CATIE Centro Agronómico Tropical de Investigación y Enseñanza
Tropical Agricultural Research and Higher Education Center

PROGRAMA DE EDUCACIÓN PARA EL DESARROLLO Y LA CONSERVACIÓN

ESCUELA DE POSGRADO

Genetic diversity and mating system analysis of *Cedrela odorata* L. (Meliaceae) populations under different human dominated landscapes and primary forests

Tesis sometida a consideración de la Escuela de Posgrado, Programa de Educación para el Desarrollo y la Conservación del Centro Agronómico Tropical de Investigación y Enseñanza como requisito para optar por el grado de:

Magister Scientiae en Manejo y Conservación de Bosques Naturales y Biodiversidad

Por

Luis Gustavo Hernández Sánchez

Turrialba, Costa Rica, 2008

Esta tesis ha sido aceptada en su presente forma por el Programa de Educación para el Desarrollo y la Conservación y la Escuela de Posgrado del CATIE, y aprobada por le Comité Consejero del Estudiante como requisito parcial para optar por el grado de:

***Magister Scientiae* en Manejo y Conservación de Bosques Tropicales y Biodiversidad**

FIRMANTES:

Carlos Navarro Ph.D.
Consejero Principal

Elizabeth S. Johnson Ph.D.
Miembro Comité Consejero

Fernando Casanoves Ph.D.
Miembro Comité Consejero

Diego Delgado M.Sc.
Miembro Comité Consejero

Stephen Cavers Ph.D.
Miembro Comité Consejero

Glenn Galloway Ph.D.
Director Programa de Educación y Decano de la Escuela de POSGRADO

Luis Gustavo Hernández Sánchez
Candidato

DEDICATION

I dedicate this document to God, which makes everything for a reason, and makes meaningful all events in my life. Thanks God for given me a second chance to finish all what I once had started.

Dedico este documento a Dios, que hace todo por un motivo y hace que todos los eventos en mi vida tengan sentido. Gracias Dios por darme una segunda oportunidad para terminar lo que una vez inicie.

ACKNOWLEDGMENTS

I will like to thank God for giving me the chance to finish this important phase of my life. Thanks for giving me the knowledge and health required! Also, thanks for putting in my way the right people and organizations, who helped and advised me through all this process.

Quiero agradecerle a toda mi familia por el apoyo incondicional a lo largo de este proceso, en especial a Kathy, quien me apoyó de novios y ahora de esposos. Resultó largo, pero al final se pudo!

Gracias a mis padres por ser siempre un ejemplo en mi vida y darme la motivación para la búsqueda de la superación y el estudio continuo. Ana, Carlos y Mario, se les agradece el apoyo de hermanos que siempre me han dado, por estar ahí cuando se les necesita, y por ser siempre unos hermanos muy unidos, poco expresivos, pero si muy amorosos.

Un especial agradecimiento a mi profesor, jefe, compañero y amigo Carlos Navarro, por todo el apoyo brindado a lo largo no solo de la maestría, sino también de los cinco años que laboré en CATIE como asistente de investigación de varios de sus proyectos. Gracias por darme la oportunidad de adentrarme en el campo de la diversidad genética, la domesticación de especies, la difícil *Hypsipyla grandella*, las colectas de frutos y los ensayos de procedencia y progenies, entre otros. Es mucho lo que me llevo, esperando que también haya podido contribuir un poco.

Special thanks to Stephen Cavers who wisely guide me through all the process. from my first lab training with microsatellites in 2004 at the CEH, Edinburgh, up to the culmination of this document. Thanks for all your help and support, hoping that this is only one of many other projects that we will make in our professional lives. Taking advantage of the English paragraph, kind regard to the staff of CEH who also gave me their support and friendship, thanks Katy, Bob, Julia, Jane, Rob, and so many others.

I really appreciate the support from my advising committee. Liz Johnson and Fernando Casanoves thanks a lot for giving me great advice and friendly unconditional support.

También quisiera darle las gracias a mis compañeros de trabajo de CATIE, quienes a través de mis años laborados en esta institución me enseñaron muchísimo, además de brindarme su invaluable amistad, gracias Leo Coto y Manuel Sojo, Chente, Carlos Navarro, Jonathan, Jeoffrey, Caro, Gera, y en general todo el Departamento de Bosques. También

quisiera mencionar muchos otros compañeros de CATIE que hicieron mi estancia una experiencia única y en donde materialicé hermosas amistades, ahí va … Tony (Gym), Birra, Fanal, Macho, Diego y Hugo (par de morados), Jonathan Cornelius (un investigador brillante, con grandes cualidades como persona, excelente jefe y amigo), par de Gustavos (comedor y Gym), Orocu y Llua (biblio), Chang, Chelsia, todo el personal de posgrado, Luis, Leo y Tomás (tres excelentes conserjes y tres grandes personas), Marco, Antonio, Diermaid, Paulo y Gerardo (los mejores compañeros de casas turrialbeñas 2000-2005), y muchos otros que se me escapan o ya no están en CATIE.

Agradecer también a mis compañeros de maestría, en especial a José González y Víctor Meza, quienes no solo me han brindado su invaluable amistad, sino que también brindaron compañía a lo largo de todo este proceso. No pueden faltar Milton, Paco, René, Basil, Edwin, Efra y Mau, un abrazo gallos!

Este trabajo no hubiera podido llevarse a cabo sin la ayuda de las fuentes de financiamiento. Un especial agradecimiento al Ministerio de Ciencia y Tecnología de Costa Rica (MICIT), así como a la Consejo Nacional de Investigación, Ciencia y Tecnología (CONICIT) por el financiamiento brindado para elaborar el presente trabajo de tesis, así como por la ayuda financiera durante todo el proceso de la maestría. Este trabajo de tesis fue parte del proyecto SEEDSOURCE, el cual a su vez es financiado por le EC dentro del FP6 INCO (contrato 003708).

BIOGRAPHY

Gustavo is a Forestry Engineer from the Instituto Tecnológico de Costa Rica (ITCR), and had worked at CATIE with genetic diversity and tree domestication research projects since he graduated as a Bachelor in 2000. His work has been primary related with forest management, ecology and genetics, working in several tropical countries. His current work is related with ecological and conservation genetics of tropical forests biodiversity and possible impacts of forest management and climate change on biodiversity. Currently, he works for the INISEFOR (Instituto de Investigación y Servicios Forestales) from the National University with the group of Integrated Forest Management. He is also the Terrestrial Research Coordinator for Proyecto de Conservación de Aguas y Tierras (ProCAT International), a nonprofit organization working in Colombia and Costa Rica. When not in the field collecting genetic samples or measuring permanent plots, Tavo can be found fighting for Karate, Thai boxing and other martial arts championships defending his black belt.

www.procat-conservation.org

CONTENT

ABSTRACT

Genetic variability is an important condition for species to successfully face present challenges and survive while adapting and reproducing to meet future environmental changes. Another crucial aspect of conservation genetics for Neotropical species in fragmented landscapes is their mating systems strategies such as outcrossing and inbreeding, and conservation genetics then deals with how to maintain these systems. *Cedrela odorata* also known as Spanish cedar is an important tropical timber tree species which distributes widely across the continent and it's considered one of the main species for timber production. The objectives of the present study were: *i)* to assess the genetic diversity of *C. odorata* in primary forests and human dominated landscapes in Mesoamerica; *ii)* to determine and compare outcrossing rates and levels of inbreeding for the species among different levels of human dominated landscapes; and *iii)* to evaluate the impact of fragmentation and mother tree isolation on the genetic diversity and outcrossing parameters. Microsatellite markers for the species were developed to assess the genetic diversity and mating systems. Leave samples were collected from provenance and progeny trials established at CATIE, Turrialba. High gene diversity estimates were found for *C. odorata* progeny arrays, although no significant differences were found between the northern and central lineages, neither the progeny arrays of isolated mother trees nor the ones in groups or clusters. Average observed heterozigosity was 0.78 and 20.56 was the average number of alleles overall the samples. High multi-locus outcrossing rates (t_m) were obtained for both isolation levels, 1.00 ± 0.08 and 1.06 ± 0.06 for isolation level 1 and 3, respectively, suggesting absence of self-fertilization. Isolation level analysis revealed significant differences for correlation of paternity and single locus inbreeding coefficient of maternal parents, which was higher for isolated families. An average of 5.0 mature trees donated pollen to isolated mother trees, while 7.7 pollen donors were found for mother trees within a ratio of 100 meters. *C. odorata* progenies from isolated trees have higher levels of bi-parental inbreeding, fewer pollen donors, potentially leading to a reduction of mechanisms for selective abortion among different pollen donors, and leading to a decrease in vigor of progenies. This information is significant for decision makers of the management and conservation of the species, since fertility of individual trees will depend on

the degree of isolation, human dominated habitats and availability of pollinators that may flight long distances between trees. Habitat degradation caused by selective logging has reduced the species density at natural forests, leading to common scenarios with small fragmented forests with few remnant *C. odorata* trees. Importance of isolated trees as stepping stones connectors between fragmented forests must be taken into account for future harvesting permits, where landscape management might be a priority for the species.

Key words: Spanish cedar, microsatellites, SSRs, heterozigosity, outcrossing rates, inbreeding, isolation levels, isolated trees.

RESUMEN

La variabilidad genética es una condición importante para las especies, ya que les permite enfrentar los retos presentes y sobrevivir, además les da capacidad de adaptación a cambios ambientales futuros. Los sistemas de cruzamiento son también un aspecto crucial para la genética de la conservación en especial para especies en paisajes fragmentados, por lo que esta ciencia debe tratar de mantener estos procesos a largo plazo. *Cedrela odorata* también conocida como cedro amargo es una especie tropical maderable de gran importancia la cual tiene un amplio rango de distribución en el continente, y es considerada una de las especies principales para producción de madera. Los objetivos del presente estudio fueron: i) evaluar la diversidad genética de *C. odorata* en bosques primarios y paisajes intervenidos en Mesoamérica; ii) determinar y comparar tasas de exocruza y niveles de endogamia para la especie entre diferentes niveles de aislamiento en paisajes intervenidos; iii) evaluar el impacto de la fragmentación y aislamiento de árboles madre sobre la diversidad genética y parámetros de exocruza. Se desarrollaron marcadores de microsatélites para evaluar la diversidad genética y los sistemas de cruzamiento. Se colectaron muestras de hojas de diferentes ensayos de proveniencia y progenie establecidos en CATIE, Turrialba. Se estimó una alta diversidad genética para los arreglos de progenies de la especie, sin embargo no se encontraron diferencias significativas entre los linajes del norte y centro, así como tampoco se encontraron diferencias significativas para arboles madre aislados y aquellos organizados en grupos o *clusters*. La heterocigocidad promedio observada fue de 0,78 y el promedio de alelos en todas las muestras fue de 20,56. Se encontraron altas tasas de exocruza para los multilocus (t_m) entre ambos niveles de aislamiento, 1,00 ± 0,08 y 1,06 ± 0,06 para niveles de aislamiento de 1 y 3 respectivamente, lo cual sugiere ausencia de autofertilización. El análisis de aislamiento reveló diferencias significativas en la correlación de paternidad y el coeficiente de endogamia de *single locus* (t_s) de progenitores maternos, el cual a su vez fue mayor para familias aisladas. Un promedio de 5.0 arboles maduros donaron polen a árboles madres aislados, mientras que 7.7 donadores fueron determinados para arboles madre dentro de un radio de 100 m. Las progenies de *C. odorata* de árboles aislados tiene mayores niveles de endogamia biparental y pocos donadores de polen, conduciendo potencialmente a una reducción de

mecanismos para abortos selectivos dentro de los diferentes donadores de polen, y a la vez a una disminución del vigor de las progenies. Esta información resulta importante para tomadores de decisiones sobre el manejo y conservación de esta especie, ya que la fertilidad de árboles individuales va a depender de los niveles de aislamiento, hábitats intervenidos y disponibilidad de polinizadores que podrían volar grandes distancias entre árboles. La degradación de hábitat causada por la tala selectiva ha reducido la densidad de la especie en bosques naturales, generando que los escenarios comunes sean bosques pequeños fragmentados con pocos individuos remanentes de la especie. La importancia de los árboles aislados como "*stepping stones*" para la conectividad entre bosques fragmentados deben ser tomados en cuenta para futuros permisos de aprovechamiento, donde el manejo del paisaje debe ser una prioridad para las especies.

Palabras clave: Cedro amargo, microsatélites, SSRs, heterocigosidad, estimadores de exocruza, endogamia, niveles de aislamiento, árboles aislados.

TABLES

FIGURES

ABBREVIATIONS

AFLP: Amplified fragment length polymorphism

bp: base pairs

F: single locus inbreeding coefficient of the maternal parents

H_o: observed heterozygosity

H_e: expected heterozygosity

HWE: Hardy-Weinberg equilibrium

Masl: meters above sea level

N_a: number of alleles

N_e: number of effective alleles

PCR: Polymerase Chain Reaction

P_r : exclusion probability

RAPD: Randomly amplified polymorphic DNA

RLFP: Restriction fragment length polymorphisms

r_p: correlation of paternity

r_s: correlation of selfing among families

SSRs: simple sequence repeats; microsatellites

t: outcrossing rate

s: selfing rate

t_m*:* multilocus outcrossing rate

t_s*:* singlelocus outcrossing rate

1 INTRODUCTION

Loss of biological diversity in humid tropical forests is mainly caused by the impact of different human activities (Kattan 2002). The situation in Mesoamerican forests is the same as that for the rest of Latin America, where livestock, agriculture, management of forest resources and demographic growth have played a very important role in the processes of deforestation. Fragmented forests and isolated trees are now so abundant and frequent that they practically dominate Latin American landscapes, with the exception of Tikal National Park in Guatemala and the Amazon Forest in South America, which are the only remaining large areas of continuous forests.

Kattan (2002) mentions that one of the main consequences of deforestation is the creation of fragmented landscapes, where some remnants of the original forest with variable forms and sizes, are immersed in a diverse number of human dominated habitats. Large scale fragmentation of ecosystems could affect the physical environment and climate at a local and regional scale. Besides physical effects, fragmentation of habitats could also result in loss of genetic diversity both at the species and population levels. In the worst scenario, species extinction can occur. In the long term, Namkoong *et al.* (2002) affirm that genetic erosion induces species extinction and ecosystem loss, and also reduces the possibility of using genetic variability for economic benefits and ecological restoration.

Species need genetic variability to successfully face present challenges and survive while adapting and reproducing to meet future environmental changes. Genetic diversity is necessary for maintenance of evolutionary potential and adaptability of local populations and species in general. Conservation of genetic diversity is an essential element for the maintenance of all levels of biodiversity, which are valued because of their existence and utility. However, genetic variability is difficult to measure directly, and its scale may vary depending on the size and connectivity between populations along with their history and landscape context. Furthermore, genetic variability in ecosystem dynamics makes levels of genetic diversity difficult to predict (Namkoong *et al*. 2002).

Genetic consequences of population fragmentation are strongly influenced by gene flow between the forest fragments. In general, restrictions in gene flow produce higher levels of inbreeding and genetic diversity loss between fragments. Furthermore, gene flow might be affected by the number of fragments, population structure, distance between fragments and the

dispersion characteristics of the species. In the long term, genetic differentiation and major risk of extinctions may occur (Young and Boyle 2000, Frankham *et al.* 2003).

Reductions in density of mature forests will directly decrease a species' population size, and potentially affect gene dispersal, and its mating system. Thus, genetic drift is a threat to the maintenance of genetic variation, particularly for small and isolated populations if many trees are removed during harvesting of mature forests. Special attention should be given to tropical forests, since the consequences of silvicultural management practices on genetic structures of tree populations apparently are most severe and detrimental in these forests (Finkeldey and Ziehe 2004).

Cloutier *et al.* (2007) categorize the directly measurable population genetic effects of selective logging into three: *i)* selective harvesting may cause an immediate loss of genetic diversity as a consequence of removal of adult trees; or a loss in the seed progeny generation as a consequence of reduced availability of pollen donors, *ii)* a reduction in the number of individuals available to donate pollen, *iii)* harvesting could modify the genetic structure of reproductive trees, possibly leading to an increased proportion of mating among unrelated individuals in the population.

Boshier (2000) highlights the importance of reducing the possibility of inbreeding and maintaining diversity in naturally outcrossing tree species for their long term viability. However, maintenance of breeding system flexibility will be a priority for species that naturally combine outcrossing and inbreeding.

Genetic-marker based estimates of mating systems in tropical trees have shown that they are predominantly outcrossing with extensive gene flow (Hall *et al.* 1994, Boshier *et al.* 1995, Hall *et al.* 1996, James *et al.* 1998, Rocha and Aguilar 2001, Muluvi *et al.* 2004). However, temporal variation in flowering within and among breeding seasons may restrict the levels of outcrossing and gene flow (Murawski and Hamrick 1991). Such variability may further aggravate the effects of deforestation, forest fragmentation (Hall *et al.* 1996), and tree isolation.

The mating system in plant populations is influenced by a variety of genetic and environmental factors (Clegg 1980) over a hierarchy of levels (*e.g.* seed, fruit inflorescence, and tree) (Boshier 2000). Furthermore, in many cases, the observed mating patterns are likely to result from diverse interactions between several factors, rather than one particular factor (Boshier 2000).

Plant mating systems are also affected by changes in the landscape, such as selective logging, deforestation and habitat destruction, all of which can modify the population density, demographic structure, phenology and abundance, diversity and behavior of the pollinator community (Lowe *et al.* 2004, Ward *et al.* 2005).

Most outcrossing angiosperms have bisexual flowers, a condition from which self-pollination can evolve directly through the modification of self-incompatibility or other floral traits that prevent self-pollination (Schoen *et al.* 1997 cited by Muluvi *et al.* 2004). As well as floral morphology, mating systems may be sensitive to spatial and temporal structure within tree populations (Boshier 2000), plant density, population size (Clegg 1980, Boshier 2000), type and abundance of pollination vector (Boshier 2000), flower color, size of floral displays, and anther-stigma separation (Muluvi *et al.* 2004).

Accurate fine-scale measures of the mating system can be obtained using molecular markers. Most recent studies have estimated outcrossing rates as population averages (Ritland 2002), while in this study estimates of outcrossing rates will be focused as averages for progeny arrays derived from mother trees under different levels of human dominated landscapes: continuous forests and isolated trees.

This study focused on the determination of estimates of outcrossing and other population genetic parameters for the tropical tree species *Cedrela odorata* in primary forests and different man-made landscapes. Also known as Spanish cedar, *Cedrela odorata* is one of the most commercially important species of the Meliaceae family in the neotropics. It is a semi-deciduous tree, native from northern México to South America and also found in the Antilles (Holdridge *et al.* 1997). Broadly valued for its high quality wood, *C. odorata* has been used for construction, furnishings and handcrafts, among other things. Wild populations have been severely reduced because of selective logging (Zamora 2000) to the point that the species is now threatened at a provenance level (Patiño 1997).

1.1 Objectives of the study

1.1.1 General objective

The goal of this project is to evaluate the genetic variability, and mating systems of *C. odorata* trees growing in different levels of isolation: i) isolated mother trees (no trees of the same species closer than 500 m) ii) and mother trees in clusters or associated with more than two trees within a radius of less than 100 m.

1.1.2 Specific objectives

To accomplish the general objective, the project had these three specific objectives:

- To assess the genetic diversity of *C. odorata* in primary forests and different man-made landscapes in México, Guatemala, Honduras, and Costa Rica.
- To determine and compare outcrossing rates and levels of inbreeding for the species among different levels of human dominated landscapes.
- To evaluate the impact of fragmentation and mother tree isolation on the genetic diversity and outcrossing parameters of the species.

1.2 Hypotheses

H1: The molecular marker technique, simple sequence repeats (SSRs) allows the determination of genetic diversity, outcrossing rates and levels of inbreeding for *C. odorata*.

H2: There are significant differences in the genetic variability estimates between the populations of *C. odorata* at different levels of isolation.

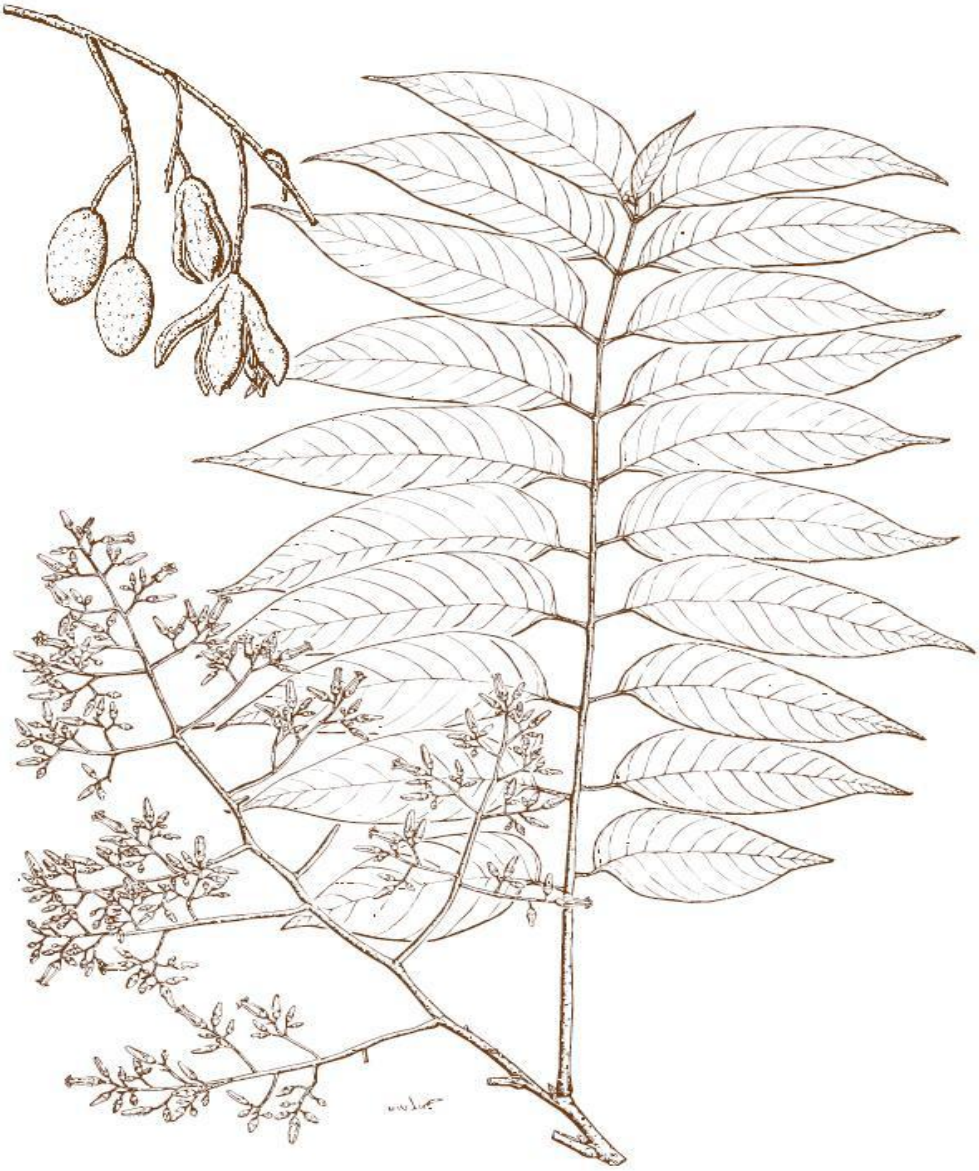
H3: There are significant differences in the outcrossing rates between the populations of *C. odorata* at different levels of isolation.

2 LITERATURE REVIEW

2.1 Description of *Cedrela odorata* and its habitat

Mesoamerica is home to more than 5000 endemic plants, around 1.7% of the world's plant species (Myers *et al.* 2000) which, as it is also experiencing an accelerated rate of habitat loss, make it one of the world's biodiversity hotspots. *Cedrela odorata* is widely distributed throughout Mesoamerica, and is found naturally in these Life Zones: tropical moist forest, subtropical moist forest and tropical dry forest. Its natural distribution range goes from northern Mexico to Bolivia and northern Argentina, and the Caribbean (Cordero and Boshier 2003). Although the species is widespread geographically, it is not common throughout moist tropical forests. In Mesoamerica, it continues to be severely impacted by exploitation (Cintron 1990, cited in Navarro 2002), particularly because of the substantial loss of tropical dry forest. Tropical dry forest is a highly threatened habitat in Mesoamerica, and has been lost through extensive forest destruction during recent decades, mainly for the establishment of cattle ranching and agriculture activities.

Cedrela odorata L., Meliaceae (Spanish cedar) is a large, semi-deciduous, tropical tree, up to 40 meters in height and 2 meters in diameter at breast height (dbh), with its crown, ample and sparse, emerging above the main forest canopy. Leaves are paripinnate, with 5-11 pairs of leaflets, lanceolate to ovate, measuring 5-16 cm long (Figure 1). It is a functionally monoecious species, with small, pale, unisexual green flowers with a garlic odor, grouped together in racemes of 30-50 cm at the end of the branches (Holdridge *et al.* 1997, Zamora 2000, Cordero and Boshier 2003). Flowers have small, white petals and are probably pollinated by small moths (Bawa *et al*. 1995). In Costa Rica, flowering occurs annually from November to December, and fruits are mature between January and March. Meanwhile in Mesoamerica, phenology varies through the region, in general flowering occurs between November and February, and fruits are mature between January and April. Fruits are dehiscent woody capsules with five divisions, each containing 30-40 seeds, green at the beginning and dark brown at maturity. The seed is plane, oval–shaped, around half centimeter long, brown color and wing-shaped at the end (Asociación Costarricense para el Estudio de Especies Forestales Nativas 1994, Cordero and Boshier 2003). The species is light-demanding and exhibits rapid seedling growth in areas of high light such as roadsides or gaps in the forest.



Source: Cordero and Boshier (2003)

Figure 1. Morphology of C. odorata Top left: Fruit capsule, closed and opened. Center: compound paripinnate leaf with florescence.

Tree plantations are an alternative for sustainable production of the species. It has great potential in reforestation because of characteristics like fast growth, easy management in the nursery, adaptability to different soils, climatic conditions and high growth rate in agroforestry systems (Navarro 2002).

The attack of the shoot-borer *Hypsipyla grandella* is a serious problem in nurseries and plantations of Spanish cedar. The problem increases during the rainy season (Cordero and

Boshier 2003), mainly because trees are in a growing period and new fresher shoots are available. This pest reduces growth, increases costs of maintenance and weeding, and induces bifurcations with consequent loss of apical dominance of the tree and value of the timber (Hilje and Cornelius 2001).

The most well known product from *C. odorata* is a very attractive and valuable high quality wood, which has several uses: construction, interior decorations, furniture, and boat construction, among others (Cordero and Boshier 2003). Spanish cedar wood is reddish, durable, with a density of 0.37 – 0.45 gr/cm3, easy to work, and without problems during the drying processes. It possesses an excellent dimensional stability and durability; it is categorized as slightly light to slightly heavy (Asociación Costarricense para el Estudio de Especies Forestales Nativas 1994).

The wood price is one of the highest in the wood market in all countries from Central America. The prices fluctuate considerably through the different stages of the commercialization chain. The highest price is reached when the wood is sawn or has an aggregated value, which in the first case has a price of 400-700 US$ (Cordero and Boshier 2003).

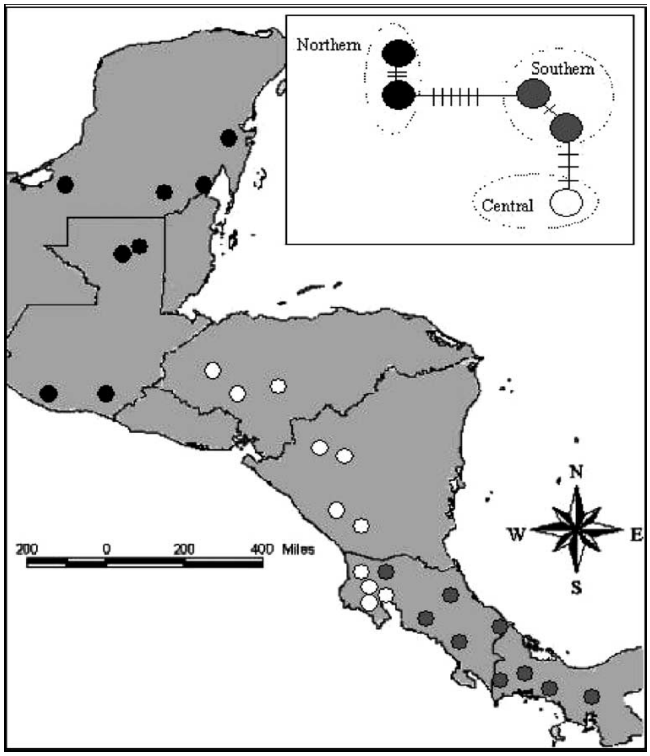
The genomic size of *C. odorata* is around 90 Mb, estimated by flow cytometry (Wilson *et al.* 2001). Chromosome number can be either 50 or 56 (2n) for different chromosome races (Styles and Koshla 1976, cited in Patiño 1997).

2.2 Phylogeography, molecular and quantitative variability

Several studies assessing different aspects of the genetic structure across the wide-ranging populations of *C. odorata* facilitated important findings for the development of the present research. Information from chloroplast, total genome, and quantitative characters helped to establish the sampling strategy of this study, which at the end will provide important additional information about the outcrossing rates of the species at different levels of human disturbance and lineages.

Cavers *et al*. (2003b) studied the chloroplast DNA phylogeography of *C. odorata* in Mesoamerica with samples from 29 populations in México, Guatemala, Honduras, Nicaragua, Costa Rica, and Panama. They characterized five haplotypes which were phylogenetically grouped into three lineages: Northern, Central, and Southern (Figure 2). Northern and Central

lineages were the most genetically distant from each other (nine mutations), followed by Northern and Southern with six mutations, and Central and Southern with three mutations. This pattern probably reflects ancient colonization processes that have occurred over thousands to millions of years, where Central lineages are genetically more related to Southern populations and more differentiated from Northern populations.

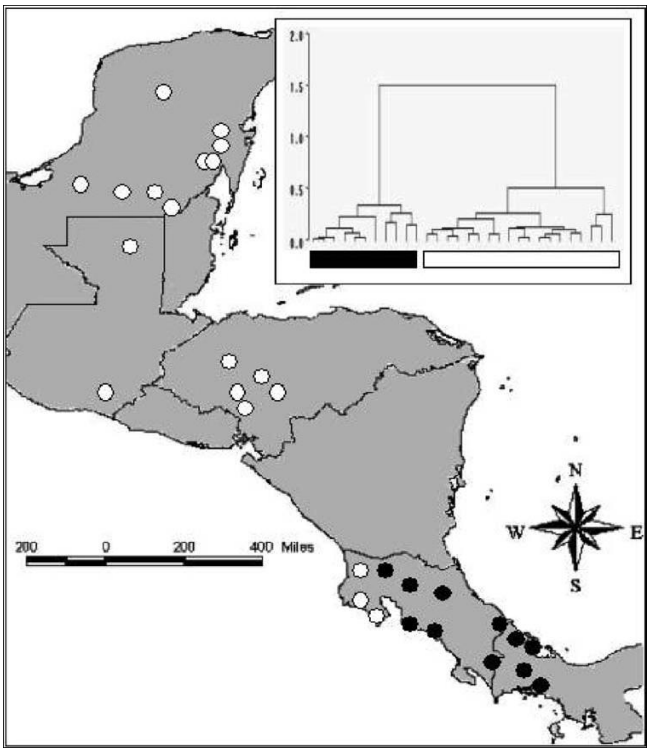


Source: Cavers *et al*. 2003a

Figure 2. Map of the populations and distributions of haplotypes of Cedrela odorata sampled in Mesoamerica. At the top: minimum spanning tree for the five haplotypes identified.

Populations of Northern and Central lineages are located in regions with longer periods of dry season, five to seven dry months (Table 1) and showed quantitative variation in provenance trials when they were compared to the Southern lineage (Figure 3). The map shows geographic origin of provenances, circle color (black/white) indicates grouping of provenance in cluster analysis. The cluster diagram shows the divisions of provenances into two major clusters-populations from west Panamá and southeast Costa Rica (Southern lineage), and those from México, Guatemala, Honduras and northwest Costa Rica (Northern and Central lineages). A standardized measure of among-population differentiation, Q_{ST}, was

estimated using quantitative genetic characters, resulting in differentiation of Northern and Central populations from Southern populations (Navarro 2002).



Source: Cavers *et al*. 2004

Figure 3. Distribution of quantitative variation based on 17 trials measured on seedlings grown in a common garden.

Similar findings emerged from a comparison between morphological variation and site of population origin in mesic and dry habitats in Costa Rica (Navarro *et al.* 2002). Genetic diversity within and between populations of this species in Costa Rica was identified with morphological divergence between northern and southern populations.

RAPD and AFLP analyses were done for Costa Rica and also found significant differentiation between populations (Gillies *et al*. 1997, Cavers *et al.* 2003a). Estimates of variance components from RAPDs analysis indicated highly significant differentiation between populations from North Pacific and Atlantic/South Pacific regions (35.3%), while most of the total variance was attributed to variation within populations (65.1%). On the other hand, AFLP analysis identified differentiation between populations from the northwest region and those from the east/southwest of Costa Rica as the principal source of variation (83.5%), although within these groupings most variation was also distributed within populations. An

additional subdivision between eastern and southwestern populations, indicating restricted gene flow across the Costa Rica's central mountain ranges (52.6%) was found. Although analysis was not carried out for populations from all Mesoamerica, a part-analysis for AFLP variation of a population in Mexico showed similarities with northwest populations of Costa Rica.

*Table 1. Passport data of Cedrela odorata populations established in 1999 at CATIE's farm in association with coffee (coordinates with associat*ed climatic vari*ables)*

Country	Location	Altitude Masl	Rain Mm	Rain Start	Rain End	NDM	Latitude (°N)	Longitude (°W)	Families
México	Xpujil	150	1094.0	Jun	Oct	7	18.54	90.14	11, 112, 115
México	Reforma Bacalar	15	1094.0	Jun	Oct	7	18.85	88.67	134, 139,187, 192, 194
México	Tres Garantías	300	1094.0	Jun	Oct	7	18.12	89.14	144, 146, 147
México	Tulum-FCP	30	1094.0	Jun	Oct	7	19.35	88.01	168, 170, 171, 175 - 177, 180, 182, 185
Guatemala	Los Esclavos	737	2834.0	May	Oct	6	14.25	90.28	32, 33, 35, 36, 38, 39, 312, 314, 317, 319
Guatemala	Tikal	250	1366.7	May	Nov	5	17.22	89.61	341, 343 - 346, 349, 351 – 355
Honduras	La Paz	726	891.0	May	Oct	6	14.15	87.61	44 – 46, 48, 410, 412 – 414
Honduras	Taulabe	633	912.0	May	Oct	6	14.83	88.10	422, 423, 426 – 430
Honduras	Cedros	555	912.0	May	Oct	6	14.66	87.30	446 - 449, 451, 454, 456, 457, 459
Honduras	Meambar	595	912.0	May	Oct	6	14.83	88.10	467, 469, 470, 472
Costa Rica	Cañas	100	2273.6	May	Nov	5	10.32	85.04	662, 663, 699, 6270
Costa Rica	Upala	250	2558.3	May	Jan	3	10.86	85.02	6177, 6189
Costa Rica	San Carlos	90	4574.1	April	Feb	1	10.47	84.58	683
Costa Rica	Cóbano	20	2896.8	May	Nov	5	9.65	85.12	6110, 6112, 6114
Costa Rica	Talamanca	20	2812.0	April	Nov	4	9.65	82.79	6121, 6123, 6125
Costa Rica	Guápiles	250	4465.8	May	April	0	10.19	83.79	6141, 6145
Costa Rica	Hojancha	350	2232.3	May	Nov	5	10.07	85.40	6105, 6108, 6166, 6176, 6101, 6103
Costa Rica	Pacífico Sur	40	4817.7	May	April	0	8.61	82.88	6207, 6213
Costa Rica	Pérez Zeledón	600	2934.5	April	Nov	4	9.34	83.65	6232, 6240, 6274
Panamá	Almirante	10	3319.0	April	Dec	3	9.39	82.56	71-715
Panamá	Guanaca	150	2620.0	April	Nov	4	8.59	82.23	745, 747, 752, 766, 768

NDM = number of dry months

Grey rows represent populations with high number of dry months

Source: Navarro *et al.* (2004)

2.3 Genetic diversity

Population genetics can be understood as the study of naturally occurring genetic differences between organisms; these differences are known as genetic variation. Genetic variation can occur at three hierarchical levels: *i)* within populations, *ii)* between populations of the same species, and *iii)* between different species (Conner and Hartl 2004). Similarly, Lowe *et al*. (2004) separate genetic variation in three components: i) genetic diversity (the amount of genetic variation); ii) genetic differentiation (the distribution of genetic variation among populations); and iii) genetic distance (the amount of genetic variation between pairs of populations).

Understood as the heritable variation of genes within biological organisms, genetic diversity may be measured at the individual, population, and species level (Lowe *et al.* 2004). At the individual level, genetic diversity is defined by the relative allele frequencies, also known as allelic richness (*A*), which are the different expressions of the same gene (locus), that are inherited from the father and the mother. The allelic diversity is the mean number of alleles per locus, and may be calculated per population or per species. Genetic diversity is also described using polymorphism, and average heterozygosity. The proportion of polymorphic loci (*P*) is calculated dividing the number of polymorphic loci between the total number of loci sampled. Whereas the average heterozygosity (*H*) is the sum of the proportions of heterozygotes at all loci, divided between the total number of loci sampled. In the case of populations, genetic diversity is defined by the type of alleles present, and the relative frequencies, and is expressed at a population level or between different populations (Nason 2002, Frankham *et al*. 2004).

Genetic diversity preserves a reservoir of response conditions to the changing environment, allowing adaptation and survival. Species maintain genetic variation as a product of survival and reproduction under the challenges of a heterogeneous landscape and past environmental changes. Conservation of genetic diversity is a necessary precondition for future evolution and adaptation of populations and entire species, and is therefore a key element in the maintenance of all other levels of biodiversity (Namkoong *et al*. 2002).

Loss of genetic variation threatens the capacity of species and ecosystems to adapt to changing environments. In general, loss of genetic variation is difficult to measure directly, and its effect on population and ecosystem dynamics is often very difficult to understand, consequently its loss is easily ignored until it is too late to recover or restore (Namkoong *et al.*

2002). More drastically, genetic erosion induces species extinctions and ecosystem loss, discarding the chance of using genetic variation for economic gain and ecological restoration. Commonly, social and economic pressures on the forests prevail over the ecological needs, having a direct influence over the genetic diversity at all its levels. Forest management practices, fire, grazing and unintended side effects of climate change may result in reductions of genetic variation, due to changes in basic evolutionary processes such as selection, drift, migration and mating.

2.4 Mating systems

In plants, the term mating system refers to the degree to which individuals are self fertilizing, outcrossing, clonal, or a combination of this systems. Studies of plant mating systems had been developed frequently, mainly because mating is directly or indirectly related to genetic changes in space and time (Clegg 1980). Research on plant mating attempts to determine which plant mates with which other plant, and the causes and reasons that determine why mating patterns become modified. Most theoretical models of mating-system evolution have focused on the fitness consequences of selfing and outcrossing (Barrett and Harder 1996). The mating system determines the composition of the male and female pools, the extent to which germplasm is exchanged between individuals (outcrossing rates), and the rates of immigration and emigration (Namkoong *et al*. 2002).

Outbreeding, self-fertilization, and clonal reproduction are three different mating systems that plants have for reproduction. Species with different mating systems typically require different genetic management strategies: for this reason it is very important to define the mating system for each species (Frankham *et al.* 2004).

To understand mating systems, it is necessary to study the categorization of the species sexual system. Trees can be classified in three main groups: *i*) dioecious: all individuals in a population are male or female; *ii*) hermaphrodite: individuals have both male and female function, and may have either monoecious (single sex) or hermaphrodite (both sexes) flowers; *iii*) monoecious: hermaphrodite individuals in which anthers and gynoecia occur in different flowers, male and female functions are separated. These mechanisms reduce the proportion of self-pollination, therefore the probability that inbreeding may occur (Boshier 2000).

2.4.1 Functions of mating systems

The mating system is of great importance for plants in determining the genetic structure and evolutionary potential of natural populations, mainly because it establishes the pattern of uniting gametes forming the next generation (Allard 1975 in Liengsiri *et al.* 1998). Gregorius (2002) cites three fundamental functions of mating systems: *i*) generative reproduction; *ii*) selection for participation in generative reproduction; *iii*) combination of genetic information (genes) into genotypes. These functions determine the adaptational capacities of populations which can be realized during the transition from one generation to the next (Gregorius 2002).

2.4.2 Major determinants of plant mating systems

There are several factors that in conjunction may affect mating systems, varying the outcrossing rates of individual trees and/or populations. A short list of categories of factors that affect plant mating systems is presented as follows (Boshier 2000, Gregorius 2002):

i) Spatial relations: Spatial distribution of individuals, consequence of their pollen dispersal characteristics. Genetic structuring within populations is common, especially for taxa with wind-dispersed seed.

ii) Temporal relations: refers to temporally varying spatial distribution patterns and activities or behavior, age of sexual maturity, timing of flowering and time-dependent expressions of reproductively relevant phenotypes of males and females. Flower synchrony between neighbors may increase the predisposition towards inbreeding.

iii) Phenotypic relations: some examples are prezygotic incompatibility or isolation mechanisms, including biochemical or physiological agents, and morphological barriers such as flowering phenology.

iv) Ecological conditions: availability and selectivity of pollinators, physical barriers to pollen dispersal, species composition. Mating patterns may vary depending on the particular characteristics of pollen vector; this variation can take place not only from tree to tree but also within the tree canopy.

v) Genealogical relations: with the exception of self-fertilization, such relations are rarely direct determinants of mating systems.

vi) Density relations: changes in local flowering densities and the spatial patterns of flowering individuals could produce annual variation and differences between the outcrossing rates of several tree species.

2.4.3 Outcrossing rate estimates and assumptions of the mixed mating model

Evolutionary genetics includes three main study areas: adaptation studies, measurement and description of variation, and research of the processes of genetic combination at a population level (Clegg 1980). Combination and transmission of genetic information is mediated through the mating systems, which are an important determinant of the genetic structure and evolutionary potential of natural populations because they establish the pattern of uniting gametes forming the next generation (Allard 1975, Clegg 1980).

The mixed mating model separates progeny into those resulting from inbreeding (both selfing and related mating) and those produced by outcrossing with a random sample of pollen from the total population. Assumptions of the mixed mating model are: i) outcrossing rate is uniform for all maternal parents, and pollen genotypes are distributed uniformly over all eggs; ii) for each maternal parent, progeny genotype classes are independent, identically distributed and multinomial random variables; iii) alleles at different loci segregate independently; iv) genetic markers are not affected by selection or mutation between the time of mating and progeny evaluation (Clegg 1980, Schoen and Clegg 1984).

Departures of the assumption of a uniform pollen genotype distribution may occur due to different reasons, for example, temporal or spatial heterogeneity in the distribution of genotypes in the population, and correlation among the outcrossed pollen types received by individual maternal parents (Schoen and Clegg 1984). In insect pollinated plants like *C. odorata*, correlation could be a factor if the families assayed for marker genotyped are collected from one or a limited number of multi-seeded fruit, specially, if the pollen deposited on a stigma derives from few previously visited trees, resulting in many full sibs progenies. This was not the case for the present study, since seed collection was done from high numbers of fruits (>50). Spatial structuring of genotypes in the population may also contribute to correlation among mating events if it modifies the fertilization probabilities by the various pollen types (Schoen and Clegg 1984).

Outcrossing rates theoretically range from $t = 0$ (complete selfing) to $t = 1$ (outcrossed to a random sample of the populations pollen pool). Values significantly lower than $t = 1$

indicates a degree of inbreeding, which may be caused by selfing or mating between individuals that are related (effective selfing). When selfing occurs, all loci are affected equally, while in mating between relatives, mean single locus value (t_s) significantly lower than the multilocus estimate (t_m) indicates effective, rather than actual selfing (Boshier 2000).

Mating parameters are estimated based on the assumption that each seed crop of a tree is made up of a proportion, *t*, of outcrossed seeds, and a proportion, 1 - *t* = *s*, of self fertilized seeds. It also assumes that each tree receives a sample of pollen with the allele frequencies representative of the population. Another assumption is that gene frequencies and outcrossing rates in the pollen pool are constant from one tree to another in the same population (Rocha and Aguilar 2001).

Any increment in the number of homozygotes in the progeny, in relation to the totally outcrossed progeny, is due to selfing. For this reason *s* is an estimator of the effective selfing rate, which includes true selfing and biparental inbreeding (Ritland 1989).

2.5 Molecular marker technique: Microsatellites (SSR's)

New technological developments have expanded the range of DNA polymorphism assays for genetic mapping, marker assisted plant breeding, genome fingerprinting, and for investigating genetic relatedness. These technologies include restriction fragment length polymorphisms (RLFP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), and simple sequence repeat polymorphisms or microsatellites (SSR). These methods detect polymorphism by assaying subsets of the total amount of DNA sequence variation in a genome (Powell *et al.* 1996).

These markers differ in the type and amount of variability they express, suitability for each particular question, and in the ease and costs of their development and application. Molecular markers display different amounts of variation, and different modes of inheritance: they may be either codominant or dominant (meaning heterozygotes can or cannot be distinguished, respectively), and may or may not be subjected to genetic recombination (Ouborg *et al.* 1999).

A comparison of these four markers done by Powell *et al.* (1996) demonstrated that all four types of marker assays have different properties: microsatellites have the highest expected heterozygosity, while AFLP are characterized by a very high multiplex ratio, RAPDs are

intermediate in heterozygosity and multiplex ratio, while RFLPs have moderate heterozygosity. Depending on the type of research that will be developed, determining the molecular marker used will depend on the level of information gathered from the marker, the ease of genotyping, and the genomic coverage.

Data on levels of polymorphism based on SSRs have dominated mammalian genome research. Microsatellites are likely to have major positive impact on plant genetics since they provide a highly informative, codominant, PCR-based assay which is compatible with the requirements of plant breeding and population genetic analyses.

The methodology to obtain SSR markers briefly includes the following steps. First, build a library of small genomic fragments (300 to 500 *bp*) for the organism of interest, by digestion of total genomic DNA by restriction enzymes. These fragments are selected for the presence of microsatellites (library enrichment), using synthetic probes that are complementary to the more common repetitive elements (*i.e.* poli-GT for mammals where elements CA are more common, or poli-AT for plants where the more common elements are TA). Individual fragments are then copied by cloning and sequenced. For each unique sequence, specific pair of primers is designed, which are carefully selected at each side of the microsatellite (usually using computer programs). Finally, the sequence is amplified by PCR to verify ease of amplification, the clarity of the fingerprint and the level of polymorphism (Ferreira and Grattapaglia 1998).

Microsatellites are recognized for their great utility in population and conservation biology studies. In the case of population studies, by looking at the variation of microsatellites in populations, inferences can be made about population structures and differences, genetic drift, genetic bottlenecks and gene flow. On the other hand for conservation biology, microsatellites can be used to detect sudden changes in population, effects of population fragmentation and interaction of different populations. Microsatellites are useful in identification of new and incipient populations.

2.5.1 Definition of microsatellites

Microsatellites, alternatively known as simple sequence repeats (SSR), short tandem repeats (STR), or variable number tandem repeats, are short repeated sequence motifs generally consisting of repeat units of 1-6 base pairs (*bp*) in length, e.g. $(TG)_n$ or $(AAT)_n$

(Hoelzel 1998). In other words, a microsatellite consist of a specific sequence of DNA base or nucleotides, which contains mono, di, tri, or tetra repeats. For example,

AAAAAAAAAAA would be referred to as $(A)_{11}$

GTGTGTGTGTGT would be referred to as $(GT)_6$

CTGCTGCTGCTG would be referred to as $(CTG)_4$

ACTCACTCACTC would be referred to as $(ACTC)_3$

Originally microsatellites were used as tandem repeat units, although at the present time higher repeat units of 3 or 4 *bp* are commonly used (Cavers 2008, personal communication) as in the current study.

They are highly abundant in eukaryotic genomes, but also occur in prokaryotes at lower frequencies. They seldom include more than about 70 repeat units and are distributed throughout the genome (Schlötterer 1998). Microsatellites are hypervariable and, as they are codominant (both alleles seen) and highly reproducible, they are ideal for population genetic studies, as well as more advanced applications such as genome mapping (Dayanandan *et al*. 1998).

The most frequent repetitive elements in mammals are extensions of dinucleotides CA and TG (Hamada *et al*. 1982, cited by Ferreira and Grattapaglia 1998). In plants, presence of microsatellites was observed in 34 species, where the most common repetitive element was the dinucleotide AT (Ferreira and Grattapaglia 1998). Nowadays molecular markers based on microsatellites have been developed to create genetic maps for some annual crops like soybean, corn, and wheat. There are many studies of microsatellites in trees, including not only boreal, but also tropical species (Chase *et al*. 1996, White and Powell 1997a, Lowe *et al*. 2002, Roth *et al*. 2003).

2.5.2 Microsatellite polymorphisms

Microsatellite sequences change (mutate) over time, observed as variation in the number of repeat units. For example, one allele may have seven repeats of a CT motif, whilst another allele has eight repeats. In a population there may exist many alleles (up to 70 or 80) at a single locus, with each allele having a different length. An individual who is homozygous for a locus will have the same number of repeats on both chromosomes, whereas a heterozygous individual will have a different number of repeats on each chromosome. The regions surrounding the microsatellite locus, called the flanking regions, may still have the

same sequence. This is important because the flanking regions can therefore be used to design PCR primers when amplifying microsatellite loci, and can be conserved across genera or sometimes even families.

Changes in repeat numbers caused by an intra-molecular mutation mechanism are called DNA slippage. The most common mutations are changes of a single repeat unit, which allows microsatellite mutations to be interpreted as a very good approximation of a stepwise mutation process (Schlötterer 1998).

Slippage occurs during DNA replication, when the polymerase either leaves out or adds too many repeat units. The result is that the new strand has a different number of repeats than the parent strand. This might explain small changes in numbers of repeats (adding or subtracting one or just a few repeats). It also explains how microsatellite loci could be generated in the first place; it is likely that sequences including two or three repeats are randomly distributed throughout the genome. Slippage could them amplify these short repeat sequences into many repeats over successive generations. Certainly, the effectiveness of the mismatch repair system would also play an important role in microsatellite mutation rate (Schlötterer 1998).

Another hypothesis that explains how microsatellites mutate is the unequal crossing-over during meiosis, which is thought to explain more drastic changes in numbers of repeats. In this case, one of the chromosomes obtains too many repeats during the crossing-over, while the other chromosome obtains too few repeats.

2.5.3 Detection of microsatellite markers

Regions that contain repetitive simple sequences are amplified individually with PCR technique, using two specific primers (20-30 *bp*), which are complementary to the unique sequences that the microsatellite targets. The amplified segments are highly polymorphic, due to the presence of different number of simple repetitive elements. This way, each "isle" microsatellite, independent from the repetitive element (CA, TG, ATG, etc.), is a highly variable, multiallelic, genetic locus with a great amount of informative content. Each amplified segment of different size (generally of several tens to hundreds of base pairs) represents one allele different of the same locus (Ferreira and Grattapaglia 1998).

Detection of the microsatellite sequences via PCR is done with polyacrylamide or high resolution agarose gels, because significant resolving power is necessary to separate segments

that differ in small number of base pairs, depending in the amount of nucleotides of the repetitive element in the microsatellite. Bands could be visualized directly by staining with ethidium bromide or silver nitrate, via x-rays using marked primers with radioisotopes in the PCR reaction, or by laser-induced fluorescence using a fluorescence sequencer and appropriately tagged primers.

Each microsatellite locus is analysed individually. It is possible to evaluate more than one locus at the time when the alleles of each locus have different sizes and migrate to separate zones of the gel. This method (named multiplex) uses more than one pair of primers simultaneously in the same PCR reaction (Ferreira and Grattapaglia 1998).

2.5.4 Advantages and disadvantages

Microsatellite markers possess the highest polymorphism information content, due to their codominant expression and the multiallelic properties (Ferreira and Grattapaglia 1998). Microsatellites are very frequent and distributed randomly, allowing a complete covering of any eukaryote genome, although processing of large numbers of loci is extremely time-consuming, both in the identification and optimisation of loci and in their screening, differently to ALFP, which covers the whole genome very easily without needing to know information sequence en advance.

The great advantage of microsatellite analysis is the large number of polymorphisms that the method reveals. Furthermore, the ability of the method to differentiate individuals when a combination of loci is examined makes the technique very useful for gene-flow assessments, cultivar identification, and paternity analyses (Gillet 1999). Since the method is DNA-based, this brings advantages such as the facility to use dried leaf or cambium material.

Once microsatellite primers have been identified, screening of material using the technique is fairly inexpensive. Furthermore, cross-species amplification of microsatellites means that identification of suitable microsatellite primers may not be necessary in closely related taxa (Gillet 1999). For example White and Powell (1997b) tested cross-species amplification for 11 microsatellite loci with DNA from 11 Meliaceae species, where successful amplification of microsatellite loci indicates that a high level of sequence conservation exists within the primer regions of these Meliaceae species. However, transfer between species often results in reduced levels of polymorphism.

The major limitation of the microsatellites technique is the great effort necessary for development of the markers. Specialized personal and sophisticated equipment are needed for the technique, although as new molecular strategies for isolation of carrier clones for microsatellites become increasingly available, and as methodologies for sequencing become easier, automatic, and economically accessible, this technique will be used more frequently for a larger number of species.

Gillet (1999) mentions some problems associated to microsatellites into three broad categories, practical, data, and analysis problems (which in our case will be treated as genetic population analysis problems), which are summarized as follows:

i) Practical problems

Screening for microsatellites: Unless useful primers have been designed in previous studies, it is necessary to screen an organism for microsatellites.

Slippage: This can be a significant problem when analysing mono- and di-nucleotide repeats. During the amplification process the thermo-polymerase can "slip", leading to production of differently sized products that differ by approximately 1-5 repeat units from the expected product. Such products are usually less intense than the desired product, so in practice can usually be discounted. However, if the products of a heterozygous individual overlap then it is sometimes difficult to differentiate the true and slippage products (Ciofi *et al*. 1998 in Gillet 1999).

Additional practical problems: Haberl and Tautz (1999) and Chavarriaga *et al*. (1998) highlight a potential problem with microsatellites run on automatic sequencing gels and automatically sized. They found that the "called" product sizes differed from the exact product sizes. They recommended that exact sizes could only be determined by allele sequencing, determining "real size" and then using these as internal standards on a gel. Inaccurate allele identification may also be caused by the tendency of *Taq* polymerase to add an adenosine nucleotide to the 3'-end of the amplified product (Ciofi *et al*. 1998). The addition is determined in a template- and marker-specific manner, which may not be a problem if the extra nucleotide is always or never added.

ii) Data problems

Homology: This is the greatest problem facing the use of SSRs in phylogenetic analysis. Microsatellite analyses assume that co-migrating fragments are homologous, whereas there are few *a priori* reasons to assume this. Furthermore, non-homology can be divided into

that which occurs within the SSR flanking and the SSR repeat regions. Since no large-scale tests of SSR homology have taken place in plants, it is difficult to estimate the percentage of bands in a microsatellite survey that are non-homologous. However, there is the potential for a serious problem, and the inclusion of non-homologous fragments in an analysis is likely to bias the results and break the assumptions of a phylogenetic analysis.

Ujino *et al*. (1998) pointed out another homology problem that occurs when analyzing compound repeats. If a SSR with the sequence 5'-$(CT)_{10}CA(CT)_8$-3' is considered and a second allele is 2 bp longer, without sequencing, it is impossible to tell which repeat has increased in size, *i.e.* 5'-$(CT)_{\underline{11}}CA(CT)_8$-3' or 5'-$(CT)_{10}CA(CT)_{\underline{9}}$-3'. One would expect a greater percentage of fragments to be non-homologous if the repeat being analyzed were compound. Ujino *et al*. (1998) recommend that only simple repeats be used in order to limit errors in genotype identification.

The third and most problematic homology uncertainty is within the repeat unit. That is, whether two fragments that co-migrate are identical by descent or just identical in state. To some extent, this problem depends upon the mutation rate of the repeats. If it is low then the probability that a mutation is unique and similar alleles are identical by descent is high. In contrast if mutation rate is high then the probability increases that two co-migrating alleles are just identical in state and non-homologous. There is no simple answer to this issue; generally it is simply assumed that the stepwise mutation model (SMM) is a good way of describing the evolution of the SSRs, when studying closely related populations.

Null Alleles: Mutations in the binding region of one or both of the microsatellite primers may inhibit annealing that may result in the reduction or loss of the PCR product. Such products are termed null alleles and are comparable to the null alleles identified by allozymes in their effects.

Null alleles may be manifested as fewer heterozygotes than expected in a randomly mating population or by the appearance of "empty" lanes. That is, in a heterozygote of two different microsatellite alleles, if one of these alleles cannot be amplified due to primer annealing difficulties, then the phenotype (on the SSR gel) will appear as a single banded homozygote. Null alleles are also responsible for mismatches between parent-offspring pairs, *i.e.* the offspring do not amplify an allele that is present in the parents. However, good analytical approaches to identifying the presence of null alleles have now been developed and are widely used.

iii) Population genetic problems

The high number of alleles per locus of microsatellite markers may cause some bias in the diversity and differentiation estimates, mainly due to increased heterozygosity estimates (Lowe *et al.* 2004). Differences between expected and observed heterozygosity estimates indicates deviations from Hardy-Weinberg equilibrium genotype frequencies, which are highly informative since allow to detect inbreeding, population fragmentation, migration, am selection in natural populations. Allele frequencies may also be affected by linkage disequilibrium, or null allele presence, it is therefore important that both are identified and accounted for in calculations of F_{ST} and related statistics.

Deviations from Hardy-Weinberg Equilibrium: When genotypic frequencies in a real population differ significantly from HWE, it suggests that one or more assumptions of HWE have been violated and therefore potentially important evolutionary or ecological process caused the deviation (Conner and Hartl 2004). Immigration, selection and non-random mating will all lead to deviations from HWE. For example, inbreeding is one form of non-random mating that reduces the frequency of heterozygotes compared to random mating. Other forms of non-random mating are the assortive and dissortive mating, which lead to increased homozygosity compared to HWE frequencies, and increased heterozygosity, respectively. Allele frequencies also diverge in isolated populations due to chance and selection, for example, fragmented populations with restricted gene flow may show deficiencies of heterozygotes compared to the expected estimates. Similarly genotype frequencies may change under selection, when a genotype is at selective disadvantage it contributes proportionally less offspring to the next generation than other genotypes, known as loss of fitness (Yeh 2000, Frankham *et al.* 2002, Conner and Hartl 2004).

Null allele treatment: As mentioned previously null alleles are the result of mutations that prevent PCR amplification, such as sequence changes within one of the primer binding sites. Genotyping errors are caused when null allele presence is not properly detected. Null alleles could only be detected when they are frequent enough to occur in homozygous forms. Nowadays there are programs available to check microsatellite data for null alleles, Micro-Checker is one of those. It helps to detect when one or more alleles fail to amplify during PCR reaction (null alleles), errors that occur during the interpretation of sequence of microsatellite allele data. Detection of null alleles may help the researcher to initially explain deviations from HWE different from the ones mentioned above.

Linkage disequilibrium: It occurs when two alleles from different genes on the same chromosome are associated in different individuals at a greater frequency than the one expected at random association, where the fate of an allele will be affected by that of neighboring loci. When analyzing microsatellites loci, it is important that they are in different chromosomes or at least very distant (well dispersed across the genome) to insure the absence of linkage disequilibrium between loci.

2.5.5 Gel interpretation

Schlötterer (1998) determined that the most common picture of a PCR-amplification microsatellite allele is not a single band, but a ladder of bands. Usually, the most intense band is observed at the expected size of the allele. The additional bands, called stutter bands, are usually smaller than the original allele and different in length by multiple repeat units. Most likely, stutter bands are the result of in vitro DNA slippage during PCR amplification. The decreasing intensity of stutter bands on the gel is a reflection of the probability distribution of DNA slippage events, with larger size deviations being less likely. Rates of *in vitro* DNA slippage have been shown to decrease with repeat unit length: more stutter bands occur for dinucleotide repeats than for trinucleotide or tetranucleotide repeats.

In addition to stutter bands, many microsatellites show an additional band above the expected allele size. This is the result of the terminal transference activity of Taq DNA polymerase, which adds an A to the PCR product. However, this terminal transference activity is polymerase and PCR primer dependent. It is important to take into account that secondary products of various types can produce more complex patterns than the ones described above.

2.5.6 Specialized software for microsatellite analysis

Freely available population genetics software are downloadable for internet to perform most (if not all) statistical analyses necessary to the ecological interpretation of molecular marker data in plants. A search in the web for "population genetics software" will lead to some very useful pages, *i.e.* http://www.biology.lsu.edu/general/software.html. Some of the programs that can be used for these studies are: GENEPOP, FSTAT, GENETIX, POPGENE, GDA, Alerquin, TFPGA, GenAlex, Micro-Checker, among others (Bacles 2004, personal

communication[1], Cavers 2005, personal communication[2]). Test for genetic diversity (Heterozygosity, allelic diversity), independence of loci and population structure in subdivided populations (F-statistics) could be obtained from these programs.

[1] Bacles, C. 2004. Software for molecular genetic analysis (interview). Edinburgh, UK, Centre for Ecology and Hydrology.

[2] Cavers, S. 2005. Software to analyse genetic diversity (e-mail). Edinburgh, UK, Centre for Ecology and Hydrology.

3 MATERIAL AND METHODS

Present work methodology has been divided into four phases: sampling strategy, field sampling, laboratory analyses, and data analyses. A description of these phases is presented.

3.1 Sampling strategy

Phylogeographic, molecular, and quantitative data obtained in previous studies for *C. odorata* was used to determine the sampling strategy. Due to the wide range of distribution of the species, its high genetic variability for growth rates and morphological characteristics between mesic and xeric provenances, and significant molecular differentiation in Mesoamerica the study only used populations of the Northern and Central lineage of the species, which includes populations from México, Guatemala, Honduras Nicaragua and northern pacific region of Costa Rica.

A total of 12 *C. odorata* populations from Northern and Central lineages were chosen to be assessed with microsatellite markers, in particular six populations from the Northern lineage and six from the Central lineage (Figure 4).

Populations considered in this study are representative of drier Mesoamerican regions (Table 2).

Each population was composed of *C. odorata* families under two levels of isolation:

i) isolated mother trees: no trees of the same species closer than 500 m, also called isolation level 1.

ii) mother trees in clusters: trees in clusters or associated with more than two trees within a radius of less than 100 m, also called isolation level 3.

A third isolation level was discarded, semi-isolated trees also called isolation level 2 (more than two trees within a radius of 100 m and 500 m), since none of the evaluated families was included in that category.

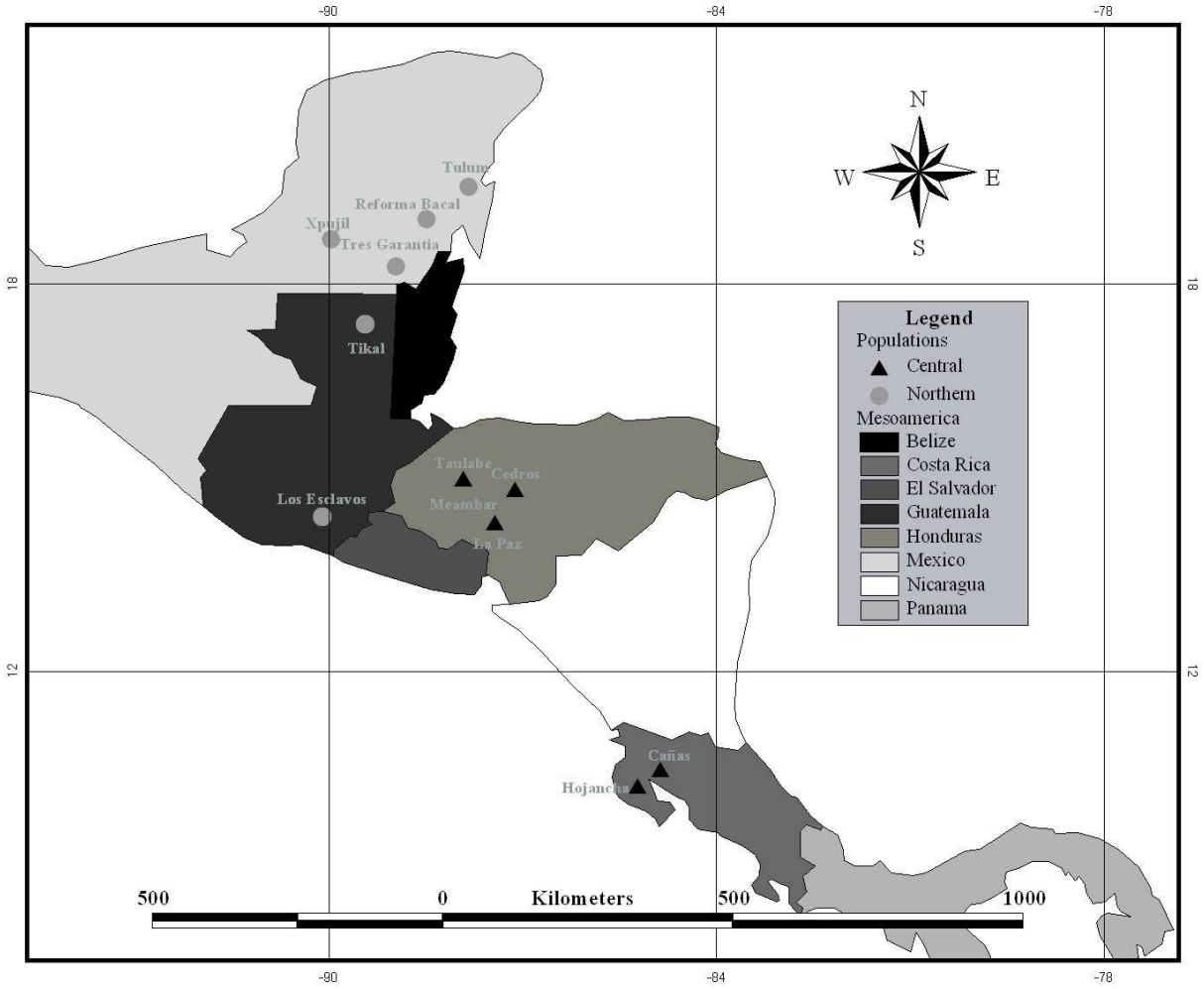


Figure 4. Map of the populations and distributions of haplotypes of Cedrela odorata families sampled in Mesoamerica. Map shows geographic origin of provenances, white circles indicate Northern lineage and black triangles indicate Central lineage.

3.2 Field sampling

Leaf collections were completed at the progeny and provenance trial of *C. odorata* located at CATIE's farm in Turrialba, Costa Rica. This trial was established in October 1999 in combination with coffee plantations by the personnel of the Forest Genetic Diversity project, which was financed by the USDA (United States Department of Agriculture). Seed materials were collected from México, Guatemala, Honduras, Costa Rica and Panamá during 1999 by members of the project, as a conservation strategy for the species in Mesoamerica. DNA genetic studies, conservation germplasm banks and provenance and progeny trials were carried out by the project.

Table 2. Passport data of Cedrela odorata populations and families included in this study

Country	Location	Altitude	Rain			NDM	Latitude	Longitude	Families
		Masl	Mm	Start	End		(°N)	(°W)	
México	Xpujil	150	1094.0	Jun	Oct	7	18.54	90.14	11, 112, 115
México	Reforma Bacalar	15	1094.0	Jun	Oct	7	18.85	88.67	139
México	Tres Garantías	300	1094.0	Jun	Oct	7	18.12	89.14	144, 146, 147
México	Tulum	30	1094.0	Jun	Oct	7	19.35	88.01	168, 170, 176, 177, 182, 185
Guatemala	Los Esclavos	737	2834.0	May	Oct	6	14.25	90.28	32, 33, 35, 36, 38, 39, 312, 314, 317, 319
Guatemala	Tikal	250	1366.7	May	Nov	5	17.22	89.61	341, 343, 344, 345, 346, 349, 352, 353, 354, 355, 356
Honduras	La Paz	726	891.0	May	Oct	6	14.15	87.61	44, 45, 46, 48, 412, 413, 414
Honduras	Taulabe	633	912.0	May	Oct	6	14.83	88.10	422, 423, 426, 427, 428, 429, 430
Honduras	Cedros	555	912.0	May	Oct	6	14.66	87.30	446, 447, 448, 449, 451, 454, 456, 457, 459
Honduras	Meambar	595	912.0	May	Oct	6	14.83	88.10	467, 469, 470, 472
Costa Rica	Cañas	100	2273.6	May	Nov	5	10.32	85.04	662, 663, 699, cr 227
Costa Rica	Hojancha	350	2232.3	May	Nov	5	10.07	85.40	6101, 6103, 6105, 6108, 6176

NDM = number of dry months

The provenance and progeny trial was established in combination with coffee under three coffee conditions: i) just planted (1-month old); ii) with total pruning (10 years old) and iii) coffee in production (5 years old). Two systems of cultivation were used: i) Spanish cedar planted between coffee rows and ii) Spanish cedar within coffee row planting. The experimental design of the trial was made with 115 selected families from a collection (progenies from single trees) of 21 provenances. The families were distributed randomly in 17 blocks, in plots of one or two trees (Navarro 2002).

Leaves from individual trees were collected and dried with silica gel to prevent DNA degradation. Ideally, sampling consisted of: 10 progenies per family; 20 families per population, out of a total of 12 populations; six populations from the Northern lineage and six from the Central lineage. Nevertheless, the final number of progenies and families collected was inferior, due to mortality and leafless trees during the collection at the progeny and provenance trials (Table 3).

Table 3. Number of families and progenies per population, grouped by isolation level

		Families per Population			Progenies per population		
Population	Lineage	Isolation 1	Isolation 3	Total Families	Isolation 1	Isolation 3	Total Progenies
Cañas	Central	4	0	4	27	0	27
Cedros	Central	0	9	9	0	87	87
Hojancha	Central	5	0	5	37	0	37
La Paz	Central	7	0	7	54	0	54
Los Esclavos	Northern	10	0	10	72	0	72
Meambar	Central	1	3	4	5	28	33
Reforma	Northern	1	0	1	10	0	10
Taulabe	Central	7	0	7	65	0	65
Tikal	Northern	0	11	11	0	89	89
Tres Garantías	Northern	3	0	3	27	0	27
Tulum	Northern	5	1	6	41	10	51
Xpujil	Northern	3	0	3	29	0	29
Total		46	24	70	367	214	581

The number of progenies per family ranged between ten and five trees (Table 4). Families with less than five progeny samples were discarded from the study.

Table 4. Number of progenies per family collected at the provenance and progeny trial of Cedrela odorata during February 2006

Progenies	Families	Families
10	23	112, 115, 139, 144, 176, 182, 414, 422, 426, 427, 429, 430, 446, 447,448, 449, 456, 457, 459, 469, 341, 343, 344
9	14	11, 146, 170, 412, 428, 454, 470, 472, 32, 317, 346, 354, cr227, 6108
8	10	147, 177, 45, 48, 451, 35, 312, 314, 352, 355
7	13	168, 185, 44, 46, 33, 36, 345, 349, 662, 6101, 6103, 6105, 6176
6	4	423, 39, 353, 663
5	6	413, 467, 319, 38, 356, 699

Although the seed collection for the establishment of the provenance and progeny trials was carried out in 1999, as background and statistical support of the current research the methodology and data collected is presented.

C. odorata seeds were collected from different types of sites chosen according to geographical and socioeconomic criteria including: topography, geology, soil type, vegetation, land use, human population density, agricultural activity, transportation, and infrastructure facilities. This information was used by the researchers to define the sampling sites and estimate the likely extent of within-species variation, based on their heterogeneity. Optimal time for seed collection was determined with local informants and field visits (Navarro 2002).

Pollination biology and seed dispersal information were taken into account to determine the minimum distance between trees and populations that reduces the possibility of collecting seed from related or inbred trees. The minimum collecting distance between trees in a population was set at 100 m, which is the distance of maximum flight recorded for Spanish cedar seeds (Navarro 2002).

The spatial coordinates of each tree were recorded during the collection using a hand held GPS. Additional information was recorded: population name, collector's name, date of collection, country, province, address, climatic data (precipitation, temperature, number of dry months), slope percentage, site position (valley, slope, top of the slope, etc.) altitude, Holdridge life zone, land uses (primary forest, secondary forest, pasture, agroforestry system), associated species, and characteristics related to the trees (dap, height, form and phenological data).

3.3 Laboratory analyses

Molecular biology work with microsatellites was done at Edinburgh, Scotland, in the laboratory of the Centre for Ecology and Hydrology (CEH), while genetic diversity and mating systems analysis of microsatellite datasets was carried out at CATIE, Costa Rica.

DNA was extracted using DNeasy Plant mini kits (QIAGEN), from silica-gel-dried leaf material. Specific *C. odorata* microsatellites were developed by A. Buonamici and B. Vendramin at the Plant Genetic Institute, Consiglio Nazionale delle Ricerche (CNR), in Florence, Italy. Primary screening for polymorphism and optimization of PCR conditions was carried out using leaf material from four families of 10 half-sib saplings by G. Hernández, K. Walker and S. Cavers at the NERC Centre for Ecology and Hydrology, in Edinburgh, Scotland, described in Hernandez *et al* 2008 (Annex 1). During screening, to maximise the potential for detection of polymorphism, families were chosen from different, widely-separated provenances: *i*) Tulum and Xpujil from Mexico; *ii*) Cedros from Honduras and *iii*) Hojancha from Costa Rica.

All microsatellites were amplified using 25.0 ul PCR reactions consisting of the following: 2.0 ul template DNA, 17.9 ul H_2O, 2.0 ul primers, 2.0 ul 10X buffer, 0.5 ul dNTPs, 0.4 ul BSA, 0.2 ul Taq DNA polymerase (New England Biolabs). Reactions were run on a Hybaid MBS thermocycler to the following protocol: 1 min at 94 °C, then 30 cycles of 1 min

at 94 °C, 1 min at 55 °C, and 1min at 72 °C, and finally 5 min at 72 °C. Optimal PCR reaction conditions for each of the polymorphic pairs were determined by testing a range of annealing temperatures (55.0 up to 70.3 °C) and different template DNA concentrations (1:5, 1:10, 1:20 and 1:1). The sequencer (LI-COR 4800 IR2) is capable of screening samples at two fluorescence wavelengths, so in general, PCR reactions included two microsatellites from each channel, one from IRD 700 and the other from IRD 800, however in the case of multiplexing the total aliquot of primer reaction was double since four microsatellite primers were included in the same PCR reaction (e.g. Ced41-Ced44 + Ced61a-Ced65). The proportion of each primer in the total primer reaction was not always 50:50, in the case of multiplex Ced41-44 it was 30:70, with 1.4 ul of Ced44 and 0.6 ul of Ced41 in the PCR reaction. For combination Ced131 (IRD 700) and Ced95 (IRD 800), 1.4 ul of Ced95 and 0.6 ul of Ced131 were included. In all cases the forward primer was labelled with either IRD 700 or 800 fluorescent labels.

PCR products were separated on 6% polyacrylamide gel (25 cm, MWG Biotech), and visualized using a LI-COR 4800 IR2 automated genotyper. PCR products were run out alongside a microSTEP DNA size standard (Microzone Limited) and fragment sizes of five microsatellites were scored using SAGA™ software, while four microsatellites were scored using Gene Profiler.

Polymorphic loci were initially chosen for the PCR amplification, while separation of fragments with polyacrylamide gels and visualization using LICOR 4800 IR2 automated genotyper was done for the total samples of the study. For the selection of polymorphic loci three characteristics were considered: ease in scoring, high standing allelic diversity, and easiness for multiplexing.

A total of 9 microsatellite loci with high levels of polymorphism were selected from a battery of 26 microsatellite loci (Table 5). A quality control screening step was carried out to validate the data obtained from the microsatellites, resulting in the final selection of nine loci for the analysis: Ced2, Ced18, Ced41, Ced44, Ced54, Ced61a, Ced65, Ced95 and Ced131. The allele size of the 9 selected loci ranged from 70 to 270 bp, with a minimum difference per locus of 20 bp at Ced131a, and a maximum difference of 60 bp at locus Ced61b and Ced95 (Annex 2). The quality control screening step consisted in re-scoring a subset of genotypes, checking scoring (Microchecker v2.2.3, van Oosterhout *et al.* 2004), testing for homozygote excess patterns consistent with null alleles (MICRO-CHECKER), indirect evidence and explicit tests for null alleles (MICRO-CHECKER and GENALEX), checking for evidence of linkage

disequilibrium (Genepop v3.3, Raymond & Rousset), and checking that populations were in Hardy-Weinberg Equilibrium (GenAlex and Genepop).

Locus Ced131a was discarded from the analysis because of linkage disequilibrium association with locus Ced131. The left primer sequence for Ced131 was the same as the right primer sequence of Ced131a, resulting in two neighboring microsatellite sites. Ced131 was selected over Ced131a because Ced131 presented more polymorphism and ease scoring. For the complete set of loci, no linkage disequilibria between loci were observed (P>0.05), assuming only a single individual per family across all populations.

Table 5. Primer sequences and characterization of 9 microsatellite markers isolated from Cedrela odorata

Locus	Repeat motif	Primer Sequences (5' - 3')	Allele Size Range (bp)	T_{ann} (°C)	GenBank Accession no.
Ced2	$(GA)_{20}$	**F:**TTTGCTTTGAGAAACCTTGT* **R:**AACTTTCGAATTGGTTAAGG	130-170	55	EF413962
Ced18	$(GA)_{23}$	**F:**CAAAGACCAAGATTTGATGC* **R:**ACTATGGGTGGCACAACTAC	130-150	55	EF413963
Ced41	$(TC)_{18}$	**F:**TCATTCTTGGATCCTGCTAT* **R:**GTGGGAAAGATTGTGAAGAA	120-160	55.5	EF413964
Ced44	$(TG)_{14}(AG)_{17}$	**F:**ACTCCATTAACTGCCATGAA* **R:**ATTTTCATTCCCTTTTAGCC	180-240	55.5	EF413965
Ced54	$(GA)_{15}(AG)_6G(GA)_5$	**F:**GATCTCACCCACTTGAAAAA* **R:**GCTCATATTTGAGAGGCATT	180-230	55	EF413966
Ced61a	$(TG)_{10}$	**F:**CAATCAAACCAAAAATGGAT* **R:**GCAAATTAACCAGAAAAACG	240-270	55.5	EF413967
Ced65	$(GA)_7(CA)_{14}$	**F:**GAGTGAGAAGAAGAATCGTGATAGC* **R:**GAGGTTCGATCAGGTCTTGG	160-200	55.5	EF413968
Ced95	$(CT)_{17}(AC)_{13}$	**F**:ATTTTCATTCCCTTTTAGCC* **R:**TTATCATCTCCCTCACTCCA	190-250	55	EF413969
Ced131	$(CT)_{16}$	**F:**CTCGTAATAATCCCATTCCA* **R:**GGAGATATTTTTGGGGTTTT	80-120	55	EF413970

* Indicates fluorescently labelled primer. Abbreviations are: base pairs (bp), Annealing temperature (T_{ann}).

3.4 Data analyses

For the data analyses each isolation level was composed of different populations from the Northern and Central lineage. Ideally a minimum of 20 families per population and 10 progenies per family was target for the analysis, although phenology of the species during the collection of leaves, number of families per population established at the trials, and DNA amplification at progeny level limited the final sampling.

Populations were characterized on the level of human disturbance of the site where the mother trees were collected. Isolated remnant trees and those located in fragmented forest

landscapes with no trees of the same species closer than 500 m were grouped as isolation level 1, while families located in conserved primary forests and mother trees in clusters or associated with more than two trees within a radius of less than 100 m were grouped as isolation level 3. No mother trees in the intermediate category were used in the current study, which correspond to semi-isolated trees (isolation level 2), which grouped mother trees where the nearest tree of the same species was located within a radius of more than 100 m but less than 500 m.

Levels of diversity, outcrossing rates and inbreeding parameters will be calculated for the species to determine the effects of landscape context and tree isolation degree (continuous forest, fragmented forests, isolated trees) on mating systems.

3.4.1 Genetic diversity analyses

Standard genetic diversity parameters were estimated for the species using the complete dataset; isolation and lineage levels estimates were also estimated. Numbers of alleles (N_a) were calculated, while numbers of effective alleles (N_e), observed heterozygosity (H_o) and expected heterozygosity (H_e), exclusion probability (P_r) were estimated with Genepop and GenAlex, while null allele rate was estimated with Microchecker v2.2.3 (van Oosterhout et al. 2004).

Proportion of polymorphic loci (P)

Genetic diversity was quantified in terms of proportion of polymorphic loci; where

$$P = \frac{\text{Number of polymorphic loci}}{\text{Total number of loci sampled}}$$

A gene (locus) was considered polymorphic if the frequency of the more common allele was less than 0.95.

It is important to consider that the parameter of polymorphism is based on allele distribution, and not of genotypes, in this way a population may have all the homozygous genotypes for a locus but the locus may still be polymorphic.

Expected heterozygosity

This parameter refers to the expected frequency of heterozygous genotypes at a certain locus (monomorphic and polymorphic), assuming Hardy-Weinberg equilibrium. Nei (1973) was the first to use the term gene diversity, and calculated the expected heterozygosity across the total species (H_T), measured as follows:

$$H_T = 1-\sum_{i=1}^{i=k} p^2$$

where p_i is the frequency of the *i*th of *K* alleles across all populations of the study. Genetic diversity at a population level is calculated for each locus and population as follows:

$$H_{ep} = 1-\sum_{i=1}^{i=k} p^2$$

where *p* is the frequency of the *i*th allele at the *k*th locus in each population and the value is averaged over all populations. Average of H_{ep} values is calculated across all loci among all populations to obtain a mean estimate of the genetic diversity at a population level. For a locus with two alleles at frequencies *p* and *q*, the expected herozygosity is calculated as:

$$H_e = 2pq$$

If allele frequency is constant through all the populations, genetic diversity measures at a species and population level must be the same. However, populations vary in their allele frequency, hence each population has a unique subgroup of the total diversity observed at the species level. For this reason, genetic diversity is frequently lower at a population level than at a species level.

Deviations of the observed distribution of genotypes from the expected distribution under Hardy-Weinberg equilibrium assumptions will be tested using a χ^2 (chi-squared) statistic, in this way deviations could be expressed as a fixation index:

$$F = 1-\frac{H_o}{H_e}$$

where H_o is the observed number of heterozygotes (Wright 1969 cited by Hall *et al.* 1994). The number of alleles needed within a locus to maintain the level of heterozygosity will be determined as:

$$A_e = \frac{1}{\sum p^2}$$ (Hall *et al.* 1994)

3.4.2 Mating system analyses

The mating system was analyzed under the mixed mating model, using the program MLTR version 3.1 of Ritland (2002). The default settings were used since there were no initial values a priori estimated for the species. From progeny array data, the program simultaneously estimates the multilocus outcrossing rates (t_m), the mean single locus outcrossing rate (t_s), the mating among relatives rate (t_m-t_s), the single locus inbreeding coefficient (Wright's fixation index) of the maternal parents (F), the multilocus correlation of paternity (fraction of siblings that share the same father) (r_p), the correlation of selfing among families (normalized variance of selfing) (r_s), and variances of the above parameters using the bootstrap method, where the progeny array (within families) was the unit of sampling. Standard errors for the parameters were calculated from 1000 bootstraps. The program calculates for each locus a χ^2 statistic to test the null hypothesis that the number of observed progeny individuals for each genotype class from each maternal genotype plant did not differ from the expected number under the mixed mating model.

Two mating system analyses from progeny arrays were performed: *i)* the first consisted in contrasting progeny arrays from isolated mother trees at pasture lands (isolation level 1) vs. those sampled from mother trees in groups or clustered at natural forests (isolation level 3); *ii)* the second consisted in obtaining mating system estimates at a family level.

4 RESULTS

4.1 Data validation

Several genotyping errors are potential sources of deviation from the Hardy-Weinberg proportions, for example: allelic dropout, short allele dropout, stuttering caused by slippage during PCR-amplification, and null alleles (van Oosterhout *et al.* 2004). Population genetic analyses could be biased due to genotyping errors, since they lead to heterozygote deficiency estimates. Data validation checks for these possible genotyping errors to control for their influence on the estimation of deviations from panmixia because of deficiencies or excesses of particular genotypes.

A total of 9 polymorphic loci were chosen to analyze all samples in the study. Selection of polymorphic loci was based on three characteristics: ease scoring, high standing allelic diversity, and easiness for multiplexing. A quality control screening step was carried out to validate the data obtained from the microsatellites, resulting in the final selection of nine loci for the analysis: Ced2, Ced18, Ced41, Ced44, Ced54, Ced61a, Ced65, Ced95 and Ced131. Furthermore, null allele presence, linkage disequilibrium, and inbreeding were tested to validate the data obtained.

Null allele presence

Null allele presence is caused by changes in the microsatellite primer sequence, resulting in the reduction or loss of the PCR product. Presence of null alleles is reflected as lower values of observed heterozygosity (*Ho*). Across all samples, null allele presence was detected in all loci, but only at rates > 0.05 at Ced65, Ced61a, and Ced41, with an average of 0.06 for the nine loci. The analysis indicated homozygote excess at all loci, except for Ced2. There was no evidence for large allele dropout at any locus. In addition, stuttering was detected for the majority of the locus, as indicated by the highly significant shortage of heterozygote genotypes with alleles of one repeat unit difference. Only Ced54 and Ced2 did not display stuttering.

A Wahlund effect (due to combination of samples from differentiated populations) may explain the homozygote excess across loci, since the analysis combined samples from

different widely-distributed sites from Southern México to Northern Costa Rica. When the null allele analysis was done at a population level, scoring errors, and null alleles presence were only present at some particular loci, in some particular populations.

But, these results might be treated cautiously due to the sampling unit of the present study. Different from most population genetic studies that deal with adult trees from native populations of a single species, our sampling consisted of progeny arrays from different adult mother trees, at different populations, through the wide range of distribution of the species studied, which are basically characterized by the isolation level of the mother trees. Due to the sampling unit of this work, population inferences from our results should be treated cautiously since we are dealing with neither established regeneration, nor a sample of adult trees from different populations.

Knowing this, when analyzing at a population level, no large allele dropout was found for any of the locus, in any population. Scoring errors due to stuttering were present mainly at two loci: Ced65 and Ced61a (Table 6). The population with most loci with stuttering problems was Tikal, at Ced65 and Ced61. On the other hand, Reforma, Cedros, Taulabe, Cañas, and Hojancha did not display scoring errors due to stuttering.

Table 6. Presence of scoring errors due to stuttering at nine microsatellite markers of Cedrela odorata in 12 populations from México, Guatemala, Honduras and Costa Rica

Population	**Ced65**	**Ced2**	**Ced54**	**Ced18**	**Ced44**	**Ced41**	**Ced131**	**Ced61a**	**Ced95**
Xpujil						X			
Reforma									
Tres Garantías									X
Tulum	X								
La Paz	X								
Taulabe									
Cedros									
Meambar	X								
Los Esclavos								X	
Tikal	X							X	
Cañas									
Hojancha									

X represent presence of scoring errors due to stuttering.

Homozygote excess was detected in almost all populations but not at all loci, two loci presented higher proportion of populations with homozygote excess, Ced65 and Ced61a

(Table 7), the same two loci with the lowest observed heterozygote coefficients (Table 6). Presence of null alleles was detected in five of the nine loci, Ced65, Ced54, Ced41, Ced61a, and Ced95.

Table 7. Presence of homozygote excess and null alleles at each locus of Cedrela odorata in 12 populations from México, Guatemala, Honduras and Costa Rica

	Ced65	**Ced2**	**Ced54**	**Ced18**	**Ced44**	**Ced41**	**Ced131**	**Ced61a**	**Ced95**
Xpujil						X		X	
Reforma									
Tres Garantías						X		X	X
Tulum	X								
La Paz	X							X	
Taulabe			X						
Cedros								X	
Meambar	X		X						
Los Esclavos						X		X	
Tikal	X							X	
Cañas								X	
Hojancha	X							X	

X represent presence of homozygote excess and null alleles.

Deviations from HWE that were not explained by the presence of stuttering, homozygote excess, and null alleles may then be explained by heterozygote deficiencies caused by subpopulation structure, also known as Wahlund effect (Hernández *et al.* 2007), or by other genetic forces, for example mating system changes.

Linkage disequilibrium

Referred to as the non-random association of alleles among loci, linkage disequilibrium becomes a consideration in the analysis of genetic data when genes are very close together at the same chromosome or when recombination rates are very low (Frankham *et al.* 2002; Lowe 2004). No linkage disequilibria were observed between the nine loci used for the current study (Table 8). [In fact, one case of linkage disequilibrium was observed but this was at the microsatellite development phase, when 10 polymorphic microsatellites where initially chosen (Annex 2). At this stage, loci Ced131 and Ced131a presented linkage disequilibria because they were sharing a primer sequence, therefore were located in the same chromosome and very close together, causing linkage disequilibrium. Locus Ced131a was discarded from the

analysis, although it is important to point out that for this test only one individual per family was used. An interpretation error may occur when analysis includes the complete progeny arrays, linkage disequilibrium occurs because different loci are expected to behave similarly, since half sibs share at least 25% of loci on average (one of two alleles from the mother) (Annex 3).

If microsatellites are mapped in the genome, one may easily discard linked loci by selecting genes that are located on different chromosomes. However, although *C. odorata* microsatellites were not mapped, they were randomly derived from whole genome digests, and the absence of linkage disequilibria indicates that loci are located on different chromosomes or scattered enough to be independent.

*Table 8. G*enotypic *disequilibrium test for all populations including one individual per family, P-value for each locus pair across all the populations (Fisher method)*

Locus pair	Chi Square	df	P-value	Locus pair	Chi Square	df	P-value
Ced131 - Ced95	0.00	2	1.00	Ced61a - Ced18	2.02	4	0.73
Ced131 - Ced65	0.00	10	1.00	Ced41 - Ced18	6.04	4	0.20
Ced95 - Ced65	0.00	2	1.00	Ced44 - Ced18	3.77	4	0.44
Ced131 - Ced61a	0.79	4	0.94	Ced131 - Ced54	0.00	4	1.00
Ced95 - Ced61a	0.00	2	1.00	Ced95 - Ced54	0.00	2	1.00
Ced65 - Ced61a	0.00	6	1.00	Ced65 - Ced54	4.04	6	0.67
Ced131 - Ced41	0.00	4	1.00	Ced61a - Ced54	Not possible		
Ced95 - Ced41	0.00	2	1.00	Ced41 - Ced54	7.14	4	0.13
Ced65 - Ced41	2.66	4	0.62	Ced44 - Ced54	4.52	2	0.10
Ced61a - Ced41	0.00	2	1.00	Ced18 - Ced54	6.70	8	0.57
Ced131 - Ced44	0.00	2	1.00	Ced131 - Ced2	0.00	6	1.00
Ced95 - Ced44	Not possible			Ced95 - Ced2	0.00	2	1.00
Ced65 - Ced44	0.00	2	1.00	Ced65 - Ced2	4.81	8	0.78
Ced61a - Ced44	0.00	2	1.00	Ced61a - Ced2	0.00	6	1.00
Ced41 - Ced44	0.00	2	1.00	Ced41 - Ced2	0.00	4	1.00
Ced131 - Ced18	1.50	8	0.99	Ced44 - Ced2	0.00	4	1.00
Ced95 - Ced18	Not possible			Ced18 - Ced2	0.00	6	1.00
Ced65 - Ced18	0.22	8	1.00	Ced54 - Ced2	0.00	6	1.00

The null hypothesis (H_0) is: "*Genotypes at one locus are independent from genotypes at the other locus*". H_0 is rejected for small p-values (< 0.05).

Inbreeding

Single locus F-statistics were estimated to evaluate if deviations from Hardy-Weinberg equilibrium may be explained by inbreeding, or non random mating due to population structure. Moderate genetic differentiation (F_{ST}) was found for all loci, with a mean value of 0.08 (Table 9). Low values for inbreeding coefficient (F_{IS}) were obtained for the majority of the loci, except Ced65 and Ced61a, which presented higher values; similarly occurred for the overall inbreeding coefficient (F_{IT}).

Table 9. Singlelocus F-statistics estimates for Cedrela odorata in 12 populations from México, Guatemala, Honduras and Costa Rica

Locus	F_{IS}	F_{ST}	F_{IT}
Ced131	0.00	0.10	0.11
Ced95	0.01	0.07	0.08
Ced65	0.14	0.08	0.21
Ced61a	0.26	0.06	0.30
Ced41	0.09	0.08	0.16
Ced44	-0.03	0.08	0.05
Ced18	0.00	0.07	0.07
Ced54	0.04	0.08	0.11
Ced2	-0.02	0.07	0.05
Mean	**0.06**	**0.08**	**0.13**

F-statistics were estimated as in Weir and Cockerham 1984.

Low inbreeding coefficient estimates indicate small reductions in heterozygosity due to non-random mating within populations. When expected heterozygosity equals observed heterozygosity, population will be panmitic and F_{ST} and F_{IS} values will be zero (Balloux and Lugon-Moulin 2002). Mean F-statistic estimates are useful since among a group of populations each locus may segregate and drift in frequency independently and differently from the other loci.

The mean overall inbreeding coefficient (F_{IT}) was of 0.13, which describes the reduction of heterozygosity within individuals relative to the total population due to non-random mating within populations and effects of random drift among populations.

4.2 Genetic diversity and heterozygosity

A total of nine highly polymorphic microsatellites were used to study the genetic diversity of *C. odorata*. The number of individuals genotyped at each locus varied between 468 and 381 progenies (Table 10). Allelic diversity (*Na*) across loci was 20.56, which corresponds to the number of alleles averaged across loci. The number of alleles per locus varied between loci, ranging from 12 alleles at locus Ced131 up to 30 alleles at locus Ced44. Three markers were very informative to genotype the species presenting the highest number of alleles, Ced54 and Ced95 with 28 alleles, and Ced44 with 30 alleles; and presenting the highest probabilities of paternity exclusion (Pr (Ex_1) > 0.70). The proportion of polymorphic loci at the species level was 0.81, which corresponds to 21 polymorphic loci from 26 loci sampled (Annex 2).

Means of observed heterozygosity (*Ho*) and expected heterozygosity (*He*) were 0.78 and 0.88, respectively. Average levels of observed heterozygosity (*Ho*) (0.78) were consistent across loci, although slightly different for three loci, Ced44 with a higher value (0.90), and Ced61a and Ced65 with smaller values (0.62 and 0.64, respectively). Expected heterozygosity (*He*) coefficients were always relatively higher than the observed (*Ho*) in all loci.

Table 10. Basic descriptive statistics of nine microsatellite markers isolated from Neotropical tree Cedrela odorata

Locus	N	Na	Ne	Ho	He	F	Pr (Ex_1)	Pr (Ex_2)	Null rate
Ced2	396	17	8.15	0.85	0.88	0.03	0.60	0.75	0.01
Ced18	401	19	6.40	0.80	0.84	0.05	0.54	0.71	0.03
Ced41	438	21	10.96	0.78	0.91	0.14	0.69	0.82	0.07
Ced44	451	30	13.99	0.90	0.93	0.03	0.75	0.86	0.02
Ced54	405	28	15.64	0.85	0.94	0.09	0.77	0.87	0.05
Ced61a	420	13	7.54	0.62	0.87	0.29	0.58	0.73	0.14
Ced65	468	17	4.99	0.64	0.80	0.20	0.45	0.63	0.10
Ced95	381	28	11.60	0.86	0.91	0.06	0.70	0.83	0.04
Ced131	415	12	5.80	0.75	0.83	0.10	0.50	0.68	0.05
Average		20.56	9.45	0.78	0.88	0.11	0.62	0.76	0.06

Abbreviations are: number of individuals (*N*), number of alleles (N_A), effective alleles (*Ne*), observed (H_O) and expected (H_E) heterozygosity, fixation index (*F*), and exclusion probability Pr(Ex - for first parent $_1$ and second parent $_2$).

The pattern of allele frequency distribution was uniform in almost all the microsatellites evaluated (Figure 5). Allele frequency distributions of loci Ced18, Ced65, and Ced131 showed a low number of alleles with high frequencies. On the other hand, the rest of the loci presented more than two alleles sharing the highest frequencies. In general,

frequencies for individual alleles at all loci were generally low, with only 1.6% of the alleles having values greater than 0.25 at a locus, which is the case for loci Ced18, Ced65 and Ced131.

Average values of heterozygosity were not significantly different between Northern and Central lineages, indicating similar gene diversity in both lineages. Means of observed multilocus heterozygosity (*Ho*) and expected multilocus heterozygosity (*He*) in the Northern lineage were 0.79 and 0.87, respectively, while for the Central lineage were 0.78 and 0.86 (Table 11). Number of alleles shows the same pattern at Northern and Central lineages, 18 and 17, respectively.

Table 11. Basic descriptive average statistics of nine microsatellite markers isolated from populations of Cedrela odorata from Northern and Central lineages

Lineage	N	Na	Ne	Ho	He
Northern	213	18	8.83	0.79	0.87
Central	206	17	8.55	0.78	0.86

Abbreviations are: number of individuals (*N*), number of alleles (*N*a), effective alleles (*Ne*), observed (*H*o) and expected (*H*e) heterozygosity.

When comparing genetic diversity between the isolation levels, results were almost identical (no significant differences) to the ones found from comparing Northern and Central lineages (Table 12). Observed heterozygosity (*Ho*) was the same for both isolation levels, 0.77, while expected heterozygosity (*He*) was 0.88 and 0.87. Higher numbers of alleles were found for families under isolation level 1 (isolated mother trees), which is explained by the difference of individuals between isolation levels.

Table 12. Basic descriptive average statistics of nine microsatellite markers isolated from populations of Cedrela odorata from populations with different levels of isolation

Isolation level	N	Na	Ne	Ho	He
Isolation 1	268	20	9.34	0.77	0.88
Isolation 3	157	17	8.42	0.77	0.87

Abbreviations are: number of individuals (*N*), number of alleles (*N*a), effective alleles (*Ne*), observed (*H*o) and expected (*H*e) heterozygosity.

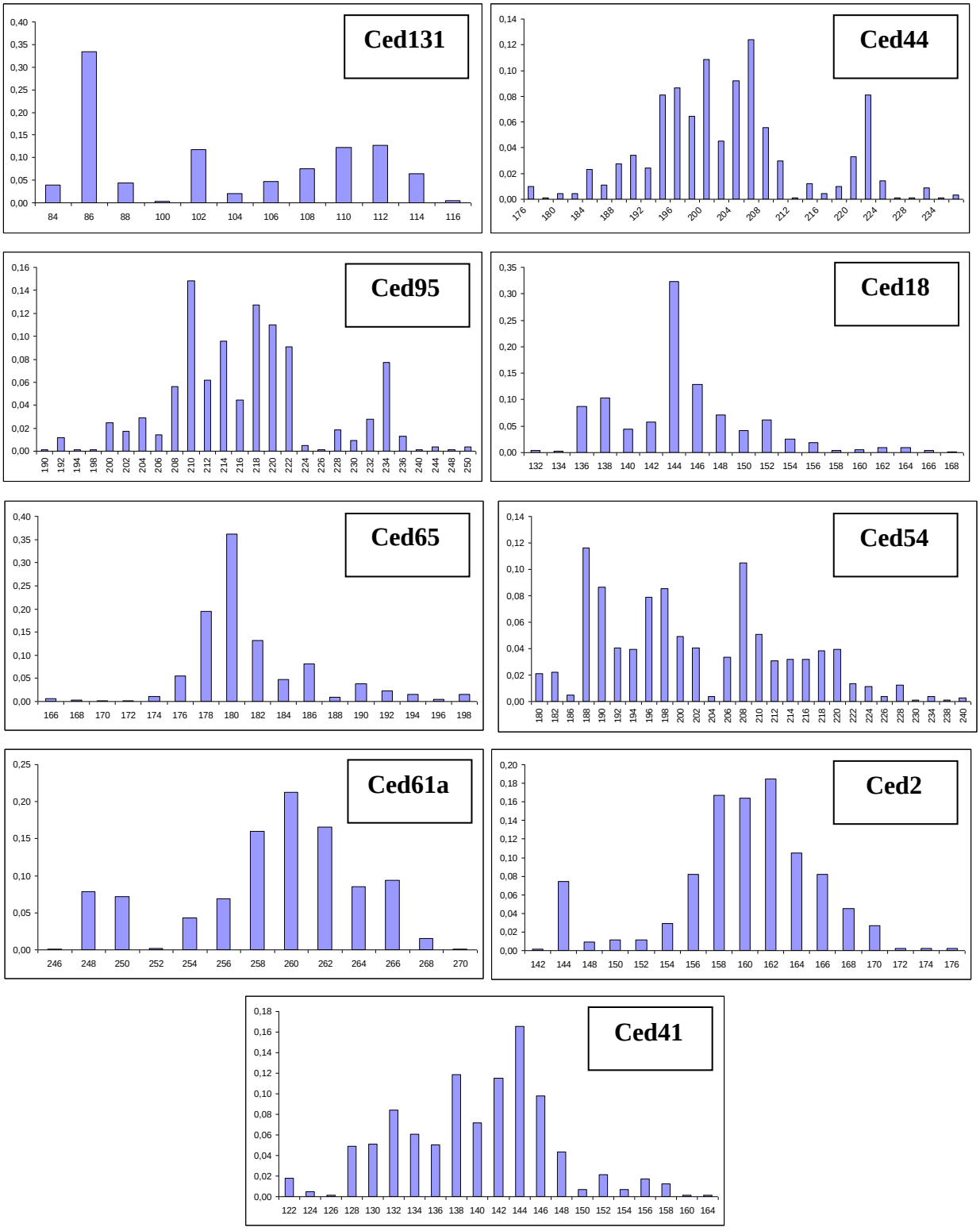


Figure 5. Patterns of allele frequency distribution for nine microsatellite loci of Cedrela odorata. X axes, are the allele sizes in base pairs; Y axes, are the allele frequencies.

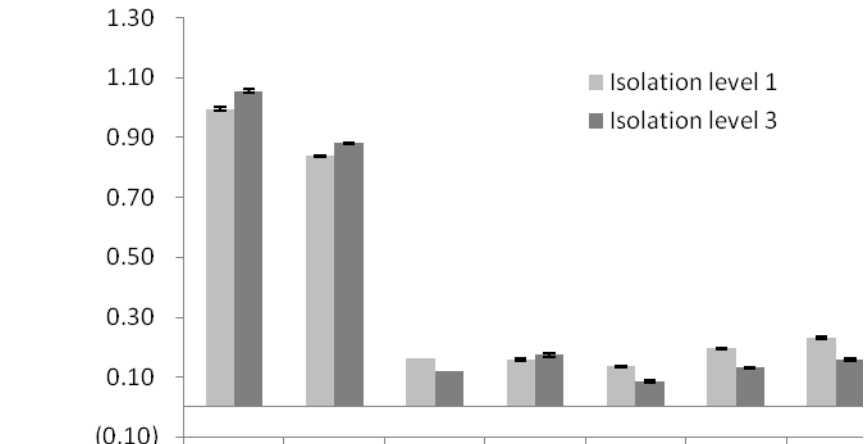
4.3 Mating system analysis

High multilocus outcrossing rates (t_m) were obtained for both isolation levels, 1.00 ± 0.08 and 1.06 ± 0.06 for isolation level 1 and 3, respectively, suggesting absence of self-fertilization (Figure 6). Singlelocus outcrossing rates (t_s) were of 0.84 ± 0.02 and 0.88 ± 0.02 for isolation level 1 and 3, respectively. No significant differences for mating system estimates were found between *C. odorata* families growing under different isolation levels. In both isolation levels singlelocus outcrossing estimates were smaller than multilocus estimates, therefore single locus selfing rates are higher than the multilocus selfing rates, indicating apparent biparental inbreeding (mating between close related parents).

The proportion of self fertilized (*s)* trees was zero for both isolation levels which is calculated as 1 - *t* = *s*. Singlelocus selfing rates do not have error bars because they are derived from the outcrossing rates provided by the program MLTR. Correlation of paternity (r_p) estimates were significantly different between the isolation levels, a trend of higher values was found for isolated mother trees (0.20 ± 0.03) when compared to mother trees in groups or clustered at natural forests (0.13 ± 0.03). Similarly for estimates of singlelocus inbreeding coefficient of maternal parents (*F*), which was higher for the isolation level 1 (0.14 ± 0.04) compared to isolation level 3 (0.08 ± 0.04) although not statistically different.

An estimate of the number of fathers contributing to the fertilization of gametes and viable siblings is obtained with the inverse function of the correlation of paternity r_p (notation: r_p^{-1}). Thus, for the isolation level 1 there is an average of five trees donating pollen to the isolated mother trees, while for the isolation level 3, there is an average of 7.7 trees.

The correlation of selfing among families (r_s) indicates the extent of variation among families at each isolation level for the selfing rate. In other words, if selfing occurred in a group of families or in the other extreme, if it is generalized at all families. In this case, since no apparent selfing occurred, r_s estimates do not apply.



	tm	ts	ss	tm-ts	F	rp	rs
Isolation level 1	1.00	0.84	0.16	0.16	0.14	0.20	0.23
Isolation level 3	1.06	0.88	0.12	0.17	0.08	0.13	0.16

Figure 6. Mating system estimates of isolation level 1 and 3 of Cedrela odorata. Error bars are presented for each parameter with 1.96 standard errors. Abbreviations are: multilocus outcrossing rate (tm), singlelocus outcrossing rate (ts), singlelocus selfing rate (ss), singlelocus inbreeding coefficient of maternal parents (F), correlation of paternity (rp), and correlation of selfing among families (rs).

Estimates for *C. odorata* mating system using overall data were very similar to the ones presented by isolation level (Table 13). Since singlelocus selfing rate (s_s) was higher than multilocus selfing rate (s_m), biparental inbreeding might be occurring at *C. odorata* populations.

Table 13. Mating system estimates for overall data of C. odorata

tm	1.00	(0.08)
ts	0.84	(0.02)
sm	0.00	
ss	0.16	
tm-ts	0.16	(0.08)
F	0.13	(0.03)
rp	0.18	(0.02)
rs	0.23	(0.04)

Family level estimates of multilocus and singlelocus outcrossing rates were calculated for the data set, results are presented for each isolation level (Figure 7 and Figure 8). Multilocus outcrossing rates were higher than singlelocus rates for the majority of the families.

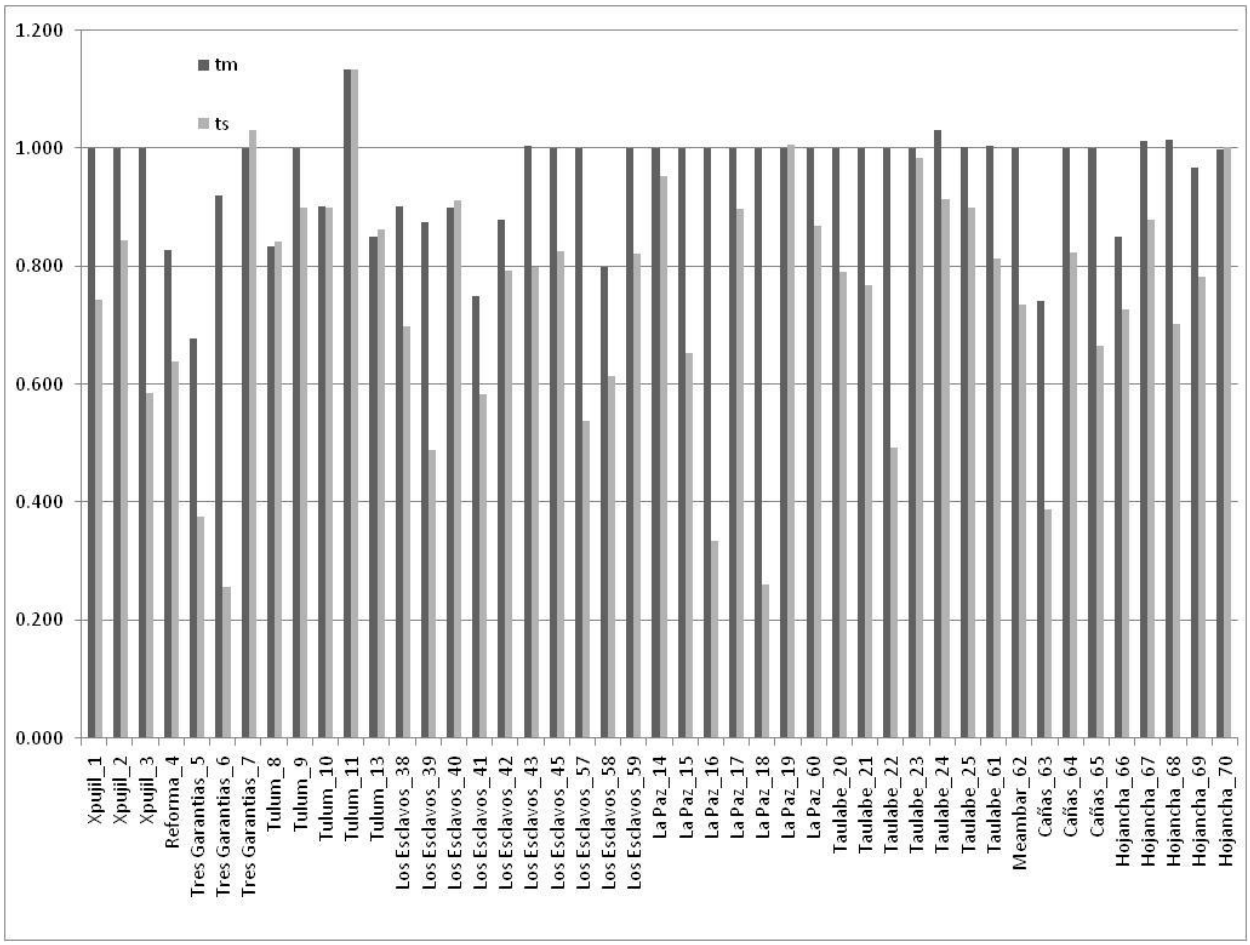


Figure 7. Family level estimates of single and multi locus outcrossing for Cedrela odorata families with isolation level 1.

Multilocus outcrossing mean values for families under isolation level 1 were not consistently close to 1.0, except for families from Xpujil, La Paz, Taulabe and Meambar (Figure 7). On the other hand, singlelocus outcrossing values were distributed more widely across families. Families under isolation level 3 presented higher numbers of families with multilocus outcrossing rates closer to 1.0, while singlelocus outcrossing rates ranged from 0.3 to 1.0 (Figure 8).

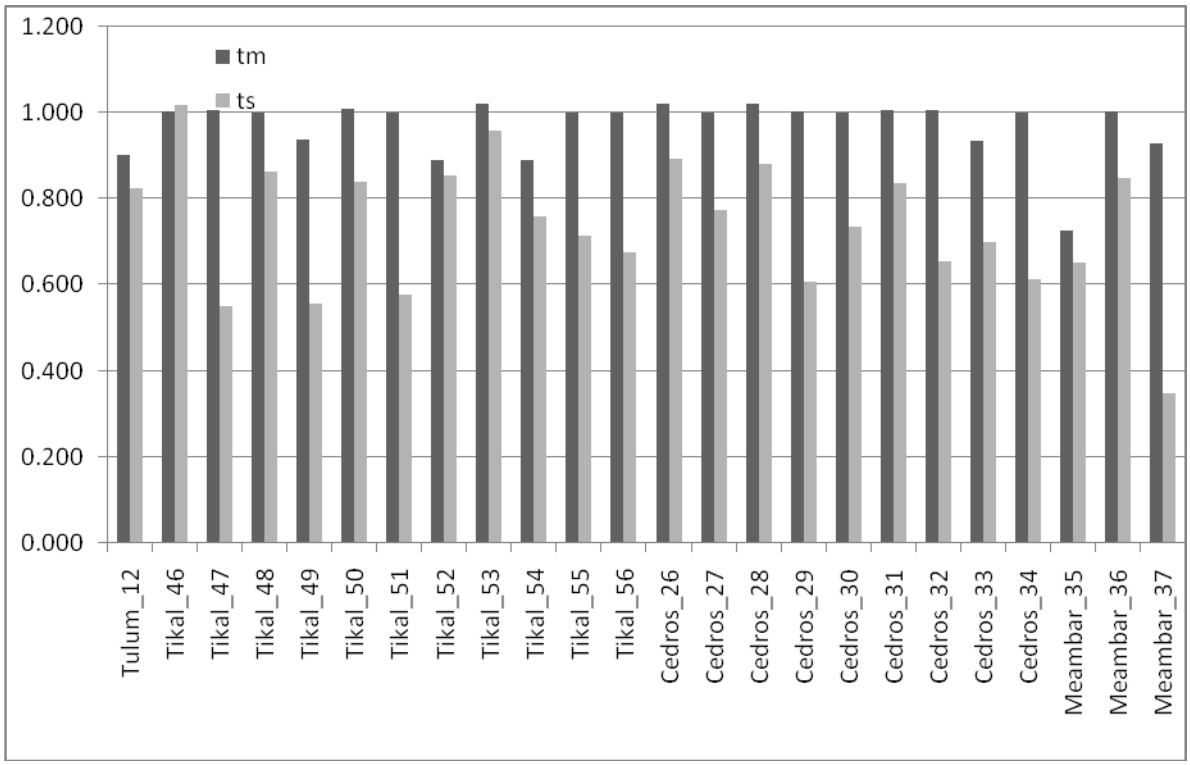
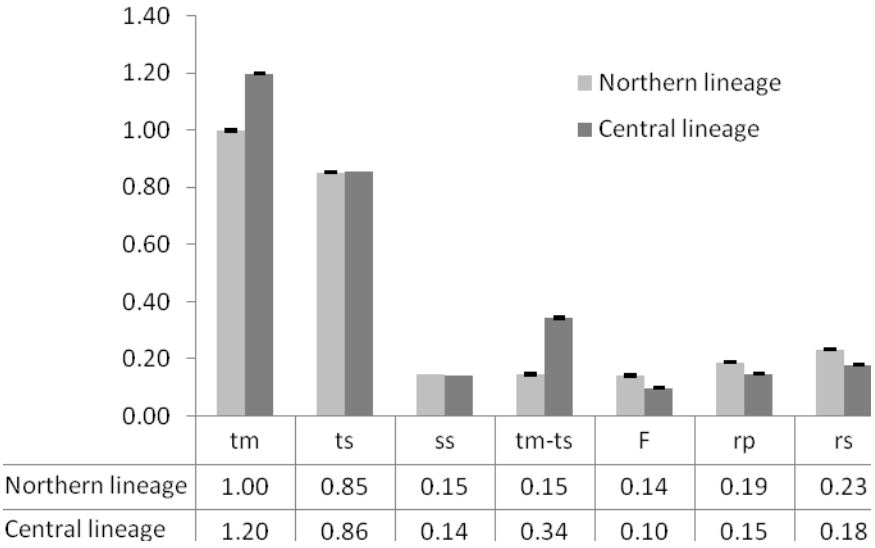


Figure 8. Family level estimates of single and multi locus outcrossing for Cedrela odorata families with isolation level 3.

Mating system analysis at a lineage level did not showed differences between Northern and Central lineage (Figure 9). On the other hand, when mating system analysis was performed at a regional level for each isolation level, important differences were observed between and among regions for the isolation level estimates (Table 14). Isolated families (isolation 1) presented similar high estimates of multilocus and singlelocus outcrossing rates, corroborating no selfing for the species when trees are no closer than 500 meters at a regional level. A similar result was found for trees grouped in clusters or within 100 meters ratio (Isolation 3) at Guatemala and Honduras-Costa Rica families; although multilocus outcrossing rate was 0.90 for Mexico, it is explained due to a small number of families under the category isolation 3 for that region, only 1. Costa Rican and Honduran populations were grouped in the same region since there were no families with the isolation level 3 in Costa Rican populations. Inbreeding levels for maternal parents are higher at isolated families for the three regional groups, and smaller for Mexican families when compared to other regions, despite the

isolation level. Interestingly, Mexican families presented higher correlation of paternity estimates and lesser number of pollen donors for both isolation levels.



	tm	ts	ss	tm-ts	F	rp	rs
Northern lineage	1.00	0.85	0.15	0.15	0.14	0.19	0.23
Central lineage	1.20	0.86	0.14	0.34	0.10	0.15	0.18

Figure 9. Mating system estimates of isolation level 1 and 3 of Cedrela odorata. Error bars are presented for each parameter with 1.96 standard errors. Abbreviations are: multilocus outcrossing rate (tm), singlelocus outcrossing rate (ts), singlelocus selfing rate (ss), singlelocus inbreeding coefficient of maternal parents (F), correlation of paternity (rp), and correlation of selfing among families (rs).

Table 14. Mating system estimates of isolation level 1 and 3 of Cedrela odorata families grouped at a regional level

Regional Level	Mexico		Guatemala		Honduras-Costa Rica	
	Isolation 1	Isolation 3	Isolation 1	Isolation 3	Isolation 1	Isolation 3
tm	0.98 (0.07)	0.90 (0.00)	1.20 (0.00)	1.01 (0.08)	1.20 (0.05)	1.20 (0.00)
ts	0.84 (0.04)	0.96 (0.01)	0.94 (0.03)	0.92 (0.03)	0.87 (0.03)	0.95 (0.03)
sm	0.02	0.10	-0.20	-0.01	-0.20	-0.20
ss	0.16	0.04	0.06	0.08	0.13	0.06
tm-ts	0.15 (0.06)	0.06 (0.00)	0.26 (0.03)	0.09 (0.08)	0.33 (0.06)	0.26 (0.03)
F	0.06 (0.05)	-0.06 (0.00)	0.20 (0.13)	0.04 (0.07)	0.11 (0.05)	0.03 (0.06)
rp	0.20 (0.05)	0.20 (0.01)	0.12 (0.05)	0.10 (0.03)	0.13 (0.03)	0.12 (0.04)
rs	0.25 (0.07)	0.20 (0.01)	0.13 (0.04)	0.08 (0.04)	0.03 (0.03)	0.12 (0.04)
Number of pollen donors	4.88	5.00	8.33	10.10	7.81	8.55

Values between parentheses are the standard deviations. Abbreviations are the same as Figure 9.

5 DISCUSSION

5.1 Impact of tree isolation on genetic diversity and heterozygosity

The high diversity levels observed for *C. odorata* match the expected diversity levels for an outcrossing, widespread tropical tree species, although at a population level, these results could be higher from the ones expected in natural *C. odorata* populations, since progeny arrays should represent a very wide sample of the natural gene pool relative to that normally sampled for population studies of adult trees. No significant differences were found between the Northern and Central lineage genetic diversity estimates, indicating that lineage is not a factor that may affect findings at the isolation level analysis.

Progeny arrays from isolated mother trees presented the same observed heterozygosity estimate (0.77) as the ones from mother trees growing in groups or near other trees from the same species at natural forest. From this, it could be said that there is no loss of genetic diversity for populations where trees are scattered in pasture lands, agricultural lands, or other human dominated landscapes, for minimum distances of 500 meters between isolated trees. Apparently, natural pollinators from the species are capable of flying long distances, and are still present in human dominated landscapes. *C. odorata* flowering events are frequented by small moths, bees and many other kind of small insects (personal observation), and possibly effective pollination might be carried out by insects capable of flying long distances. Pollination studies for the species should be developed to understand the complex ecological interactions that generate gene diversity and mating patterns.

Allele frequencies were not significantly different between the two isolation levels, except for the number of private alleles (Figure 10). Progeny arrays from isolated trees presented a higher number of private alleles due to a higher number of families and progenies sampled for this isolation level. For example all Costa Rican families are under the category of isolation level 1, consequently all private alleles from Costa Rican population will be assigned to this level.

When comparing current results with previous genetic studies using microsatellite markers for different species of the Meliaceae family, *C. odorata* shows higher genetic diversity than other species, gauged by higher values for expected and observed heterozygosity, and allelic diversity (Table 15), again the fact that progeny arrays have been sampled for the present study must be taken into consideration. *C. odorata* populations from

Northern and Central Mesoamerica have higher genetic diversity than related Meliaceae species, *Swietenia macrophylla*, *Swietenia humilis*, and *Carapa guianensis*. High diversity levels could be associated with dry ecotype populations, which are well represented in the study by northern and central lineage provenances, both with five or more dry months. According to Chase *et al.* (1995), dry ecotype populations may have higher diversity levels than humid ecotypes possibly because of a greater variety of habitats in dry ecosystems, where variable availability of water resources creates higher habitat heterogeneity.

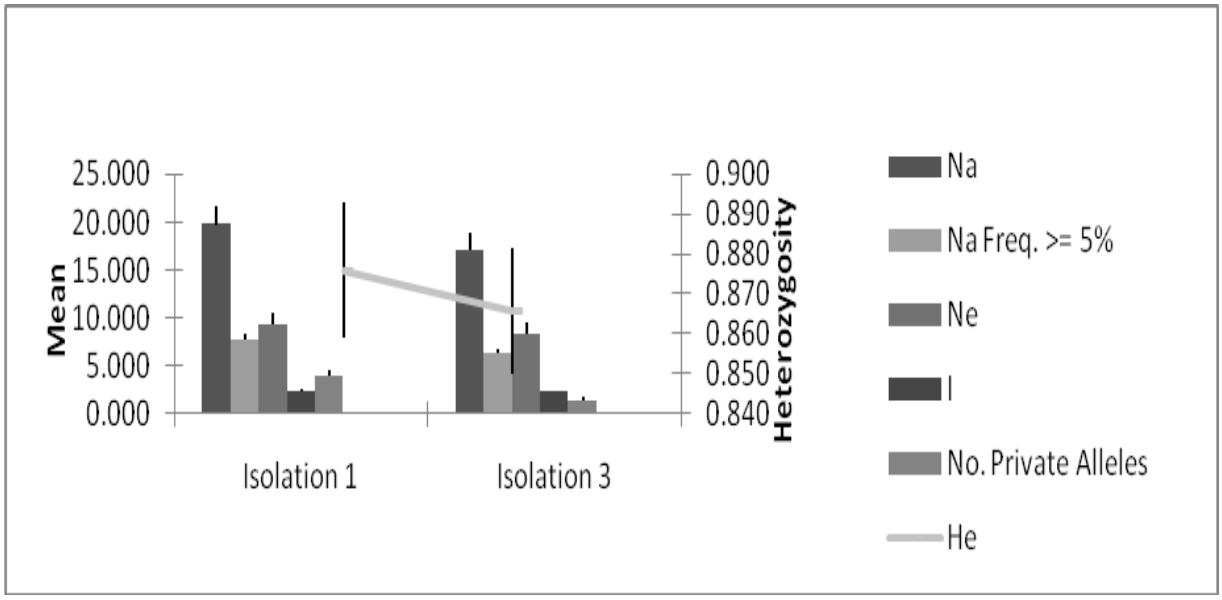


Figure 10. Allelic patterns of Cedrela odorata *across the different isolation levels.*

Previous AFLP study of *C. odorata* had shown differences between dry and humid provenances of Costa Rica, while similar findings were observed with *Cordia alliodora* (Chase *et al.* 1995) and *Swietenia macrophylla* (Lowe *et al.*2003), also in Costa Rican populations. According to these studies, future surveys in southern lineage populations might present lower levels of diversity from the ones observed in the present study.

Similar heterozygosity values and numbers of alleles were found when compared with the African Meliaceae, *Entandrophragma cylindricum* (Garcia *et al.* 2004), which could validate the idea that dry populations may have high genetic diversity levels, since *E. cylindricum* is also found in this kind of habitat. On the other hand, *Swietenia macrophylla* from Brazilian Amazon, a humid ecotype, also had similar heterozygosity values to the ones obtained for *C. odorata*, but in this case, different from the reasoning that genetic diversity is

higher in dry ecotypes than in humid ecotypes, since populations from *Swietenia macrophylla* from the north region and west coast of Costa Rica and Mesoamerica have lower genetic diversity values. Probably, range-edge effects. overexploitation, selective logging, and habitat loss in Mesoamerican populations of *Swietenia macrophylla* are the main factors influencing this reduction of genetic diversity.

Table 15. Basic descriptive statistics for isolated microsatellite markers (SSRs) in Meliaceae family studies of Latin America and one in Africa

Species	Region	No SSRs	N	Na	Ho	He	Reference
Swietenia macrophylla	Mesoamerica	7	287	13.0	0.56	0.66	Novick *et al.*(2003)
Swietenia macrophylla	Costa Rica: North Region	3	72	2.67	0.15	0.31	Lowe *et al.* (2003)
	Costa Rica: West Coast	3	74	5.00	0.52	0.52	
Swietenia macrophylla	Brazilian Amazon	8	194	9.50	0.75	0.78	Lemes *et al.* (2003)
Swietenia humilis	Honduras Pacific Coast	10	171	9.70	0.49	0.45	White and Powell (1997)
Carapa guianensis	Brazilian Amazon	5	12	4.20	0.57	0.69	Vinson *et al.* (2005)
Carapa guianensis	Brazilian Amazon	6	1241	9.33	0.67	0.71	Cloutier *et al.* (2007)
Cedrela odorata (Current study)	Northen and Central Mesoamerica	9	451	20.56	0.78	0.88	From Table 8
Entandrophragma cylindricum	Cameroon	7	186	22.10	0.77	0.85	García *et al.* (2004)

Abbreviations are: number of individuals (N), number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities.

High levels of genetic variation are also expected in species with large distribution ranges, high levels of outcrossing, high fecundities, wind pollination systems, long generation times and that occur in habitats representing later stages of succession (Hamrick *et al.* 1979, cited by Ratnam and Boyle 2000). Such characteristics describe *C. odorata* species except for wind pollinated, since the species is pollinated by small moths (Bawa *et al.* 1995) and its seeds are wind dispersed.

Low genetic diversity is associated with endangered species and populations when compared to non-endangered species with large population sizes. A comparison of microsatellite genetic diversity levels between these two categories suggests that endangered species posses about half the genetic diversity of non-endangered species (Frankham *et al.* 2002). Considering that *C. odorata* is included in the IUCN Red List of Threatened Species as

vulnerable, and the population size of the species has been severely reduced because of selective logging, habitat loss and degradation, and extraction for wood and charcoal, the low genetic diversity does not apply. However, this does not deny the fact that the species has been overexploited for its high quality wood through its distribution range. Probably characteristics of the species such as wide distribution range, fast growth, pollination by insects capable of flying long distances, wind dispersed seeds, and outcrossing, combined with social and economical factors such as people valuing *C. odorata* trees, letting natural regeneration of the species take place, sowing in live fences, and establishing plantations and agroforestry systems, may explain such high genetic diversity levels.

Lowe *et al.* (2004), recommend treating cautiously the diversity measures when microsatellites markers are used, since the high number of alleles per locus obtained may cause some bias, due to increased levels of heterozygosity. A large number of microsatellites (nine) were used to counteract the possibility of diversity estimates bias. Comparison of diversity levels between different tree tropical species (Table 15) all assessed with microsatellite marker technique shows that *C. odorata* posses a higher genetic diversity over the other species, indicating that its high diversity is not due only because of the type of molecular marker used. To corroborate this, a further study with the current microsatellite markers should be done, using adult trees from the studied populations.

Significant phylogenetic divergence between northern and central lineages was observed in analysis of chloroplast DNA variation (Cavers *et al*. 2003) although no quantitative divergence was detectable between these two regions: the diversity (phylogenetic and quantitative) was mainly between populations of these two areas and the southern lineages (Navarro *et al.* 2005). For microsatellite data, means of observed heterozygosity (*Ho*) and the average number of alleles in Northern and Central lineage were very similar, suggesting no gene flow isolation between lineages, supporting Navarro *et al.* (2005) findings for quantitative data, and denoting that chloroplast DNA variation may be explaining ancient colonization events.

Considering the dataset as a whole, both lineages presented at all loci relatively higher expected heterozygosity (*He*) coefficients when compared to observed heterozygosity (*Ho*). Deficiency of heterozygotes in populations indicates that populations are not mating randomly (Frankham *et al.* 2002), as might be expected of the Central American *Cedrela odorata* population given the great distances involved.

5.2 Impact of tree isolation on mating system

The impact of changes in the landscapes on the reproductive biology and population genetics of *C. odorata* populations was assessed through the isolation level analysis. Habitat loss, landscape fragmentation and over exploitation of *C. odorata* trees are the main threats to reproductive patterns of the species throughout its natural growth range. Should it occur, reproductive isolation may have serious consequences on the fitness of populations, and may contribute to loss of genetic diversity at the population and species level, and cause increased inter-population structure. Furthermore, reproductive isolation is associated with reduction in the size of plant populations, reduced gene flow, elevated inbreeding, and inbreeding depression (Young *et al*. 1996; Young and Boyle 2000).

Estimates from mating system analysis confirm that *C. odorata* is an obligate outcrosser through all Northern and Central lineage populations, and possibly through Southern lineage as well. Analysis of the impact of isolation level gave more information about the mating structure of the populations and how environmental factors and the number of reproductive individuals affect the mating patterns.

Complete outcrossing was found for both isolation levels, indicating self-fertilization of *C. odorata* trees occurs very rarely if at all, and seedlings are the product of crossed pollination.

The number of pollen donor trees contributing to the fertilization did differ between the isolation levels, as indicated by the difference between the correlations of paternity estimates (r_p). A mean number of 5.0 trees are effectively contributing with pollen to mother trees separated at least 500 meters from another *C. odorata* tree (isolation 1), while mother trees within groups or close to another *C. odorata* tree within a ratio of 100 meters, receive pollen contribution from at least 7.7 trees (isolation 3).

Similar findings were observed by Rocha and Aguilar (2001) between *Enterolobium cyclocarpum* trees from pastures and trees from continuous forests, where no significant difference in the outcrossing rates (t_m) were observed. Differently from our results, they detected lower correlation of paternity estimates for progeny arrays from trees in pastures, indicating higher amount of pollen donors in pasturelands. Additionally, they found that progeny from continuous forests were more vigorous than progeny from pasture trees, suggesting that the mechanisms that regulate progeny vigor are disrupted in trees from pastures, lacking of the conditions for selective abortion among different pollen donors.

Quantitative studies for *C. odorata* in different isolation levels, indicates higher vigor for progenies from not isolated mother trees from continuous forests (Navarro *et al.* 2008). Probably for isolated trees, the higher levels of biparental inbreeding, lesser amount of pollen donors, and the reduction of mechanisms for selective abortion among different pollen donors, leads to a decrease in vigor of progeny.

Although *C. odorata* occurs at low densities in natural forests, sampled forest population of Tikal was characterized by a high density of mature trees. Cedros was another population with mother trees clustered or with trees within a ratio of 100 meter, but differently from Tikal, mother trees were located in fragmented forests, pasture lands, and even in the backyards of houses at a community. Forest site condition differences between Tikal and Cedros did not lead to differences of gene diversity or mating systems estimates between populations, which indicates that pollination events occurred normally at both sites, despite the fact that Cedros population was not a continuous forest. On the other hand, populations from isolated mother trees were all located mainly in pasturelands, cultivated lands, fragmented forests, or near communities, but families did present differences in mating system estimates, indicating that isolation level and degree of disturbance may be affecting mating process.

Other studies had revealed differences in outcrossing rates in different habitats, for example, Liengsiri *et al.* (1998) found differences in the single and multilocus outcrossing rates for populations of *Pterocarpus macrocarpus*, revealing a geographic pattern with western populations having higher outcrossing than eastern populations. They attribute low estimates of outcrossing rates in eastern populations to habitat characteristics, like degree of disturbance, density, and the distribution of flowering mature trees.

Comparisons between populations of Mesic and Xeric habitats of several grass species showed differences in the outcrossing rates (Table 16), mainly explained by environmental factors that influence the mating structure of the population (Clegg 1980). Very different from *C. odorata*, these species are annual grass plants, and predominantly selfing with very low rates of outcrossing. It is important to recognize that there are differences between primarily selfing and outcrossing species, for example that selfers usually express inbreeding depression late, while outcrossed species exhibit earlier inbreeding depression with recessive lethals that can be purged through selection (Husband and Schemske, 1996). The intention of presenting these results was to highlight how environmental factors may affect the outcrossing rates of

plant species, rather than making a comparison of outcrossing magnitudes between grasses and trees.

Table 16. Environmental influences on outcrossing rates (t) for several species

		Habitat			
		Mesic		Xeric	
Species	Research	*T*	(SE)	*T*	(SE)
Avena barbata	Marshall and Allard (1970)	0.075	(0.012)	0.014	(0.004)
Bromus mollus	Hamrick and Allard (1972)	0.021		0.005	
	Brown *et al* (1974)	0.110	(0.027)	0.007	(0.027)
Hordeum spontaneum	Brown *et al* (1978)	0.021	(0.0001)	0.004	(0.0002)

Source: Clegg (1980)

Populations were grouped and analyzed under lineages since phylogenetic studies with chloroplast DNA revealed haplotypes differentiation between three regions in Mesoamerica: northern, central, and southern lineage. Phylogenetic associations with the mating system were considered in this study, since anatomical and biochemical aspects of self incompatibility systems could be restricted phylogenetically (Gribel *et al*, 1999). Also, considering that outcrossing rates are under genetic control, and that the mating structure of most plant species is variable and subject to environmental influences (Clegg 1980), mating system at the lineage level was analyzed.

Populations from Northern lineage have higher annual precipitations and number of dry months than Central lineage populations (Table 2). Lineages also differ considerably at the landscape level, where Northern lineage populations (excluding Los Esclavos) could be characterized as an area of highland consisting of a flat surface, also known as plateau, while Central lineage populations are in lower lands but with an irregular topography. Other dissimilarities between the lineages are the incidence of deforestation and degree of disturbance of landscapes, which is higher in Central lineage populations (Navarro 2008 pers. comm.). However, despite environmental differences between sites, all populations were fully outcrossing (t_m ranged between 1.0 and 1.2) regardless of lineage.

C. odorata northern lineage presented higher estimates of correlation of paternity (r_p) and correlation of selfing among families (r_s), than central lineage, indicating that fewer number of trees are contributing on average as pollen donors in northern lineage populations (5.26), compared to central lineage (6.66). These differences might be explained by individual population or regional characteristics of the family arrays, rather than lineage level or by

population isolation levels, since each lineage had a similar amount of family arrays with isolation level 1 and isolation level 3. In the case of isolation level 1 Northern and Central lineages were represented with 22 and 24 family arrays, respectively, while both lineages had 12 family arrays with isolation level 3 (Table 3).

To gain a better understanding of the latter, a further mating analysis between geographically related populations was done to probe consistency with lineage analysis results. Populations were grouped by countries in three main groups: Mexico, Guatemala and Honduras-Costa Rica. Differences in the correlation of paternity and correlation of selfing among families were significant between Mexican populations and Guatemalan and Honduran-Costa Rican populations (Table 14). These differences might be reconsidered for further studies of *C. odorata*, since probably phylogenetic lineages might not explain mating patterns for the species, while regional environmental patterns might explain partially of the mating system differences among populations.

Multilocus estimations are statistically more efficient than singlelocus because multilocus data sets contain more info about outcrossing than is available at any one singlelocus (Muluvi *et al*. 2004). Singlelocus estimations are more sensitive to related mating other than selfing, which explain why higher singlelocus selfing rates over multilocus selfing rates indicate biparental inbreeding. When analyzing close relatives, there is a higher probability of homozygosity at some loci, which is the reason why singlelocus estimates can give additional information of how closely related a progeny array might be. For singlelocus estimates, all loci are used but individual locus estimates are obtained separately (one by one), whereas multilocus estimates consider all loci at the same time.

Although it was low, some apparent biparental inbreeding was observed in both lineages, 0.34 for Central lineage, but a lower estimate for Northern lineage, 0.15 (t_m-t_s). Difference between lineages was because Central lineage had a multilocus outcrossing rate of 1.20, which is higher than the expected (1.0) for any estimation. Ruter *et al*. (2000) explain that several factors may cause higher outcrossing estimates: first, can suggest obligate outcrossing or self incompatibility; second, may be caused by violations of the mixed mating model assumptions. There are three important assumptions: *i)* mating events are due to random outcrossing (with probability t) or self fertilization (with probability $s = 1 - t$); *ii)* gene frequency distribution among pollen should be identical over all maternal plants; and *iii)* the rate of outcrossing is independent of maternal genotype (Clegg, 1980).

The analysis of the mating system at a family level revealed higher multi and single locus outcrossing rates within families from isolation level 3, which confirms less biparental inbreeding in natural forests and sites where *C. odorata* trees are more abundant. Differences among site environmental conditions, conservation status of the landscapes at each population, and relation between local pollinators and *C. odorata* trees might be favoring families of isolation 3. High variance of singlelocus outcrossing rates were found for isolated families, which is due to mating between related parents and due to a sampling error, since there were less number of progenies per family, at isolation level 1.

One of the main factors that may be causing estimates above 1.0 could be self incompatibility of the species. Regarding the assumptions of the mixed mating model there are also violations due to sampling and population characteristics. The assumption of identical gene frequency distribution among pollen was not identical over all maternal trees since we are working with different isolation levels of mother trees, which lead to heterogeneous pollen pools.

5.3 Implication for conservation

Outcrossing rate estimates of 1 indicates that the species is completely outcrossing. Our results indicate that *C. odorata* is an obligately outcrossing species. This information is significant for decision makers of the management and conservation of the species, since fertility of individual trees will depend on the degree of isolation, human disturbance habitats and availability of pollinators that may flight long distances between trees.

Habitat degradation caused by selective logging has reduced the species density at natural forests, leading to common scenarios with small fragmented forests with few remnant *C. odorata* trees. Knowing that the species is completely outcrossing, if these trees are at an isolation level where pollinators are unable to visit another tree, 'living dead' (alive but reproductively redundant) trees are likely to be occurring in natural forests. The same situation will be happening for populations with isolated trees in pasture lands or other human dominated landscapes, where harvesting permissions are easy to obtain and do not consider any ecological, biological or genetic information of the species.

Differences determined at a regional level indicate that management and conservation of the species should also be considered at a regional level. For example, establishing

harvesting permits for isolated trees in pasturelands that consider minimum distance between trees to guarantee gene flow, might take into account that pollinators in Mexican populations apparently have lower flying distances than pollinators from Guatemalan populations, or that climate conditions and human dominated landscapes affect the local pollinators differently.

Climate change could also be considered as a possible menace for species, especially for populations in degraded landscapes with isolated trees. Changes in temperatures may affect the pollinator behavior, or in a worse scenario may increase fire frequencies that could affect directly isolated trees or pollinators.

5.4 Applications of *Cedrela odorata* microsatellite markers

For all loci paternity exclusion probabilities $Pr(Ex_1)$ and $Pr(Ex_2)$ were high, with a mean value of 0.62 and 0.76, respectively, indicating the potential of these microsatellites for future parentage analysis studies. As expected, in general $Pr(Ex_2)$ was higher than $Pr(Ex_1)$ since the probability of excluding the second parent increases when the first parent is known (Table 6). The successful application of microsatellite genotyping to paternity exclusion is based on the detection of high levels of genetic variation in *C. odorata*. The combination of different microsatellites could lead to a complete paternity exclusion, for example: an exclusion probability of 0.999 for the first parent could be obtained combining five microsatellites (C131, C95, C65, C61a, and C41), on the other hand, an exclusion probability of 0.999 for the second parent could be obtained combining four microsatellites (C131, C95, C65, and C61a).

Considering the work of cross-species amplification of SSR loci in the Meliaceae family by White and Powell (1997), where they found high levels of sequence conservation within the primer regions of Meliaceae species, the nine polymorphic microsatellites (Table 5) could be used with other closely related species like *Swietenia humilis*, *Swietenia macrophylla*, or *Carapa guianensis*.

6 CONCLUSIONS

1. Microsatellite markers developed for *C. odorata* were very informative and useful to determine the gene diversity and mating system estimates of the species.

2. High gene diversity estimates were found for *C. odorata* progeny arrays, although no significant differences were found between the northern and central lineages, neither the progeny arrays of isolated mother trees and the ones in groups or clusters.

3. *C. odorata* is an obligately outcrossed species, where no selfing occurrence was observed when adult trees were within a maximum distance of 500 meters.

4. Isolation level analysis revealed that although the species' outcrossing rates (t_m) were not significantly different from 1 at both isolation levels, significant differences were observed for correlation of paternity and single locus inbreeding coefficient of maternal parents, which was higher for isolated families.

5. It was estimated an average of five trees donated pollen to the isolated mother trees, while for the mother trees within a ratio of 100 meters, an average of 7.7 adult trees was found to be contributing with pollen.

6. Biparental inbreeding was found for both isolation levels, indicating that approximately 16% of the progenies resulted from mating between closely related parents at both isolation levels.

7. The analysis of the mating system at a family level revealed higher multi and single locus outcrossing rates within families from isolation level 3, which confirms less biparental inbreeding in natural forests and sites where *C. odorata* trees are more abundant. Differences among site environmental conditions, conservation status of the landscapes at each population, and relation between local pollinators and *C. odorata* trees might be favoring families of isolation 3.

8. *C. odorata* progenies from isolated trees have higher levels of biparental inbreeding, fewer pollen donors, potentially leading to a reduction of mechanisms for selective abortion among different pollen donors and leading to a decrease in vigor of progenies.

7 RECOMMENDATIONS

1. Further studies of genetic diversity and genetic differentiation with adult trees from natural populations of *C. odorata* should be assessed with the nine microsatellite markers developed for the present study.

2. Acknowledging the importance of isolated *C. odorata* trees in pastures, agriculture lands, and other human dominated landscapes, in the movement of pollinators in fragmented forests, harvesting permits of isolated trees in this type of landscapes should be more regulated and based on scientific and ecological information as the one generated in the present and latter studies.

3. Research on gene flow analysis of the species would be important to determine the minimum distance where pollinators could effective pollinate the isolated trees.

4. Further analysis with southern lineage families of *C. odorata* should be done to generate a complete understanding of the tree isolation and forest fragmentation implications on mating systems and genetic diversity of the species.

5. Variation of quantitative traits analyses of *C. odorata* at different isolation levels should be done to know if there is correlation between the microsatellite markers information and the quantitative data.

8 REFERENCES

Allard, RW. 1975. The mating system and microevolution. Genetics 79(suppl):115-126.

Asociación Costarricense para el Estudio de Especies Forestales Nativas. 1994. Encuentro Regional sobre Especies Forestales Nativas de la Zona Norte y Atlántica de Costa Rica. Taller de publicaciones ITCR. Memoria. Septiembre 24-25, 1992. Estación Biológica La Selva, Sarapiquí, CR. 86 p.

Balloux, F; Lugon-Moulin, N. 2002. The estimation of population differentiation with microsatellite markers. Molecular Ecology 11: 155-165..

Barret, SCH; Harder LD. 1996. Ecology and evolution of plant mating. Tree 11(2): 73-79.

Bawa, KS; Bullock, SH; Perry, DR; Covillerre, RE; Grayum, MH. 1995. Reproductive biology of tropical lowland rain forest trees. II. Pollination systems. American Journal of Botany 72: 346-356.

Boshier, DH; Chase, MR; Bawa, KS. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a Neotropical tree; mating system. American Journal of Botany 82: 476-483.

Boshier, DH. 2000. Mating systems. *In* Young, A; Boshier, D; Boyle, T. eds. Forest conservation genetics: principles and practice. Collingwood, AU. CSIRO: CABI. p. 63-79.

Cavers, S; Navarro, C; Lowe, AJ. 2003 a. A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. Conservation Genetics 4:571-580.

Cavers, S; Navarro, C; Lowe, AJ. 2003 b. Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L. (Meliaceae) in Mesoamerica. Molecular Ecology 12:1451-1460.

Cavers, S; Navarro, C; Lowe, AJ. 2004. Targeting genetic resource conservation in widespread species: a case study of *Cedrela odorata* L. Forest Ecology and Management. 197:285-294.

Chavarriaga, P; Maya, M; Bonierbale, M; Kresovich, S; Fregene, M; Tohme, J; Kochert, G. 1998. Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theoretical and Applied Genetics 97: 493-501.

Chase, M; Kessell, R; Bawa, K. 1996. Microsatellite markers for population and conservation genetics of tropical trees. American Journal of Botany 83(1): 51-57.

Cintron, BB. 1990. *Cedrela odorata* L. Cedro Hembra, Spanish Cedar, Meliaceae. Mahogany family. In: Silbéis of North America. Hardwods. Agric. Handbook, 654. Washington, DC. USDA. Vol. 2. p 250-257.

Ciofi, C; Funk, S; Coote, T; Cheeseman, D; Hammond, R; Saccher, I; Bruford, M. 1998. Genotyping with microsatellites markers. In: Karp A., Isaac P., Ingram D. Molecular Tools for Screening Biodiversity. Chapman and Hall, London, UK. p 195-201.

Clegg, MT. 1980. Measuring plant mating systems. Bioscience 30: 814-818.

Cloutier, D; Kanashiro, M; Ciampi, AY; Shoen, DJ. 2007. Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. Molecular ecology 16: 797-809.

Conner, JK; Hartl, DL. 2004. A primer of ecological genetics. Sinauer Associates, Inc. Sunderland, Massachusetts, U.S.A. 304 p.

Cordero, J; Boshier, DH. 2003. Árboles de Centroamérica: un manual para extensionistas. Oxford Forestry Institute, UK. CATIE, Turrialba, CR. 1077 p.

Cregan, P. 2002. Simple Sequence Repeat DNA Length Polymorphism (on line). USDA, Agricultural Research Service. Consulted 19 Aug. 2004. Available in http://www.nal.usda.gov/pgdic/Probe/v2n1/simple.html

Dayanandan, S; Rajora, OP; Bawa, KS. 1998. Isolation and characterisation of microsatellites in trembling aspen (*Populus tremuloides*) Theoretical and Applied Genetics 96: 950-956.

Ferreira, M; Grattapaglia, D. 1998. Introducción al Uso de Marcadores Moleculares en el Análisis Genético. BR, EMBRAPA-CENARGEN. 220 p.

Finkeldey, R; Ziehe, M. 2004. Genetic implications of silvicultural regimes. Forest Ecology and Management. 197: 231-244.

Frankham, R; Ballou, JD; Briscoe, DA. 2003. Introduction to conservation genetics. Cambridge, UK. Cambridge University Press. 617 p.

Frankham, R; Ballou, JD; Briscoe, DA. 2004. A primer of conservation genetics. Cambridge, UK. Cambridge University Press. 220 p.

García, F; Noyer, JL; Risterucci, AM; Chevallier, MH. 2004. Genotyping of mature trees of *Entandrophragma cylindricum* with microsatellites. Journal of Heredity. 95(5): 454-457.

Gillet, EM. 1999. Which DNA Marker for Which Purpose? Final compendium of the research project: Development, optimisation and validation of molecular tools for assessment of

biodiversity in forest trees. Consulted 10 Nov. 2005. Available in http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Gillies, ACM; Cornelius, JP; Newton, AC; Navarro, C; Hernández, M. 1997. Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L., assessed using RAPDs. Molecular Ecology. 6: 1133-1145.

Gregorius, HR. 2002. An integrative approach to modelling mating systems of tree populations. *In* Degen, B; Loveless, MD; Kremer, A. 2002. Modelling and Experimental Research of Genetic Processes in Tropical and Temperate Forests. 2002. Belém, BR. pp 42 -68.

Haberl, M; Tautz, D. 1999. Comparative allele sizing can produce inaccurate allele size differences for microsatellites. Molecular Ecology. 8: 1347-1350.

Hall, P; Orrell, LC; Bawa, KS. 1994. Genetic diversity and mating system in a tropical tree, *Carapa guianensis* (Meliaceae). American Journal of Botany 81(9):1104-1111.

Hall, P; Walker, S; Bawa, K. 1996. Effect of forest fragmentation on genetic diversity and mating system in a tropical tree, *Pithecellobium elegans*. Conservation Biology 10(3):757-768.

Hamada, H; Cetrino, M; Kakunaga, T. 1982. A novel repeated element with Z—DNA forming potential is widely found in evolutionarily diverse eukayotic genomes. Proc. Natl. Acad. Sci. 79: 6465-6469.

Hernández, G; Buonamici, A; Walker, K; Vendramin, GG; Navarro, C; Cavers, S. 2007. Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree. Technical Note, Conservation Genetics. 3 p.

Hilje, L; Cornelius, J. 2001. Es inmanejable *Hypsipyla grandella* como plaga forestal? Manejo Intentegrado de Plagas. Turrialba, CR. CATIE. 4 p. (Boletín Técnico no. 61).

Hoelzel, AR. 1998. Molecular Genetic Analysis of Populations, A Practical Approach. Oxford University, UK.

Holdridge, LR; Poveda LJ; Jiménez, Q. 1997. Árboles de Costa Rica. San José, CR. Centro Científico Tropical. v. 1, p. 269.

James, T; Vege, S; Aldrich, P; Hamrick, JL. 1998. Mating systems of three tropical dry forest species. Biotropica 30(4): 587-594.

Kattan, GH. 2002. Fragmentación: patrones y mecanismos de extinción de especies. *In* Guariguata, MR; Kattan, GH. eds. Ecología y conservación de bosques neotropicales. Cartago, CR. LUR. p. 561-590.

Lemes, MR; Brondani, RPV; Grattapaglia, D. 2002. Multiplexed Systems of Microsatellite Markers for Genetic Analysis of Mahogany, *Swietenia macrophylla* King (Meliaceae), a Threatened Neotropical Timber Species. The Journal of Heredity 93(4): 287-201.

Liengsiri, C; Boyle, TJB; Yeh, FC. 1998. Mating system in *Pterocarpus macrocarpus* Klus in Thailand. The American Genetic Association 89:216-221.

Litt, M; Luty, J. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle acting gene. Am. J. Hum. Genet. 44. 398-401.

Lowe, A; Goodall-Copestake, P; Caron, H; Kremer, A; Decroocq, S. 2002. A set of polymorphic microsatellites for *Vochysia ferruginea*, a promising tree for land reclamation in the Neotropics. Molecular Biology Notes 2: 153-155.

Lowe, AJ; Jourde, B; Breyne, P; Colpaert, N; Navarro, C; Wilson, J; Cavers, S. 2003. Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietenia macrophylla*). Heredity 90: 268-275.

Lowe, A; Harris, S; Ashton, P. 2004. Ecological genetics: design, analysis, and application. Malden, MA (EUA). Blackwell. 326 p.

Muluvi, GM; Sprent, JI; Odee, D; Powell, W. 2004. Estimates of outcrossing rates in *Moringa oleifera* using Amplified fragment length polymorphism (AFLP). African Journal of Biotechnology 3(2): 146-151.

Murawski, DA; Hamrick, JL. 1991. The effect of the density of flowering individuals on the mating systems of nine tropical tree species. Heredity 67: 167-174.

Myers, N; Mittermier, RA; Mittermier, CG; Da Fonseca, GAB; Kent, J. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853-858.

Nakamura, Y; Leppert, M; O'Nnell, P; Wolf, R; Holm, T; Culver, M; Martin, C; Fujimoto, E; Hoff, M; Kumlin, E; White, R. 1987. Variable number of tandem repeat (VNTR) marker for human gene mapping. Science 235: 1616-1622.

Namkoong, G; Boyle, T; El-Kassaby, Y; Eriksson, G; Gregorius, HR; Joly, H; Kremer, A; Savolainem, O; Wickneswari, R; Young, A; Zeh-Nlo, M; Prabhu R. 2002. Criteria and Indicators for Assessing the Sustainability of Forest Management: Conservation of Genetic Diversity. Forest Genetic Resources Working Papers, Working Paper FGR/37E. Forest Resources Development Service, Forest Resource Division, FAO, Rome. 25 p.

Nason, JD. 2002. La estructura genética de las poblaciones de árboles. *In* Guariguata, MR; Kattan, GH. eds. Ecología y conservación de bosques neotropicales. Cartago, CR. LUR. p. 299-327.

Navarro, C. 2002. Genetic resources of *Cedrela odorata* L. and their efficient use in Mesoamerica. Academic dissertation in forest tree breeding. Publication no. 12. Helsinki, FI. 112 p.

Navarro, C; Ward, S; Hernández, M. 2002. The tree *Cedrela odorata* (Meliaceae): A morphologically subdivided species in Costa Rica. Revista de Biología Tropical 50(1): 21-29.

Navarro, C; Montagnini, F; Hernández, G. 2004. Genetic variation of *Cedrela odorata* Linnaeus: results of early performance of provenance and families from Mesoamerica grown in association with coffee. Forest Ecology and Management 192: 217-227.

Navarro, C; Cavers, S; Lowe, A. 2008. Seed sourcing recommendations for forest restoration: tree isolation impacts progeny performance of Cedar and Mahogany in the Neotropics. Journal of sustainable forestry. (in press)

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America. 70: 3321-3323.

Novick, RR; Dick, CW; Lemes, MR; Navarro, C; Caccone, A; Bermingham, E. 2003. Genetic Structure of Mesoamerincan populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred form microsatellite analysis. Molecular Ecology 12: 2885-2893.

Ouborg, NJ; Piquot, Y; Groenendael, V. 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87: 551-568.

Patiño, F. 1997. Genetic resources of *Swietenia* and *Cedrela* in the Neotropics. Rome, Italy. United Nations Food and Agriculture Organization. 58 p.

Powell, W; Morgante, M; Andre, C; Hanafey, M; Voel, J; Tingey, S; Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2: 225-238.

Ratnam, W; Boyle, TJ. 2000. Effects of logging and other forms of harvesting on genetic diversity in humid tropical forests. In Young, A; Boshier, D; Boyle, T. eds. Forest conservation genetics: principles and practice. Collingwood, AU, CSIRO/CABI. p. 115-122.

Ritland, K. 1989. Correlated-matings in the partial selfer, *Mimulus guttatus*. Evolution 43: 848-859.

Ritland, K. 2002. Extentions of models for estimations of mating systems using *n* independent loci. Heredity 88: 221-228.

Rocha, OJ; Aguilar, G. 2001. Reproductive biology of the dry forest tree *Enterolobium cyclocarpum* (guanacaste) in Costa Rica: a comparison between trees left in pastures and trees in continuous forest. American Journal of Botany 88(9):1607-1614.

Roth Novick, R; Dick, C; Lemes, M; Navarro, C; Caccone, A; Bermingham, E. 2003. Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellites analysis. Molecular Ecology. 12: 2885-2893.

Schlötterer, C. 1998. Microsatellites. *In* Hoelzel, AR. Molecular Genetic Analysis of Populations, A Practical Approach. 2 ed. Oxford University, UK. sp.

Schoen, DJ; Clegg, MT. 1984. Estimation of mating system parameters when outcrossing events are correlated. Proc. Natl. Acad. Sci. 81: 5258-5262.

Schoen, DJ; Johnston, MO; L'Heureux, AM; Marsolais, JV. 1997. Evolutionary history of mating system in *Asinckia* (Boraginaceae) Evolution 51: 1090-1099.

Styles, BT; Khosla, PK. 1976. Cytology and reproductive biology of Melieaceae. *In*: Burley, J; Styles BT. (eds). Tropical trees: variation, breeding and conservation. Linnean Society Symposium. Series Number 2: 61-68.

Tautz, D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res. 17: 6463-6471.

Unjino, T; Kawahara, T; Tsumura ,Y; Nagamitsu, T; Yoshimuru, H; Ratnam, W. 1998. Development and polymorphism of simple sequence repeat DNA markers for *Shorea curtisii* and other Dipterocarpaceae species. Heredity 81: 422-428.

Van Oosterhout, C; Hutchinson, WF; Wills, DPM; Shipley, P. 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

Vinson, CC; Azevedo, VCR; Sampaio, I; Ciampi, AY. 2005. Developement of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. Molecular Ecology Notes 5: 33-34

Ward, M; Dick, CW; Gribel, R; Lemes, M; Caron, H; Lowe, AJ. 2005. To self, or not to self... A review of outcrossing and pollen-mediated gene flow in neotropical trees. Heredity 95:246-254

Weber, J; May, P. 1989. Abundant class of human DNA polymorphism, which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

White, G; Powell, W. 1997 (a). Isolation and characterization of microsatellites loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. Molecular Ecology. 6: 851-860.

White, G; Powell, W. 1997 (b). Cross-species amplification of SSR loci in the Meliaceae family. Molecular Ecology 6: 1195-1197.

Wilson, J; Lowe, A; Cavers, S; Navarro, C; Hernandez, M. Kremer, A; Caron, H; Labbe, P; Margis, R; Margis M; Breyne, P; Gribel, R; Lemes, M. 2001. Assessment of levels and dynamics of intra-specific genetic diversity of tropical trees. Final Scientific Report 1997-2001. European Commission. 76 p.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Wright, S. 1969. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago. US.

Wright, S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among populations. University of Chicago. US.

Yeh, FC. 2000. Population Genetics. *In* Young, A; Boshier, D; Boyle, T. eds. Forest conservation genetics: principles and practice. Collingwood, AU, CSIRO/CABI. p. 21-37.

Young, AG; Boyle, TJ. 2000. Forest Fragmentation. *In* Young, A; Boshier, D; Boyle, T. eds. Forest conservation genetics: principles and practice. Collingwood, AU, CSIRO/CABI. p. 123-134.

Zamora, N. 2000. Árboles de la Mosquitia Hondureña: descripción de 150 especies. Turrialba, CR. CATIE. p. 76-77.

9 ANNEXES

Annex 1. Technical note:

Conserv Genet
DOI 10.1007/s10592-007-9334-y

TECHNICAL NOTE

Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree

G. Hernández · A. Buonamici · K. Walker · G. G. Vendramin · C. Navarro · S. Cavers

Received: 3 April 2007 / Accepted: 15 April 2007

Abstract We describe 9 primers for amplification of microsatellite loci for the Neotropical tree *Cedrela odorata* L. (Meliaceae). Loci were isolated from an enriched library derived from a single DNA sample from a tree in Costa Rica. Levels of polymorphism were determined using samples from a large progeny trial. Across loci, the number of alleles ranged from 14 to 30. Observed heterozygosity levels ranged from 0.61 to 0.88. No linkage disequilibria were detected although some departures from Hardy-Weinberg equilibrium (HWE) were found, probably due to a Wahlund effect.

Keywords *Cedrela odorata* · Cedar · Meliaceae · Microsatellites

Cedrela odorata (Meliaceae), known as Cedro Amargo or Spanish Cedar, is a high-value species of the Mahogany family, widely distributed in the neotropics (Holdridge et al. 1997). Internationally valued for its high quality wood, *C. odorata* has been used for construction, furniture and boat building amongst other things (Cordero and Boshier 2003). As a result, heavy selective logging has severely reduced wild populations, to the point that the species is now threatened at a provenance level (Patiño 1997).

Previous studies have assessed genetic structure using AFLP and universal chloroplast DNA markers (Cavers et al. 2003a, b) and quantitative traits (Navarro et al. 2005), thoroughly describing genetic variation in the species at a landscape scale in Mesoamerica. Now, attention is turning to restoration efforts and best practice for sourcing seed. A major progeny trial is now underway in Costa Rica to examine the effects of landscape context on progeny fitness, as recent studies have shown that alteration of the forest surrounding seed trees changes gene flow patterns (e.g. Rocha and Aguilar 2001), with potentially significant consequences for inbreeding rates in progeny arrays. To enable analysis of mating system variation in these progeny arrays, a set of microsatellites were isolated for *C. odorata*. Loci were optimised for PCR and screened for polymorphism using 487 individuals in 68 families, from 12 populations distributed across Mesoamerica.

DNA was extracted from silica-gel-dried leaf material (DNeasy Plant mini kit, QIAGEN) from a single adult tree of *C. odorata* from Costa Rica. A microsatellite library enriched for di- (AG, GT, AT, GC) and trinucleotide (CAA, ATT, GCC) repeats was constructed following Edwards et al. (1996). The enriched DNA was cloned into pGEM-T vector and transformed into JM109 cells. A total of 200 clones were sequenced with an Amersham MegaBACE 1000 automated sequencer using DYEnamic ET Terminator Sequencing Kit (Amersham Biosciences).

Screening for polymorphism and optimisation of PCR conditions was carried out using leaf material from families from different, widely-separated provenances (Tulum and Xpujil, Mexico; Cedros, Honduras; Hojancha, Costa Rica) to maximise the potential for detection of polymorphism. All microsatellites were amplified using 25.0 ul PCR

G. Hernández · C. Navarro
Centro Agrónomico Tropical de Investigación y Enseñanza, Turrialba, Cartago 7170, Costa Rica

G. Hernández · K. Walker · S. Cavers (✉)
NERC Centre for Ecology and Hydrology, CEH Edinburgh, Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland, UK
e-mail: scav@ceh.ac.uk

A. Buonamici · G. G. Vendramin
Plant Genetics Institute, Consiglio Nazionale delle Ricerche (CNR), Sesto Fiorentino, Florence 50019, Italy

Table 1 Primer sequences, characterisation and basic descriptive statistics of nine microsatellite markers isolated from *Cedrela odorata*

Locus	Array	Primer sequences (5′–3′)	Allele size /bp	T (°C)	N	N_A	H_O	H_E	Pr (Ex_1)	Pr (Ex_2)	Null rate	GenBank accession no.
Ced2	$(GA)_{20}$	F:TTTGCTTTGA GAAACCTTGT*	130–170	55	400 (63)	19 (12)	0.850 (0.841)	0.882 (0.842)	0.614	0.762	0.017	EF413962
		R:AACTTTCGAATTG GTTAAGG										
Ced18	$(GA)_{23}$	F:CAAAGACCAAGA TTTGATGC*	130–150	55	403 (69)	19 (16)	0.797 (0.826)	0.844 (0.875)	0.545	0.708	0.029	EF413963
		R:ACTATGGGTGGCA CAACTAC										
Ced41	$(TC)_{18}$	F:TCATTCTTGGATCC TGCTAT*	120–160	55.5	451 (72)	21 (12)	0.745 (0.792)	0.910 (0.839)	0.693	0.819	0.100	EF413964
		R:GTGGGAAAGATTGT GAAGAA										
Ced44	$(TG)_{14}(AG)_{17}$	F:ACTCCATTAACTG CCATGAA*	180–240	55.5	456 (74)	30 (17)	0.882 (0.959)	0.931 (0.920)	0.755	0.860	0.028	EF413965
		R:ATTTTCATTCCCTTTTAGCC										
Ced54	$(GA)_{15}(AG)_6G(GA)_5$	F:GATCTCACCCACT TGAAAAA*	120–160	55	408 (70)	28 (21)	0.843 (0.886)	0.936 (0.931)	0.772	0.871	0.053	EF413966
		R:GCTCATATTTGAGA GGCATT										
Ced61a	$(TG)_{10}$	F:CAATCAAACCAAAA ATGGAT*	240–270	55.5	428 (71)	14 (10)	0.605 (0.535)	0.868 (0.843)	0.581	0.737	0.181	EF413967
		R:GCAAATTAACCAGAAA AACG										
Ced65	$(GA)_7(CA)_{14}$	F:GAGTGAGAAGAAGAAT CGTGATAGC*	160–200	55.5	468 (76)	17 (12)	0.618 (0.605)	0.802 (0.798)	0.456	0.632	0.136	EF413968
		R:GAGGTTCGATCAGG TCTTGG										
Ced95	$(CT)_{17}(AC)_{13}$	F:ATTTTCATTCCCTT TTAGCC*	190–250	55	389 (63)	28 (18)	0.918 (0.905)	0.829 (0.913)	0.715	0.834	0.043	EF413969
		R:TTATCATCTCCCTCA CTCCA										
Ced131	$(CT)_{16}$	F:CTCGTAATAATCCCAT TCCA*	80–120	55	419 (66)	14 (10)	0.745 (0.773)	0.829 (0.823)	0.507	0.677	0.053	EF413970
		R:GGAGATATTTTGGGGT TTT										
			Total cumulative exclusion probabilities						0.999	0.999		

* Indicates fluorescently labelled primer. Abbreviations are: number of individuals (*N*), number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities and exclusion probability Pr (Ex—for first parent $_1$ and second parent $_2$). Values in brackets for N, N_A, H_O, H_E are single-population estimates from the population Tikal

reactions consisting of the following: 2.0 ul template DNA, 15.4 ul H_2O, 2.0 ul primers, 2.5 ul 10× buffer, 0.5 ul dNTPs, 0.4 ul BSA, 0.2 ul Taq DNA polymerase (New England Biolabs). Reactions were run on a Hybaid MBS thermocycler to the following protocol: 1 min at 94°C, then 30 cycles of 60 s at 94°C, 60 s at 55°C, 60 s at 72°C and finally 5 min at 72°C. Optimal PCR reaction conditions for each of the polymorphic pairs were determined by testing a range of annealing temperatures (55.0–70.3°C) and different template DNA concentrations (1:5, 1:10, 1:20 and 1:1). In all cases the forward primer was labelled with either IRD 700 or 800 fluorescent label (MWG Biotech). PCR products were then separated on 6% polyacrylamide gel (25 cm), and visualized using a LI-COR 4200 IR2 automated genotyper. PCR products were run out alongside a microSTEP DNA size standard (Microzone Limited) and fragment sizes were scored using SAGA^{TM} software. Numbers of alleles were calculated, and observed and expected heterozygosity, exclusion probability (Cervus v3.0, Marshall et al. 1998) and null allele rate were estimated (Microchecker v2.2.3, van Oosterhout et al. 2004). Tests for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium, corrected for multiple comparisons and considering one individual per family, were carried out using Genepop v3.3 (Raymond and Rousset 1995). Results for effective numbers of alleles (N_A), observed (H_O) and expected heterozygosity (H_E) are reported (Table 1) for the collection as a whole and for a single population (Tikal) to demonstrate the within-population utility of the markers.

In 80% of the sequences a microsatellite motif was detected. However, about 130 sequences were discarded because of unfavourable properties for primer design. In 30 cases primers were designed, using PRIMER 3 software (Rozen and Skaletsky 2000). Of these, nine gave clear, interpretable band patterns and were polymorphic (Table 1). The numbers of alleles per locus ranged from 14 to 30 with levels of observed heterozygosity from 0.61 to 0.88. Null alleles were detected at rates of >0.05 at ced41, ced61a, ced65 and ced131. In all cases, these are most likely due to scoring errors due to stuttering, as indicated by low frequencies of heterozygote genotypes with size differences of a single repeat unit (van Oosterhout et al. 2004). Across the whole dataset, only loci ced2 and ced44 showed no departure from HWE. For some loci, this was contributed to by the presence of null alleles (highest null allele frequency estimated at locus ced61a with 18.1%, Table 1), but most probably reflects a heterozygote deficiency due to combination of samples from several, widely-distributed populations (Wahlund effect). No linkage disequilibria between loci ($P > 0.05$) were observed, suggesting that these loci should be valuable markers for population genetics and parentage analysis for *Cedrela odorata*.

Acknowledgements This work was part of the project SEEDSOURCE, funded by the EC under FP6 INCO (contract 003708). Library development, cloning and sequencing was done by CNR and optimisation at CEH by GH during his M.Sc., sponsored by CATIE and Ministerio de Ciencia y Tecnología.

References

Cavers S, Navarro C, Lowe AJ (2003a) A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. Conserv Genet 4:571–580

Cavers S, Navarro C, Lowe AJ (2003b) Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L. (Meliaceae) in Mesoamerica. Mol Ecol 12:1451–1460

Cordero J, Boshier DH (2003) Árboles de Centroamérica: un manual para extesionistas. Oxford Forestry Institute, UK, CATIE, Turrialba, CR

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758–760

Holdridge LR, Poveda LJ, Jiménez Q (1997) Árboles de Costa Rica. San José, CR, Centro Científico Tropical. vol 1, p 269

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Navarro C, Cavers S, Pappinen A, Tigerstedt P, Lowe AJ, Merilä J (2005) Contrasting quantitative traits and neutral genetic markers for genetic resource assessment of mesoamerican *Cedrela odorata*. Silvae Genet 54(6):281–292

Patiño F (1997) Genetic resources of *Swietenia* and *Cedrela* in the neotropics. United Nations Food and Agriculture Organization, Rome, Italy

Raymond M, Rousset F (1995) genepop (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rocha OJ, Aguilar G (2001). Variation in the breeding behaviour of the dry forest tree *Enterolobium cyclocarpum* (Guanacaste) in Costa Rica. Am J Bot 88:1600–1606

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molec Ecol Notes 4:535–538

Annex 2. Total of 26 microsatellite markers isolated from Cedrela odorata and tested for amplification and polymorphism.

Locus	Allele Size Range (bp)	Channel	Polymorphic/Monomorphic	Ease scoring	Multiplexing
Ced2	130-170	700	Polymorphic	yes	Ced131a+Ced2
Ced4	200-220	700	Polymorphic	yes	
Ces15	140-160	800	Monomorphic	yes	
Ced18	130-150	700	Polymorphic	yes	
Ced22	120-140	700	Polymorphic	yes	
Ced26	140-160	700	Polymorphic	no	
Ced27	200-240	700	Polymorphic	yes	
Ced36	160-200	800	Polymorphic	no	
Ced41a	180-200	700	Monomorphic	yes	
Ced41	120-160	700	Polymorphic	yes	Ced41+Ced44
Ced44	180-240	700	Polymorphic	yes	Ced41+Ced44
Ced48	180-220	700	Polymorphic	no	
Ced50	120-140	700	Polymorphic	no	
Ced54	180-230	800	Polymorphic	yes	
Ced56	100-120	800	Monomorphic	no	
Ced61a	240-270	800	Polymorphic	yes	Ced61a+Ced65
Ced61b	180-240	800	Polymorphic	yes	
Ced64	160-180	800	Polymorphic	yes	
Ced65	160-200	800	Polymorphic	no	Ced61a+Ced65
Ced72	140-160	800	Polymorphic	no	
Ced80	240-260	800	Monomorphic	yes	
Ced89	100-140	800	Monomorphic	yes	
Ced95	190-250	800	Polymorphic	yes	
Ced120	120-180	800	Polymorphic	no	
Ced131	80-120	700	Polymorphic	yes	
Ced131a	70-90	700	Polymorphic	yes	Ced131a+Ced2

Grey rows represent the 10 loci chosen for the genetic and mating system study with *Cedrela odorata* populations.

Annex 3. Genotypic disequilibrium test for all data including complete families (half sib individuals) of C. odorata. P-value for each locus pair across all the populations (Fisher method).

Locus pair			Chi Square	df	P-value
Ced131	-	Ced95	45.435	24	0.005
Ced131	-	Ced65	40.565	22	0.009
Ced95	-	Ced65	32.776	22	0.065
Ced131	-	Ced61a	Infinity	24	Highly significant
Ced95	-	Ced61a	Infinity	22	Highly significant
Ced65	-	Ced61a	Infinity	22	Highly significant
Ced131	-	Ced41	Infinity	24	Highly significant
Ced95	-	Ced41	Infinity	24	Highly significant
Ced65	-	Ced41	Infinity	22	Highly significant
Ced61a	-	Ced41	Infinity	24	Highly significant
Ced131	-	Ced44	54.493	22	0
Ced95	-	Ced44	Infinity	22	Highly significant
Ced65	-	Ced44	Infinity	22	Highly significant
Ced61a	-	Ced44	Infinity	22	Highly significant
Ced41	-	Ced44	Infinity	22	Highly significant
Ced131	-	Ced18	Infinity	22	Highly significant
Ced95	-	Ced18	Infinity	22	Highly significant
Ced65	-	Ced18	Infinity	22	Highly significant
Ced61a	-	Ced18	Infinity	22	Highly significant
Ced41	-	Ced18	44.505	22	0.003
Ced44	-	Ced18	Infinity	22	Highly significant
Ced131	-	Ced54	Infinity	24	Highly significant
Ced95	-	Ced54	Infinity	24	Highly significant
Ced65	-	Ced54	Infinity	22	Highly significant
Ced61a	-	Ced54	Infinity	24	Highly significant
Ced41	-	Ced54	37.312	24	0.041
Ced44	-	Ced54	Infinity	22	Highly significant
Ced18	-	Ced54	46.609	22	0.002
Ced131	-	Ced2	52.377	24	0.001
Ced95	-	Ced2	Infinity	22	Highly significant
Ced65	-	Ced2	23.02	22	0.401
Ced61a	-	Ced2	Infinity	24	Highly significant
Ced41	-	Ced2	Infinity	24	Highly significant
Ced44	-	Ced2	Infinity	22	Highly significant
Ced18	-	Ced2	39.676	22	0.012
Ced54	-	Ced2	28.613	24	0.235

The null hypothesis (H_0) is: "*Genotypes at one locus are independent from genotypes at the other locus*". H_0 is rejected for small p-values, usually < 0.05.

Annex 4. Summary by locus of tests for Hardy-Weinberg Equilibrium for all populations.

Summary by locus for Xpujil

Locus	ChiSquare	DF	Prob	
Ced131	21.495	28	0.804	ns
Ced95	30.564	36	0.725	ns
Ced65	27.921	21	0.142	ns
Ced61a	50.156	21	0.000	***
Ced41	41.279	28	0.051	ns
Ced44	72.483	55	0.057	ns
Ced18	17.447	15	0.293	ns
Ced54	50.600	45	0.262	ns
Ced2	32.486	28	0.255	ns

Summary by locus for Tulum

Locus	ChiSquare	DF	Prob	
Ced131	39.123	36	0.331	ns
Ced95	53.511	66	0.865	ns
Ced65	42.615	28	0.038	*
Ced61a	33.915	28	0.204	ns
Ced41	47.972	78	0.997	ns
Ced44	87.166	66	0.042	*
Ced18	49.587	66	0.934	ns
Ced54	126.451	120	0.326	ns
Ced2	49.312	36	0.069	ns

Summary by locus for Cedros

Locus	ChiSquare	DF	Prob	
Ced131	60.663	28	0.000	***
Ced95	80.534	105	0.964	ns
Ced65	36.509	15	0.001	**
Ced61a	34.088	21	0.035	*
Ced41	56.711	78	0.967	ns
Ced44	45.612	105	1.000	ns
Ced18	10.092	36	1.000	ns
Ced54	94.552	136	0.997	ns
Ced2	35.791	45	0.835	ns

Summary by locus for Tikal

Locus	ChiSquare	DF	Prob	
Ced131	52.729	45	0.200	ns
Ced95	142.863	120	0.076	ns
Ced65	160.797	66	0.000	***
Ced61a	117.051	36	0.000	***
Ced41	65.298	55	0.161	ns
Ced44	151.357	136	0.174	ns
Ced18	118.085	120	0.532	ns
Ced54	264.509	231	0.064	ns
Ced2	67.797	66	0.416	ns

Summary by locus for Reforma

Locus	ChiSquare	DF	Prob	
Ced131	12.833	10	0.233	ns
Ced95	2.880	10	0.984	ns
Ced65	19.556	15	0.190	ns
Ced61a	0.426	3	0.935	ns
Ced41	6.415	6	0.378	ns
Ced44	7.531	15	0.941	ns
Ced18	2.064	6	0.914	ns
Ced54	15.000	28	0.978	ns
Ced2	6.528	15	0.969	ns

Summary by locus for La Paz

Locus	ChiSquare	DF	Prob	
Ced131	22.060	36	0.967	ns
Ced95	108.336	78	0.013	*
Ced65	37.718	28	0.104	ns
Ced61a	73.566	45	0.005	**
Ced41	80.434	78	0.403	ns
Ced44	107.898	91	0.109	ns
Ced18	40.098	36	0.293	ns
Ced54	125.205	105	0.087	ns
Ced2	146.399	66	0.000	***

Summary by locus for Meambar

Locus	ChiSquare	DF	Prob	
Ced131	9.078	15	0.873	ns
Ced95	59.852	66	0.689	ns
Ced65	12.566	6	0.050	ns
Ced61a	42.502	28	0.039	*
Ced41	50.639	55	0.642	ns
Ced44	94.506	66	0.012	*
Ced18	29.013	28	0.412	ns
Ced54	60.864	55	0.273	ns
Ced2	35.833	36	0.476	ns

Summary by locus for Cañas

Locus	ChiSquare	DF	Prob	
Ced131	1.345	3	0.719	ns
Ced95	37.250	36	0.411	ns
Ced65	9.366	21	0.986	ns
Ced61a	12.195	6	0.058	ns
Ced41	43.902	36	0.172	ns
Ced44	27.977	36	0.828	ns
Ced18	14.750	21	0.835	ns
Ced54	6.949	10	0.730	ns
Ced2	23.063	15	0.083	ns

Summary by locus for Tres Garantias

Locus	ChiSquare	DF	Prob	
Ced131	28.030	21	0.139	ns
Ced95	35.542	21	0.025	*
Ced65	29.423	15	0.014	*
Ced61a	45.770	21	0.001	**
Ced41	55.200	21	0.000	***
Ced44	58.407	55	0.351	ns
Ced18	51.276	55	0.618	ns
Ced54	33.861	21	0.038	*
Ced2	10.874	15	0.762	ns

Summary by locus for Taulabe

Locus	ChiSquare	DF	Prob	
Ced131	33.210	36	0.602	ns
Ced95	77.446	78	0.496	ns
Ced65	40.639	28	0.058	ns
Ced61a	110.154	66	0.001	***
Ced41	94.819	120	0.957	ns
Ced44	136.213	153	0.831	ns
Ced18	47.116	45	0.386	ns
Ced54	67.829	91	0.967	ns
Ced2	38.470	36	0.358	ns

Summary by locus for Los Esclavos

Locus	ChiSquare	DF	Prob	
Ced131	56.202	21	0.000	***
Ced95	127.599	120	0.300	ns
Ced65	48.032	36	0.087	ns
Ced61a	56.430	28	0.001	**
Ced41	93.076	66	0.016	*
Ced44	212.463	153	0.001	**
Ced18	55.380	45	0.138	ns
Ced54	184.282	136	0.004	**
Ced2	54.764	45	0.151	ns

Summary by locus for Hojancha

Locus	ChiSquare	DF	Prob	
Ced131	9.889	15	0.827	ns
Ced95	64.276	66	0.537	ns
Ced65	32.863	36	0.619	ns
Ced61a	52.499	28	0.003	**
Ced41	57.977	45	0.093	ns
Ced44	87.018	91	0.599	ns
Ced18	61.447	55	0.256	ns
Ced54	23.570	28	0.704	ns
Ced2	79.333	66	0.126	ns

Annex 5. Singles and multi locus outcrossing rates for Cedrela odorata families.

	ts		*tm*			
Family	**Mean**	**SE**	**Mean**	**SE**	**Progenies**	**Isolation level**
Xpujil_1	0.74	0.14	1.00	0.00	6	Isolation level 1
Xpujil_2	0.84	0.06	1.00	0.00	8	Isolation level 1
Xpujil_3	0.59	0.22	1.00	0.00	9	Isolation level 1
Reforma_4	0.64	0.15	0.83	0.16	10	Isolation level 1
Tres Garantias_5	0.38	0.15	0.68	0.16	8	Isolation level 1
Tres Garantias_6	0.26	0.06	0.92	0.09	9	Isolation level 1
Tres Garantias_7	1.03	0.06	1.00	0.04	8	Isolation level 1
Tulum_8	0.84	0.17	0.83	0.16	6	Isolation level 1
Tulum_9	0.90	0.07	1.00	0.05	8	Isolation level 1
Tulum_10	0.90	0.12	0.90	0.10	9	Isolation level 1
Tulum_11	1.14	0.17	1.14	0.16	2	Isolation level 1
Tulum_12	0.82	0.13	0.90	0.10	10	Isolation level 3
Tulum_13	0.86	0.15	0.85	0.15	6	Isolation level 1
La Paz_14	0.95	0.07	1.00	0.06	7	Isolation level 1
La Paz_15	0.65	0.11	1.00	0.03	8	Isolation level 1
La Paz_16	0.34	0.19	1.00	0.03	6	Isolation level 1
La Paz_17	0.90	0.06	1.00	0.02	8	Isolation level 1
La Paz_18	0.26	0.27	1.00	0.02	9	Isolation level 1
La Paz_19	1.01	0.07	1.00	0.02	8	Isolation level 1
La Paz_60	0.87	0.11	1.00	0.03	4	Isolation level 1
Taulabe_20	0.79	0.09	1.00	0.02	9	Isolation level 1
Taulabe_21	0.77	0.08	1.00	0.01	10	Isolation level 1
Taulabe_22	0.49	0.11	1.00	0.02	8	Isolation level 1
Taulabe_23	0.98	0.07	1.00	0.02	7	Isolation level 1
Taulabe_24	0.92	0.09	1.03	0.03	8	Isolation level 1
Taulabe_25	0.90	0.09	1.00	0.03	5	Isolation level 1
Taulabe_61	0.81	0.18	1.01	0.06	3	Isolation level 1
Cedros_26	0.89	0.09	1.02	0.03	10	Isolation level 3
Cedros_27	0.78	0.14	1.00	0.03	6	Isolation level 3
Cedros_28	0.88	0.11	1.02	0.03	8	Isolation level 3
Cedros_29	0.61	0.10	1.00	0.02	7	Isolation level 3
Cedros_30	0.73	0.10	1.00	0.03	5	Isolation level 3
Cedros_31	0.84	0.12	1.01	0.03	7	Isolation level 3
Cedros_32	0.65	0.11	1.01	0.02	10	Isolation level 3
Cedros_33	0.70	0.16	0.93	0.09	10	Isolation level 3
Cedros_34	0.61	0.15	1.00	0.03	9	Isolation level 3
Meambar_35	0.65	0.19	0.73	0.26	7	Isolation level 3
Meambar_36	0.85	0.14	1.00	0.06	8	Isolation level 3
Meambar_37	0.35	0.08	0.93	0.08	9	Isolation level 3
Meambar_62	0.74	0.19	1.00	0.06	5	Isolation level 1
Los Esclavos_38	0.70	0.20	0.90	0.19	8	Isolation level 1
Los Esclavos_39	0.49	0.14	0.88	0.14	6	Isolation level 1
Los Esclavos_40	0.91	0.15	0.90	0.15	8	Isolation level 1
Los Esclavos_41	0.58	0.22	0.75	0.24	4	Isolation level 1
Los Esclavos_42	0.79	0.13	0.88	0.13	8	Isolation level 1
Los Esclavos_43	0.80	0.08	1.00	0.05	8	Isolation level 1
Los Esclavos_45	0.83	0.15	1.00	0.03	7	Isolation level 1
Los Esclavos_57	0.54	0.12	1.00	0.04	5	Isolation level 1
Los Esclavos_58	0.62	0.16	0.80	0.18	5	Isolation level 1
Los Esclavos_59	0.82	0.18	1.00	0.14	3	Isolation level 1
Tikal_46	1.02	0.07	1.00	0.04	10	Isolation level 3
Tikal_47	0.55	0.33	1.01	0.03	6	Isolation level 3
Tikal_48	0.86	0.09	1.00	0.02	9	Isolation level 3
Tikal_49	0.56	0.14	0.94	0.08	7	Isolation level 3
Tikal_50	0.84	0.11	1.01	0.04	7	Isolation level 3
Tikal_51	0.58	0.12	1.00	0.04	5	Isolation level 3
Tikal_52	0.86	0.13	0.89	0.11	7	Isolation level 3
Tikal_53	0.96	0.15	1.02	0.07	5	Isolation level 3
Tikal_54	0.76	0.12	0.89	0.11	9	Isolation level 3
Tikal_55	0.72	0.19	1.00	0.05	7	Isolation level 3
Tikal_56	0.68	0.09	1.00	0.04	5	Isolation level 3
Cañas_63	0.39	0.13	0.74	0.22	3	Isolation level 1
Cañas_64	0.82	0.12	1.00	0.12	4	Isolation level 1
Cañas_65	0.67	0.18	1.00	0.09	4	Isolation level 1
Hojancha_66	0.73	0.18	0.85	0.15	5	Isolation level 1
Hojancha_67	0.88	0.15	1.01	0.10	4	Isolation level 1
Hojancha_68	0.70	0.16	1.02	0.06	5	Isolation level 1
Hojancha_69	0.78	0.09	0.97	0.04	8	Isolation level 1
Hojancha_70	1.00	0.06	1.00	0.04	5	Isolation level 1