



Tesis Doctoral

**“Estudio del rol del flujo génico y de la deriva genética en
la determinación de la estructura de las poblaciones
fragmentadas de *Anadenanthera colubrina* var *cebil*
(Fabaceae, Fabales)”**

Lic. María Eugenia Barrandeguy

Directora: Dra. María Victoria García

Codirector: Dr. Rolando Rivera Pomar

Lugar de realización: Cátedra de Genética de Poblaciones y Cuantitativa. Departamento de Genética. Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones.

Doctorado en Ciencias Naturales
Facultad de Ciencias Naturales y Museo
Universidad Nacional de La Plata

Febrero de 2013 - La Plata



Tesis Doctoral

**“Estudio del rol del flujo génico y de la deriva genética en
la determinación de la estructura de las poblaciones
fragmentadas de *Anadenanthera colubrina* var *cebil*
(Fabaceae, Fabales)”**

Lic. María Eugenia Barrandeguy

a
mis padres
Diego y Susana

Agradecimientos

- A la Universidad pública, en especial a la Universidad Nacional de Misiones y a la Universidad Nacional de la Plata las cuales me han permitido acceder a mi formación de grado y de postgrado de manera gratuita.
- Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y al Comité Ejecutivo de Desarrollo e Innovación Productiva (CEDIT) de la Provincia de Misiones porque gracias a la modalidad de becas cofinanciadas me permitieron la realización de todas las instancias de la carrera de Doctorado. Un especial reconocimiento al presidente del CEDIT, Dr. Carlos Schvezov, por su visión al adherir a la modalidad de becas cofinanciadas fortaleciendo la formación de recursos humanos en la provincia de Misiones.
- Al Centro de Intercambio Académico Alemán (DAAD) por otorgarme una beca para la realización de una estadía de corta duración.
- A mis directores la Dra. María Victoria García y el Dr. Rolando Rivera Pomar por haberme dado la oportunidad de formarme junto a ellos.
- A mi directora, María Victoria, en primer lugar por haber confiado en mí y desde entonces, por brindarme su conocimiento y transmitirme la pasión por este trabajo enfrentándome día a día con nuevos desafíos.
- Al Profesor Reiner Finkeldey y a la Dra. Kathleen Prinz por haberme abierto las puertas del Instituto de Genética y Mejoramiento Forestal de la Universidad de Göttingen y por todo el apoyo recibido.
- Al Departamento de Genética, en especial a la Dra. Carina Argüelles, por permitirme el acceso al equipamiento necesario para la obtención de parte de los resultados.
- Al personal de la Reserva Horco Molle (Tucumán) y del Parque Nacional Calilegua (Jujuy) por su colaboración en la toma de muestras.
- A María Edith Barrandeguy por haber colaborado con el grupo de trabajo y transmitirme, oportunamente, la importancia de obtener el grado académico de Doctor.
- A las integrantes de la cátedra de Genética de Poblaciones y Cuantitativa: Elena, Alejandra y Verónica, por los momentos y logros compartidos, valorando la importancia de trabajar en equipo.
- Un especial agradecimiento al Dr. Alberto Fenocchio y a la Mgter. Cristina Pastori, profesores de la Licenciatura en Genética, por su constante preocupación e interés.
- A mis amigas de siempre Alejandra y Sole por estar siempre y por todos los momentos inolvidables compartidos. A Chunxia, Isolde y Silvina por su ayuda y compañía en momentos importantes durante esta etapa.
- A mis padres, Diego y Susana, por su apoyo incondicional. A Papá, un especial agradecimiento por su participación en los viajes de campaña y por ser nuestro “calificado asistente de campo”. A mis hermanos, Julia, María José y Ramiro por ser mi familia y a pesar de las distancias estar presentes.
- A Javier por haberme acompañado desde el principio y en esta última etapa desde más cerca. A la familia de Javier por su apoyo y comprensión.

INDICE DE CONTENIDOS

RESUMEN	xi
SUMMARY	xiii
SECCION I	1
1. Introducción	2
1.1 Importancia de la variabilidad genética en las poblaciones naturales.....	2
1.2 Fuerzas que moldean la variabilidad genética: Estructuración genética.....	3
1.3 Consecuencias genéticas de la fragmentación de hábitat.....	5
1.4 Contrarrestando la fragmentación: Flujo génico mediado por polen y semillas.....	7
1.5 Medidas y modelos para el estudio del flujo génico.....	8
1.6 <i>Anadenanthera colubrina</i> var. <i>cebil</i> una especie paradigmática de los bosques tropicales estacionalmente secos.....	10
1.6.1 Sistema de estudio: poblaciones naturales de <i>Anadenanthera colubrina</i> var. <i>cebil</i>	10
1.6.2 Consecuencias del cambio climático histórico sobre la distribución de <i>A. colubrina</i> var. <i>cebil</i>	18
2. Hipótesis y Objetivos Generales	21
2.1 Hipótesis.....	21
2.2 Objetivos generales.....	21
3. Fundamentación de la metodología utilizada para el logro de los objetivos	22
4. Material biológico: Area de estudio y metodología	24
SECCION II	28
1. Introducción	29
1.1 Flujo génico mediado por polen y semillas en especies forestales.....	29
1.2 El genoma nuclear vegetal y los loci microsatélites.....	30
1.3 Estudios genético-poblacionales en especies forestales.....	33
1.4 Estimación del índice de fijación (F_{ST}).....	34
2. Objetivos específicos	36
3. Materiales y Métodos	37
3.1 Desarrollo de SSR específicos para <i>Anadenanthera colubrina</i> var. <i>cebil</i>	37
3.2 Análisis de la utilidad de los marcadores desarrollados para estudios genético-poblacionales.....	40
3.3 Genotipificación de los individuos empleando SSRs.....	41
3.4 Análisis estadístico de los datos.....	41
3.4.1 Caracterización de la diversidad genética nuclear.....	41
3.4.2 Análisis de las relaciones genéticas entre los genotipos multilocus.....	42

3.4.3 Análisis de la estructura genética de las poblaciones.....	43
3.4.4 Cuantificación del flujo génico.....	48
4. Resultados.....	51
4.1 Desarrollo de SSR específicos para <i>Anadenanthera colubrina</i> var. <i>cebil</i>	51
4.2 Utilidad de los nuevos marcadores desarrollados para estudios genético - poblacionales....	51
4.3 Análisis estadístico de los datos.....	57
4.3.1 Caracterización de la diversidad genética nuclear.....	57
4.3.2. Análisis de las relaciones entre los genotipos multilocus.....	61
4.3.3 Análisis de la estructura genética de las poblaciones.....	61
4.3.4 Cuantificación de los niveles de flujo génico.....	68
5. Discusión.....	71
6. Conclusiones.....	80
SECCION III.....	81
1. Introducción.....	82
1.1 Importancia del flujo génico mediado por semillas en especies forestales.....	82
1.2 ADN citoplasmático: el genoma cloroplástico.....	82
1.3 Marcadores microsatélites de ADN cloroplástico.....	84
1.4 Utilidad de los cpSSRs para estudios genético poblacionales.....	85
2. Objetivos específicos.....	86
3. Materiales y Métodos.....	87
3.1 Genotipificación de los individuos empleando cpSSRs.....	87
3.2 Análisis estadístico de los datos.....	87
3.2.1 Caracterización de la diversidad genética cloroplástica.....	87
3.2.2 Análisis de las relaciones genéticas entre haplotipos.....	88
3.2.3 Análisis de la estructura genética cloroplástica de las poblaciones.....	89
3.2.4 Cuantificación del flujo génico histórico mediado por semillas.....	90
4. Resultados.....	92
4.1 Genotipificación de los individuos empleando cpSSRs.....	92
4.2 Análisis estadístico de los datos.....	92
4.2.1 Caracterización de la diversidad genética cloroplástica.....	92
4.2.2 Análisis de las relaciones genéticas entre los haplotipos.....	92
4.2.3 Análisis de la estructura genética poblacional.....	97
4.2.4 Cuantificación del flujo génico histórico mediado por semillas.....	102
5. Discusión.....	104
6. Conclusiones.....	111

SECCION IV	112
1. Introducción	113
<i>1.1 Marcadores moleculares nucleares y cloroplásticos para inferir los niveles de flujo génico mediado por polen y por semillas</i>	113
<i>1.2 Importancia relativa del flujo génico mediado por polen y por semillas</i>	114
<i>1.2.1 Modelo de islas</i>	114
2. Objetivos específicos	116
3. Materiales y Métodos	117
<i>3.1 Niveles de diversidad y estructuración genética nuclear y cloroplástica</i>	117
<i>3.2 Importancia relativa del flujo génico mediado por el polen y por las semillas</i>	117
<i>3.2.1 Modelo de Islas</i>	117
4. Resultados	118
<i>4.1 Comparación de los niveles de diversidad y estructuración genética nuclear y cloroplástica</i> ..	118
<i>4.2 Importancia relativa del flujo génico mediado por polen y por semillas</i>	118
<i>4.2.1 Modelo de Islas</i>	118
5. Discusión	121
6. Conclusiones	128
Proyecciones	129
BIBLIOGRAFIA	130

LISTA DE FIGURAS

SECCION I

Figura I-1: Distribución de <i>Anadenanthera colubrina</i> var. <i>cebil</i> en Argentina.....	11
Figura I-2 Características morfológicas de <i>Anadenanthera colubrina</i> (Vell.) Brenan var. <i>cebil</i> Griseb. Altschul.....	14
Figura I-3: Distribución del Bosque Tropical Estacional en Sudamérica.....	15
Figura I-4: Distribución de ejemplares de <i>A. colubrina</i> var. <i>cebil</i> y <i>A. colubrina</i> var. <i>colubrina</i> según el estudio realizado por Prado y Gibbs (1993).....	17
Figura I-5: Posible distribución histórica de los SDTF obtenida mediante modelos de paleodistribución.....	20
Figura I-6: Imágenes satelitales de los sitios considerados en este estudio.....	25

SECCION II

Figura II-1: Esquema del proceso de desarrollo de los marcadores SSRs específicos para <i>Anadenanthera colubrina</i> var. <i>cebil</i>	39
Figura II-2: Resultados de las diferentes etapas del desarrollo de los marcadores microsátélites en <i>A. colubrina</i> var. <i>cebil</i>	52
Figura II-3: Alelos revelados por los ocho marcadores SSRs polimórficos específicos para <i>A. colubrina</i> var. <i>cebil</i> identificados mediante un secuenciador analizador de fragmentos.....	54
Figura II-4: Frecuencias alélicas por locus en las regiones estudiadas.....	58
Figura II-5: Relaciones entre los individuos de <i>A. colubrina</i> var. <i>cebil</i> a partir de la información de sus genotipos multilocus.....	62
Figura II-6: Representación gráfica de los valores de Δk para los diferentes valores de k	66
Figura II-7: Determinación de la estructura genética poblacional mediante inferencia Bayesiana.....	67
Figura II-8: Determinación de la subestructuración genética dentro de cada región mediante inferencia Bayesiana.....	69

SECCION III

Figura III-1: Alelos revelados por los cuatro marcadores <i>Ccmp</i> en las poblaciones de <i>A. colubrina</i> var. <i>cebil</i> identificados mediante un secuenciador analizador de fragmentos.....	93
Figura III-2: Red de haplotipos.....	98
Figura III-3: Análisis bayesiano de la estructura genética poblacional.....	100
Figura III-4: Relaciones entre los <i>clusters</i> definidos por el análisis bayesiano de la estructura genética sin información espacial <i>a priori</i>	103

LISTA DE TABLAS

SECCION I

Tabla I-1: Descripción de las poblaciones de <i>A. colubrina</i> var. <i>cebil</i> analizadas.....	27
---	----

SECCION II

Tabla II-1: Efectividad en el proceso de desarrollo de microsatélites específicos para <i>A. colubrina</i> var. <i>cebil</i>	53
---	----

Tabla II-2: Caracterización de los nueve marcadores SSRs desarrollados para <i>A. colubrina</i> var. <i>cebil</i> (n= 69 individuos).....	55
--	----

Tabla II-3: Número de alelos, Heterocigosidad observada (H_O) y Heterocigosidad esperada (H_E) por locus en las poblaciones estudiadas.....	56
--	----

Tabla II-4: Caracterización de la diversidad genética revelada por los marcadores SSRs en las poblaciones estudiadas.....	59
--	----

Tabla II-5: Resultados de las pruebas de <i>t</i> pareadas entre los parámetros de diversidad genética..	60
---	----

Tabla II-6: Análisis de la varianza molecular (AMOVA).....	63
---	----

Tabla II-7: Estimaciones del índice de fijación F_{ST} mediante diferentes definiciones y estimación del índice de diferenciación D de Jost.....	65
---	----

SECCION III

Tabla III-1: Caracterización de los <i>loci</i> cpSSRs en <i>Anadenanthera colubrina</i> var. <i>cebil</i>	94
---	----

Tabla III-2: Descripción y frecuencias de los seis haplotipos cloroplásticos presentes en las poblaciones de <i>Anadenanthera colubrina</i> var. <i>cebil</i> analizadas.....	95
--	----

Tabla III-3: Caracterización de la diversidad genética revelada por los marcadores cpSSRs en las poblaciones estudiadas.....	96
---	----

Tabla III-4: Análisis de la varianza molecular (AMOVA).....	99
--	----

Tabla III-5: Análisis bayesiano de la estructura genética poblacional.....	101
---	-----

SECCION IV

Tabla IV-1: Comparación de los niveles de diversidad genética nuclear y cloroplástica.....	119
---	-----

Tabla IV-2: Distribución de la diversidad genética nuclear y cloroplástica en poblaciones naturales de <i>A. colubrina</i> var. <i>cebil</i>	120
---	-----

RESUMEN

Flujo génico es un término amplio que incluye todos los mecanismos que resultan en el movimiento de genes desde una población a otra. En especies vegetales, el flujo génico se realiza por medio del movimiento del polen y de las semillas. La reproducción es el proceso que transfiere un arreglo alélico a la siguiente generación, mientras que la dispersión del polen y de las semillas determinan los patrones de distribución de los genes dentro y entre las poblaciones vegetales. Se entiende como estructura genética poblacional a la cantidad de variabilidad genética y a su distribución dentro y entre poblaciones locales y entre los individuos dentro de una especie. La estructura genética resulta de la acción conjunta del flujo génico, mutación, migración, selección natural y deriva genética, los cuales operan dentro de un contexto histórico y biológico de cada especie vegetal. La distribución de la variación genética, refleja las respuestas de las especies a las fuerzas microevolutivas que se encuentran operando en tiempos presentes y pasados, pudiendo brindar información acerca de la manera en que las especies han evolucionado y la manera en la cual podrían continuar evolucionando en el futuro.

Anadenanthera colubrina var. *cebil* constituye un recurso forestal nativo Sudamericano el cual presenta una distribución discontinua en Argentina, restringiéndose su presencia a las provincias biogeográficas Paranaense y de las Yungas. La hipótesis sobre la cual se sustentó el presente trabajo sostiene que el flujo génico homogeneiza las frecuencias génicas entre las poblaciones contrarrestando los efectos de la deriva genética provocados por la fragmentación. Los objetivos generales de este trabajo fueron: indagar acerca del rol evolutivo del flujo génico y de la deriva genética como fuerzas modeladoras de las frecuencias génicas en las poblaciones fragmentadas y determinar la importancia del movimiento de los alelos a través de la dispersión de las semillas y del polen. Se emplearon marcadores moleculares selectivamente neutros de herencia biparental, microsatélites nucleares, y de herencia uniparental, microsatélites cloroplásticos.

Se estudiaron 69 individuos de *A. colubrina* var. *cebil* pertenecientes a cuatro poblaciones del Norte argentino; (1) Candelaria (n=20), (2) Santa Ana (n=16), (3) Tucumán (n=14) y (4) Jujuy (n=19); mediante el empleo de ocho marcadores microsatélites nucleares y cuatro marcadores microsatélites cloroplásticos. Se estimaron los niveles de diversidad genética, estructuración genética (AMOVA, F_{ST} , análisis basados en distancias, análisis bayesiano) y se estimó el nivel de flujo génico de manera indirecta. Estos análisis fueron realizados para ambos marcadores. Además, se analizó la proporción relativa de los niveles de flujo génico mediado por el polen y por las semillas a partir de los índices de fijación de ambos genomas.

A nivel del genoma nuclear, las poblaciones de *A. colubrina* var. *cebil*, contuvieron niveles elevados de diversidad genética ($H_{Eb}=0,756-0,841$), mientras que presentaron estructuración genética moderada ($F_{STb}=0,111$). Los individuos fueron asignados a dos *clusters* según su origen regional. El nivel de flujo génico histórico entre las poblaciones contrarrestó los

efectos de la deriva genética ($N_e m_b \approx 2$). A nivel del genoma cloroplástico las poblaciones contuvieron niveles bajos de diversidad genética ($H_m = 0,000-0,500$) y presentaron estructuración genética elevada ($F_{STm} = 0,949$). Los individuos fueron asignados a seis *clusters* denotando estructuración poblacional. El nivel de flujo génico histórico entre las poblaciones no fue suficiente para contrarrestar los efectos de la deriva genética ($Nm_m = 0,027$). Por su parte, la proporción del nivel de flujo génico mediado por polen superó en 144 veces al nivel de flujo génico mediado por semillas.

Los efectos de la fragmentación histórica sobre los Bosques tropicales estacionalmente secos se reflejarían en la estructuración genética detectada a nivel global entre las regiones en ambos genomas. El impacto antrópico reciente en la región Paranaense sería responsable de la menor diversidad y de la mayor estructuración genética de sus poblaciones a nivel del genoma nuclear. Por su parte, los niveles superiores de flujo génico entre las poblaciones de las Yungas habrían mantenido mayor diversidad y menor estructuración genética de sus poblaciones a nivel del genoma nuclear. La presencia de haplotipos ancestrales y de diversidad genética antigua en la región Paranaense explicarían la mayor diversidad y la menor estructuración genética de sus poblaciones a nivel del genoma cloroplástico. La presencia de diferentes variantes haplotípicas fijadas en cada población de la región de las Yungas, debido a la ausencia de flujo génico mediado por semillas, indicaría la acción de la deriva genética a nivel de genoma cloroplástico. La hipótesis de trabajo no se rechaza al considerarse el rol del flujo génico y de la deriva genética a nivel del genoma nuclear mientras que, a nivel del genoma cloroplástico, la deriva genética ha jugado el rol principal.

SUMMARY

Gene flow is a collective term that includes all the mechanisms that originate gene movements from one population to another. In plants, gene flow is mediated by pollen and seeds. The reproduction is the process that exchanges an allele arrangement to next generation while pollen and seeds dispersion determine postreproductive patterns of gene distribution within and among plant populations. Population genetic structure is the magnitude of genetic variability and its distribution within and among local populations and, also, among individuals in the same species. Genetic structure is a consequence of joint action of gene flow, mutation, migration, natural selection and genetic drift, which operate on historical and biological context in each plant species. Distribution of genetic variation shows the answers of the species to microevolutive forces that operate at this time and during the past, giving information regarding the mode in that the species evolve and how they could evolve in the future.

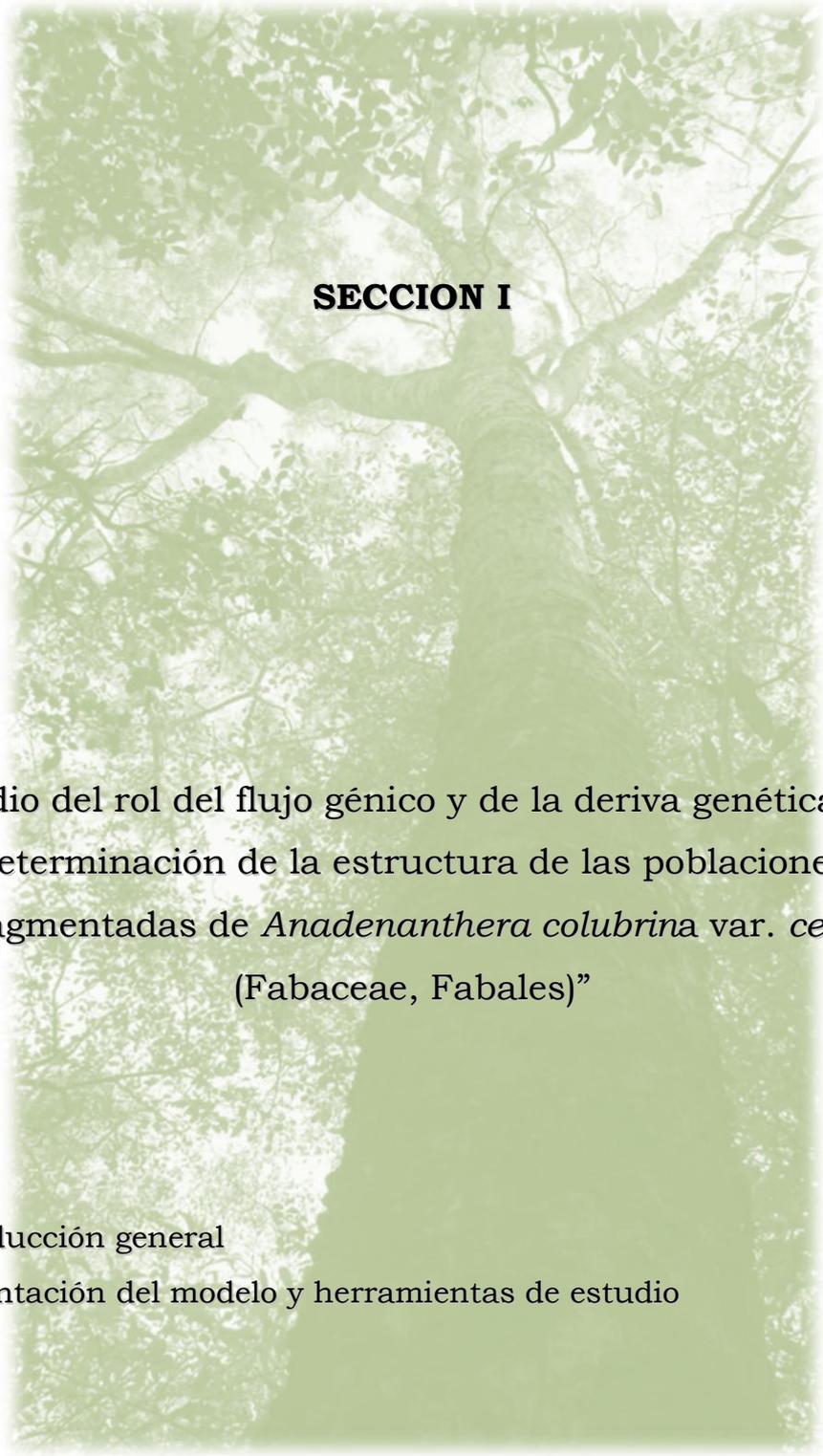
Anadenanthera colubrina var. *cebil* is a native forest resource from South America. It presents a discontinue distribution in Argentina and it is distributed in Paranaense and Yungas biogeographic provinces. The hypothesis of this work establishes that gene flow makes uniform the gene frequencies among population and, in this way it counteracts the consequences of genetic drift. Main aims were investigate about the evolutive role of gene flow and genetic drift as shaping forces of gene frequencies in fragmented populations and determine the importance of allele movement through the dispersion of seeds and pollen. Biparental inherited neutral molecular markers as nuclear microsatellites and uniparental inherited neutral molecular markers as chloroplastic microsatellites were used.

Sixty nine individuals of *A. colubrina* var. *cebil* belong to four population from the North of Argentina; (1) Candelaria (n=20), (2) Santa Ana (n=16), (3) Tucumán (n=14) y (4) Jujuy (n=19); were studied using eight nuclear microsatellites and four chloroplast microsatellites. Genetic diversity and genetic structure were evaluated (AMOVA, F_{ST} , distance and bayesian analysis) and indirect estimation of gene flow was calculated. All these analyses were designed for both markers. Also, the relative proportion of gene flow by pollen and by seeds was analyzed.

The populations of *A. colubrina* var. *cebil* showed high levels of genetic diversity ($H_{Eb}=0.756-0.841$) at nuclear genome level while that showed moderate genetic structure ($F_{STb}=0.11$). The individuals were assigned to two clusters regarding its regional origin. The level of historical gene flow between populations counteracted the effects of genetic drift ($N_e m_b \approx 2$). On the other hand, the populations showed low level of chloroplast genetic diversity ($H_m=0.000-0.500$) and showed high genetic structure ($F_{STm}=0.95$). The individuals were assigned to six clusters indicating population structure. Historical gene flow between

populations was not enough to counteract the effects of genetic drift ($Nm_m=0.027$). Also, the proportion of gene flow by pollen exceeded 144 times gene flow by seeds.

The effects of historical fragmentation on Seasonally dry tropical forest could be reflected in the hierarchical genetic structure detected between regions in both genomes at global level. Deforestation and recent fragmentation could be responsible of the low genetic diversity and high genetic structure to genome level in populations from Paranaense region. On the other hand, high levels of gene flow between Yungas populations could maintain the high genetic diversity and the low genetic structure in these populations at nuclear genome level. The presence of ancestral haplotype and ancient genetic diversity could explain the high genetic diversity and low genetic structure in the chloroplastic genome in populations from Paranaense region. The presence of different fixed haplotypes in each population from Yungas could indicate the action of genetic drift at level of chloroplast genome. The hypothesis of this work could not be rejected regarding the role of gene flow and genetic drift at nuclear genome level while that genetic drift has played the main role at chloroplast genome level.



SECCION I

“Estudio del rol del flujo génico y de la deriva genética en la determinación de la estructura de las poblaciones fragmentadas de *Anadenanthera colubrina* var. *cebil* (Fabaceae, Fabales)”

- Introducción general
- Presentación del modelo y herramientas de estudio

1. Introducción

Los procesos biológicos en su mayoría están ligados al movimiento y sus causas operan a múltiples escalas espaciales y temporales. A nivel de comunidad el movimiento de individuos de diferentes especies nos permite conocer acerca de procesos ecológicos. A nivel de población, los patrones de movimiento de un conjunto de individuos nos permiten conocer acerca del papel de procesos demográficos tales como competencia, organización social, migración y colonización, entre otros. A nivel de individuo, el estudio del movimiento proporciona conocimiento sobre el ciclo vital, comportamiento e interacción con factores bióticos y abióticos. Por último, a nivel de los genes, el movimiento es importante para la evolución y la diversificación de la vida dado que el proceso evolutivo consiste, en última instancia, en la variación espacial y temporal de las frecuencias de las diferentes variantes génicas (Robledo-Arnuncio y Gonzalez Martínez, 2009).

El movimiento de los genes dentro y entre las poblaciones naturales y su interacción con la deriva genética, la mutación, y la selección natural determinan la composición genética de una población, incluyendo esto su diversidad y estructura genética. Los marcadores moleculares permiten tomar una instantánea en un tiempo dado de esa composición para ciertos *loci* y, a partir de los métodos y teorías de la genética de poblaciones, podremos cuantificar esa variabilidad, estudiar y analizar los mecanismos que la generan y la mantienen.

1.1 Importancia de la variabilidad genética en las poblaciones naturales

La evolución ha sido definida como los cambios temporales en la composición genética de las poblaciones (Dobzhansky, 1973), siendo la variación genética un prerequisite para este proceso. Es así que, el desafío empírico central en genética de poblaciones ha sido desde siempre la medición de la variación genética bajo la razón fundamental de que esta cuantificación ayudará a revelar de que manera operan los procesos microevolutivos, tales como deriva genética, mutación, migración y selección (Avice, 2004). Luego, la estructuración poblacional determinará los patrones y la cantidad de variabilidad genética que se encuentra disponible dentro de una especie (Templeton, 2006). Los patrones de variación genética, a nivel del paisaje, reflejan las respuestas de las especies a las fuerzas microevolutivas que se encuentran operando en tiempos presentes y pasados, pudiendo brindar información acerca de la manera en que las especies han evolucionado y la manera en la cual podrían continuar evolucionando en el futuro (White *et al.*, 2007).

Durante la primera mitad del siglo XX hubo dos corrientes científicas con miradas opuestas acerca de las magnitudes de la variación genética en la naturaleza. Por un lado, se ubicó la escuela clásica que sostenía que la variación genética era baja y por otro lado estaban quienes proponían que existía un balance en la variación genética, con lo cual la mayoría de los *loci* eran polimórficos (Avice, 2004). Los descubrimientos de la existencia de grandes niveles de

variación molecular no aportaron a la resolución de la cuestión y surgió una explicación alternativa para explicar la variación molecular detectada, la denominada teoría de mutación “neutral”, la cual ha asumido un rol prominente en la genética de poblaciones impactando hasta los tiempos actuales (Avice, 2004). Bajo la teoría neutral, los alelos alternativos no confieren efectos diferenciales en la eficacia biológica de sus portadores. A un nivel intraespecífico, la teoría neutral predice que el polimorfismo molecular es mantenido por un balance entre el aporte realizado por la mutación y la pérdida al azar provocada por la deriva genética (Kimura, 1991).

La mayoría de las especies forestales poseen niveles considerables de variación genética, principalmente dentro de sus poblaciones, siendo la heterocigosidad esperada aproximadamente un 50% superior en una población de árboles forestales que la heterocigosidad esperada promedio en poblaciones de plantas anuales y de plantas perennes con ciclo de vida corto (White *et al.*, 2007). Se ha descrito un número de factores que contribuyen a los elevados niveles de diversidad genética típicamente detectada en poblaciones de árboles forestales, entre ellos se pueden nombrar el gran tamaño poblacional, la longevidad, los elevados niveles de fecundación cruzada y los elevados niveles de flujo génico entre poblaciones (White *et al.*, 2007).

Mediante el empleo de marcadores moleculares selectivamente neutros, es posible cuantificar características de la variabilidad genética a nivel de genes dentro y entre poblaciones, mediante el empleo de diferentes *loci* marcadores (Gillet *et al.*, 2005; White *et al.*, 2007). Se han propuesto diferentes niveles para medir la variación genética intraespecífica, los cuales se organizan en mediciones de la variación genética dentro de poblaciones locales (conocida como diversidad genética), en mediciones de las diferencias genéticas entre pares de poblaciones locales y en mediciones de la variación genética entre varias poblaciones locales (Gillet *et al.*, 2005).

1.2 Fuerzas que moldean la variabilidad genética: Estructuración genética

Se entiende como estructura genética poblacional a la cantidad de variabilidad genética y a su distribución dentro y entre poblaciones locales y entre los individuos dentro de una especie (Templeton, 2006). Esta definición enfatiza los patrones espaciales de la variación genética que emergen desde las reglas del intercambio gamético lo cual determina una distribución no aleatoria de la variación genética entre individuos (Sato *et al.*, 2006). Las unidades fundamentales para estudiar la variación genética son las poblaciones (Finkeldey y Hattemer, 2007), es decir, grupos de organismos de la misma especie que habitan dentro de un área geográfica suficientemente restringida como para que cualquier miembro pueda, potencialmente, aparearse con cualquier otro del sexo opuesto (Hartl y Clark, 2007). La estructura genética resulta de la acción conjunta del flujo génico, mutación, migración,

selección natural y deriva genética, los cuales operan dentro de un contexto histórico y biológico de cada especie vegetal (Heuertz *et al.*, 2003; Loveless y Hamrick, 1984).

En plantas, la distribución de los alelos en el espacio está influenciada por el flujo génico a través de las semillas y del polen, de la deriva genética, la selección natural en sus diferentes formas, la divergencia mutacional y la recombinación genética (Ouborg *et al.*, 1999).

Flujo génico es un término amplio que incluye todos los mecanismos que resultan en el movimiento de genes desde una población a otra (Slatkin, 1985). La reproducción es el proceso que transfiere un arreglo alélico a la siguiente generación, mientras que la dispersión del polen y de las semillas determinan los patrones de distribución de los genes dentro y entre las poblaciones vegetales (Loveless y Hamrick, 1984). De esta manera, puede decirse que los patrones de variabilidad genotípica (homocigosis vs heterocigosis) entre individuos dentro de una población local es altamente dependiente del sistema de apareamiento, mientras que la distribución de la variación alélica dentro y entre poblaciones locales es influenciada por el flujo génico y la deriva genética (Templeton, 2006).

Los patrones de subdivisión geográfica y los niveles de flujo génico entre poblaciones locales en una especie son importantes para comprender la evolución genética (Slatkin y Barton, 1989). El flujo génico es la mayor fuerza evolutiva homogeneizadora de la variación genética a través de las poblaciones mediante el movimiento de alelos dentro y entre poblaciones (Dutech *et al.*, 2005). Sin embargo, el flujo génico será una fuerza evolutiva siempre y cuando este ocurra entre poblaciones genéticamente diferentes (Templeton, 2006). Así, el flujo génico puede actuar como una fuerza que mantiene integrada a la especie además de influir en procesos ecológicos, entre ellos determinar la persistencia y la adaptación de las poblaciones locales, las tasas de extinción de las especies y la evolución de los patrones de distribución de las especies (Planter, 2007).

Para alelos neutrales, cuando el flujo génico se interrumpe, la deriva genética es una fuerza mucho más efectiva para producir diferenciación genética entre las poblaciones en comparación con la mutación (Slatkin y Barton, 1989). Podemos resumir los efectos del flujo génico sobre la variabilidad genética diciendo que el flujo génico reduce la variabilidad genética entre poblaciones y la aumenta dentro de una población local (Templeton, 2006) mientras que la deriva genética causa un incremento de la variabilidad genética entre poblaciones, debido a la divergencia de las frecuencias alélicas, y reduce la variabilidad genética dentro de una población, debido a la pérdida o fijación de alelos. Así, los efectos del flujo génico sobre la variabilidad dentro y entre poblaciones son opuestos a los efectos de la deriva genética (Templeton, 2006). De esta manera, los estudios de flujo génico cobran relevancia para la interpretación de los patrones microevolutivos y la estructura genética de las poblaciones (Bossart y Prowel, 1998).

Dentro de las poblaciones de árboles, el flujo génico es considerado uno de los factores fundamentales para mantener el elevado nivel de diversidad genética (Burczyk *et al.*, 2004). Para especies forestales, se ha informado que la dispersión del polen es a menudo el mayor contribuyente al flujo génico y que la estructura genética espacial es principalmente el resultado de dispersión de semilla limitada (Gonzalez-Martínez *et al.*, 2002).

1.3 Consecuencias genéticas de la fragmentación de hábitat

La fragmentación del hábitat implica la transformación de hábitats continuos en varios fragmentos remanentes espacialmente aislados, lo cual representa una seria amenaza para el mantenimiento de la biodiversidad en muchos ecosistemas terrestres (Young *et al.*, 1996).

Las predicciones acerca de las consecuencias genéticas de la fragmentación del hábitat se relacionan con la reducción del tamaño poblacional y con el incremento del aislamiento espacial de las poblaciones (Young *et al.*, 1996). Estos cambios poblacionales producen erosión de la variación genética e incrementan la divergencia interpoblacional como consecuencia de la deriva genética, incremento de la endogamia, reducción de los niveles de flujo génico y extinción de las subpoblaciones (Young *et al.*, 1996). La reducción del tamaño de los fragmentos, el incremento del aislamiento entre ellos, la pérdida de hábitat y la reducción de la densidad forestal generalmente producen como consecuencia cuellos de botella, incremento de la deriva genética, incremento de la endogamia, flujo génico reducido y efecto fundador (Andrianoelina *et al.*, 2009). Además, debido a que los efectos de la deriva genética dependen del tiempo, se esperan menores efectos de la deriva genética en aquellas especies cuyas poblaciones se encuentran aisladas por pocas generaciones (Young *et al.*, 1996).

Las poblaciones fragmentadas se encuentran en peligro de perder alelos debido a la acción de la deriva genética, siendo esas pérdidas difíciles de compensar por el ingreso de nuevos alelos a través del flujo génico mediado por polen. Como consecuencia del aislamiento se incrementa la diferenciación genética entre las poblaciones remanentes debido a los cambios no adaptativos de su estructura genética (Finkeldey y Hattemer, 2007).

En las plantas, las consecuencias genéticas de la fragmentación adquieren relevancia debido a que son organismos sésiles, poseen diferencias interespecíficas en cuanto a su longevidad, tiempo generacional y abundancia pre-fragmentación. Además, presentan diferencias en sus sistemas reproductivos, poseen la capacidad de intercambiar genes por medio del polen y de las semillas, son capaces de almacenar su material genético en las semillas y experimentan diferentes interacciones con los animales polinizadores y vectores involucrados en la dispersión los cuales a su vez, pueden verse afectados por la fragmentación (Young *et al.*, 1996). Dependiendo de los patrones de distribución de las especies, la fragmentación del hábitat puede interrumpir el rango de una especie y no afectar a otras (Finkeldey y Hattemer, 2007).

Características tales como la longevidad de los individuos, la elevada diversidad dentro de las poblaciones y las elevadas tasas potenciales de flujo génico mediado por polen confieren a las especies forestales la capacidad de hacer frente a las consecuencias negativas de la fragmentación (Andrianoelina *et al.*, 2009). Las respuestas de las especies forestales tropicales dependen de factores especie-específicos y de factores generales que pueden influenciar los patrones reproductivos a largo plazo y la conectividad genética entre las poblaciones. Entre estos factores, la abundancia de las especies y los mecanismos de dispersión de las semillas y del polen se destacan dada su importancia para la supervivencia de estas especies en hábitats disturbados (Nason y Hamrick, 1997).

La pérdida de grandes masas forestales se ha convertido en un problema de preocupación mundial (Finkeldey y Hattermer, 2007). La deforestación generalmente no es completa, resultando en un mosaico de parches de forestación remanente embebidos en un paisaje en el cual los procesos bióticos y abióticos han sido alterados por las actividades humanas (Nason y Hamrick, 1997). Por otra parte, la presión ejercida por las poblaciones humanas sobre los remanentes boscosos genera una mayor reducción de la densidad poblacional. Esta situación plantea el interrogante de conocer si los parches remanentes constituyen poblaciones en sentido genético asumiendo que la reducción de la diversidad genética en poblaciones pequeñas es más un síntoma de riesgo que su causa (Finkeldey y Hattermer, 2007).

Los disturbios en los bosques inciden en la reproducción de sus especies alterando su composición, como así también el comportamiento de los polinizadores y agentes dispersores de semillas. La fragmentación, además de afectar directamente a las poblaciones vegetales, puede afectar a los agentes dispersores ya sea por la pérdida completa de los mismos, debido a cambios en su comportamiento o a modificaciones en la composición de las asociaciones (Aldrich y Hamrick, 1998). La alteración del comportamiento de los polinizadores podría disminuir la fecundidad de las especies que dependen de ellos como vectores. Sin embargo, existen casos en los cuales los disturbios ambientales incrementan la actividad de los polinizadores promoviendo la fecundidad y el flujo génico (Dick *et al.*, 2003). Estudios genéticos revelaron que el polen de especies forestales tropicales puede dispersarse en hábitats fragmentados (Aldrich y Hamrick, 1998), con lo cual la hipótesis que sostiene que la fragmentación del hábitat puede conducir al incremento de los niveles de flujo génico interpoblacional no debe ser descartada (Young *et al.*, 1996). Muchos de los estudios de flujo génico realizados en especies forestales neotropicales que habitan ambientes disturbados han detectado niveles inesperadamente elevados de flujo génico (Dick *et al.*, 2003).

La reproducción sexual de las plantas con flores se encuentra negativamente afectada por la fragmentación del hábitat. Las especies autoincompatibles son alógamas obligadas, lo cual las hace altamente dependientes de los polinizadores para reproducirse y como

consecuencia susceptibles a las alteraciones producidas por la fragmentación al afectar esta a sus polinizadores (Aguilar *et al.*, 2006). Además, se han reportado cambios en los sistemas de apareamiento en especies autocompatibles que habitan hábitats fragmentados, aumentando los niveles de autofecundación (Aldrich y Hamrick, 1998).

1.4 Contrarrestando la fragmentación: Flujo génico mediado por polen y semillas

En especies vegetales, el flujo génico se realiza por medio del movimiento del polen y de las semillas desde una población local a otra (White *et al.*, 2007). El polen es transferido al óvulo en la flor femenina por acción del viento o por medio de vectores tales como insectos, aves o mamíferos. Las semillas son luego dispersadas desde la madre por medio de animales, por acción de la gravedad o del viento (Avise, 2004). El sistema de apareamiento es decisivo para la fusión de las gametas masculinas y femeninas. El flujo génico a través del polen es un prerrequisito para el apareamiento entre plantas de fecundación cruzada. Es por ello que el flujo génico y el sistema de apareamiento están estrechamente relacionados y son los dos factores más importantes del comportamiento reproductivo de las especies vegetales. El sistema de apareamiento determina la combinación de los alelos en los genotipos si bien no causa modificaciones en las frecuencias alélicas (Finkeldey y Hattmer, 2007). Las relaciones genéticas entre la progenie, la heterogeneidad genética espacial dentro de las poblaciones y la distribución de la diversidad genética dentro y entre las poblaciones son parámetros que se encuentran directamente influenciados por los sistemas de apareamiento. Generalmente, las especies forestales tropicales poseen sistemas de apareamiento que estimulan o aseguran la fecundación cruzada (Hamrick y Murawski, 1990).

Características propias de las especies vegetales, tales como la longitud del período de vida, el sistema de apareamiento y el tamaño de los individuos, influyen los patrones de diferenciación, tanto en la escala espacial como temporal. Debido a su gran tamaño y a su tendencia a la fecundación cruzada, las especies forestales experimentan elevados niveles de flujo génico (Linhart y Grant, 1996). Aquellas especies cuya dispersión de polen o de semillas ocurre en una escala superior a 1 km, presentan estructuras poblacionales de mayor homogeneidad a escala de paisaje, debido a las mayores distancias recorridas por sus dispersores (Linhart y Grant, 1996).

El empleo de marcadores genéticos de herencia materna y biparental en angiospermas permite diferenciar las contribuciones históricas del movimiento de las semillas y del polen sobre los niveles de flujo génico. Esta información es relevante para distinguir entre las consecuencias genéticas de la colonización mediante semillas del intercambio de genes por medio del polen entre poblaciones establecidas (Sork *et al.*, 1999).

1.5 Medidas y modelos para el estudio del flujo génico

Los métodos indirectos para estimar los niveles de flujo génico emplean las frecuencias alélicas presentes en muestras tomadas desde diferentes poblaciones (Slatkin, 1985) y se basan principalmente en el análisis de la distribución de los alelos en las poblaciones para inferir acerca de los niveles o patrones de flujo génico entre ellas (Planter, 2007). Son indirectos porque el flujo génico es inferido desde sus consecuencias sobre las frecuencias genotípicas empleando modelos genéticos poblacionales (Broquet y Petit, 2009). El punto de partida para la mayoría de los métodos indirectos es una lista de frecuencias alélicas para diferentes *loci* (Slatkin y Barton, 1989).

Por su parte, los métodos directos, los cuales parten de observaciones de la dispersión de individuos o gametas (Slatkin y Barton, 1989) o de análisis de parentesco, cuantifican el patrón de movimiento de los genes mediante la determinación de parentales o el origen poblacional de los individuos (Planter, 2007; Broquet y Petit, 2009). Sin embargo, las medidas directas de dispersión no siempre pueden ser consideradas como estimadoras de flujo génico ya que miden la migración en un área determinada, la cual no siempre puede estar conteniendo a una población y en un período dado el cual no siempre representa una escala evolutiva, sumado a que la dispersión no necesariamente refleja el movimiento de los genes (Planter, 2007).

El flujo génico puede ser cuantificado mediante el parámetro m , el cual describe el movimiento de cada gameta o individuo independientemente del tamaño poblacional (Slatkin y Barton, 1989). Sin embargo, debido a que el flujo génico requiere tanto movimiento como reproducción, m no es simplemente una medida de la dispersión de individuos o gametas entre poblaciones locales, sino que m representa una compleja interacción entre los patrones de dispersión y el sistema de apareamiento (Templeton, 2006).

Bajo el modelo de islas de Wright (1931, 1951), el cual asume que un gran número de subpoblaciones de igual tamaño intercambian individuos o gametas entre ellas de manera constante, m indica la probabilidad de que cada gameta sea un inmigrante (Slatkin y Barton, 1989). El modelo de islas es central en genética de poblaciones ya que permite la derivación de ecuaciones analíticas para los efectos del flujo génico y otras fuerzas evolutivas sobre la dinámica de la varianza genética (Broquet y Petit, 2009).

Las consecuencias evolutivas del flujo génico pueden ser cuantificadas mediante el número efectivo de migrantes, $N_e m$ (Wright, 1931). Este estimador permite cuantificar el flujo génico en términos del número de gametas o de individuos que se mueven e históricamente fue el primer parámetro que pudo ser estimado desde datos moleculares (Broquet y Petit, 2009). Bajo el modelo de islas Nm permite determinar si la acción de la deriva genética por sí sola puede producir diferenciación genética sustancial entre poblaciones; si Nm es mayor a uno entonces el flujo génico es suficiente para prevenir diferenciación sustancial por acción de la deriva genética (Slatkin y Barton, 1989). Sin embargo, aunque Nm no supere a la unidad aún

existe variación entre las poblaciones y la acción de la deriva genética no llevará a la fijación de alelos alternativos (Slatkin y Barton, 1989). Así, el principal desafío no es la estimación de Nm sino la comprensión de la estructura poblacional real dado que las poblaciones naturales no son estables en su demografía ni en sus patrones de migración. Las variaciones en los niveles de flujo génico existente, generación tras generación, pueden llevar a pensar al flujo génico como un proceso estocástico. Sin embargo, los cambios en las condiciones ambientales experimentadas por las poblaciones a lo largo de largas escalas de tiempo, podrían indicar que los procesos no son estacionarios, lo cual implicaría que las estimaciones promedio a largo plazo no sean constantes (Slatkin y Barton, 1989).

El número efectivo de migrantes es aún un método popular para la estimación del flujo génico desde datos genéticos debido a su sencilla estimación a pesar de múltiples problemas asociados a ella (Freeland, 2005). El principal cuestionamiento surge del modelo de islas en el cual se encuentra sustentada su estimación ya que el mismo depende de supuestos biológicamente poco probables, como el que sostiene que una proporción m de gametas son extraídas de cada población y distribuidas al azar en todas las poblaciones locales restantes, sin tener en consideración la localización relativa de las mismas (Templeton, 2006). Además, el modelo de islas asume que no actúa la selección ni la mutación y que cada población persiste indefinidamente en equilibrio de migración y deriva genética (Whitlock y McCauley, 1999). Las poblaciones son capaces de alcanzar y permanecer en equilibrio migración – deriva genética únicamente cuando el tamaño efectivo y las tasas de migración permanecen más o menos constantes. Es por ello que este equilibrio es perdido como consecuencia de diversos procesos demográficos tales como expansión de rango reciente, fragmentación del hábitat y cuellos de botella poblacionales (Freeland, 2005).

Debido a que las estimaciones indirectas del flujo génico se basan en la relación matemática entre la estructura genética y la tasa de flujo génico, se asume implícitamente que las propiedades ecológicas de las poblaciones desde las cuales se han obtenido los datos genéticos coinciden con los supuestos poco realistas del modelo teórico sobre el cual se sustenta esta relación matemática. Sin embargo, Whitlock y McCauley (1999) sugirieron que las mediciones de la estructura genética son valiosas por sí mismas y que las transformaciones de estas mediciones a estimaciones cuantitativas de flujo génico suelen ser innecesarias y erróneas.

Para evitar esta distorsión, existen métodos alternativos que permiten la asignación de los individuos a grupos o *clusters* empleando sus genotipos multilocus y, de esta manera, pueden ser empleados para agrupar localidades en unidades panmícticas sobre la base de sus perfiles alélicos y de su geografía (Beerli y Palczewski, 2010). Así, estas aproximaciones permiten identificar migrantes a partir de genotipos multilocus sin emplear modelos que incluyan supuestos de equilibrio migración – deriva genética ya que brindan información acerca de la migración reciente a partir del desequilibrio observado en los genotipos multilocus de los

migrantes o de los descendientes recientes de ellos. En comparación con las estimaciones indirectas del flujo génico histórico estos métodos incluyen pocos supuestos aunque son informativos acerca de los patrones recientes de migración. Sin embargo, ambas aproximaciones son complementarias ya que proveen información acerca del flujo génico a diferentes escalas de tiempo (Wilson y Rannala, 2003).

1.6 Anadenanthera colubrina var. cebil una especie paradigmática de los bosques tropicales estacionalmente secos

1.6.1 Sistema de estudio: poblaciones naturales de Anadenanthera colubrina var. cebil

Anadenanthera colubrina (Vell.) Brenan var. *cebil* (Griseb) Altschul pertenece al orden Fabales, familia Fabaceae, subfamilia Mimosoideae, tribu Mimoseae (Cialdella, 2000). Es una especie forestal nativa de América del Sur y se la conoce vulgarmente como curupay, cebil colorado o angico.

El género *Anadenanthera* se halla descrito por la combinación de los siguientes caracteres morfológicos: inflorescencia globosa, dehiscencia de las vainas a lo largo de una única sutura, semillas casi circulares no aladas que carecen de endosperma y distribución estrictamente americana (Altschul, 1964). Dentro del género *Anadenanthera* se reconocen dos especies: *A. peregrina* y *A. colubrina*, las cuales se diferencian por unos pocos caracteres morfológicos diagnósticos basados principalmente en la textura de las vainas. La distribución geográfica está parcialmente solapada entre ambas especies, siendo la distribución de *A. peregrina* más estrecha y corresponde a un rango desde el sudeste de Brasil hasta las Antillas Mayores, mientras que el patrón de distribución de *A. colubrina* es más amplio, extendiéndose desde la Caatinga hacia el Noroeste y el Suroeste de Brasil, en confluencia con los ríos Paraná y Paraguay; en el Sudoeste de Bolivia y Noroeste de Argentina; y esporádicamente se extiende dentro del valle andino de Perú o costa Oeste del Ecuador (Prado, 2000).

La distribución geográfica de *A. colubrina var. cebil* en Argentina se restringe al N del país, encontrándose representada en las provincias de Misiones, Corrientes, Catamarca, Tucumán, Jujuy y Salta. Los especímenes colectados en Catamarca representan el límite de distribución respecto a la latitud S para esta especie (Altschul, 1964) (Figura I-1). Los paisajes en los que se encuentra esta especie incluyen: riberas de ríos con suelos fértiles, bosques, laderas de montañas, campos y el interior de montañas ribereñas, en altitudes de hasta 1250 metros sobre el nivel del mar (Altschul, 1964).

Las especies del género *Anadenanthera* presentan largos períodos de adaptación a condiciones de sequía y a diferentes altitudes. Estos árboles son semidecíduos y aparentemente tolerantes a largos períodos de sequía con lluvias esporádicas, las cuales inician la producción de follaje nuevo (Altschul, 1964). Las especies de este género no parecen hibridizar en la naturaleza, aunque sus períodos de floración se solapan y son simpátricas. *A. colubrina* presenta

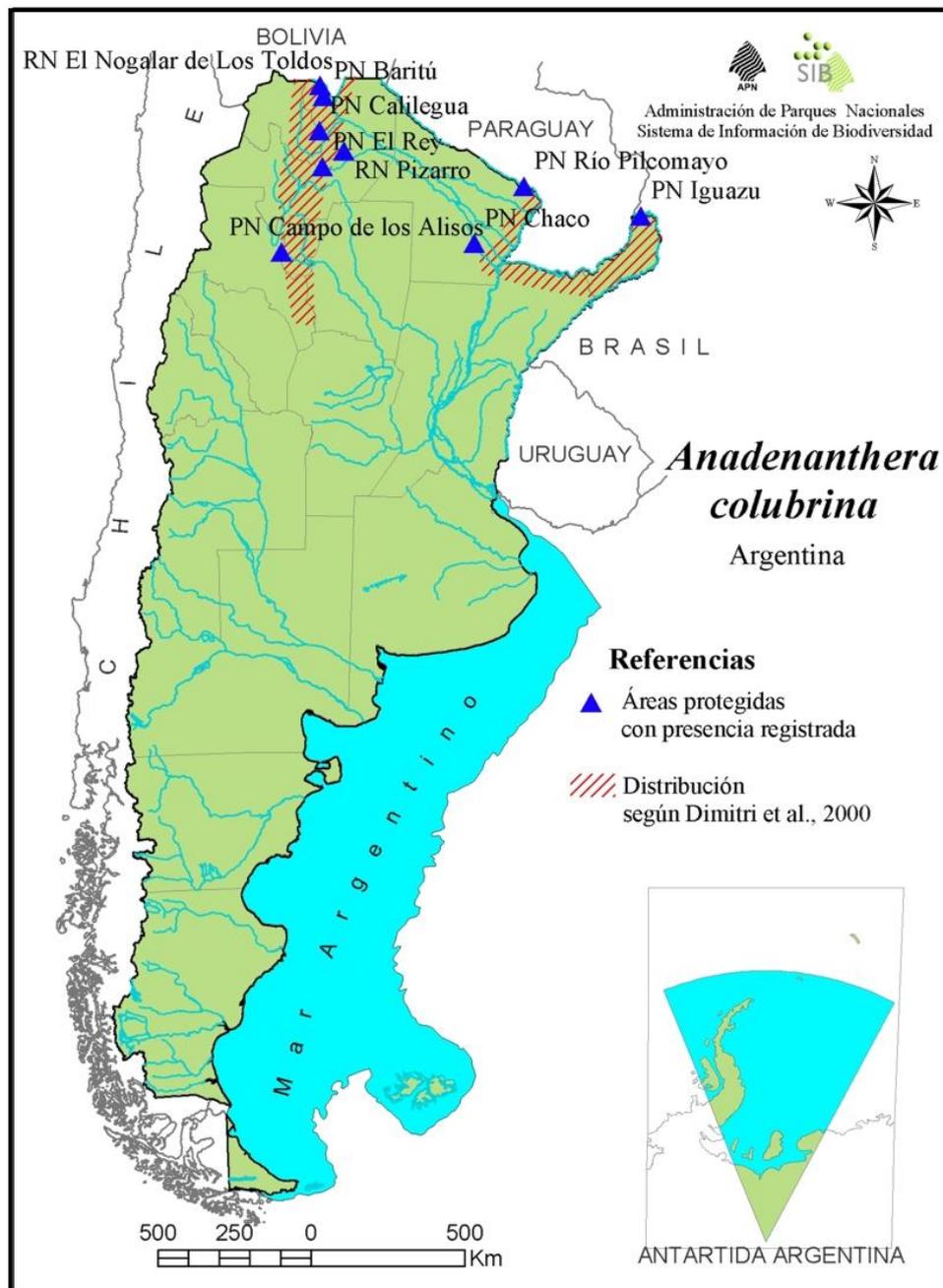


Figura I-1: Distribución de *Anadenanthera colubrina* var. *cebil* en Argentina según Dimitri *et al.* (2000). Imagen tomada de Administración de Parques Nacionales, Sistema de Información de Biodiversidad disponible en http://www.sib.gov.ar/ficha/PLANTAE*Anadenanthera*colubrina

una mayor variación que *A. peregrina* lo que podría estar relacionado a una distribución más reciente de *A. colubrina* como así también, a que sería una especie más joven (Altschul, 1964). Por su parte, *A. colubrina* presenta dos variedades: *A. colubrina* var. *cebil* y *A. colubrina* var. *colubrina* las que se pueden distinguir por una serie de características morfológicas correlacionadas con una distribución geográfica particular, siendo *A. colubrina* var. *cebil* la que presenta una distribución más extendida. Las diferencias morfológicas involucradas entre estas variedades no son absolutas y dentro de una serie de especímenes se pueden identificar tendencias, haciendo que sean consideradas variedades y no especies distintas (Altschul, 1964).

Los estudios citogenéticos realizados en muestras provenientes de la reserva forestal ubicada en el estado San Paulo (Brasil) por Gibbs e Ingram (1982) determinaron para *A. colubrina* var. *cebil* un número cromosómico de $2n=26$ mientras que Oliveira Alves y Valdinar Carvalho Custodio (1989), establecieron un número cromosómico de $2n=24$. Por su parte, Honfi y Daviña (1997), estudiando individuos de la ciudad de Posadas establecieron para esta especie un número cromosómico de $2n=26$.

Desde el punto de vista botánico, se describe *A. colubrina* var. *cebil* como un árbol de gran altura, que puede alcanzar hasta 35 metros (Cialdella, 2000). El fuste se presenta generalmente recto, cilíndrico y raras veces cuadrilátero. Por ser una especie exclusivamente heliófita se presenta en el dosel arbóreo o de forma emergente. La forma de la copa es variable dependiendo de las condicionantes de su entorno vegetal (Justiniano y Fredericksen, 1998). Posee un tronco de entre 30-50 cm de diámetro a la altura de pecho, siendo su corteza de 2-5 cm de color grisácea provista con proyecciones en forma de mamelones especialmente desde la base (Altschul, 1964). Las hojas son compuestas dispuestas en espiral, bipinnadas con 18 a 20 pares de pinnas opuestas, cada una formada por 30 a 60 folíolos finos, sésiles y opuestos (Justiniano y Fredericksen, 1998). El raquis y las raquillas son pubescentes, presentando una glándula alargada en el pecíolo por debajo del primer par de pinnas, además presenta de 1-3 glándulas similares pero más pequeñas por debajo de los últimos pares de pinnas. Estas glándulas exudan una sustancia atractiva para los insectos y posiblemente se halla relacionada a la polinización (Altschul, 1964). Las flores son blanco amarillentas, hermafroditas, actinomorfas y sésiles. El cáliz y la corola son acampanados, presentando 5 lóbulos triangulares. Los estambres son 10, con filamentos libres y exsertos. El ovario es súpero y oblongo (Justiniano y Fredericksen, 1998). Las inflorescencias están constituidas por 35 a 50 flores sobre un receptáculo globoso oblongo. Las inflorescencias se producen sobre un pedúnculo y se agrupan en fascículos, los cuales forman racimos dispuestos en las axilas de las hojas compuestas y subterminales (Altschul, 1964). El fruto es una vaina larga algo angosta, achatada, con una superficie glabra entre coriácea y leñosa de color castaño oscuro, y presenta una longitud de 10 a 32 cm. Es unilocular y presenta un borde ondulado que se ensancha entre las divisiones de cada semilla. La dehiscencia de las vainas ocurre a lo largo de una única sutura (Altschul, 1964)

y las semillas se precipitan al suelo por gravedad (Justiniano y Fredericksen, 1998). Las semillas son discoideas, aplanadas, coriáceas, de color café-negruzco, lustrosas, de 12 a 20 mm de diámetro. El número de semillas por fruto oscila entre 8 y 16. Se ha descrito que las semillas de *A. colubrina* var. *cebil* germinan entre los dos y tres días de ser sembradas (Altschul, 1964), siendo su capacidad germinativa muy elevada superando el 80% (Justiniano y Fredericksen, 1998).

A. colubrina var. *cebil* es una especie alógama (Ciadella, 2000). La floración generalmente se manifiesta al final de la época seca e inicio de la húmeda, entre Octubre y Noviembre. El *A. colubrina* var. *cebil* se caracteriza por el intenso olor que desprenden sus flores para atraer a los polinizadores, principalmente las abejas, con lo cual su polinización es entomófila. Los frutos tardan 10 meses en desarrollarse hasta su diseminación, observándose picos de maduración en Octubre (Justiniano y Fredericksen, 1998). La dispersión de semillas es autocórica-anemocórica (Abraham De Noir *et al.*, 2002). Luego de la dehiscencia de las vainas, el viento ayuda a la diseminación de las semillas a corta distancia mediante el movimiento de los frutos (Justiniano y Fredericksen, 1998). Además, las semillas al ser delgadas, circulares y livianas pueden ser dispersadas fácilmente por los ríos, por las inundaciones y aún por el viento. Su germinación rápida podría representar una adaptación a las inundaciones repentinas (Altschul, 1964).

En relación a la anatomía de la madera, *A. colubrina* var. *cebil* posee una madera dura, pesada y de textura fina, con fibras entrelazadas. Es de color rojiza donde la albura se encuentra muy diferenciada del duramen y los anillos de crecimiento son claramente aparentes (Altschul, 1964). Desde el punto de vista económico el valor de la madera está dado por su dureza y ha sido utilizada como energía, principalmente para leña y carbón, ya que es un excelente combustible de alto poder calorífico (Justiniano y Fredericksen, 1998). Las ventajosas cualidades de esta especie nativa, tales como elevado poder germinativo, tolerancia a alta exposición solar, alta tasa de regeneración en zonas alteradas, capacidad de adaptarse a ambientes con disturbios, rápido crecimiento, capacidad de rebrote luego de ser cortado, elevada fijación de nitrógeno en el suelo y la ausencia de depredadores la transforman en una especie ideal para reforestar áreas degradadas (Justiniano y Fredericksen, 1998). En la Figura I-2 se presentan imágenes de algunos de caracteres morfológicos de *A. colubrina* var. *cebil*.

A. colubrina var. *cebil* forma parte de los Bosques tropicales estacionalmente secos (*Seasonally Dry Tropical Forests – SDTF*) (Figura I-3). Estos bosques secos se encuentran en áreas de suelos fértiles donde las precipitaciones son inferiores a 1600 mm/año presentando, al menos, cinco o seis meses de precipitaciones inferiores a 100 mm. En estas áreas la vegetación es principalmente decidua presentando un ecosistema dominado por árboles, en el cual predominan Leguminosas y Bignoniáceas, y donde los pastos están poco representados (Pennington *et al.*, 2000).



Figura I-2: Características morfológicas de *Anadenanthera colubrina* (Vell.) Brenan var. *cebil* Griseb. Altschul. A) Individuo adulto, B) Frutos y hoja compuesta, C) Glándula alargada en el pecíolo, D) Tronco grisáceo con mamelones, E) Inflorescencias globosa en racimos, F) Semillas discoidales aplanadas, y G) Corte transversal y longitudinal de la madera (todas las imágenes son de propia autoría y fueron tomadas en diferentes salidas de campaña).

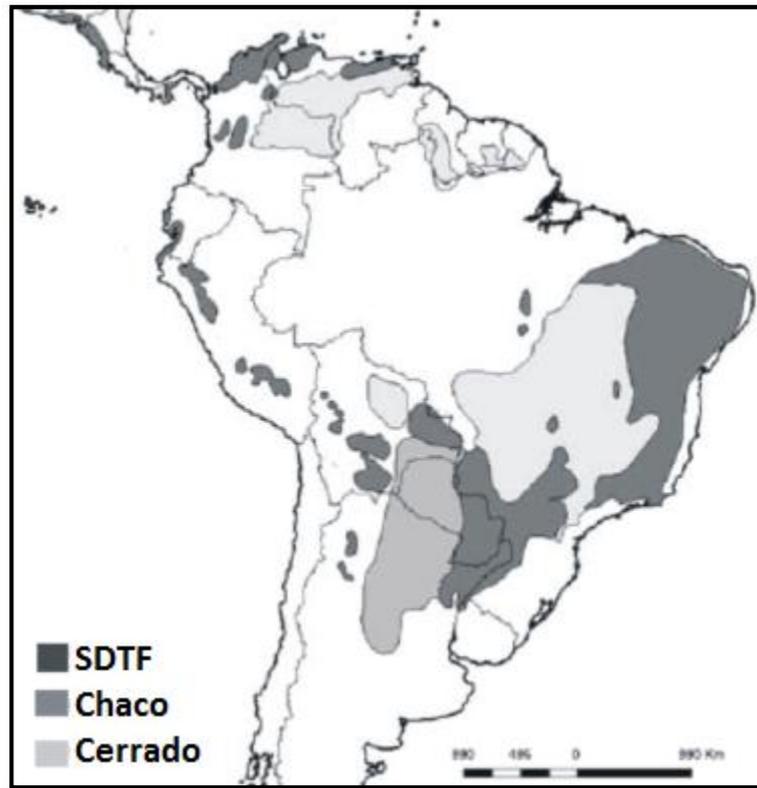


Figura I-3: Distribución del Bosque Tropical Estacional en Sudamérica. Además se diferencian las formaciones del Chaco y Cerrado. Imagen adaptada desde Pennington *et al.* (2000).

A. colubrina var. *cebil* es considerada por Prado y Gibbs (1993) como una de las especies paradigmáticas entre las especies involucradas en los ciclos migratorios de expansión y retracción ocurridos durante el Pleistoceno. Pueden definirse tres núcleos para la distribución de esta especie: 1) el núcleo Caatingas, en el NE de Brasil, donde su presencia es muy frecuente, 2) el núcleo Misiones, que ocurre a lo largo de los sistemas de los ríos Paraguay y Paraná en el NE de Argentina, E de Paraguay y SE de Brasil, donde su presencia es común pero raramente dominante y 3) el núcleo del Pedemonte Subandino, en el SO de Bolivia y NO de Argentina, en el cual es una de las cinco especies más frecuentes en estos bosques (Figura I-4).

Estos bosques tropicales se encuentran amenazados ya que estas regiones se hallan localizadas en uno de los mejores suelos para la agricultura en la región tropical de Sudamérica y mantienen una mayor población que la selva en los trópicos (Prado, 2000). En Argentina, las Yungas constituyen una unidad de particular interés. Aunque representada por una estrecha franja de hábitat que ocupa solo el 1% del territorio argentino (Brown *et al.*, 2002), albergan una diversidad largamente reconocida que, junto a la Selva Paranaense, incluye más del 50% de la diversidad biológica del territorio nacional (Brown *et al.*, 2001). Ambos núcleos sufrieron la extracción de especies maderables que produjo la degradación y el empobrecimiento del bosque. Además de la actividad forestal, otros procesos tales como el avance de la frontera agropecuaria, la mecanización del transporte, la construcción de caminos y de represas y el establecimiento de poblaciones favorecieron la eliminación de la cubierta boscosa y la fragmentación del hábitat natural (Tortorelli, 2009).

Actualmente en la Provincia de Misiones persisten unas pocas y pequeñas poblaciones fragmentadas de *A. colubrina* var. *cebil* en los extensos pajonales del sur de la Provincia. La tala y la quema redujeron la superficie del bosque, el cual hoy subsiste en forma de parches o fragmentos en sitios de difícil acceso, tales como las laderas del cerro Santa Ana, parcelas de propiedad fiscal y a lo largo de arroyos (Fontana, 1993). El típico paisaje de los “campos”, en el que se encuentran las poblaciones de *A. colubrina* var. *cebil*, está constituido por comunidades de pajonales y pastizales, salpicados por isletas de bosques en un relieve de suaves ondulaciones (Fontana, 1993). Un origen antropogénico de las isletas de bosque puede ser inferido debido a su presencia sin preferencias por una ubicación determinada, a la afinidad entre los suelos del bosque y del pajonal y al avance continuo de la vegetación arbórea sobre las comunidades de pajonales (Fontana, 1993). Además, *A. colubrina* var. *cebil* está presente en el Norte de la provincia dentro del Parque Nacional Iguazú.

En la provincia biogeográfica de las Yungas, la elevada deforestación ocurrió por el avance de la frontera agropecuaria. La explotación azucarera representó la mayor actividad antes de los '80, mientras que el cultivo de soja representa la principal actividad desde los '80 en adelante. Actualmente esta actividad produce la pérdida de 10.000 ha al año (Brown, 2009). Las Yungas, en sentido estricto, se caracterizan por un fuerte gradiente altitudinal. Como

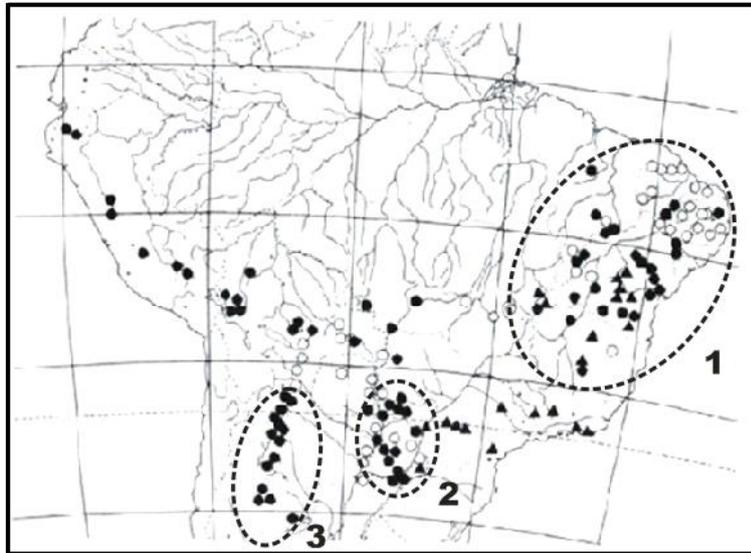


Figura I-4: Distribución de ejemplares de *Anadenanthera colubrina*: ● var. *cebil* y ▲ var. *colubrina* según el estudio realizado por Prado y Gibbs (1993). En el mapa se indican los núcleos de distribución de esta especie: Núcleo 1) Caatingas, Núcleo 2) Misiones y Núcleo 3) Pedemonte Subandino. Imagen adaptada desde Prado y Gibbs (1993).

respuesta a este gradiente su vegetación se organiza en pisos o franjas de vegetación. *A. colubrina* var. *cebil* es una especie dominante en la Selva pedemontana, la cual ocupa un rango altitudinal entre los 400 y 700 msnm en el pedemonte y las serranías de escasa altitud. Además se encuentra presente en la Selva montana la cual ocupa las laderas de las montañas entre los 700 y los 1500 msnm.

1.6.2 Consecuencias del cambio climático histórico sobre la distribución de A. colubrina var. *cebil*

Los primeros fósiles de la familia Fabaceae corresponden a la subfamilia Caesalpinioideae y datan del Cretácico tardío. Por su parte, los fósiles de polen y microfósiles correspondientes a las subfamilias Mimosoideae y Papilionoideae no han sido documentados antes del Eoceno (Pennington *et al.*, 2004). Estudios filogenéticos datan el origen de los SDTF entre 20 y 30 millones de años antes del presente, mientras que el registro fósil data su origen entre 12 y 13 millones de años (Collevatti *et al.*, 2012). Asimismo, el registro fósil sugiere que la mayoría de las especies del neotrópico deben haberse originado durante los pasados 65 millones de años de la era Cenozoica (Pennington *et al.*, 2004).

Evidencias desde el registro polínico, datos climatológicos y genéticos han sido empleados para determinar la importancia de las fluctuaciones climáticas históricas y los cambios en la vegetación sobre la diversificación en el Neotrópico. Sin embargo, la magnitud de la influencia de dichas fluctuaciones sobre la evolución de los distintos grupos taxonómicos se encuentra aún en discusión (Werneck *et al.*, 2011).

Los SDTF en Sudamérica forman un arco alrededor de la cuenca Amazónica (Figura I-3), con sus extremos posicionados en el dominio de Caatingas en el Norte de Brasil y las costas del Caribe en Colombia y Venezuela, mientras que entre estos extremos se presentan como núcleos aislados en la periferia del Dominio Chaqueño, en el pedemonte Subandino, en los valles interandinos y en las costas del Pacífico (Caetano *et al.*, 2008). La distribución disyunta actual de los SDTF neotropicales podría representar refugios modernos para las especies que habitan estos bosques. En el neotrópico, el clima glacial fue extremadamente seco, lo cual sugiere que las áreas actuales disponibles para el crecimiento de los SDTF se encuentren probablemente en su mínima expresión debido a que al presente período interglaciar húmedo (Pennington *et al.*, 2004). Los SDTF ocurren en parches disyuntos (conocidos como núcleos) dispersos a lo largo del Neotrópico ya que las especies que lo componen crecen en áreas con una marcada estacionalidad. Esta distribución fragmentada ha persistido a través de un período de tiempo suficiente, en términos evolutivos, como para haber influenciado la evolución y a los componentes biogeográficos de los linajes vegetales (Pennington *et al.*, 2009).

Las evidencias fósiles y climatológicas indican que los SDTF constituyen un bioma antiguo que data del Eoceno medio. Los estudios de filogenia molecular avalan la antigüedad geológica y la constitución de clados geográficamente confinados de las especies que integran

los SDTF (Pennington *et al.*, 2009). De esta manera, los núcleos persistentes y con limitada dispersión podrían resultar en reducidos tamaños efectivos poblacionales constituidos por especies antiguas (Pennington *et al.*, 2009).

Prado y Gibbs (1993) postularon la existencia de un Arco Pleistocénico para explicar la extensa distribución disyunta de varias especies arbóreas. Pennington *et al.* (2000) además sugirieron que las especies de los SDTF podrían haber penetrado la cuenca Amazónica durante los períodos glaciares. De manera que la distribución disyunta actual resulta de la fragmentación de esa distribución más extensa de los SDTF. La explicación alternativa a esta amplia distribución continental de los SDTF invoca a eventos raros de dispersión a larga distancia (Caetano *et al.*, 2008). Otra hipótesis plantea la posible expansión de las sabanas tropicales, presentes en el Cerrado de Brasil y en el Chaco del Norte Argentino, oeste de Paraguay y Sudeste de Bolivia. A pesar de que bioma también crece en áreas con reducidas precipitaciones y elevada estacionalidad, el Cerrado se caracteriza por presentar una abundante capa herbácea xeromórfica tolerante al fuego y sus suelos presentan baja fertilidad. Esta hipótesis puede ser descartada porque además los paleoecólogos no detectaron granos de polen de gramíneas en la cuenca Amazónica (Pennington *et al.*, 2004).

Werneck *et al.* (2011), empleando modelos de paleodistribución y validación palinológica, describieron la posible distribución histórica de los SDTF como una distribución restringida durante el último evento máximo glacial (LGM) seguida de una expansión que se inició en el Holoceno temprano. Luego de esa fase de expansión, la distribución de los SDTF habría experimentado fluctuaciones menores hasta alcanzar la distribución actual (Figura I-5). Según estos autores la distribución actual de los SDTF es más extensa que en el Pleistoceno en contradicción con la teoría del Arco Pleistocénico. Estos mismos autores proponen que las disyunciones y endemismos pueden haberse originado tanto por eventos de dispersión antiguos como por un periodo de expansión en el Terciario seguido por una fragmentación previa al LGM. Sin embargo, a pesar de que ambas teorías difieren en cuanto a los tiempos en que ocurrieron los eventos ambas sostienen patrones de distribución históricos más extensos (Werneck *et al.*, 2011).

Estudios filogeográficos en *Astronium urundeuva* (Caetano *et al.*, 2008) y *Tabebuia impetiginosa* (Collevatti *et al.*, 2012) sugieren que su distribución actual se corresponde con un relictos de una distribución ancestral más amplia a partir de evidencias de expansión de rango en el Pleistoceno. Sin embargo, Collevatti *et al.* (2012) a partir de técnicas de modelado de distribución de especies y simulaciones coalescentes, afirman que la respuesta de los SDTF a los cambios climáticos del Cuaternario es altamente compleja y podría diferir entre las especies.

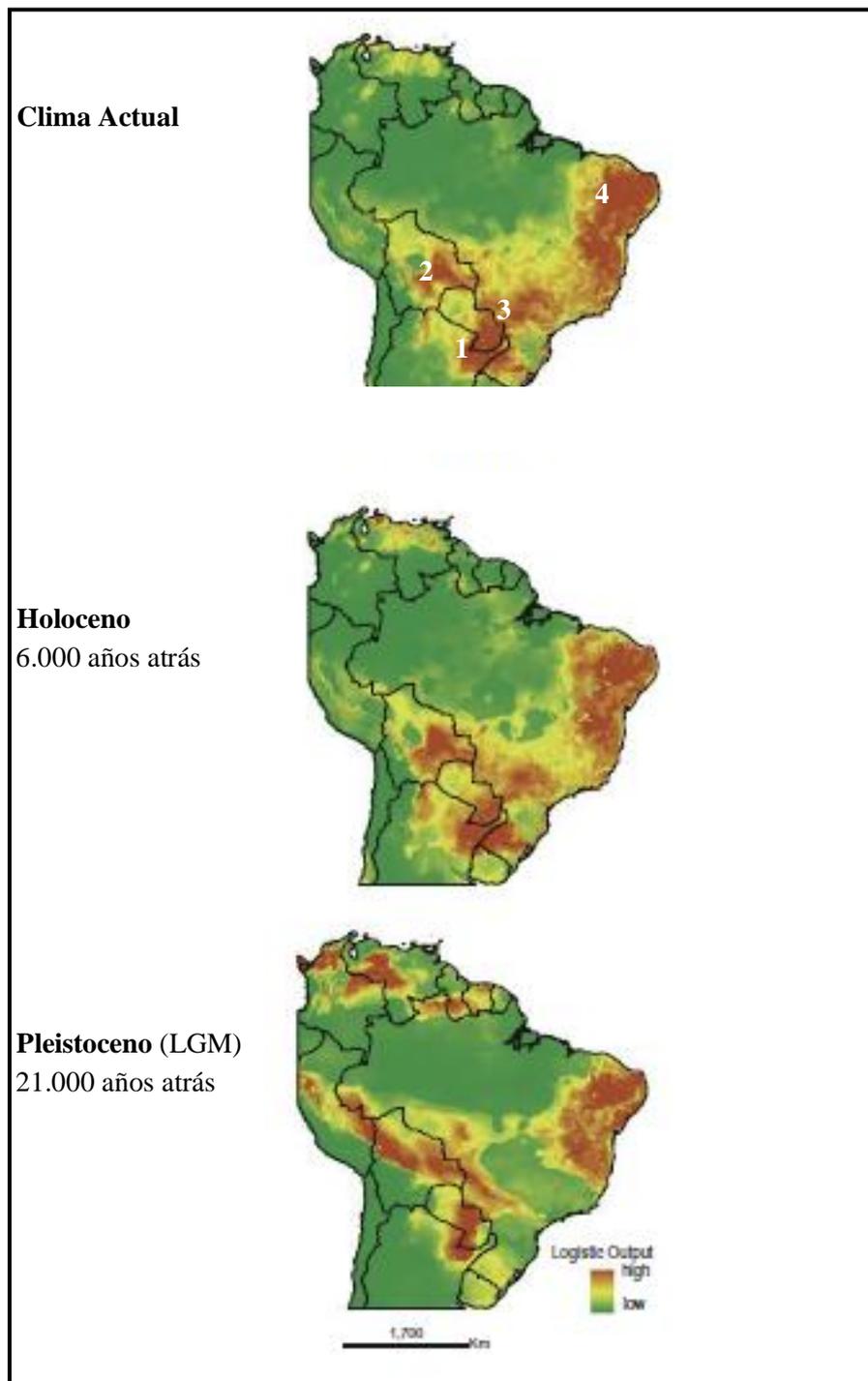


Figura I-5: Posible distribución histórica de los SDTF obtenida mediante modelos de paleodistribución en tres escenarios: clima actual, Holoceno y último máximo glacial (LGM). En anaranjado se representan áreas con elevada estabilidad y en amarillo áreas inestables dentro de los SDTF. Las áreas estables desde LGM se corresponden con refugios. **1:** Núcleo Misiones; **2:** Chiquitano; **3:** Mato Grosso y **4:** Caatinga. Imagen adaptada desde Werneck *et al.* (2011).

2. Hipótesis y Objetivos Generales

2.1 El trabajo de tesis se apoya en la siguiente hipótesis:

“El flujo génico homogeneiza las frecuencias génicas entre las poblaciones contrarrestando los efectos de la deriva genética provocados por la fragmentación”.

2.2 Objetivos generales

- Indagar acerca del rol evolutivo del flujo génico y de la deriva genética como fuerzas modeladoras de las frecuencias génicas en las poblaciones fragmentadas.
- Determinar la importancia del movimiento de los alelos a través de la dispersión de las semillas y del polen.

3. Fundamentación de la metodología utilizada para el logro de los objetivos: Empleo de marcadores moleculares

Se conoce como marcador molecular a cualquier fragmento específico de ADN que puede corresponder a regiones codificantes o no del genoma. (Ferreira y Grattapaglia, 1996) y es representativo de diferencias a nivel genómico (Agarwal *et al.*, 2008). En la medida que se conozca su comportamiento de acuerdo a las leyes mendelianas de la herencia, un marcador molecular puede adicionalmente ser definido como un marcador genético, por ello se dice que los datos moleculares proveen información genética (Ferreira y Grattapaglia, 1996). Los ensayos moleculares no sólo revelan características detalladas del ADN sino que además revelan los estados variables de esas características, poseen bases genéticas y presentan modos de transmisión que pueden ser especificados (Avisé, 2004). Un aspecto importante de los datos moleculares es que permiten una comparación directa de los niveles relativos de diferenciación genética, esencialmente entre cualquier *taxa*, conociéndose esto como “criterio genético universal” (Avisé, 2004). Los marcadores moleculares ofrecen ventajas sobre las alternativas convencionales basadas en el fenotipo, ya que contrariamente a los datos morfológicos los datos moleculares son estables y detectables en todos los tejidos sin estar relacionados con el desarrollo, diferenciación, crecimiento, o estado de defensa de la célula y no son influenciados por los efectos ambientales (Marcucci Poltri, 2005; Agarwal *et al.*, 2008). El rastreo con marcadores moleculares se basa en la medición de la variación en *loci* específicos para revelar los niveles de diversidad genética (Ayad *et al.*, 1995). El conocimiento de las bases moleculares precisas y el modo de transmisión hereditaria de un polimorfismo genético es crucial para la interpretación apropiada de los marcadores moleculares en un contexto poblacional (Avisé, 2004).

En la segunda mitad del siglo XX, el interés de los estudios de ecología molecular y de evolución molecular radicaba en el entendimiento del rol funcional y la significancia adaptativa de la variación genética. Es así, que antes de la mitad de los años '80 se buscaba entender la manera en la cual la selección natural operaba a nivel de las proteínas y del ADN desmereciendo la aplicación de las moléculas como “meros” marcadores genéticos. De manera gradual, los marcadores moleculares fueron apreciados por mérito propio, debido a su riqueza empírica y conceptual, en estudios del comportamiento de los organismos, de la historia natural y de las relaciones filogenéticas (Avisé, 2004). Actualmente, los estudios con marcadores moleculares han brindado una importante contribución al entendimiento del flujo génico, hibridación, estructura poblacional, deriva genética y sistemas de fecundación (González-Martínez *et al.*, 2006). La mayoría de los estudios que utilizan marcadores moleculares pueden ser interpretados como una estimación de las relaciones genéticas a lo largo de un continuo jerárquico de divergencia evolutiva que va desde reciente a distante, siendo ejemplos los estudios de identidad genética, parentesco genético, pedigrís en poblaciones locales, afinidades

genealógicas entre poblaciones geográficas de una especie, divergencia genética, entre otros. Los diferentes tipos de ensayos con marcadores moleculares proveen información genética adecuada para diferentes horizontes temporales siendo el desafío desarrollar y emplear métodos apropiados para dar respuesta a diferentes preguntas biológicas (Avise, 2004).

El tipo ideal de marcador genético debe ser medible de manera confiable, presentar *loci* altamente variables, ser codominante, y debe estar densamente distribuido en el genoma. Los marcadores microsatélites, SSRs (*Simple Sequence Repeat*) o STR (*Short Tandem Repeat*), reúnen todos estos requisitos (Karhu, 2001). Resultan marcadores informativos y prácticos ya que proveen información acerca de la cantidad y distribución de la diversidad genética y de los procesos que determinan la estructura y la variación dentro y entre poblaciones naturales (Provan *et al.*, 2001; Frankham *et al.*, 2004).

Los SSRs son repeticiones monótonas de motivos nucleotídicos cortos de 1 a 6 pares de bases. Estos elementos repetitivos se encuentran distribuidos de manera intercalada en los tres genomas eucariotas: núcleo (SSRs), mitocondrias (mtSSRs) y cloroplastos (cpSSRs) (Tautz y Renz, 1984). Los distintos alelos SSRs son generados principalmente mediante mecanismos de adición y sustracción de repeticiones simples que ocurren con igual probabilidad (Schlotterer y Tautz, 1992), y raramente se encuentran en regiones codificantes (Karhu, 2001).

Entre las desventajas del uso de marcadores moleculares están principalmente el entrenamiento profesional necesario para hacer uso de estas herramientas ya que es considerablemente alto y los costos económicos son elevados (Avise, 2004). Sin embargo, los avances tecnológicos han contribuido en gran medida con el perfeccionamiento y la disponibilidad de las herramientas moleculares haciéndolas técnicamente más simples y rentables (Agarwal *et al.*, 2008).

Los vegetales presentan una variedad importante de modos de herencia (biparental, materna y/o paterna) y muchos de sus patrones reproductivos permiten realizar estudios genéticos que no son posibles de realizar en otros organismos (Linhart y Grant, 1996).

En este trabajo se emplearon marcadores microsatélites del cloroplasto materna (cpSSRs) y microsatélites nucleares (SSRs) los cuales brindan información diferencial acerca de los niveles de flujo génico mediado por polen y semillas. Las tasas de mutación diferenciales, los diferentes niveles de ploidía y la ausencia o presencia de recombinación hacen de estos marcadores una herramienta valiosa para el estudio de los efectos de la fragmentación histórica y de la fragmentación contemporánea sobre las poblaciones, permitiendo contrastar el efecto relativo de la acción de la deriva genética en relación al flujo génico.

4. Material biológico: Área de estudio y metodología

El área de estudio se situó en el N argentino donde se delimitaron cuatro poblaciones naturales de *A. colubrina* var. *cebil*, dos de ellas localizadas al Noreste en la provincia biogeográfica Paranaense y las otras dos al Noroeste en la provincia biogeográfica de las Yungas (Figura I-6 A). La distancia lineal que separa estas dos provincias supera los 1000 km.

Provincia biogeográfica Paranaense: Se halla ubicada en la provincia de Misiones, la cual desde un punto de vista fitogeográfico pertenece al dominio Amazónico (Cabrera, 1971). Además, la provincia Paranaense se encuentra dentro de una formación más extensa, denominada Bosque Atlántico del Alto Paraná (Di Bitetti, 2003). Siguiendo la descripción propuesta por Prado (2000) acerca de la distribución de *A. colubrina* var. *cebil*, las poblaciones analizadas son representativas del núcleo Misiones. El área de estudio se localiza al S de la provincia de Misiones dentro del departamento Candelaria, y se extiende desde una zona de elevaciones que corresponden al cerro Santa Ana en dirección hacia el río Paraná. La zona que se consideró se ubica en la sabana que conforma la zona de los “campos”. Los suelos de esta zona son bien drenados, extremadamente ácidos, con muy baja dotación de nutrientes (Ligier *et al.*, 1985). El clima característico es cálido y húmedo con temperaturas medias entre los 20 y 21° C, siendo los inviernos suaves y los veranos cálidos con lluvias frecuentes (Cabrera, 1971). Las diferencias de altitud entre las poblaciones consideradas en esta región no son marcadas, registrándose elevaciones entre los 100 y 150 msnm. La distancia lineal que separa las poblaciones es de 18 km. En esta provincia biogeográfica se estudiaron dos poblaciones, una localizada en Candelaria y otra en Santa Ana.

1) *Población Candelaria:* Se ubica en el suroeste de la provincia de Misiones a orillas del río Paraná dentro de una propiedad privada. En esta zona crecen los pajonales con predominancia de *Aristida jubata* en suelos lateríticos profundos zona que Martínez Crovetto denominó “Distrito de los Campos” (Fontana, 1993).

2) *Población Santa Ana:* Se ubica en una propiedad fiscal en las laderas del cerro del mismo nombre en una región de pajonales de *Elyonurus muticus* entre isletas de bosque que coincide con la zona que Martínez Crovetto denominó “Distrito del Urunday” (Fontana, 1993).

Provincia biogeográfica de las Yungas: Las poblaciones estudiadas pertenecen a la ecoregión de las Selvas pedemontanas y a las Selvas montanas dependiendo de la altitud a la cual se encuentren los individuos. En esta provincia biogeográfica se estudiaron dos poblaciones, una localizada en la provincia de Tucumán y otra en la provincia de Jujuy las cuales se encuentran en las regiones Sur y Norte de la subdivisión ecológica de las Yungas (Cabrera, 1994; Dimitri, 1997). De esta manera, este estudio no comprende el gradiente latitudinal de distribución presentado por la especie en esta región.

3) *Población Tucumán:* Se ubica dentro de la Reserva Experimental Horco Molle. Esta reserva se localiza en una zona de lomadas de rumbo submeridional que constituyen la prolongación

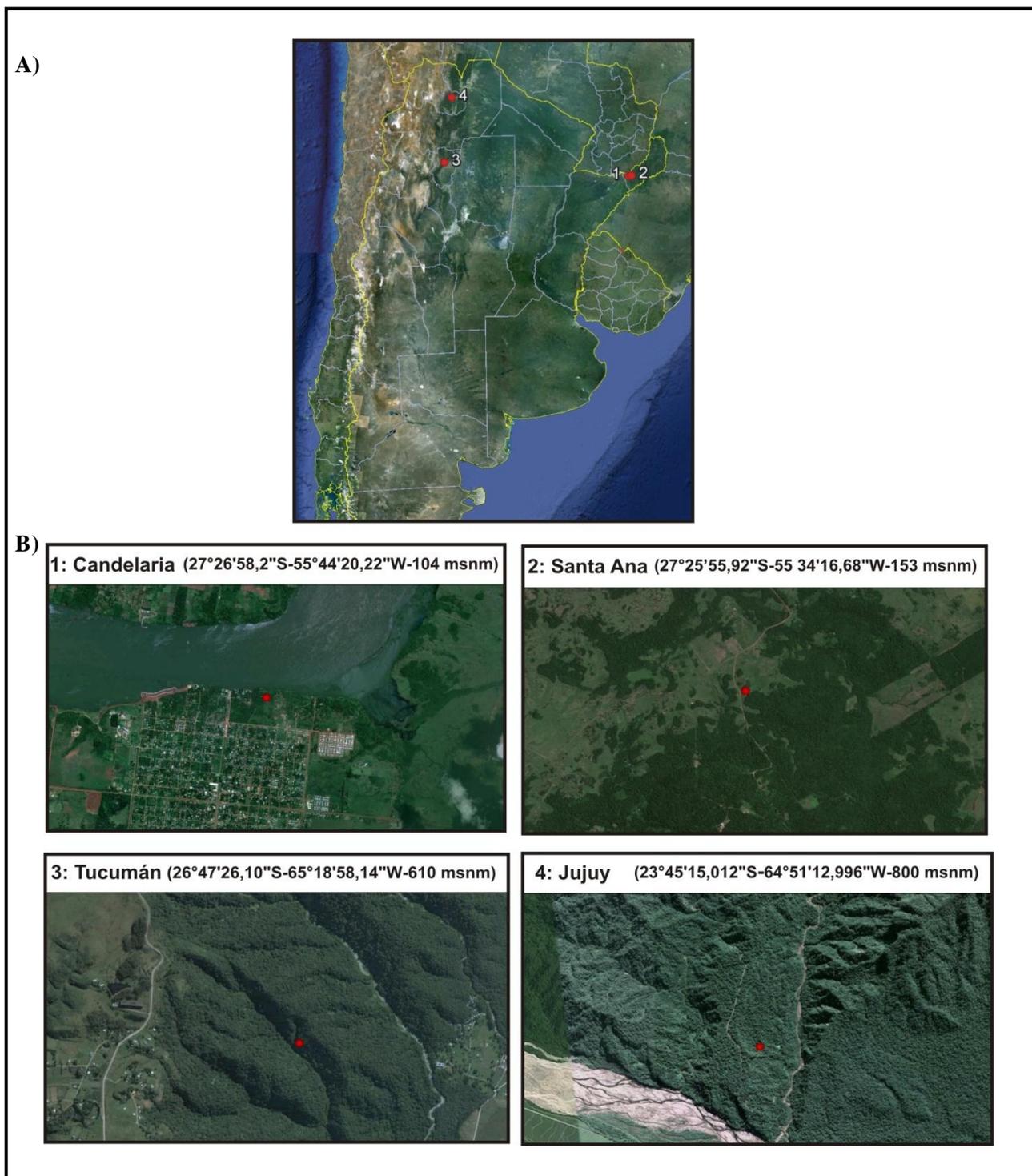


Figura I-6: Imágenes satelitales de los sitios considerados en este estudio. En **A)** se ubican las cuatro poblaciones analizadas en el N argentino: 1: Candelaria, 2: Santa Ana, 3: Tucumán y 4: Jujuy. 1 y 2 se encuentran en núcleo Paranaense, mientras que 3 y 4 en el núcleo Pedemonte Subandino. En **B)** se muestran fotos satelitales tomadas a una distancia de 5,5 km y los datos de geoposición y altitud correspondientes al primer individuo colectado en cada población. Imágenes satelitales obtenidas desde *Google earth (Image © 2012 Digital Globe)*.

norte de las llamadas “Lomas de Imbaud” sobre el pie oriental de la Sierra de San Javier. Se encuentra en la zona de Horco Molle en el Departamento Yerba Buena a unos 15 km de la ciudad de San Miguel de Tucumán (Richard, 2000). La zona se caracteriza por la presencia de lomadas de baja altura. El clima es templado y húmedo, con inviernos secos y veranos lluviosos. Siendo la temperatura media anual de 18°C (Richard, 2000). Su vegetación corresponde a la Selva Pedemontana (Cabrera y Willink, 1980).

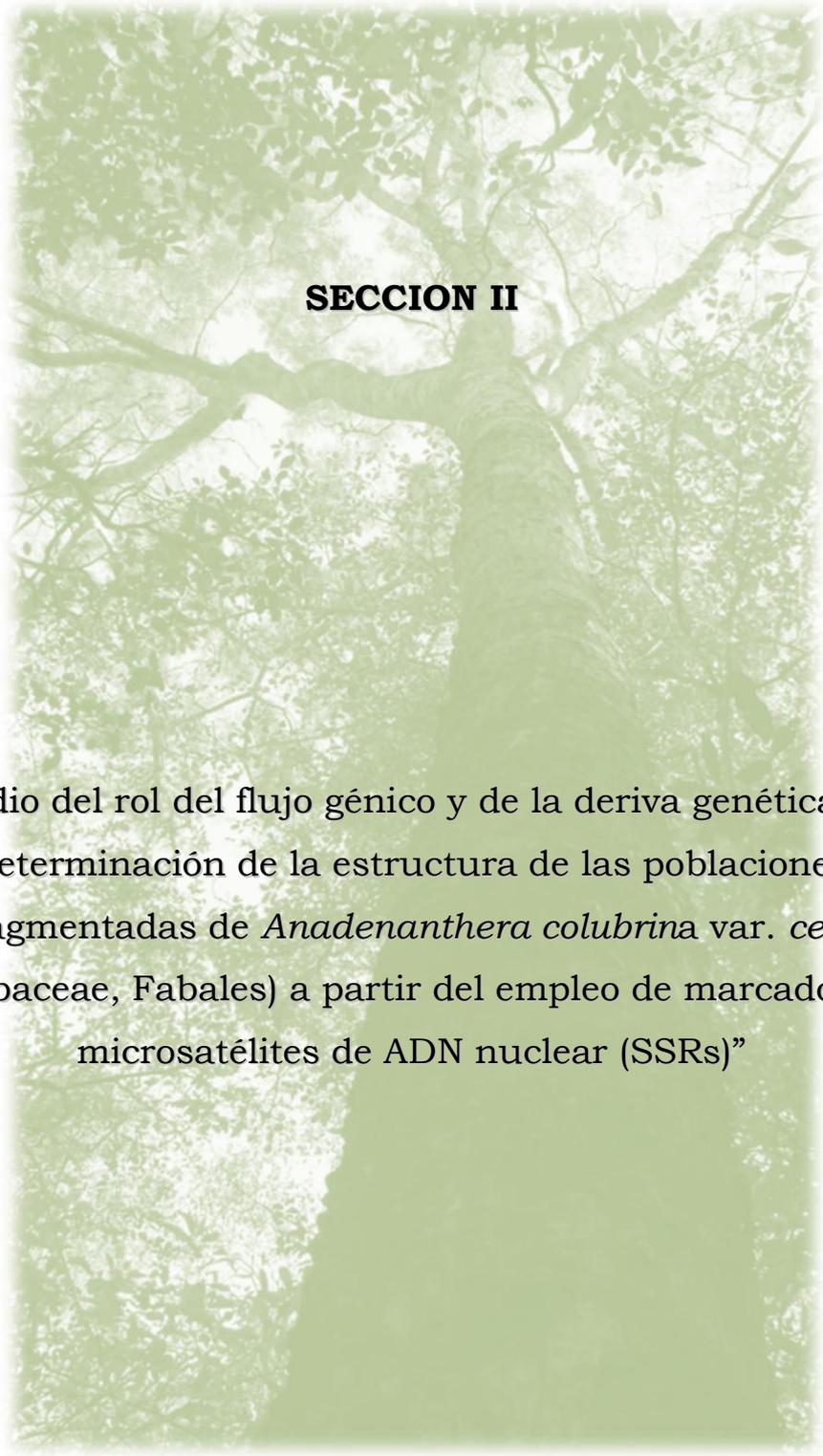
4) *Población Jujuy*: Se ubica dentro del Parque Nacional Calilegua. Este Parque se encuentra ubicado sobre el faldeo oriental de las Serranías de Calilegua y abarca una superficie de 76.000 ha que protege principalmente la franja altitudinal de Selva Montana. La Selva Pedemontana ocupa sectores marginales del Parque en las áreas más bajas, ya que la mayor superficie de este ecosistema ha quedado fuera del área protegida debido a su valor potencial como tierra de cultivo. Este Parque posee laderas muy abruptas, donde se producen regularmente deslizamientos generando un paisaje dominado por *Tipuana* y *Anadenanthera* (Brown *et al.*, 2002). El clima es de tipo subtropical con estación seca, caracterizado por inviernos templados y veranos lluviosos y calurosos. Las diferencias altitudinales de la zona provocan modificaciones en este patrón climático general. En el paisaje del Parque predominan las montañas escarpadas, pliegues, fallas, quebradas y valles.

En las imágenes satelitales de cada población (Figura I-6 B) puede verse el estado actual que presenta cada población en cuanto a la subdivisión poblacional y a la transformación de su hábitat.

Se colectaron hojas jóvenes de 10 individuos adultos en cada unidad de muestreo ubicados a una distancia de separación superior a 3 m de radio entre sí. Como consecuencia de la reducida densidad poblacional se muestrearon seis individuos en Santa Ana B, nueve en Tucumán E, cinco en Tucumán F y nueve en Jujuy H, con lo cual se analizó un total de 69 individuos (Tabla I-1). Las muestras fueron almacenadas en bolsas de cierre hermético conteniendo sílica gel, para su deshidratación y conservación. Los individuos fueron georeferenciados mediante el sistema de posicionamiento global (GPS) Garmin modelo eTrex Legend™

Tabla I-1: Descripción de las poblaciones de *A. colubrina* var. *cebil* analizadas.

Región	Población	N° de individuos	Latitud (S)	Longitud (O)	Altitud (msnm)
Paranaense	(1) Candelaria	20	27° 26' 58,20''	55° 44' 20,22'	104
	(2) Santa Ana	16	27° 25' 55,92''	55° 34' 16,68''	153
Yungas	(3) Tucumán	14	26° 47' 26,10''	65° 18' 58,14''	610
	(4) Jujuy	19	23° 45' 15,01''	64° 51' 12,99''	800



SECCION II

“Estudio del rol del flujo génico y de la deriva genética en la determinación de la estructura de las poblaciones fragmentadas de *Anadenanthera colubrina* var. *cebil* (Fabaceae, Fabales) a partir del empleo de marcadores microsatélites de ADN nuclear (SSRs)”

1. Introducción

El nivel de intercambio de genes entre las poblaciones se encuentra determinado por múltiples factores, entre los cuales se incluye el tamaño, la densidad y la forma de las poblaciones donantes y receptoras, el sistema de apareamiento de la especie, las características de la vegetación circundante, los alcances de la dispersión y los niveles de producción del polen y de las semillas, el comportamiento de los polinizadores y de los vectores de semillas, como así también la distancia entre las poblaciones. Además, estos factores pueden variar en tiempo y en espacio (Finkeldey y Hattemer, 2007).

1.1 Flujo génico mediado por polen y semillas en especies forestales

El polen puede ser dispersado en masa a través de grandes distancias, pero el flujo génico ocurre sólo cuando el polen llega a una población de destino y se produce la fertilización (White *et al.*, 2007). Por su parte, la zoogamia o polinización por animales, es la forma dominante en las especies forestales tropicales. A pesar de estar involucrados numerosos animales en la dispersión del polen, los insectos son los polinizadores primarios en la mayoría de los árboles tropicales (Dick *et al.*, 2003) en tanto que la polinización por el viento resulta ineficiente en estas especies (Hamrick y Murawski, 1990).

Los procesos de regeneración y de recolonización de nuevos hábitats requieren del movimiento de la información genética a través de las semillas (Finkeldey y Hattemer, 2007). En árboles forestales gran parte del conocimiento acerca de la dispersión de las semillas surge de los estudios de marcado y recaptura. Estos estudios directos, muestran que la dispersión de semillas es altamente leptocúrtica, es decir, que la gran mayoría de las semillas son dispersadas cerca del árbol madre y únicamente un pequeño porcentaje es distribuido a distancias mayores (White *et al.*, 2007). De esta manera, el potencial del flujo génico mediado por semillas parece ser mucho menor que el del flujo génico por polen. Sin embargo, los métodos directos para el estudio del flujo génico por semilla subestiman su aporte debido a que cada semilla lleva consigo los alelos maternos y paternos desde la población de origen, es decir, cada semilla inmigrante produce un efecto doble sobre el flujo de genes en relación a cada grano de polen inmigrante, el cual lleva únicamente los alelos paternos (White *et al.*, 2007).

Los mecanismos de dispersión, tanto gamética como cigótica, influyen sobre los niveles de flujo génico y sobre la estructura genética de las poblaciones vegetales. En adición, la historia evolutiva de las especies también juega un papel fundamental en la distribución de la variación genética (Avice, 2004). El estudio del movimiento de los genes por medio de las semillas y del polen es posible mediante el empleo de marcadores moleculares de herencia biparental. Así, es posible estimar los niveles y los patrones actuales del flujo génico y además, evaluar el impacto de los niveles de flujo génico histórico sobre la estructura genética de las

poblaciones. Cuando el flujo génico es limitado, es posible que las poblaciones diverjan en cuanto a sus frecuencias alélicas (White *et al.*, 2007).

Diversos estudios en los cuales se evalúa la estructura genética de las poblaciones demuestran que niveles elevados de diversidad genética son mantenidos dentro de las poblaciones de diferentes árboles tropicales y que las diferencias encontradas entre las especies no se encuentran asociadas con el sistema de apareamiento. Sin embargo, la distribución de la variación genética entre poblaciones y las estimaciones de los niveles de flujo génico dependen en igual medida del sistema de apareamiento y de los mecanismos de dispersión de las semillas de la especie (Hamrick y Murawski, 1990).

1.2 El genoma nuclear vegetal y los loci microsatélites

En el genoma nuclear de las plantas, así como en el de otros eucariotas superiores, existe una cantidad considerable de ADN no codificante (Lodish *et al.*, 2002). Una elevada proporción de ese ADN, aparentemente no funcional, consiste de motivos repetidos y es considerado ADN basura (Weising *et al.*, 2005). Los elementos de ADN repetitivo, dependiendo de la organización genómica, pueden ser clasificados como repeticiones intercaladas o repeticiones en tándem. Las repeticiones intercaladas, también llamado ADN móvil, son secuencias de cientos a miles de pares de bases con repeticiones moderadas que se encuentran presentes en múltiples sitios por todo el genoma, mientras que las repeticiones en tándem están restringidas a una menor cantidad de *loci* y consisten en arreglos desde 1-5 Kb (Lodish *et al.*, 2002). Las repeticiones en tándem pueden ser clasificadas de acuerdo a la longitud y al número de copias de la unidad básica de las repeticiones como así también por su localización en el genoma como: ADN satélite, minisatélites y microsatélites. El ADN satélite se encuentra principalmente en regiones subteloméricas y centroméricas en forma de heterocromatina y consiste en motivos de entre 100 y 300 pb repetidos entre 1000 y 100.000 veces. Los minisatélites son motivos de una longitud intermedia (entre 10 y 60 pb) con un menor grado de repetición que el ADN satélite y por último los microsatélites, corresponden a motivos muy cortos con un menor grado de repetición para un locus particular, sin embargo pueden consistir de motivos idénticos y ser encontrados en miles de *loci* genómicos (Weising *et al.*, 2005). Los microsatélites pueden ser estudiados mediante el desarrollo de cebadores complementarios a las regiones flanqueantes de la región repetitiva, donde las secuencias son relativamente conservadas (Finkeldey y Hattemer, 2007). Mediante el empleo de estos cebadores específicos para un locus dado es posible amplificarlos mediante la Reacción en Cadena de la Polimerasa (PCR). Los productos de amplificación son separados y sus tamaños definidos para identificar los diferentes alelos presentes en un locus polimórfico (Weising *et al.*, 2005). Los microsatélites han sido clasificados según el arreglo de los nucleótidos de los motivos repetidos como: simples perfectos, simples imperfectos, compuestos perfectos y

compuestos imperfectos. Las repeticiones perfectas son repeticiones en tándem de un único motivo de repetición, mientras que las repeticiones imperfectas están interrumpidas por motivos no repetidos. En los microsatélites compuestos, dos motivos de repetición están presentes juntos en varias configuraciones (Kalia *et al.*, 2011).

Los microsatélites son considerados por muchos autores como secuencias selectivamente neutrales distribuidas de manera aleatoria o casi aleatoria en el genoma (Li *et al.*, 2002). La abundancia de los microsatélites se halla relacionada con su tasa de mutación, la cual es elevada pudiendo llegar a 10^{-2} - 10^{-6} eventos por locus por generación y, aunque el proceso mutacional muestra diferencias entre especies, tipos de repetición, *loci* y alelos, generalmente el polimorfismo de los *loci* microsatélites se manifiesta a partir de cambios en el número de repeticiones. Existen dos mecanismos mutacionales que podrían explicar esas elevadas tasas de mutación: el deslizamiento de la polimerasa durante la transcripción y la recombinación (Li *et al.*, 2002). La génesis de los microsatélites consiste en un proceso evolutivo dinámico y complejo donde el balance entre el deslizamiento de la polimerasa, durante la transcripción, y las mutaciones puntuales son importantes en la determinación de la distribución equilibrada de los microsatélites y de su estabilidad. El deslizamiento de la polimerasa favorece al crecimiento en su longitud, mientras que las mutaciones puntuales rompen o interrumpen los arreglos de repeticiones largas en dos o más microsatélites de menor longitud (Kalia *et al.*, 2011).

La mayoría de las aplicaciones biológicas de los marcadores microsatélite o SSR (*Simple Sequence Repeat*) no necesitan del conocimiento de los mecanismos mutacionales exactos para cada locus, sin embargo varios estadísticos basados en las frecuencias alélicas (por ejemplo F_{ST} y R_{ST}) se encuentran sustentados sobre un modelo mutacional explícito (Selkoe y Toonen, 2006). Entre los modelos mutacionales se encuentra el modelo de alelos infinitos (*Infinite Allele Model* - IAM), el cual considera que cada evento mutacional crea un alelo nuevo que no estaba presente en la población. Este modelo es tradicionalmente elegido para los análisis de genética de poblaciones debido a que es el más simple y general. Por otra parte, el modelo de mutación por pasos (*Stepwise Mutation Model* – SMM) considera que cada alelo es generado por la adición y/o sustracción de motivos de repetición a los alelos ancestrales presentes en la población. Este modelo es específico para los microsatélites (Selkoe y Toonen, 2006). Sin embargo, se sabe que ocurren eventos de mutación puntual y recombinación en las regiones microsatélites. Los estadísticos empleados en el modelo SMM son muy sensibles a las violaciones del modelo, siendo generalmente más robustos y confiables los estadísticos empleados en el modelo IAM (Selkoe y Toonen, 2006).

En plantas, los microsatélites fueron clonados por primera vez en 1991, mientras que regiones microsatélites fueron descritas como marcadores por primera vez en 1992 en *Glycine max* por Akkaya *et al.* (1992). En estudios posteriores se determinó que la frecuencia promedio de microsatélites en el genoma nuclear de las plantas es de un microsatélite cada 6,3 kb, lo cual

es equivalente a lo establecido para mamíferos (Weising *et al.*, 2005). En cuanto a la composición de los motivos de repetición en el genoma nuclear de las plantas los motivos más frecuentes consisten en (A)_n, (AT)_n, (GA)_n y (GAA)_n siendo menos frecuente los microsatélites compuestos por motivos de tri, tetra o penta nucleótidos (Weising *et al.*, 2005).

Los marcadores SSR han ganado considerable importancia en los estudios genéticos en plantas debido a sus numerosos atributos, entre los cuales se incluyen su hipervariabilidad, su naturaleza multialélica, su herencia codominante, su reproducibilidad, su abundancia y cobertura extensiva por todo el genoma, además cada locus posee localización específica en el genoma y son susceptibles de ser automatizados en el laboratorio (Kalia *et al.*, 2011).

A pesar de sus múltiples cualidades, las ventajas de su utilización y la amplia aplicabilidad de los marcadores SSRs en estudios genéticos en vegetales, su desarrollo sigue siendo el principal cuello de botella para su utilización en la mayoría de las especies. Esto ocurre debido a que es necesario el aislamiento de SSRs *de novo* para la mayoría de las especies que están siendo estudiadas por primera vez (Kalia *et al.*, 2011). Tradicionalmente, los SSRs son aislados desde librerías genómicas parciales de la especie de interés mediante el rastreo de la existencia de regiones repetitivas en las secuencias obtenidas desde cientos de clones. La construcción de la librería y el rastreo de las regiones repetitivas es un trabajo laborioso, tedioso y de elevado costo económico (Kalia *et al.*, 2011). En las plantas, la estrategia más utilizada es el aislamiento de SSRs desde librerías genómicas enriquecidas con motivos repetidos; constituyendo esta la aproximación más ventajosa en el aislamiento de SSRs en varias especies vegetales (Kalia *et al.*, 2011). Los SSRs nucleares desarrollados para una especie pueden ser transferidos hacia especies cercanamente emparentadas, generalmente dentro del mismo género (Finkeldey y Hattmer, 2007).

El diseño de SSRs de utilidad para estudios genéticos requiere de la identificación de una región del genoma con una tasa de mutación suficientemente elevada para detectar polimorfismo intraespecífico y esta región debe estar localizada en una posición adyacente a una región con una tasa de mutación baja que permita la hibridación de cebadores para la mayoría de los individuos de la especie. Cuando las regiones flanqueantes a la región SSR no son lo suficientemente conservadas pueden producirse alelos nulos, es decir falta de amplificación de ciertos alelos en ciertos individuos. Una baja tasa de alelos nulos puede tener un impacto indetectable en muchos tipos de análisis, sin embargo su presencia puede ser un serio problema en análisis de parentesco (Selkoe y Toonen, 2006).

Otra desventaja asociada a los microsatélites surge de la identificación de los alelos polimórficos a partir de la estimación del tamaño de los fragmentos en pares de bases lo cual puede originar homoplasia. Este fenómeno consiste en que alelos de igual tamaño pueden no ser idénticos ya sea en su secuencia (homoplasia detectable) o ser idénticos en secuencia pero no ser idénticos por origen (homoplasia indetectable), es decir, pueden haberse originado desde

linajes distintos. La homoplasia puede reducir la diversidad alélica visible en las poblaciones y podría provocar una sobreestimación en las mediciones de flujo génico cuando las tasas mutacionales son elevadas. Sin embargo, es importante resaltar que la homoplasia indetectable afecta a las diferentes clases de marcadores moleculares (Selkoe y Toonen, 2006).

1.3 Estudios genético-poblacionales en especies forestales

En plantas, los SSRs son aplicados tradicionalmente en estudios de diversidad genética, evolutivos y de genética de poblaciones. Además, están siendo utilizados en investigaciones aplicadas que incluyen análisis genómico, mapeo de genes y selección asistida por marcadores, entre otras (Kalia *et al.*, 2011). En especies forestales, la disponibilidad de marcadores altamente polimórficos como los SSRs son particularmente útiles para los estudios de flujo génico y de los sistemas de apareamiento (Finkeldey y Hattemer, 2007).

El análisis de los patrones de variación en caracteres genéticos ambientalmente estables, requiere de la comprensión del mecanismo de transmisión del carácter desde los padres a su prole. Dos consideraciones resultan particularmente importantes en este contexto, la primera es conocer si para el tipo de marcador que estamos empleando la información genética se transmite desde ambos padres o solamente desde uno de ellos (herencia) y la segunda es conocer si el marcador empleado permite la identificación de uno o dos alelos por locus en la expresión observada del carácter (fenotipo) (Finkeldey y Hattemer, 2007). Los SSRs nucleares son heredados desde ambos parentales (herencia biparental) y son codominantes, con lo cual en los organismos diploides cada locus SSR se encuentra representado por dos alelos, cada uno heredado desde un parental (Finkeldey y Hattemer, 2007). Un individuo puede definir un genotipo homocigota si posee alelos idénticos, los cuales son identificados como una única banda en los geles de electroforesis o un único pico en los gráficos generados por el secuenciador analizador de fragmentos, o puede definir un genotipo heterocigota si posee alelos alternativos para un locus dado y son identificados como dos bandas o dos picos dependiendo de la técnica empleada para resolver los alelos.

Diferentes genes y diferentes regiones del genoma pueden presentar historias genealógicas diferentes debido a la recombinación, a la selección y a la deriva genética. Cada locus marcador sólo es representativo de una pequeña región del genoma, por ello, confiar en un único locus para estimar parámetros poblacionales crea una elevada tasa de error, de esta manera es recomendable incorporar el análisis de la mayor cantidad posible de *loci*. Esto provee una vía precisa y estadísticamente poderosa para comparar individuos y poblaciones (Selkoe y Toonen, 2006). Así, mediante el empleo de SSR nucleares se puede genotipificar a los individuos muestreados de una especie y la combinación de genotipos para distintos *loci* permite establecer genotipos multilocus. Empleando estos datos genéticos, y a partir de las teorías y ecuaciones desarrolladas en genética de poblaciones, es posible estimar parámetros e

identificar los efectos y consecuencias de los procesos microevolutivos sobre los patrones de la variación genética dentro y entre poblaciones. Las estimaciones indirectas de los niveles de flujo génico a partir de SSR nucleares se basan en las mediciones de la variación en las frecuencias génicas entre las poblaciones e infieren los niveles de flujo génico efectivo, sin embargo no permiten distinguir el flujo génico por polen del flujo génico por semillas (Nason y Hamrick, 1997).

Las especies forestales se encuentran entre los organismos más variables genéticamente tanto en regiones templadas y boreales del planeta como en las regiones tropicales. Por su parte, ciertas especies tropicales se caracterizan por niveles más bajos de variación dentro de poblaciones mientras que otras presentan una elevada diversidad, con lo cual no es posible predecir los niveles de variación genética de una especie sobre la base de sus características de historia de vida (Finkeldey y Hattemer, 2007). Las especies endémicas son, en promedio, menos variables dentro de poblaciones que las especies con amplia distribución. Además, las especies poco comunes, por ejemplo aquellas que habitan bosques tropicales en baja densidad, son menos variables que las especies que se encuentran en mayor densidad. De todas maneras, la heterocigosidad esperada es mayor para muchas especies forestales aún para las que se presentan a baja densidad (Finkeldey y Hattemer, 2007).

La estructura genética es el resultado de complejas interacciones entre diferentes factores evolutivos en las poblaciones naturales (Finkeldey y Hattemer, 2007). El grado de diferenciación genética entre poblaciones es marcadamente superior para especies tropicales que para especies de zonas templadas (Finkeldey y Hattemer, 2007). Además, las especies forestales tropicales, cuya dispersión de semillas es realizada por agentes abióticos como el viento o la gravedad, presentan una mayor diferenciación genética que aquellas cuya dispersión de semillas ocurre por animales (Finkeldey y Hattemer, 2007).

1.4 Estimación del índice de fijación (F_{ST})

El índice de fijación F_{ST} es una de las mediciones de la estructura poblacional más comúnmente utilizada en estudios de genética evolutiva. Este índice mide la relación entre el flujo génico y la deriva genética y permite conocer cómo influye esta relación sobre la estructura genética (Templeton, 2006). Valores elevados de F_{ST} indican que existe poca variación genética en las poblaciones locales en relación a la población total, mientras que valores bajos de F_{ST} indican mucha variación local en relación a la total (Templeton, 2006).

La distribución de la variación genética neutral entre poblaciones está establecida por el balance entre la deriva genética, la cual promueve a la diferenciación entre poblaciones, y el flujo génico, el cual reduce o atenúa la divergencia (Broquet y Petit, 2009).

En la mayoría de las especies forestales, aplicando marcadores de origen biparental, se han observado valores bajos de F_{ST} los cuales indican que el flujo génico en el pasado ha sido intenso (White *et al.*, 2007).

Por otro lado, el índice F_{ST} se puede definir en términos de varianza en las frecuencias alélicas entre poblaciones locales, midiendo así la reducción de la variabilidad alélica al pasar de un nivel jerárquico de estructuración poblacional superior a uno inferior (Hartl y Clark, 2007). Además, este índice de fijación depende del tamaño y de la historia poblacional, pero es independiente de aspectos tales como los relacionados al esquema de muestreo, al número de alelos por locus, al número de individuos muestreados por población y/o al número de poblaciones muestreadas (Weir y Cockerham, 1984).

También, el índice F_{ST} puede ser definido en términos de identidad por descendencia a partir del tiempo de coalescencia de los genes (Slatkin, 1991). Así, este índice puede ser visto como una descripción del tiempo relativo hasta el antecesor más reciente para alelos tomados dentro y entre poblaciones (Whitlock, 2011). Si el flujo génico es restringido entre poblaciones locales, el tiempo de coalescencia de dos genes tomados al azar desde una población local será menor al tiempo de coalescencia de dos genes tomados al azar desde diferentes poblaciones locales de una misma especie (Templeton, 2006). De esta manera, valores elevados indican que el tiempo hasta el ancestro común de dos alelos tomados desde diferentes poblaciones es mayor al tiempo hasta el ancestro común de dos alelos tomados desde la misma población (Slatkin, 1991). Por ello, el índice F_{ST} así estimado puede considerarse una medición del grado en el cual las poblaciones son históricamente independientes (Holsinger, 1999) y provee una medición del aislamiento evolutivo de las poblaciones, ya que las propiedades de este índice no dependen del modelo de islas o de supuestos demográficos (Whitlock, 2011). El índice de fijación estimado de esta manera es llamado índice de fijación coalescente y permite analizar la historia demográfica de las poblaciones sin ser afectado por la mutación (Whitlock, 2011).

Finalmente, puede estimarse un índice de fijación F_{ST}^i para cada población local empleando un enfoque bayesiano jerárquico. Este índice mide el nivel de divergencia de la población local i respecto de la población total. El modelo considerado para realizar estas inferencias asume que las poblaciones subdivididas se encuentran en equilibrio migración-deriva genética donde la proporción de migrantes puede diferir entre las poblaciones locales pero todos los migrantes provienen del mismo pool de migrantes (Foll y Gaggiotti, 2006).

2. Objetivos específicos

- Desarrollar marcadores microsatélites nucleares específicos para *A. colubrina* var. *cebil*.
- Caracterizar la diversidad genética nuclear en poblaciones naturales de *A. colubrina* var. *cebil* del Norte argentino.
- Analizar la estructura genética en las poblaciones fragmentadas de *A. colubrina* var. *cebil* a partir de marcadores microsatélites de ADN nuclear.
- Indagar acerca del rol evolutivo del flujo génico y de la deriva genética como fuerzas modeladoras de las frecuencias génicas en las poblaciones fragmentadas.

3. Materiales y Métodos

3.1 Desarrollo de SSR específicos para *Anadenanthera colubrina* var. *cebil*

La hipervariabilidad, codominancia y elevada repetitividad hacen de los marcadores microsatélites (SSRs) una herramienta útil para los estudios genéticos poblacionales, sin embargo presentan una gran desventaja ya que deben ser aislados *de novo* en especies estudiadas por primera vez (Zane *et al.*, 2002). La obtención de SSRs nucleares puede ser una tarea costosa en cuanto a tiempo y esfuerzo dado que los métodos tradicionales consisten en la búsqueda de regiones repetidas en tándem dentro de librerías genómicas mediante la aplicación de pruebas apropiadas (Zane *et al.*, 2002). Esta es la principal razón por la cual las especies nativas, las especies no modelo y las especies de bajo interés económico, generalmente carecen de marcadores SSRs específicos dificultando de esta manera la realización de estudios genéticos poblacionales.

El desarrollo de marcadores SSRs *de novo* específicos para *A. colubrina* var. *cebil* se llevó a cabo durante una pasantía en el Instituto de Genética y Mejoramiento Forestal de la Universidad Georg August en Göttingen (Alemania), esta pasantía corresponde a una de las actividades complementarias de postgrado realizadas como requisito para optar el grado de Doctora en Ciencias Naturales. La obtención de los SSRs específicos se realizó siguiendo el protocolo descrito por Fisher y Bachmann (1998) modificado por Prinz *et al.* (2009). Se construyeron dos librerías de ADN genómico enriquecidas con motivos microsatélites para obtener los marcadores SSRs específicos para *A. colubrina* var. *cebil*. Ambas librerías se construyeron empleando ADN genómico de un individuo proveniente de Jujuy.

Se extrajo ADN genómico total empleando el kit de extracción DNeasy[®] – Plant minikit (Qiagen, Hilden, Germany). El ADN extraído fue digerido con la enzima de restricción *RsaI* y se ligaron pares de adaptadores a los extremos romos de los fragmentos generados por esta enzima. Los productos de restricción-ligación fueron desnaturalizados a altas temperaturas e hibridados con oligonucleótidos biotinilados.

En la librería A se empleó un único oligonucleótido (GA)₁₀ y una temperatura de hibridación de 60°C; mientras que en la librería B se empleó un pool equimolar de oligonucleótidos (GAA)₈, (CAA)₈ y (CA)₁₀ (un picomol de cada uno) y una temperatura de hibridación de 74°C.

La captura de los oligos marcados con biotina hibridados a los fragmentos de ADN se realizó mediante el empleo de perlas magnéticas recubiertas con estreptavidina (Invitrogen, Carlsbad, CA). Los fragmentos capturados fueron amplificados mediante la reacción en cadena de la polimerasa (*Polimerase Chain Reaction* – PCR) utilizando las secuencias de los adaptadores como cebadores. El proceso de enriquecimiento incluyendo la hibridación, captura y amplificación, fue repetido una vez más para cada librería.

Los productos de amplificación fueron purificados y luego ligados en un vector pCR[®]2.1 - TOPO[®] (Invitrogen). Estos vectores fueron empleados para transformar células competentes de *Escherichia coli* One shot[®] TOP10 (Invitrogen). Se seleccionaron las colonias positivas, se repicaron y el ADN fue amplificado y secuenciado en un secuenciador automático ABI Prism[®] 3100 (Applied Biosystems, Forster City, CA) (Instituto de Genética y Mejoramiento Forestal de la Universidad Georg August) (Figura II-1).

En total, se secuenciaron 204 clones positivos desde ambas librerías, 62 de las 97 secuencias que presentaron regiones microsátélites mostraron elevada calidad y sus regiones flanqueantes presentaron longitud suficiente para el posterior diseño de los cebadores específicos.

Se diseñaron 30 pares de cebadores empleando el programa Primer3 versión 2.2.3 (Rozen y Skaletsky, 2000). Para ello se ensayaron las características habituales, a saber 20 a 80 % de contenido de CG y temperatura de hibridación entre 57 y 63°C, priorizándose los pares de cebadores que generaron fragmentos de una longitud entre 80 y 200 pb dependiendo del motivo de repetición del microsátélite y de la longitud total del fragmento secuenciado desde los clones.

Se realizaron pruebas de funcionalidad para los 30 pares de cebadores diseñados. Las reacciones de amplificación se realizaron en un volumen final de 15 µl empleando 0,5 ng/µl de ADN genómico, 1x de Buffer Hot Start, 2,5 mM de Cl₂Mg, 0,2 mM de cada dNTP, 1 U de enzima ADN polimerasa Hot Start (5 U/µl FirePol Solis BioDyne, Tartu, Estonia) y 0,33 pmol de cada cebador. La reacción de amplificación fue realizada en un termociclador con gradiente (Biometra, Göttingen, Alemania) empleando un programa de tipo *touchdown*. Se ensayaron las siguientes condiciones de ciclado: desnaturalización inicial: 94°C por 15 min, 10 ciclos de desnaturalización 1 min a 94°C, hibridación 1 min en un rango de temperatura de 60°C a 50°C con una reducción de 1°C en cada ciclo y elongación por 1 min a 72 °C seguidos por 29 ciclos similares empleando una temperatura de hibridación de 50°C, finalizando con una elongación por 20 min a 72°C.

Los productos de amplificación fueron resueltos en geles de agarosa al 1,5% teñidos con Bromuro de Etidio y se definieron los tamaños de los alelos visualmente mediante la comparación con un marcador de peso molecular de 100 pb como referencia. Dieciséis *loci* mostraron bandas únicas, bien definidas y del tamaño esperado. Los 14 *loci* restantes mostraron patrones de amplificación inespecíficas.

Los productos de amplificación de los 16 *loci* fueron clonados, resecuenciados y alineados con los fragmentos originales obtenidos en el proceso de enriquecimiento. Las secuencias obtenidas desde 15 de estos *loci* mostraron complementariedad con las secuencias originales mientras que la restante no lo hizo.

Una muestra de 20 individuos fue genotipificada empleando los 15 SSRs desarrollados para *A. colubrina* var. *cebil*. Para ello, los cebadores fueron resintetizados incorporando una

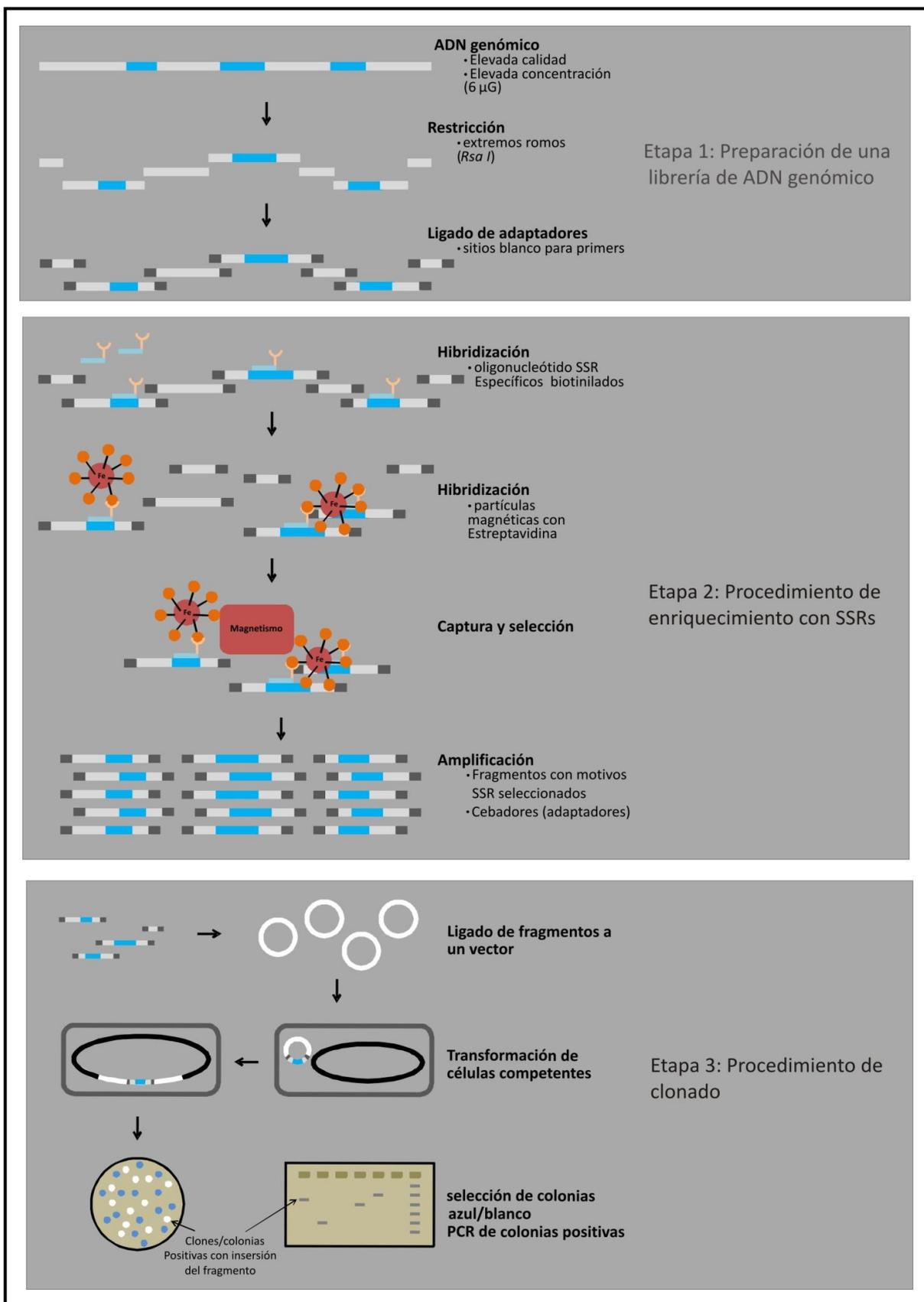


Figura II-1: Esquema del proceso de desarrollo de los marcadores SSRs específicos para *Anadenanthera colubrina* var. *cebil*.

marca fluorescente (FAM=azul o HEX=verde) a su extremos 5' y las condiciones de amplificación fueron mismas a las empleadas para las pruebas de funcionalidad. Los *loci* *Ac28.3* y *Ac157.1* fueron amplificados sin emplear el programa *touchdown*, siendo la temperatura de hibridación 65°C. Se asignó el tamaño de los fragmentos empleando el secuenciador ABI Prism® 3100 (Applied Biosystems) mediante el empleo del programa GENESCAN™ empleando un marcador de peso estándar GS 500 ROX™ (Applied Biosystems). Seis de los 15 *loci* mostraron patrones inconsistentes presentando tres o cuatro picos por ello sólo se consideraron nueve *loci* para los análisis posteriores. Las secuencias de estos *loci* fueron incluidas en la base pública de datos *GenBank* con los siguientes números de acceso: JQ086537, JQ086538, JQ086539, JQ086540, JQ086541, JQ086542, JQ086543, JQ086544, JQ806379. Los cebadores *Ac41.1* y *Ac172.1* y los cebadores *Ac34.3*, *Ac48.1* y *Ac11.2* pudieron ser amplificados de manera conjunta en una única reacción (PCR-*multiplex*) debido a que los tamaños de los alelos de estos *loci* presentaron rangos de tamaños diferentes o pudieron ser distinguidos empleando diferentes marcas fluorescentes. En el locus *Ac48.1* el alelo de 171 pb se registró una única vez, por ello algunos individuos debieron ser amplificados de manera independiente para confirmar los genotipos asignados. De los nueve cebadores desarrollados, ocho fueron polimórficos en los individuos analizados mientras que el restante fue monomórfico.

3.2 Análisis de la utilidad de los marcadores desarrollados para estudios genético-poblacionales

El número total de alelos y la heterocigosidad observada y esperada por locus fue determinada empleando el programa Genalex 6.41 (Peakall y Smouse, 2006) disponible en http://www.anu.edu.au/BoZo/GenAlEx/new_version.php. El equilibrio de Hardy Weinberg por locus para cada población fue testado mediante pruebas de chi-cuadrado empleando el programa PopGen versión 1.32 disponible en <http://www.ualberta.ca/~fyeh/> (Yeh y Boyle, 1997). La presencia de alelos nulos y/o posibles problemas en la genotipificación de los individuos por locus fue analizada para cada población empleando el programa Microchecker versión 2.2.3 (Van Oosterhout *et al.*, 2004) disponible en <http://www.microchecker.hull.ac.uk/>. El desequilibrio de ligamiento fue testado para todos los pares de *loci* en cada población y se realizó un test de probabilidad empleando el algoritmo de Montecarlo basado en Cadenas de Markov (*Markov Chain Monte Carlo* - MCMC). Para finalizar, se realizó una prueba global entre las poblaciones mediante el método de Fisher empleando el programa GENEPOP versión 4.0.10 (Raimond y Rousset, 1995).

3.3 Genotipificación de los individuos empleando SSRs

El ADN fue extraído empleando el kit DNeasy[®] – Plant minikit (Qiagen). Se emplearon ocho pares de cebadores universales desarrollados para este estudio: *Ac34.3*, *Ac48.1*, *Ac11.2*, *Ac28.3*, *Ac157.1*, *Ac41.1*, *Ac172.1* y *Ac162.1*.

Las reacciones de amplificación se realizaron en un volumen final de 15 µl empleando 0,5 ng/µl de ADN, 1x de Buffer Hot Start, 2,5 mM de Cl₂Mg, 0,2 mM de cada dNTP, 1 U de enzima ADN polimerasa Hot Start (5 U/µl FirePol Solis BioDyne, Tartu, Estonia) y 0,33 pmol de cada cebador. La reacción de amplificación fue realizada en un termociclador con gradiente (Biometra, Göttingen, Alemania). Las condiciones de ciclado empleadas fueron: Desnaturalización inicial: 94°C por 15 min, 10 ciclos de desnaturalización 1 min a 94°C, hibridación 1 min en un rango de temperatura de 60 °C a 50°C con una reducción de 1°C en cada ciclo y elongación por 1 min a 72 °C seguidos por 29 ciclos similares empleando una temperatura de hibridación de 50 °C.

El tamaño de los fragmentos fue asignado empleando el secuenciador ABI Prism[®] 3100 (Applied Biosystems) mediante el empleo del programa GENESCAN[™]. Para ello se empleó un marcador de peso estándar GS 500 ROX[™] (Applied Biosystems).

3.4 Análisis estadístico de los datos

3.4.1 Caracterización de la diversidad genética nuclear

Se calcularon las frecuencias alélicas por locus para ambas regiones y se identificaron los alelos únicos por región.

Se estimaron diferentes parámetros para caracterizar la diversidad genética nuclear:

Número promedio de alelos por locus (N_A): también conocida como multiplicidad génica y corresponde al promedio de los distintos alelos identificados para cada locus en cada población (Gillet *et al.*, 2005).

Número efectivo de alelos por locus (N_E): permite determinar el número de alelos presentes en altas frecuencias o alelos prevalentes. Fue calculado de la siguiente manera $N_E = \frac{1}{\sum_1^n x_i^2}$ donde x_i corresponde a las frecuencias de alélicas (Gillet *et al.*, 2005).

Número de alelos únicos (N_{AU}): corresponde al promedio de los alelos detectados en una única población para cada locus (Barton y Slatkin, 1986; Slatkin y Barton, 1989).

Heterocigosidad observada (H_O): corresponde a la frecuencia promedio de heterocigotas (Hartl y Clark, 2007).

Heterocigosidad esperada (H_E): corresponde al índice de diversidad génica de Nei $h = 1 - \sum_1^n x_i^2$; donde x_i corresponde a las frecuencias alélicas (Nei, 1987). Permite conocer la proporción de *loci* heterocigotas por individuo esperados en equilibrio de Hardy-Weinberg (Trapnell y Hamrick, 2004). La diversidad genética cuantificada mediante la heterocigosidad

esperada (H_E) permite conocer la variación genética a nivel de individuos (Young *et al.*, 1996). Esta medida corresponde a la proporción de individuos heterocigotas en un locus dado, estando limitada por la multiplicidad génica, por las frecuencias alélicas y por los niveles de endogamia (Young *et al.*, 1996; Gillet *et al.*, 2005). Estas estimaciones se realizaron empleando el programa GenAlEx versión 6.41 (Peakall y Smouse, 2006).

También se estimaron los siguientes parámetros:

Riqueza alélica (R): Permite predecir la multiplicidad alélica esperada si las muestras fueran de igual tamaño (Gillet *et al.*, 2005).

Riqueza de alelos únicos (R_{AU}): Permite predecir el número de alelos únicos esperados si dos muestras fueran de igual tamaño (Szpiech *et al.*, 2008). Las estimaciones de riqueza alélica permiten independizarse del sesgo que puede introducir el tamaño de la muestra ya que tanto el número promedio de alelos por locus, el número efectivo de alelos por locus como el número promedio de alelos únicos por locus dependen del tamaño de la muestra (Gillet *et al.*, 2005).

Para ambos cálculos se empleó el método de rarefacción. Su implementación requiere seleccionar un tamaño poblacional estándar (g o tamaño de rarefacción) el cual debe ser igual o inferior al tamaño de la población de menor tamaño (Szpiech *et al.*, 2008) por ello se consideró $g=14$ dado que Tucumán presenta el menor tamaño muestral igual a 14 individuos. Estas estimaciones se realizaron empleando el programa ADZE versión 1.0 (Szpiech *et al.*, 2008) disponible en <http://www.stanford.edu/group/rosenberglab/adzeDownload.html>.

Para determinar si existen diferencias estadísticamente significativas entre los parámetros de diversidad estimados entre las poblaciones se aplicaron pruebas de t pareadas (Nei, 1987; Ndiade-Borobou *et al.*, 2010). Se realizaron comparaciones pareadas por locus y para el promedio de todos los *loci* implementando el programa de aplicación web GraphPad disponible en <http://www.graphpad.com/quickcalcs/ttest1.cfm>.

3.4.2 Análisis de las relaciones genéticas entre los genotipos multilocus

Se construyó una matriz de distancia genotípica individuo x individuo ($N \times N$) para calcular la distancia genética existente entre los individuos tomados de a pares.

La matriz se construyó considerando un locus simple con los siguientes alelos: i -th, j -th, k -th y l -th, y se definieron las distancias cuadráticas como: $d^2(ii, ii) = 0$, $d^2(ij, ij) = 0$, $d^2(ii, ij) = 1$, $d^2(ij, ik) = 1$, $d^2(ij, kl) = 2$, $d^2(ii, jk) = 3$, y $d^2(ii, jj) = 4$. Luego las distancias genéticas fueron sumadas a través de los *loci* bajo el supuesto de independencia empleando el programa GenAlEx versión 6.41 (Peakall y Smouse, 2006).

A partir de esta matriz de distancia genotípica se construyó un árbol no enraizado empleando el algoritmo Neighbor-Joining (NJ) mediante la implementación del programa Mega versión 4 (Tamura *et al.*, 2007) disponible en <http://www.megasoftware.net/>.

3.4.3 Análisis de la estructura genética de las poblaciones

La estructuración genética medida a partir del empleo de marcadores neutrales es una consecuencia de la magnitud de la deriva genética que se encuentra operando, del tamaño efectivo y/o del grado conjunto de aislamiento geográfico o ecológico de cada población local (Foll y Gaggiotti, 2006).

3.4.3.1 Distribución de la diversidad genética entre y dentro de poblaciones

El Análisis de la Varianza Molecular (AMOVA) describe la distribución de la variación genética dentro y entre las poblaciones testando su significancia estadística. Este análisis parte de la hipótesis nula de ausencia de estructuración de la población (Stewart y Excoffier, 1996) siendo su idea central convertir una matriz de distancia entre individuos en su equivalente análisis de varianza (Huff *et al.*, 1993).

El AMOVA permite estimar los índices de la estructura genética usando información del contenido alélico de los haplotipos y sus frecuencias reflejando la correlación de la diversidad haplotípica a diferentes niveles jerárquicos de subdivisión (Excoffier *et al.*, 1992).

La significancia de los componentes de covarianza asociados con los diferentes niveles de estructura genética, ya sea entre regiones, entre poblaciones dentro de regiones o dentro de poblaciones es testada mediante procedimientos de permutación no paramétricos (Excoffier *et al.*, 1992).

El análisis de las diferencias en los contenidos alélicos entre haplotipos es efectuado a partir de una matriz de distancia obtenida mediante la aplicación del índice de distancia Euclidiana entre todos los pares de haplotipos. Se asume que este análisis se desarrolla sobre segmentos no recombinantes. Para N individuos se identifican S sitios polimórficos. Un haplotipo es definido como una combinación de alelos SSRs y puede ser considerado como un vector de S dimensiones (Excoffier *et al.*, 1992).

La diferencia entre dos haplotipos h_j y h_k es definida como $(p_j - p_k)$. Cada sitio polimórfico contribuye con información adicional, sin ser necesariamente independientes evolutivamente. Se define una distancia métrica Euclidiana (δ_{jk}^2) entre los haplotipos h_j y h_k como:

$$\delta_{jk}^2 = (p_j - p_k)' W (p_j - p_k)$$

donde W es una matriz de diferencias para los distintos sitios. Esta matriz W puede tomar diversas formas, dependiendo del análisis que quiera realizarse. Si se supone que todos los sitios son independientes e igualmente informativos, $W = I$, es decir será una matriz de identidad, donde la distancia métrica es igual al resultado de las diferencias entre los alelos SSRs (Excoffier *et al.*, 1992).

Para particionar la matriz de distancia se considera un sistema genético haploide donde las distancias inter-haplotípicas son idénticas a las distancias entre individuos. Se organiza un

grupo de N individuos desde I poblaciones en una matriz de distancia D^2 , particionada en una serie de submatrices correspondientes a las subdivisiones particulares.

Cuando los individuos son agrupados en poblaciones, y las poblaciones agrupadas en regiones definidas *a priori* sobre un criterio no genético, corresponde el modelo lineal:

$$P_{jig} = p + a_g + b_{ig} + c_{jig}$$

donde P_{jig} representa al individuo j ($j = 1, \dots, N_{ig}$) en la i población ($i = 1, \dots, I_g$) y p es la esperanza no conocida de P_{jig} , los efectos son a para regiones, b para poblaciones y c para individuos dentro de poblaciones. Se asume que los efectos son aditivos, aleatorios, no correlacionados y que están asociados a componentes de varianza (desvíos cuadráticos esperados) σ_a^2 , σ_b^2 y σ_c^2 , respectivamente.

Para una distribución jerárquica de N individuos en distintos niveles, se puede particionar la suma total de los desvíos cuadráticos en sus componentes de variación de la siguiente forma:

$$SSD_{(Total)} = SSD_{(WP)} + SSD_{(AP/WP)} + SSD_{(AG)}$$

donde $SSD_{(WP)}$ = es el desvío cuadrático esperado dentro de poblaciones, $SSD_{(AP/WP)}$ = es el desvío cuadrático esperado entre poblaciones dentro de regiones y $SSD_{(AG)}$ = es el desvío cuadrático esperado entre regiones. Considerando que W indica “dentro” (*within*), A indica “entre” (*among*), P = poblaciones y G = regiones. Para este modelo lineal, corresponde el siguiente diseño de AMOVA:

Fuente de variación	Grados de libertad	Suma de cuadrados (SSD)	Cuadrados medios esperados
Entre regiones	G-1	SSD(AG)	$n \sigma_a^2 + n \sigma_b^2 + \sigma_c^2$
Entre poblaciones dentro de regiones	P-G	SSD(AP/WG)	$n \sigma_b^2 + \sigma_c^2$
Dentro de poblaciones	N-P	SSD(WP)	σ_c^2
Total	N-1	SSD(T)	$\sigma^2 T$

Los coeficientes n de la tabla representan los tamaños promedio de las muestras de cada nivel jerárquico, teniendo en cuenta los tamaños de muestra desiguales.

Se testa σ_c^2 y F_{ST} por permutación de los haplotipos entre subpoblaciones y entre poblaciones. Este análisis fue desarrollado utilizando el software Arlequín 3.5 (Excoffier y Lischer, 2010) disponible en <http://cmpg.unibe.ch/software/arlequin35/>.

Este análisis fue realizado de manera global empleando dos modelos definidos atendiendo diferentes niveles de subdivisión jerárquica para analizar la distribución de la variación genética:

Modelo 1: Entre regiones; entre poblaciones dentro de regiones y dentro de poblaciones

Modelo 2: Entre regiones; entre poblaciones dentro de regiones; dentro de poblaciones y dentro de individuos.

A partir de este último modelo se estimó el índice F_{IS} , el cual permite inferir estructuración genética como consecuencia de los efectos de la endogamia, es decir del cruzamiento entre individuos más emparentados que la media de la población (Hartl y Clark, 2007).

3.4.3.2 Estimación del índice de fijación F_{ST}

El índice de fijación de Wright F_{ST} definido en términos de varianza fue calculado para todos los *loci* (F_{ST} global) a partir del AMOVA según la ecuación

$$F_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2} \text{ (Weir y Cockerham, 1984)}$$

donde la varianza molecular total (σ_T^2) es la sumatoria de los componentes de covarianza debido a diferencias haplotípicas dentro de poblaciones (σ_c^2), los componentes de covarianza debido a diferencias haplotípicas entre diferentes poblaciones dentro de la misma región (σ_b^2), y los componentes de covarianza debido a diferencias haplotípicas entre poblaciones (σ_a^2). La significancia estadística del índice de fijación fue testada con un valor obtenido mediante 1023 permutaciones de los haplotipos dentro de subpoblaciones y entre regiones para el 95% de confianza ($p < 0,05$). Además, bajo esta definición se estimó el índice de fijación para cada región tomando a las poblaciones como unidad jerárquica menos inclusiva.

El índice de fijación tiene un mínimo teórico de cero, el cual indica ausencia de divergencia genética y un máximo teórico de uno, que indica fijación de alelos alternativos en diferentes poblaciones (Hartl y Clark, 2007).

El índice de fijación F_{ST} definido en términos de identidad por descendencia fue estimado para las poblaciones tomadas de a pares según la ecuación: $F_{ST} = \frac{\bar{t} - \bar{t}_0}{\bar{t}}$ (Slatkin, 1991) donde \bar{t}_0 es el tiempo de coalescencia promedio de dos genes tomados al azar desde la misma subpoblación, mientras que \bar{t} es el tiempo de coalescencia promedio de dos genes tomados al azar de la especie en su totalidad (Templeton, 2006). El índice de fijación para el total de las poblaciones se estimó a partir del promedio de los valores de a pares (F_{ST} global) y se estimó el índice de fijación para cada región tomando a las poblaciones como unidad jerárquica menos inclusiva.

Estos análisis fueron realizados empleando el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

El índice F_{ST} por población fue estimado mediante inferencia bayesiana. Para la estimación de los parámetros del modelo se emplea una combinación del algoritmo de Montecarlo basado en Cadenas de Markov (*MCMC*) y el algoritmo de Montecarlo basado en

cadena reversible de Markov (*Reversible Jump Markov Chain Monte Carlo - RJMCMC*), algoritmos que permiten estimar la probabilidad *a posteriori* de los modelos y sus parámetros (Foll y Gaggiotti, 2006). Para la estimación de los índices de fijación por población se empleó el programa BIMr Versión 1.1 (Faubet y Gaggiotti, 2008) disponible en <http://www-leca.ujf-grenoble.fr/logiciels.htm>.

Para la implementación del modelo se empleó una matriz de distancia geográfica entre poblaciones, estos datos son introducidos en el análisis a través de la distribución *a priori* de los índices F_{ST} específicos para cada población, mientras que los datos genéticos son introducidos en las funciones de probabilidad. Para la estimación de la probabilidad *a posteriori* la cadena de Markov Montecarlo (MCMC) fue corrida por un total de 8×10^6 iteraciones donde las primeras 10^6 correspondieron al período de *burn-in*, es decir un período previo al muestreo que le permite a la cadena alcanzar el período estacionario. Además, se corrieron 20 pruebas piloto (*pilot runs*) de una longitud de 10^5 para contribuir a alcanzar el estado estacionario de la cadena. Las restantes 5×10^6 iteraciones presentaron una frecuencia de muestreo en la cadena (*thinning interval*) de 100, lo cual da un tamaño de muestreo (*sample size*) igual a 5×10^4 para cada análisis. El grado de incertidumbre de los estimadores fue medido mediante el establecimiento de los límites del menor intervalo que contenga el 95% de los valores (*Highest Probability Density Interval – HPDI*). Cada corrida fue repetida 5 veces de manera paralela para evitar problemas de convergencia. Para obtener las estimaciones del parámetro se seleccionó la corrida que presentó la menor desviación.

El índice F_{ST} puede ser sobreestimado en caso de presencia de alelos nulos en los *loci* analizados (Chapuis y Estoup, 2007) y puede ser subestimado cuando la heterocigosidad dentro de las poblaciones es elevada (Jost, 2008) situación frecuente en las especies forestales al emplear SSRs nucleares. Así, se controló el efecto potencial de la presencia de alelos nulos sobre la diferenciación genética mediante el cálculo de los valores de F_{ST} empleando el método de exclusión de alelos nulos (*excluding null allele - ENA*) implementado en el programa FreeNA (Chapuis y Estoup, 2007) disponible en <http://www.ensam.inra.fr/URLB/>.

La presencia de elevados niveles de diversidad genética puede ocultar la existencia de estructuración genética debido a la distorsión que puede provocar en la estima del índice F_{ST} tradicionalmente utilizado (Jost, 2008) el cual es una medida explícita de la diferenciación relativa entre poblaciones (Whitlock, 2011) y es independiente de los niveles de diversidad genética (H_S). Este análisis se realizó empleando el programa SMOGD versión 1.2.5 (Crawford, 2010) disponible en <http://www.ngcrawford.com/django/jost/>.

Las estimaciones de F_{ST} se basan en un número limitado de individuos tomados desde algunas poblaciones los cuales son genotipificados para un número reducido de *loci*. Debido a

que F_{ST} es básicamente la relación entre dos varianzas, este índice difícilmente es estimado de manera exacta sin contar con gran cantidad de datos (Whitlock y McCauley, 1999).

3.3.3.3 Determinación de la estructura genética mediante análisis bayesiano

Esta aproximación involucra el muestreo de ADN desde los miembros de un número potencial de grupos genéticos o *clusters* y la estimación de las frecuencias alélicas en cada *cluster* para una serie de *loci* no ligados. A partir de las frecuencias alélicas es posible calcular la probabilidad del origen de un genotipo dado en cada *cluster*. De esta manera, individuos de origen desconocido pueden ser asignados a *clusters* de acuerdo a esta probabilidad.

Los estudios poblacionales parten, generalmente, de la definición subjetiva del número de poblaciones. En el presente estudio se definió el número de poblaciones en base a la localización geográfica de los individuos muestreados. Sin embargo, resulta de interés el conocer si las poblaciones definidas bajo criterios subjetivos representan poblaciones en términos genéticos. Debido a esto, los métodos de asignación resultan de utilidad para confirmar si la clasificación subjetiva es consistente con la información genética y si resulta apropiada para el estudio de otras cuestiones de interés (Pritchard *et al.*, 2000). Este método puede ser aplicado a varios tipos de marcadores (microsatélites o SSRs, polimorfismos de los fragmentos de restricción o RFLPs y polimorfismos de base única o SNPs). Entre sus supuestos asume que los *loci* presentan equilibrio de ligamiento y condiciones de equilibrio de Hardy-Weinberg dentro de poblaciones (Pritchard *et al.*, 2000)

Es posible definir el número de *clusters* y asignar individuos, probabilísticamente a estos, implementando métodos de agrupamiento mediante inferencia bayesiana. Se asume un modelo en el cual existen k *clusters*, y cada uno de ellos está caracterizado por determinadas frecuencias alélicas en cada locus (Pritchard *et al.*, 2000). Mientras que los estimadores tradicionales en genética de poblaciones pueden ser calculados empleando cálculos analíticos simples, los métodos actuales de análisis dependen en gran medida del poder computacional, esta dependencia se acentúa principalmente, para los métodos basados en *maximum likelihood* o inferencia bayesiana (Pearse y Crandall, 2004).

La estructura genética poblacional fue estudiada mediante algoritmos de agrupamiento bayesiano para determinar el número más probable de *clusters* (k) en los cuales se agrupan los individuos empleando únicamente información genética (genotipos multilocus) asignando individuos a estos *clusters*. Para este análisis se empleó un modelo de tipo *admixture* (*admixture ancestry model*), el cual asume que la constitución genética de un individuo puede originarse desde diferentes poblaciones y se consideró que las frecuencias alélicas son independientes entre las poblaciones (*Independent frequency model*). Este análisis requiere de dos etapas: la primera atiende cuestiones relacionadas a la selección del modelo y a la determinación del número k más apropiado para la interpretación de los datos, en tanto que la segunda consiste en

el análisis del agrupamiento de los individuos dentro de cada grupo genético para el número k inferido.

Para estimar el valor de k se asignaron valores desde 1 hasta 6 con 10 repeticiones independientes para cada valor de k . Se obtuvieron los resultados a partir de corridas de MCMC de 750.000 iteraciones luego de un período de *burn-in* de 500.000 iteraciones. Para la determinación de la longitud del período de *burn-in* (m) se estimó el $\log[Pr(X|P^{(m)}, Q^{(m)})]$ y otros estadísticos en una serie de ensayos, empleando distintos valores de m , hasta establecer el tiempo requerido por la cadena para alcanzar un estado aproximadamente estacionario. Cuando no se detectaron diferencias sustanciales entre distintas corridas para el mismo valor de k se consideró estar en condiciones de obtener estimaciones precisas. Para el desarrollo de este análisis se empleó el programa STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) disponible en http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/html/structure.html.

Para estimar el número real de k se implementó el método de Evanno. Evanno *et al.* (2005) propusieron una medida *ad hoc* Δk , la cual representa la tasa de cambio de segundo orden del $\ln[Pr(X|k)]$ con respecto a k y provee un estimador del número de agrupamientos o *clusters*. Mediante estudios de simulación, Evanno *et al.* (2005) demostraron que la estimación de k a partir de la probabilidad *a posteriori*, $Pr(X|k)$, a menudo no coincide con el número real de k , por el contrario el valor de k que presentó el mayor Δk coincidió con el valor real de k . En el presente estudio, el número real de k fue obtenido mediante la estimación de los valores de Δk para los diferentes valores de k ($k=1-6$). Para ello, se partió de los resultados obtenidos mediante 10 corridas independientes del programa STRUCTURE 2.3.3. Las estimaciones de Δk fueron realizadas empleando la aplicación web STRUCTURE HARVESTER versión 0.6.92 (Earl y vonHoldt, 2011) disponible en <http://taylor0.biology.ucla.edu/structureHarvester/>. Además, se analizó la estructura genética mediante inferencia bayesiana para cada región de manera independiente para conocer si existe subestructuración genética dentro de las regiones analizadas.

3.4.4 Cuantificación del flujo génico

3.4.4.1 Cuantificación del flujo génico histórico

Los métodos indirectos para la cuantificación de los niveles de flujo génico permiten inferir el número de migrantes que han sido intercambiados históricamente por generación ($N_e m$) a partir del nivel de diferenciación genética existente entre poblaciones (Burczyk *et al.*, 2004). Slatkin y Barton (1989) propusieron distintos métodos para estimar la importancia relativa del flujo génico y de la deriva genética empleando la distribución de las frecuencias alélicas. Entre ellos se encuentra el método basado en el estadístico F_{ST} de Wright y el método basado en alelos raros (Slatkin y Barton, 1989).

Método basado en el estadístico F_{ST} : Wright (1951) demostró que, bajo el modelo de islas, $F_{ST} \approx \frac{1}{1+4N_e m}$; si $m \ll 1$. Desde esta fórmula podemos estimar $N_e m = \frac{1}{4} \left(\frac{1}{F_{ST}} - 1 \right)$. La mayoría de los estudios en los cuales se cuantifica el flujo génico publicados hasta la fecha están basados en las relaciones entre Nm y F_{ST} (Broquet y Petit, 2009).

Debido a las diferentes concepciones subyacentes a las estimas del índice de fijación, ya sea en términos de varianzas en las frecuencias alélicas o en términos de identidad por descendencia, es posible estimar el número efectivo de migrantes basado en el tamaño efectivo en términos de varianza o en el tamaño efectivo en términos de identidad por descendencia, respectivamente. De esta manera, estas estimaciones reflejarán la magnitud de la fragmentación o de la identidad por descendencia sobre los niveles de flujo génico.

Se estimó el nivel de flujo génico global $N_{ev}m$ a partir de la relación $F_{ST} \approx \frac{1}{1+4N_{ev}m}$, donde N_{ev} es el tamaño efectivo en términos de varianza (*variance effective size*) el cual permite describir los efectos del error de muestreo y/o de la subdivisión poblacional como consecuencia de la deriva genética a partir de la medición de la varianza en las frecuencias alélicas a través de las generaciones (Templeton, 2006).

Se estimó el nivel de flujo génico global $N_{ef}m$ a partir de la relación $F_{ST} \approx \frac{1}{1+4N_{ef}m}$, donde N_{ef} es el tamaño efectivo en términos de endogamia (*inbreeding effective size*) el cual permite describir la acumulación de identidad por descendencia (\bar{F}) en una población como consecuencia de la deriva genética (Templeton, 2006).

La estimación de los niveles de flujo génico desde datos genéticos, debe ser realizada con cautela, principalmente porque estos métodos asumen supuestos acerca de las poblaciones que son poco probables en la naturaleza. Withlock y McCauley (1999) cuestionaron el cálculo de $N_e m$ de manera indirecta a partir del F_{ST} aunque consideraron que esta estimación indirecta puede ser válida cuando la escala espacial es pequeña, la tasa de migración es relativamente alta, el tamaño de la muestra y el número de *loci* son elevados y cuando la pregunta biológica real sea la estimación de los niveles de flujo génico a partir de $N_e m$.

Método basado en alelos raros: Slatkin (1981) desarrolló un método para estimar los valores promedios de Nm a partir de muestras de alelos tomados desde varias poblaciones, donde $\bar{p}(1)$ es la frecuencia promedio de alelos que son encontrados en una única población. Mediante estudios de simulación bajo el modelo de islas, Slatkin (1985) describió la relación $\log_{10}[\bar{p}(1)]$ de manera aproximadamente lineal con $\log_{10}(Nm)$, así $\log_{10}[\bar{p}(1)] \approx a \log_{10}(Nm) + b$, donde a y b dependen del número de individuos muestreados en cada población. Barton y Slatkin (1986) desarrollaron una teoría analítica a partir de la cual demostraron que $\bar{p}(1)$ es análogo a F_{ST} por lo cual esta sería una medida de la dispersión en la distribución de las frecuencias génicas. La mayoría de los alelos encontrados en una única muestra han estado

presentes por un tiempo suficiente para alcanzar un cuasi-equilibrio entre migración y deriva genética (Barton y Slatkin, 1986).

Se calculó Nm usando el método de alelos raros, incluyendo una corrección para el tamaño de la muestra, tanto a nivel global como a nivel regional implementando el programa Genepop versión 4.0.10 (Raymond y Rousset, 1995) disponible en <http://www.genepop.curtin.edu.au/>.

En cuanto a las dificultades presentadas por este método, debido a que asume fundamentalmente los mismos supuestos que el F_{ST} se debe tener las mismas precauciones a la hora de interpretar los resultados (Whitlock y McCauley, 1999).

3.4.4.2 Cuantificación del flujo génico reciente

La identificación de posibles inmigrantes o descendientes de inmigrantes recientes puede ser realizada incluyendo información *a priori* acerca de la localización geográfica de los individuos al análisis bayesiano. Este análisis asume que cada individuo presenta una alta probabilidad de ser originario de la región geográfica en la cual fue muestreado y permite, con una baja probabilidad, que sea un inmigrante o un descendiente de un inmigrante reciente (Pritchard *et al.*, 2000). Así, el empleo de la información geográfica mejora la precisión de la asignación de los individuos a los *clusters* y mejora la estimación de probabilidad *a posteriori*. Además, esta aproximación permite, mediante comparación con el análisis que no emplea información geográfica *a priori*, corroborar si los grupos genéticamente definidos, coinciden con los grupos geográficos (Pritchard *et al.*, 2000).

Se testaron distintos valores de la tasa de inmigración ν para el cálculo de la probabilidad de que un individuo sea un inmigrante. El análisis fue realizado para valores de ν iguales a: 0,005; 0,05 y 0,1, dado que ha sido sugerido emplear valores dentro del rango 0,001-0,1 (Pritchard *et al.*, 2000). Las demás características del modelo empleado para este análisis coinciden con las descritas en el punto 3.2.3.3. El valor de k fue calculado para un rango $k=1-6$ con 10 repeticiones para cada valor de k y se aplicó el método de Evanno para la identificación del valor real de k a partir de Δk . Estos análisis se realizaron empleando el programa STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) y la aplicación web STRUCTURE HARVESTER versión 0.6.92 (Earl y von Holdt, 2011).

4. Resultados

4.1 Desarrollo de SSR específicos para *Anadenanthera colubrina* var. *cebil*

En la Figura II-2 se presentan los resultados obtenidos en las distintas etapas del desarrollo de los marcadores SSRs específicos para *A. colubrina* var. *cebil*. Si bien se secuenció un número similar de insertos en cada librería (~100) sólo un bajo porcentaje de los insertos de la librería B contenían regiones SSRs. A su vez, la calidad de las secuencias provenientes de esta librería fue menor, con lo cual tan sólo el 20% de las secuencias provenientes de esta librería pudieron ser utilizadas para el diseño de cebadores (Tabla II-1). El proceso de desarrollo de cebadores de *novo* fue exitoso, presentando la librería compuesta por motivos (GA)₁₀, rendimientos superiores. Estos resultados pueden explicarse por la temperatura de hibridación empleada durante el proceso de captura o por la prevalencia de motivos (GA)_n en el genoma de esta especie. Al utilizar una mezcla de oligonucleótidos, como es el caso de la librería B, debe seleccionarse una temperatura de hibridación óptima para todos ellos. La mayoría de los insertos de la librería B secuenciados no contenían regiones SSRs, con lo cual, la temperatura de hibridación empleada podría haber sido muy elevada. En relación a la composición del genoma de *A. colubrina* var. *cebil*, se sabe que en plantas los motivos (GA)_n ocurren con mayor frecuencia (Weising *et al.*, 2005) y en coincidencia con esto, la librería A compuesta por motivos (GA)₁₀ fue la más exitosa.

Se diseñaron 29 pares de cebadores desde la librería A y sólo un par de cebadores desde la librería B. Los 30 pares de cebadores fueron empleados en las pruebas de funcionalidad. El 50% de los cebadores diseñados desde la librería A fueron funcionales y un 60% de ellos permitieron genotipificar claramente a los individuos (Tabla II-1). Así, se obtuvieron nueve pares de cebadores de los cuales 8 resultaron polimórficos en las poblaciones analizadas (Figura II-2 y Tabla II-2).

4.2 Utilidad de los nuevos marcadores para estudios genético - poblacionales

Los nueve marcadores SSRs desarrollados presentaron un número promedio de alelos por locus por población de 9,719 siendo las heterocigosidades observada y esperada promedio por locus de 0,725 y 0,791, respectivamente (Tabla II-3). Algunos *loci*, en determinadas poblaciones, no presentaron las proporciones esperadas en equilibrio Hardy - Weingberg. Además, en algunas poblaciones, se detectó para estos *loci* un exceso de homocigosis debido a la presencia de alelos nulos. Todos los *loci* se encontraron en equilibrio de ligamiento, a excepción de los *loci* *Ac28.3-Ac162.1* que presentaron desequilibrio de ligamiento entre ellos.

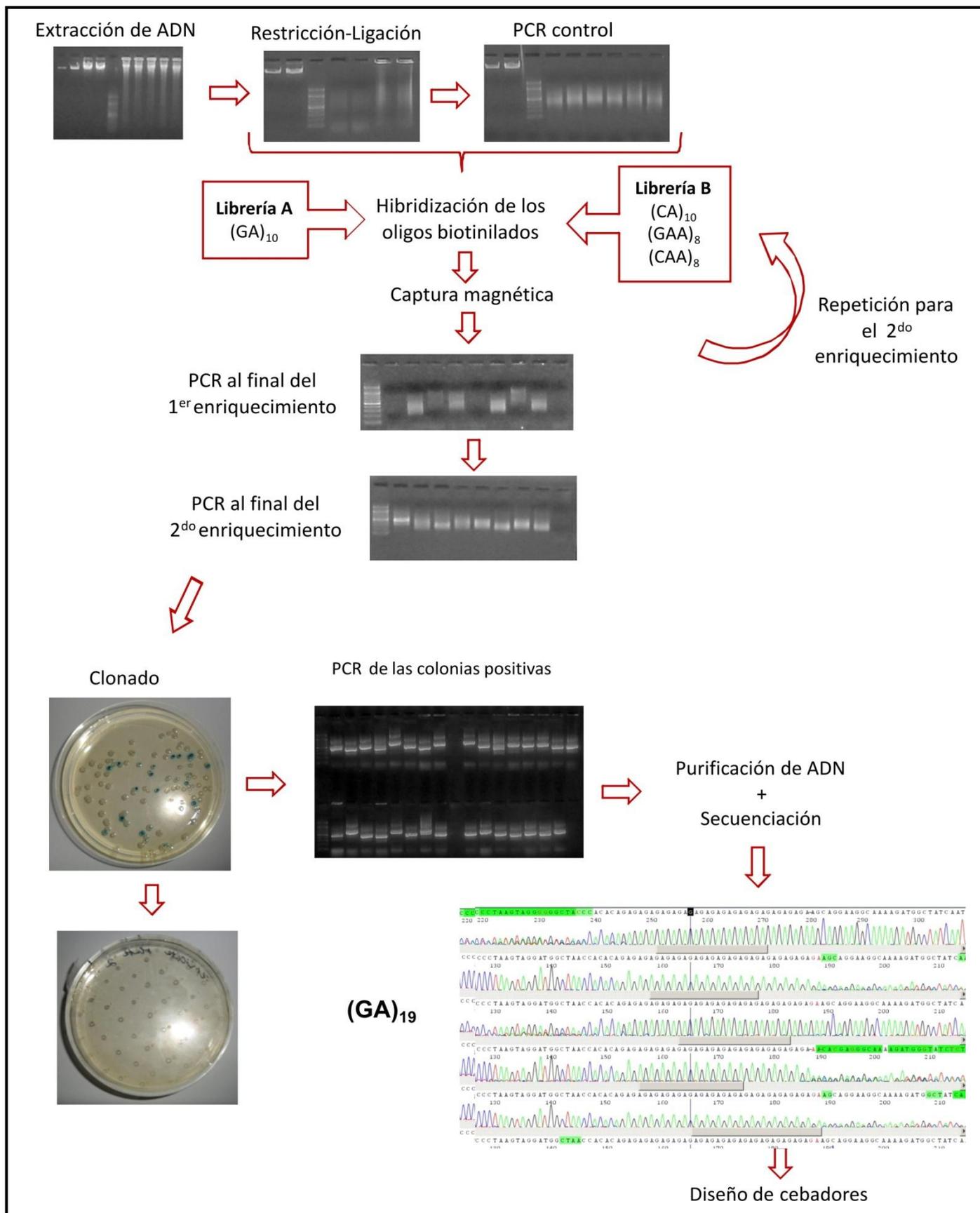


Figura II-2: Resultados de las diferentes etapas del desarrollo de los marcadores microsatélites en *A. colubrina* var. *cebil*.

Tabla II-1: Efectividad en el proceso de desarrollo de microsatélites específicos para *A. colubrina* var. *cebil*.

Librería	Oligonucleótidos	Eficiencia del Enriquecimiento			Eficiencia del diseño de cebadores			Funcionalidad	
		N° de insertos secuenciados	N° de insertos con SSRs	Secuencias de alta calidad con SSRs	Secuencias con regiones flanqueantes óptimas	N° de pares de cebadores diseñados	N° de loci amplificados claramente	N° de loci claramente genotipificados	Loci polimórficos
A	(GA) ₁₀	106	77 (73%)	56 (73%)	84%	29 (52%)	15 (51.7%)	9 (60%)	8 (88%)
B	(GAA) ₈ (CAA) ₈ (CA) ₁₀	98	20 (20.5%)	6 (30%)	100%	1 (17%)	0%	-	-
Total		204	97 (47.6%)	62 (64%)	85.5%	30 (48.4%)	15 (50%)	9 (60%)	8 (60%)

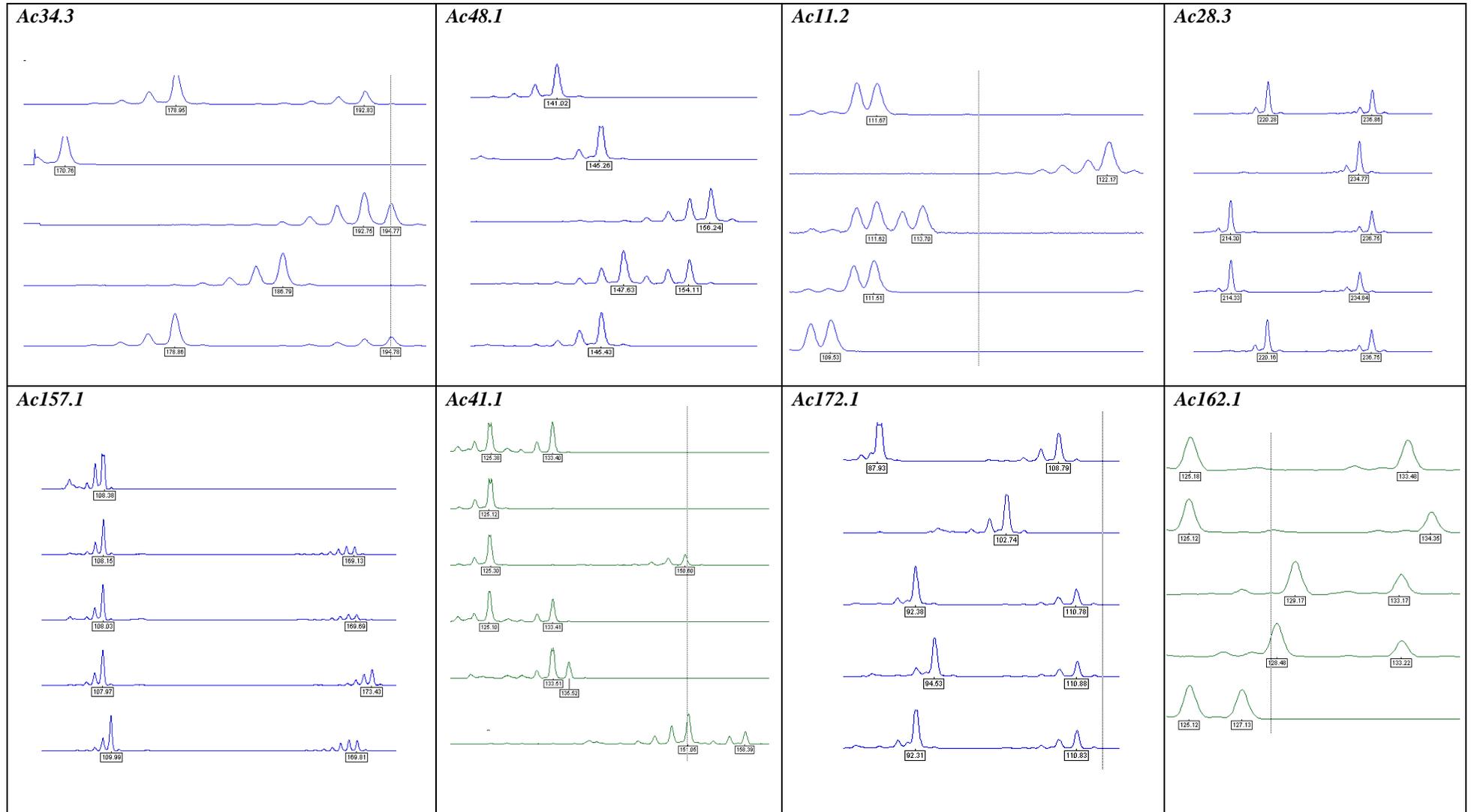


Figura II-3: Alelos revelados por los ocho marcadores SSRs polimórficos específicos para *A. colubrina* var. *cebil* identificados mediante un secuenciador analizador de fragmentos.

Tabla II-2: Caracterización de los nueve marcadores SSRs desarrollados para *A. colubrina* var. *cebil* (n= 69 individuos).

Locus	Secuencia de los cebadores (5'-3')	Motivo de repetición	Tamaño (bp)	T _A (°C)	Número de acceso GenBank
<i>Ac34.3*</i>	F: ccattctacaacacgcaagtg R: cctccaattcctccaactcc	(CT) ₂₁ (GT) ₃	171-203	TD 60 to 50°C PCR Multiplex	JQ086537
<i>Ac48.1*</i>	F: gagaccaaccacacgagtt R: ggtgtaattccataactctctctctg	(GA) ₃₀	129-171		JQ086538
<i>Ac11.2*</i>	F: caagcgtttctgatattattg R: ttgccattcctatttagtatga	(GT) ₁₁	110-124		JQ086539
<i>Ac28.3*</i>	F: gagcagccatgtttggagta R: cccacttctgccttgctatt	(GAA) ₄ X ₃₅ (GA) ₂₅	207-277	65°C	JQ086540
<i>Ac157.1*</i>	F: ccaccctccattttatttatct R: cagaaagaaccacggcaac	(CT) ₁₅ (CA) ₃ (CT) ₆ (CA) ₁₀	106-201	65°C	JQ086541
<i>Ac41.1**</i>	F: acgagctccacattcatgc R: gagccagtttcggttgaagg	(TC) ₁₀ (AC) ₇	125-158	TD 60 to 50°C PCR Multiplex	JQ086542
<i>Ac172.1*</i>	F: tctaaattacgtggagaaaacgaa R: tgacaggaccctcaccatgt	(A) ₅ G(A) ₄ (CT) ₁₉	87-122		JQ086543
<i>Ac162.1*</i>	F: tgtatgtgtaatatggaagtgc R: gcagtgcattgaccacctt	(GA) ₂₄	114-170	TD 60 to 50°C	JQ086544
<i>Ac29.2**</i>	F: gccagtgtgatgatatctgc R: tcaagtagtgcttcaactcc	(GT) ₁₅ (GA) ₂₈	327	TD 60 to 50°C	JQ806379

T_A: Temperatura de hibridación (TD: *Touchdown*)

*cebador *Forward* etiquetados con FAM su extremo 5'

**cebador *Forward* etiquetados con HEX su extremo 5'

Tabla II-3: Número de alelos, Heterocigosidad observada (H_O) y Heterocigosidad esperada (H_E) por locus en las poblaciones estudiadas.

Población	Locus	Número de alelos	H_O	H_E
Candelaria	Ac34.3	12	0,850	0,853
	Ac48.1	9	0,800	0,818
	Ac11.2	5	0,500	0,696
	Ac28.3	10	0,450	0,858
	Ac157.1	14	1,000	0,873
	Ac41.1	11	0,737	0,637
	Ac172.1	8	0,450	0,765
	Ac162.1	10	0,800	0,744
Santa Ana	Ac34.3	10	0,875	0,840
	Ac48.1	9	0,688	0,760
	Ac11.2	5	0,375	0,594
	Ac28.3	8	0,438	0,803
	Ac157.1	13	0,875	0,793
	Ac41.1	11	0,813	0,736
	Ac172.1	9	0,563	0,686
	Ac162.1	9	1,000	0,836
Tucumán	Ac34.3	10	1,000	0,862
	Ac48.1	10	0,636	0,872
	Ac11.2	6	0,571	0,673
	Ac28.3	14	0,786	0,888
	Ac157.1	8	0,714	0,768
	Ac41.1	6	0,714	0,684
	Ac172.1	9	0,786	0,852
	Ac162.1	6	0,778	0,679
Jujuy	Ac34.3	12	0,722	0,897
	Ac48.1	11	0,438	0,859
	Ac11.2	5	0,474	0,717
	Ac28.3	12	0,917	0,892
	Ac157.1	12	0,833	0,826
	Ac41.1	8	0,842	0,717
	Ac172.1	15	0,895	0,917
	Ac162.1	14	0,895	0,906
Promedio (por locus)		9,719	0,725	0,791

4.3 Análisis estadístico de los datos

4.3.1 Caracterización de la diversidad genética nuclear

Los marcadores específicos desarrollados para *A. colubrina* var. *cebil* presentaron entre 8 y 28 alelos por locus. Los ocho marcadores SSRs definieron un total de 152 alelos considerando a las cuatro poblaciones estudiadas, de los cuales 41 alelos (27%) estuvieron presentes únicamente en la región Paranaense, 46 alelos (30%) estuvieron presentes en la región Yungas, en tanto que los 65 alelos restantes (43%) estuvieron presentes en ambas regiones (Figura II-4).

La diversidad genética fue caracterizada mediante diferentes parámetros, siendo Jujuy la población que presentó los mayores valores para todos ellos (Tabla II-4). El número de alelos por locus promedio presentó un rango de 8,625 a 11,125 alelos, mientras que el número efectivo de alelos promedio presentó un rango de 4,479 a 7,692 alelos. En todas las poblaciones el número efectivo de alelos fue inferior al número de alelos observado, indicando que los alelos no presentaron frecuencias similares en las poblaciones analizadas dado que existen alelos a baja frecuencia. Los alelos presentes en una única población, conocidos como alelos únicos, presentaron un rango comprendido entre 1,125 y 3,125 (Tabla II-4). Las pruebas de *t* pareadas mostraron que para el análisis por locus Jujuy presentó 3 *loci* con los mayores valores en varios parámetros de diversidad genética resultando estas diferencias estadísticamente significativas, entre ellos se destacan la riqueza alélica y la riqueza de alelos únicos ya que estos estimadores son independientes del tamaño muestral. Al realizarse este análisis para los valores promedio de todos los *loci*, tanto el número promedio de alelos por locus, como el número efectivo de alelos, la heterocigosidad esperada, la riqueza alélica y la riqueza de alelos únicos presentaron valores mayores resultando estas diferencias estadísticamente significativas (Tabla II-5).

En promedio, todos los *loci* presentaron valores elevados de H_E en todas las poblaciones ($\bar{H}_E=0,791$) y una elevada multiplicidad génica promedio ($\bar{N}_A=9,719$), siendo Jujuy la población que presentó el mayor valor de H_E ($H_E=0,841$). Jujuy presentó mayor diversidad genética ($R=7,698$ y $R_{AU}=2,559$). Además, esta misma población presentó el mayor valor de riqueza de alelos únicos ($R_{AU}= 2,56$) (Tabla II-4).

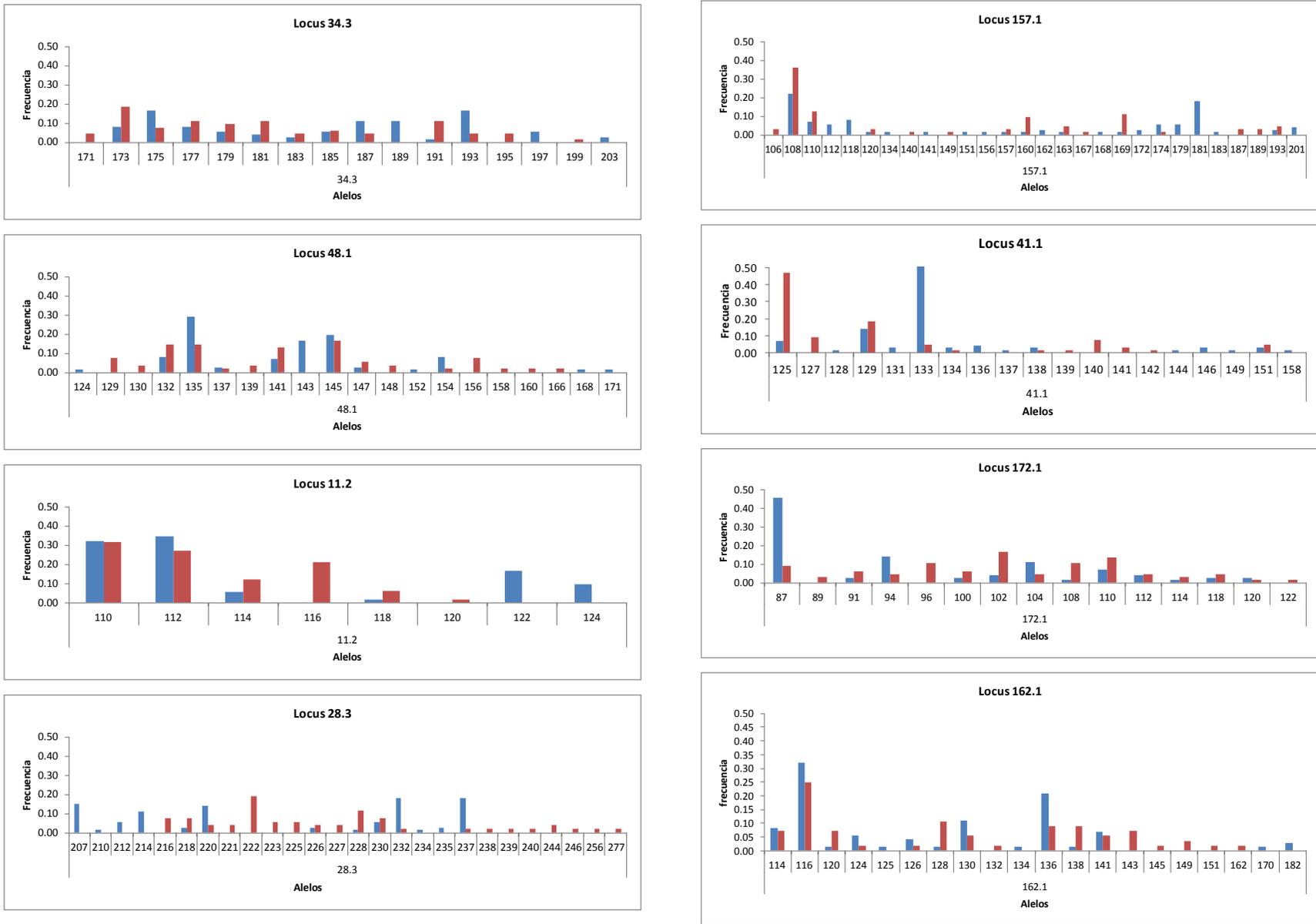


Figura II-4: Frecuencias alélicas por locus en las regiones estudiadas. Regiones: ■ Paranaense y ■ Yungas.

Tabla II-4: Caracterización de la diversidad genética revelada por los marcadores SSRs en las poblaciones estudiadas. Se resaltan los valores más elevados por población.

Región	Población	N	N_A	N_E	N_{AU}	H_O	H_E	R	R_{AU}
Paranaense	Candelaria	20	9,875	5,166	2,500	0,698	0,780	6,388	1,829
	Santa Ana	16	9,250	4,479	1,125	0,703	0,756	6,353	1,255
Yungas	Tucumán	14	8,625	5,548	1,625	0,748	0,785	6,573	1,607
	Jujuy	19	11,125	7,692	3,125	0,752	0,841	7,698	2,559
Promedio			9,719	5,721	2,094	0,725	0,791	6,753	1,813

N=tamaño de la población, N_A=número promedio de alelos por locus, N_E=número efectivo de alelos por locus, N_{AU}=número promedio de alelos únicos por locus, H_O=heterocigosidad observada, H_E=heterocigosidad esperada, R=riqueza alélica y R_{AU}=riqueza de alelos únicos.

Tabla II-5: Resultados de las pruebas no paramétricas, pruebas de t pareadas entre los parámetros de diversidad genética por locus y para el promedio de los 8 *loci*. En la tabla A) se presentan las poblaciones de la región Paranaense y en la tabla B) las poblaciones de la región Yungas. Entre paréntesis se indican los valores de p para las estimas estadísticamente significativas, en verde los valores menores y en rojo los valores mayores.

A	Candelaria							Santa Ana						
	Locus	N _a	N _e	N _{AU}	H _o	H _e	R	R _{AU}	N _a	N _e	N _{AU}	H _o	H _e	R
34.3	12	6.780	1.000	0.850	0.853	7,765	1,246	10	6.244	0.000	0.875	0.840	7,056	0,935
48.1	9	5.479	2.000	0.800	0.818	6,383	1,258	9	4.163	2.000	0.688	0.760 (0.031)	6,471	0,984
11.2	5	3.292	1.000	0.500	0.696	4,125	1,137	5	2.462	0.000	0.375 (0.040)	0.594 (0.015)	4,088	0,197
28.3	10	7.018	4.000	0.450	0.858	7,136	2,694	8	5.069 (0.042)	0.000 (0.020)	0.438	0.803 (0.019)	6,321	0,349 (0.019)
157.1	14	7.843 (0.020)	6.000	1.000	0.873 (0.044)	8,213	4,398	13	4.830	4.000	0.875	0.793	7,523	3,440
41.1	11	2.756	3.000	0.737	0.637 (0.038)	5,709	2,096	11	3.793	2.000	0.813	0.736	6,547	1,999
172.1	8	4.255	0.000	0.450	0.765	5,671	0,258	9	3.180	0.000	0.563	0.686	6,058	0,898
162.1	10	3.902	3.000	0.800	0.744	6,104	1,543	9	6.095	1.000	1.000 (0.039)	0.836	6,726	1,249
Todos los loci	9.875	5.166	2.500	0.698	0.78	6.388	1.829	9.250	4.479	1.125	0.703	0.756	6.353	1.255
Locus	Tucumán							Jujuy						
	N _a	N _e	N _{AU}	H _o	H _e	R	R _{AU}	N _a	N _e	N _{AU}	H _o	H _e	R	R _{AU}
34.3	10	7.259	1.000	1.000	0.862	7,538	0,800	12	9.672 (0.010)	2.000	0.722	0.897 (0.019)	8,791 (0.023)	1,914 (0.020)
48.1	10	7.806	3.000	0.636	0.872	8,211	3,272	11	7.111	3.000	0.438 (0.031)	0.859	7,557	2,44
11.2	6	3.063	1.000	0.571	0.673	4,389	0,962	5	3.539 (0.026)	0.000	0.474	0.717	4,338	0,266
28.3	14	8.909	6.000	0.786	0.888	9,268	4,078	12	9.290	4.000	0.917	0.892	9,166	3,822
157.1	8	4.308	1.000	0.714	0.768	5,978	1,800	12	5.735	5.000	0.833	0.826	7,486	3,572
41.1	6	3.161	1.000	0.714	0.684	4,647	0,952	8	3.539	4.000	0.842	0.717	5,435	2,973
172.1	9	6.759	0.000	0.786	0.852	7,258	0,599	15 (0.003)	12.033 (0.020)	2.000	0.895	0.917	9,717 (0.019)	2,132 (0.014)
162.1	6	3.115	0.000	0.778	0.679	5,294	0,394	14 (0.042)	10.618 (0.020)	5.000	0.895	0.906	9,092 (0.018)	3,555 (0.019)
Todos los loci	8.625	5.548	1.625	0.748	0.785	6.573	1.607	11.125 (0.035)	7.692 (0.014)	3.125	0.752	0.841 (0.017)	7.698 (0.003)	2.559 (0.027)

4.3.2. Análisis de las relaciones entre los genotipos multilocus

A partir de un árbol no enraizado es posible visualizar las relaciones existentes entre los individuos teniendo en consideración la distancia genética entre sus genotipos multilocus (Figura II-5). Los genotipos se agrupan coincidiendo con el origen geográfico de los individuos.

Diecisiete de los veinte individuos de Candelaria se encuentran agrupados principalmente en el Grupo 1, mientras que doce de los dieciséis individuos de Santa Ana se encuentran agrupados principalmente en el Grupo 3. Por su parte, la mayoría de los individuos de Tucumán y de Jujuy se agruparon en el Grupo 2 (once de los catorce individuos de Tucumán y trece de los diecinueve individuos de Jujuy). La mayor heterogeneidad estuvo presente en el Grupo 3 donde se agruparon individuos originarios de las cuatro poblaciones, si bien este grupo estuvo integrado en mayor proporción por individuos de Santa Ana (Figura II-5).

4.3.3 Análisis de la estructura genética de las poblaciones

4.3.3.1 Distribución de la diversidad genética dentro y entre poblaciones

El AMOVA permitió conocer la partición de la variación genética en los diferentes niveles jerárquicos de subdivisión poblacional, los cuales fueron establecidos a partir de criterios no genéticos. En el Modelo 1 se establecieron tres niveles jerárquicos, siendo el nivel dentro de poblaciones el que presentó el mayor porcentaje de variación (88,94%), el porcentaje de variación restante estuvo distribuido de manera proporcional entre los otros dos niveles: entre poblaciones dentro de regiones y entre regiones (~5,5%) (Tabla II-6 A). En el Modelo 2 se adicionó un nivel jerárquico menos inclusivo que considera a la fuente de variación dentro de individuos, es decir la variación en la composición alélica de los genotipos dentro de cada individuo. Bajo este modelo, el nivel dentro de individuos contuvo el mayor porcentaje de variación (82,26%) (Tabla II-6 B). Este AMOVA permitió la estimación del índice F_{IS} el cual resultó estadísticamente significativo a un nivel del 95% de confianza y su magnitud fue de 0,08. Este valor indica que los efectos de la endogamia son bajos en estas poblaciones ya que no se ha incrementado la homocigosis. Resultado que además concuerda con los elevados niveles de variación detectados en el nivel dentro de individuos. Además, valores bajos de F_{IS} indican que en las poblaciones presentan principalmente fertilización cruzada o alogamia con niveles bajos de endogamia.

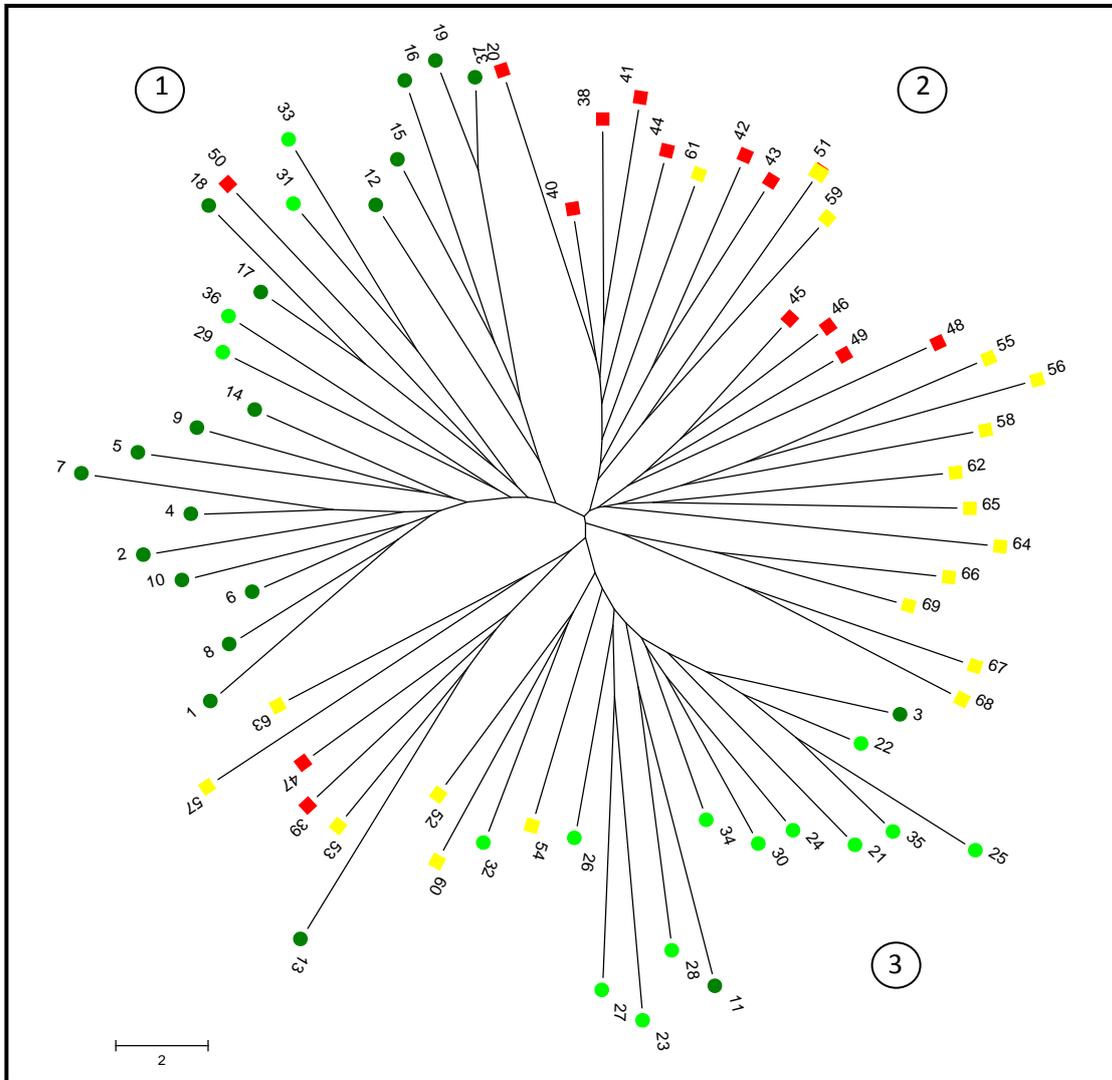


Figura II-5: Relaciones entre los individuos de *A. colubrina* var. *cebil* a partir de la información de sus genotipos multilocus. Poblaciones: ● Candelaria, ● Santa Ana, ■ Tucumán y ■ Jujuy. Grupos: 1 (izquierda), 2 (derecha) y 3 (inferior).

Tabla II-6: Análisis de la varianza molecular: **A)** Modelo 1, sin nivel individual y **B)** Modelo 2, con nivel individual.

A)

Fuente de Variación	Grados de libertad	Suma de cuadrados	Componentes de varianza	Porcentajes de variación
Entre regiones	1	14,600	0,123 <i>Va</i>	5,59
Entre poblaciones dentro de regiones	2	12,048	0,120 <i>Vb</i>	5,48
Dentro de poblaciones	134	261,403	1,951 <i>Vc</i>	88,94
Total	137	288,051	2,193	

B)

Fuente de Variación	Grados de libertad	Suma de cuadrados	Componentes de varianza	Porcentajes de variación
Entre regiones	1	14,600	0,123 <i>Va</i>	5,59
Entre poblaciones dentro de regiones	2	12,048	0,116 <i>Vb</i>	5,27
Entre individuos dentro de poblaciones	65	136,903	0,151 <i>Vc</i>	6,88
Dentro de Individuos	69	124,500	1,804 <i>Vd</i>	82,26
Total	137	288,051	2,193	

4.3.3.2 Estimación del índice de fijación F_{ST}

El índice F_{ST} se estimó a partir de diferentes aproximaciones y empleando diferentes metodologías. Por su parte, los valores de este índice inferidos tanto desde la varianza en las frecuencias alélicas como desde la identidad por descendencia presentaron valores estadísticamente significativos $F_{ST}=0,111$ y $F_{ST}=0,093$, respectivamente. Estos valores indican que existe estructuración genética moderada entre las poblaciones. El índice de fijación F_{ST} por región, calculado tanto en términos de la varianza en las frecuencias alélicas como en términos de identidad por descendencia, presentó valores estadísticamente significativos para ambas regiones e indicó presencia de estructura genética moderada en la región Paranaense y una baja estructuración genética en la región de las Yungas (Tabla II-7).

Se determinó el nivel de diferenciación de cada población respecto al conjunto de las poblaciones estimando el índice de fijación de cada población local mediante inferencia bayesiana. Todos los índices F_{ST} por población presentaron intervalos de confianza relativamente estrechos indicando estimaciones confiables de estos índices y un rango similar. De esta manera, podría decirse que estas poblaciones presentaron niveles similares de diferenciación respecto al complemento, excepto Jujuy, la cual muestra un rango comprendido entre valores de diferenciación menores (Tabla II-7).

Dado que algunos *loci* presentaron alelos nulos en algunas de las poblaciones se implementó el cálculo del índice F_{ST} mediante la exclusión de los mismos. A partir de este análisis pudo determinarse si la inclusión de los alelos nulos genera distorsión en las estimaciones de los parámetros poblacionales. El índice F_{ST} obtenido excluyendo los *loci* con alelos nulos no presentó diferencias respecto al índice F_{ST} calculado incluyendo estos *loci* siendo los resultados 0,08 y 0,07, respectivamente. Dado este resultado, se mantuvieron todos los *loci* en los análisis posteriores para aumentar la cobertura del genoma (Tabla II-7).

A pesar de que los valores de F_{ST} fueron estadísticamente significativos y revelaron presencia de estructura genética moderada en las poblaciones estudiadas, el índice D de Jost mostró valores más elevados ($D_{global}=0,370$). A nivel regional, la región Paranaense presentó un mayor valor de diferenciación genética respecto al presentado por la región de las Yungas ($D=0,214$ y $D=0,124$, respectivamente (Tabla II-7).

4.3.3.3 Determinación de la estructura genética poblacional mediante análisis bayesiano

Se estableció el número de *clusters* genéticos (k), es decir, el número de poblaciones desde el punto de vista genético mediante un análisis bayesiano. Implementando del método de Evanno (Evanno *et al.*, 2005) se determinó el número real de k a partir del máximo valor de Δk . En la Figura II-6 A se presenta el gráfico de Δk respecto a k y es posible identificar a $k=2$ como número real de *clusters* genéticos a nivel global ya que presentó el mayor valor de Δk . En la Figura II-7 A se muestra la representación gráfica de estos 2 *clusters*, donde cada color representa un *cluster* genético. Los individuos fueron asignados a cada *cluster* de manera

Tabla II-7: Estimaciones del índice de fijación F_{ST} mediante diferentes definiciones y estimación del índice de diferenciación D de Jost.

*estadísticamente significativo a un nivel del 95%

Definición empleada	Nivel		Índice de fijación	Referencia
Varianza en las frecuencias alélicas	Global		0,111*	(Cockerham y Weir, 1987)
	Regional	Paranaense	0,075*	
		Yungas	0,039*	
Identidad por descendencia	Global		0,093*	(Slatkin, 1991)
	Regional	Paranaense	0,080*	
		Yungas	0,040*	
Inferencia Bayesiana	Por población	Candelaria	0,080 [#] [0,033;0,150]	(Faubet y Gaggiotti, 2008)
		Santa Ana	0,066 [#] [0,024;0,129]	
		Tucumán	0,065 [#] [0,023;0,129]	
		Jujuy	0,032 [#] [0,009;0,084]	
Exclusión de alelos nulos	Global con alelos nulos		0,079	(Chapuis y Estoup, 2007)
	Global sin alelos nulos		0,076	
D de Jost	Global		0,370	(Jost, 2008)
	Regional	Paranaense	0,214	
		Yungas	0,124	

[#]intervalo *HDPI*

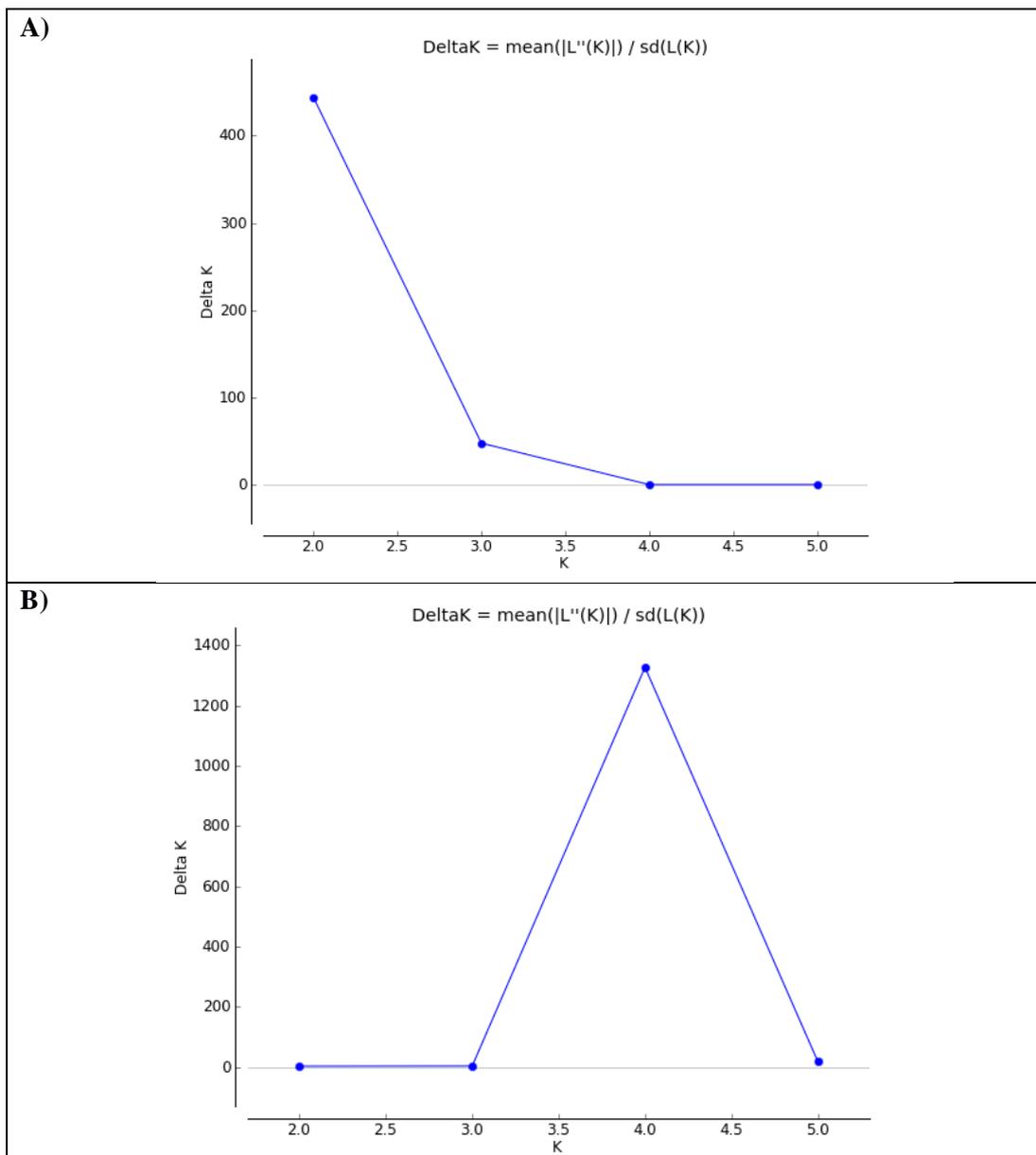


Figura II-6: Representación gráfica de los valores de Δk calculado para los diferentes valores de k . A) modelo sin información geográfica *a priori* y B) modelo empleando información geográfica *a priori* ($v=0,05$).

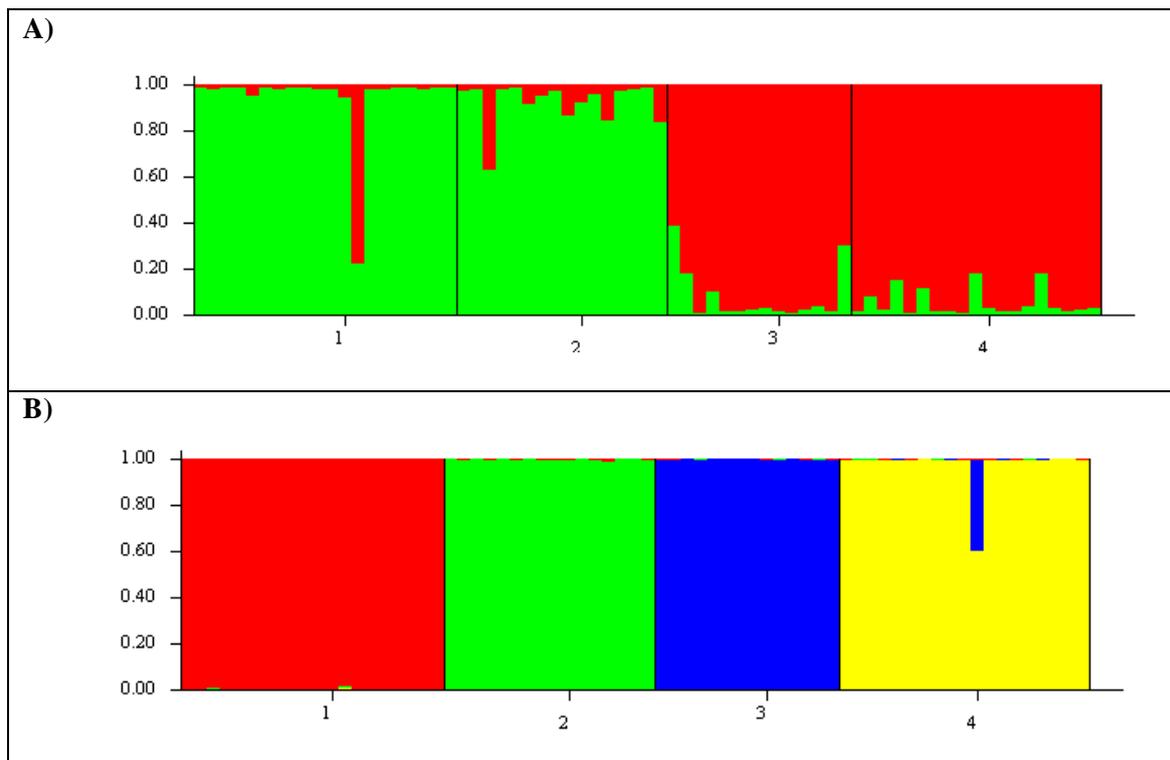


Figura II-7: Determinación de la estructura genética poblacional mediante inferencia Bayesiana (cada color representa un *cluster* genético). A) k=2 modelo sin información geográfica *a priori* y B) k=4 modelo empleando información geográfica *a priori* ($v=0,05$). Población: 1: Candelaria, 2: Santa Ana, 3: Tucumán y 4: Jujuy.

coincidente con su región de origen.

En el análisis independiente de cada región, la región Paranaense presentó dos *clusters* genéticos ya que $k=2$ presentó el mayor valor de Δk . La asignación de los individuos a estos dos *clusters* coincide con sus poblaciones de origen, Candelaria y Santa Ana (Figura II-8 A). Por el contrario, en la región Yungas no se indentificaron *clusters* genéticos (Figura II-8 B), indicando la ausencia de estructuración genética entre estas poblaciones.

4.3.4 Cuantificación de los niveles de flujo génico

4.3.4.1 Cuantificación de los niveles de flujo génico histórico

Método basado en el estadístico F_{ST} : El número efectivo de migrantes basado en el tamaño efectivo en términos de varianza en las frecuencias alélicas presentó un valor de $N_{ev}m=2,010$; mientras que el número efectivo de migrantes basado en el tamaño efectivo en términos de identidad por descendencia presentó un valor de $N_{ej}m=2,902$. Ambos resultados superaron la unidad lo cual indicaría que el flujo génico ha contrarrestado los efectos de la deriva genética. Puede inferirse que la fragmentación poblacional ha tenido mayor incidencia en la varianza en las frecuencias alélicas ya que $N_{ej}m$ mostró un valor mayor a $N_{ev}m$. Por su parte a nivel regional, la región Paranaense presentó un nivel de flujo génico histórico menor que la región Yungas, $N_{ev}m=3,083$ y $6,160$, respectivamente.

Método basado en alelos raros: El número efectivo de migrantes estimado a partir de la frecuencia de alelos raros en las poblaciones presentó un valor a nivel global de $Nm=2,540$ siendo la frecuencia de alelos raros a nivel global $p(1)=0,055$. Por su parte, a nivel regional Nm presentó un valor de $2,246$ y $2,041$ para las regiones Paranaense y Yungas, respectivamente. Siendo la frecuencia de alelos raros $p(1)=0,059$ y $0,064$ para las regiones Paranaense y Yungas, respectivamente.

4.3.4.2 Cuantificación de los niveles de flujo génico reciente

El flujo génico reciente, hasta dos generaciones previas, puede ser analizado mediante inferencia bayesiana incorporando información geográfica *a priori* en el análisis de la estructura genética. Para la determinación del número k se emplearon tres valores diferentes en relación a la tasa de migración ($v= 0,005$; $0,05$ y $0,1$). Dado que no se detectaron diferencias entre los resultados obtenidos según los diferentes valores de v , el análisis se continuó con los resultados obtenidos para la tasa de migración intermedia ($v=0,05$). Mediante la estimación de Δk se identificaron cuatro *clusters* genéticos al incorporar información geográfica *a priori* para el cálculo de la probabilidad *a posteriori* mediante inferencia bayesiana (Figura II-7 B). Al incorporar la información geográfica *a priori* los individuos provenientes de cada población definieron *clusters* genéticos relativamente aislados, ya que la mayoría de los individuos fueron asignados según las localidades geográficas que habitan. En la Figura II-7 B se observa que los límites de los *clusters* coinciden con los límites de cada localidad geográfica destacándose que

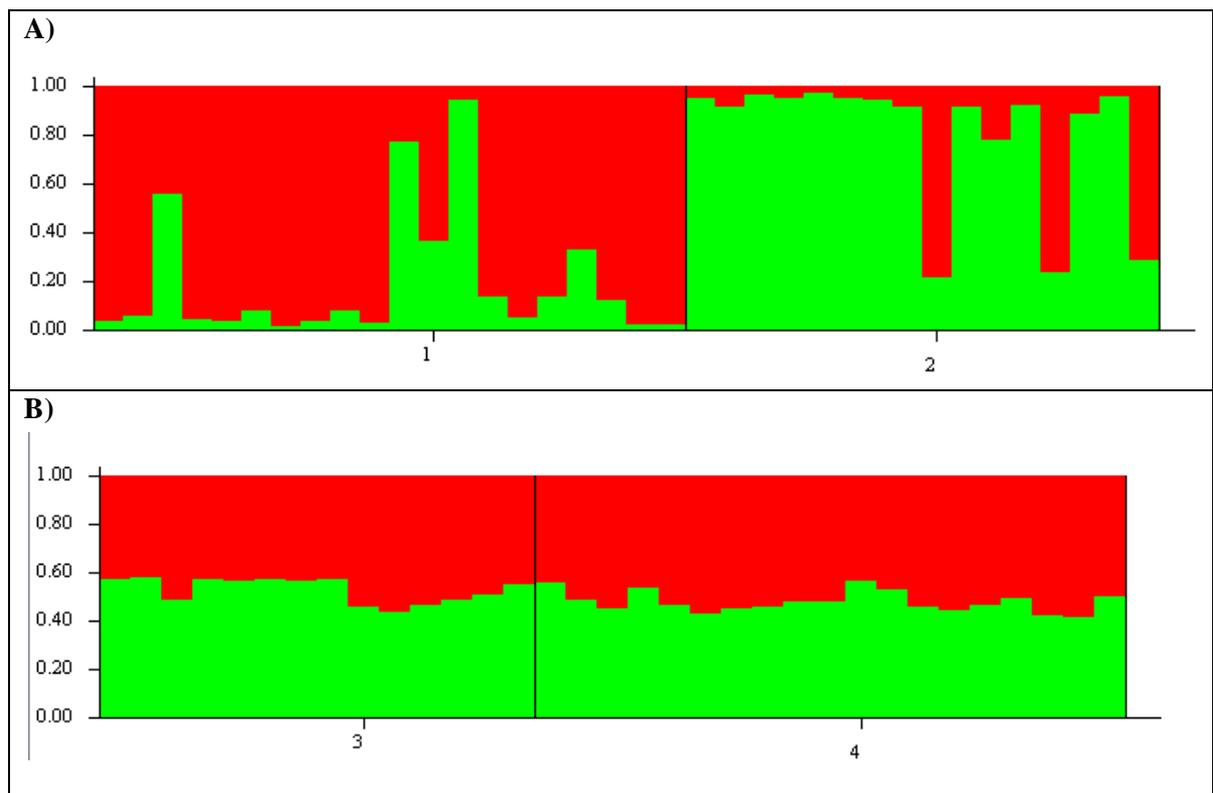


Figura II-8: Determinación de la subestructuración genética dentro de cada región mediante inferencia Bayesiana (cada color representa un *cluster* genético). A) Región Paranaense: k= 2 y B) Región Yungas k=2. Población: 1: Candelaria, 2: Santa Ana, 3: Tucumán y 4: Jujuy.

la migración reciente entre las poblaciones estudiadas es prácticamente nula, dado que sólo un genotipo de Jujuy fue asignado al *cluster* integrado por individuos de Tucumán. Este genotipo corresponde al individuo 61 de Jujuy, presentando una probabilidad de 0,71 de ser inmigrante.

5. Discusión

El presente estudio tiene como objetivo analizar la acción del proceso microevolutivo dispersivo sobre los niveles y la distribución de la diversidad genética entre las poblaciones buscando dar respuesta a los siguientes interrogantes ¿Qué roles juegan el flujo génico y la deriva genética como fuerzas antagónicas responsables de la distribución de las frecuencias alélicas dentro y entre las poblaciones naturales de *A. colubrina* var. *cebil*?, ¿Cuál es la importancia relativa de la dispersión de las semillas y del polen sobre el movimiento de los alelos en esta especie?, y ¿Cuáles son las consecuencias genéticas de la distribución fragmentada en estas poblaciones? atendiendo a las diferencias temporales implícitas en estas estimaciones.

La diversidad genética poblacional consiste en la sumatoria de toda la variación genética presente entre los individuos dentro de una población (Young *et al.*, 1996). La primera estima de la variación genética dentro de una población es el número de variantes genéticas o alelos presentes en un locus determinado, lo cual también es conocido como multiplicidad génica (Gillet *et al.*, 2005). En general, todos los *loci* fueron altamente polimórficos en las poblaciones estudiadas, presentando valores entre 5 y 15 alelos por locus por población y entre 8 y 28 alelos por locus al considerar el total de individuos estudiados. La población Jujuy presentó el mayor promedio del número de alelos por locus siendo las diferencias respecto a las demás poblaciones estadísticamente significativas. Debido a que la multiplicidad y la abundancia pueden variar independientemente, la variación genética puede ser expresada como un número efectivo. Si los alelos se presentan en igual frecuencia, el número efectivo de alelos será igual a la multiplicidad génica, en cambio si los alelos no presentan distribuciones de frecuencias uniformes, el número efectivo de alelos será una función decreciente (Gillet *et al.*, 2005). En las cuatro poblaciones estudiadas el número efectivo de alelos fue menor al número de alelos, indicando este resultado que en todas las poblaciones existen alelos presentes a baja frecuencia. Dado que el número de alelos detectado en una población depende de su tamaño, no es recomendable comparar la multiplicidad génica entre poblaciones locales de distinto tamaño. El concepto de riqueza alélica permite predecir la multiplicidad génica esperada en el caso de que dos muestras tengan el mismo tamaño. Las estimaciones de riqueza alélica son empleadas para comparar el número de alelos entre muestras que difieren en tamaño, situación común en estudios genéticos poblacionales en especies forestales (Gillet *et al.*, 2005). La diversidad genética puede ser caracterizada por parámetros como la riqueza alélica (R) y la diversidad génica (H_e), conocida generalmente como heterocigosidad esperada (Young *et al.*, 1996). Así, mediante la estimación de riqueza alélica y la diversidad génica se distingue a Jujuy como la población más diversa. Además, esta población presentó una elevada riqueza de alelos únicos. Valores elevados de alelos únicos en una población podrían proveer evidencias de una pérdida de alelos raros en las demás poblaciones por reducción de su tamaño (Pakkad *et al.*, 2008).

La fragmentación del hábitat podría causar la pérdida de diversidad genética en las poblaciones naturales ya que la reducción del tamaño poblacional causa cuellos de botella debido a que los individuos remanentes sólo portan una muestra del *pool* génico original. Seguido a esta pérdida de diversidad genética original, las poblaciones remanentes que permanecen reducidas y aisladas por varias generaciones continúan perdiendo diversidad alélica debido a los efectos de la deriva genética (Young *et al.*, 1996). Dado que los efectos de la deriva genética dependen del número de generaciones durante las cuales las poblaciones presentan un número reducido de individuos, las poblaciones con tiempos generacionales cortos, como las plantas herbáceas, deberían mostrar mayor pérdida de diversidad genética que las especies con prolongados tiempos generacionales, como los árboles. Sin embargo, estudios empíricos realizados en tres especies herbáceas y una especie de *Eucaliptus* demostraron que la pérdida de diversidad genética fue principalmente debida a los cuellos de botella más que a la acción subsecuente de la deriva genética (Young *et al.*, 1996). Dado que Jujuy presentó mayor diversidad genética y mayor riqueza de alelos únicos esto podría estar indicando un mayor tamaño efectivo y un menor impacto de la fragmentación reciente en esta población.

La distribución de la diversidad genética entre los diferentes niveles jerárquicos mostró que el mayor porcentaje de variación estuvo contenido en el nivel jerárquico menos inclusivo en los dos modelos empleados. Es decir, dentro de poblaciones en el Modelo 1 y dentro de individuos en el Modelo 2. Este resultado es frecuente en especies forestales debido a ciertas características de sus poblaciones que contribuyen a mantener elevados niveles de diversidad genética en ellas. Entre estas características se cuenta el gran tamaño poblacional, la longevidad de los individuos, los elevados niveles de alogamia y/o elevados niveles de flujo génico entre poblaciones (White *et al.*, 2007). Sumado a estas características, elevados niveles de polimorfismo son esperables cuando se analiza la diversidad genética mediante marcadores de tipo microsatélites nucleares ya que presentan elevada tasa de mutación, generando un aumento en el valor de multiplicidad génica y además al ser neutrales, sus frecuencias no estarán determinadas por acción de la selección natural (Kimura, 1968). El Modelo 2 permitió estimar el índice de endogamia (F_{IS}) a partir de la fuente de variación dentro de individuos, la cual contiene la variación genotípica presente en ellos. El índice F_{IS} refleja el aumento de la homocigosis como consecuencia de la endogamia. En las poblaciones estudiadas este índice presentó un valor estadísticamente significativo aunque alcanzó un valor bajo ($F_{IS}=0,08$). En *A. colubrina* var. *cebil* pueden esperarse niveles elevados de fecundación cruzada ya que se asume que es una especie principalmente alógama. La alogamia es el sistema de apareamiento típico de las especies forestales tropicales, aunque se han reportado ciertos niveles de autofecundación en estas especies, variando los niveles de alogamia y autofecundación de un año a otro (Finkeldey y Hattemer, 2007). Además, a partir de los elevados niveles de heterocigosis mantenidos en las poblaciones estudiadas ($H_O=0,703-0,780$) podría afirmarse que el sistema de apareamiento de

esta especie es principalmente alógama. Dado que los niveles de heterocigosidad presentaron valores elevados y el índice F_{IS} presentó un valor bajo, puede decirse que las poblaciones estudiadas no presentarían endogamia.

La presencia de alelos nulos no incidió sobre el valor del índice F_{ST} estimado. Debido a que para algunos *loci* en ciertas poblaciones se detectó homocigosis en exceso deben considerarse como causas posibles la presencia de alelos nulos, el sistema de apareamiento, el tamaño efectivo poblacional reducido, el efecto Wahlund (ya que algunas poblaciones pueden presentar subestructuras espaciales o temporales) o ser consecuencia de errores de tipo metodológicos (por ejemplo errores de muestreo o presencia de datos perdidos en exceso) (Muller *et al.*, 2009).

El índice de fijación global fue estimado según la varianza en las frecuencias alélicas y según la identidad por descendencia, arrojando valores de $F_{ST}= 0,11$ y $0,09$, respectivamente. Ambos valores fueron estadísticamente significativos y revelaron presencia de estructura genética moderada. Debido a que el nivel de flujo génico estimado considerando el tamaño efectivo en términos de varianza en las frecuencias alélicas es consecuencia de los efectos de la fragmentación (Templeton, 2006) es posible priorizar los efectos de la fragmentación histórica sobre la estructura genética de las poblaciones de *A. colubrina* var. *cebil* estudiadas. Los niveles de flujo génico interpoblacional se reducen como consecuencia de la pérdida de poblaciones asociadas a la fragmentación, al aumento del aislamiento y a los cambios entre la superficie ocupada por las poblaciones remanentes (Young *et al.*, 1996).

Los niveles de estructuración genética detectados ($F_{ST}\approx 0,10$) coinciden con los niveles de diferenciación esperados en especies tropicales en las cuales el polen es dispersado a mayores distancias y cuya dispersión de semillas es realizada por medios abióticos como el viento o la gravedad, ya que estas especies presentan mayor estructuración genética ($F_{ST}= 0,138$) en relación a las especies forestales tropicales cuya dispersión de semillas es realizada por animales ($F_{ST}=0,05$) (Finkeldey y Hattemer, 2007). Esta diferencia en el mecanismo de dispersión de las semillas destaca la acción del flujo génico mediado por semillas sobre la estructuración genética.

La estructuración genética poblacional revelada por los marcadores neutrales es una consecuencia de la magnitud de la deriva genética a la cual cada población local ha sido expuesta, del tamaño efectivo poblacional y del grado de aislamiento geográfico y/o ecológico (Foll y Gaggiotti, 2006). Sin embargo, el estudio de la estructuración genética poblacional es realizado tradicionalmente empleando estimaciones globales (tal como el índice de fijación F_{ST}) las cuales ignoran las diferencias entre las poblaciones en relación a la intensidad de la deriva genética. Por ello, resulta apropiado basar la estimación de los parámetros y las variables específicas para cada población local mediante aproximaciones bayesianas que permiten conocer los índices F_{ST} por población local teniendo en consideración que la intensidad de la

deriva genética puede variar entre las poblaciones (Foll y Gaggiotti, 2006). Así, la estimación del índice F_{ST} específico para cada población permite establecer cuan diferenciada está cada población local en términos de su composición genética en relación al complemento, donde valores elevados de F_{ST} indicarán que la distribución de las frecuencias alélicas en la población local es completamente diferente a la población como un todo (Foll y Gaggiotti, 2006). De acuerdo con las estimaciones de los índices de fijación de cada población es posible inferir niveles similares de diferenciación genética. Sin embargo, Jujuy presentó un rango entre valores de menor magnitud lo cual podría implicar un tamaño efectivo mayor un menor efecto de la deriva genética en esta población.

El índice de fijación global ($F_{ST}=0,11^*$) indica presencia de estructura genética moderada. Debido a la distribución disyunta de *A. colubrina* var. *cebil* en el Norte argentino este valor podría representar la diferenciación existente entre las regiones, la cual se habría producido por un evento de fragmentación histórica de estas poblaciones a nivel regional en concordancia con la hipótesis postulada por diferentes autores que sostiene que los SDTF en los tiempos glaciares presentaban una distribución más extensa que luego se fragmentó (Prado y Gibbs, 1993; Pennington *et al.*, 2004; Caetano *et al.*, 2008; Werneck *et al.*, 2011). Sin embargo, estos autores no coinciden en los tiempos en los cuales estos eventos sucedieron. La región Paranaense presentó moderada estructuración genética según las estimas de los índices F_{ST} considerando la varianza en las frecuencias alélicas y la identidad por descendencia, a pesar de la menor distancia geográfica entre las poblaciones estudiadas. Mientras que, la región de las Yungas presentó baja estructuración genética a partir de las mismas estimaciones. La mayor estructuración genética de la región Paranaense, a pesar de la menor distancia geográfica entre sus poblaciones y a pesar de presentar probablemente mayor estabilidad histórica, podría ser interpretada como una consecuencia de la fragmentación reciente en las poblaciones causada por el impacto antrópico, el cual podría haber aumentando el aislamiento de las poblaciones remanentes debido a la eliminación de posibles poblaciones intermedias y/o haber afectado a la dinámica poblacional de las especies polinizadoras. Sin embargo a partir de este estudio no es posible obtener más conclusiones al respecto y deja planteado un interrogante al respecto. Por el contrario, en la región de las Yungas la reducida estructuración genética, a pesar de la distancia geográfica entre las poblaciones estudiadas, podría interpretarse que los efectos de la deriva genética en sus poblaciones son contrarrestados por la acción del flujo génico. Además, los efectos de la deriva genética podrían ser reducidos también como consecuencia del mayor tamaño efectivo poblacional. Debe recordarse que *A. colubrina* var. *cebil* se encuentra distribuida a lo largo de las Yungas principalmente en la selva pedemontana, con lo cual el flujo génico mediado por el polen podría mantener interconectadas a una red de poblaciones locales. Estas poblaciones podrían intercambiar migrantes de manera efectiva y así mantener mayores niveles de diversidad genética presentando, como consecuencia, niveles reducidos de

estructuración genética. Slatkin (2005) empleó el término “*ghost population*” o población fantasma en referencia a cualquier población que no ha sido muestreada pero que está conectada por migración a poblaciones que sí lo han sido.

Atendiendo a que las poblaciones de las Yungas se encuentran en áreas protegidas, la mayor diversidad genética presentada por Jujuy podría ser consecuencia del estado de conservación, dado que esta población se ubica en el Parque Nacional Calilegua el cual se encuentra rodeado de poblaciones de *A. colubrina* var. *cebil* pudiendo existir flujo génico principalmente por medio del polen entre la población considerada y las poblaciones vecinas. Por otro lado, en la región Paranaense ambas poblaciones se encuentran bajo distintos regímenes de manejo siendo más intensa la historia de aprovechamiento y el impacto sufrido por sus poblaciones (Di Bittetti *et al.*, 2003).

El valor del índice F_{ST} es altamente dependiente de la diversidad genética dentro de las poblaciones, siendo este índice siempre inferior a H_S y su rango estrecho cuando H_S es grande. Por ejemplo, cuando H_S es igual a 0,9, valor comúnmente detectado en *loci* SSR, el valor máximo de F_{ST} es 0,1 (Meirmans y Hedrick, 2011). Se debe tener presente que las comparaciones de los índices F_{ST} pueden resultar inconsistentes cuando son estimados considerando poblaciones locales dentro de diferentes regiones de la distribución de la especie las cuales difieren en cuanto a los niveles de diversidad genética (Meirmans y Hedrick, 2011). Jost (2008) desarrolló una buena aproximación para estimar los niveles de diferenciación genética en estos casos basando su estadístico D en el número efectivo de alelos. Sin embargo, D se encuentra afectado en mayor medida por la tasa de mutación haciendo a este estimador menos adecuado para aplicaciones prácticas como la estimación de tasas de migración (Meirmans y Hedrick, 2011). Además, D no es una función del tamaño poblacional local N y por ello no describe los efectos de la deriva genética local (Whitlock, 2011). A pesar de ello, D permite una medición adecuada de la diferenciación en las frecuencias alélicas entre poblaciones mientras que F_{ST} resulta adecuado para describir la influencia de eventos demográficos y de procesos microevolutivos sobre la distribución de la variación genética (Meirmans y Hedrick, 2011, Whitlock, 2011). El posible efecto de la elevada diversidad en la estimación de estos índices fue evaluado mediante el índice D . A nivel global se obtuvo un valor marcadamente superior al valor de F_{ST} ($D=0,370$) el cual, posiblemente sea una mejor aproximación a la diferenciación genética existente entre estas poblaciones. Al igual que el F_{ST} , las poblaciones de la región Paranaense presentaron mayor de diferenciación genética que las poblaciones de la región Yungas.

El análisis mediante el uso de más de una medida de diferenciación permite una evaluación del análisis de la diversidad desde diferentes aproximaciones. Sin embargo, resultan más adecuados aquellos estimadores que permitan dar respuesta a preguntas biológicas, con lo cual para los estudios intraespecíficos el índice F_{ST} es el estimador preferido (Whitlock, 2011).

Es por ello que, a pesar de la propuesta de nuevas aproximaciones, F_{ST} continúa siendo considerado una estimación útil para estudios comparativos del flujo génico (Meirmans y Hedrick, 2011). Así, el índice F_{ST} es tradicionalmente empleado en estos estudios y permite la comparación de las estimaciones realizadas en *A. colubrina* var. *cebil* respecto a otras especies.

A partir del árbol construido desde las distancias genéticas entre los genotipos multilocus de *A. colubrina* var. *cebil* en el presente estudio, es posible identificar grupos de individuos y establecer relaciones en cuanto al origen geográfico de los mismos. Algunos grupos incluyen individuos pertenecientes a diferentes poblaciones. El individuo 50, perteneciente a Tucumán, se ubica en el grupo integrado principalmente por individuos del núcleo Paranaense (Figura II-5 Grupo 1). Debido a que en este método los grupos son identificados desde la topología del árbol, no es posible establecer si el individuo 50 es un inmigrante o si ha sido ordenado en un grupo distinto por azar. También, desde la topología de este árbol, puede distinguirse que las poblaciones de la región Paranaense definieron dos grupos diferentes, (Figura II-5 Grupo 1 y Grupo 3). Por su parte, la mayoría de los individuos de la región Yungas formaron un grupo (Grupo 2). Esta mayor heterogeneidad genética presente entre las poblaciones de la región Paranaense es coincidente con los resultados obtenidos desde los índices F_{ST} por región los cuales indicaron mayor nivel de estructuración genética entre las subpoblaciones de la región Paranaense. Tanto los métodos basados en distancias como los métodos basados en modelos no asumen una estructura genética predeterminada. Debido a la falta de poder estadístico para testar los grupos identificados desde la topología del árbol se toma al método basado en distancias como un análisis exploratorio pudiendo aplicarse luego métodos basados en modelos los cuales permiten realizar inferencias con poder estadístico.

Mediante la aplicación de métodos basados en modelos se estableció que, a nivel global, los individuos de *A. colubrina* var. *cebil* se agrupan definiendo dos *clusters* ($k=2$), quedando agrupadas en estos dos *clusters* las poblaciones pertenecientes a cada región (Figura II 7-A). Los individuos de las poblaciones Candelaria y Santa Ana integran un *cluster* y los individuos de las poblaciones Tucumán y Jujuy integran el *cluster* restante. Esta diferenciación se podría interpretar como una consecuencia de la fragmentación histórica que llevó al aislamiento geográfico entre ambas regiones donde la distancia geográfica estaría actuando como una barrera al flujo génico.

Evanno *et al.* (2005) han demostrado que los algoritmos que subyacen al programa STRUCTURE permiten identificar grupos de individuos correspondientes al nivel jerárquico superior, es decir, permiten detectar el nivel más elevado de estructuración genética, mientras que los subgrupos creados por la mejor asignación de individuos a estos grupos permiten identificar los niveles inferiores de estructuración genética. De acuerdo con esto, se realizó el análisis de la estructuración genética de cada región, coincidente con cada *cluster* identificado por el análisis global, de manera independiente con el objetivo de detectar presencia de

subestructuración genética dentro de las regiones. Los resultados de este análisis fueron diferentes para ambas regiones. La región Paranaense mostró subestructuración genética siendo sus individuos asignados a dos *clusters*. Los genotipos asignados a estos *clusters* coincidieron en gran medida con sus poblaciones de origen. Por su parte, la región de las Yungas no presentó subestructuración genética. Estos resultados, concuerdan con las inferencias realizadas desde los índices de fijación regionales y los agrupamientos observados en el árbol de genotipos.

Respecto a la cuantificación del flujo génico histórico, el número efectivo de migrantes global superó la unidad ($Nm \approx 2$), indicando, desde un punto de vista histórico, que en las poblaciones estudiadas los efectos del flujo génico son suficientes para contrarrestar los efectos de la deriva genética. La ecuación empleada para el cálculo de Nm desde el índice de fijación, muestra la manera en que los procesos interactúan y determinan los patrones de estructuración genética: a medida que $\frac{1}{N_e}$ incrementa la deriva genética se hace más fuerte y F_{ST} aumenta ya que existe menor variación dentro de las poblaciones locales y mayor diferenciación entre ellas. Por otro lado, a medida que m aumenta el flujo génico se hace más fuerte y F_{ST} decrece debido a que a mayor variación dentro de las poblaciones locales será menor la diferencia entre ellas (Templeton, 2006). Lo más importante en esta ecuación es que, niveles mínimos de flujo génico pueden producir que dos poblaciones se comporten como un único linaje evolutivo. Así, $Nm=1$ constituye un punto de inflexión en la relación entre F_{ST} y Nm . Cuando aumenta el número de migrantes a valores mayores a la unidad, el valor de F_{ST} se reduce lentamente. Sin embargo, cuando Nm es menor a la unidad, el valor de F_{ST} aumenta rápidamente. De esta manera, $Nm=1$ marca una transición importante desde el punto de vista biológico en la relación entre flujo génico y deriva genética. Sólo se necesita un único migrante en promedio, por generación, para que el flujo génico domine sobre la deriva genética, permitiendo que las poblaciones muestren elevada homogeneidad genética entre ellas (Templeton, 2006). Kimura y Weiss (1964) demostraron que cuando $Nm > 4$, los efectos del flujo génico superan en gran medida a los efectos de la deriva genética y que las poblaciones locales son efectivamente panmícticas. En contraste, cuando $Nm \ll 1$ la deriva genética domina al flujo génico y existe una gran diferenciación genética entre las poblaciones. El valor de Nm estimado para las poblaciones de *A. colubrina* var. *cebil* se encuentra entre estos dos extremos, denotando que niveles moderados de ambos procesos se encuentran presentes en las regiones consideradas. Los métodos indirectos miden el nivel de diferenciación genética entre las poblaciones y permiten inferir el número de migrantes que han sido intercambiados históricamente por generación ($N_e m$) (Burczyk *et al.*, 2004). Los valores obtenidos de $N_e m$ indicarían la existencia de flujo génico entre las regiones a nivel histórico, esto apoyaría a las teorías que sostienen que los SDTF presentaron una distribución más extensa en el pasado (Prado y Gibbs, 1993; Werneck *et al.*, 2011). Por otro lado, el método basado en alelos raros propuesto por Slatkin (1985) estima Nm a

partir de una muestra de los datos, lo cual podría brindar una estimación diferente del número de migrantes intercambiados por generación además de proveer un medio para probar si las estimaciones son consistentes entre ellas (Slatkin y Barton, 1989). Slatkin y Barton (1989) recomiendan el empleo de F_{ST} para las estimaciones indirectas de Nm por sobre el método basado en alelos raros ya que este último es más sensible a errores de genotipificación y además requiere de la presencia de suficiente cantidad de alelos únicos en las poblaciones. La estimación del nivel de flujo génico global a partir del método basado en alelos raros arrojó un valor de $Nm=2,540$, el cual supera a la unidad y es similar al estimado desde F_{ST} ($Nm=2,010$). A nivel regional, las estimas de Nm mediante el método de alelos raros presentaron valores similares entre ambas regiones. En tanto que las mismas estimas basadas en F_{ST} presentaron valores superiores que las estimas mediante alelos raros y la región de las Yungas presentó un valor de Nm que duplica el valor obtenido en la región Paranaense. La estimación de Nm desde ambos métodos indirectos en una situación ideal, es decir muestreos sin errores de poblaciones en equilibrio bajo neutralidad, deberían ser coincidentes. Sin embargo, cuando estos supuestos son violados estas pueden diferir (Yamimachi e Innan, 2012). La mayoría de los alelos presentes en una única población suelen ser raros ya que presentan una baja frecuencia, dado que es de esperar que los alelos raros sean más recientes las estimaciones de Nm desde alelos raros serán más sensibles a la migración reciente. Por su parte, las estimaciones de F_{ST} se basan en las heterocigosidades, las cuales son determinadas en gran medida por los alelos comunes pudiendo esperarse que los alelos presentes a mayores frecuencias sean más antiguos, con lo cual las estimaciones de Nm desde F_{ST} deberían reflejar el flujo génico en un período de tiempo relativamente prolongado (Yamamichi e Innan, 2012). A partir de esto, se podría inferir que las poblaciones de la región Yungas serían recientes y que los niveles de flujo génico estimados desde F_{ST} serían representativos del flujo génico histórico entre estas poblaciones siendo este de mayor importancia que el flujo génico reciente.

Los métodos indirectos tienen la ventaja de incorporar los efectos de todos los componentes históricos del flujo génico y generar un promedio de su variación a través del tiempo (Planter, 2007). Sin embargo, estos métodos presentan algunas deficiencias (Burczyk *et al.*, 2004) a pesar de la existencia de datos empíricos y de simulación que indican que son robusto a algunas desviaciones de los supuestos básicos. La estimación del número efectivo de migrantes desde F_{ST} fue criticada debido a que el modelo de islas sobre el cual se basa es poco realista (Withlock y McCauley, 1999; Broquet y Petit, 2009). Resulta difícil asignar separadamente los efectos del flujo génico y de la deriva genética sobre la estructura genética. La mayoría de los grupos de poblaciones naturales muy probablemente no se encuentran en equilibrio flujo génico-deriva genética el cual es uno de los supuestos del modelo de islas sobre el cual se sustentan las aproximaciones indirectas tradicionales. Por ello, los efectos del flujo génico y de la deriva genética pueden ser confundidos en el producto Nm (Hutchinson y

Templeton, 1999). El modelo de islas es una abstracción conveniente que separa los efectos opuestos de la deriva genética y del flujo génico, y aunque F_{ST} no siempre es una estimación exacta, es poco probable que esté sobreestimado en varios órdenes de magnitud. Los resultados empíricos y desde simulaciones indican que el uso del índice F_{ST} para estimar Nm es útil y sencillo. Además, Yamamichi e Innan (2012) sostienen que las estimaciones de Nm de manera indirecta son soluciones teóricas simples que proveen una comprensión de los niveles de flujo génico, que son útiles en muchas situaciones y que hacen contribuciones significativas en ecología y evolución desde la década del 90.

Se testaron los niveles de flujo génico reciente mediante la incorporación de información *a priori* de la localización de los individuos al análisis bayesiano. Bajo este modelo $k=4$ es el número real de *clusters* genéticos, donde los individuos asignados a estos cuatro *clusters* coinciden con sus orígenes geográficos (Figura II-3 B). A excepción del individuo 61, el cual presentó una elevada probabilidad de ser un inmigrante de la población Tucumán, la totalidad de los individuos analizados no resultaron inmigrantes ni descendientes recientes de estos. A pesar de que las poblaciones de la región Yungas no presentaron estructuración genética tampoco mostraron intercambio reciente de migrantes. Esto se podría explicar por la gran distancia geográfica entre ambas poblaciones lo cual indicaría la necesidad de considerar poblaciones localizadas entre las dos poblaciones estudiadas de manera de analizar si existen relaciones entre la distancia geográfica y la distancia genética empleando el modelo de aislamiento por distancia. A partir de los resultados presentados, las poblaciones de la región Paranaense tampoco han intercambiado migrantes en las generaciones previas a pesar de su menor distancia geográfica, pudiendo atribuir a los niveles de flujo génico reciente entre las poblaciones de la región Paranaense un efecto nulo como fuerza que contrarresta a los efectos de la deriva genética en las poblaciones mientras que en las Yungas no debería ser descartado.

Las poblaciones de *A. colubrina* var. *cebil* estudiadas se encuentran estructuradas genéticamente, siendo los niveles de flujo génico histórico superiores a los niveles de flujo génico reciente. A nivel intrapoblacional, las poblaciones se comportarían como unidades panmícticas, ya que los individuos de las distintas poblaciones fueron asignados a *clusters* homogéneos coincidentes con sus poblaciones de origen al considerar la información acerca de su posición geográfica.

6. Conclusiones

- Las poblaciones de *A. colubrina* var. *cebil* estudiadas presentaron niveles elevados de diversidad genética nuclear siendo Jujuy la población que presentó mayor diversidad genética.
- Las poblaciones de *A. colubrina* var. *cebil* estudiadas presentaron una estructuración genética moderada y una mayor diferenciación entre las regiones.
- Los niveles de flujo génico histórico contrarrestaron a los efectos de la deriva genética.
- Los niveles de flujo génico reciente entre las poblaciones analizadas fueron nulos. En la región Yungas podría ser consecuencia de la elevada distancia geográfica que separa a las poblaciones consideradas mientras que en la región Paranaense podría estar indicando un elevado aislamiento reciente entre las poblaciones.



SECCION III

“Estudio del rol del flujo génico y de la deriva genética en la determinación de la estructura de las poblaciones fragmentadas de *Anadenanthera colubrina* var. *cebil* (Fabaceae, Fabales) a partir del empleo de marcadores microsatélites de ADN cloroplástico (cpSSRs)”

1. Introducción

Los patrones de dispersión de las semillas modelan la composición y estructura genética de las poblaciones vegetales. Las especies con niveles bajos de flujo génico por semillas poseen elevada probabilidad de presentar niveles considerables de heterogeneidad genética entre poblaciones locales, mientras que, las especies con flujo génico intensivo mediado por semillas presentan bajos niveles de estructura genética espacial (Hamrick *et al.*, 1993).

1.1 Importancia del flujo génico mediado por semillas en especies forestales

Tanto en angiospermas como en gimnospermas los procesos de regeneración y colonización de nuevos hábitats requieren del movimiento de diásporas e involucran el transporte de la información genética a través de las semillas. Las diásporas son frutos completos, partes de frutos que contienen semillas o bien las propias semillas (Finkeldey y Hattemer, 2007). Las semillas son dispersadas por medios bióticos o abióticos, siendo el agua y el viento los principales agentes abióticos mientras que entre los principales agentes bióticos se encuentran los animales. Además, debe destacarse el rol que juega el hombre en la distribución actual de las plantas y en los patrones de variación genética dentro de las especies (Finkeldey y Hattemer, 2007).

La dispersión limitada de las semillas resulta en complejas relaciones entre los patrones de distribución espacial de la variación genética y las relaciones de parentesco. En general, los árboles vecinos suelen ser genéticamente más similares que los árboles espacialmente distantes de la misma población formando así “estructuras familiares” (Finkeldey y Hattemer, 2007).

1.2 ADN citoplasmático: el genoma cloroplástico

El genoma de las plantas, como el de todos los eucariotas, está compartimentalizado, y la constitución del genoma de las organelas difiere de la del genoma nuclear, lo cual tiene profundas consecuencias en su función y evolución (Greiner *et al.*, 2011). El ADN cloroplástico (ADNcp) revela otra historia respecto al ADN nuclear. Este genoma es transmitido por vía materna en la mayoría de las angiospermas (Corriveau y Coleman, 1988), biparentalmente sólo en algunas especies y por vía paterna en la mayoría de las gimnospermas (Avisé, 2004). En angiospermas, al ser heredado por vía materna, se dispersa únicamente a través de las semillas (Hamilton, 1999) alcanzando niveles de diferenciación mayores que aquellos *loci* con herencia biparental (Fontaine *et al.*, 2004). En angiospermas el ADNcp generalmente exhibe niveles considerables de estructura genética espacial, estando los mayores niveles de variación distribuidos entre poblaciones geográficas (McCauley, 1995). Se ha comprobado que la información proveniente del ADN de las organelas es de gran utilidad para diferentes estudios evolutivos. El ADNcp ha sido ampliamente utilizado para estudios a nivel poblacional, ya que cuando este es heredado por vía materna se dispersa únicamente por semillas y la distribución

espacial de las variantes polimórficas del ADNcp contrasta con aquella obtenida para los marcadores biparentalmente heredados pudiendo, de esta manera ser de utilidad para el estudio de influencia relativa de la dispersión de las semillas y del polen sobre el flujo génico total (McCauley, 1995).

El ADN cloroplástico (ADNcp) es una molécula circular cuyo tamaño oscila entre 115 a 165 kb. Esta variación en el tamaño entre especies se debe principalmente al grado de reiteración de una larga repetición invertida que incluye genes para las subunidades de ARN ribosómico (Avise, 2004) El genoma cloroplástico de la mayoría de las plantas terrestres tiene una estructura cuatripartita que consta de dos repeticiones invertidas (IR), que separan dos regiones de copia única (SC), la región de copia única grande y pequeña (LSC y SSC, respectivamente) (Ravi *et al.*, 2008). Este genoma se halla en múltiples copias en cada célula siendo su organización estructural conservada en plantas superiores (Ravi *et al.*, 2008) presentando una evolución lenta (Zhan *et al.*, 2011). Las mutaciones más frecuentes en el ADNcp son las mutaciones puntuales y las inserciones/delecciones (*indels*) en regiones no codificantes (Palmer *et al.*, 1988). Además, el ADNcp carece de recombinación y debido a esto es posible interpretar los patrones de variación observados en los diferentes *loci* en conjunto y asignar haplotipos particulares de ADNcp a los diferentes individuos. La interpretación de la diversidad genética dentro y entre poblaciones se basa, convencionalmente, en la variación entre los haplotipos más que en la variación entre *loci* particulares (Finkeldey y Hattermer, 2007).

La tasa de cambio de la estructuración genética a través del flujo génico es afectada por el modo de transmisión del genoma. En comparación con los genomas de herencia biparental y los de herencia paterna, los marcadores heredados por vía materna detectan una marcada diferenciación genética entre poblaciones (Finkeldey y Hattermer, 2007), ya que normalmente las semillas se distribuyen a distancias más cortas que el polen (Zhan *et al.*, 2011). Además, al tratarse de un genoma haploide, el tamaño efectivo poblacional en plantas hermafroditas de fecundación cruzada, es la mitad que el correspondiente a genomas nucleares diploides (Provan *et al.*, 2001). De este modo, el flujo génico entre poblaciones de pequeño tamaño posee una menor efectividad para contrarrestar los efectos de la deriva genética en *loci* transmitidos vía materna (Petit *et al.*, 2005).

De esta manera, los marcadores del genoma cloroplástico son buenos indicadores de los efectos de posibles cuellos de botella, efecto fundador y deriva genética (Vendramin *et al.*, 1998; Provan *et al.*, 1999).

A pesar de las ventajosas cualidades del empleo de marcadores específicos del genoma cloroplástico, la mayor desventaja que presentan estos marcadores se relaciona a la baja tasa de mutación presente en este genoma, lo cual representa una limitación en cuanto a los niveles de variación intraespecífica capaz de ser detectados. Un marcador genético posee valor potencial para el estudio de los niveles de flujo génico dependiendo del grado de polimorfismo que

presente en la escala espacial a la cual es estudiado el flujo génico (McCauley, 1995). El hallazgo de marcadores microsatélites en el genoma cloroplástico (cpSSRs) ofrece nuevas oportunidades para el estudio de la variación citoplásmica intraespecífica aportando elevados niveles de resolución (Pakkad *et al.*, 2008)

1.3 Marcadores microsatélites de ADN cloroplástico

Los marcadores microsatélites de ADN cloroplástico (cpSSRs) consisten en motivos cortos repetidos en tándem de 8 a 15 veces (Navascués y Emerson, 2005), donde las repeticiones mononucleóticas A:T son las más abundantes (Deguilloux *et al.*, 2004). El principal mecanismo que origina los cpSSRs, debido a la ausencia de recombinación del genoma cloroplástico, es el desplazamiento erróneo de la polimerasa, en una o las dos hebras de ADN, durante la replicación (Karhu, 2001). Este desplazamiento particular o *slippage* produce la disociación transitoria de la cadena de ADN que se está replicando seguida de una hibridación fuera de fase, generando, de esta manera, la inserción o delección de una unidad de repetición (Ellegreen, 2004).

Hay dos características principales que diferencian los microsatélites cloroplásticos de los nucleares: los SSRs cloroplásticos tienen herencia uniparental y todos los *loci* cpSSRs constituyen un grupo de ligamiento, por cuanto el cromosoma cloroplástico es una molécula haploide no recombinante. De esta forma, la combinación de alelos para los diferentes *loci* cpSSRs constituyen los haplotipos cloroplásticos (Navascués y Emerson, 2005). Además, desde el punto de vista evolutivo son clonales y presentan una baja tasa de evolución (Navascués y Emerson, 2005). La tasa de mutación en las regiones cpSSRs es más elevada que la tasa de mutación en cualquier otra región del genoma cloroplástico (Navascués y Emerson, 2005) sin embargo, es menor que la tasa de mutación de los SSRs nucleares (Vendramin *et al.*, 1998; Deguilloux *et al.*, 2004).

Los marcadores de genomas de herencia uniparental permiten el estudio de la historia evolutiva de las poblaciones (Vendramin *et al.*, 1998). Los marcadores cpSSRs polimórficos a nivel intraespecífico constituyen herramientas útiles para el estudio de la estructura genética poblacional resultante del flujo génico mediado por semillas (Hamilton, 1999).

Al igual que en los microsatélites nucleares, la variación en *loci* cpSSRs puede ser detectada a partir de su amplificación empleando cebadores complementarios a las secuencias conservadas que flanquean el locus microsatélite de interés (Struss y Plieske, 1998). Sin embargo, a diferencia de los las regiones flanqueantes de los SSRs nucleares, el elevado grado de conservación de las secuencias en el genoma cloroplástico en plantas superiores tiene implicancias prácticas en los estudios genéticos. Entre ellas, la más importante es el facilitar la obtención de amplicones mediante el uso de cebadores interespecíficos (Weising y Gardner, 1999). En la actualidad, se cuenta con cebadores universales desarrollados para el estudio de la

variación del genoma cloroplástico (Hansen *et al.*, 2005; Deguilloux *et al.*, 2004). Estos cebadores universales son secuencias de ADN homólogas a las regiones codificantes altamente conservadas del genoma cloroplástico (Chung y Staub, 2003) y posibilitan la amplificación de intrones y espaciadores intergénicos cloroplásticos en un amplio rango de taxones de plantas superiores (Provan *et al.*, 2001). Taberlet (1991) promovió el uso de cebadores cloroplásticos a través de especies y géneros vegetales, con vistas al análisis de la variación genética intra e interespecífica (Heinze, 2007). Actualmente, se cuenta con cebadores universales para coníferas (Vendramin *et al.*, 1997), gramíneas (Provan *et al.*, 2001) y dicotiledóneas (Weising y Gardner, 1999).

1.4 Utilidad de los cpSSRs para estudios genético poblacionales

La información proveniente del genoma cloroplástico se ha convertido en una herramienta de uso corriente para los análisis filogenéticos, y ha sido empleado en el estudio de la dinámica de las zonas híbridas (McCauley, 1995). Además, el ADN cloroplástico es de gran utilidad para estudios poblacionales. Cuando este genoma es heredado por vía materna es dispersado únicamente por las semillas, por ello la tasa de migración del ADNcp (m_e) en angiospermas es potencialmente inferior a la tasa de migración para los genes nucleares (m). Cuando esto ocurre los valores de F_{ST} para el ADN cloroplástico pueden ser marcadamente superiores a aquellos de los genes nucleares (McCauley, 1995).

La estructura genética de las poblaciones es consecuencia de procesos evolutivos que se encuentran operando a través de largos períodos de tiempo. El cambio climático global del pasado, no solamente afectó a la distribución de las especies sino que también afectó a la variación genética dentro de las especies. Los eventos globales más importantes con consecuencias considerables sobre la distribución de las especies y sobre la distribución de la variación genética dentro de ellas fueron los cambios climáticos del Cuaternario. Los *loci* marcadores del genoma cloroplástico, de herencia uniparental y dispersados por semilla, típicamente revelan grupos genéticos distintos asociados con refugios glaciales y rutas de colonización (Tollefsrud *et al.*, 2009).

El análisis de la cantidad y distribución de la diversidad genética cloroplástica contribuye al conocimiento de la contribución del flujo génico histórico mediado únicamente por semillas en las poblaciones fragmentadas de *A. colubrina* var. *cebil*. Además, la baja tasa de evolución del genoma cloroplástico permite a partir de los patrones espaciales de distribución de la diversidad genética cloroplástica hacer inferencias acerca de la historia evolutiva de estas poblaciones al contrastar los patrones observados con las teorías biogeográficas propuestas para los Bosques tropicales estacionalmente secos.

2. Objetivos específicos

- Caracterizar la diversidad genética cloroplástica presente en las poblaciones naturales de *A. colubrina* var. *cebil* del Norte Argentino.
- Analizar la estructura genética en las poblaciones fragmentadas de *A. colubrina* var. *cebil* a partir de marcadores microsatélites de ADN cloroplástico.
- Indagar acerca del rol evolutivo del flujo génico mediado por semilla y de la deriva genética como fuerzas modeladoras de las frecuencias haplotípicas del genoma cloroplástico en las poblaciones fragmentadas.

3. Materiales y Métodos

3.1 Genotipificación de los individuos empleando cpSSRs

El ADN fue extraído empleando el kit de extracción DNeasy® – Plant minikit (Qiagen). Se emplearon cuatro pares de cebadores universales desarrollados por Weising y Gardner (1999) *Ccmp2*, *Ccmp4*, *Ccmp5* y *Ccmp7*. Estos cebadores fueron sintetizados incorporando una marca fluorescente (FAM=azul o HEX=verde) a su extremo 5'.

Las reacciones de amplificación se realizaron en un volumen final de 15 µl empleando 0,5 ng/µl de ADN, 1x de Buffer Hot Start, 2,5 mM de Cl₂Mg, 0,2 mM de cada dNTP, 1 U de enzima ADN polimerasa Hot Start (5 U/µl FirePol Solis BioDyne, Tartu, Estonia) y 0,33 pmol de cada cebador. La reacción de amplificación fue realizada en un termociclador con gradiente (Biometra, Göttingen, Alemania) empleando un programa de tipo *touchdown*. Las condiciones de ciclado empleadas fueron: Desnaturalización inicial: 94°C por 15 min, 10 ciclos de desnaturalización 1 min a 94°C, hibridación 1 min en un rango de temperatura de 60 °C a 50°C con una reducción de 1°C en cada ciclo y elongación por 1 min a 72 °C seguidos por 29 ciclos similares empleando una temperatura de hibridación de 50 °C.

El tamaño de los fragmentos fue asignado empleando el secuenciador ABI Prism® 3100 (Applied Biosystems) mediante el empleo del programa GENESCAN™. Para ello se empleó un marcador de peso estándar GS 500 ROX™ (Applied Biosystems).

3.2 Análisis estadístico de los datos

3.2.1 Caracterización de la diversidad genética cloroplástica

Porcentaje de loci polimórficos (P): corresponde al porcentaje de *loci* que resultaron polimórficos en cada población.

Número promedio de alelos por locus (N_A): también conocido como multiplicidad génica y corresponde al promedio de los distintos alelos identificados para cada locus en cada población (Gillet *et al.*, 2005).

Número efectivo de alelos por locus (N_E): fue calculado de la siguiente manera: $N_E = \frac{1}{\sum_1^n x_i^2}$

donde x_i corresponde a las frecuencias alélicas (Gillet *et al.*, 2005).

Número de alelos únicos (N_{AU}): corresponde al promedio de los alelos presentes en una única población para cada locus (Barton y Slatkin, 1986; Slatkin y Barton, 1989).

Índice de diversidad génica de Nei (h): fue calculado de la siguiente manera $h = \sum_1^n x_i^2$ (Nei, 1987); donde x_i corresponde a las frecuencias de alélicas.

Número de haplotipos (N_H): corresponde al número de haplotipos definidos en cada población. Los haplotipos fueron definidos a partir de la combinación de alelos para los diferentes *loci*.

Número de haplotipos únicos (N_{HU}): corresponde al número de haplotipos presentes en una única población.

Número de haplotipos compartidos (N_{HC}): corresponde al número de haplotipos compartidos entre las poblaciones.

Índice de diversidad haplotípica de Nei (H_{Ecp}): Fue calculado de la siguiente manera $H_E = \{[n/(n-1)][1-\sum p_i^2]\}$ (Nei, 1987) donde n es el número de individuos analizados en una población y p_i es la frecuencia del haplotipo i -ésimo en esa población.

Estas estimaciones fueron realizadas empleando el programa GenAlEx versión 6.41 (Peakall y Smouse, 2006) disponible en http://www.anu.edu.au/BoZo/GenAlEx/new_version.php.

3.2.2 Análisis de las relaciones genéticas entre haplotipos

Se construyó una red a partir del número de alelos diferentes entre todos los pares de haplotipos para graficar las distancias genéticas y de esta manera inferir la divergencia evolutiva entre los mismos. Una red de haplotipos permite conocer las relaciones genéticas existentes entre haplotipos ya que conecta a los haplotipos más cercanos entre sí de acuerdo a su similitud genética basándose en el número de mutaciones presentes y en los sitios diferentes entre todos los pares de haplotipos. Esta clase de redes muestra la variación en los alelos de los haplotipos debida a los eventos mutacionales que se han producido en la evolución, pero no indican la orientación temporal de dichos cambios (Templeton, 2006). Además, es posible relacionar esta red con información acerca de la localización geográfica de los haplotipos y acerca de la población de origen de cada variante.

Los haplotipos, definidos por la combinación de alelos para *loci* específicos, fueron considerados como las unidades taxonómicas operativas (*Operational taxonomic unit* - OTUs). Todos los enfoques basados en distancias se basan en una matriz OTU x OTU, que consiste en las distancias genéticas entre tales unidades estimadas de a pares. Para n OTUs, se tiene $n(n-1)/2$ distancias. Para la construcción de la red se emplearon algoritmos diseñados para moléculas no recombinantes. Así, se empleó el algoritmo Median-Joining el cual es de utilidad para el análisis de datos multiestado. Una red mediana (*median network*) consiste en nodos y conexiones entre estos nodos. Un vector mediano (*median vector*) consiste en un haplotipo hipotético el cual es requerido para conectar a los haplotipos existentes dentro de la red de manera parsimoniosa. Sin el vector mediano podría no existir conexión cercana entre los haplotipos existentes en el conjunto de datos. Los vectores medianos pueden ser interpretados de manera biológica como haplotipos existentes pero no muestreados o como haplotipos ancestrales extintos (Bandelt *et al.*, 1999).

Este análisis, empleando el algoritmo Median Joining fue desarrollado aplicando el programa Network versión 4.6.1.0 (Bandelt *et al.*, 1999) disponible en <http://www.fluxus-engineering.com/sharnet.htm>. En la red de haplotipos queda representada la frecuencia de cada

variante haplotípica mediante el tamaño de los círculos y la población de origen de cada variante haplotípica mediante diferentes colores.

3.2.3 Análisis de la estructura genética cloroplástica de las poblaciones

3.2.3.1 Distribución de la diversidad genética cloroplástica entre y dentro de poblaciones

Se realizó un análisis de la varianza molecular (AMOVA) para conocer la distribución de la variación genética en los diferentes niveles de subdivisión jerárquicos definidos mediante criterios no genéticos. El punto de partida para este análisis es una matriz de distancia euclidiana entre haplotipos.

En este estudio se consideraron tres niveles de subdivisión jerárquicos: entre regiones; entre poblaciones dentro de regiones y dentro de poblaciones.

Este análisis se desarrolló utilizando el software Arlequín 3.5 (Excoffier y Lischer, 2010) disponible en <http://cmpg.unibe.ch/software/arlequin35/>.

3.2.3.2 Estimación del índice de fijación F_{ST}

El índice de fijación de Wright para la población total (F_{ST} global) definido en términos de varianza fue computado a partir del AMOVA. La significancia estadística de este índice fue calculada mediante 1023 permutaciones al azar. Este análisis fue realizado empleando el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

El índice de fijación F_{ST} definido en términos de identidad por descendencia fue estimado para las poblaciones tomadas de a pares (F_{ST} de a pares) de igual manera que en el punto 3.3.3.2 de la sección II. El índice de fijación para el total de las poblaciones se estimó a partir del promedio de los valores de a pares (F_{ST} global). La significancia estadística fue testada mediante 110 permutaciones al azar. Este análisis fue realizado empleando el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

El índice de fijación tiene un mínimo teórico de 0, el cual indica ausencia de divergencia genética y un máximo teórico de 1, que indica fijación de alelos alternativos en diferentes poblaciones (Hartl y Clark, 2007).

3.2.3.3 Determinación de la estructura genética mediante análisis bayesiano

La estructuración genética entre poblaciones puede ser representada mediante grupos o *clusters* cuya composición alélica o haplotípica, a partir de los *loci* considerados, difieran estadísticamente (Corander *et al.*, 2008a).

Para inferir la estructura genética poblacional se realizó un análisis bayesiano usando un modelo de agrupamiento de tipo *mixture* (*Non-spatial genetic mixture analysis*) para *loci* ligados implementado por el programa BAPS versión 5.3 (Corander *et al.*, 2006; Corander *et al.*, 2008a). A partir de este análisis se determinó el número más probable de *clusters* en los

cuales se agrupan los individuos empleando únicamente información genética (haplotipos) y se asignaron los individuos integrantes a estos *clusters*. Bajo este modelo el número de poblaciones es tratado como un parámetro desconocido y la partición entre *clusters* como el parámetro de principal interés, sin probabilidades diferenciales *a priori* (Corander *et al.*, 2003). Además, se implementó un modelo de agrupamiento espacial de individuos (*Spatial genetic mixture analysis*) (Corander *et al.*, 2008a) ya que el modelo espacial mejora el poder estadístico para detectar la estructura poblacional cuando se dispone de un bajo número de *loci* (Corander *et al.*, 2008a). De esta manera, se testó el agrupamiento no espacial y espacial de los individuos para diferentes números de *clusters* con el objetivo de determinar cuál es el número de agrupamientos que presenta la mayor probabilidad *a posteriori*. Se testó el agrupamiento con un número de *clusters* de 1 a 6 ($k = 1-6$) realizando 10 repeticiones de cada valor de k .

Para determinar las relaciones existentes entre los *clusters* identificados por el análisis bayesiano, se construyeron dendrogramas usando la técnica de agrupamiento *Neighbor Joining* a partir de las matrices de distancia genética de Kullback-Leibler (KL) (Kullback y Leibler, 1951) las cuales fueron obtenidas desde los *clusters*. El índice KL puede ser empleado como una medida de distancia genética relativa entre la distribución de dos conjuntos de datos (Kasturi *et al.*, 2003). Esta medida de distancia genética fue empleada dado que no se pudo emplear el índice de distancia de Nei ya que la población Tucumán presentó un haplotipo único, lo cual llevó a que el índice de similitud resultara igual a 0 y la distancia genética ($D = -\ln 0$) infinita al estimar la distancia genética de Nei entre grupos de ambas regiones,. Dado que el estimador de distancia de Kullback-Leibler interpreta la cantidad ($0 \times \ln 0$) como cero, las distancias genéticas pueden ser representadas en un dendrograma.

Estos análisis fueron desarrollados utilizando el programa BAPS versión 5.3 (Corander *et al.*, 2006; Corander *et al.*, 2008b) disponible en <http://web.abo.fi/fak/mnf//mate/jc/software/baps.html>.

3.2.4 Cuantificación del flujo génico histórico mediado por semillas

Los métodos indirectos para estimar los niveles de flujo génico permiten inferir el número de migrantes que han sido intercambiados históricamente por generación ($N_e m$) a partir del nivel de diferenciación genética existente entre poblaciones (Burczyk *et al.*, 2004). Al estudiar *loci* de genomas haploides con herencia uniparental en una especie monoica, el tamaño efectivo se ve reducido en un factor de dos debido a que posee una copia de cada alelo.

Entonces, $F_{ST} = \frac{1}{(2Nm+1)}$, de manera que $Nm = \frac{1}{2} \left(\frac{1}{F_{ST}} - 1 \right)$ (Templeton, 2006).

Debido a las diferentes concepciones subyacentes a las estimas del índice de fijación, ya sea en términos de varianzas en las frecuencias alélicas o en términos de identidad por descendencia, es posible estimar el número efectivo de migrantes basado en el tamaño efectivo en términos de varianza o en el tamaño efectivo en términos de identidad por descendencia,

respectivamente. De esta manera, estas estimaciones reflejarán la magnitud de la fragmentación o de la identidad por descendencia sobre los niveles de flujo génico.

Se estimó el nivel de flujo génico global $N_{ev}m$ a partir de la relación $N_{ev}m = \frac{1}{2} \left(\frac{1}{F_{ST}} - 1 \right)$, donde N_{ev} es el tamaño efectivo en términos de varianza (*variance effective size*) el cual permite describir los efectos del error de muestreo y/o de la subdivisión poblacional como consecuencia de la deriva genética a partir de la medición de la varianza en las frecuencias alélicas a través de las generaciones (Templeton, 2006). Esta relación permite estimar $N_{ev}m$ global.

Se estimó el nivel de flujo génico global $N_{ef}m$ a partir de la relación $N_{ef}m = \frac{1}{2} \left(\frac{1}{F_{ST}} - 1 \right)$, donde N_{ef} es el tamaño efectivo en términos de endogamia (*inbreeding effective size*) el cual permite describir la acumulación de identidad por descendencia (\bar{F}) en una población como consecuencia de deriva genética (Templeton, 2006). Así, mediante esta relación se estimó $N_{ef}m$ entre pares de poblaciones y global.

4. Resultados

4.1 Genotipificación de los individuos empleando cpSSRs

Los cuatro marcadores resultaron polimórficos para las poblaciones estudiadas. En la Figura III-1 se presentan algunos de los alelos de cada *loci* cpSSRs, en tanto que en la Tabla III-1 se resume el número de alelos en cada *loci* y el rango de tamaño de los mismos. Dada la naturaleza haploide del genoma cloroplástico cada individuo presenta un alelo por locus cpSSR.

4.2 Análisis estadístico de los datos

4.2.1 Caracterización de la diversidad genética cloroplástica

Los marcadores generaron entre dos y cuatro alelos por locus alcanzando un total de 11 alelos al considerarse los cuatro *loci*. La combinación de alelos entre los diferentes *loci* determinó seis haplotipos (Tabla III-2). A excepción de Tucumán, que resultó monomórfica para los cuatro marcadores, las tres poblaciones restantes presentaron un 25% de *loci* polimórficos (Tabla III-3). Cada población presentó en promedio un alelo por locus mientras que en el total de los individuos cada locus presentó en promedio tres alelos por locus (Tabla III-3). El número efectivo de alelos por locus tomo valores cercanos a la unidad y resultó menor al número de alelos por locus en la mayoría de las poblaciones debido a la presencia de alelos poco frecuentes en esas poblaciones. La excepción fue la población Santa Ana, la cual presentó un número de alelos por locus similar al número efectivo de alelos por locus debido a que los alelos presentes en esta población fueron igualmente frecuentes. Por su parte, la población Jujuy presentó mayor número de alelos únicos mientras que Santa Ana no presentó alelos únicos y resultó la población más polimórfica. Santa Ana además, presentó mayor diversidad génica ($h=0,12$) y mayor diversidad haplotípica ($H_{Ep}=0,50$).

En la región Paranaense ambas poblaciones presentaron dos haplotipos cada una (Candelaria HA y HB; Santa Ana HB y HC) siendo uno de ellos compartido entre las poblaciones (HB). En la región de las Yungas las poblaciones presentaron valores inferiores de diversidad génica y haplotípica. En esta región, Tucumán presentó un solo haplotipo (HD) el cual además, fue único y del mismo modo, Jujuy presentó dos haplotipos únicos (HE y HF) (Tabla III-2).

4.2.2 Análisis de las relaciones genéticas entre los haplotipos

En la red de haplotipos es posible identificar a los cuatro haplotipos más frecuentes. Entre ellos, únicamente el haplotipo HB fue compartido entre las dos poblaciones de la región

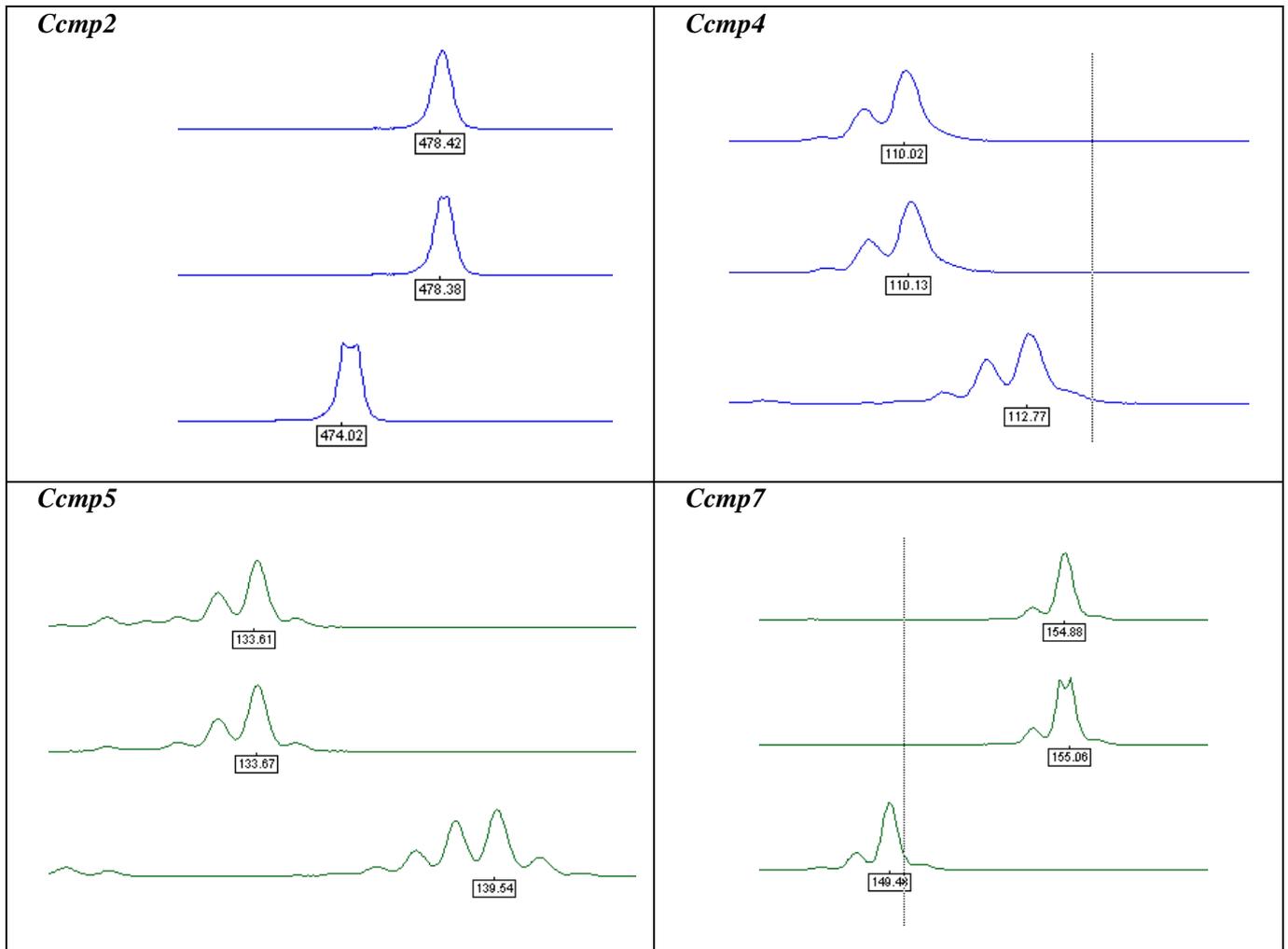


Figura III-1: Alelos revelados por los cuatro marcadores *Ccmp* en las poblaciones de *A. colubrina var. cebil* identificados mediante un secuenciador analizador de fragmentos.

Tabla III-1: Caracterización de los *loci* cpSSRs en *Anadenanthera colubrina* var. *cebil*

Locus	Secuencias cebadores (5'-3')	Posición en <i>N. tabacum</i>	Localización en <i>N. tabacum</i>	Motivo de repetición en <i>N. tabacum</i>	Tamaño de los fragmentos en <i>N. tabacum</i>	Tamaño de los fragmentos en <i>A. colubrina</i>	N _a en <i>A. colubrina</i>
<i>Ccmp2</i> *	F: gatcccgacgtaatcctg R: atcgtaccgaggggtcgaat	8609	5' trnS	(A) ₁₁	189	478 – 491	3
<i>Ccmp4</i> *	F: aatgctgaatcgaycaccta R: ccaaaatattbgaggactct	12872	intron atpF	(T) ₁₃	126	109 – 112	2
<i>Ccmp5</i> **	F: tgtccaatatcttctgtcatt R: aggtccatcggacaattat	16950 16977	3' rps2	(C) ₇ (T) ₁₀ (T) ₅ C(A) ₁₁	121	133 – 140	4
<i>Ccmp7</i> **	F: caacatataccactgtcaag R: acatcattattgtactcttc	57339	región intergénica atpB–rbcL	(A) ₁₃	133	150 – 155	2

*cebador *Forward* etiquetados con FAM su extremo 5'

**cebador *Forward* etiquetados con HEX su extremo 5'

Tabla III-2: Descripción y frecuencias de los seis haplotipos cloroplásticos presentes en las poblaciones de *Anadenanthera colubrina* var. *cebil* analizadas.

Haplotipo	Ccmp2	Ccmp4	Ccmp5	Ccmp7	Población				Total
					Candelaria	Santa Ana	Tucumán	Jujuy	
HA	478	112	133	155	18	-	-	-	18
HB	478	112	134	155	02	10	-	-	12
HC	478	109	134	155	-	06	-	-	06
HD	491	109	140	150	-	-	14	-	14
HE	474	109	140	150	-	-	-	18	18
HF	474	109	139	150	-	-	-	01	01
Total					20	16	14	19	69

Tabla III-3: Caracterización de la diversidad genética revelada por los marcadores cpSSRs en las poblaciones estudiadas.

Región	Población	N	P (%)	N _A	N _E	N _{AU}	h	N _H	N _{HU}	N _{HC}	H _{Ecp}
Paranaense	Candelaria	20	25	1,250	1,055	1,000	0,045	2	1	1	0,400
	Santa Ana	16	25	1,250	1,221	0,000	0,117	2	1	1	0,500
Yungas	Tucumán	14	0	1,000	1,000	1,000	0,000	1	1	0	0,000
	Jujuy	19	25	1,250	1,028	2,000	0,025	2	2	0	0,094
Total		69	100	2,750	1,076*	0,250*	0,047*	6	-	-	0,249*

*valor promedio por población, en negrita se resaltan los mayores valores por población.

N=tamaño de la población, P= porcentaje de *loci* polimórficos, N_A=número promedio de alelos por locus, N_E=número efectivo de alelos por locus, N_{AU}=número promedio de alelos únicos por locus, *h*=índice de diversidad génica de Nei, N_H=número de haplotipos, N_{HU}=número de haplotipos únicos N_{HC}=número de haplotipos compartidos y H_{Ecp}=índice de diversidad haplotípica de Nei.

Paranaense. Los haplotipos de la región Paranaense se encuentran muy cercanos en la red lo cual indicaría que también lo están genéticamente. A diferencia, los haplotipos de las poblaciones de la región de las Yungas fueron propios de cada población y se ubicaron distanciados en la red (Figura III-2). Además, fueron necesarios dos vectores medianos para relacionar de manera parsimoniosa a los haplotipos existentes.

Esta red de haplotipos refleja presencia de estructuración genética en estas poblaciones. A su vez, los haplotipos de la región de las Yungas son propios de cada población y se encuentran distantes entre sí en la red reflejando subestructuración en la región de las Yungas.

4.2.3 Análisis de la estructura genética poblacional

4.2.3.1 Distribución de la diversidad genética cloroplástica entre y dentro de poblaciones

El AMOVA indicó que el 69,69% de la variación estuvo contenido en el nivel entre regiones, mientras que nivel entre poblaciones dentro de regiones, contuvo el 25,10% de variación. Las poblaciones de la región Yungas podrían ser las principales responsables de estos niveles de variación dentro de las regiones. Por su parte, el nivel dentro de poblaciones presentó un nivel de variación bajo, aproximadamente de 5% (Tabla III-4). Estos resultados también reflejan estructuración de la variación genética principalmente a nivel regional.

4.2.3.2 Estimación del índice de fijación F_{ST}

El índice F_{ST} arrojó valores significativos y elevados, tanto al ser definido en términos de varianza en las frecuencias alélicas como al ser definido en términos de identidad por descendencia 0,95* y 0,91*, respectivamente. Estos valores indican niveles extremos de estructuración genética, casi al límite de la fijación de distintos haplotipos en las diferentes poblaciones.

4.2.3.3 Determinación de la estructuración genética poblacional mediante análisis bayesiano

El análisis bayesiano permitió identificar *clusters* genéticos con elevada probabilidad de representar unidades panmícticas. En el análisis no espacial, $k=6$ presentó la mayor probabilidad *a posteriori*. Estos seis *clusters* se hallan integrados en mayor proporción por individuos pertenecientes a una población geográfica (Figura III-3 A, Tabla III 4 A), destacándose el *cluster* 5 el cual se diferencia dentro de la población Santa Ana. Por otra parte, dos individuos de la población Candelaria (individuos 19 y 20) tuvieron mayor probabilidad de integrar el *cluster* 2, el cual se halla integrado en mayor proporción por individuos de Santa Ana. Dentro del *cluster* 4, integrado en mayor proporción por individuos de Jujuy, existe un solo individuo (individuo 59) que forma un *cluster* diferente (*cluster* 6). El dendrograma construido presentó dos agrupamientos mayores que representan las relaciones existentes

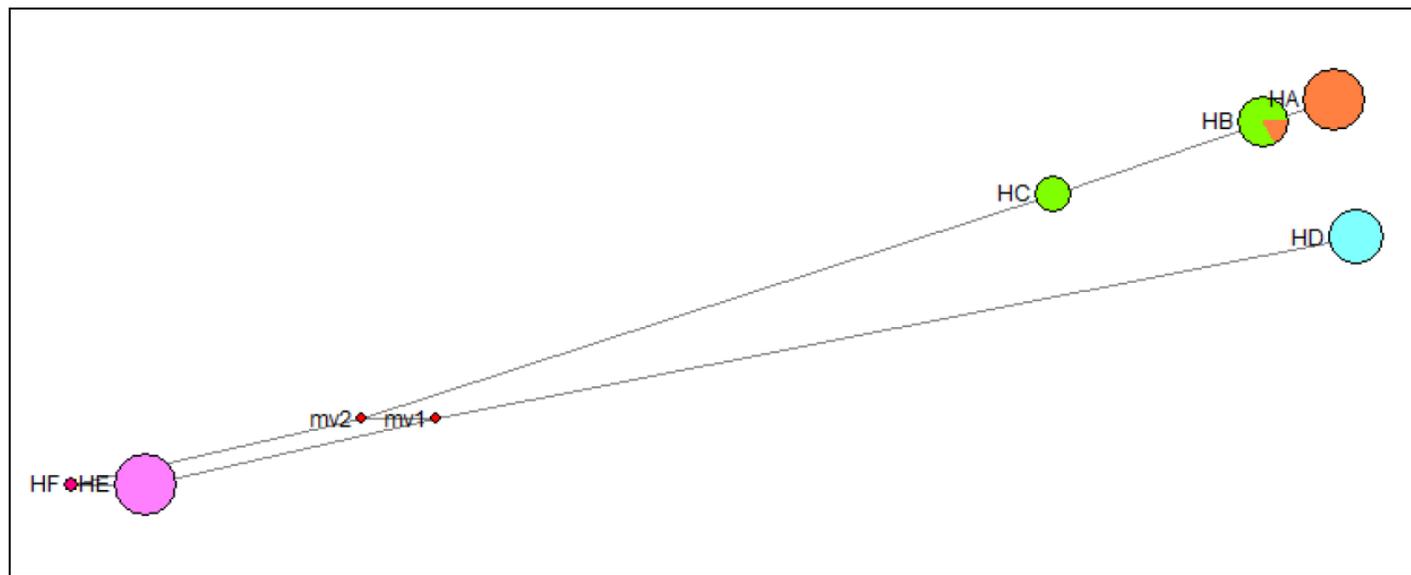


Figura III-2: Red de haplotipos. Cada círculo representa un haplotipo mientras que su tamaño representa la frecuencia. Los colores representan las poblaciones de origen. ■ Candelaria, ■ Santa Ana, ■ Tucumán y ■ Jujuy.
 mv1 y mv2: Vectores medianos 1 y 2.

Tabla III-4: Análisis de la varianza molecular (AMOVA).

Fuente de Variación	Grados de libertad	Suma de cuadrados	Componentes de varianza	Porcentajes de variación
Entre regiones	1	54,593	1,337 Va	69,690
Entre poblaciones dentro de regiones	2	16,533	0,482 Vb	25,110
Dentro de poblaciones	65	6,497	0,010 Vc	5,200
Total	68	77,623	1,919	

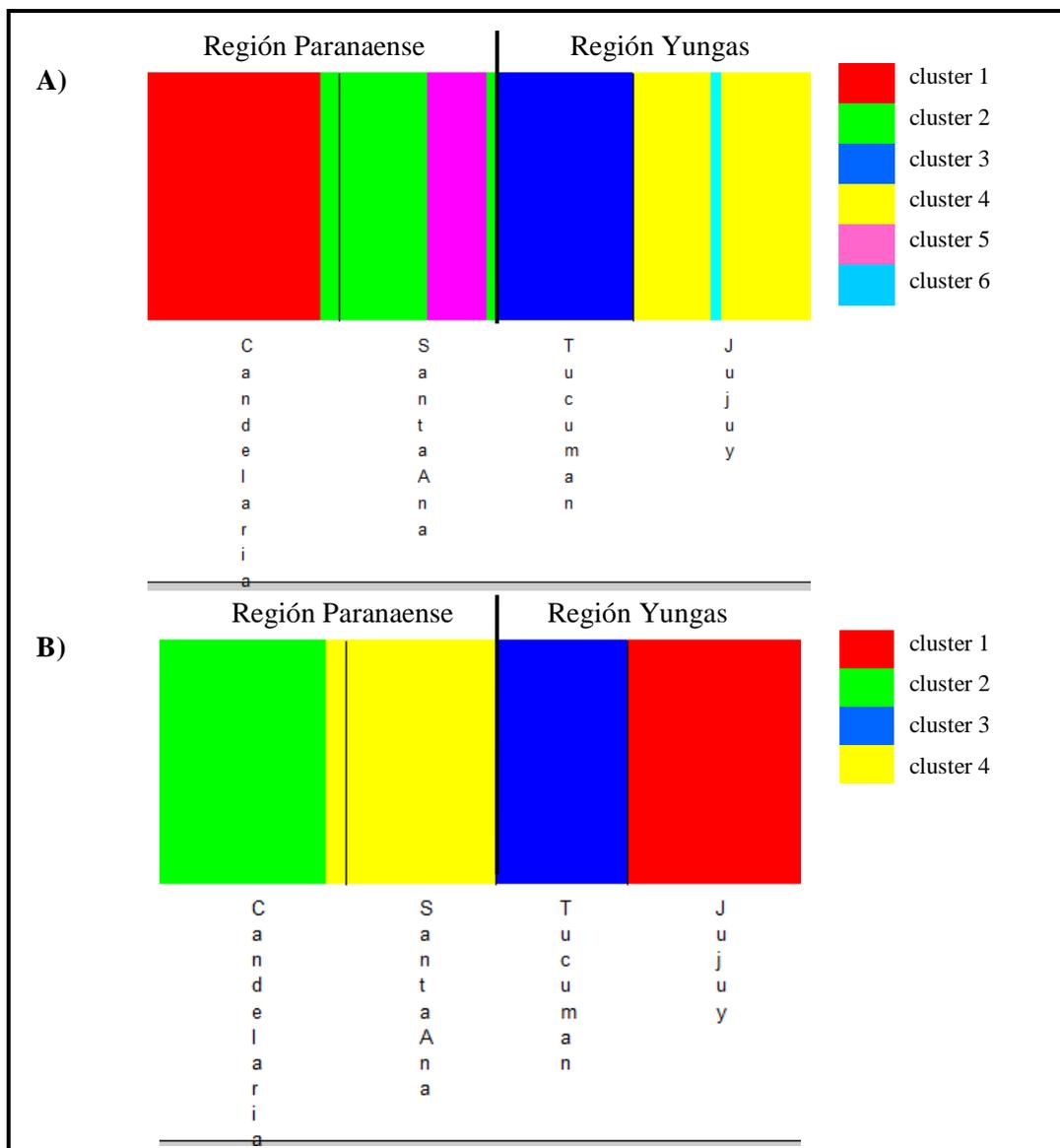


Figura III-3: Análisis bayesiano de la estructura genética poblacional. Modelo de tipo *mixture* para *loci* ligados a nivel de individuos: A) análisis no espacial y B) análisis espacial.

Tabla III-5: Análisis bayesiano de la estructura genética poblacional. Asignación de los individuos a cada *cluster*: A) análisis no espacial y B) análisis espacial.

A)

<i>Cluster</i>	N° de individuos	Población de origen de los individuos			
		Candelaria	Santa Ana	Tucumán	Jujuy
1	18	18 (100%)	0	0	0
2	12	2 (17%)	10 (83%)	0	0
3	14	0	0	14 (100%)	0
4	18	0	0	0	18 (100%)
5	6	0	6 (100%)	0	0
6	1	0	0	0	1 (100%)
Total	69	20	16	14	19

B)

<i>Cluster</i>	N° de individuos	Población de origen de los individuos			
		Candelaria	Santa Ana	Tucumán	Jujuy
1	19	0	0	0	19 (100%)
2	18	18 (100%)	0	0	0
3	14	0	0	14 (100%)	0
4	18	2 (11%)	16 (89%)	0	0
Total	69	20	16	14	19

En negrita se indican los valores que corresponde a la población con mayor número de individuos en cada *cluster*

entre estos seis *clusters* (Figura III-4). El primer agrupamiento contuvo a los *clusters* 3, 4 y 6, cada uno de ellos integrado en su totalidad por individuos pertenecientes a la región de las Yungas, mientras que el segundo agrupamiento contuvo a los *clusters* 1, 2 y 5, cada uno de ellos integrado en su totalidad por individuos pertenecientes a la región Paranaense. Así, el agrupamiento de los *clusters* a partir de su similitud genética marcó la separación entre las regiones.

En el análisis espacial, es decir, al incluir la localización geográfica como información *a priori* para el cálculo de las probabilidades, $k=4$ presentó la mayor probabilidad *a posteriori* (Figura III-3 B, Tabla III-4 B). Los *clusters* 1 y 3 se encuentran integrados por individuos de las poblaciones Jujuy y Tucumán, respectivamente. Por su parte, los *clusters* 2 y 4 se encuentran integrados principalmente por individuos pertenecientes a las poblaciones Candelaria y Santa Ana, respectivamente. Sin embargo, los individuos 19 y 20 pertenecientes a la población Candelaria continúan siendo asignados al *cluster* 4 integrado por individuos de Santa Ana.

4.2.4 Cuantificación del flujo génico histórico mediado por semillas

El número efectivo de migrantes basado en el tamaño efectivo en términos de varianza en las frecuencias alélicas presentó un valor de $N_{ev}m=0,03$; mientras que el número efectivo de migrantes basado en el tamaño efectivo en términos de identidad por descendencia presentó un valor de $N_{ef}m=0,05$. Ambos resultados no superaron la unidad lo cual indicaría que el nivel de flujo génico histórico mediado por semillas no es suficiente para contrarrestar los efectos de la deriva genética. Puede inferirse que la fragmentación poblacional ha tenido mayor incidencia en la varianza en las frecuencias alélicas ya que $N_{ef}m$ mostró un valor mayor a $N_{ev}m$.

Además, entre las poblaciones de diferentes regiones y entre las poblaciones de la región Yungas se evidencia que el flujo génico mediado por semillas se ha interrumpido completamente ya que las poblaciones no comparten haplotipos y la deriva genética ha llevado a la fijación de diferentes haplotipos.

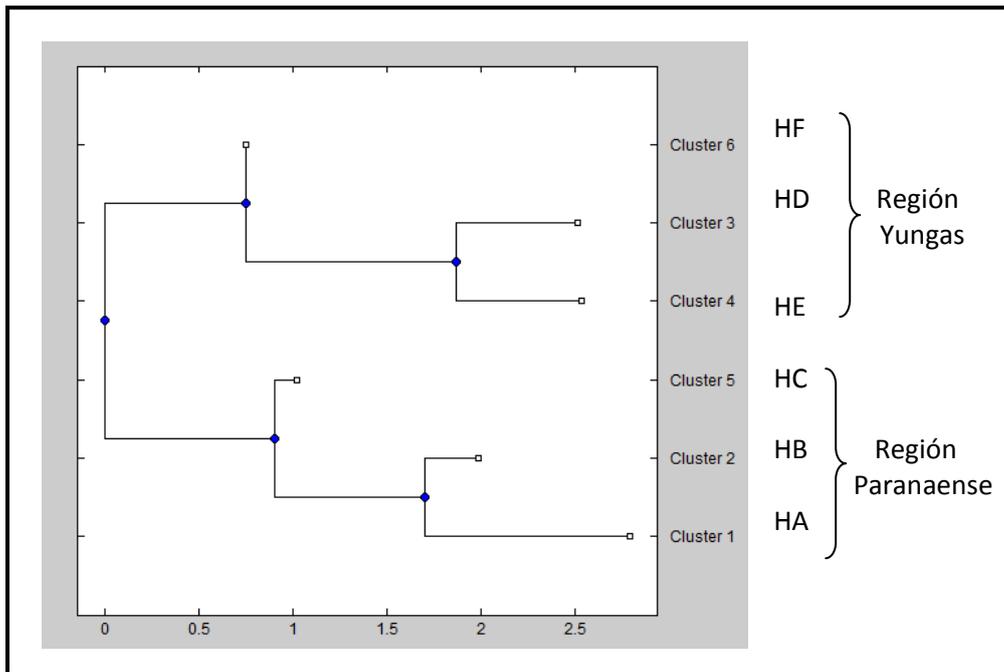


Figura III-4: Relaciones entre los *clusters* definidos por el análisis bayesiano de la estructura genética sin información espacial *a priori*. Se indican los haplotipos que definieron cada *cluster* y las regiones a las que pertenecen.

5. Discusión

La herencia materna del genoma cloroplástico en *A. colubrina* var. *cebil* permite el empleo de marcadores cpSSRs para inferir los niveles de flujo génico histórico mediado por semillas. Debido a que el tamaño efectivo del genoma cloroplástico representa la mitad del tamaño efectivo del genoma nuclear, los efectos de la deriva genética serán más pronunciados sobre el primero. Por estos motivos, a nivel del genoma cloroplástico la estructura genética poblacional se verá afectada en mayor medida por la acción de la deriva genética, mientras que el rol del flujo génico, como fuerza que contrarresta a los efectos de la deriva genética, dependerá de los mecanismos de dispersión de semillas de la especie. Además, los marcadores del genoma cloroplástico brindan información útil acerca de la evolución histórica de las poblaciones debido a su tasa de evolución lenta, de manera que la distribución geográfica contemporánea de los haplotipos cloroplásticos reflejará en mayor medida las consecuencias de los eventos históricos. Sumado a lo anteriormente expuesto, los niveles reducidos de flujo génico mediado por semilla y la falta de recombinación en este genoma, harán que las poblaciones necesiten más tiempo para borrar las huellas del pasado. Por el contrario, las diferencias entre poblaciones se incrementarán por la acción más intensa de la deriva genética debido al menor tamaño efectivo poblacional presentado por este genoma.

Los patrones de dispersión de semillas pueden moldear la composición y la estructura genética de las poblaciones vegetales. Especies con un flujo de semillas limitado probablemente poseen heterogeneidad genética entre la descendencia de distintos individuos, mientras que especies con dispersión de semillas extensiva presentan menor estructura genética (Hamrick *et al.*, 1993). La estructuración de la variación genética citoplasmática también se encuentra afectada por la interacción de la dispersión de las semillas con otros procesos ecológicos y genéticos. Los patrones de deposición de las semillas, la dispersión del polen, la densidad de los adultos, la selección de micro hábitat y varios aspectos de la ecología de la especie podrían tener efectos significativos sobre los patrones de variación genética dentro de las especies (Hamrick *et al.*, 1993). Mientras que, tanto la dispersión del polen como la dispersión de semillas determinan el flujo génico en plantas, la dispersión de semillas es de mayor importancia debido a que permite a las especies colonizar hábitats y por lo tanto influye en la dinámica de las poblaciones (Ndiade-Bourobou *et al.*, 2010). Las especies que producen semillas con bajo potencial de dispersión generalmente poseen menor diversidad genética en el ADN cloroplástico y elevados niveles de diferenciación genética poblacional (Pakkad *et al.*, 2008). Los resultados del presente trabajo muestran que estas generalizaciones se cumplen en las poblaciones de *A. colubrina* var. *cebil* consideradas. Los individuos estudiados presentaron baja diversidad genética en los cuatro *loci* cpSSR analizados. Se identificaron seis haplotipos, de los cuales sólo uno fue compartido entre las poblaciones de la región Paranaense, mientras que las poblaciones de las Yungas presentaron haplotipos únicos en cada población. Santa Ana fue la

población más diversa ($h=0,117$ y $H_E=0,50$). A pesar de que los marcadores cpSSRs revelaron niveles reducidos de variación dentro de las poblaciones, resultaron eficaces para detectar la variación entre poblaciones y entre regiones.

En la red de haplotipos es posible identificar a los cuatro haplotipos más frecuentes, siendo los haplotipos de las poblaciones Tucumán y Jujuy característicos de estas, mientras que los haplotipos de las poblaciones Candelaria y Santa Ana se ubicaron próximos en la red, reforzando esta relación la presencia de un haplotipo compartido (HB). El haplotipo HB podría ser considerado un haplotipo ancestral ya que se encuentra a mayor frecuencia y es compartido por las poblaciones de la región Paranaense. Además, a pesar de la reducida distancia geográfica entre las poblaciones de esta región, las mismas presentan haplotipos propios que se diferencian entre sí por un bajo número de eventos mutacionales. Esto podría reflejar los efectos de la fragmentación reciente causada por el impacto antrópico en esta región. La proximidad genética y geográfica entre haplotipos en la región Paranaense podría ser explicada atendiendo a dos hipótesis alternativas. Una de ellas considera la existencia de flujo génico mediado por semillas entre estas poblaciones, mientras que la segunda hipótesis postula la existencia de una distribución continua ancestral que incluye a estas dos poblaciones. Debido a que las poblaciones de la región Paranaense están separadas entre sí por una distancia geográfica lineal de aproximadamente 17 km, es poco probable la existencia de flujo génico mediado por semillas entre estas poblaciones, por ende, la segunda hipótesis sería la explicación más plausible para lo observado. Debe destacarse que el mecanismo de dispersión de semillas presentado por *A. colubrina* var. *cebil* es de tipo autocoria-anemocoria, lo cual conduce a que la descendencia se establezca con una distribución cercana a la planta madre.

Por su parte, los haplotipos de la región de las Yungas se diferenciaron por un elevado número de eventos mutacionales de los haplotipos de la región Paranaense. Las grandes distancias existentes entre los haplotipos estarían indicando aislamiento geográfico antiguo, y podría relacionarse esta diferenciación con la distancia geográfica que separa estas poblaciones, siendo la distancia lineal entre las Yungas y la región Paranaense mayor a 1000 km. La fijación del haplotipo HD en la población Tucumán, sería la consecuencia actual de la deriva genética actuando tras largos períodos de aislamiento.

Todos los procesos microevolutivos son capaces de producir cambios en las frecuencias alélicas sólo cuando existe variación genética. La fijación de haplotipos diferentes en grandes áreas de la misma población, sin la ayuda de la acción de la selección, indicaría que no se encuentran operando otros procesos que modifican las frecuencias alélicas (Wright, 1943). A medida que las poblaciones de una especie se fragmentan, el flujo génico se reduce y las poblaciones locales pueden presentar temporalmente variantes alélicas o haplotípicas particulares fijadas. La naturaleza aleatoria de la deriva genética hace poco probable que las poblaciones locales presenten las mismas variantes fijadas. De esta manera, si una población se

fragmenta en pequeñas subpoblaciones aisladas, estas evolucionaran en diferentes complementos alélicos y fijarán en última instancia alelos diferentes (Gillet *et al.*, 2005) dado que, en ausencia de mutación y de flujo génico, la deriva genética indefectiblemente tiende a la fijación en todos los *loci*.

A pesar de que las poblaciones de la región Paranaense no se encuentran en áreas protegidas aún mantienen niveles de diversidad haplotípica moderados ($H_E=0,45$), pudiendo reflejar diversidad genética antigua. Por su parte, las poblaciones consideradas en la región de las Yungas se encuentran en áreas protegidas aunque, desde el punto de vista del genoma cloroplástico, su diversidad genética haplotípica es baja ($H_E=0,047$). Estos resultados indican procesos de aislamiento antiguo, donde el flujo génico restringido mediado por semilla, sumado al menor tamaño efectivo y a la evolución lenta del genoma cloroplástico, ha llevado a la fijación de haplotipos como consecuencia de la deriva genética.

En cuanto a la distribución de la variación genética entre y dentro de poblaciones se encontró que alrededor del 70% de la variación puede ser atribuible a la diferenciación entre las regiones, mientras que el 25% de la variación total estuvo contenido entre poblaciones dentro de regiones. Las diferencias genéticas entre las poblaciones de la región Yungas podrían ser las principales responsables del porcentaje de variación dentro de las regiones ya que, a pesar de encontrarse en la misma región, cada población posee haplotipos únicos fijados.

El índice F_{ST} presentó valores elevados estadísticamente significativos, siendo que la estructuración medida a partir de la definición de varianza en las frecuencias alélicas fue superior a los niveles de estructuración definidos en términos de identidad por descendencia ($F_{ST}=0,95$ y $0,91$, respectivamente), esto indicaría que existe mayor efecto de la fragmentación sobre la distribución de la diversidad genética. Valores de F_{ST} cercanos a uno indican la fijación de diferentes haplotipos en las distintas poblaciones. *A. colubrina* var. *cebil* presenta una estructuración similar a *Caesalpinia echinata* ($F_{ST}=0,91$) cuyas semillas, al igual que las de *A. colubrina* var. *cebil*, son dispersadas por gravedad (autocoria) (Lira *et al.*, 2003). Respecto a los estudios de la distribución de la diversidad genética cloroplástica realizados en otras especies de los SDTF, Caetano *et al.* (2008) y Caetano y Naciri (2009) mediante secuencias de ADN cloroplástico estudiaron poblaciones naturales de *Astronium urundeuva* ($F_{ST}=0,97$) y *Geoffroea spinosa* ($F_{ST}=0,99$), respectivamente. A partir de sus resultados, sugirieron ausencia de flujo génico mediado por semillas y una divergencia marcada entre las poblaciones probablemente prepleistocénica.

En *A. colubrina* var. *cebil* el grado de estructuración fue reflejado además por los *clusters* identificados mediante el análisis bayesiano. Mediante este análisis se determinó mayor probabilidad de existencia de seis *clusters*, los cuales pueden ser considerados poblaciones desde el punto de vista genético. Cada uno de estos *clusters* quedó definido por uno de los seis haplotipos identificados en estas poblaciones. El dendrograma agrupó a los *clusters* según la

región de origen de los haplotipos que los definen reflejando la máxima diferenciación existente entre las regiones (Figura III-4). Además, al comparar los niveles de similitud genética entre los *clusters* es posible identificar similitudes genéticas similares entre los *clusters* dentro de cada región. Es decir, el nodo que agrupa a los *clusters* 3 (HD) y 4 (HE) denota una similitud genética similar a la presentada por el nodo que agrupa a los *clusters* 1 (HA) y 2 (HB). Esta observación es relevante debido a que, dentro de la región Paranaense, puede esperarse estos niveles elevados de similitud, ya que las poblaciones se encuentran geográficamente cercanas. Sin embargo, en la región de las Yungas, la mayor distancia geográfica entre las poblaciones consideradas permitiría suponer menor similitud genética. La pertenencia de ambas poblaciones a el núcleo Pedemonte subandino implicaría una dinámica histórica similar con haplotipos ancestrales compartidos.

Los resultados del modelo espacial de análisis refuerzan la estructuración de la variación haplotípica, ya que al incorporar información *a priori* acerca de la localización geográfica de los individuos, se establecieron cuatro *clusters* cada uno de ellos integrado en su mayoría por individuos pertenecientes a una única población, siendo la excepción los individuos 19 y 20 de Candelaria ya que fueron asignados al *cluster* integrado por individuos de Santa Ana. Estos resultados demuestran que el rol del flujo génico reciente mediado por semillas sería nulo.

En general, las poblaciones tienen historias complejas que consisten en fases de expansión y declive, y los patrones de flujo génico entre ellas pueden fluctuar en tiempo y espacio (Corander *et al.*, 2008b). Cuando las partes de una población se vuelven relativamente aisladas unas de otras, las fuerzas estocásticas remodelan la composición genética de los individuos a lo largo de las generaciones (Corander *et al.*, 2008b).

Werneck *et al.* (2011) estudiaron la distribución histórica de los SDTF mediante modelos de paleodistribución y evidencias palinológicas, prediciendo que las mayores áreas de distribución actuales de los SDTF representan áreas de estabilidad a largo plazo, concluyendo que los núcleos de SDTF poseen zonas de estabilidad histórica y actúan como refugios tanto históricos como actuales de estos bosques (áreas de color rojo en la Figura I-5). Además, han identificado zonas temporalmente inestables (áreas de color amarillo Figura I-5) en las cuales se espera una colonización más reciente, y consecuentemente, deberían retener señales genéticas de expansiones poblacionales y presentar niveles de diversidad genética menores. Siguiendo estos mapas y teorías, podríamos considerar que las poblaciones de la región Paranaense, pertenecientes al núcleo Misiones, serían poblaciones más antiguas y estables que han actuado como refugios. Esto explicaría el mayor número de haplotipos detectado y el haplotipo compartido entre las poblaciones Candelaria y Santa Ana. En relación a las poblaciones de la región de las Yungas, pertenecientes al núcleo Pedemonte Subandino, serían poblaciones cuya colonización ha sido más reciente lo cual explicaría la presencia de haplotipos propios en cada población. Además, Tucumán presenta un único haplotipo (HD), esta menor diversidad y

elevada similitud genética entre los haplotipos de la región de las Yungas podría coincidir con la localización en un área probablemente inestable según Werneck *et al.* (2011). Sería de interés para futuros estudios incorporar muestras a lo largo del gradiente latitudinal en esta región para testar esta hipótesis.

Actualmente, existen controversias en cuanto al impacto de las fluctuaciones climáticas del Cuaternario sobre la distribución de los SDTF (Prado y Gibbs, 1993; Pennington *et al.*, 2004; Caetano *et al.*, 2008; Pennington *et al.*, 2009, Werneck *et al.*, 2011; Collevatti *et al.*, 2012). Estos estudios en su mayoría coinciden en la posibilidad de una mayor extensión antigua de los SDTF seguida de un evento de vicarianza, que llevó a la fragmentación en la distribución de las especies que integran este bosque. Sin embargo, no existe consenso en cuanto a la época en la cual se produjeron estos hechos. Por su parte algunos autores sostienen que ocurrió en el Cuaternario más precisamente en el máximo glacial Pleistocénico (LGM) (Prado y Gibbs, 1993; Caetano *et al.*, 2008; Collevatti *et al.*, 2012) mientras que otros sostienen que los eventos del Cuaternario están siendo sobredimensionados y que posiblemente el mayor impacto sobre este bosque fue producido por los cambios ocurridos en el Terciario (Pennington *et al.*, 2009; Werneck *et al.*, 2011). Werneck *et al.* (2011) en discrepancia con Prado y Gibbs (1993) proponen un escenario alternativo de expansión de los SDTF durante el Terciario tardío, seguidos de una fragmentación durante el LGM y una expansión gradual secundaria hacia el sur desde principios del Holoceno argumentando que probablemente el clima del LGM ha sido extremadamente seco y frío para estos bosques.

Los resultados obtenidos en el presente trabajo muestran marcadas diferencias entre las regiones analizadas concordando con un aislamiento geográfico antiguo. Además, dentro de la región de las Yungas se observaron menores niveles de diversidad genética cloroplástica, mayores niveles de diferenciación junto a niveles elevados de similitud genética entre sus haplotipos los cuales representarían eventos de colonización posteriores, aunque no es posible hacer inferencias acerca del tiempo de divergencia entre las regiones, ni entre las poblaciones dentro de las mismas debido al marcador empleado.

Los niveles bajos de diversidad genética y los niveles elevados de estructuración genética cloroplástica en las poblaciones naturales de *A. colubrina* var. *cebil* analizadas concuerdan con los patrones informados por Lira *et al.* (2003) quienes estudiaron poblaciones naturales de *Caesalpinia echinata* Lam., una especie tropical que se distribuye en la costa brasileña de los SDTF. Estos autores propusieron que los reducidos niveles de diversidad genética citoplasmática y los elevados niveles de diferenciación entre regiones sugieren flujo génico limitado por medio de las semillas a través de prolongados períodos de tiempo. Del mismo modo, en *A. colubrina* var. *cebil* el número efectivo de migrantes no superó la unidad ($N_e m = 0,03$) lo cual implica que el rol del flujo génico histórico, mediado por semillas, ha sido

limitado y que la deriva genética ha jugado un rol principal sobre la estructura genética cloroplástica superando de esta manera a los efectos del flujo génico.

Quiroga *et al.* (2012) estudiaron poblaciones naturales de *Podocarpus parlatorey* distribuidas latitudinalmente en la provincia biogeográfica de las Yungas analizando una región intergénica del ADN cloroplástico. En dicho estudio identificaron haplotipos comunes y niveles de diversidad genética similares en el rango completo de distribución de esta especie interpretándolos como resultado de flujo génico histórico continuo durante los períodos fríos del Pleistoceno y del Holoceno temprano. Sin embargo, han detectado haplotipos únicos en los extremos de su distribución, interpretando a estos haplotipos como una señal de aislamiento contemporáneo en concordancia con la estructura genética detectada por marcadores del genoma nuclear. La herencia del genoma cloroplástico en *P. parlatorey* es por vía paterna, de manera que la dispersión del ADNcp se realiza a través del polen implicando que las distancias de dispersión sean mayores. *P. parlatorey* habita el estrato de la selva montana de las Yungas (desde 1200 hasta 3000 msnm) y es una especie tolerante al frío, lo cual le habría permitido sobrevivir a las eras glaciares sin movimientos latitudinales. Sin embargo, los períodos cálidos actuales, combinados con los disturbios antrópicos, han incrementado la fragmentación de estos bosques los cuales tienden a migrar a elevaciones superiores (Quiroga *et al.*, 2012). Este estudio permite reforzar la importancia de la vía de dispersión del genoma cloroplástico en los árboles sobre la distribución de la diversidad haplotípica en sus poblaciones, como así también, se podrían postular diferentes impactos de los cambios climáticos históricos en las especies de los SDTF entre las cuales se cuenta *A. colubrina* var. *cebil*.

Se detectó un patrón geográfico de distribución de la diversidad haplotípica en las poblaciones estudiadas donde las poblaciones de la región Yungas poseen haplotipos propios y diferentes a los de las poblaciones de la región Paranaense. Esta distribución haplotípica resultaría de la acción de la deriva genética debido a que este proceso microevolutivo actúa modificando las frecuencias haplotípicas de manera azarosa incrementando la varianza entre las poblaciones y disminuyendo la varianza dentro de ellas. De esta manera, las regiones cuentan con haplotipos propios que posiblemente representan diferentes acervos genéticos ancestrales.

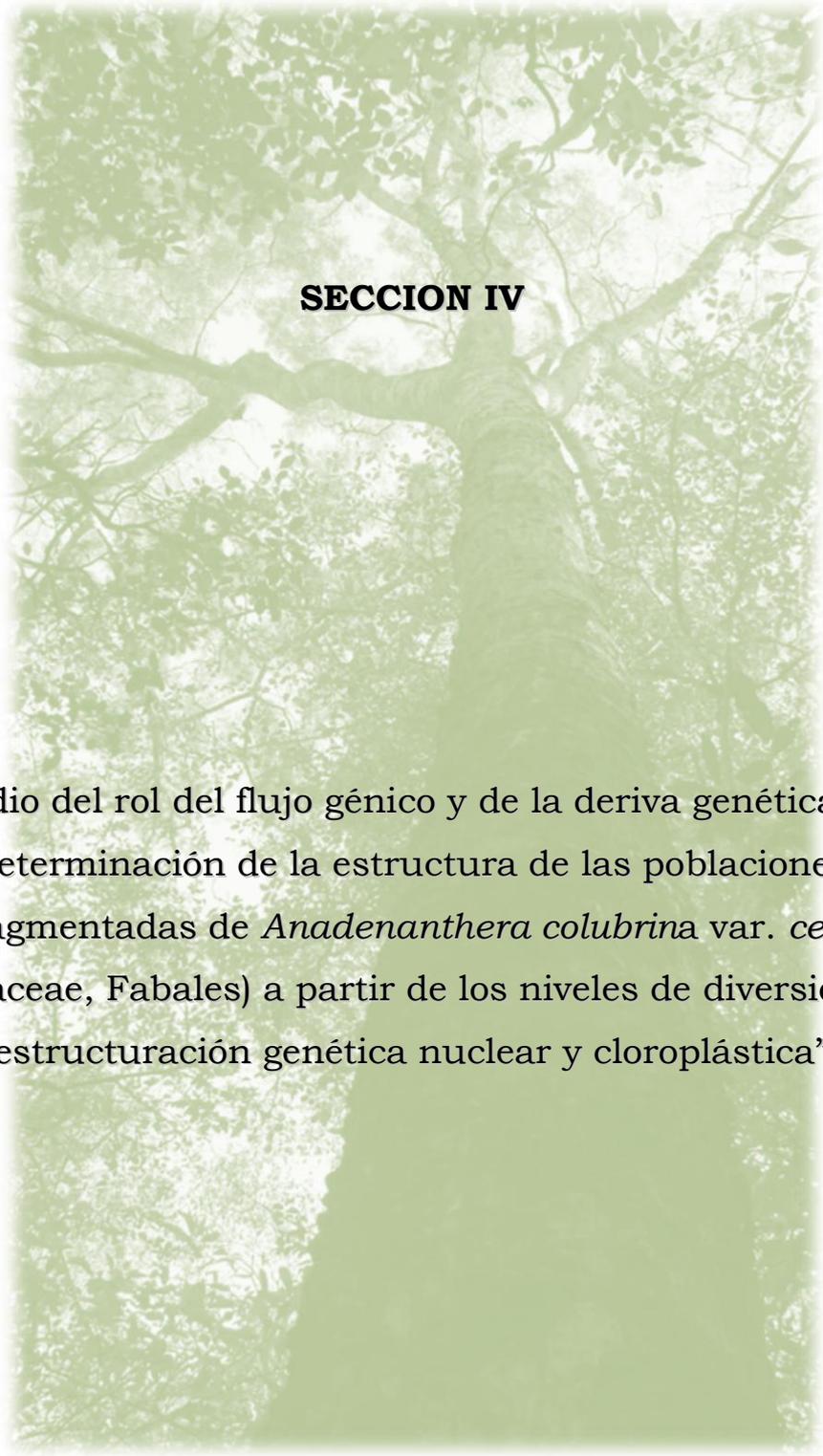
Los resultados denotan una elevada estructuración genética entre regiones junto con niveles reducidos de variación dentro de las poblaciones. Es poco probable que ocurra dispersión de semillas contemporánea entre poblaciones o entre regiones debido a la distancia geográfica que las separa. De esta manera, a nivel global los mayores niveles de variación de las frecuencias haplotípicas se encuentran entre las regiones Paranaense y de las Yungas los cuales serían consecuencia de la historia biogeográfica de estas poblaciones. Pennington *et al.* (2009) demostró que la distribución fragmentada de los SDTF ha persistido por un tiempo evolutivo suficientemente prolongado como para haber afectado a los patrones de distribución de las especies leñosas, a sus relaciones filogenéticas y a sus estructuras genéticas poblacionales. Este

estudio demostraría que la presencia de haplotipos cloroplásticos característicos y distantes genéticamente en los núcleos Misiones y Pedemonte Subandino sería consecuencia de una prolongada persistencia como entidades genéticamente aisladas y de la ausencia de flujo génico mediado por semillas.

A. colubrina var. *cebil* es un recurso nativo de relevancia dadas sus particularidades biológicas y utilitarias. Su distribución es peculiar, como resultado de los procesos biogeográficos producidos por los cambios climáticos históricos, los cuales han dejado su impronta sobre la distribución de la variabilidad genética.

6. Conclusiones

- Las poblaciones presentaron niveles reducidos de diversidad genética citoplasmática.
- Las poblaciones presentaron fuerte estructuración genética, encontrándose variantes haplotípicas fijadas en algunas poblaciones.
- Los individuos fueron asignados a diferentes *clusters* genéticos en concordancia con su constitución haplotípica definiendo dos grupos según su región de origen.
- A nivel histórico la deriva genética ha tenido mayor relevancia que el flujo génico.
- La distribución geográfica contemporánea de los haplotipos en ambas regiones sería consecuencia de eventos históricos. En la región de las Yungas la distribución geográfica sería resultante de eventos de colonización históricos y se corresponderían con una región inestable del Núcleo Pedemonte Subandino en tanto que, en la región Paranaense, los haplotipos se corresponderían con una constitución genética ancestral de una región estable del núcleo Misiones.



SECCION IV

“Estudio del rol del flujo génico y de la deriva genética en la determinación de la estructura de las poblaciones fragmentadas de *Anadenanthera colubrina* var. *cebil* (Fabaceae, Fabales) a partir de los niveles de diversidad y estructuración genética nuclear y cloroplástica”

1. Introducción

El flujo génico entre poblaciones vegetales puede ocurrir mediante dos vías. La primera involucra la dispersión del polen hacia una población diferente, la fertilización exitosa de un óvulo por ese polen y finalmente el establecimiento de la semilla resultante en ese sitio. Además, el flujo génico también puede ocurrir mediante la dispersión de la semilla y el establecimiento exitoso de esa semilla inmigrante en la nueva población (Ennos, 1994). El flujo génico por medio de semillas y del polen se encuentran entre los principales determinantes de la estructura genética poblacional en árboles de bosques tropicales y templados (Ndiade-Bourobou *et al.*, 2010). Por ello, una completa descripción del flujo génico debe incluir la evaluación de la importancia relativa del polen y de las semillas como agentes intermediarios del flujo génico (Ennos, 1994).

1.1 Marcadores moleculares nucleares y cloroplásticos para inferir los niveles de flujo génico mediado por polen y por semillas

Los análisis genéticos poblacionales requieren de marcadores polimórficos adecuados que permitan dilucidar detalles a fina escala de la estructura genética espacial y reconstruir los patrones de flujo génico (Pakkad *et al.*, 2008). La inferencia de los niveles de flujo génico mediados por el polen y por las semillas desde los patrones de estructuración genética requiere de marcadores moleculares que permitan diferenciar los efectos del movimiento de las semillas y del polen (Oddou-Muratorio *et al.*, 2001). Los análisis comparativos mediante el empleo de microsatélites nucleares y cloroplásticos se han convertido en un enfoque ampliamente utilizado para estos propósitos debido a que proveen información complementaria y a menudo contrastante sobre la estructura genética, diferenciación y flujo génico dentro y entre poblaciones (Pakkad *et al.*, 2008). El empleo combinado de marcadores del genoma cloroplástico junto con marcadores del genoma nuclear permite estudiar la acción de los procesos demográficos que actúan a diferentes escalas de tiempo, debido a que las organelas presentan modos de herencia diferentes al genoma nuclear, además de un tamaño efectivo poblacional menor y tasas de mutación diferentes (Tollefsrud *et al.*, 2009). Los genes de las organelas pueden presentar grados de diferenciación genética diferentes a los genes nucleares, simplemente debido a diferencias en el tamaño efectivo. Si el ADNcp haploide es heredado por vía materna el tamaño efectivo poblacional será considerablemente inferior al del genoma nuclear (McCauley, 1995). Sin embargo, la disparidad de los valores de F_{ST} entre ambos genomas es mayor a la esperada únicamente por las diferencias en cuanto al tamaño efectivo, destacando los efectos del flujo génico mediado por el polen en la reducción del valor de F_{ST} medido desde marcadores nucleares (McCauley, 1995). Cuando la tasa de flujo génico por medio del polen excede a la tasa de dispersión de las semillas y la herencia del ADNcp es estrictamente materna, los valores de F_{ST} de equilibrio para los polimorfismos del ADNcp

pueden ser drásticamente mayores a los valores de F_{ST} de equilibrio para los genes nucleares (McCauley, 1995). El conocimiento de la tasa mutacional en los *loci* SSR resulta de importancia porque determina los niveles de variabilidad dentro de las poblaciones e influye sobre las estimas de estructuración poblacional. La tasa de mutación de los *loci* SSRs es superior a cualquier otra región del genoma, sin embargo, el menor número de alelos informado para los *loci* cpSSRs comparado a los *loci* SSRs nucleares sugiere una menor tasa de mutación para los primeros (Provan *et al.*, 1999). Estas consideraciones constituyen el marco teórico de la estimación de la contribución relativa del polen y de las semillas al flujo génico. Ennos (1994) demostró que la proporción de flujo por polen/flujo por semillas es una función simple de los valores de F_{ST} para los genes nucleares y cloroplásticos. En angiospermas, la herencia biparental y materna puede resultar en valores de F_{ST} contrastantes cuando hay ciertos niveles de movimiento de polen entre las poblaciones (McCauley, 1995). Por lo tanto, es de esperar que los niveles de diferenciación poblacional varíen entre los marcadores de herencia uniparental y biparental para el mismo conjunto de poblaciones (Ennos, 1994). Poblaciones que han fijado un haplotipo en el genoma citoplasmático no necesariamente muestran baja diversidad en los marcadores nucleares (Tollefsrud *et al.*, 2009).

Existen características del polimorfismo cloroplástico y supuestos del modelo empleado en estas estimaciones que deben ser consideradas cuidadosamente al momento de interpretar los resultados. Entre ellas, caben destacar el supuesto de neutralidad selectiva, el supuesto de herencia uniparental y la posible acción de procesos demográficos que pueden producir efectos que no son considerados en el modelo de equilibrio flujo génico-deriva genética (McCauley, 1995). Sin embargo, los polimorfismos a nivel de ADN representan generalmente la variación en regiones no codificantes y la herencia materna de las organelas en las angiospermas es en general una regla, con lo cual sus efectos serían nulos. Por su parte, el supuesto de equilibrio demográfico, a pesar de que generalmente simplifica las relaciones entre las medidas de flujo génico y estructura genética, es poco probable en la naturaleza. Las consecuencias genéticas de los eventos de colonización pueden ser diferenciales para los genomas nuclear y cloroplástico, pudiendo una nueva población ser monomórfica para el genoma cloroplástico y presentar elevada diversidad para el genoma nuclear (McCauley, 1995).

La combinación de marcadores de herencia materna y biparental, además, facilita el estudio de procesos evolutivos históricos y de procesos recientes (Muller *et al.*, 2009) proveyendo una visión complementaria acerca de los patrones de flujo génico (Andrianoelina *et al.*, 2009).

1.2 Importancia relativa del flujo génico mediado por polen y por semillas

1.2.1 Modelo de islas

Las diferencias en cuanto a los niveles de estructuración genética detectada con los marcadores de herencia uniparental y biparental es función de los niveles relativos del flujo génico mediado por el polen y por las semillas (Ennos, 1994).

En el modelo empleado para realizar este análisis se asumen especies diploides y hermafroditas. Además, considera que la especie se encuentra distribuida en un número infinito de islas, cada una de tamaño N , relacionadas por un flujo constante de genes y apareamiento aleatorio (Ennos, 1994).

El grado de diferenciación genética bajo el modelo de equilibrio migración-deriva es función del tamaño efectivo y de la tasa de migración. Cabe esperar que los valores de F_{ST} obtenidos desde marcadores con modos de herencia contrastantes no serán independientes resultando de importancia establecer la relación entre estos índices (Ennos, 1994). Debido a que los marcadores heredados por vía materna sólo migran a través de las semillas, en equilibrio migración-deriva genética se espera que la magnitud del índice de fijación para el genoma de herencia materna (F_{STm}) exceda la magnitud del índice de fijación estimado para el genoma de herencia biparental (F_{STb}) (Ennos, 1994).

Cuando se tiene en cuenta la historia poblacional, los contrastes entre la estructuración genética cloroplástica y la estructuración genética nuclear pueden ser informativos en cuanto a los procesos de colonización, como así también en cuanto al rol del flujo génico mediado por el polen en relación al flujo génico mediado por las semillas en las mediciones tradicionales de flujo génico (McCauley, 1995).

2. Objetivos específicos

- Analizar los niveles de diversidad y de estructuración genética nuclear y cloroplástica en las poblaciones fragmentadas de *A. colubrina* var. *cebil*.
- Determinar la importancia relativa de los niveles de flujo génico a través del polen y de las semillas en estas poblaciones.

3. Materiales y Métodos

3.1 Niveles de diversidad y estructuración genética nuclear y cloroplástica

Se analizó, de manera comparativa, a los niveles de diversidad y estructuración genética nuclear y cloroplástica estimados desde los marcadores nuSSRs y cpSSRs, respectivamente.

3.2 Importancia relativa del flujo génico mediado por el polen y por las semillas

3.2.1 Modelo de Islas

Cuando los niveles de flujo génico mediado por semillas entre poblaciones es menor al flujo génico mediado por polen, es posible estimar la tasa relativa de flujo génico mediado por polen y por semillas entre poblaciones a partir de los niveles de estructuración genética estimados desde marcadores de herencia biparental (nuSSRs) y desde marcadores de herencia materna (cpSSRs), mediante la siguiente ecuación:

$$\frac{\text{Flujo por polen}}{\text{Flujo por semillas}} = \frac{\left[\left(\frac{1}{F_{STb}} - 1 \right) - 2 \times \left(\frac{1}{F_{STm}} - 1 \right) \right]}{\left(\frac{1}{F_{STm}} - 1 \right)}$$

donde F_{STb} y F_{STm} corresponden al índice de fijación estimado desde marcadores nucleares y cloroplásticos, respectivamente (Ennos, 1994).

4. Resultados

4.1 Comparación de los niveles de diversidad y estructuración genética nuclear y cloroplástica

En cuanto a diversidad genética, el genoma nuclear presentó mayores niveles de diversidad ($H_E=0,791$) en relación a los niveles de diversidad presentes en el genoma cloroplástico ($h=0,047$ y $H_{Ecp}=0,249$) (Tabla IV-1).

Las poblaciones mostraron patrones opuestos de diversidad genética según el marcador considerado. Así, la población Jujuy reveló la mayor diversidad genética en cuanto al genoma nuclear ($H_E=0,841$), mientras que la población Santa Ana resultó la más diversa en cuanto al genoma cloroplástico ($h=0,117$ y $H_{Ecp}=0,500$) (Tabla IV-1).

Los resultados en cuanto a la estructuración genética mostraron un patrón similar. Los cpSSRs revelaron elevada estructuración genética y niveles limitados de flujo génico mediado por semillas mientras que los nuSSRs revelaron estructuración genética moderada y niveles de flujo génico total que superaron la unidad (Tabla IV-2).

El análisis bayesiano a partir de los genotipos multilocus del genoma nuclear asignó a los individuos a dos *clusters*, identificando un patrón de estructuración regional. Por su parte, el análisis bayesiano a partir de los haplotipos cloroplásticos asignó a los individuos a seis *clusters*, identificando un patrón de estructuración poblacional. Estos resultados denotan presencia de mayor estructuración genética cloroplástica en estas poblaciones (Tabla IV-2).

4.2 Importancia relativa del flujo génico mediado por polen y por semillas

4.2.1 Modelo de Islas

Los índices de fijación F_{ST} globales obtenidos desde los marcadores de herencia materna y biparental arrojaron valores $F_{ST(b)} = 0,111$ y $F_{ST(m)} = 0,948$, respectivamente. Así, la tasa relativa de flujo génico mediado por polen superó 144 veces a la tasa de flujo génico mediado por semillas denotando que el flujo génico mediado por polen es mayor al flujo génico mediado por semillas.

Tabla IV-1: Comparación de los niveles de diversidad genética nuclear y cloroplástica

Región	Población	N	nuSSRs				cpSSRs			
			N _{Ab}	N _{Eb}	H _O	H _{Eb}	N _{Am}	N _{Em}	h _m	H _{Em}
Paranaense	Candelaria	20	9,875	5,166	0,698	0,780	1,250	1,055	0,045	0,400
	Santa Ana	16	9,250	4,479	0,703	0,756	1,250	1,221	0,117	0,500
Yungas	Tucumán	14	8,625	5,548	0,748	0,785	1,000	1,000	0,000	0,000
	Jujuy	19	11,125	7,692	0,752	0,841	1,250	1,028	0,025	0,094
	Total*	69	9,719*	5,721*	0,725*	0,791*	2,750	1,076*	0,047*	0,249*

*valor promedio por población, en negrita se resaltan los mayores valores por población.

N=tamaño de la población, N_A=número promedio de alelos por locus, N_E=número efectivo de alelos por locus, H_O=heterocigosidad observada, H_{Eb}=heterocigosidad esperada, N_{Am}=número promedio de alelos por locus cpSSR, N_{Em}=número efectivo de alelos por locus cpSSR, h_m=índice de diversidad génica de Nei, y H_{Em}=índice de diversidad haplotípica de Nei.

Tabla IV-2: Distribución de la diversidad genética nuclear y cloroplástica en poblaciones naturales de *A. colubrina* var. *cebil*.

Métodos	Análisis	nuSSRs	cpSSRs
No espaciales	Distribución de la variación dentro y entre poblaciones	AMOVA	Mayor variación dentro de poblaciones (88,94%)
	Índice de fijación	Global	$F_{STb} = 0,111^*$
		Regional	Paranaense $F_{STb} = 0,075^*$ Yungas $F_{STb} = 0,039^*$
	Flujo génico histórico	$N_{e,m}$	$N_{e,m_b} = 2,010$
	Método basado en distancias genéticas para el establecimiento de grupos genéticos	Arbol de genotipos multilocus (nuSSRs) y red de haplotipos (cpSSRs)	Tres grupos integrados principalmente por individuos de: (1) Candelaria, (2) Yungas y (3) Santa Ana.
	Método basado en modelos para establecer <i>clusters</i> genéticamente homogéneos	Inferencia Bayesiana sin información geográfica a priori	Análisis <i>Admixture</i> 2 <i>clusters</i> genéticos, (estructuración mayor a nivel de regiones)
Inferencia Bayesiana con información geográfica a priori		Análisis <i>Adixture</i> 4 <i>clusters</i> genéticos (coinciden con los orígenes poblacionales)	
Espaciales			Análisis <i>Mixture</i> 6 <i>clusters</i> genéticos (subestructuración poblacional)
			Análisis <i>Mixture</i> 4 <i>clusters</i> genéticos (coinciden e con los orígenes poblacionales)

*valor estadísticamente significativo para el 95% de confianza.

5. Discusión

En las plantas con semillas las relaciones entre flujo génico y estructura genética no son simples debido a que los genes poseen dos vías potenciales de intercambio durante dos estadios de vida diferentes, las semillas y el polen. El entendimiento de las contribuciones relativas del polen y de las semillas al flujo génico total ha sido el principal desafío de los genetistas que estudian poblaciones vegetales (McCauley, 1997). Los estudios poblacionales a nivel intraespecífico, basados en datos moleculares obtenidos desde diferentes genomas, proveen una oportunidad para comparar el movimiento del material genético transportado por el polen en relación al movimiento del material genético transportado por las semillas (Grauke *et al.*, 2011).

Estudios teóricos y empíricos han mostrado que, generalmente, el ADN de las organelas exhibe niveles más elevados de estructuración genética en relación al ADN nuclear (Grauke *et al.*, 2011). Los resultados obtenidos en las poblaciones naturales de *A. colubrina* var. *cebil* concuerdan con los resultados obtenidos en estudios previos basados en marcadores del ADNcp en especies forestales, debido a que en la mayoría de ellos se detectó una marcada estructuración genética entre poblaciones y reducidos niveles de diversidad genética dentro de las poblaciones en comparación con los marcadores nucleares (Oddou-Muratorio *et al.*, 2001). La estructuración genética del genoma nuclear generalmente es menos significativa debido a que el efecto de la deriva genética es dos veces menor (genoma diploide) y el flujo génico se ve incrementado por medio de la dispersión del polen (Ndiade-Bourobou *et al.*, 2010).

En el análisis de varianza molecular, los cpSSRs revelaron mayor variación entre regiones mientras que los SSRs nucleares revelaron mayor variación dentro de las poblaciones. Las diferencias en relación a las tasas de mutación y al tamaño efectivo de los genomas nuclear y cloroplástico explicarían este resultado. Los niveles de flujo génico también podrían explicarlo ya que en las plantas los niveles de flujo génico en *loci* heredados de manera biparental es mayor a los niveles de flujo génico en *loci* de herencia materna debido a que los *loci* biparentalmente heredados se transmiten a través del polen y de las semillas y los *loci* maternos se transmiten únicamente a través de las semillas (McCauley, 1997).

En este estudio, los niveles de estructuración genética son diferentes según el genoma considerado. Si se considera que el efecto del menor tamaño efectivo del genoma cloroplástico incrementa la estructuración genética en dos veces respecto a la estructuración genética del genoma nuclear, la diferencia entre los índices de fijación F_{ST} entre ambos marcadores ($F_{STb}=0,11$ vs $F_{STm}=0,95$, respectivamente) resulta en una estructuración cloroplástica 8 veces mayor a la estructuración genética nuclear, con lo cual las diferencias en cuanto a los mecanismos de dispersión, serían causantes de las diferencias entre los niveles de flujo génico por medio de las semillas y del polen.

En cuanto a los niveles de diversidad genética nuclear presentes en las poblaciones de *A. colubrina* var. *cebil* ($H_{Enu}=0,756-0,841$) fueron similares a los niveles detectados en

Dalbergia monticola cuya polinización es entomófila ($H_{Enu}=0,64-0,81$, Andrianoelina *et al.*, 2009). Sin embargo, la diversidad genética presente en las poblaciones naturales de *A. colubrina* var. *cebil* superó a los niveles de diversidad estimados en otras especies tropicales con polinización entomófila, como por ejemplo en las especies *Baillonella toxisperma* Pierre ($H_{Enu}=0,56-0,58$; Ndiade-Bourobou *et al.*, 2010), *Aucoumea klaineana* ($H_{Enu}=0,38-0,55$; Born *et al.*, 2008), *Vouacapoua americana* ($H_{Enu}=0,47-0,51$; Dutech *et al.*, 2002); *Distemonanthus benthamianus* ($H_{Enu}=0,470-0,567$; Debout *et al.*, 2011); *Pterocarpus officinalis* Jacq ($H_{Enu}=0,24 - 0,59$, Muller *et al.*, 2009). En general, los estudios realizados con isoenzimas revelaron que los árboles mantienen niveles superiores de diversidad genética dentro de sus poblaciones y e inferiores entre poblaciones, siendo que los valores reducidos de estructuración genética de la mayoría de los árboles son consistentes con la presencia de mayores niveles de flujo génico y flujo génico a larga distancia (Hamrick, 2004).

En cuanto a los niveles de diversidad genética cloroplástica las poblaciones de *A. colubrina* var. *cebil* estudiadas presentaron un total de 6 haplotipos ($H_{cp}=0,000-0,500$). Estos resultados representan menores niveles de diversidad genética cloroplástica en *A. colubrina* var. *cebil* en comparación con *Pterocarpus officinalis* Jacq donde se identificaron 10 haplotipos siendo la dispersión de las semillas efectuada por medio de la gravedad y por flotación ($H_{cp}=0,000-0,680$) (Muller *et al.*, 2009). En *Dalbergia monticola* se detectaron 18 haplotipos siendo la dispersión de sus semillas por medio de la gravedad y por animales ($H_{cp}=0,00-0,80$) (Andrianoelina *et al.*, 2009). En *Baillonella toxisperma* Pierre se detectaron 13 haplotipos y la dispersión de sus semillas se realiza por medio de animales ($H_{ecp}=0,702-0,836$) (Ndiade-Bourobou *et al.*, 2010).

Considerando estos resultados en conjunto podría decirse que *A. colubrina* var *cebil* presentó un nivel de diversidad genética nuclear mayor a otras especies tropicales aunque presentó un nivel menor de diversidad genética cloroplástica lo cual podría deberse, principalmente, a su dispersión de semillas restringida ya que la dispersión de las mismas en *A. colubrina* var. *cebil* es de tipo autocoria-anemocoria.

La estructuración genética del genoma cloroplástico es elevada mientras que la estructuración genética nuclear es moderada, lo cual se corresponde con un nivel de flujo génico mediado por semillas inferior al nivel de flujo génico mediado por polen y semillas ($N_{ev}m_b=2,010$ vs $N_{ev}m_m=0,027$, respectivamente). Además, sabiendo que el punto de inflexión en este estimador corresponde a $N_e m=1$, puede inferirse que, a nivel del genoma nuclear, el flujo génico es suficiente para contrarrestar los efectos de la deriva genética mientras que, a nivel del genoma cloroplástico, el flujo génico no contrarresta los efectos de la deriva genética.

Las poblaciones fragmentadas que presentan mayores niveles de estructuración genética en marcadores heredados por vía materna y menores niveles en marcadores de herencia

biparental probablemente presentan flujo génico entre los fragmentos en tiempos posteriores a la fragmentación (Hamrick, 2004).

Las poblaciones de árboles forestales se ven expuestas a numerosos cambios ambientales que pueden moldear su constitución genética. Estos cambios pueden ser antiguos y haber ocurrido hace 100.000 años, como las glaciaciones, o pueden ser recientes y haber ocurrido en la última década dentro de la vida de un árbol adulto (Hamrick, 2004).

Cuestiones relacionadas a la historia del ecosistema de los Bosques tropicales estacionalmente secos fueron analizadas en dos especies forestales propias de estos bosques: *Astronium urundeuva* (Anacardiaceae) y *Geoffroea spinosa* (Leguminosae) (Naciri *et al.*, 2006). Mediante el empleo de SSR nucleares en ambas especies y secuencias de regiones no codificantes del ADNcp de *A. urundeuva* estudiaron poblaciones del núcleo Misiones y del núcleo Pedemonte Subandino. En cuanto a los niveles de estructuración genética nuclear presente entre estas regiones el índice de fijación F_{ST} fue de 0,064 en *A. urundeuva* y de 0,151 en *G. spinosa*. Las diferencias en los niveles de estructuración genética entre estas especies fueron atribuidas tanto a las diferencias en cuanto a las capacidades de dispersión de estas especies como a las consecuencias de los eventos históricos (Naciri *et al.*, 2006). Debido a que *A. urundeuva* dispersa sus semillas por acción del viento y *G. spinosa* por medio de pequeños vertebrados, se esperan mayores niveles de dispersión en *A. urundeuva* mientras que, considerando los eventos históricos, la mayor separación entre las poblaciones de *G. spinosa* junto a los mayores niveles de estructuración genética indicarían eventos de fragmentación más antiguos en esta especie. En relación a la estructuración genética en *A. urundeuva*, los niveles fueron superiores en el genoma cloroplástico ($F_{STm}=0,83 > F_{STb}=0,064$) resultados que fueron adjudicados a una mayor dispersión del polen en relación a las semillas. En *A. urundeuva* se identificaron seis haplotipos entre los cuales el más frecuente, considerado más antiguo, estuvo localizado en el núcleo Misiones (Naciri *et al.*, 2006). Debido a que los niveles de estructuración genética nuclear en *A. colubrina* var. *cebil* fueron similares a los niveles de estructuración genética en las poblaciones del núcleo Misiones y Pedemonte Subandino de *G. spinosa* es de esperar que estas especies compartan una misma historia y/o similares relaciones en cuanto a la importancia del flujo génico por medio del polen en relación al flujo génico por medio de las semillas. Por su parte, *A. colubrina* (Vell.) Brenan fue incluida entre las especies candidatas para el estudio de los Bosques tropicales estacionalmente secos debido a su amplia distribución en los mismos, su definición taxonómica clara, la facilidad de ser reconocida en el campo, su abundancia relativa, su especificidad de hábitat y su baja probabilidad de dispersión por el hombre (Naciri *et al.*, 2006).

La fragmentación de una población continua de árboles puede, potencialmente, interrumpir diferentes procesos biológicos y evolutivos pudiendo modificar, además, su propia composición genética (Hamrick, 2004). Los efectos de la fragmentación se pueden dividir en

efectos a corto y a largo plazo. Los efectos a corto plazo ocurren durante la fragmentación o luego de una o dos generaciones y dependen de los niveles de estructuración genética preexistentes. Por su parte, los efectos a largo plazo ocurren luego del transcurso de varias generaciones posteriores a la misma y pueden modificar la composición genética de los fragmentos individuales de manera considerable. Si el flujo génico entre los fragmentos es limitado, la diversidad genética se verá estructurada por acción de la deriva genética, la diversidad genética dentro de los fragmentos decrecerá, mientras que la estructuración genética entre los fragmentos se incrementará. Los niveles de flujo génico entre las poblaciones son determinantes para la predicción de los efectos de la fragmentación (Hamrinck, 2004).

De acuerdo con lo anteriormente expuesto, los efectos a largo plazo de la fragmentación en las poblaciones de *A. colubrina* var *cebil* pueden ser identificados a nivel global en ambos genomas. La conformación de dos *clusters* genéticos en el análisis bayesiano, coincidentes con las regiones geográficas estudiadas, podría ser resultado de fragmentación antigua, la cual produjo profundos cambios en la composición genética nuclear. A nivel del genoma cloroplástico, también se identificaron diferencias haplotípicas entre las regiones, ya que no se han detectado haplotipos compartidos entre las mismas. Sin embargo, debido al menor tamaño efectivo de este genoma y a los menores niveles de flujo génico por medio de las semillas, los efectos de la deriva genética han sido pronunciados llevando a la fijación de variantes haplotípicas en dos de las cuatro poblaciones estudiadas. La dispersión de semillas en *A. colubrina* var. *cebil* ocurre por gravedad o por el movimiento de las vainas por acción del viento, lo cual indica que el flujo de semillas es altamente restringido y que, posiblemente, esta especie presentara estructuración genética cloroplástica previa a la fragmentación.

Por su parte, a nivel del genoma nuclear, los niveles de flujo génico histórico intrarregional son mayores. La región de las Yungas presentó niveles reducidos de estructuración genética lo cual implica flujo génico efectivo mediado por polen. Ya que el mismo, según Caetano *et al.* (2008) tiene un efecto homogeneizador que diluye la estructuración genética resultante del flujo por medio de las semillas. Por su parte, la región Paranaense presentó estructuración genética moderada, indicando ausencia de flujo génico mediado por polen y mayor aislamiento genético entre sus poblaciones.

Ennos (1994) demostró que es posible estimar las tasas relativas de los niveles de flujo génico por medio del polen y por medio de las semillas entre poblaciones vegetales a partir de la comparación de los valores de F_{ST} estimados empleando marcadores de herencia biparental y marcadores heredados por vía materna. En *A. colubrina* var. *cebil* el resultado obtenido para esta proporción, indica que el flujo génico mediado por el polen supera 144 veces al flujo génico mediado por las semillas. La consideración del valor estimado para esta proporción de manera aislada carece de sentido biológico, resultando aparente la importancia de esta

estimación al considerarse de manera comparativa, como por ejemplo, entre especies con diferentes historias de vida (McCauley, 1997).

En los estudios realizados en siete especies forestales, la tasa de flujo génico mediado por el polen fue mayor a la tasa de flujo génico mediado por las semillas, aunque existieron diferencias sustanciales en las proporciones de ambos niveles entre las especies (Ennos, 1994). *Quercus petraea* presentó el mayor contraste, superando el nivel de flujo génico mediado por polen aproximadamente en 200 veces al nivel de flujo génico mediado por las semillas. Este contraste puede deberse al sistema reproductivo de la especie, la cual se caracteriza por ser alógama con polinización anemófila y flujo restringido de semillas a través de aves y roedores (Ennos, 1994). Por su parte, Oddou-Muratorio *et al.* (2001) estudiando las poblaciones naturales de *Sorbus torminalis* obtuvieron una proporción de aproximadamente 2, la cual constituye una de las menores proporciones observadas en especies leñosas. Esta especie se caracteriza por presentar polinización entomófila, dispersión de semillas de tipo endozoocora y una dinámica poblacional típica de las metapoblaciones con eventos de extinción y recolonización. Esta combinación de caracteres de historia de vida, resulta en reducidos niveles de diferenciación genética entre poblaciones, especialmente en el genoma cloroplástico, debido a múltiples eventos de colonización (Oddou-Muratorio *et al.*, 2001). *A. colubrina* var. *cebil* presentó una proporción igual a 144, la cual supera ampliamente a la proporción observada en *S. torminalis*. A diferencia de esta última especie, *A. colubrina* var. *cebil* es una especie alógama, su polinización es entomófila y el mecanismo de dispersión de la semilla es de tipo autocoria-anemocoria, características que agudizan las diferencias en cuanto a la capacidad de dispersión del polen y de las semillas.

Las interpretaciones de la proporción de flujo génico por medio de polen y por medio de semillas en *A. colubrina* var. *cebil* se basa en el modelo de islas. Es probable que esta proporción esté sobreestimada y no sea un fiel reflejo de la proporción real. El modelo de islas posee supuestos con bajas probabilidades de cumplirse en las poblaciones naturales, como ser tamaños poblacionales constantes e iguales y tasas de migración idénticas entre poblaciones locales (Ennos, 1994). Además, este modelo asume equilibrio demográfico el cual simplifica las relaciones entre las medidas de flujo génico y las de estructuración genética. Estrictamente hablando, este equilibrio es poco probable en la naturaleza ya que las poblaciones fluctúan en tamaño y su distribución puede variar a lo largo del tiempo (McCauley, 1995). Las consecuencias genéticas de los eventos de colonización pueden ser diferentes para el genoma nuclear y para el genoma cloroplástico. En el caso de una población colonizada por un fruto con múltiples semillas, asumiendo herencia materna, la nueva población será uniforme para el genoma cloroplástico pero puede expresar considerable variación nuclear. Más aún, luego de las perturbaciones, el genoma nuclear y el cloroplástico alcanzarán el equilibrio a diferentes tasas. Por todo esto, las diferencias entre los F_{ST} nuclear y cloroplástico que son explicadas asumiendo

el equilibrio por las diferencias en cuanto a los niveles de flujo génico pueden ser confundidas por las influencias diferenciales de los eventos de colonización (McCauley *et al.*, 1995). De esta manera, cuando se tiene en cuenta la historia poblacional los contrastes cloroplásticos y nucleares sobre las medidas tradicionales de flujo génico pueden ser informativos en relación a los procesos de colonización como así también acerca de los roles relativos del movimiento del polen y de las semillas (McCauley *et al.*, 1995).

El estudio de los genomas heredados por vía materna y biparental reflejó diferentes patrones de distribución de la diversidad genética en las poblaciones fragmentadas de *A. colubrina* var. *cebil* estudiadas. Se detectó estructuración entre regiones como consecuencia de la fragmentación histórica en ambos genomas y mayor estructuración genética a nivel del genoma cloroplástico conjuntamente con menores niveles de flujo génico por medio de las semillas. Debe destacarse, la importancia intrarregional a nivel histórico del rol del flujo génico mediado por el polen, el cual mantuvo niveles elevados de diversidad genética dentro de las poblaciones. Sin embargo, las diferencias en cuanto a la diversidad y estructuración genética a nivel del genoma nuclear detectados en la región Paranaense y en la región de las Yungas no pudieron ser explicadas a partir de las distancias geográficas entre las poblaciones.

Caetano *et al.* (2008) estudiaron poblaciones de *Astronium urundeuva* empleando nueve SSRs nucleares y dos espaciadores cloroplásticos, y explicaron los patrones definidos en una de las regiones de los SDTF estudiadas concluyendo que en el genoma cloroplástico aún persisten las huellas de la colonización. Sin embargo, en el genoma nuclear, estas huellas se han ido borrando gradualmente debido al intercambio genético entre las poblaciones, el cual se ve favorecido, probablemente, por una distribución continua del SDTF en esa área. Una interpretación similar podría ser realizada en las poblaciones de *A. colubrina* var. *cebil* de la región de las Yungas ya que según Wernek *et al.* (2011) esta área presentaría mayor inestabilidad y tendría una colonización más reciente. Además, los resultados obtenidos en el presente trabajo reflejaron niveles contrastantes de estructuración y diversidad genética entre ambos genomas, presentando baja diversidad y elevada estructuración genética cloroplástica, y elevada diversidad y baja estructuración genética nuclear. El genoma cloroplástico aún refleja el pasado histórico en el cual los eventos de colonización en la región de las Yungas tienen alta probabilidad de haber ocurrido. Mientras que en el genoma nuclear, debido al flujo génico a través del polen se habrían homogeneizado las frecuencias alélicas y reducido los niveles de estructuración genética en esta región.

Por su parte, la región Paranaense presentó elevada estructuración genética cloroplástica y mayor estructura genética nuclear. Según Wernek *et al.* (2011), esta área presentaría mayor estabilidad con lo cual, se podría pensar que entre sus poblaciones existen barreras recientes tanto para el flujo por semillas como para el flujo por polen. Estas barreras

podieron haberse originado en el impacto antrópico reciente el cual ha producido una reducción del tamaño efectivo de estas poblaciones y ha aumentado el aislamiento entre ellas.

Los árboles constituyen modelos de estudio particulares debido a que son organismos sésiles y longevos. Factores tales como los caracteres de historia de vida, los mecanismos de dispersión del polen y de las semillas, la dinámica poblacional, las características del paisaje actual y los eventos biogeográficos históricos, influyen y determinan los patrones actuales de distribución de la variación genética en sus poblaciones.

El polen y las semillas son los responsables del movimiento de los alelos en las poblaciones. El polen es haploide en tanto que las semillas son diploides lo cual determina, en caso de que las tasas de flujo génico sean las mismas, que el polen contribuya con un tercio y las semillas con dos tercios al flujo génico nuclear total. En la mayoría de las especies forestales, la tasa de flujo génico por polen supera a la tasa de flujo génico mediado por semillas. Estudiar los niveles de flujo génico mediante marcadores de herencia materna permite conocer el movimiento de las semillas, las cuales contribuyen en mayor medida al flujo génico nuclear total. Estas diferencias en cuanto al contenido genético y a las tasas de flujo génico contribuyen, de manera predecible, a la magnitud de la diferenciación genética poblacional.

La diversidad genética es esencial para asegurar el proceso evolutivo de adaptación de los recursos forestales a las condiciones ambientales cambiantes y a los requerimientos sociales (Young *et al.*, 1996; Inza *et al.*, 2012). Afortunadamente, el estudio de la variación genética interpoblacional está siendo valorado como una herramienta valiosa para guiar los esfuerzos de conservación y para el diseño de áreas prioritarias de conservación de especies (Quiroga y Prémoli, 2007, Inza *et al.*, 2012). La elevada diversidad genética nuclear presente en las poblaciones estudiadas destaca la importancia de estas poblaciones como reservorios de diversidad genética. Por su parte, la presencia de variantes de ADNcp propias en la región Paranaense, en áreas que no se encuentran protegidas y que corresponden a áreas de estabilidad histórica, merecen una especial atención.

6. Conclusiones

- Las poblaciones fragmentadas de *A. colubrina* var. *cebil* presentaron mayores niveles de diversidad genética nuclear en relación a la diversidad genética cloroplástica, de acuerdo a las diferencias en cuanto al balance entre los roles del flujo génico y de la deriva genética en ambos genomas.
- Debido a las diferencias en el tamaño efectivo y en los niveles de flujo génico en ambos genomas la estructuración genética de las poblaciones fragmentadas de *A. colubrina* var. *cebil* fue menor en el genoma nuclear en relación al genoma cloroplástico.
- Los efectos de la fragmentación histórica sobre los Bosques tropicales estacionalmente secos se reflejarían en la estructuración genética detectada a nivel global entre las regiones en ambos genomas.
- El impacto antrópico en la región Paranaense sería responsable de la mayor estructuración genética de sus poblaciones a nivel del genoma nuclear. Por su parte, los niveles superiores de flujo génico entre las poblaciones de las Yungas habrían mantenido mayor diversidad genética y menor estructuración genética de sus poblaciones a nivel del genoma nuclear.
- La presencia de haplotipos ancestrales y diversidad genética antigua en la región Paranaense explicarían la mayor diversidad y la menor estructuración genética de sus poblaciones a nivel del genoma cloroplástico. La presencia de diferentes variantes haplotípicas fijadas en cada población de la región de las Yungas, debido a la ausencia de flujo génico mediado por semillas, indicaría procesos de colonización antiguos y la acción de la deriva genética a nivel del genoma cloroplástico.
- La hipótesis sobre la cual se sustentó el presente trabajo no se rechaza al considerarse el rol del flujo génico y de la deriva genética a nivel del genoma nuclear mientras que, a nivel del genoma cloroplástico, la deriva genética ha jugado el rol principal.

Proyecciones

Este trabajo constituye la primera aproximación al conocimiento del estado actual de la diversidad genética en poblaciones argentinas de *A. colubrina* var. *cebil*. Como en todo estudio genético poblacional el conocimiento de la diversidad genética actual de las poblaciones permite hacer inferencias acerca de los procesos y eventos que operaron sobre ellas en el pasado para desde allí poder predecir los futuros niveles de diversidad genética en estas poblaciones. Es así, que se plantean nuevos abordajes al estudio de poblaciones en esta especie. Por un lado, podría aumentarse el conocimiento del pasado histórico mediante estudios filogeográficos y por otro, debería ahondarse sobre las consecuencias del impacto antrópico y fragmentación reciente sobre las poblaciones, en especial en la región Paranaense.



- Abraham de Noir, F., S. Bravo y R. Abdala (2002) Mecanismos de dispersión de algunas especies leñosas nativas del chaco occidental y serrano. *Quebracho Revista de Ciencias Forestales* 9:140-150.
- Agarwal, M., N. Shrivastava y H. Padh (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27:617-631.
- Aguilar, R., L. Ashworth, L. Galetto y M.A. Aizen (2006) Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. *Ecology Letters* 9: 968–980.
- Akkaya M. S., A. A. Bhagwat y P. B. Cregan (1992) Length Polymorphisms of Simple Sequence Repeat DNA in Soybean. *Genetics* 132(4):1131-1139.
- Aldrich, P.R. y J.L. Hamrick (1998) Reproductive Dominance of Pasture Trees in a Fragmented Tropical Forest Mosaic. *Science* 281:103-105.
- Altschul S. von R. (1964) A taxonomic study of the genus *Anadenanthera*. *Contributions of the Gray Herbarium of Harvard University* 193:3-65.
- Andrianoelina, O., B. Favreau, L. Ramamonjisoa y J.M. Bouvet (2009) Small effect of fragmentation on the genetic diversity of *Dalbergia monticola*, an endangered tree species of the eastern forest of Madagascar, detected by chloroplast and nuclear microsatellites. *Annals of Botany* 104(6):1231-1242.
- Avise, J.C. (2004) *Molecular Markers, Natural history and Evolution*. Segunda Edición. Sinauer Associates. Sunderland, Massachusetts 684 pág.
- Ayad, W.G., T. Hodgkin, A. Jaradat y V.R. Rao editors (1995) *Molecular genetics techniques for plant genetic resources*. Report of an IPGRI workshop 137 pág.
- Bandelt H.J., P. Forster y A. Rohl (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution* 16(1):37-48.
- Barton, N.H. y M. Slatkin (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56:409-415.
- Berli, P. y M. Palczewski (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185(1): 313-26.
- Born, C., F. Kjellberg, M. Chevallier, H. Vignes, J. Dikangadissi, J. Sanguié, E. J. Wickings y M. Hossaert-Mckey (2008) Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree *Aucoumea klaineana*. *Proceedings of the Royal Society* 275:2171-2179.
- Bossart, D.L. y D.P. Prowell (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution* 13:202-206.
- Broquet, T. y E.J. Petit (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution and Systematics* 40:193-216.

- Brown A.D., A. Grau, T. Lomáscolo y N.I. Gasparri (2002) Una estrategia de conservación para las selvas subtropicales de montaña (yungas) de Argentina. *Ecotropicos* 15(2):147-159.
- Brown A.D., H.R. Grau, L.R. Malizia y A. Grau (2001) Argentina editado por Kappelle M. y A.D. Brown. Bosques nublados del Neotrópico. Instituto Nacional de Biodiversidad, Santo Domingo de Heredia, Costa Rica. 704 pág.
- Brown, A.D. (2009) Las selvas pedemontanas de las Yungas: manejo sustentable y conservación de la biodiversidad de un ecosistema prioritario del noroeste Argentino. Editado por Brown A.D., P.G. Blendinger, T. Lomáscolo y P. García Bes. Selva pedemontana de las Yungas: Historia natural ecología y manejo de un ecosistema en peligro. Ediciones del subtrópico, Tucumán 487 pág.
- Burczyk, J., S.P. DiFazio y W.T. Adams (2004) Gene flow in forest trees: How far do genes really travel? *Forest Genetics* 11(2-3): 1-14.
- Cabrera, A.L. (1971) Fitogeografía de la Republica Argentina. *Boletín Sociedad Argentina de Botánica* 14(1-2): 1-42.
- Cabrera, A.L. y A. Willink (1980) Biogeografía de América Latina. Monografía 13. Programa Regional de Desarrollo Científico y Tecnológico, Organización de los Estados Americanos 122 pág.
- Cabrera, A.L. (1994) Regiones Fitogeográficas Argentinas, 2nd edn. ACME Buenos Aires.
- Caetano S., D. Prado, R.T. Pennington, B.A. OliveiraFolho, R. Spichiger e Y. Naciri (2008) The history of Seasonally Dry Tropical Forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology* 17:3147–3159.
- Caetano, S. e Y. Naciri (2009). Insights on the biogeography of Seasonally Dry Tropical Forest: Spatial distribution of the chloroplast DNA polymorphism in two Neotropical trees *Astronium urundeuva* (Anacardiaceae) and *Geoffroea spinosa* (Fabaceae). In Seasonally Dry Tropical Forests. R. Dirzo, H. Mooney, G. Ceballos Eds. Stanford University Press, Stanford.
- Chapuis, M. y A. Estoup (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24:621-31.
- Chung, S.M. y J.E. Staub (2003) The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theoretical and Applied Genetics* 107:757-767.
- Cialdella, A.M. (2000) Flora Fanerogámica Argentina. Proflora. Fascículo 67: Fabaceae Subfamilia Mimosoideae 1-10.
- Cockerham, C.C. y B.S. Weir (1987) Correlations, descent measures: Drift with migration and mutation. *Evolution* 84:8512-8514.

- Collevatti, R.G., L.C.Terribile, M.S. Lima-Ribeiro, J.C. Nabout, G.de Oliveira, T.F. Rangel, S.G. Rabelo, J.A.F. Diniz-Filho (2012) A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology* 21:5845-5863.
- Corander J., J. Sirén y E. Arjas (2008a) Bayesian spatial modelling of genetic population structure. *Computational Statistics* 23:111-129.
- Corander J., P. Marttinen y S. Mäntyniemi (2006) Bayesian identification of stock mixtures from molecular marker data. *Fishery Bulletin* 104:550-558.
- Corander J., P. Marttinen, J. Sirén y J. Tang (2008b) Enhanced bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9(539):1-14.
- Corander, J., P. Waldmann y M.J. Sillanpää (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163:367-374.
- Corriveau, J.L. y A.W. Coleman (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75(10):1443-1458.
- Crawford, N.G. (2010) SMOGD: Software for the measurement of genetic diversity. *Molecular Ecology* 10(3):556-557.
- Debout, G.D., J.L. Doucet y O.J. Hardy (2011) Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity* 106:88-99.
- Deguilloux, M.F., M.H. Pemonge y R. Petit (2004) Use of chloroplast microsatellites to differentiate oak populations. *Annals of Forest Science* 61:825-830.
- Di Bitetti, M.S., G. Placci y L.A. Dietz (2003): Una visión de biodiversidad para la ecorregión del Bosque Atlántico del Alto Paraná: diseño de un paisaje para la conservación de la biodiversidad y prioridades para las acciones de conservación. Washington, D.C., World Wildlife Fund 154 pagpág.
- Dick, C.W., G. Etchelecu y F. Austerlitz (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by natives insects and Africans honneybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology* 12:753-764.
- Dimitri, M.J., R.F.J. Leonardis y J.S. Biloni (2000) El nuevo libro del árbol. Manufacturas Editoriales Litográficas. S.A., 4a Edición, Madrid.
- Dobzhansky, T. (1973) Nothing in biology make sense except in the light of evolution. *American Biology Teacher* 35:125-129.
- Dutech, C., J. Seiter, P. Petronelli, H.I. Joly y P. Jarne (2002) Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Molecular Ecology* 11:725-738.

- Dutech, C., V.L. Sork, A.J. Irwin, P.E. Smouse y F.W. Davis (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, *Quercus lobata* (Fagaceae). *American Journal of Botany* 92(2):252-261.
- Earl, D.A. y B.M. vonHoldt (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2): 359-361.
- Ellegren, H. (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews* 5:435-445.
- Ennos, R.A. (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250-259.
- Evanno G., S. Regnaut y J. Goudet (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620.
- Excoffier, L. y H. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.
- Excoffier, L., P. Smouse y J. Quattro (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Faubet, P. y O.E. Gaggiotti (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* 178: 1491-1504.
- Ferreira, M. y D. Grattapaglia (1996) Introducao ao uso de marcadores moleculares em análise genética 220 pagpág.
- Finkeldey, R. y H.H. Hattmer (2007) Tropical forest genetics. Springer Verlag Edition 315 pagpág.
- Fisher, D. y K. Bachmann (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *Biotechniques* 24: 796-802.
- Foll, M. y O. Gaggiotti (2006) Identifying the Environmental Factors That Determine the Genetic Structure of Populations. *Genetics* 174: 875-891.
- Fontaine, C., P. Lovett, H. Sanou, J. Maley y J. Bouvet (2004) Genetic diversity of the shea tree (*Vitellaria paradoxa* C. F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* 93: 639-648.
- Fontana, J.L. (1993) Los pajonales Mesófilos y Hidrófilos del Sur de Misiones (Argentina). Composición Florística Hábitat y sindinámica. Unité d'Ecologie et de Biogéographie Faculté des Sciences Université Catholique de Louvain 220 pag.
- Frankham, R., J.D. Ballou y D.A. Briscoe (2004) *Conservations Genetics*. Cambridge University Press 617 pag.
- Freeland, J.R. (2005) *Molecular Ecology*. Wiley-Liss 388 pag.

- Gibbs P.E. y R. Ingram (1982) Notes RBG Edinburg 40(2):399-407.
- Gillet, E., D. Gömöry y L. Paule (2005) Measuring genetic variation within and among populations at marker loci. En Conservation and Management of forest genetic resource in Europe. Editado por Geburek, T. y J. Turok. Arbora Publishers. 693 pág.
- González-Martínez, S.C., K. V. Krutovsky y D. B. Neale (2006) Forest-tree population genomics and adaptive evolution. *New phytologist* 170(2): 227-238.
- González-Martínez, S.C., S. Gerber, M. T. Cervera, J. M. Martínez-Zapater, L. Gil y R. Alía (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. *Theoretical and Applied Genetics* 104:1290-1297.
- Grauke, L.J., M.A. Mendoza-Herrera, A.J. Miller y B.W. Wood (2011) Geographic patterns of genetic variation in native pecans. *Tree Genetics & Genomes* 7(5):917-932.
- Greiner, S., U. Rauwolf, J. Meurer y R.G. Herrmann (2011) The role of plastids in plant speciation. *Molecular Ecology* 20:671-691.
- Hamilton, M.B. (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8:521-523.
- Hamrick J.L., Murawski D.A. y Nason J.D. (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* 107(108):281-297.
- Hamrick, J.L. (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management* 197: 323-335.
- Hamrick, J.L. y D.A. Murawski (1990) The breeding structure of tropical tree populations. *Plant Species Biology* 5:157-165.
- Hansen, O.K., E.D. Kjaer y G.G. Vendramin (2005) Chloroplast microsatellite variation in *Abies nordmannia* and simulation of causes of low differentiation among populations. *Tree Genetics & Genomes* 1:116-123.
- Hartl, D.L. y A.G. Clark (2007) Principles of population genetics. Sinauer Associates, Inc Publishers 652 pág.
- Heinze, B. (2007) A database of PCR primers for the chloroplast genomes of higher plants. *Plants Methods* 3:4.
- Heuertz M., X. Vekemans, J.F. Hausman, M. Palada y O.J. Hardy (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology* 12: 2483-2495.
- Holsinger, K.E. (1999) Genetic diversity in geographically structured populations: A Bayesian perspective. *Hereditas* 130:245-255.
- Honfi, A.I. y J.R. Daviña (1997) El cariotipo de *Anadenanthera colubrina* var *cebil* y de *Parapiptadenia rígida* (Leguminosae – Mimosoideae) *Bonplandia* 9:277-280.

- Huff, D.R., R. Peakall y P.E. Smouse (1993) RAPD variation within and among natural populations of outcrossing buffalograss (*Buchloe dactyloides* (Nutt.) Engelm.). *Theoretical and Applied Genetics* 86(8):927-934.
- Hutchinson, D.W. y A.R. Templeton (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898-1914.
- Inza M.V., N. Zelener, L. Fornes y L.A. Gallo (2012) Effect of latitudinal gradient and impact of logging in genetic diversity of *Cedrel lilloi* along the Argentina Yungas Rainforest. *Ecology and Evolution* 2(11):2722-2736.
- Jost, L. (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* 17(18):4015-4026.
- Justiniano, M.J. y T.S. Fredericksen (1998) Ecología y silvicultura de especies menos conocidas Curupá *Anadenanthera colubrina* (Vell. Conc.) Benth. Mimosoideae. Proyecto de Manejo Forestal Sostenible (BOLFOR) 31 pág.
- Kalia, R.K., M.K. Rai, S. Kalia, R. Singh y A.K. Dhawan (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177:309-334.
- Karhu, A. (2001) Evolution and applications of pine microsatellites. Academic dissertation. Faculty of Science. University of Oulu. Oulu. University press 52 pág.
- Kasturi, J., R. Acharya y M. Ramanathan (2003) An information theoretic approach for analyzing temporal patterns of gene expression. *Bioinformatics*. Oxford University Press 19 (4):449-458.
- Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* 217:624-626.
- Kimura, M. (1991) Recent developments of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proceedings of the National Academy of Sciences of the United States of America* 88: 5969-5973.
- Kimura, M. y G.H. Weiss (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.
- Kullback, S. y R.A. Leibler (1951) On information and sufficiency. *The Annals of Mathematical Statistics* 22:79-86.
- Li, Y., A.B. Korol, T. Fahima, A. Beiles y E. Nevo (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* 11:243-2465.
- Ligier, HD, HR Matteio, HL Polo y JR Rosso (1985) Atlas de suelos de la República Argentina. Tomo II. Secretaría de Agricultura, Ganadería y Pesca. Proyecto PNUD ARG 85/019. 154 pág.
- Linhart, Y.B. y M.C. Grant (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* 27:237-77.

- Lira, C.F., S.R.S. Cardoso, P. C. G. Ferreira, M.A. Cardoso y J. Provan (2003) Long-term population isolation in the endangered tropical tree species *Caesalpinia echinata* Lam. revealed by chloroplast microsatellites. *Molecular Ecology* 12:3219-3225.
- Lodish, H., A. Berk, S. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell (2002) *Biología celular y molecular*. Cuarta Edición. Editorial Panamericana Buenos Aires, Argentina 1084 pág.
- Loveless, M.D. y J.L. Hamrick (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology & Systematic* 15:65-95.
- Marcucci Poltri, S. (2005) *Marcadores Moleculares aplicados a Programas de Mejoramiento Genético de Eucalyptus*. Editado por Secretaría de Agricultura. Ganadería, Pesca y Alimentos, Mejores árboles para más forestadores ISBN: 987-9184-47-5 241 pág.
- McCauley, D.E. (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution* 10(5):198-202.
- McCauley, D.E. (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity* 88:257-263.
- Meirmans, P.G. y P.W. Hedrick (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* 11(1):5-18.
- Muller, F., M. Voccia, A. Ba y J.M. Bouvet (2009) Genetic diversity and gene flow in a Caribbean tree *Pterocarpus officinalis* Jacq.: a study based on chloroplast and nuclear microsatellites. *Genética* 135:185-198.
- Naciri, Y., S. Caetano, R.T. Pennington, D. Prado y R. Spichiger (2006) Population Genetics and inference of ecosystem history: An example using two neotropical seasonally dry forest species. En *Neotropical Savannas and Seasonally dry forest: Plant diversity, biogeography, and conservation*. Editado por Pennington, P., G.P. Lewis y J.A. Ratter. The Systematics Association Special Volume Series 69. Taylor & Francis Group, London, New York 504 pág.
- Nason, J.D. y J.L. Hamrick (1997) Reproductive and Genetic Consequences of Forest Fragmentation : Two Case Studies of Neotropical Canopy Trees. *Journal of Heredity* 88:264-276.
- Navascués, M. y B.C. Emerson (2005) Chloroplast microsatellites: measure of genetic diversity and the effect of homoplasy. *Molecular Ecology* 14:1333-1341.
- Ndiade-Bourobou, D., O.J. Hardy, B. Favreau, H. Moussavou, E. Nzengue, A. Mignots y J.M. Bouvet (2010) Long distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. *Molecular Ecology* 19:4949-4962.

- Nei, M. (1987): Molecular evolutionary genetics. Columbia University Press, New York 512 pág.
- Oddou-Muratorio, S., R.J. Petit, B. Le Guerroue, D. Guesnet y B. Demesure (2001) Pollen versus Seed mediated gene flow in a scattered forest tree species. *Evolution* 55(6):1123-1135.
- Oliveira Alves, M.A y A. Valdinar de Carvalho Custodio (1989) *Revista Brasileira de Genética* 12:81-92.
- Ouborg, NJ, Y Piquot y MV Groenendael (1999) Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* 87:551-668.
- Pakkad, G., S. Ueno, H. Yoshimaru (2008) Genetic differentiation of *Quercus semiserrata* Roxb. In northern Thailand revealed by nuclear and chloroplast microsatellite markers. *Forest Ecology and Management* 255:1067-1077.
- Palmer, J.D., R.K. Jansen, H.J. Michaels, M.W. Chase y J.R. Manhart (1988) Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden* 75:1180-1206.
- Peakall, R. y P.E. Smouse (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288 - 295.
- Pearse, D.E. y K.A. Crandall (2004) Beyond F_{ST} : Analysis of population genetic data for conservation. *Conservation Genetics* 5:585-602.
- Pennington, R.T., D.E. Prado y C.A. Pendry (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography* 27: 261-273.
- Pennington, R.T., M. Lavin, D.E. Prado, C.A. Pendry, S.K. Pell (2004) Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society of London* 359:515-537.
- Pennington, R.T., M. Lavin y A.Oliveira-Filho (2009) Woody Plant Diversity. Evolution, and Ecology in the Tropics: Perspectives from Seasonally Dry Tropical Forests. *Annual Review of Ecology Evolution and Systematics* 40:437-457.
- Petit, R.J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini y G.G. Vendramin (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14:689-701.
- Planter, E.A. (2007) Flujo génico: métodos para estimarlo y marcadores moleculares. En *Ecología molecular* Editado por Eguiarte, L., V. Souza y X. Aguirre. México 594 pagpág.
- Prado, D. (2000) Seasonally dry forest of tropical South America: from forgotten ecosystems to a new phytogeography unit. *Edinburgh Journal of Botany* 57(3):437-461.
- Prado, D.E. y P.E. Gibbs (1993) Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden* 80:902-927.

- Prinz, K., S. Schie, T. Debener, I. Hensen y K. Weising (2009) Microsatellite markers for *Spergularia media* (L.) C. Presl. (Caryophyllaceae) and their cross-species transferability. *Molecular Ecology Notes* 9:1424-1426.
- Pritchard, J.K., M. Stephens y P. Donnelly (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Provan, J., N. Soranzo, N.J. Wilson, D. Golstein y W. Powell (1999) A Low Rate For Chloroplast Microsatellites. *Genetics* 153:943-947.
- Provan, J., W. Powell y P.M. Hollingsworth (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16 (3):142-147.
- Quiroga, M.P. y A.C. Premoli (2007) Genetic patterns in *Podocarpus parlatorei* reveal the long-term persistence of cold-tolerant elements in the southern Yungas. *Journal of Biogeography* 34:447-455.
- Quiroga, M.P., S. Pacheco, L. Malizia y A.C. Premoli (2012) Shrinking Forests under Warming: Evidence of *Podocarpus parlatorei* (pino del cerro) from the Subtropical Andes. *Journal of Heredity* 103(5):682-691.
- Ravi, V., J.P. Khurana, A.K. Tyagi y P. Khurana (2008) An update on chloroplast genomes. *Plant Systematic and Evolution* 271:101-122.
- Raymond, M. y F. Rousset (1995) GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *Journal of Heredity* 86:248-249.
- Reynolds, J., B.S. Weir y C.C. Cockerham (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* 105:767-779.
- Richard, E. (2000) Reserva Experimental Horco Molle. Pág: 249-262 En Manejo de fauna silvestre en Amazonia y Latinoamérica. Editado por Cabrera E., C. Mercolli y R. Resquin. Fundación Moisés Bertoni y University of Florida. Paraguay 578 pág.
- Robledo-Arnuncio, J.J. y S.C. Gonzalez-Martínez (2009) Marcadores moleculares y ecología del movimiento. *Ecosistemas* 18(1):44-51.
- Rozen, S. y H. Skaletsky (2000) Primer3 on the www for general users and for biologist programmers. En *Bioinformatics methods and protocols: Methods in molecular biology*. Editado por Krawetz S y S. Misener. Humana Press, Totowa, New Jersey, USA 365-386.
- Sato T., M. Nagoshi, S. Mor, K. Watanabe y Y. Kano. (2006) Long-term population fluctuation and present status of Kirikuchi char (*Salvelinus leucomaenis japonicus*). *Japanese Journal of Conservation Ecology* 11:13-20.
- Schlötterer, C. y D. Tautz (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20, 211-215.
- Selkoe K.A. y R.J. Toonen (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9:615-629.

- Slatkin, M. (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393-430.
- Slatkin, M. (1991) Inbreeding coefficients and coalescence times. *Genetic Resources* 58: 167–175.
- Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-462.
- Slatkin, M. (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* 14: 67 - 73.
- Slatkin, M. y H.E. Arter (1991). *Spatial Autocorrelation Methods in Population Genetics*. *The American Naturalist* 138(2): 499-517.
- Slatkin, M. y N.H. Barton (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43(7):1349-1368.
- Sork, V.L., J. Nason, D.R. Campbell y J.F. Fernandez (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution* 14(6):219-224.
- Stewart, C.N. y L. Excoffier (1996) Assessing population genetic structure and variability with RAPD data: Application to *Vaccinium macrocarpa* (American cranberry). *Journal of Evolutionary Biology* 9:153-171.
- Struss, D. y J. Plieske (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. *Theoretical and Applied Genetics* 97:308-315.
- Szpiech, Z.A., M. Jakobsson y Rosenberg N.A. (2008) *ADZE*: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24(21):2498-2504.
- Taberlet, P., L. Gielly, G. Pautou y J. Bouvet (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17:1105-1109.
- Tamura, K., J. Dudley, M. Nei y S. Kumar (2007) *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. *Molecular Biology and Evolution* 24:1596-1599.
- Tautz, D. y M. Renz (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acid Research* 12(10):4127-4138.
- Templeton, A.R. (2006) *Population genetics and microevolutionary theory*. Wiley-Liss Publication 705 pág.
- Tollefsrud, M.M., J.H. Sønstebø, C. Brochmann, Ø. Johnsen, T. Skrøppa y G.G. Vendramin. (2009) Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. *Heredity* 102(6):549-562.
- Tortorelli, L.A. (2009) *Maderas y bosques argentinos*. Buenos Aires, Argentina. Orientación Gráfica Editora 1111 pág.

- Trapnell, D.W. y J.L. Hamrick (2004) Partitioning nuclear and chloroplast variation at multiple spatial scales in the neotropical epiphytic orchid, *Laelia rubescens*. *Molecular Ecology* 13:2655–2666.
- Van Oosterhout, C., Hutchinson W.F., Wills D.P.M. y Shipley P. (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4:535-538.
- Vendramin, G.G., M. Anzidei, A. Madaghiele y G. Bucci (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. As revealed by chloroplast microsatellites. *Theoretical and Applied Genetics* 97:456-463.
- Vendramin, G. y B. Ziegenhagen (1997) Characterization and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome* 40:857-864.
- Weir, B.S. y C.C. Cockerham (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6): 1358-1370.
- Weising, K. y R. Gardner (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphism in chloroplast genomes of dicotyledonous. *Genome* 42:9-19.
- Weising, K., H. Nybom, K. Wolff y G. Kahl (2005) *DNA Fingerprinting in Plants Principles, Methods, and Applications Second Edition*. CRC Press Taylor & Francis Group Estados Unidos 470 pág.
- Werneck, F.P., G.C. Costa, G.R. Colli, D.E. Prado y J.W. Sites (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography* 20:272-288.
- White, T.W., W.T. Adams y D.B. Neale (2007) *Forest Genetics*. CAB International Publishing. Cambridge USA 702 pág.
- Whitlock M.C. y D.E. McCauley (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity* 82:117-125.
- Whitlock, M.C. (2011) G_{ST} and D do not replace F_{ST} . *Molecular Ecology* 20:1083-1091.
- Wilson, G.A. y B. Rannala (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177-1191.
- Wright, S. (1931) Evolution in Mendelian Populations. *Genetics* 16: 97-159.
- Wright, S. (1943) Isolation by Distance. *Genetics* 28(2):114-138.
- Wright, S. (1951) The genetical structure of populations. *Annals of Eugenetics* 15:323-354.
- Yamamichi, M. e Innan (2012) Estimating the migration rate from genetic variation data. *Heredity* 108:362-363.
- Yeh, F.C. y T. J.B. Boyle (1997): Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* 129:157.

- Young, A., T. Boyle y T. Brown (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* 11:413-418.
- Zane, L., L. Bargelloni y T. Patarnello (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11:1-16.
- Zhan, Q.Q., J.F. Wang, X. Gong y H. Peng (2011) Patterns of chloroplast DNA variation in *Cycas debaoensis* (Cycadaceae): conservation implications. *Conservation Genetics* 12:959-970.