Tepetates, Veracruz, México.

Pulido-Salas MT, Equihua M, Pérez-Vázquez A. 2013. *Inga jinicuil* (jinicuil) como árbol frutal nativo: valor comercial, calidad de frutos y valor cultural como determinantes para permanecer o ser eliminados. *In*: Reyes-Cervantes H, Ibarra-Contreras M, Loya-Monares N, eds. *Matemáticas, Estadística y Medio Ambiente*. Puebla: Benemérita Universidad Autónoma de Puebla. 54-65.

Richardson JE, Fay MF, Cronk QCB, Bowman D, Chase MW. 2000. A phylogenetic analysis of Rhamnaceae using *rbc*L and *trn*L-F plastid DNA sequences. *American Journal of Botany* **87**: 1309-1324.

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**: 2242-2245.

Ricker M, Hernández HM, Sousa M, Ochoterena H. 2013. Tree and tree-like species of Mexico: Asteraceae, Leguminosae, and Rubiaceae. *Revista Mexicana de Biodiversidad* **84**: 439-470.

Robledo-Martínez JD. 2003. *Caracterización genética de germoplasma de macadamia (Macadamia spp.)*. Tesis de Maestría. Colegio de Postgraduados, Montecillo, Estado de México, México.

Rollo A, Lojka B, Honys D, Mandák B, Chia Wong JA, Santos C, Costa R, Quintela-Sabarís C, Ribeiro MM. 2016. Genetic diversity and hybridization in the two species *Inga ingoides* and *Inga edulis*: potential applications for agroforestry in the Peruvian Amazon. *Annals of Forest Science* **73**: 425-435.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Schierenbeck KA, Skupski M, Lieberman D, Lieberman M. 1997. Population structure and genetic diversity in four tropical tree species in Costa Rica. *Molecular Ecology* **6**: 137-144.

Soltis DE, Soltis PS, Doyle JF. (eds.) 1998. *Molecular Systematic of Plants II: DNA sequencing*. New York: Kluwer Academy Publishers.

Sousa MS. 1993. El género *Inga* (Leguminosae: Mimosoideae) del sur de México y Centroamérica, estudio previo para la Flora Mesoamericana. *Annals of the Missouri Botanical Garden* **80**: 223-269.

Sousa M. 2001. *Inga*. *In*: Stevens WD, Ulloa Ulloa C, Pool A, Montiel OM, eds. *Flora de Nicaragua Vol. 2, Angiospermas (Fabaceae-Oxalidaceae)*. St. Louis: Monographs in Systematic Botany from the Missouri Botanical Garden, 1467-1483.

Sousa M. 2009. Adiciones al género *Inga* (Ingeae, Mimosoideae, Leguminosae) para la Flora Mesoamericana. *Acta Botanica Mexicana* **89**: 25-41.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105-1109.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Vázquez-Yanes C, Batis-Muñoz AI, Alcocer-Silva MI, Gual-Díaz M, Sánchez-Dirzo C. 1999. Árboles y arbustos nativos potencialmente valiosos para la restauración ecológica y la reforestación. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Ciudad de México.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In*: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols a guide to methods and applications*. San Diego: Academic Press, 315-322

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph. Thesis, University of Texas at Austin, Austin, Texas.