



**UNIVERSIDAD DE MURCIA**

**FACULTAD DE BIOLOGÍA**

**Tesis doctoral**

**Estudio de la evolución del espaciador ribosomal intergénico 45S (IGS45S) y otras familias de ADN repetido en plantas, mediante técnicas moleculares y citogenéticas**

**D. Jose Antonio Galián Megías**

**2014**



Jose Antonio Galián Megías

Tesis doctoral 2014

Estudio de la evolución del espaciador ribosomal intergénico 45S (IGS45S) y otras familias de ADN repetido en plantas, mediante técnicas moleculares y citogenéticas

PROGRAMA DE DOCTORADO EN BIODIVERSIDAD Y GESTIÓN AMBIENTAL

Director: Josep Antoni Rosselló Picornell

(Profesor titular Universidad de Valencia)

Directora: Marcela Rosato

(Doctora Investigadora Instituto Cavanilles de Biodiversidad y Biología Evolutiva)

Tutor Académico Universidad de Murcia: José Sebastián Carrión García

(Catedrático de Universidad)



**a Pep y a Marcela,  
y a todos los demás, pero sobre todo a ellos.**



# ÍNDICE

**INTRODUCCIÓN GENERAL**    pág. 1

**PRINCIPALES OBJETIVOS Y CONCLUSIONES**    pág. 13

**COROLARIO**    pág. 16

**BIBLIOGRAFÍA**    pág. 18

**ARTÍCULO I**    pág. 31

**Amplification, contraction and genomic spread of a satellite DNA family (E180) in Medicago (Fabaceae) and allied genera (Annals of Botany 109: 773–782, 2012)**

**ARTÍCULO II**    pág. 43

**Partial Sequence Homogenization in the 5S Multigene Families May Generate Sequence Chimeras and Spurious Results in Phylogenetic Reconstructions (Systematic Biology 63(2): 219-230, 2014)**

**ARTÍCULO III**    pág. 59

**Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm Ginkgo biloba (Heredity 108(6): 640-646, 2012)**





## Introducción general

Dentro de los genomas las secuencias de ADN codificante pueden suponer menos del 2% del contenido total de ADN, siendo gran parte del 98% restante ADN repetido. Hasta no hace mucho este ADN mayoritario se consideraba carente de función o “ADN basura”, debido a la falta de datos sobre su utilidad para la célula. Pero cada vez son las referencias que atribuyen a este ADN repetido diversas funciones relacionadas con la arquitectura del genoma, la correcta condensación y mantenimiento de la estructura de los cromosomas en las diferentes fases del ciclo celular y el reconocimiento de los cromosomas homólogos durante la profase I de la meiosis (Csink and Henikoff, 1998; Ugarković and Plohl, 2002, Hemleben et al. 2007). Este ADN repetido a su vez puede clasificarse según la longitud y la organización de las repeticiones dentro del genoma:

Las secuencias repetidas en tándem están constituidas por una secuencia determinada que se repite consecutivamente una detrás de otra en un número variable de veces. Estas regiones repetitivas, repartidas por el genoma, tienen una localización cromosómica preferentemente en los centrómeros y los telómeros. Atendiendo a la unidad de repetición estas secuencias se denominan:

- Satélites. La unidad de repetición oscila entre 100 y 10000 nucleótidos, típico de centrómeros.
- Minisatélites o polimorfismos VNTR (*variable number of tandem repeat*). La unidad de repetición es menor, alrededor de 7-100 nucleótidos. Así según el número de repeticiones de esta secuencia base, los polimorfismos en longitud VNTR varían entre 500 y 10000 nucleótidos.
- Microsatélites o polimorfismos STR (*short tandem repeat*). La unidad de repetición está formada por 2-6 nucleótidos, y según el número variable de

repeticiones los polimorfismos de longitud STR oscilan entre 100 y 500 nucleótidos.

Tanto mini- como microsátélites se han venido usando, mediante análisis por PCR o RFLP, en pruebas forenses, de paternidad y parentesco, y asesoramiento genético mediante análisis de ligamiento. Ambos tipos de repetición pueden incluir pequeñas variaciones de la secuencia base.

Además de este ADN repetido en tándem, existe otra clase de ADN repetitivo con secuencias dispersas a lo largo de todo el genoma con posibilidad de transposición. Las más conocidas son:

- Las secuencias SINE (*short interspersed nuclear elements*), la más conocida es la secuencia *Alu* humana, de unos 300 pares de bases, repetida unas 750000 veces, representando de un 3 a un 5% del total de ADN.
- Las secuencias LINE (*long interspersed nuclear elements*), siendo la más conocida la L1 en humanos, de unas 6 kilobases (Kb) de largo repetida unas 10000 veces, representando un 3% del genoma. Se localizan en las bandas G y además presentan un promotor para la ARN polimerasa II. Los genes próximos a LINEs presentan una expresión disminuida, y el 80% de genes humanos presentan alguna secuencia L1 en sus intrones.

Por último y más importante en cuanto al contenido de esta tesis, están las familias multigénicas de ADN ribosomal, que codifican para el ARN ribosómico que forma el armazón de los ribosomas celulares. Estrictamente hablando forman un ADN repetitivo mixto, porque forma largos tándem de cientos a miles de unidades de repetición y en muchas especies se encuentra disperso en numerosos cromosomas dentro de cada célula.

El argumento central de esta tesis será por tanto, el conocimiento de la evolución de diversas familias de ADN repetido en plantas, atendiendo a su distribución cromosómica, estructura y variabilidad molecular, además de contrastar los hallazgos con los diversos modelos sobre su organización general y comportamiento evolutivo, y evaluaremos su utilidad como herramientas manejables en estudios filogenéticos. Hablaremos brevemente ahora, para dar introducción a cada uno de los trabajos de esta tesis, de las familias de ADN repetido así como de las especies de plantas en las que hemos llevado a cabo dichos estudios.

### *ADN satélite*

El ADN satélite es un componente genómico teóricamente presente en todos los organismos eucarióticos. El movimiento de este ADN satélite altamente repetido es un elemento importante en la evolución y organización del genoma en las plantas. Con los proyectos de secuenciación del genoma nuclear cerca de su consecución, se ha hecho mucho más fácil investigar la diversidad, abundancia y distribución cromosómica de las secuencias de ADN repetido en plantas. Gracias a estos avances, junto con otros resultados obtenidos por otras técnicas convencionales (centrifugación en gradiente de CsCl, bandeó C de citogenética clásica, Southern y dot blot análisis, citogenética molecular) se ha podido vislumbrar una imagen global de la estructura del genoma de plantas. Las secuencias de ADN repetido forman la mayor parte del genoma nuclear, y estas secuencias constan de muchas clases de motivos repetidos que van desde dos hasta más de diez mil nucleótidos (Kubis et al., 1998; Heslop-Harrison 2000; Macas et al., 2007). El ADN satélite (ADNsat) consta de secuencias de ADN altamente repetidas y se organiza en largas cadenas de miles a millones de unidades repetidas en tándem. Estas

largas cadenas representan un gran parte del ADN que forma la heterocromatina constitutiva, de alto contenido A+T y replicación tardía (Ugarković and Plohl, 2002). Diferentes secuencias de ADN satélite pueden coexistir en los genomas, formando lo que se ha venido definiendo como *librería de ADNsat* (Fry and Salser, 1977; Mestrovic et al., 1998). El ADN satélite se suele plegar en estructuras de orden superior [lo que se sugiere de su capacidad intrínseca de inducir curvaturas en el ADN], algo que puede jugar un papel fundamental en el empaquetamiento del ADN y proteínas asociadas de la heterocromatina (Ugarković and Plohl, 2002). A pesar de lo relativamente bien que se conoce el papel de algunos ADNsat (centromérico y telomérico) en la estabilización de la estructura cromosómica y los mecanismos de división celular, su significancia biológica global no está del todo clara (Csink and Henikoff, 1998). Las secuencias de ADNsat evolucionan a través de procesos evolutivos propuestos como el *molecular drive model* (Dover 2002). De esta manera, las reorganizaciones del ADN llevan a una alta tasa de similitud intraespecífica (homogeneización) de aquellas secuencias en tándem pertenecientes a la misma familia de ADNsat. Sin embargo, algunos estudios han revelado que hay un equilibrio entre la homogeneización y la persistencia o generación de variantes en el ADNsat. Este equilibrio podría generar suficiente divergencia entre las secuencias como para causar un aislamiento reproductivo entre las diferentes líneas evolutivas de una especie, causando en último término la especiación. Sobre esta base se ha hipotetizado que las familias de ADNsat están sometidas a presión selectiva y limitaciones evolutivas (Ugarković and Plohl, 2002). Con esto, se acepta de manera general que la similitud entre las secuencias, la distribución genómica y el número de copias del ADNsat son características especie-específicas, y no se esperaría encontrar concordancia de alguno de estos parámetros entre diferentes especies, más si no están muy estrechamente relacionadas (Kubis et al., 1998; Suárez-Santiago et al.,

2007). Sin embargo, algunas de las secuencias satélite han permanecido inalterables durante largos periodos evolutivos (Abad et al., 1992; Heikkinen et al., 1995). Aunque la evolución de las familias de ADNsat es un elemento importante en la organización y evolución del genoma de las plantas, la mayoría de los esfuerzos han ido encaminados hacia el aislamiento, la caracterización de la secuencia y la variación en número de copias de las familias de ADNsat de especies modelo o agronómicamente importantes y en especies estrechamente relacionadas (Contento et al., 2005; Macas et al., 2006; Hemleben et al., 2007; Ambrožová et al., 2011). Por el contrario, es poco lo que se sabe acerca de su distribución y organización dentro de los genomas y su evolución desde un punto de vista filogenético.

*Medicago* es un género dentro de las leguminosas (Fabaceae) que incluye especies económicamente importantes como *M. sativa* (alfalfa), *M. scutellata* (medicago caracol) y *M. lupulina* (medicago negro), además de otras como la especie modelo de la biología del género *M. truncatula*. Pertenece al clado Vicioide (Sanderson and Wojciechowski, 1995), un grupo monofilético que comprende las tribus Cicereae, Trifolieae y Vicieae, el género *Galega* (Galegeae).

El ADNsat E180 consta de un monómero rico en AT de 185-189 pb de longitud. Este ADNsat constituye cerca del 1% del genoma de *M. sativa*, aproximadamente  $1.8 \times 10^5$  copias (Xia and Erickson, 1993). Calderini et al. (1997) clonaron una secuencia similar (C300) en *M. caerulea*. El alto grado de similaridad entre los monómeros de E180 y C300 (96%) y el idéntico número de copias por genoma estimado, sugiere que ambas secuencias son en realidad variantes pertenecientes a la misma familia de ADNsat (por consenso la familia E180). En el mismo trabajo Calderini et al. (1997) encontró la presencia de esta familia de ADNsat en otros miembros del complejo *M. sativa* (*M. sativa*, *M. caerulea* and *M. falcata*) pero no así en la especie leñosa de medicago *M.*

*arbórea*. La pequeña muestra empleada en este trabajo (ya que *Medicago* es un género que incluye 85 especies repartidas en 12 secciones ampliamente reconocidas, Small and Jomphe 1989) puede no reflejar la verdadera distribución de la familia de ADNsat E180 en *Medicago*. Aunque se han propuesto diversas hipótesis, todos los análisis (con secuencias nucleares, plastidiales y mitocondriales) muestran que el complejo *M. sativa* pertenece a un clado derivado dentro del género (Bena et al., 1998, b; Downie et al., 1998; Bena, 2001; Maureira-Butler et al., 2008; Steele et al., 2010). Por lo tanto, si la familia de ADNsat E180 es específica del complejo *sativa*, debería haber aparecido recientemente en la evolución. En el trabajo de esta tesis dedicado al estudio del E180 en el género *Medicago*, caracterizamos la presencia y la distribución genómica en 70 especies tanto de *Medicago*, como de los géneros filogenéticamente relacionados *Trigonella* y *Melilotus* y el menos relacionado *Trifolium*; (todos ellos incluidos en clado Vicioide; Wojciechowski et al., 2000), usando las técnicas de PCR e hibridación *in situ* fluorescente (FISH, por sus siglas en inglés).

### *Genes de ADN ribosomal 5S*

Las familias multigénicas son conjuntos de genes parálogos descendientes por procesos de amplificación de genes ancestrales ortólogos. Dichas familias están compuestas por cientos o miles de copias (supuestamente idénticas) dispuestas en tándem y separadas por espaciadores intergénicos. Todas ellas pueden estar reunidas en un simple locus o repartidas en varios loci. A su vez, estos loci pueden localizarse en el mismo cromosoma, o estar repartidos en varios (King and Stanfield 1990).

El concepto de evolución concertada, el cual postula que genes miembros dentro una familia repetida no evolucionan independientemente unos de otros, fue propuesto para

dar coherencia a la evolución homogénea observada en ciertas secuencias dispuestas en tándem, lo cual contrastaba con el modelo clásico que postula que los cambios se acumulan en las secuencias incluso si éstas se encuentran en el mismo sitio cromosómico (Brown et al. 1972, Zimmer et al. 1980, Dover 1982). La evolución concertada es una hipótesis biológica muy extendida que explica la eliminación de las mutaciones a lo largo de estas regiones de repetición, protegiendo la estabilidad estructural minimizando las variaciones de productos clave que son requeridos sin grandes cambios y en gran cantidad para el correcto desarrollo de su función dentro las células (reviews in Ohta 1980; Arnheim 1983; Dover et al. 1993; Elder and Turner 1995; Liao 1999).

Aunque los estudios de las familias multigénicas han dado muchas oportunidades a los biólogos evolucionistas para evaluar los procesos de evolución molecular a diferentes escalas sistemáticas, su uso no está exento de retos técnicos y analíticos. Entre los más citados están los fenómenos de *PCR selection and drift* (Wagner et al.1994) y los resultados filogenéticos espurios producidos por parálogos divergentes dentro de un mismo genoma (Mayol and Rosselló 2001, Álvarez and Wendel 2003).

En eucariotas, una de las secuencias más abundantes de ADN repetido en tándem pertenece a la familia multigénica de ADN ribosomal 5S (ADNr 5S). Esta familia consiste en una región altamente conservada de unas 120 pb que se alterna con un espaciador de longitud variable (Long and David 1980). La región codificante del ADNr 5S codifica para el ARN 5S, que junto con los ARN 5.8S y 25S, más unas 50 proteínas ribosomales, constituyen la subunidad grande de los ribosomas citoplasmáticos eucarióticos. Dado que los genes de ARNr 5S están presentes en un número muy elevado por genoma nuclear, son de pequeño tamaño y las secuencias codificantes han sido muy conservadas durante la evolución, es fácil el uso de primers

de PCR universales para comparar una amplia variedad de organismos (por ejemplo las angiospermas; Cox et al. 1992), lo que ha contribuido a su extendido uso en medicina forense, la identificación de especies, reconstrucciones filogenéticas y evolutivas, y análisis de evolución molecular a lo largo de casi todo el árbol de la vida (por ejemplo Hori et al. 1985, Sastri et al. 1992, Baker et al. 2000, Favia et al. 2000, Baum and Bailey 2004, Doveri and Lee 2007, Sucher and Carles 2008).

Sin embargo se hace evidente la diversidad intragenómica que existe para las unidades de ARNr nuclear (ARNrn) en eucariotas, ya que la mayoría de trabajos con primers universales utilizan técnicas de clonación para aislar secuencias de un pool heterogéneo de PCR (Cronn et al. 1996, Sajdak et al. 1998, Fernández-Tajes and Méndez 2009). Por tanto, la asunción de que la familia multigénica 5S es un ejemplo paradigmático de familia sujeta a estricta evolución concertada está siendo puesta en duda. Los trabajos que muestran variabilidad dentro de un mismo tándem, la detección de supuestos pseudogenes, y la presencia de transcritos de ARN 5S divergentes (incluso en ribosomas maduros, Dimarco et al. 2012) han puesto de manifiesto un escenario mucho más complejo. Consecuentemente está ganando fuerza la hipótesis de que los mecanismos de birth-and-death (Nei and Rooney 2005) guían los mecanismos de evolución de esta familia ribosomal a largo plazo, quizá con el efecto combinado de la evolución concertada (Fujiwara et al. 2009).

Cómo toda esta diversidad genética está particionada dentro de los genomas (dentro de locus, entre loci o ambas cosas) ha sido rara vez referida en la literatura (Cloix et al. 2000, 2002). Esto es debido sin duda al inusual alto número de copias de unidades 5S por tándem (de cientos a miles), lo que hace imposible amplificarlo por técnicas de PCR convencional. De hecho hay relativamente pocas herramientas de análisis para evaluar la estructura y diversidad de los loci de ARNr 5S, a no ser que éstos estén compuestos



por unidades dispersas o asociadas a otros genes, tales como los de ADNrn 45S (Drouin and Moniz de Sá 1995, Drouin et al. 1992, Cross and Rebordinos 2005, Galián et al. 2012).

Si revisamos muchos de los estudios filogenéticos y evolutivos que han usado el ADNrn 5S hasta ahora, nos damos cuenta de que solo una pequeña proporción de los mismos usaron secuencias enteras de la región codificante (Trontin et al. 1999; Fallistocco et al. 2007). En la mayoría de los casos se usan primers universales anclados en la región codificante y se analizan amplicones formados por zonas parciales de dos regiones codificantes contiguas separadas por la región espaciadora, mucho más variable. Estas secuencias son frecuentemente usadas directamente en los análisis evolutivos (Volkov et al. 2001). Más todavía, las secuencias parciales de las regiones codificantes obtenidas por el procedimiento anterior, son normalmente concatenadas creando genes quiméricos *in silico* (formados por la mitad de un gen 5S más la otra mitad de un gen 5S contiguo), los cuales son directamente usados en los análisis (Kellogg and Appels, 1995, Cronn et al. 1996, Ko and Henry, 1996, Crisp et al. 1999, Benabdelmouna et al. 2001, Liu et al. 2003, Saini and Jawali 2009). Sorprendentemente todavía no han sido evaluados los riesgos potenciales que tienen para las reconstrucciones filogenéticas y los patrones de evolución los análisis basados en estos genes quiméricos.

#### *Espaciador intergénico del ADN ribosomal nuclear 45S (IGS del ADNrn 45S)*

Los ribosomas son estructuras celulares ubicuas en el citoplasma, matriz mitocondrial y cloroplástica, de organismos vivos. Están universalmente formados por dos subunidades de distinto tamaño, las cuales están constituidas por moléculas de ARN ribosomal (ARNr) y entre 60 y 80 ribonucleoproteínas que ayudan al correcto plegamiento de los

ARNs (Lafontaine and Tollervey 2001). Dado el papel clave que desempeñan los ribosomas en la síntesis de proteínas, los genes de ARN ribosomal constituyen una de las familias de ADN repetido en tándem más abundantes y conservadas de eucariotas (Long and David, 1980). El ADN ribosomal se divide, de forma general en eucariotas, en dos familias de genes: el ADNr 5S y el ADNr45S (también llamado 35S, Komarova et al., 2008). La unidad de repetición del ADNr 5S consiste en unas 120 pb de gen junto con un espaciador intergénico (IGS) que normalmente tiene de 100 a 900 pb de longitud (Sastri et al., 1992). La familia de ADNr, 45S, localizada en la región organizadora nucleolar, y consiste en tres regiones codificantes, 18S, 5.8S y 25S/26S/28S (25S y 26S en plantas y 28S en animales), dos espaciadores intergénicos que se transcriben y separan los tres genes anteriores (ITS1 e ITS2), y un largo espaciador (IGS) que separa unidades contiguas de todo lo anterior (Rogers and Bendich, 1987; Jorgensen and Cluster, 1988). Ambas familias de ADNr (45S y 5S) suelen presentarse en loci independientes en la mayoría de eucariotas, sin embargo, la integración de los genes 5S dentro de la secuencia de ADNr 45S se ha referenciado en diferentes grupos evolutivos (Maizels, 1976; Batts-Young and Lodish, 1978; Vahidi et al., 1991; Capesius, 1997; Sone et al., 1999; Vitturi et al., 2002; Bergeron and Drouin, 2008, García et al. 2009). En estos casos los genes 5S están incluidos dentro del espaciador IGS, entre los genes 18S y 25S, aunque se han descrito organizaciones mucho más raras (Drouin and Moniz de Sa', 1995).

#### Caso de estudio en *Ginkgo biloba*

Durante mucho tiempo, las especies de plantas vasculares se caracterizaron por presentar ambas familias de ADNr en sitios cromosómicos diferentes, a diferencia de

las especies más antiguas de plantas terrestres (briófitos; Sone et al., 1999). Recientemente, sin embargo, se ha demostrado inequívocamente, mediante el empleo de técnicas de citogenética molecular, Southern, PCR y secuenciación, que los genes 5S pueden estar incluidos dentro de la secuencia de ADN<sub>r</sub>45S en angiospermas (Asteraceae; García et al., 2009; García et al., 2010). El estudio detallado dentro de Asteraceae ha revelado que cerca del 25% de las especies de esta familia tienen ligadas ambas familias en un sola de ADN<sub>r</sub> 45S-5S (García et al. 2010). Los autores de este estudio sugieren que la integración del 5S dentro del 45S ha ocurrido probablemente una sola vez en Asteraceae, pero que puede haber ocurrido repetidamente en otras ramas del árbol evolutivo de plantas. Recientemente, Wicke et al. (2011) evaluaron la organización del ADN<sub>r</sub> nuclear en plantas, y sus resultados concluyeron que todas las plantas con semillas que analizaron experimentalmente presentaban la organización separada, incluidas 10 gimnospermas, el grupo basal de las angiospermas Nymphaeales, y 10 especies más en las que extrajeron las secuencias de los espaciadores IGS del 45S de las bases de datos. Estos resultados daban a entender que la organización asociada (45S-5S) en las plantas con semilla era exclusiva de la familia Asteraceae, una familia derivada de angiospermas, y que la asociación de ambas familias ribosomales ocurrió probablemente hace 7-10 millones de años (García et al., 2010). Algunos de los resultados obtenidos por Wicke et al. (2011) en plantas con semilla son ciertamente intrigantes. La yuxtaposición de las señales fluorescentes de sondas de ADN específicas del ADN ribosomal 45S y 5S en FISH en cromosomas o en núcleos en interfase, ha sido siempre usada como una evidencia de la asociación de ambas familias ribosomales. En plantas con semillas, la colocación de las señales de FISH del 45S Y 5S se ha observado además de en Asteraceae, en otras angiospermas como *Tulipa* (Mizuochi et al., 2007), *Rhoeo* (Golczyk et al., 2005), *Linum* (Muravenko et al., 2004), *Silene* (Široký

et al., 2001), *Brassica* (Snowdon et al., 2000; Hasterok et al., 2001), así como en dos géneros de gimnospermas: *Ginkgo* (Nakao et al., 2005) y *Podocarpus* (Murray et al., 2002). Sin embargo, dados los límites de resolución del FISH clásico (entre 2000 y 10000 Kb; Figueroa and Bass, 2010), las limitaciones de esta técnica pueden generar artefactos visuales y dar como solapados dos loci que en realidad estén muy cercanos (pero separados) en la hebra del cromosoma.

*Ginkgo biloba* L. es el género de plantas con semilla más antiguo que se conoce, y ya fue descrito por Charles Darwin como fósil viviente debido a su aislamiento filogenético, su rara distribución, y su marcado parecido morfológico con sus congéneres mesozoicos y cenozoicos (Major, 1967; Royer et al., 2003). *G. biloba* es el único superviviente del clado Ginkgoales, otrora un clado diverso y dominante de gimnospermas (con hasta 6 familias y 19 géneros; Tralau, 1968) que existió desde el Pérmico (hace más de 250 millones de años) y alcanzó su máxima diversidad durante el Jurásico y el Cretácico temprano (Royer et al., 2003). Otros autores, sobre la base del parecido con el fósil *G. adiantoides* han sugerido su conespecificidad y que *G. biloba* en realidad existió desde el Cretácico tardío al Mioceno medio (Royer et al., 2003).

## Principales objetivos y conclusiones de los trabajos

### *ADN satélite*

Nuestros principales objetivos fueron (1) evaluar la robustez y sensibilidad de las dos técnicas moleculares para detectar familias de ADN repetido, (2) obtener nuevos datos sobre la evolución genómica (cariotípica) del género *Medicago* usando ADNsat como marcador, y (3) evaluar el rastro filogenético dejado por una familia de ADNsat en un género tan particular como *Medicago*, donde se tiene la evidencia de que complejos patrones de hibridación han jugado un papel fundamental en su historia evolutiva (Maureira-Butler et al., 2008).

Nuestros resultados demuestran que la detección tanto por técnicas moleculares como citogenéticas de la familia repetida E180 es altamente reproducible a través del género, y que los patrones cromosómicos sirvieron para diferenciar complejos de especies estrechamente relacionados (son taxón-específicos). Sin embargo, el uso de la familia E180 como marcador filogenético en *Medicago* debe considerarse con precaución. Su amplificación parece haberse producido a través de episodios evolutivos recurrentes e independientes, tanto en especies de *Medicago* anuales como perennes, así como en clados basales y derivados.

### *Genes de ADNrn 5S*

En el segundo trabajo de esta tesis, comparamos la integridad de la secuencia (longitud, motivos universales conservados y estructura secundaria) y el comportamiento filogenético de secuencias 5S codificantes completas en comparación con las

quiméricas (concatenadas) de un grupo de genes de ADNrn 5S obtenido de varias especies estrechamente relacionadas.

Los resultados que se desprenden sugieren poderosamente que la homogeneización de secuencias no está operando, ni siquiera dentro de un mismo tándem, en la región codificante del ADNrn 5S, la cual había sido tradicionalmente considerada como un ejemplo de secuencia altamente conservada. De esta manera, la generalizada práctica de concatenar secuencias de genes 5S puede incrementar la diversidad haplotípica, distorsionando los patrones de evolución génica y conduciendo a relaciones haplotípicas incorrectas en algunas reconstrucciones evolutivas.

#### *Estudio del espaciador intergénico (IGS) del ADNr 45S de G. biloba*

Dados los contradictorios resultados mostrados por dos equipos diferentes, en este trabajo intentamos discernir si las familias de ADNrn 45S y 5S estaban o no combinadas en el mismo locus en la gimnosperma *G. biloba*, usando técnicas de citogenética molecular y secuenciación de ADN del IGS del cistrón ribosomal 45S.

Los resultados obtenidos dan luz a las observaciones contradictorias que se han venido vertiendo sobre la organización cromosómica y estructural de las familias de ADNr 5S y 45S en *G. biloba*. Nuestro trabajo muestra que la única organización existente en esta especie es la asociada (demostrándose por primera vez en gimnospermas), donde el gen 5S aparece insertado dentro del IGS 45S, en clara concordancia con los trabajos previos de citogenética molecular (Nakao et al., 2005). Además, nuestros análisis sobre la funcionalidad de las variantes encontradas, sugieren poderosamente que las variantes detectadas mediante PCR clásica por Wicke et al. (2011), y que carecen del gen inserto 5S, pertenecen en realidad a un determinado tipo de pseudogenes, lo que condujo a la

conclusión errónea de que los genes de ADNr 45S y 5S en *G. biloba* no estaban ligados. De tal manera, cuando se pensaba que la asociación de ambas familias ribosomales en plantas con semilla había ocurrido de manera puntual y de forma tardía (en la derivada familia Asteraceae; García et al., 2009 y 2010) nuestros resultados evidencian la necesidad de reevaluar la evolución de la organización de las familias ribosomales a través del árbol evolutivo de las plantas.

## Corolario de la tesis

El estudio de las familias de ADN repetido ha supuesto desde siempre un reto tanto técnico como analítico, debido al alto número de copias por genoma haploide, y por el nivel de variabilidad encontrado en muchas de ellas. Tal hecho ha relegado históricamente a aquellas familias más variables de ADN repetido en tándem (ADN minisatélite, microsatélite... etc.) al rastreo de la huella genética de poblaciones o incluso individuos, casi nunca preguntándose o analizando los porqués de tal nivel de variabilidad y su significación funcional para la célula, si es que la tuviera. Por otro lado, las familias de ADN repetitivo disperso (LINEs, SINEs... etc.) provenientes en su mayoría de familias de virus atemperados que han permanecido en el genoma, y que suponen en algunos casos un gran porcentaje de la cantidad total de ADN nuclear, se observaron con precaución tras su descubrimiento por ser secuencias enormemente variables (de un nivel igual o superior a los ADN satélites) y con una movilidad tan alta (p.ej., los retrotransposones) que hacía prácticamente imposible detectar rastros genéticos sin el temor a obtener resultados espurios. De un tiempo a esta parte el estudio de estas secuencias está tomando un nuevo ímpetu (en nuestro mismo laboratorio ya se tienen datos sobre secuencias de retrotransposición en diversas especies, algunas de ellas asociadas a secuencias de IGS 45S) debido a la cantidad de información que pueden aportar (se ha visto que son más informativos que los AFLP en algunos estudios; en la revisión de Schulman et al., 2012) y a que se conoce que su activación dentro de los genomas esta mediada por elementos supresores y activadores que intervienen en la respuesta celular a diferentes niveles de estrés. Es decir, en cierta medida están relacionados directamente con la funcionalidad celular. Por último, y casi como columna sobre la que se han vertebrado todos los trabajos de esta tesis, están las



familias de ADN ribosomal (45S y 5S), siendo familias de ADN repetido mixtas, ya que están en tándem y también están dispersas por el genoma (en el caso de que haya más de un locus). Estas familias habían venido siendo uno de los pocos ejemplos de ADN repetido en los que, a pesar de estar en un alto número de copias, se mantenía una homogeneidad de secuencia intragenómica que facilitaba, tanto técnica, analítica como conceptualmente, su estudio. Ya hemos visto que ha tal hecho se le llamó evolución concertada (Arheim 1983), y gracias a los estudios llevados a cabo con las regiones codificantes tanto del 5S como del 45S y con los espaciadores transcritos ITS1 e ITS2 del cistrón 45S, se convirtió en un modelo casi universalmente aceptado. Pero cada vez más, nuevos trabajos están minando los cimientos de este paradigma en las familias ribosomales, y esta tesis es un ejemplo claro de ello. La enorme variabilidad encontrada en los genes de ADN<sub>r</sub> de las especies presentadas en nuestros trabajos ha hecho cambiar nuestra concepción a la hora de afrontar futuros estudios en las familias multigénicas, convencidos de que el conocimiento actual de tales genes debe ser reevaluado. El método científico asegura que las hipótesis puedan ser contrastadas y modificadas si ello fuera necesario. Ya en el tercer trabajo de esta tesis contrastamos una hipótesis planteada por dos equipos de investigación cuyos resultados eran contradictorios. Como resultado evidenciamos la necesidad de replantearse la organización cromosómica de ambas familias de ADN<sub>r</sub> en el reino Plantae, además de ser los primeros en referenciar la inserción del 5S en el IGS 45S en gimnospermas. Futuros trabajos en este campo, y sobre las familias de ADN repetido en general pondrán de manifiesto, aún más, la enorme cantidad de información contenida en sus secuencias que, a diferencia de los genes comunes de copia simple, aparecen multiplicadas de cientos a miles de veces dentro de los genomas eucariotas.

## Bibliografía

**Abad J.P., Carmena M., Baars S., et al. 1992.** Dodeca satellite: a conserved G+C rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. Proceedings of the National Academy of Sciences, USA, 89: 4663–4667.

**Álvarez I., Wendel J.F. 2003.** Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetic and Evolution. 29:417–434.

**Ambrožová K., Mandáková T., Bureš P. et al. 2011.** Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*. Annals of Botany 107: 255–268.

**Arnheim N. 1983.** Concerted evolution of multigene families. In: Nei, M., Koehn, R., editors. Evolution of Genes and Proteins. Sinauer, Sunderland, MA, p. 38–61.

**Baker W. J., Hedderson T. A., Dransfield J. 2000.** Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA rpS16 intron sequence data. Molecular Phylogenetic and Evolution. 14:195–217.

**Batts-Young B., Lodish H.F. 1978.** Triphosphate residues at the 50 ends of rRNA precursor and 5S RNA from *Dictyostelium discoideum*. Proceedings of the National Academy of Sciences. USA 75: 740–744.

**Baum B.R., Bailey L.G. 2004.** The origin of the A genome donor of wheats (*Triticum*: Poaceae) – a perspective based on the sequence variation of the 5S DNA gene units. Genome Research and Crop Evolution. 51:183-196.

**Bena G. 2001.** Molecular phylogeny supports the morphologically based taxonomic transfer of the “medicoid” *Trigonella* species to the genus *Medicago* L. *Plant Systematics and Evolution*. 229: 217–236.

**Bena G., Jubier M.-F., Olivieri I., Lejeune B. 1998.** Ribosomal external and internal transcribed spacers: combined use in the phylogenetic analysis of *Medicago* (Leguminosae). *Journal of Molecular Evolution*. 46: 299–306.

**Benabdelmouna A., Abirached-Darmency M., Darmency H. 2001.** Phylogenetic and genomic relationships in *Setaria italica* and its close relatives based on the molecular diversity and chromosomal organization of 5S and 18S-5.8S-25S rDNA genes. *Theoretical and Applied Genetics*. 103:668-677.

**Bergeron J., Drouin G. 2008.** The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungal species. *Current Genetics*. 54: 123–131.

**Brown D.D., Wensink, P.C., Jordan E. 1972.** A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: The evolution of tandem genes. *Journal of Molecular Biology*. 63:57–73.

**Calderini O., Pupilli F., Paolocci F., Arcioni S. 1997.** A repetitive and species-specific sequence as a tool for detecting the genome contribution in somatic hybrids of the genus *Medicago*. *Theoretical and Applied Genetics*. 95: 734–740.

**Capesius I. 1997.** Analysis of the ribosomal RNA gene repeat from the moss *Funaria hygrometrica*. *Plant Molecular Biology*. 33: 559–564.

**Cloix C., Tutois S., Mathieu O., Cuvillier C., Espagnol M-C., Picard G., Tourmente S. 2000.** Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. *Genome Research*. 10:679-90.

**Cloix C., Tutois S., Yukawa Y., Mathieu O., Cuvillier C., Espagnol M-C., Picard G., Tourmente S. 2002.** Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Research*. 12:132-144.

**Cox A.V., Bennett M.D., Dyer T.A. 1992.** Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*. 83:684-690.

**Contento A., Heslop-Harrison J.S., Schwarzacher T. 2005.** Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenetics and Genome Research*. 109: 34–42.

**Crisp M.D., Appels R., Smith F.M., Keys W.M.S. 1999.** Phylogenetic evaluation of 5S ribosomal RNA gene and spacer in the *Callistachys* group (Fabaceae: Mirbelieae). *Plant Systematic and Evolution*. 218: 33-42.

**Cronn R.C., Zhao X., Paterson A.H., Wendel J.F. 1996.** Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution*. 42:685–705.

**Cross I., Rebordinos L. 2005.** 5S rDNA and U2snRNA are linked in the genome of *Crassostrea angulata* and *Crassostrea gigas* oysters: does the (CT)<sub>n</sub>. (GA)<sub>n</sub> microsatellite stabilize this novel linkage of large tandem arrays? *Genome*. 48:1116-1119.

**Csink A.K., Henikoff S. 1998.** Something from nothing: the evolution and utility of satellite repeats. *Trends in Genetics*. 14: 200–204.

**Dimarco E., Cascone E., Bellavia D., Caradonna F. 2012.** Functional variants of 5S rRNA in the ribosomes of common sea urchin *Paracentrotus lividus*. *Gene*. 508: 21-25.

**Dover G.A. 1982.** Molecular drive: a cohesive mode of species evolution. *Nature*. 299:111–117.

**Dover G.A., Linares A.R., Bowen T., Hancock J.M. 1993.** Detection and quantification of concerted evolution and molecular drive. *Methods in Enzymology*. 224:525–541.

**Dover G. 2002.** Molecular drive. *Trends in Genetics*. 18: 587–589.

**Doveri S., Lee D. 2007.** Development of sensitive crop-specific polymerase chain reaction assays using DNA: applications in food traceability. *Journal in Agricultural and Food Chemistry*. 55:4640-4644.

**Downie S.R., Downie D.S.K., Rogers E.J., Zujewski H.L., Small E. 1998.** Multiple independent losses of the plastid rpoC1 intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Canadian Journal of Botany*. 76: 791–803.

**Drouin G., Moniz de Sá M. 1995.** The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Molecular Biology and Evolution*. 12: 481-493.

**Elder J.F., Turner B.J. 1995.** Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology*. 70:297–320.

**Falstocco E., Passeri V., Marconi G. 2007.** Investigations of 5S rDNA of *Vitis vinifera* L.: sequence analysis and physical mapping. *Genome*: 50:927-938.

**Favia G., Cancrini G., Ricci I., Bazzocchi C., Magi M., Pietrobelli M., Genchi C., Bandi C. 2000.** 5S ribosomal spacer sequences of some filarial parasites: comparative analysis and diagnostic applications. *Molecular and Cellular Probes* 14: 285-290.

**Fernández-Tajes J., Méndez J. 2009.** Two different size classes of 5S rDNA units coexisting in the same tandem array in the razor clam *Ensis macha*: is this region suitable for phylogeographic studies? *Biochemical Genetics*. 47:775-788.

**Figuroa D.M., Bass H.W. 2010.** A historical and modern perspective on plant cytogenetics. *Briefings in Functional Genomics*. 9: 95–102.

**Fry K., Salsler W. 1977.** Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*. 12: 1069–1084.

**Fujiwara M., Inafuku J., Takeda A., Watanabe A., Fujiwara A., Kohno S., Kubota S. 2009.** Molecular organization of 5S rDNA in bitterlings (Cyprinidae). *Genetica*. 135:355–365.

**Galián J.A., Rosato M., Rosselló J.A. 2012.** Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba*. *Heredity*. 108:640-646.

**García S., Lim K.Y., Chester M., Garnatje T., Pellicer J., Vallès J. et al. 2009.** Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Cromosoma*. 118: 85–97.

**García S., Panero J.L., Siroky J., Kovarik A. 2010.** Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC Plant Biology*. 10: 176.

**Golczyk H., Hasterok R., Joachimiak A.J. 2005.** FISH-aimed karyotyping and characterization of Renner complexes in permanent heterozygote *Rhoeo spathacea*. *Genome*. 48: 145–153.

**Hasterok R., Jenkins G., Langdon T., Jones R.N., Maluszynska J. (2001).** Ribosomal DNA is an effective marker of *Brassica* chromosomes. *Theoretical and Applied Genetics*. 103: 486–490.

**Heikkinen E., Launonen V., Muller E., Bachmann L. 1995.** The pVB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. *Journal of Molecular Evolution*. 41: 604–614.

**Hemleben V., Kovarik A., Torres-Ruiz R.A., Volkov R.A., Beridze T. 2007.** Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and Biodiversity*. 5: 277–289.

**Heslop-Harrison J.S. 2000.** Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *The Plant Cell*. 12: 617–635.

**Hori H., Lim B.-L., Osawa S. 1985.** Evolution of green plants as deduced from 5S rRNA sequences. *Proceedings of the National Academy of Sciences, USA* 82:820-823.

**Jorgensen R., Cluster P. 1988.** Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies useful at several taxonomic levels. *Annals of the Missouri Botanical Garden*. 75: 1238–1247.

**Kellogg E.A., Appels R. 1995.** Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics*. 140:325–43.

**King R.C., Stansfield W.D. 1990.** A dictionary of genetics, 4<sup>th</sup> Ed. New York: Oxford University Press.

**Ko H.L., Henry R.J. 1996.** Specific 5S ribosomal RNA primers for plant species identification in admixtures. *Plant Molecular Biology Reporter*. 14:33-43.

**Komarova N.Y., Grimm G.W., Hemleben V., Volkov R.A. 2008.** Molecular evolution of 35S rDNA and taxonomic status of *Lycopersicon* within *Solanum* sect. Petota. *Plant Systematic and Evolution*. 276: 59–71.

**Kovarik A., Dadejova M., Lim Y.K., Chase M.W., Clarkson J.J., Knapp S. et al. 2008.** Evolution of rDNA in *Nicotiana* allopolyploids: A potential link between rDNA homogenization and epigenetics. *Annals of Botany*. 101: 815-823.

**Kubis S., Heslop-Harrison J.S., Desel C. and Schmidt T. 1998.** The genomic organization of non-LTR retrotransposons (LINEs) from three Beta species and five other angiosperms. *Plant Molecular Biology*. 36: 821–831.

**Lafontaine D.L.J., Tollervey D. 2001.** The function and synthesis of ribosomes. *Nature Review Molecular Cell Biology*. 2: 514–520.

**Leitch I.J., Bennett M.D. 2004.** Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*. 82: 651-663.

**Liao D. 1999.** Concerted evolution: molecular mechanisms and biological implications. *The American Journal of Human Genetics*. 64:24-30.

**Liu Z.-L., Zhang D., Wang X.-Q., Ma X.-F., Wang X.-R. 2003.** Intragenomic and interspecific 5S rDNA sequence variation in five Asian pines. *American Journal of Botany*. 90:17-24.



**Long E.O., David I.D. 1980.** Repeated genes in eukaryotes. *Annual Review of Biochemistry*. 49:727-764.

**Macas J., Navrátilová A., Koblížková A. 2006.** Sequence homogenization and chromosomal localization of VicTR-B satellites differ between closely related *Vicia* species. *Chromosoma*. 115: 437–447.

**Macas J., Neumann P., Navrátilová A. 2007.** Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics*. 8: 427.

**Maizels N. 1976.** *Dictyostelium* 17S, 25S, and 5S rDNAs lie within a 38000 base pair repeated unit. *Cell*. 9: 431–438.

**Major R.T. 1967.** The Ginkgo, the most ancient living tree. *Science* 157: 1270–1273.

**Maluszynska J., Heslop-Harrison J.S. 1993.** Molecular cytogenetics of the genus *Arabidopsis* – in situ localization of rDNA sites, chromosome numbers and diversity in centromeric heterochromatin. *Annals of Botany*. 71:479–484

**Maureira-Butler I.J., Pfeil B.E., Muangprom A., Osborn T.C., Doyle J.J. 2008.** The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*. 57: 466–482.

**Mayol M., Rosselló J.A. 2001.** Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution*. 19:167–176.

**Mestrovic N., Plohl M., Mravinac B., Ugarković D. 1998.** Evolution of satellite DNAs from the genus *Palorus* experimental evidence for the ‘library’ hypothesis. *Molecular Biology and Evolution*. 15: 1062–1068.

**Mizuochi H., Marasek A., Okazaki K. 2007.** Molecular cloning of *Tulipa fosteriana* rDNA and subsequent FISH analysis yields cytogenetic organization of 5S rDNA and 45S rDNA in *T. gesneriana* and *T. fosteriana*. *Euphytica*. 155: 235–248.

**Muravenko O.V., Amosova A.V., Samatadze T.E., Semenova O.Y., Nosova I.V., Popov K.V. et al. 2004.** Chromosome localization of 5S and 45S ribosomal DNA in the genomes of *Linum* L. species of the section *Linum* (Syn. *Protolinum* and *Adenolinum*). *Russian Journal of Genetics*. 40: 193–196.

**Murray B.G., Friesen N., Heslop-Harrison J.S. 2002.** Molecular cytogenetic analysis of *Podocarpus* and comparison with other gymnosperm species. *Annals of Botany*. 89: 483–489.

**Nakao Y., Taira T., Horiuchi S., Kawase K., Mukai Y. 2005.** Chromosomal difference between male and female trees of *Ginkgo biloba* examined by karyotype analysis and mapping of rDNA on the chromosomes by fluorescence in situ hybridization. *Journal of the Japanese Society for Horticultural Science*. 74: 275–280.

**Nei M., Rooney A. P. 2005.** Concerted and birth-and-death evolution of multigene families. *Annual Review in Genetics*; 39:121–152.

**Ohta T. 1980.** Evolution and variation of multigene families. Berlin: Springer-Verlag.

**Pires J.C., Lim K.Y., Kovarik A., Matyasek R., Boyd A., Leitch A.R., Leitch I.J., Bennett M.D., Soltis P.S., Soltis D.E. 2004.** Molecular cytogenetic analysis of recently evolved *Tragopogon* (Asteracea) allopolyploids reveal a karyotype that is additive of the diploid progenitors. *American Journal of Botany*. 91:1022–1035.

- Pontes O., Neves N., Silva M., Lewis M.S., Madlung A., Comai L. et al. 2004.** Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proceedings of the National Academy of Sciences USA*. 101: 18240–18245.
- Rogers S.C., Bendich A.J. 1987.** Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Molecular Biology*. 9: 509–520.
- Rooney, A.P. and Ward, T.J. 2005.** Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. *Proceedings of the National Academy of Sciences USA*. 102: 5084–5089.
- Royer D.L., Hickey L.J., Wing S.L. 2003.** Ecological conservatism in the ‘living fossil’ *Ginkgo*. *Paleobiology*. 29: 84–104.
- Saini A., Jawali N. 2009.** Molecular evolution of 5S rDNA region in *Vigna* subgenus *Ceratotropis* and its phylogenetic implications. *Plant Systematic and Evolution*. 280:187-206.
- Sajdak S.L., Reed K.M., Phillips R.B. 1998.** Intraindividual and interspecies variation in the 5S rDNA of coregonid fish. *Journal of Molecular Evolution*. 46 680-688.
- Sanderson M.J., Wojciechowski M.F. 1995.** Molecular phylogenetic analysis of a temperate legume clade (Fabaceae). *American Journal of Botany (Supplement)* 82: 159.
- Sastri D.C., Hilu K., Appels R., Lagudah E.S., Playford J., Baum B.R. 1992.** An overview of evolution in plant 5S DNA. *Plant Systematic Evolution*. 183:169–181.

**Schulman A. H., Flavell A. J., Ellis T. H. N. 2004.** The application of LTR retrotransposons as molecular markers in plants. *Methods in molecular biology (Clifton, N.J.)*, 260, 145–73.

**Šíroky J., Lysák M.A., Doležel J., Kejnovský E., Vyskot B. 2001.** Heterogeneity of rDNA distribution and genome size in *Silene spp.* *Chromosome Research*. 9: 387–393.

**Small E., Jomphe M. 1989.** A synopsis of the genus *Medicago* (Leguminosae). *Canadian Journal of Botany* 67: 3260–3294.

**Snowdon R.J., Friedt W., Koehler A., Koehler W. 2000.** Molecular cytogenetic localization and characterization of 5S and 25S rDNA loci for chromosome identification in oilseed rape (*Brassica napus* L.). *Annals of Botany*. 86: 201–204.

**Sone T., Fujisawa M., Takenaka M., Nakagawa S., Yamaoka S., Sakaida M. et al. 1999.** Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. *Plant Molecular Biology*. 41: 679–685.

**Steele K.P., Ickert-Bond S.M., Zarre S., Wojciechowski M.F. 2010.** Phylogeny and character evolution in *Medicago* (Leguminosae): evidence from analyses of plastid TRNK/MATK and nuclear GA3OX1 sequences. *American Journal of Botany*. 97: 1142–1155.

**Suárez-Santiago V.N., Blanca G., Ruiz-Rejón M., Garrido-Ramos M.A. 2007.** Satellite-DNA evolutionary patterns under a complete evolutionary scenario: the case of *Acrolophus* subgroup (*Centaurea* L., Compositae) from the western Mediterranean. *Gene*. 404: 80–92.

**Sucher N.J., Carles M.C. 2008.** Genome-based approaches to the authentication of medicinal plants. *Planta Medica*. 74:603–623.

- Tralau H. 1968.** Evolutionary trends in the genus *Ginkgo*. *Lethaia*. 1: 63–101.
- Trontin J.-F., Grandemange C., Favre J.-M. 1999.** Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix decidua* Mill.) and Japanese larch (*Larix kaempferi* (Lamb.) Carr.). *Genome*. 42: 837-848.
- Ugarković D, Plohl M. 2002.** Variation in satellite DNA profiles – causes and effects. *EMBO Journal*. 21: 5955–5959.
- Vahidi H., Purac A., Leblanc J.M., Honda B.M. 1991.** Characterization of potentially functional 5S rRNA-encoding genes within ribosomal DNA repeats of the nematode *Meloidogyne arenaria*. *Gene*. 108: 281–284.
- Vitturi R., Colomba M.S., Pirrone A.M., Mandrioll M. 2002.** rDNA (18S–28S and 5S) colocalization and linkage between ribosomal genes and (TTAGGG)(n) telomeric sequence in the earthworm *Octodrilus complanatus* (Annelida, Oligochaeta, Lumbricidae), revealed by single- and double-color FISH. *Journal of Heredity*. 93: 279–282.
- Volkov R.A., Zanke C., Panchuk I.I., Hemleben V. 2001.** Molecular evolution of 5S rDNA of *Solanum* species (sect. *Petota*): application for molecular phylogeny and breeding. *Theoretical in Applied Genetics*. 103:1273-1282.
- Wagner A., Blackstone N., Cartwright P., Dick M., Misof B., Snow P., Wagner G.P., Bartels J., Murtha M., Pendleton J. 1994.** Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology*. 43:250-261.

**Wicke S., Costa A., Muñoz J., Quandt D. 2011.** Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Molecular Phylogenetics and Evolution*. 61: 321–332.

**Wojciechowski M.F., Sanderson M.J., Steele K.P., Liston A. 2000.** Molecular phylogeny of the “Temperate Herbaceous Tribes” of Papilionoid legumes: a supertree approach. In: Herendeen PS, Bruneau A. eds. *Advances in legume systematics*. Kew, UK: Royal Botanic Gardens, 9, 277–298.

**Xia X., Erickson L. 1993.** An AT-rich satellite DNA sequence E180 in alfalfa (*Medicago sativa*). *Genome*. 36: 427–432.

**Zimmer E.A., Martin S.L., Beverley S.M., Kan Y.W., Wilson A.C. 1980.** Rapid duplication and loss of genes coding for the alpha-chains of hemoglobin. *Proceedings of the National Academy of Sciences USA*. 77:2158-21162.

# ARTÍCULO I

## **Amplification, contraction and genomic spread of a satellite DNA family (E180) in *Medicago* (Fabaceae) and allied genera**

**(Annals of Botany 109: 773–782, 2012)**

*Background and Aims* Satellite DNA is a genomic component present in virtually all eukaryotic organisms. The turnover of highly repetitive satellite DNA is an important element in genome organization and evolution in plants. Here we assess the presence and physical distribution of the repetitive DNA E180 family in *Medicago* and allied genera. Our goals were to gain insight into the karyotype evolution of *Medicago* using satellite DNA markers, and to evaluate the taxonomic and phylogenetic signal of a satellite DNA family in a genus hypothesized to have a complex evolutionary history.

*Methods* Seventy accessions from *Medicago*, *Trigonella*, *Melilotus* and *Trifolium* were analysed by PCR to assess the presence of the repetitive E180 family, and fluorescence *in situ* hybridization (FISH) was used for physical mapping in somatic chromosomes.

*Key Results* The E180 repeat unit was PCR-amplified in 37 of 40 taxa in *Medicago*, eight of 12 species of *Trigonella*, six of seven species of *Melilotus* and in two of 11 *Trifolium* species. Examination of the mitotic chromosomes revealed that only 13 *Medicago* and two *Trigonella* species showed FISH signals using the E180 probe. Stronger hybridization signals were observed in subtelomeric and interstitial loci than in the pericentromeric loci, suggesting this satellite family has a preferential genomic location. Not all 13 *Medicago* species that showed FISH localization of the E180 repeat were phylogenetically related. However, nine of these species belong to the phylogenetically derived clade including the *M. sativa* and *M. arborea* complexes.

*Conclusions* The use of the E180 family as a phylogenetic marker in *Medicago* should be viewed with caution. Its amplification appears to have been produced through recurrent and independent evolutionary episodes in both annual and perennial *Medicago* species as well as in basal and derived clades.

<http://aob.oxfordjournals.org/content/109/4/773.full>

# ARTÍCULO II

## **Partial Sequence Homogenization in the 5S Multigene Families May Generate Sequence Chimeras and Spurious Results in Phylogenetic Reconstructions**

**(Systematic Biology 63(2): 219-230, 2014)**

*Abstract.* — Multigene families have provided opportunities for evolutionary biologists to assess molecular evolution processes and phylogenetic reconstructions at deep and shallow systematic levels. However, the use of these markers is not free of technical and analytical challenges. Many evolutionary studies that used the nuclear 5S rDNA gene family rarely used contiguous 5S coding sequences due to the routine use of head-to-tail polymerase chain reaction primers that are anchored to the coding region. Moreover, the 5S coding sequences have been concatenated with independent, adjacent gene units in many studies, creating simulated chimeric genes as the raw data for evolutionary analysis. This practice is based on the tacitly assumed, but rarely tested, hypothesis that strict intra-locus concerted evolution processes are operating in 5S rDNA genes, without any empirical evidence as to whether it holds for the recovered data. The potential pitfalls of analysing the patterns of molecular evolution and reconstructing phylogenies based on these chimeric genes have not been assessed to date. Here, we compared the sequence integrity and phylogenetic behavior of entire versus concatenated 5S coding regions from a real data set obtained from closely related plant species (*Medicago*, Fabaceae). Our results suggest that within arrays sequence homogenization is partially operating in the 5S coding region, which is traditionally assumed to be highly conserved. Consequently, concatenating 5S genes increases haplotype diversity, generating novel chimeric genotypes that most likely do not exist within the genome. In addition, the patterns of gene evolution are distorted, leading to incorrect haplotype relationships in some evolutionary reconstructions.

<http://sysbio.oxfordjournals.org/content/63/2/219.abstract>



# ARTÍCULO III

## **Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba***

**(Heredity 108(6): 640-646, 2012)**

*Abstract.* — In seed plants, the colocalization of the 5S loci within the intergenic spacer (IGS) of the nuclear 45S tandem units is restricted to the phylogenetically derived Asteraceae family. However, fluorescent in situ hybridization (FISH) colocalization of both multigene families has also been observed in other unrelated seed plant lineages. Previous work has identified colocalization of 45S and 5S loci in *Ginkgo biloba* using FISH, but these observations have not been confirmed recently by sequencing a 1.8 kb IGS. In this work, we report the presence of the 45S–5S linkage in *G. biloba*, suggesting that in seed plants the molecular events leading to the restructuring of the ribosomal loci are much older than estimated previously. We obtained a 6.0 kb IGS fragment showing structural features of functional sequences, and a single copy of the 5S gene was inserted in the same direction of transcription as the ribosomal RNA genes. We also obtained a 1.8 kb IGS that was a truncate variant of the 6.0 kb IGS lacking the 5S gene. Several lines of evidence strongly suggest that the 1.8 kb variants are pseudogenes that are present exclusively on the satellite chromosomes bearing the 45S–5S genes. The presence of ribosomal IGS pseudogenes best reconciles contradictory results concerning the presence or absence of the 45S–5S linkage in *Ginkgo*. Our finding that both ribosomal gene families have been unified to a single 45S–5S unit in *Ginkgo* indicates that an accurate reassessment of the organization of rDNA genes in basal seed plants is necessary.

<http://www.nature.com/hdy/journal/v108/n6/full/hdy20122a.html>