

UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE GENÉTICA



**EVOLUCIÓN DEL ADN REPETIDO
EN EL GENOMA VEGETAL**

TESIS DOCTORAL

MARÍA ESTER QUESADA DEL BOSQUE
Granada, 2013

Editor: Editorial de la Universidad de Granada
Autor: María Ester Quesada del Bosque
D.L.: GR 385-2014
ISBN: 978-84-9028-782-8

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Departamento de Genética



TESIS DOCTORAL

**EVOLUCIÓN DEL ADN REPETIDO
EN EL GENOMA VEGETAL**

MARÍA ESTER QUESADA DEL BOSQUE

Granada, julio de 2013

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Departamento de Genética

**EVOLUCIÓN DEL ADN REPETIDO
EN EL GENOMA VEGETAL**

Memoria para optar al grado de Doctor
en Biología presentada por
María Ester Quesada del Bosque

DIRECTORES:

Dr. Manuel A. Garrido Ramos

Dra. Inmaculada López Flores

Granada, julio de 2013

Parte de los resultados que se presentan en esta Memoria han sido presentados en los siguientes congresos:

- **Quesada del Bosque ME, Suárez-Santiago VN, Garrido-Ramos MA.** Evolución del ADN satélite en la subfamilia Carduoideae (Compositae).

XXXVII Congreso de la Sociedad Española de Genética (SEG).
Torremolinos (Málaga). Septiembre, 2009

- **Quesada del Bosque ME, Navajas-Pérez R, Panero JL, Fernández-González A, Garrido-Ramos MA.** Evolución del ADN satélite en la planta dioica *Rumex hastatulus* (Polygonaceae).

XXXVII Congreso de la Sociedad Española de Genética (SEG).
Torremolinos (Málaga). Septiembre, 2009.

Asimismo, parte de los resultados de este trabajo se recogen en los siguientes artículos:

- **Navajas-Pérez R, Quesada del Bosque ME, Garrido-Ramos MA.** Effect of location, organization, and repeat-copy number in satellite-DNA evolution. *Mol Genet Genomics*, 2009; 282(4): 395-406. doi: 10.1007/s00438-009-0472-4.

- **del Bosque ME, Navajas-Pérez R, Panero JL, Fernández-González A, Garrido-Ramos MA.** A satellite DNA evolutionary analysis in the North American endemic dioecious plant *Rumex hastatulus* (Polygonaceae). *Genome*, 2011; 54(4): 253-60. doi: 10.1139/g10-115.

- **Quesada del Bosque ME, López-Flores I, Suárez-Santiago VN, Garrido-Ramos MA.** Differential spreading of HinfI satellite-DNA variants during radiation in Centaureinae. *Annals of Botany*, 2013 (revision enviada).

- **Quesada del Bosque ME et al.** Analysis of HinfI satellite DNA variants in the tribe Cardueae, en preparación.

La financiación del presente trabajo ha corrido a cargo del siguiente proyecto de investigación:

- Evolución de los cromosomas sexuales y de los mecanismos de determinación sexual en plantas: el género *Rumex* (Polygonaceae) como modelo de estudio. Ministerio de Educación y Ciencia (CGL2006-00444). 2006-2009.

AGRADECIMIENTOS

En el devenir de la vida se suceden muchos acontecimientos, situaciones y personas que nos van dejando huella como ha sido la experiencia vivida en el Departamento de Genética. La llegada a este departamento tenía un objetivo, realizar una tesis doctoral. Como toda meta no ha estado exenta de momentos difíciles, insatisfacción, inseguridad, incertidumbre que, finalmente, fueron sustituidos por satisfacción, superación y motivación. Aunque el final parecía lejano se ha alcanzado la meta. Durante este largo camino de subidas y bajadas se han cruzado muchas personas a las que les estoy enormemente agradecida por lo que me han aportado a nivel profesional y personal.

En primer lugar, quiero expresar mi gratitud al Dr. Manuel Ángel Garrido Ramos y a la Dra. Inmaculada López Flores por brindarme la oportunidad de realizar con ellos este proyecto científico. Gracias por haber dedicado parte de vuestro tiempo y esfuerzo durante estos años y haberme permitido alcanzar mi meta.

Mi andadura investigadora comenzó en la Estación Experimental del Zaidín, donde por circunstancias de la vida, tuve la gran suerte de poder acceder a sus laboratorios y empezar a conocer de cerca la profesión del científico. Mi más sincero agradecimiento al Dr. Juan de Dios Alché Ramirez, a la Dra. Mariam Sahrawy Barragán y a todos los miembros de sus grupos de investigación por abrir las puertas de sus laboratorios a una novel investigadora. Allí, tanto en la casa blanca como amarilla comenzaron mis primeras inquietudes y curiosidades en el ámbito de la investigación. Además, allí también conocí las alegrías y los sinsabores de esta profesión y, que con constancia, esfuerzo, ilusión y perseverancia se podían alcanzar muchas logros que, en un principio, piensas que no vas a ser capaz. Sin embargo, tras estos años de doctorado, te das cuenta de que nada es imposible, que esa palabra no debería seguir existiendo en mi vocabulario. Porque todo es susceptible de ser alcanzado por muy lejos que parezca. Gracias por todo lo que me enseñasteis y las destrezas que adquirí día tras día entre pipetas y cubetas.

Gracias al Dr. Víctor Suárez del Departamento de Botánica por su ayuda para conseguir parte del material analizado. También agradezco al Herbario de la Universidad de Granada permitirme acceder a sus colecciones de pliegos, y el trato agradable y cordial del personal en todo momento.

Gracias a todos los profesores del Departamento de Genética por su acogida cuando llegué por primera vez a este departamento, por sus buenos consejos tanto científicos como personales y por haberme arrancado una gran sonrisa cuando más lo necesitaba, haciéndome la rutina del laboratorio mucho más entrañable y entretenida. Y, como no, por la inestimable ayuda de algunos de ellos en la puesta a punto de algunas técnicas de laboratorio.

Durante estos años viviendo entre pipetas, electroforesis, PCRs y extracciones de ADN tuve la grandísima oportunidad de poder trabajar con el grupo de la Dra. Asli Silahtaroglu del “Wilhelm Johannsen Centre for Functional Genome Research, Faculty of Health Sciences, Department of Cellular and Molecular Medicine” de Copenhague (Dinamarca). Mi más sincero agradecimiento por haberme acogido con tanto cariño sintiéndome como si estuviera en mi propia casa. Gracias por todo lo que aprendí con aquel equipo de profesionales daneses. Me gustaría también agradecer enormemente a mis compañeros de Østerbro Kollegiet por aquellos momentos tan maravillosos fuera del trabajo, de ricas cenas y de muchas risas y buenos ratos.

Pero durante todos estos años, estos logros científicos han ido de la mano de grandes personas que me han acompañado sin dudarlos en ningún momento. Muchísimas gracias a los becarios, antiguos alumnos internos, y técnicos del Departamento de Genética por su apoyo incondicional, por todos los buenos y grandes ratos que hemos pasado en el trabajo y fuera de él. Especialmente, quiero expresar mi enorme gratitud a mis vecinas de laboratorio, a “las niñas”. Gracias por la dosis de energía que me dabais todos los días en el desayuno y esos buenos ratos de charla en el laboratorio. Sin vosotros estoy segura de que este proyecto científico no lo habría desarrollado con tanta motivación.

Además, este proyecto no podría haberlo desarrollado sin el apoyo continuado e incondicional de la gente que siempre ha confiado y ha creído en mí sin dudarlos en ningún momento. Los que siempre han creído en que era capaz de esto y mucho más, los que más me quieren y siempre me han apoyado. Infinitas gracias a mi familia, amigos y a todos los que me estiman y aprecian. Muchísimas gracias a mis padres por traerme a este mundo para vivir momentos de alegría, satisfacción y superación como este. Gracias por haberme dado tanto, por haberme inculcado una educación y a intentar aprobar con cum laude la tesis doctoral más difícil que cualquier persona puede realizar que es entender la vida. Y que por muy bueno que seas en tu trabajo, por muchos reconocimientos que recibas nunca debes olvidar que lo más importante es intentar ser buena persona. No puedo olvidar a la mejor compañera de hazañas que se puede tener, a mi hermana Elenita, parte de esta tesis es tuya. Mi más profundo agradecimiento por todo tu inconmesurable apoyo, por todos los momentos vividos y por estar siempre a mi lado. Muchísimas gracias también a mis otros hermanos y al resto de los miembros de mi familia, en especial a mis tíos Elena y Ramón porque gracias a vosotros me he podido evadir de las largas jornadas de laboratorio y las larguísimas horas frente al ordenador. Llegado el final de estas líneas, y no por ello menos importante, justo todo lo contrario, gracias a las nuevas incorporaciones a la familia, a los recién llegados y a los que les quedan meses por llegar por impregnarme de alegría, ilusión y de tanta ternura.

MUCHÍSIMAS GRACIAS A TODOS

A mis padres

A mis hermanos

ABREVIATURAS

A	Adenina
ADN	Ácido desoxirribonucleico
ARNr	ARN ribosómico
ARN	Ácido ribonucleico
ARNasa A	Ribonucleasa A
ARNt	ARN transferente
BLAST	<i>Basic Local Alingment Search Tool</i>
°C	Grados centígrados
dNTPs	Desoxi-nucleótidos-trifosfato
EMBL	<i>European Molecular Biology Laboratory</i>
ETS	Espaciador transcrito externo (<i>External-Transcribed Spacer</i>)
EDTA	Ácido etilen diamino tetraacético
g	Gramo
H ₂ O _{dd}	Agua ultrapura
HOR	Repeticiones de orden superior (<i>Higher-Order Repeats</i>)
ITS	Espaciador transcrito interno (<i>Intergenic Transcribed Spacer</i>)
kb	Kilo bases
LB	Medio de cultivo Luria Bertani
LINEs	Elementos dispersos largos (<i>Long Interspersed Element</i>)
LTR	Repeticiones terminales largas (<i>Long Terminal Repeats</i>)
M	Moles/litro
matK	Gen codificador para una maturasa localizado en un intrón del gen cloroplastidial trnK
MITE	Elemento transponible con secuencias invertidas repetidas en miniatura (<i>Miniature Inverted-repeat Transposable Element</i>)
mg	Miligramo (10 ⁻³ gramos)
ml	Mililitro
ML	Máxima verosimilitud (<i>Maximum Likelihood</i>)
mM	MiliMolar
NCBI	National Center for Biotechnology Information
ndhF	Gen codificador para la subunidad F de la NADP dehidrogenasa

ng	Nanogramos (10^{-9} gramos)
NJ	<i>Neighbor joining</i>
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa (<i>Polimerase Chain Reaction</i>)
pmol	pico moles
qPCR	PCR cuantitativa
rbcL	Gen codificador de la subunidad mayor de la ribulosa-1,5-bis-fosfato carboxilasa/oxigenasa
π	Diversidad nucleotídica
rpm	revoluciones por minuto
seg	Segundo
siARN	ARN de interferencia sintéticos
SINE	Elementos dispersos cortos (<i>Short Interspersed Element</i>)
SOC	Medio de cultivo (<i>Super Optimal Broth with Catabolic represor</i>)
T	Timina
<i>Taq</i>	Polimerasa de ADN de <i>Termus aquaticus</i>
TE	Elemento transponible
TBE	Tris-borato-EDTA
Tris	2-amino-2-hidroximetil-1,3-propanodiol
TrnL/F	Gen codificador para el ARN transferente de la Leucina/
U	Unidad de enzima
μg	Microgramo (10^{-6} gramos)
μl	Microlitro (10^{-6} litros)
μM	Micromolar

RESUMEN

Este trabajo pretende analizar en profundidad la evolución del ADN satélite en el genoma de plantas, para lo cual hemos estudiado dos sistemas que permiten estudiar la influencia de factores tanto intrínsecos (ubicación, organización y número de copias) como extrínsecos (evolución del grupo biológico) en el proceso de *molecular drive* que conduce al patrón de evolución concertada.

Uno de estos sistemas biológicos lo componen las especies dioicas del género *Rumex* (familia *Polygonaceae*), en las que la familia de ADN satélite RAE180 se ha acumulado diferencialmente y muestra un patrón de distribución distinto en las diferentes especies. Mientras en las especies XX/XY el ADN satélite RAE180 se localiza en un único locus autosómico, en las especies XX/XY₁Y₂ además de estar presente en un locus autosómico pequeño se ha amplificado masivamente en los cromosomas Y. En este trabajo hemos encontrado que en las especies XX/XY, la comparación entre secuencias RAE180 de especies relacionadas muestra un patrón general de evolución concertada. Por el contrario, en las especies XX/XY₁Y₂ tanto en el locus autosómico como en el ligado al cromosoma Y, la variabilidad ancestral se ha mantenido, y se observan tasas bajas de homogeneización y de evolución concertada de las secuencias. Así, este estudio demuestra que los mecanismos moleculares de intercambio no recíproco son factores claves en el proceso de *molecular drive*; las familias de ADN satélite localizadas en la región no recombinante de los cromosomas Y muestran tasas bajas de evolución concertada, de modo que la variabilidad intraespecífica aumenta sin que se produzca al mismo tiempo un aumento de la divergencia entre especies. Esto contrasta con lo que ocurre en los loci que se recombinan libremente, que evolucionan de manera concertada y muestran diferenciación genética entre especies, como es el caso del locus autosómico de las especies XX/XY. No obstante, puede ocurrir alternancia entre periodos evolutivos de cambio rápido en las secuencias, y periodos evolutivos de estancamiento en los que la variabilidad se mantiene debido a la acción reducida de los mecanismos moleculares de intercambio no recíproco, como ha ocurrido en el locus autosómico de las especies XX/XY₁Y₂, lo cual podría depender a su vez del número de copias y de los procesos implicados en la amplificación de estas secuencias.

Entre las especies de este grupo, la especie norteamericana *Rumex hastatulus* muestra dos razas cromosómicas, una con un sistema de cromosomas sexuales XX/XY (raza de Texas), y otra con un sistema cromosómico sexual complejo XX/XY₁Y₂ (raza de Carolina del Norte). Los cromosomas sexuales en los individuos de la raza de Carolina del Norte son aún eucromáticos, y no

han acumulado secuencias de ADN satélite, lo que contrasta con el resto de especies dioicas del género *Rumex* con cromosomas XX/XY₁Y₂. En *R. hastatulus*, hemos detectado la existencia de tres subfamilias RAE180. No obstante, mientras que en la raza de Texas la subfamilia TX1/NC1 es la más abundante, en la raza de Carolina del Norte la subfamilia más abundante es la TX2/NC2. El análisis filogenético de las secuencias RAE180 de *R. hastatulus* más las de otras cinco especies dioicas del género *Rumex* apoya la hipótesis de la biblioteca, según la cual asumimos que en el genoma ancestral de las especies dioicas de *Rumex* se encontraban presentes un conjunto de variantes RAE180 a partir del cual cada una de ellas se amplificó diferencialmente en diferentes linajes. La amplificación diferencial del ADN satélite RAE180 en cada linaje, en diferentes momentos evolutivos, y en diferentes posiciones cromosómicas, dio lugar a distintos patrones evolutivos de sus secuencias.

El otro sistema que hemos estudiado lo constituyen las especies pertenecientes a la tribu Cardueae (familia Asteraceae). Dentro de la tribu Cardueae, la subtribu Centaureinae es un grupo modelo excelente para analizar el ADN satélite y evaluar la influencia que la biología y/o la evolución de los diferentes linajes han tenido en la evolución del ADN satélite. Los análisis filogenéticos de Centaureinae reflejan dos fases de radiación, que han dado lugar a dos grupos de géneros. Además, en los diferentes linajes se observan modos de evolución distintos, que se ven reflejados tanto en la morfología de los individuos como en las secuencias de ADN. Nosotros hemos encontrado hasta ocho subfamilias diferentes de la familia de ADN satélite Hinfl, identificadas por la presencia de posiciones diagnósticas compartidas por todas las secuencias pertenecientes a un mismo grupo. Las subfamilias V a VIII se encuentran predominantemente en los géneros más antiguos (primera fase de radiación dentro de la subtribu, Oligoceno-Mioceno tardíos), aunque algunas copias de estas secuencias se encuentran también presentes en algunas especies de los géneros derivados. Las subfamilias I a IV se han amplificado mayoritariamente en las especies del clado derivado (segunda fase de radiación, Plioceno a Pleistoceno), aunque las repeticiones de estas subfamilias están presentes también en las especies más antiguas. Los árboles filogenéticos no agrupan las repeticiones por afinidad taxonómica, sino que las especies se agrupan dependiendo de la subfamilia a la que pertenecen. Se observa evolución concertada en las subfamilias Hinfl presentes en los géneros más antiguos, mientras que no se aprecia diferenciación genética ni entre los grupos de radiación reciente, ni entre los grupos con una historia de hibridación recurrente, llegando incluso a coexistir varias subfamilias dentro de una misma especie. Nuestros resultados sugieren que las ocho subfamilias debieron estar presentes en el ancestro común de todas las especies de Centaureinae, y que cada una de ellas se amplificó diferencialmente en los distintos géneros durante las dos fases de radiación según el modelo evolutivo de la biblioteca. Además, diferentes vías de especiación han dado lugar a diferentes patrones de evolución de las secuencias en distintos

linajes. Así, la historia evolutiva de cada grupo de Centaureinae se ve reflejada en la evolución del ADN satélite. Nuestros datos refuerzan el valor de las secuencias de ADN satélite como marcadores de los procesos evolutivos que ocurren en los grupos biológicos.

La clasificación de la tribu Cardueae es complicada, ya que existe controversia en la delimitación de la tribu, en la delimitación de subtribus y en la delimitación de los géneros. La clasificación actual reconoce a las subtribus Carlininae, Cardopatiinae y Echinopsinae como subtribus basales mientras que las subtribus Carduinae y Centaureinae, a las que pertenecen más del 90% de las especies de la tribu, serían derivadas. Esta clasificación no resulta totalmente satisfactoria teniendo en cuenta la probable parafilia de la subtribu Carduinae. En este trabajo hemos analizado si la familia de ADN satélite *HinfI* se encuentra conservada en la tribu Cardueae, y si el análisis de su evolución podría arrojar luz sobre la delimitación de las subtribus y de los géneros de esta tribu y su actual clasificación. Analizando toda la tribu, además de las ocho subfamilias mencionadas más arriba, hemos encontrado una adicional, sumando un total de nueve subfamilias diferencialmente distribuidas en diferentes especies de cada subtribu. Los árboles filogenéticos agrupan a las secuencias según la subfamilia a la que pertenecen. No obstante, cuando se comparan secuencias *HinfI* del mismo tipo, las distancias genéticas y los árboles filogenéticos muestran diferenciación entre los géneros y entre las subtribus en todos los casos excepto cuando se comparan secuencias *HinfI* de algunas especies de Centaureinae entre sí, y cuando se comparan secuencias de algunas especies de la subtribu Carduinae con especies de la subtribu Centaureinae. Con estos datos, podemos concluir que la familia de ADN satélite *HinfI* se encuentra conservada en el genoma de las especies de la tribu Cardueae, la que cuenta con mayor número de especies dentro de la familia Asteraceae, y que data de hace 29-24 m.a. En el ancestral común de todas las especies de Cardueae existiría una biblioteca de nueve subfamilias que se habrían amplificado diferencialmente en los distintos linajes con el consiguiente reemplazamiento de una variante por otra en las distintas especies. Este proceso conllevaría posteriormente la divergencia gradual entre repeticiones de una subfamilia compartida por un grupo de especies, lo que conduciría a la evolución concertada a una tasa que sería variable entre los diferentes linajes. Estas diferentes tasas de evolución serían las responsables de la escasa diferenciación entre especies del mismo género en algunas comparaciones. Desde un punto de vista filogenético, estos resultados apoyan la actual clasificación en subtribus y la delimitación de los géneros de Cardueae, y apunta a la parafilia de la subtribu Carduinae.

ÍNDICE

I. INTRODUCCIÓN GENERAL	
I.1. ADN repetido en el genoma eucariota	3
I.1.1. ADN satélite	4
I.1.1.1. Evolución del ADN satélite.....	8
II. OBJETIVOS	15
III. MATERIAL Y MÉTODOS	
III.1. Material empleado	19
III.2. Extracción de ADN genómico	21
III.3. Aislamiento de ADN satélite	22
III.4. Clonación de los productos de PCR	23
III.5. Purificación de ADN plasmídico	25
III.6. Secuenciación	26
III.7. Análisis de las secuencias	27
III.8. Cuantificación del satélite RAE180 mediante PCR cuantitativa (qPCR)	29
IV. ANÁLISIS DEL ADN SATÉLITE RAE180 EN EL GÉNERO <i>Rumex</i>	
IV.1. Introducción	33
IV.1.1. Sistemas de determinación del sexo en plantas.....	33
IV.1.2. El género <i>Rumex</i>	34
IV.1.3. Familias de ADN satélite en el género <i>Rumex</i> L.	37
IV.2. Resultados y Discusión	39
IV.2.1. Identificación, aislamiento y caracterización de la familia de ADN satélite RAE180 en el género <i>Rumex</i>	39
IV.2.2. Análisis filogenético de las secuencias RAE180	39
IV.2.3. Análisis de la evolución de las secuencias RAE180 de acuerdo a su localización cromosómica.....	41
IV.2.4. Caracterización de las secuencias RAE180 y diferenciación genética entre las dos razas cromosómicas de <i>Rumex hastatulus</i>	49

V. ANÁLISIS DEL ADN SATÉLITE <i>Hinfl</i> EN LA SUBFAMILIA CARDUOIDEAE (ASTERACEAE)	
V.1. Introducción	59
V. 1.1. Familia Asteraceae	59
V.1.2. Tribu Cardueae Cass	61
V.1.3. Subtribu Centaureinae	62
V.1.4. Géneros estudiados de la subtribu Centaureinae	64
V.1.5. Subtribu Carduinae	70
V.1.6. Subtribu Echinopsinae	72
V.1.7. Subtribu Cardopatiinae	73
V.1.8. Subtribu Carlininae	74
V.2. Resultados	75
V.2.1. Subtribu Centaureinae	75
V.2.2. Subtribus Carlininae, Echinopsinae y Carduinae	83
V.3. Discusión	90
V.3.1. Subtribu Centaureinae	90
V.3.2. Subtribus Carlininae, Echinopsinae y Carduinae	96
VI. CONCLUSIONES	103
VII. BIBLIOGRAFÍA	109
VIII. ANEXOS	
ANEXO I	123
ANEXO II	125
ANEXO III	127
ANEXO IV	131
ANEXO V	137
IX. PUBLICACIONES	
Navajas-Pérez <i>et al.</i> , 2009. Mol Genet Genom 282: 395-406.	143
Quesada del Bosque <i>et al.</i> , 2011. Genome 54: 253-260	145
Quesada del Bosque <i>et al.</i> , 2013	147
Quesada del Bosque <i>et al.</i> , 2014	149

ÍNDICE DE FIGURAS

Figura 1.	Filogenia molecular del género <i>Rumex</i> basada en secuencias nucleares y cloroplastidiales.....	35
Figura 2.	Filogenia molecular de las especies dioicas del género <i>Rumex</i> basada en la familia de ADN satélite RAE180	40
Figura 3.	Árbol de secuencias RAE180 (neighbor-joining) de las especies <i>R. suffruticosus</i> y <i>R. acetosella</i>	43
Figura 4.	Árbol de secuencias RAE180 (neighbor-joining) de las especies <i>R. acetosa</i> y <i>R. papillaris</i>	43
Figura 5.	Alineamiento múltiple de secuencias representativas RAE180 de las dos razas de <i>Rumex hastatulus</i>	51
Figura 6.	Árbol de secuencias de las subfamilias RAE180 en la raza de Texas (TX) de <i>R. hastatulus</i>	52
Figura 7.	Árbol de secuencias de las subfamilias RAE180 en la raza de Carolina del Norte (NC) de <i>Rumex hastatulus</i>	52
Figura 8.	Árbol de secuencias de las subfamilias RAE180 en las razas de Texas y Carolina del Norte de <i>R. hastatulus</i>	54
Figura 9.	Secuencia consenso de las tres subfamilias RAE180 en las razas de Texas y Carolina del Norte de <i>R. hastatulus</i>	55
Figura 10.	Relaciones filogenéticas entre especies dioicas del género <i>Rumex</i>	56
Figura 11.	Árbol filogenético de la familia Asteraceae construido mediante inferencia bayesiana empleando 10 loci cloroplastidiales como marcadores.....	60
Figura 12.	Reconstrucción tentativa de la filogenia de los clados basal y derivado que incluyen a los principales géneros de la subtribu Centaureinae	67

Figura 13. Representación esquemática del árbol filogenético del género <i>Cheirolophus</i>	68
Figura 14. Filogenia del género <i>Rhaponticum</i> y <i>Leuzea</i>	69
Figura 15. Filogenia de la tribu Cardueae basada en las secuencias e ITS, trnL-trnF, and matK.....	72
Figura 16. Alineamiento de las secuencias consenso de las distintas subfamilias del ADN satélite Hinfl en los diferentes géneros de la subtribu Centaureinae	77
Figura 17. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias Hinfl de la Subtribu Centaureinae	80
Figura 18. Secuencias Hinfl pertenecientes a la subfamilia III de especies de los géneros <i>Carthamus</i> y <i>Phonus</i>	81
Figura 19. Secuencias Hinfl pertenecientes a la subfamilia V de especies de los géneros <i>Rhaponticum</i> y <i>Klasea</i>	82
Figura 20. Secuencias Hinfl pertenecientes a la subfamilia VI de especies del género <i>Volutaria</i>	83
Figura 21. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias Hinfl de las subtribus Carlininae y Centaureinae	84
Figura 22. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias Hinfl de las subtribus Echinopsinae y Centaureinae	87
Figura 23. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias Hinfl de las subtribus Carduinae y Centaureinae	88
Figura 24. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias Hinfl de las subtribus Carlininae, Echinopsinae, Carduinae y Centaureinae de la tribu Cardueae.....	89

ÍNDICE DE TABLAS

Tabla 1	Especies analizadas del género <i>Rumex</i>	19
Tabla 2	Especies analizadas de la subfamilia Carduoideae	20
Tabla 3	ADN satélite RAE180 en el genoma de cuatro especies dioicas del género <i>Rumex</i>	42
Tabla 4	Análisis comparativo interespecífico del ADN satélite RAE180	43
Tabla 5	Análisis de la variación entre cada subfamilia RAE180 y de la divergencia entre subfamilias en la raza de Texas de <i>R. hastatulus</i>	52
Tabla 6	Análisis de la variación entre cada subfamilia RAE180 y de la divergencia entre subfamilias en la raza de Carolina del Norte de <i>R. hastatulus</i>	52
Tabla 7	Distancias genéticas entre las razas de Texas y Carolina del Norte de <i>R. hastatulus</i>	53
Tabla 8	Distribución por géneros de las subfamilias HinfI en la subtribu Centaureinae	78
Tabla 9	Distribución por géneros de las subfamilias HinfI en la subtribus Carlininae, Echinopsinae y Carduinae	97
Tabla 10	Variabilidad y divergencia medias de las secuencias <i>HinfI</i> en las especies de <i>Centaurea</i> analizadas	123
Tabla 11	Variabilidad y divergencia medias entre los grupos de secuencias <i>HinfI</i> en las especies de <i>Carduncellus</i> analizadas	123
Tabla 12	Variabilidad y divergencia medias de las secuencias <i>HinfI</i> en las especies del grupo <i>Carthamus-Phonus</i> analizadas	123
Tabla 13	Variabilidad y divergencia medias las secuencias <i>HinfI</i> en las especies de <i>Crupina</i> analizadas	123
Tabla 14	Variabilidad y divergencia medias de las secuencias <i>HinfI</i> en las especies de <i>Rhaponticoides</i> analizadas	124
Tabla 15	Variabilidad y divergencia medias las secuencias <i>HinfI</i> en las especies de <i>Klasea</i> analizadas	124
Tabla 16	Variabilidad y divergencia medias las secuencias <i>HinfI</i> en las especies de <i>Rhaponticum</i> analizadas	124
Tabla 17	Variabilidad y divergencia medias las secuencias <i>HinfI</i> en las especies de <i>Cheirolophus</i> analizadas	124

Tabla 18	Variabilidad y divergencia medias las secuencias HinfI en las especies de <i>Volutaria</i> analizadas	124
Tabla 19	Variabilidad y divergencia medias entre las tres subfamilias HinfI en las especies de <i>Centaurea</i> analizadas.....	125
Tabla 20	Variabilidad y divergencia medias entre las subfamilias HinfI en las especies de <i>Carduncellus</i> analizadas	125
Tabla 21	Variabilidad y divergencia medias entre los grupos de secuencias HinfI en las especies de <i>Crupina</i> analizadas.....	125
Tabla 22	Variabilidad y divergencia medias entre las subfamilias I, II y III HinfI en las especies de <i>Rhaponticoides</i> analizadas	125
Tabla 23	Variabilidad y divergencia medias entre los grupos de secuencias HinfI en las especies del género <i>Rhaponticum</i> analizadas.....	125
Tabla 24	Variabilidad y divergencia medias entre los grupos de secuencias HinfI en las especies de <i>Cheirolophus</i> analizadas	126
Tabla 25	Variabilidad y divergencia medias entre las subfamilias HinfI en las especies de <i>Volutaria</i> analizadas	126
Tabla 26	Variabilidad y divergencia medias entre las secuencias de la subfamilia II de HinfI entre los géneros de las subtribus Centaureinae, Carduinae, Carlininae y Echinopinae analizados.....	127
Tabla 27	Variabilidad y divergencia medias de las secuencias de la subfamilia III de HinfI entre los géneros de las subtribus Centaureinae y Carduinae analizados	127
Tabla 28	Variabilidad y divergencia medias de las secuencias de la subfamilia IV de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados.....	128
Tabla 29	Variabilidad y divergencia medias de las secuencias de la subfamilia V de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados.....	128
Tabla 30	Variabilidad y divergencia medias de las secuencias de la subfamilia VII de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados.....	128
Tabla 31	Variabilidad y divergencia medias de las secuencias de la subfamilia VIII de HinfI entre los géneros de las subtribus Centaureinae y Carduinae analizados	128
Tabla 32	Variabilidad y divergencia medias de las secuencias de la subfamilia IX de HinfI entre los géneros de la subtribu Carduinae analizados.....	128

I.1. ADN repetido en el genoma eucariota

El genoma de las especies eucariotas está compuesto por secuencias de ADN de copia única y por una importante cantidad de secuencias repetidas diferentes, que pueden clasificarse como secuencias moderada o altamente repetidas según su grado de repetitividad (Britten y Kohne, 1968). Además, las secuencias de ADN repetido pueden definirse de acuerdo a otros criterios como su organización en el genoma (distribución dispersa o en tándem), o su función (López-Flores y Garrido-Ramos, 2012). Atendiendo a este último criterio, aunque las secuencias de ADN repetido engloban varios tipos de secuencias que se transcriben y codifican para ARNs (como los genes que codifican los distintos ARNs ribosómicos) o proteínas (genes de la actina o tubulina, por ejemplo), la mayoría del ADN repetido del genoma ha sido considerado tradicionalmente como "ADN basura" sin función conocida, como ocurre con parte del ADN satélite (López-Flores y Garrido-Ramos, 2012).

Las secuencias de ADN que se localizan de modo disperso pueden tener un grado de repetitividad moderado como sucede con las secuencias SINEs (*short interspersed elements* o elementos dispersos cortos), retrogenes, retropseudogenes y diversas familias de genes formadas por copias dispersas. Además, existen secuencias altamente repetidas entre las que destacan los elementos transponibles (TEs) como los transposones de ADN y retrotransposones (principalmente los retrotransposones con repeticiones terminales largas (LTR) y retrotransposones no virales o LINEs (*long interspersed elements* o elementos dispersos largos) (López-Flores y Garrido-Ramos, 2012).

Entre las secuencias de ADN que se disponen en tándem, se encuentran secuencias de ADN moderadamente repetidas, como por ejemplo los genes de las tubulinas, los genes que codifican a ARNs ribosómicos, varias familias de genes que codifican proteínas, secuencias teloméricas y minisatélites. Además, las secuencias de ADN con este modo de organización pueden presentar un elevado grado de repetitividad en el genoma, miles e incluso millones de veces repetidas a lo largo del genoma, como las secuencias de los microsátélites y el ADN satélite, incluido el ADN centromérico (López-Flores y Garrido-Ramos, 2012).

Esta Tesis Doctoral se centra en el estudio del ADN satélite y de su evolución en vegetales. Es por ello que, en las siguientes secciones de esta Introducción, nos vamos a centrar en este tipo de secuencias repetidas en tándem.

I.1.1. ADN satélite

El ADN satélite está compuesto por secuencias de ADN altamente repetido organizadas en tándem, siendo el principal constituyente de la heterocromatina constitutiva. Se localiza mayoritariamente en las regiones pericentroméricas y subteloméricas de los cromosomas eucariotas, aunque también puede encontrarse a nivel intercalar en autosomas o constituyendo parte de los cromosomas sexuales (Plohl *et al.*, 2012). Los monómeros (unidades de repetición) se diferencian en el tamaño, pudiendo estar compuestos por docenas de nucleótidos (en ocasiones son de longitud tan pequeña como la de los mini o microsátélites) o llegar a tener hasta varios cientos de pares de bases. En el genoma de una especie pueden estar presentes diferentes ADNs satélites, y especies relacionadas pueden compartir una colección o biblioteca de secuencias de ADN satélite de diferentes familias. Los ADNs satélite de animales y plantas se diferencian no solo en la longitud de los monómeros, sino también en la secuencia nucleotídica, número de copias en el genoma, complejidad en la secuencia y localización cromosómica.

La longitud habitual de los monómeros tanto en especies animales como vegetales varía entre 150-180 pb y 300-360 pb, la longitud de una doble hélice de ADN necesaria para el ensamblaje de mono o dinucleosomas (Sharma y Raina, 2005; Plohl *et al.*, 2008). No obstante, existen excepciones a esta regla general, como por ejemplo el ADN satélite III humano formado por monómeros de 5 pb o muchos de los ADNs satélites identificados en *Drosophila* (Ugarković, 2005), el ADN satélite de la mosca doméstica con monómeros de 24 pb y el de *Scilla siberica* con monómeros de 35 pb. En el otro extremo, el ADN satélite de la hormiga *Monomorium subopacum* está formado por monómeros de 2,5 kb, o el del centeno con unidades de repetición de 3,9 kb de longitud (Langdon *et al.*, 2000; Sharma y Raina, 2005; Palomeque y Lorite, 2008). Además, dentro de una misma especie, las unidades de repetición no son exactamente idénticas sino que presentan polimorfismos en su secuencia. De hecho, el ADN satélite representa uno de los tipos de secuencias de ADN más variables, y por tanto uno de los componentes más dinámicos de los genomas eucariotas. Las nuevas variantes se originan por la propagación de las mutaciones entre las unidades de repetición, mediante mecanismos moleculares de intercambio no recíproco, y la posterior fijación en la población gracias a la reproducción sexual, que conduce a la evolución concertada (Dover, 1982), tal y como se explicará detalladamente en el apartado "Evolución del ADN satélite".

No obstante, hay diferentes niveles de diversidad intraespecífica de las unidades monoméricas del ADN satélite dependiendo de distintos factores que afectan a la evolución de estas secuencias. Además, la variación en las secuencias a menudo da lugar a la aparición de nuevas subfamilias de ADN satélite que se homogenizan de modo independiente (Navajas-

Pérez *et al.*, 2006). La homogeneización intraespecífica está asociada con una rápida divergencia entre las secuencias repetidas de diferentes especies, cuyas unidades de repetición son habitualmente muy diferentes en su secuencia de ADN, originándose así ADN_s satélites específicos de especie o de género (Barragán *et al.*, 2003). Por otra parte, existe el caso contrario, el de familias de ADN satélite conservadas entre las especies de una familia, incluso entre familias de un orden, o ADN_s satélites ancestrales cuya secuencia se ha conservado durante más de 90 millones de años, como por ejemplo las familias de ADN satélite PstI y HindIII descritas en esturiones (Robles *et al.*, 2004).

Como se ha indicado anteriormente, diferentes ADN_s satélite pueden coexistir en un mismo genoma, presentando distinto número de copias debido a la amplificación diferencial a partir de una biblioteca común compartida con otras especies relacionadas. Un ejemplo son los ADN_s satélite I y II caracterizados en el escarabajo *Tribolium madens*, que constituyen el 30% y el 4% respectivamente de su genoma (Mravinac y Plohl, 2007). En una biblioteca, habitualmente, una o unas pocas familias de ADN satélite pueden estar presentes en un número alto de copias (denominándose ADN_s satélites mayoritarios), mientras que las restantes familias que forman parte de esa biblioteca están representadas en menor número (ADN_s satélites minoritarios). Cada una de las diferentes familias de ADN satélite de una biblioteca pueden ser consideradas unidades evolutivas independientes, con una dinámica propia de variación en el número de copias y en su secuencia nucleotídica. Por este motivo, un ADN satélite mayoritario en una especie (o genoma) puede tener un ADN satélite homólogo minoritario en otra especie relacionada con la anterior (Meštrović *et al.*, 1998; Pons *et al.*, 2004). Un ejemplo que ilustra este hecho es el que ocurre en las especies de escarabajo del género *Pimelia*, en el que el ADN satélite PIM357 constituye entre el 25% y el 45% del genoma de 26 especies analizadas (Bruvo *et al.*, 2003).

Por otro lado, se han descrito monómeros de ADN satélite con distintos niveles de complejidad en su estructura. Así, algunas familias de ADN satélite presentan elementos estructurales que se han asociado con una función determinada, mientras que otras familias están compuestas únicamente por repeticiones simples de secuencias de ADN. Ejemplos del primer caso son los ADN_s satélites centroméricos de humanos (ADN satélite α -humano), murino (ADN satélite γ murino) y de aves. Los monómeros del ADN satélite α -humano tienen un tamaño de 171 pb de longitud, y presentan un motivo de 17 pb denominado caja CENP-B (Ugarcovick, 2005) donde se une la proteína centromérica B (CENP-B) importante para el reclutamiento de CENP-A (variante centromérica de la histona H3-CenH3) durante la formación *de novo* del centrómero y la correcta formación del nucleosoma centromérico (Stimpson *et al.*, 2010). En aves, existen motivos de 3-10 pb que están conservados en las

secuencias de los ADN_s satélites centroméricos de seis especies pertenecientes a diferentes familias. Estos motivos se han asociado con la curvatura de la hélice de ADN descrita en muchas familias de ADN satélite diferentes que presentan series de A+T y díadas (Ugarković, 2005; Lorite *et al.*, 2013).

La complejidad de las unidades monoméricas puede deberse también a la propia formación de la unidad de repetición. Hay unidades monoméricas originadas por eventos de duplicación y subsiguiente divergencia de una secuencia inicial o motivo corto, como por ejemplo se ha descrito en diferentes ADN_s satélites de animales (Garrido-Ramos *et al.*, 1995; Bruvo *et al.*, 2003; De la Herrán *et al.*, 2001) y plantas (Navajas-Pérez *et al.*, 2005; Macas *et al.*, 2010). Además, en regiones con baja tasa de recombinación, los monómeros tienden a formar repeticiones de orden superior o HORs (del inglés *Higher-Order Repeats*), que son el resultado de la homogeneización conjunta de variantes monoméricas adyacentes. Se forman así repeticiones más largas, existiendo una mayor similitud entre unidades HORs que dentro de ellas, generándose de este modo una nueva estructura repetitiva, una unidad de repetición de orden superior (Plohl *et al.*, 2008). Como ejemplos pueden citarse el ADN satélite del escarabajo *Pholeuon proserpinae* que se organiza en una HOR de 532 pb de longitud formada por dos tipos de monómeros de 266 pb de longitud cada uno, o el ADN satélite α del centrómero humano. Por ejemplo, el cromosoma 7 humano presenta dos HORs formadas por dos subfamilias divergentes, una HOR formada por seis monómeros y otra formada por un dímero. En ambos casos, las HORs son altamente homogéneas (98.7% en el escarabajo y 97-100% en el cromosoma 7 humano), mientras que las subunidades dentro de ellas muestran una menor identidad (96,6% en el escarabajo y 72% de media en el humano) (Plohl *et al.*, 2008). Finalmente, también se ha sugerido que los elementos transponibles y algunos genes repetidos podrían contribuir a la formación y difusión de ADN_s satélites. Algunos ejemplos son el elemento transponible tipo *MITE* (*Miniature Inverted-repeat Transposable Element*) que se ha descrito como potencial elemento generador de una familia de ADN satélite en moluscos bivalvos (López-Flores *et al.*, 2004), un elemento tipo *mariner* que parece haber tomado parte en la expansión de una familia de ADN satélite entre los cromosomas de hormigas del género *Messor* (Palomeque *et al.*, 2006; Palomeque y Lorite, 2008), copias duplicadas del ADN_r 5S que podrían constituir el origen del ADN satélite 5SHindIII del pez neotropical *Hoplias malabaricus* (Vicari *et al.*, 2010), o de un gen ARN_t ancestral que ha sido propuesto como probable origen de secuencias cortas repetidas en tándem en plantas superiores (Benslimane *et al.*, 1986).

Con respecto a la localización en posiciones específicas de los cromosomas, los ADN_s satélites pueden clasificarse como secuencias centroméricas, pericentroméricas, subteloméricas

o intercalares. Entre los ADN_s satélite centroméricos, el más conocido y mejor caracterizado es el ADN satélite α humano citado anteriormente, cuya unidad de repetición tiene 171 pb de longitud y se encuentra conservado en primates. Otros ADN_s satélites centroméricos han sido caracterizados por ejemplo en los genomas de peces (Garrido-Ramos *et al.*, 1995), de la cabra y otros rumiantes (Nijman *et al.*, 2001) y en insectos (Palomeque y Lorite, 2008). Tanto en animales como en plantas, los ADN_s satélite centroméricos se consideran componentes esenciales de los centrómeros, como es el caso del satélite α humano o las familias CentO y CentC del arroz y del maíz (Buscaino *et al.*, 2010). En animales, se han descrito ADN_s satélite subteloméricos por ejemplo en peces de la familia Sparidae (Garrido-Ramos *et al.*, 1998) y en la coquina *Donax trunculus* (Petrovič *et al.*, 2009). En plantas, diferentes familias de ADN_s satélites subteloméricos/intercalares han sido descritos en muchas especies como *Silene latifolia* (ADN satélite subtelomérico de 313 pb de longitud monomérica, Garrido-Ramos *et al.*, 1999), *Aegilops tauschii* (familia de ADN satélite subtelomérico con monómero de 340 pb de longitud presente en géneros de la tribu Triticeae), y en las especies cultivada y silvestre del centeno *Secale cereale*, *S. montanum*, *S. silvestre*, and *S. africanum* (familia de ADN satélite subtelomérico e intersticial que se encuentra ampliamente distribuida en la tribu Triticeae y en el género Avena) (Sharma y Raina, 2005). La especie *S. cereale* presenta además en su genoma otra familia de ADN satélite subtelomérica (con monómero de 350 pb de longitud y que constituye el 2,5% de su genoma) y la familia pSc250 (con unidad de repetición de 550 pb y que constituye el 1% de su genoma) (Sharma y Raina, 2005). Aparte de los ADN_s satélites localizados en regiones cromosómicas determinadas, hay otras familias de ADN satélites específicas de cromosomas concretos como es el caso de la familia MCSAT localizada en un único par cromosómico de la especie *Muscari comosum* (de la Herrán *et al.*, 2001), o familias de ADN satélite localizadas únicamente en cromosomas sexuales o cromosomas B. Un ejemplo de ADN_s satélite específicos de cromosomas sexuales son los identificados en *Drosophila melanogaster*: el ADN satélite de 359 pb de tamaño monomérico está localizado en la región heterocromática del cromosoma X, mientras que la heterocromatina del cromosoma Y contiene tres familias de ADN satélite de tamaño monomérico menor (de 5 a 7 pb de unidad monomérica) (Palomeque y Lorite, 2008). En plantas, la familia de ADN satélite RAYSI se encuentra localizada exclusivamente en el cromosoma Y de diferentes especies del género *Rumex* (Navajas-Pérez *et al.*, 2006), mientras que el ADN satélite RAE180 se ha acumulado diferencialmente entre las especies de este mismo género, abundando en el cromosoma Y de las especies con determinación sexual de tipo XX/XY₁Y₂ (Navajas-Pérez *et al.*, 2009). Ejemplos de ADN_s satélites localizados en los cromosomas B son la familia pSsP216 específica de la especie *Drosophila subsilvestris* (Gutknecht *et al.*, 1995) y la familia de ADN satélite pericentromérica del saltamontes *Eyprepocnemis plorans* (Abdelaziz *et al.*, 2007).

En cuanto a la posible función que puede tener el ADN satélite, los estudios más recientes cuestionan la hipótesis inicial que consideraba a este tipo de ADN repetido como ADN basura o ADN egoísta que simplemente se acumula en los genomas. Esta consideración inicial se fundamentaba principalmente en la elevada variabilidad que presentan sus secuencias, y en la ausencia de productos génicos transcritos a partir de ellas. Sin embargo, su ubicuidad en el genoma de prácticamente todos los organismos eucariotas y su localización preferencial en determinadas regiones cromosómicas contrastan con esa aparente ausencia de función. En la actualidad, entre las funciones que potencialmente se consideran que podrían atribuirse al ADN satélite se incluyen la participación en el establecimiento y mantenimiento de estados de la cromatina facilitando su ensamblaje, influencia en la expresión génica y contribución en la regulación epigenética, por ser en ocasiones las repeticiones de ADN satélite una fuente generadora de siARNs (Ugarković, 2005; Pezer *et al.*, 2012).

I.1.1.1. Evolución del ADN satélite

El ADN satélite, al igual que otras secuencias de ADN repetidas, presenta un modo de evolución particular denominado evolución concertada (Plohl *et al.*, 2010, 2012). Este modelo evolutivo propone que las unidades de una familia de ADN repetido no evolucionan de manera independiente sino en conjunto, como una entidad única, debido a que intercambian información entre ellas, favoreciendo así la transferencia horizontal de las mutaciones entre las unidades de repetición de una familia (homogeneización de las secuencias de una misma familia). Posteriormente, las mutaciones se propagan mediante transferencia vertical (reproducción) a los individuos que forman parte de la población (fijación de las secuencias en la población). El resultado es que las unidades de repetición que constituyen una familia de ADN repetido en una especie son muy similares entre sí, mientras que se diferencian ampliamente de las unidades repetidas homólogas presentes en especies relacionadas.

Dover (2002) definió *molecular drive* como el proceso que conduce a la evolución concertada a través de la consecución de las dos etapas descritas anteriormente, la homogeneización y la fijación de las unidades de repetición de una familia determinada de ADN repetido. Los mecanismos de intercambio no recíproco entre moléculas de ADN (entrecruzamiento desigual, conversión génica, replicación por círculo rodante e intercambio mediado por transposones) permiten la propagación de las nuevas variantes monoméricas (unidades de repetición mutadas) dentro de la familia a la que pertenece. El entrecruzamiento desigual y la conversión génica son considerados los dos motores principales a nivel molecular de la evolución concertada. En el entrecruzamiento desigual ocurre intercambio desigual entre

dos unidades cromosómicas que mantienen un apareamiento deslizado favorecido por la existencia de repeticiones homólogas en tándem, entre dos cromátidas hermanas o dos cromátidas homólogas durante la mitosis o durante la meiosis. El resultado del entrecruzamiento desigual es una alteración en el número de copias en una cromátida con respecto a la otra, ya que se trata de una recombinación no recíproca que origina la multiplicación de la secuencia repetida en una cromátida a costa de la delección correspondiente en la otra cromátida. La conversión génica es el fruto de la resolución en regiones de heterocigosis del ADN heterodúplex que se origina como consecuencia inherente a todo proceso de recombinación genética. En este caso, no se altera el número de copias de la secuencia repetida, sino que lo que ocurre es que una se "transforma" en la otra debido a la actuación de los mecanismos de reparación celular que corrigen los apareamientos erróneos. Se ha propuesto que secuencias motivo palindrómicas, como la caja CENP-B, podrían estar implicadas en el proceso de homogeneización por conversión génica (Lorite *et al.*, 2002). Adicionalmente, se han propuesto otros mecanismos homogeneizadores, que al igual del entrecruzamiento desigual, estarían implicados también en la formación y amplificación de tándemes repetitivos. Así, se ha demostrado la existencia de moléculas de ADN extracromosómicas circulares originadas a partir de escisiones una región de ADN repetido en tándem. Este tipo de moléculas circulares se replicarían activamente mediante un mecanismo que seguiría el modelo de replicación por círculo rodante. Los largos concatámeros que se generan serían posteriormente insertados en la misma región de la que proceden o en una nueva localización. Por último, mecanismos de transposición podrían explicar también el origen de tándemes repetitivos así como su amplificación y expansión por el genoma. En este caso, también habría que incluir el origen y amplificación de un tándem por entrecruzamiento desigual entre secuencias transponibles. Las mutaciones se fijarán en la población al transmitirse a los descendientes mediante la reproducción sexual.

La evolución concertada sería por tanto un proceso evolutivo gradual constituido por ciclos de mutación, homogeneización de las mutaciones y fijación de las mutaciones en la población tras la reproducción de los individuos, que en conjunto da lugar a una homogeneización intraespecífica y a una rápida divergencia de las secuencias de ADN satélite entre grupos de organismos aislados reproductivamente (Plohl *et al.*, 2012). De este modo, el proceso de *molecular drive* relaciona la filogenia con la divergencia en las secuencias de ADN satélite (Dover, 1982), si bien esta relación no siempre se establece fielmente ya que el proceso de *molecular drive* depende de diferentes factores intrínsecos y extrínsecos (Luchetti *et al.* 2003, 2006; Robles *et al.* 2004; Mravinac *et al.* 2005; Navajas-Pérez *et al.* 2005b, 2009; Suárez-Santiago *et al.* 2007; Pérez-Gutiérrez *et al.*, 2012).

La variabilidad global de las secuencias monoméricas de un genoma es una característica compleja que depende de la conservación genómica y la divergencia de los ADN_s satélite, y de los patrones de distribución y de homogeneización dentro de las variantes, la presión selectiva a la que están sometidas, así como en el modo de reproducción y de diversos factores poblacionales (Plohl *et al.*, 2010, 2012). Por tanto, la evolución concertada puede enlentecerse debido a la localización del ADN satélite, a su organización y a su número de copias (Navajas-Pérez *et al.* 2005, 2009), a restricciones funcionales (Mravinac *et al.* 2005), o a diferentes factores biológicos (Luchetti *et al.* 2003, 2006; Robles *et al.* 2004; Suárez-Santiago *et al.* 2007). En ausencia de restricciones selectivas y biológicas, la tasa de evolución concertada de la secuencia de una familia de ADN satélite dependería básicamente del tiempo de divergencia entre especies (Pérez-Gutiérrez *et al.*, 2012).

Aún así, la elevada tasa de cambios en el curso de la evolución concertada no es la única causa posible que puede explicar la diversidad entre familias mayoritarias de ADN satélite entre especies relacionadas (Plohl *et al.*, 2012). Además del potencial de cambio en la secuencia, los ADN_s satélite están cambiando constantemente en número de copias debido a la expansión y contracción de la serie de unidades monoméricas. Puesto que varias familias o subfamilias de ADN satélite pueden coexistir en un mismo genoma, la evolución de ADN_s satélites específicos puede estar dirigida por el cambio en el número de copias de monómeros dentro de una biblioteca compartida por un grupo de especies (Plohl *et al.*, 2012). De acuerdo con la hipótesis de la biblioteca, la existencia de una familia de ADN satélite específica de especie es el resultado de su expansión y/o contracción dentro del *pool* de secuencias compartidas por genomas relacionados (Fry and Salser, 1977). En definitiva, las bibliotecas de ADN satélite representan una fuente permanente de secuencias que pueden expandirse de modo independiente dando lugar a una familia de ADN satélite mayoritario en cada especie cambiando rápidamente el perfil genómico de ADN satélite (Plohl *et al.*, 2012).

De este modo, y de acuerdo de nuevo a la hipótesis de la biblioteca, distintos táxones relacionados comparten una biblioteca de diferentes secuencias conservadas de ADN satélite (diferentes familias de ADN satélite, pero también variantes monoméricas o subfamilias de ADN satélite), que pueden expandirse diferencialmente en cada taxón. La variación en las secuencias puede estancarse durante largos periodos evolutivos a causa de una acción reducida de los mecanismos de recombinación no recíproca por lo que las diferentes variantes podrían mantenerse como una biblioteca a partir de la cual cualquiera de las secuencias puede ser expandida diferencialmente en cada taxón con el subsiguiente reemplazamiento de una variante por otra en diferentes especies (Pons *et al.*, 2004; Mravinac *et al.* 2002; Meštrović *et al.* 2006). Un ejemplo de esto ha sido descrito en el género *Palorus*, en el que el estudio de las

repeticiones de familias de ADN satélite mayoritario en cada especie reveló su presencia en número bajo de copias en otras especies del género pudiendo comprobarse una elevada conservación interespecífica de sus secuencias, además de la ausencia completa de cualquier mutación diagnóstica de especie (Meštrović *et al.*, 1998). Esta hipótesis ha sido confirmada además de en insectos (Meštrović *et al.*, 1998; Mravinac *et al.*, 2002; Cesari *et al.*, 2003; Pons *et al.*, 2004) y en anfibios (Feliciello *et al.*, 2005, 2006).

Aunque existen muchos estudios relacionados con el análisis del ADN satélite, son muchas las incógnitas que hoy día persisten sobre los factores que influyen en su evolución. En esta Tesis Doctoral, se ha pretendido contribuir al conocimiento de la dinámica evolutiva del ADN satélite en vegetales, realizando dos estudios diferentes. En el primero, hemos estudiado el efecto de la localización cromosómica en la evolución de este tipo de secuencias repetidas mediante el análisis de una familia de ADN satélite conservada en las especies dioicas del género. En el segundo estudio, hemos analizado los patrones de evolución del ADN satélite poniendo a prueba las teorías evolutivas aceptadas en la actualidad frente a un escenario evolutivo complejo como es el de la tribu Cardueae de la subfamilia Carduoideae (Asteraceae). Para ello, hemos analizado 736 secuencias de una familia de ADN satélite conservada en la tribu, aisladas a partir del genoma de 55 especies pertenecientes a 20 géneros diferentes.

El objetivo general planteado en este trabajo es el de contribuir al conocimiento de la dinámica evolutiva del ADN satélite en vegetales mediante el análisis de la influencia que en ella tienen tanto factores intrínsecos al propio genoma como factores extrínsecos.

Para abordar este objetivo general, se plantearon los siguientes objetivos específicos:

Primero. Se pretende determinar el efecto de la localización cromosómica en la evolución de este tipo de secuencias repetidas mediante el análisis de una familia de ADN satélite, conservada en las especies dioicas del género *Rumex*, que presenta una distribución diferente en diferentes especies: autosomas en especies dioicas con sistema de determinación sexual XX/XY y cromosomas sexuales en especies dioicas con sistema de determinación sexual XX/XY₁Y₂.

Segundo. Se pretende determinar el efecto que en la evolución del ADN satélite tiene una historia evolutiva compleja como es la de la tribu Cardueae de la subfamilia Carduoideae (Asteraceae). De la misma manera, pretendemos determinar si el estudio del ADN satélite en este grupo puede arrojar algo de luz en su clasificación. La clasificación de esta tribu es complicada debido a una gran diversidad morfológica y a su elevado número de especies, existiendo gran controversia con respecto la delimitación de la tribu, la delimitación de los límites entre sus subtribus y la delimitación de sus géneros.

III.1. Material empleado

El material vegetal utilizado para la extracción de ADN corresponde a hojas procedentes de 6 especies del género *Rumex* (familia Polygonaceae) y 55 especies pertenecientes a diferentes géneros de la subfamilia Carduoideae (familia Asteraceae).

En el caso del género *Rumex*, el estudio se ha centrado en especies dioicas. En particular, hemos analizado tanto especies con un sistema de cromosomas sexuales simple XX/XY, con un mecanismo de determinación del sexo basado en la presencia de un cromosoma Y activo, como especies con un sistema de cromosomas sexuales complejo XX/XY₁Y₂, cuyo mecanismo de determinación del sexo se basa en la relación X:A.

El material de las diferentes especies se ha obtenido de poblaciones naturales como se detalla en la Tabla 1. Se obtuvo ADN procedente de individuos de ambos sexos.

Tabla 1. Especies analizadas del género *Rumex*

Especie	Sistema cromosómico de determinación del sexo	Procedencia
<i>Rumex acetosa</i> L.	XX/XY ₁ Y ₂	Capileira, Granada (España)
<i>Rumex papillaris</i> Boiss & Reut.	XX/XY ₁ Y ₂	Capileira, Granada (España)
<i>Rumex acetosella</i> L.	XX/XY	Vollubilis (Marruecos)
<i>Rumex suffruticosus</i> J. Gay ex Willk.	XX/XY	Puerto de Navacerrada, Segovia (España)
<i>Rumex hastatulus</i> (NC) Baldwin ex Ell.	XX/XY ₁ Y ₂	Carolina del Norte (EEUU)
<i>Rumex hastatulus</i> (TX) Baldwin ex Ell.	XX/XY	Texas (EEUU)

(NC) Raza de Carolina del Norte; (TX) Raza de Texas.

En el caso de la subfamilia Carduoideae, hemos estudiado varios géneros pertenecientes a cada una de las cuatro subtribus integrantes de la subfamilia: Centaureinae, Carduinae, Carlininae y Echinopsinae descritas en la Tabla 2. La mayoría de los ejemplares con los que hemos trabajado proceden de pliegos de herbario registrados en las colecciones del Herbario de la Universidad de Granada (Tabla 2).

Tabla 2. Especies analizadas de la subfamilia Carduoideae

Táxones	Localidad/Referencia
Subtribu Centaureinae	
<i>Centaurea</i> L.	
<i>Centaurea</i> subg. <i>Acrocentron</i>	
<i>C. clementei</i> Boiss.	ESP: Cádiz, Grazalema/GDAC 5102
<i>C. granatensis</i> Boiss.	ESP: Granada, Huétor Santillán/ GDA 24948
<i>C. saxifraga</i> Coincy	ESP: Granada, Zújar/GDA 49332
<i>Centaurea</i> subg. <i>Cyanus</i>	
<i>C. cyanus</i> L.	ESP: Ctra. Ávila-Salamanca/ GDAC 23553
<i>Centaurea</i> subg. <i>Centaurea</i>	
<i>Centaurea</i> sect. <i>Acrolophus-Phalolepis</i> (Cass.) Dostál	
<i>C. alba</i> L. subsp. <i>alba</i>	ESP: Ávila, Romacastaña, Sierra del Águila/GDAC 24948
<i>C. corymbosa</i> Pourret	FRA: Narbonne, La Clape/BC
<i>Centaurea</i> sect. <i>Willkommia</i> Blanca	
<i>C. boissieri</i> D.C. subsp. <i>willkommii</i> (Schultz Bip. ex Willk.) Dostál	ESP: Murcia, Sierra Esuña/GDAC 6649
<i>C. delicatula</i> Breitw. and Podlech	TUN: Djebel Chambi/MPU
<i>C. jaemensis</i> Degen and Debeaux in Degen	ESP: Jaén, Pozo Alcón, Embalse de la Bolera/GDAC 6724
<i>Carthamus</i> L.	
<i>C. tinctorius</i> L.	ESP: Sevilla/ GDAC 13709
<i>C. lanatus</i> L.	ESP: Córdoba, Sierra Morena, Los Patalos/GDAC 39253
<i>Phomis</i> Hill	
<i>P. arborescens</i> (L.) G. López	ESP: Granada, Sierra de Lújar/GDAC 18987
<i>Carduncellus</i> Adans.	
<i>C. caeruleus</i> (L.) C. Presl	ESP: Málaga, de Alfarnate a Colmenar/ GDA 49305
<i>C. calvus</i> Boiss. & Reut.	MAR: Beni Hosmar, Mantis Borgues/GDA 3998
<i>C. cuatrecasasii</i> G. López	ESP: Granada, Pedro Martínez, Mencal, Cerro Matías/ GDA 25041/
<i>C. dianthus</i> Webb	ESP: Alicante, Mangó, cerca de Denia/GDAC 4193/
<i>C. hispanicus</i> Boiss. ex DC.	ESP: Granada, Sierra Sagra/ GDAC 20480
<i>C. mitissimus</i> (L.) DC.	ESP: León, Solana del Puerto de Ventana/ GDAC 13047
<i>Rhaponticoides</i> Vaill.	
<i>R. alpina</i> (L.) M.V. Agab. & Greuter *	ESP: Jaén, Sº de las Villas/GDAC 55133
<i>R. linarensis</i> (Lázaro Ibiza) M.V. Agab. & Greuter *	ESP: Palencia, Dueñas/GDA 13180
<i>Rhaponticum</i> Vaill.	
<i>R. acule</i> (L.) DC.	MAR: Región de Mante Arruit, Marruecos Español/ GDA 5265
<i>R. berardioides</i> Batt.	MAR: Talasentant / GDA 4998
<i>R. coniferum</i> (L.) DC.	ESP: Granada, Pantano de Cubillas/GDA 42407
<i>R. centauroides</i> (L.) Holub	ESP: Lérida, Valle de Arán Puerto de la Bonaigua/GDAC 24874
<i>R. exaltatum</i> (Willk.) Greuter	ESP: Ávila, Hoyocasero/ GDAC 34999
<i>Klasea</i> Cass.	
<i>K. flavescens</i> (L.) Holub subsp. <i>flavescens</i>	ESP: Granada, Lanjarón, Bco. de la Chuca/ GDA 54612
<i>K. flavescens</i> subsp. <i>leucantha</i> (Cav.) Cantó & Rivas Mart.	ESP: Granada, Cerro Gordo/ GDAC 10730
<i>Crupina</i> (Pers.) DC.	
<i>C. crupinastrum</i> (Maris) Vis.	ESP: Granada, Alhama de Granada, Venta de La Paloma/GDA54194
<i>C. vulgaris</i> Cass.	ESP: Córdoba, Sº Marena, Barranco del Arenoso/ GDAC 42205
<i>Cheirolophus</i> Cass.	
<i>C. falcipectus</i> Montelongo & Moraleda	ESP: Gran Canaria, Jardín Botánico Viera y Clavijo/GDAC 44225
<i>C. intybaceus</i> (Lam) Dostál	ESP: Almería, Níjar, Las Presillas Altas/ GDA 25217
<i>C. sempervirens</i> (L.) Pomel	ESP: Granada, Sº de Cázulas, cueva de Fune/ GDA 52606
<i>C. sventenii</i> (A. Santos) G. Kunkel	ESP: Gran Canaria, Jardín Botánico Viera y Clavijo/ GDAC 44224
<i>C. teydis</i> (Buch) G. López	ESP: Tenerife, Sta. Cruz Tenerife, Arico, circo de Las Cañadas del Teide/ GDA 41753
<i>Volutaria</i> Cass.	
<i>V. crupinoides</i> (Desf.) Maire	MAR: entre Midar y Nador, próxima a Driouch/GDAC 39797
<i>V. lippii</i> (L.) Maire	ESP: Almería, proximidades de Alsodux/ GDA 52154
<i>V. muricata</i> (L.) Maire	ESP: Málaga, carretera a Canillas de Aceituno/ GDAC 42882
<i>V. tubuliflora</i> (Murb.) Sennen	ESP: Almería, Tabernas/ GDAC 43411
Subtribu Carduinae	
<i>Carduus</i> L.	
<i>C. bourgeanus</i> Boiss. & Reut.	ESP: Granada, Maracena ^a
<i>C. pycnocephalus</i> L.	ESP: Granada, carretera Íllora-Montefrío, Arroyo del Charcón/GDA 16589
<i>Cirsium</i> Mill.	
<i>C. echinatum</i> (Desf.) DC.	ESP: Granada, Iznalloz, Cueva del Agua/ GDA 48724
<i>C. pyrenaicum</i> (Jacq.) All.	ESP: Almería, Bayárcal, Arroyo Anchuelo/ GDAC 42831
<i>Galactites</i> Moench.	
<i>G. duriaei</i> Durieu	ESP: Alicante, Benidorm/ GDA 14547
<i>Jurinea</i> Cass.	
<i>J. fontkeri</i> Cuatrec.	ESP: Granada
<i>J. humilis</i> (Desf.) DC.	ESP: Huesca, Sierra de Guara, Nueno, Nocito/GDA 53602
<i>Onopordum</i> L.	
<i>O. nervosum</i> Boiss.	ESP: Granada, Montefrío, cerca del Puente Romano, Arroyo del Chorro / GDA 54200
<i>Silybum</i> Vaill.	ESP: Granada, Sierra de Orce / GDA 54283
<i>S. eburneum</i> Coss. & Durieu	ESP: Granada, Maracena ^a
<i>S. marianum</i> (L.) Gaertn.	
Subtribu Carlininae	
<i>Arctylis</i> L.	
<i>A. leyserooides</i> (Desf.) Maire	ESP: Almería, Alboloduy, Rambla de los yesos/ GDA 52208
<i>A. humilis</i> L.	ESP: Granada, Baza, curva de la Muerte/ GDA 49405
<i>Carlina</i> L.	
<i>C. acaulis</i> L.	ESP: Santander, Picos de Europa, La Colladina, Macizo de Peña Vieja/ GDA 6331
<i>C. salicifolia</i> (L.f.) Cav.	ESP: Tenerife, Santa Cruz de Tenerife, cerca de Arona/ GDA 41903
Subtribu Echinopsinae	
<i>Echinops</i> L.	
<i>E. sphaerocephalus</i> L.	ESP: Granada, Cogollos Vega, cerca del Cortijo de los Asperones / GDA 21763
<i>E. strigosus</i> L.	ESP: Málaga, proximidades de Estepona/ GDA 50299

Abreviaturas: ESP, España; FRA, Francia; MAR, Marruecos; GDA/GDAC, Herbario de la Universidad de Granada; BC, Herbario de Barcelona; MPU, Herbario de la Universidad de Montpellier. ^aMaterial recogido de la naturaleza. * Especies consideradas sinónimas.

III.2. Extracción de ADN genómico

El ADN genómico de los diferentes individuos de cada una de las especies analizadas se extrajo utilizando el kit comercial *DNAeasy Plant Mini Kit* (Qiagen), a partir de hojas recolectadas en poblaciones naturales, o de pliegos registrados en el Herbario de la Universidad de Granada. El material procedente de pliegos del Herbario fue hidratado con H₂O_{dd} durante toda una noche antes de proceder a la extracción de ADN según el protocolo suministrado por el kit y detallado a continuación:

1. Se maceraron de 0,05 a 0,1 g de material foliar fresco y aproximadamente 0,02 g para el material de herbario en presencia de nitrógeno líquido con la ayuda de un mortero de porcelana.
2. El material pulverizado se transfirió a microtubos de 1,5 ml a los que se añadieron 400 µl de tampón de lisis (*buffer API*) y 4 µl de ARNasa A (100 mg/ml). Para conseguir una mezcla homogénea se agitó cada tubo fuertemente con un vórtex para después dejar en incubación durante 10 minutos a una temperatura de 65 °C. Durante el transcurso de esta incubación se invirtió el tubo repetidas veces evitando la acumulación de material en el fondo del tubo y optimizar la lisis celular.
3. Se añadieron 130 µl de tampón para precipitar las proteínas (*buffer AP2*) mezclándolo por inversión suavemente para incubarlo a continuación al menos 5 minutos en hielo.
4. Tras la incubación, se centrifugó 5 minutos a 14.000 rpm. El sobrenadante obtenido se transfirió a una columna (*Quiashredder Mini spin column*) que se centrifugó durante 2 minutos a 14.000 rpm.
5. El sobrenadante se transfirió a un tubo nuevo de 1.5 ml evitando tocar el precipitado. Una vez recogido el volumen, se añadieron 1.5 volúmenes de tampón AP3/E y se mezcló mediante pipeteo.
6. La mezcla se transfirió a columnas *DNAeasy Mini spin column* (volumen máximo 650 µl) que se centrifugó durante 1 minuto a 8.000 rpm. Se descartó el sobrenadante y se repitió el proceso hasta pasar todo el volumen de la muestra a través de la columna.
7. A continuación se transfirió la columna a un tubo de 2 ml, y se añadieron 500 µl de tampón de lavado (*buffer AW*). Se centrifugó 1 minuto a 8.000 rpm y se eliminó el sobrenadante.

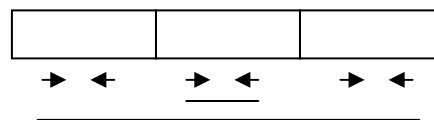
8. Se agregaron nuevamente 500 µl de tampón de lavado (*buffer AW*) y se centrifugó 2 minutos a 14.000 rpm.
9. La columna se transfirió a un tubo de 1,5 ml y se añadieron 50 µl de tampón de elución (*buffer AE*) directamente sobre la membrana. Durante 5 minutos se incubó a temperatura ambiente y, finalmente, se centrifugó durante 1 minuto a 8.000 rpm.

La integridad del ADN genómico se comprobó mediante electroforesis en geles horizontales de agarosa al 1% teñidos con SYBR-safe® (Invitrogen), que se fotografiaron bajo luz ultravioleta utilizando el sistema de documentación de geles Gel Doc XR (Bio-Rad).

III.3. Aislamiento de ADN satélite

Las distintas familias de ADN satélite se aislaron mediante amplificación por PCR utilizando cebadores diseñados previamente a partir de las unidades monoméricas RAE180 del género *Rumex*, Rae180a 5'-TCATCGAACTTCATTCAT-3'/Rae180b 5'-TATAGTAATATCTCGAT-C-3' (Navajas-Pérez et. al., 2005b), y *Hinfl* de la Subfamilia Carduoideae, CenHinf1 5'-GCTTCG-TTTTGATAGTTCGTGG-3'/CenHinf2 5'-TAACTTTTGCTACGGGAGTCCG-3' (Suárez-Santiago et al., 2007).

Cada pareja de cebadores se diseñó en una zona interna conservada del monómero, de modo que cada amplificado puede constar de medio monómero o de un monómero completo y dos medios monómeros a ambos lados, correspondientes a las secuencias de repetición adyacentes, como se observa en el esquema de la derecha. Se han analizado siempre monómeros completos a partir del segundo tipo de amplificados, prescindiendo de los monómeros incompletos adyacentes.



Representación esquemática de los sitios de anclaje de los cebadores. Cada una de las cajas representa un monómero, y las flechas indican la posición y orientación de los cebadores. Las líneas inferiores muestran el tamaño de los amplificados.

Las reacciones de PCR se llevaron a cabo en un termociclador Applied Biosystems (GeneAmp 2700), y se realizaron en un volumen final de 50 µl conteniendo 15-20 ng de ADN. Para las especies del género *Rumex* la mezcla de la reacción de PCR contuvo además 10 mM Tris-HCl pH 8,3; 5 mM KCl; 1,5-2 mM MgCl₂; 0,2 mM de cada dNTP (Roche); 0,2 µM de cada uno de los cebadores (Rae180a/Rae180b) y 1,25 U de Taq polimerasa (Bioline). La reacción de amplificación se llevó a cabo con los siguientes parámetros: desnaturalización inicial a 94 °C durante 5 minutos; 35 ciclos de desnaturalización a 94 °C durante 1 minuto,

unión de los cebadores a 55 °C durante 1 minuto y elongación de la cadena a 72 °C durante 1 minuto; por último, elongación final a 72 °C durante 10 minutos. En el caso de especies de la tribu Cardueae, la mezcla de reacción de PCR contuvo 10 mM Tris-HCl pH 8,3; 5 mM KCl; 0,75-3 mM MgCl₂; 0,2 mM de cada dNTP (Roche); 0,2 μM de cada uno de los cebadores (CenHinf1/ CenHinf2) y 2,25 U de Taq polimerasa (Biotools). Las condiciones de amplificación fueron: desnaturalización inicial a 94 °C durante 5 minutos; 40 ciclos de desnaturalización a 94 °C durante 1 minuto, unión de los cebadores a 55 °C durante 1 minuto y elongación de la cadena a 72 °C durante 1 minuto; por último, elongación final a 72 °C durante 10 minutos.

Los resultados de las reacciones de amplificación se analizaron mediante electroforesis en geles de agarosa al 1,5% teñidos con SYBR safe[®] (Invitrogen), y los productos amplificados se purificaron utilizando el kit comercial *GFX[®] PCR DNA and Gel Band Purification Kit* (GE Healthcare) según el protocolo que se describe a continuación:

1. Se escindió del gel de agarosa la banda discreta de ADN generada en la reacción de amplificación utilizando un bisturí estéril, y se dispuso en un microtubo de 1,5 ml.
2. Se añadieron 10 μl de tampón de captura (*Capture buffer*) por cada 10 mg de peso de la banda de agarosa. Se mezcló vigorosamente y se incubó durante 15 minutos a 60 °C.
3. La disolución obtenida se transfirió a una columna *GFX Microspin[™]* dispuesta sobre un microtubo colector de 1,5 ml y se incubó a temperatura ambiente durante 1 minuto. A continuación, se centrifugó a 13.000 rpm durante 30 segundos.
4. Los restos de agarosa disueltos se decantaron del tubo colector y se colocó de nuevo la columna sobre dicho tubo. Se añadieron 500 μl de tampón de lavado (*Wash buffer*) y se centrifugó a 13.000 rpm durante 30 segundos.
5. La columna se transfirió a un microtubo de 1,5 ml, y se añadieron 25 μl de H₂O ultrapura sobre la matriz de la columna. Se incubó 2-3 minutos a temperatura ambiente.
6. Finalmente, el ADN purificado se recuperó centrifugando la columna a 13.000 rpm durante 2 minutos.

III.4. Clonación de los productos de PCR

Una vez purificados los fragmentos obtenidos por PCR a partir de geles de agarosa, las unidades monoméricas del ADN satélite se ligaron al vector TOPO-TA utilizando el kit comercial *TOPO-TA Cloning[®] Kit for Sequencing* (Invitrogen). Incluye un vector plasmídico denominado pCR[®]4-TOPO[®] diseñado para la ligación de productos de PCR, que se caracteriza por estar linearizado y provisto de un extremo 3'dT unido covalentemente a una Topoisomerasa, que

cataliza la unión inserto-vector. La complementariedad de bases entre el inserto y el vector se produce debido al extremo final 3'-dA que las ADN polimerasas añaden a todos los productos de PCR (por su actividad nucleotidil transferasa terminal).

La reacción de ligación se llevó a cabo en un tubo de 0,2 ml en el que se añadieron:

6 µl de producto purificado (15-20 ng/ul)
1 µl de solución salina
1,5 µl de vector TOPO-TA

La mezcla se incubó durante un mínimo de 30 minutos a temperatura ambiente.

A continuación, se utilizaron de 3 a 6 µl de la reacción de ligación para transformar células bacterianas competentes de la cepa DH5α-T1^R de *Escherichia coli* proporcionadas con el kit comercial de ligación. El protocolo seguido fue el recomendado por la casa comercial:

1. Se tomó el volumen de la ligación y se transfirió a la suspensión de células competentes de la cepa DH5α-T1^R, mezclando suavemente.
2. La mezcla se sometió a un choque térmico mediante incubación durante 20 minutos en hielo, durante 30 segundos en un baño a 42°C y nuevamente en hielo durante 2 minutos.
3. Las bacterias transformadas se crecieron en 250 µl de medio SOC (triptona 2%, extracto de levadura 0.5%, NaCl 10 mM, KCl 2.5 mM, MgCl₂ 10 mM, MgSO₄ 10 mM, Glucosa 20 mM) durante 1 hora y media a 37°C con agitación suave y en posición horizontal para una mejor aireación.
4. Finalmente, se sembraron 50-100 µl del cultivo de bacterias transformadas en placas de Petri con medio LB sólido (peptona 10%, extracto de levadura 5%, NaCl 5%, agar 15%) con Ampicilina (0,05 mg/ml), y se incubaron a 37°C durante 16 horas.

La presencia de inserto en los clones recombinantes crecidos en las placas de Petri se comprobó mediante PCR de colonias, utilizando las parejas de cebadores T3/T7 (T3: 5'-ATTA-ACCCTCACTAAAGGGA-3'; T7: 5'-TAATACGACTCACTATAGGG-3') o M13F/M13R (M13F: 5'-GTAAAACGACGGCCAG-3'; M13R: 5'-CAGGAAACAGCTATG-AC-3'). La mezcla de la reacción de PCR se preparó con los siguientes reactivos: 10 mM Tris-HCl pH 8.3, 5 mM KCl, 2 mM MgCl₂, 0,2 µM de cada dNTP (Roche), 3,2 µM de cada uno de los cebadores (T3/T7), y 1 unidad de Taq polimerasa (Bioline) en un volumen final de 50 µl. La reacción de amplificación se llevó a cabo un termociclador Applied Biosystems (GeneAmp 2700) con los siguientes parámetros: desnaturalización inicial a 94 °C durante 3 minutos; 30 ciclos de desnaturalización a 94 °C durante 15 segundos, unión de los cebadores a 55 °C durante 20

segundos y elongación de la cadena a 72 °C durante 2 minutos; por último, elongación final a 72 °C durante 7 minutos.

El resultado se observó en geles de agarosa, y los clones recombinantes se crecieron en medio LB líquido (peptona 10%, extracto de levadura 5%, NaCl 5%) a 37°C durante 16 horas.

III.5. Purificación de ADN plasmídico

La extracción y purificación de los plásmidos recombinantes se realizó siguiendo el protocolo del kit comercial “Purificación Estándar de Plásmidos” (Dominion). Los pasos seguidos se describen a continuación:

1. Las bacterias crecidas en medio LB líquido se recogieron en microtubos de 1,5 ml y se centrifugaron a 13.000 rpm durante 2 minutos.
2. Se decantó el sobrenadante y se resuspendió el precipitado bacteriano en 100 µl de tampón de resuspensión (*Buffer S-I*) mediante agitación vigorosa.
3. A continuación, se añadieron 100 µl de tampón de lisis (*Buffer S-II*), mezclando por inversión, y 100 µl de tampón de precipitación, mezclando nuevamente por inversión.
4. La mezcla se centrifugó 10 minutos a 13.000 rpm. El sobrenadante se recuperó y se transfirió a una columna de purificación colocada sobre un tubo colector limpio.
5. Se añadieron 500 µl de tampón de unión, mezclando por inversión y se dejó reposar a temperatura ambiente unos 2 minutos. La mezcla se transfirió a una columna (*DNA prep*) y se centrifugó a 5.500 rpm durante 2 minutos.
6. Se eliminó el sobrenadante, y se añadieron a la columna 700 µl de tampón de lavado. Se centrifugó 5.500 rpm durante 2 minutos.
7. La columna se colocó nuevamente en el vial y se añadieron 500 µl de isopropanol frío. Se centrifugó a 5.500 rpm 2 minutos. Para desechar los restos de isopropanol, se repitió la centrifugación a una velocidad de 13.000 rpm.
8. Finalmente, la columna se colocó en un microtubo de 1,5 ml y se eluyó el ADN plasmídico con 25 µl de tampón EB previamente calentado a 65°C centrifugando a 13.000 rpm durante 1 minuto.

III.6. Secuenciación

Las reacciones de secuenciación se realizaron según el método de Sanger *et al.* (1977), utilizando el kit comercial *ABI Prism[®] Big Dye[®] Terminator Cycle Sequencing* (Applied Biosystems). Las dos cadenas de los insertos ligados en plásmidos recombinantes fueron secuenciadas utilizando el secuenciador automático multicapilar *ABI PRISM[®] 3100-Avant Genetic Analyzer* (Applied Biosystems) del Centro de Instrumentación Científica de la Universidad de Granada, según las instrucciones del fabricante. Usamos parejas de cebadores universales cuyas secuencias diana se encuentran presentes en el vector de clonación TOPO (T3/T7 o M13F/M13R).

La reacción de secuenciación se llevó a cabo utilizando 200-300 ng de ADN plasmídico, 2 µl del tampón de secuenciación 5x, 3,2 pmol del cebador y 1 µl de la mezcla BDT (*Big Dye[®] Terminator*). Las condiciones de amplificación consistieron en una incubación a 95°C durante 3 minutos, 30 ciclos de desnaturalización a 95°C durante 20 segundos, alineamiento a 50°C durante 10 segundos y elongación a 60°C durante 4 minutos.

Las reacciones de secuenciación se precipitaron tras incubarlas durante 15 minutos con una mezcla de acetato sódico y etanol (62,5 µl de etanol 95%, 3µl de acetato sódico 3M y 14,5 µl de agua ultrapura estéril por muestra) mediante centrifugación durante 20 minutos a 13.000 rpm a 4 °C. A continuación se decantó el sobrenadante y el precipitado se lavó con 250 µl de etanol 70% y tras centrifugar durante 15 minutos a 13.000 rpm en frío las muestras se dejaron secar en oscuridad a 37 °C. Los productos secuenciados se resuspendieron en 20 µl de formamida.

En el caso del estudio llevado a cabo en el género *Rumex*, las secuencias RAE180 utilizadas en los análisis presentados en esta memoria proceden de estudios previos en el caso de *R. acetosa* y *R. papillaris* (Navajas-Pérez *et al.*, 2005b), así como de *R. acetosella*, *R. suffruticosus* y *R. intermedius* (Navajas-Pérez, 2005; Navajas-Pérez *et al.*, 2007). El aislamiento de las secuencias de *R. hastatulus* se ha realizado durante el desarrollo de la presente memoria. Los números de acceso de estas secuencias en la base de datos EMBL son AM398600-AM398639, AJ580328-AJ580351, AJ580453-AJ580469, AJ634478-AJ634526, y FN564362-FN5643696 (Navajas-Pérez *et al.*, 2009; Quesada del Bosque *et al.*, 2011), en el primer caso, y FN561567-FN561625 (raza de Carolina del Norte), y FN563834-FN563909 (raza de Texas) en el caso de *R. hastatulus* (Quesada del Bosque *et al.*, 2011).

En el caso del estudio realizado en las especies de la subfamilia *Carduoideae* los números de acceso a la base de datos EMBL de las 55 secuencias seleccionadas del estudio

previo (Suárez-Santiago *et al.* (2007a) corresponden a los siguientes números de acceso a la base de datos EMBL: AM712738-AM712751, AM712761-AM712774, AM712803-AM712809, AM712815-AM712821, y AM712828-AM712840. Los números de acceso a la base de datos EMBL para las secuencias restantes analizadas en este estudio son HF571538-HF571985 (subtribu Centaurineae), estando pendientes de asignación los de las Subtribus Carlininae, Echinopsinae y Carduinae.

III.7. Análisis de las secuencias

Los cromatogramas se visualizaron utilizando el programa Chromas v. 1.45 (Connor McCarthy, 1996-1998). Se cotejaron las secuencias obtenidas con las existentes en la base de datos de ADN de EMBL, EBI y GenBank, mediante el programa interactivo en red BLAST (Altschul *et al.*, 1997) del NCBI. Las secuencias se alinearon con la ayuda del software Clustal X (Thompson *et al.*, 1997) que utiliza el algoritmo Clustal (Higgins *et al.*, 1988, 1994).

Para la caracterización de las secuencias RAE180 de *R. hastatulus*, se determinó la variación dentro de la especie/raza (medida como diversidad nucleotídica, π , el número medio de diferencias nucleotídicas por sitio entre dos secuencias; Nei, 1987) y las diferencias entre especies/razas (medidas como el número medio de sustituciones nucleotídicas por sitio entre especies, Dxy; Nei, 1987) utilizando el programa informático “satDNA Analyzer” (Navajas-Pérez *et al.*, 2007).

Con el mismo programa se analizó el patrón de variación en cada posición nucleotídica entre las unidades monoméricas de una especie y entre los monómeros de diferentes especies. Dentro de cada especie, el programa clasifica cada sitio como monomórfico (el mismo nucleótido en todas las repeticiones de las secuencias monoméricas de esa especie) o polimórfico (el nucleótido varía entre repeticiones). Cuando se comparan secuencias de diferentes especies, el programa diferencia los sitios polimórficos entre polimorfismos no compartidos y polimorfismos compartidos (cuando el mismo nucleótido está presente en dos especies). Estas posiciones polimórficas compartidas son consideradas ancestrales, previas a la divergencia de las dos especies (Navajas-Pérez *et al.* 2005b). Por otro lado, los sitios polimórficos no compartidos son autapomorfías, y representan diferentes estados de transición durante el proceso de homogeneización de las secuencias y divergencia interespecífica. Bajo la asunción de la evolución concertada como un proceso dependiente del tiempo, los estados de transición esperados durante la propagación de una variante monomérica hacia su fijación se puede definir de acuerdo con el modelo de Strachan *et al.* (1985). Este es un método de diferenciación de la variación mediante el análisis de los patrones de variación de nucleótidos en

cada sitio considerado de modo independiente entre todas las repeticiones de una familia de secuencias determinada cuando se comparan las especies por pares (Strachan *et al* 1985; Pons *et al.*, 2002; Navajas-Pérez *et al* 2007a). Según este modelo, se diferencian seis estados (clases I-VI) en la propagación de una variante monomérica a lo largo de la familia y de la especie. Brevemente, la posición (sitio) de clase I representa la completa homogeneidad a lo largo de todas las unidades monoméricas comparadas entre un par de especies, mientras que las clases II, III y IV representan estados intermedios en los cuales una de las especies presenta un polimorfismo. La frecuencia de la nueva variante nucleotídica en el sitio considerado es baja en la clase II e intermedia en la clase III, mientras que la clase IV representa sitios en los que la mutación ha reemplazado al nucleótido original en la mayoría de los miembros de la familia de ADN repetido de la otra especie. La clase V representa sitios diagnósticos, en los que la variante está completamente homogeneizada y fijada en todas las repeticiones de una especie mientras que las repeticiones de la otra especie mantienen el nucleótido original. La clase VI representa un paso más con respecto a los sitios diagnósticos, es decir, nuevas variantes que aparecen en algunos miembros de la familia de ADN repetido en un sitio divergente entre las dos especies.

En el caso del ADN satélite RAE180 de *Rumex* el análisis filogenético se realizó mediante el método de neighbor-joining utilizando el programa MEGA v.5 (Tamura *et al.*, 2011). La selección del modelo para el cálculo de distancias genéticas (Kimura gamma) se realizó según el criterio del programa Modeltest v. 3.6 (Posada y Crandall 1998), que lleva a cabo un test jerárquico de probabilidad entre 56 modelos diferentes de sustitución de ADN. Como soporte de los tests realizados se utilizó el método de remuestreo con reemplazamiento *bootstrap* (1000 réplicas) (Felsenstein, 1985).

En el caso de las especies de la subfamilia Carduoideae, las filogenias se infirieron mediante los métodos de máxima probabilidad (ML) y análisis bayesiano. Se utilizó el programa MEGA v.5 para seleccionar el modelo de sustitución en secuencias de ADN (Tamura *et al.*, 2011). Se eligió el modelo con menor valor en el parámetro AIC (*Akaike Information Criterion*) (GTR+G). El análisis de ML se llevó a cabo utilizando el programa MEGA v.5 (Tamura *et al.*, 2011). Los valores de bootstrap se calcularon con 1.000 réplicas (Felsenstein, 1985). Las probabilidades posteriores bayesianas (PPs) se estimaron utilizando el programa MrBayes v3.2 (Ronquist *et al.*, 2012). El grado de la tasa de variación a lo largo de los sitios para particiones de datos individuales fue estimado mediante el parámetro (*s*) de la distribución gamma. Cada análisis bayesiano se inició con árboles de partida al azar, y se realizó durante un millón de generaciones con frecuencias de muestreo de árboles en la iteración número 500. Para todos los análisis, la varianza de las secuencias fraccionadas fue de 0,01. El factor de reducción de escala potencial (PSFR) tuvo un valor cercano a 1,0 para todos los parámetros del modelo de

sustitución. La fracción de los valores muestreados descartados se fijó en 0,25. Las probabilidades posteriores de 0,95-1,00 se consideraron estadísticamente significativas. Los árboles filogenéticos se editaron utilizando el programa FigTree, disponible en la dirección <http://tree.bio.ed.ac.uk/software/figtree>.

III.8. Cuantificación del satélite RAE180 mediante PCR cuantitativa (qPCR)

El porcentaje relativo de repeticiones RAE180 en los diferentes genomas de las especies del género *Rumex* se determinó mediante PCR cuantitativa (qPCR). Las condiciones de temperatura de cebado y concentración de cloruro de magnesio en la PCR fueron optimizadas para cada ensayo. Las reacciones de PCR estándar se llevaron a cabo en un volumen final de 25 microlitros, con ADN genómico a una concentración de 90 ng/μl, 0.2 mM dNTPs, 200 mM de cada cebador (Rae180a and Rae180b), la concentración optimizada de MgCl₂, y 1U de Taq polimerasa. Las condiciones de amplificación fueron una desnaturalización inicial a 95°C durante 10 min, 40 ciclos de amplificación consistentes en desnaturalización a 95°C durante 15 seg, temperatura optimizada de cebado durante 30 seg, y extensión a 72°C durante 30 seg. Las reacciones de qPCR se llevaron a cabo utilizando el kit comercial SensiMix/dT Kit, que incluye una referencia interna (ROX) (Quantace). Cada reacción de qPCR requirió 12.5 μl SensiMix/dT, cebadores sentido y antisentido (Rae180a and Rae180b) a una concentración final optimizada de 10 pmol, 1 μl SYBR Green I solution 50X, 90 ng/μl de ADN genómico, y agua ultrapura estéril hasta completar un volumen final de 25 μl. Las condiciones de amplificación en este caso fueron una desnaturalización inicial a 95°C durante 10 min, para una activación de la ADN polimerasa, seguida de 40 ciclos de desnaturalización a 95°C durante 15 seg, cebado a 56,7°C durante 30 seg, y extensión a 72°C durante 30 seg. Las reacciones de qPCR se llevaron a cabo en el termociclador para PCR a tiempo real "Real-Time PCR Chromo 4 Thermocycler" (BioRad). La cinética de la qPCR determina, para cada pocillo correspondiente a una muestra, el Ct, el ciclo en el que la fluorescencia acumulada en el pocillo sobrepasa el umbral correspondiente a varias veces las desviaciones estándar de la fluorescencia basal. La representación gráfica de los valores Ct frente al logaritmo de la cantidad de ADN diana muestra una relación lineal, lo que permite una cuantificación relativa simple mediante comparaciones con una curva estándar derivada de la amplificación en la misma placa, de diluciones seriadas de una muestra de ADN de referencia. La curva estándar se genera a partir de diluciones seriadas de una disolución de plásmido recombinante (50 ng/μl) que contiene una unidad de repetición RAE180 amplificada con los cebadores Rae180a and Rae180b. Todas las muestras, incluyendo los estándares externos y el control negativo de la PCR (sin ADN diana), se amplificaron por triplicado. La eficiencia de la amplificación para cada cebador fue

determinada a partir de la pendiente de la curva estándar; solo los cebadores con una pendiente de la curva estándar de alrededor de -3,3 fueron utilizados para la cuantificación. Cuando se analizaron mediante qPCR, las diluciones produjeron un conjunto de curvas estándar que fueron utilizadas para calcular el valor de la pendiente usando el programa "Opticon3 software" (BioRad).

IV.1. Introducción

IV.1.1. Sistema de determinación del sexo en plantas

A diferencia de los animales, el hermafroditismo es el sistema reproductivo más frecuente en vegetales. Tan sólo un 6% aproximadamente de las plantas con flores son dioicas. Aun así, la dioecia se extiende en un amplio rango taxonómico. Las especies dioicas existen en un 7.1% de los géneros y en un 38% de las familias de angiospermas (Renner y Ricklefs, 1995). Muchas de ellas se encuentran en géneros en los que la mayoría de sus miembros son hermafroditas, como sucede en el género *Rumex*.

Entre los distintos sistemas de determinación del sexo destaca el sistema cromosómico, predominante en la mayoría de especies vegetales que, a su vez, está ampliamente extendido entre los animales. Sin embargo, los cromosomas sexuales han evolucionado en un número limitado de angiospermas dioicas y sólo unos pocos han sido caracterizados a nivel citológico y molecular (Charlesworth, 2002). Estos cromosomas portadores de genes que determinan el sexo, en muchas especies, se pueden diferenciar por su tamaño y forma a nivel microscópico denominándose heteromórficos. Si, por el contrario, no se visualiza ninguna diferencia morfológica entre ellos, se conocen como homomórficos. En la mayoría de los casos descritos, los machos son los que portan los cromosomas sexuales heteromórficos, es decir, son el sexo heterogamético y las hembras el homogamético. Al contrario de lo que sucede en animales, la presencia de cromosomas sexuales heteromórficos es rara en vegetales y su aparición es relativamente reciente (Ruiz-Rejón, 2004; Navajas-Pérez, 2005).

El sistema cromosómico de determinismo sexual más común consiste en la presencia de un sistema de cromosomas sexuales heteromórficos XX/XY en el que el control genético está basado en la presencia de un cromosoma Y activo (Ruiz-Rejón, 2004). Está muy bien representado en diversos géneros de plantas entre los que se incluyen *Silene*, *Asparagus* y *Rumex*, en particular, en la especie *R. acetosella*. Sin embargo, en algunas especies del género *Rumex* existe otra alternativa, un sistema múltiple de cromosomas sexuales XX/XY₁Y₂, en las que el control genético se basa en el balance entre el número de cromosomas X y el número de juegos autosómicos (relación X:A) (Ainsworth *et al.*, 1999). En *Rumex* este sistema cromosómico de determinación del sexo se encuentra en especies como *R. acetosa* (Ruiz Rejón, 2003; Navajas-Pérez, 2005).

IV.1.2. El género *Rumex* L.

El género *Rumex* pertenece a la familia Polygonaceae y comprende, aproximadamente, unas 200 especies. Este género es un buen ejemplo de la variedad de sistemas reproductivos y sistemas cromosómicos de determinación del sexo, ya que se encuentran todos ellos representados en las especies que lo integran. En este grupo coexisten especies hermafroditas, polígamas, ginodioicas, monoicas y dioicas. Asimismo, como se ha comentado anteriormente, las especies dioicas presentan sistemas de cromosomas sexuales en distintos estadios evolutivos, simples (XX/XY) y complejos (XX/XY₁Y₂), e incluso sistemas sin cromosomas sexuales diferenciados (Navajas-Pérez *et al.*, 2005a, 2007a).

La clasificación de las especies pertenecientes al género *Rumex* se ha modificado en diversas revisiones realizadas según los criterios empleados en cada sistema de clasificación. La más reciente es de Navajas-Pérez *et al.* (2005a), que analizaron filogenéticamente el grupo a partir de secuencias ITS y cloroplastidiales (Figura 1). Según esta filogenia molecular, el género *Rumex* consta de tres clados filogenéticos íntimamente asociados a sus sistemas reproductivos. Un grupo basal compuesto por especies hermafroditas, un segundo clado constituido por especies polígamas y ginodioicas, y un tercer clado que comprende especies dioicas (Navajas-Pérez *et al.*, 2005a).

El clado de las especies dioicas, en el que se centra nuestro trabajo, comenzó a divergir del resto de las especies del género hace unos 15-16 millones de años (Navajas-Pérez *et al.* 2005a). Está compuesto por todas las especies dioicas euroasiáticas y americanas del género, con dos linajes bien diferenciados por sus sistemas cromosómicos de determinación del sexo, el sistema simple XX/XY y el sistema complejo XX/XY₁Y₂. Ambos linajes se separaron hace unos 12-13 millones de años (Navajas-Pérez *et al.* 2005). Dentro del grupo de especies XX/XY un caso particular es la especie *R. hastatulus*. Esta especie endémica del norte de América, presenta dos razas cromosómicas, la raza de Texas con un sistema cromosómico de determinación del sexo XX/XY y la raza de Carolina del Norte con un sistema XX/XY₁Y₂. Según esta nueva filogenia molecular, existe un paralelismo entre la evolución de los sistemas reproductivos y la evolución cariotípica del género. De este modo, se habría producido una reducción cromosómica desde el nivel x=10 de las especies hermafroditas, pasando por estadios intermedios x =9 de especies hermafroditas, polígamas y ginodioicas hasta el nivel x=8 y x=7, característico de especies dioicas (Figura 1).

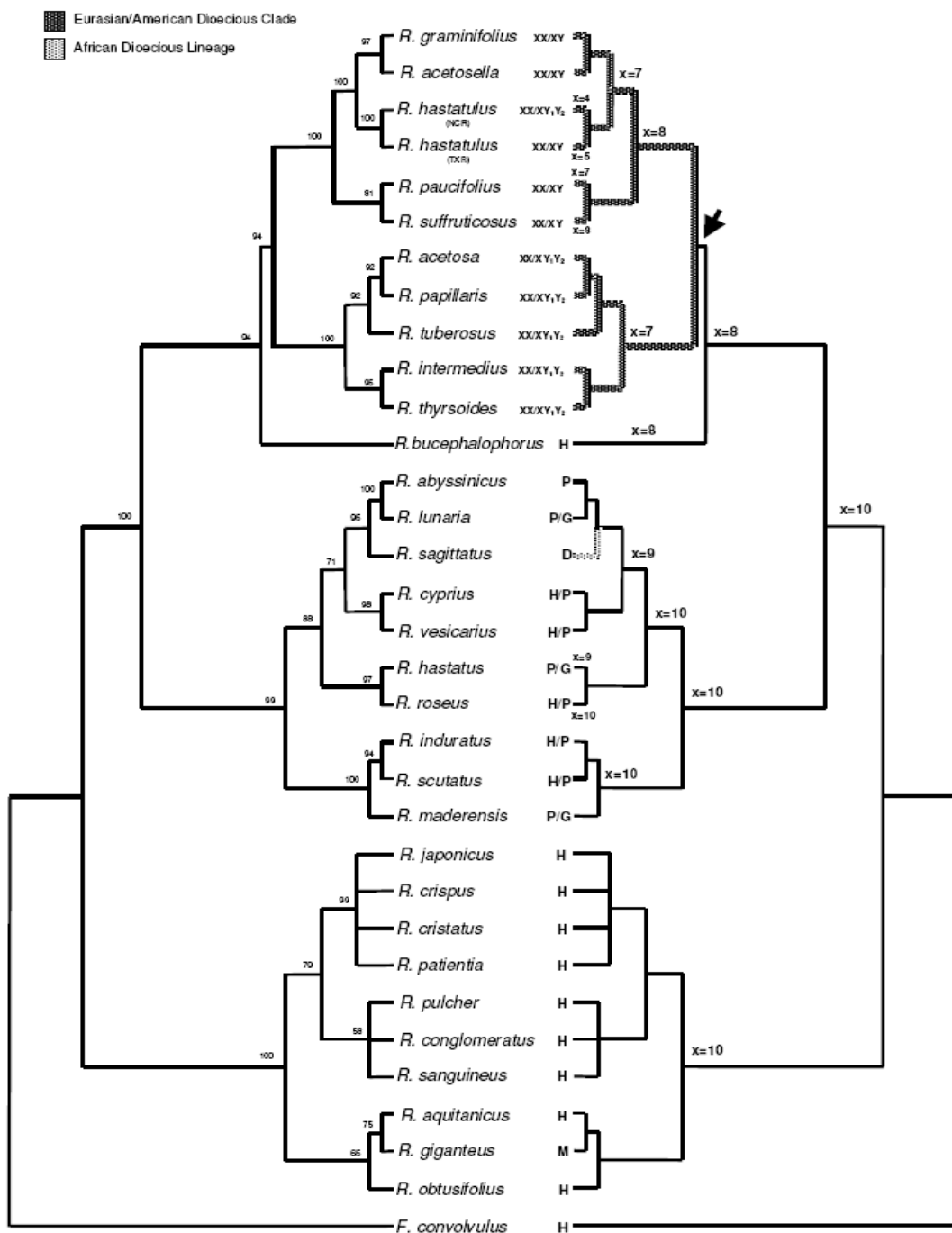


Figura 1. Filogenia molecular del género *Rumex* basada en secuencias nucleares y cloroplastidiales. El árbol muestra la correlación entre la filogenia y la evolución de los sistemas reproductivos (D: dioecia, G: ginodioecia, H: hermafroditismo, M: monoecia, P: poligamia) y de los números cromosómicos básicos. Este árbol se construyó según el criterio de máxima verosimilitud combinando las secuencias nucleares (ITS-1 e ITS-2) y cloroplastidiales (intrón trnL y espaciador intergénico trnL-trnF) analizadas por Navajas-Pérez *et al.*, (2005a). Las especies dioicas del género *Rumex*, en el que centra esta Tesis Doctoral, se subdivide, a su vez, en dos grupos monofiléticos que comparten un ancestro común. Mientras que uno de los clados está compuesto por especies con un sistema cromosómico de determinación del sexo complejo XX/X₁Y₂, el otro, consta de especies cuyo sistema cromosómico de determinación del sexo es simple XX/YY. (Figura tomada de Navajas-Pérez *et al.*, 2005a).

El grupo de especies XX/X₁Y₂ consta de especies como *R. acetosa*, *R. papillaris*, *R. intermedius* y *R. thyrsoides*, especies en las que las hembras son 2n= 12 + XX y los machos 2n= 12 + X₁Y₂. En ellas, la determinación del sexo se basa en el balance entre el número de cromosomas X y el número de juegos autosómicos (relación X:A) (Ainsworth *et al.*, 1999). Por otro lado, dentro del grupo de especies XX/X₁Y, se incluyen *R. hastatulus*, *R. acetosella*, *R. suffruticosus*, *R. graminifolius* y *R. paucifolius*, en los que la determinación del macho se basa en la presencia de un cromosoma Y activo (Figura 1).

El género *Rumex* es muy interesante para estudiar la evolución de los cromosomas sexuales, porque es el único donde se encuentran representados todos los estadios evolutivos, desde una pareja de cromosomas sexuales indiferenciados y no heteromórficos como en *R. sagittatus* (Degraeve, 1976) a estadios de diferenciación más avanzados con una amplia región no recombinante entre los cromosomas sexuales heteromórficos y con el cromosoma Y degenerado como en *R. acetosa* y, otras especies relacionadas XX/X₁Y₂, pasando por estadios intermedios de diferenciación genética de cromosomas sexuales heteromórficos con una pequeña región en la que la recombinación ha sido suprimida (como en las especies XX/X₁Y, tal es el caso de *R. acetosella* y *R. suffruticosus*) (Cuñado *et al.*, 2007).

Dentro del grupo de especies dioicas del género *Rumex* con un sistema cromosómico de determinación del sexo simple, la especie *R. hastatulus* resulta especialmente interesante. Se encuentra distribuida en el sudeste de Norteamérica siendo endémica de esta región, y presenta dos razas cromosómicas alopátricas, morfológicamente indistinguibles. Se las conoce por denominaciones relacionadas con su distribución geográfica: la raza de Texas y la raza de Carolina del Norte. La raza de Texas está integrada por individuos con un sistema cromosómico de determinación del sexo XX/X₁Y en el que las hembras son XX y los machos son X₁Y. Es diploide con un número de cromosomas 2n=10 y un número cromosómico básico x=5. La raza de Carolina del Norte habría evolucionado secundariamente, mediante reordenaciones cromosómicas, desde un sistema de cromosomas sexuales simple a uno complejo en el que las hembras presentan un cariotipo con dos cromosomas X (XX) mientras que los machos son X₁Y₂ (Smith, 1969). Esta raza también es diploide con un número cromosómico básico x=4. En esta raza las hembras son 2n=8 (6 autosomas + XX) y los machos son 2n=9 (6 autosomas + X₁Y₂). Esta variación implica un cambio en el mecanismo de determinación sexual dado que, al igual que ocurre en el resto de especies XX/X₁Y de *Rumex*, la raza de Texas presenta un mecanismo de determinación basado en la presencia o ausencia de un cromosoma Y activo en el genoma del individuo, mientras que al igual que ocurre en la especie X₁Y₂, la raza de Carolina del Norte tiene un mecanismo de determinación sexual basado en la relación entre el número de cromosomas X y el número de juegos haploides del complemento cromosómico del

individuo (relación X:A) (Smith, 1969). Curiosamente, el mismo cambio evolutivo desde un sistema cromosómico XX/XY a uno XX/XY₁Y₂ y desde un mecanismo de determinación basado en un Y activo a un mecanismo basado en la relación X:A, ha ocurrido independientemente en dos linajes diferentes, uno en el ancestro de especies euroasiáticas y americanas XX/XY₁Y₂ y otro en *R. hastatulus* (Smith, 1969; Degraeve, 1976; Navajas-Pérez *et al.*, 2005a).

IV.1.3. Familias de ADN satélite en el género *Rumex* L.

En el género *Rumex* se han caracterizado tres familias de ADN satélite: RAE730 (Shibata *et al.*, 2000a), RAYSI (Shibata *et al.*, 1999) y RAE180 (Shibata *et al.*, 2000b).

La familia RAE730, con unidades de repetición de 730 pb, está presente en segmentos heterocromáticos de una pareja cromosómica autosómica. La familia RAYSI la constituyen unidades monoméricas de repetición de 920 pb, con una localización exclusiva en el cromosoma Y, dentro de la que se han descrito dos subfamilias denominadas RAYSI-S y RAYSI-J (Navajas-Pérez *et al.*, 2005b, 2006). Ambas familias, RAE730 y RAYSI, se han acumulado específicamente en especies dioicas portadoras de sistemas XX/XY₁Y₂ como *R. acetosa*, *R. papillaris* y *R. intermedius* (Navajas-Pérez *et al.*, 2005a, 2006, 2009), interviniendo RAYSI activamente en la degeneración del cromosoma Y. Por el contrario, no se ha detectado la implicación de RAYSI en la degeneración de los cromosomas sexuales de especies con sistemas de cromosomas sexuales XX/XY como *R. acetosella* y *R. suffruticosus* (Navajas-Pérez *et al.*, 2006; Cuñado *et al.*, 2007). RAE730 y RAYSI presentan un origen común (Navajas-Pérez *et al.*, 2005b). Se originaron a partir de una secuencia inicial con un tamaño de unos 120 pb (Navajas-Pérez *et al.*, 2005b).

Las unidades de repetición de la familia RAE180, objeto de estudio en esta memoria, constan de 183 pb aproximadamente. Esta familia está presente en el genoma de todas las especies dioicas euroasiáticas y americanas del género *Rumex* indistintamente del sistema cromosómico de determinación del sexo (XX/XY o XX/XY₁Y₂), estando ausente en el resto de especies del género. Esta distribución tan particular pone de manifiesto, por un lado, que el clado de las especies dioicas es monofilético y, por otro lado, que la familia RAE180 se originó aproximadamente hace 15-16 millones de años en un ancestro común a todas las especies dioicas (Navajas-Pérez *et al.*, 2005b).

Dentro del grupo de las especies dioicas, la familia RAE180 se ha acumulado diferencialmente en el genoma de las especies del género *Rumex*, mostrando diferentes patrones

de distribución. Mientras que en las especies dioicas con un sistema cromosómico de determinación del sexo XX/XY RAE180 se localiza únicamente en un pequeño locus autosómico (Cuñado *et al.*, 2007), en las especies con un sistema cromosómico de determinación del sexo XX/X₁Y₂ es un constituyente mayoritario de la heterocromatina de los cromosomas Ys (Cuñado *et al.*, 2007), favoreciendo así la degeneración de los mismos, acumulándose además en un pequeño locus de una pareja de autosomas. Respecto a su origen, la familia RAE180 no muestra un origen en común ni ningún tipo de parentesco con ninguna de las otras dos familias de ADN satélite, RAYSI o RAE730 (Navajas-Pérez *et al.*, 2005b).

El patrón evolutivo de una familia de ADN satélite difiere en función de su localización cromosómica, entre otros factores. Así, Navajas-Pérez *et al.* (2005b) demostraron que en ausencia de recombinación, una familia de ADN satélite presente en los cromosomas Ys tiene menor tasa de evolución que otra que está localizada en los autosomas. Al comparar las secuencias pertenecientes a las familias de ADN satélite RAE180 y RAYSI, presentes en los cromosomas Y, entre las especies *R. acetosa* y *R. papillaris*, se observaron tasas de cambio y de evolución concertada menores que cuando se compararon entre ambas especies las secuencias de la familia RAE730, localizada en autosomas (Navajas-Pérez *et al.*, 2005b). Este resultado se atribuyó a la ausencia de recombinación entre los cromosomas sexuales. De este modo, se estimó que las tasas de cambio y de evolución concertada de la familia RAE180 y RAYSI en los cromosomas Y son aproximadamente la mitad de los que presentan la familia RAE730 de localización autosómica (Navajas-Pérez *et al.*, 2005b, 2007b).

Toda vez que la familia de ADN satélite RAE180 se localiza en unas especies dioicas en los cromosomas Y además de en un pequeño locus autosómico, mientras que en otras lo hace en autosomas, sería muy interesante demostrar la hipótesis anterior mediante la comparación de la evolución que sigue el mismo tipo de secuencias repetidas dependiendo de su localización. Siendo interesante además analizar la influencia que en dicha evolución tiene el hecho de que existan grandes diferencias entre el número de copias existentes en los loci autosómicos y los loci de cromosomas sexuales. Por otro lado, se hace necesario comprobar si estas mismas diferencias se mantienen entre las dos razas de *R. hastatulus* que presentan sistemas de cromosomas sexuales diferentes, así como determinar la evolución global de RAE180 en el grupo. Todos estos aspectos forman parte de este estudio, presentándose a continuación sus resultados y conclusiones.

IV.2. Resultados y Discusión

IV.2.1. Identificación, aislamiento y caracterización de la familia de ADN satélite RAE180 en el género *Rumex*

Tal como se indica en Material y Métodos, las secuencias RAE180 utilizadas en los análisis presentados en esta memoria proceden de estudios previos en el caso de *Rumex acetosa* y *R. papillaris* (Navajas-Pérez *et al.*, 2005b), así como de *R. acetosella*, *R. suffruticosus* y *R. intermedius* (Navajas-Pérez, 2005; Navajas-Pérez *et al.*, 2007). El aislamiento de las secuencias de *R. hastatulus* se ha realizado durante el desarrollo de la presente memoria. Además, se han realizado nuevas extracciones de ADN y aislamiento de secuencias RAE180 tanto para obtención de sondas como para la cuantificación de esta familia de ADN satélite en el genoma de cada especie, y se han hecho los análisis que nos han permitido comparar la evolución de las distintas secuencias según su localización cromosómica y su representación en el genoma.

En total hemos analizado 309 secuencias de RAE180 a partir de 3 especies dioicas de *Rumex* con un sistema cromosómico de determinación sexual XX/XY (*R. acetosella*, *R. suffruticosus*, *R. hastatulus*) y de 3 especies con un sistema XX/XY₁Y₂ de determinación sexual (*R. acetosa*, *R. papillaris*, *R. intermedius*).

IV. 2.2. Análisis filogenético de las secuencias RAE180

La Figura 2 muestra un árbol filogenético obtenido a partir de una selección de secuencias RAE180 de las especies analizadas de *Rumex*. Este árbol muestra la agrupación de las secuencias en dos clados. El primero incluye secuencias de las especies *R. hastatulus* y *R. acetosella*. A su vez, este clado se subdivide en tres grupos de secuencias, uno compuesto por repeticiones de *R. acetosella* junto con secuencias de la raza Texas y Carolina del Norte de *R. hastatulus* y otros dos, compuestos también por secuencias de *R. hastatulus*. En el caso de las secuencias RAE180 de *R. hastatulus*, el árbol refleja la agrupación de sus secuencias por subfamilias (ver más adelante la definición de estas subfamilias). El segundo clado lo integran secuencias de *R. acetosa*, *R. papillaris* y *R. intermedius*. Las secuencias de *R. suffruticosus* se agrupan en este segundo clado independientes de las del resto de especies.

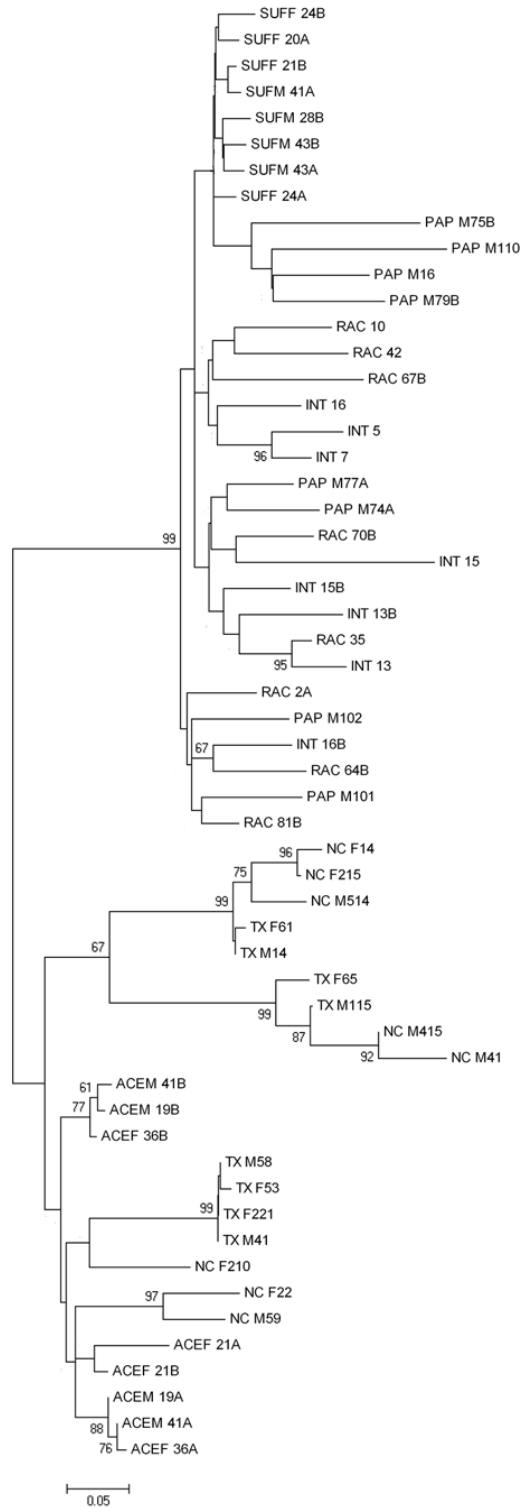


Figura 2. Filogenia molecular de las especies dioicas del género *Rumex* basada en la familia de ADN satélite RAE180. El árbol muestra la relación de las repeticiones RAE180 entre las especies dioicas del género *Rumex* analizadas. (SUF) *R. suffruticosus*, (PAP) *R. papillaris*, (INT) *R. intermedius*, (RAC) *R. acetosa*, (NC) *R. hastatulus* Carolina del Norte, (TX) *R. hastatulus* Texas, (ACE) *R. acetosella*, (M) macho, (F) hembra. El primer número de la secuencia corresponde al número del individuo analizado y los siguientes al número de repetición según un código establecido durante la clonación. No aparecen todas las secuencias analizadas, sólo algunas seleccionadas de cada especie. Los números sobre los nodos representan valores de *bootstrap*.

El árbol filogenético basado en las secuencias del ADN satélite RAE180 difiere del obtenido a partir de otros marcadores nucleares (Figura 1, Introducción). La mayoría de las secuencias no se agrupan por afinidad taxonómica, sino por tipo de repetición y según el sistema de determinación sexual (Figura 2). Las ocho secuencias de la especie *R. suffruticosus* constituyen la excepción en ambos casos (se agrupan por afinidad taxonómica, y son del tipo XX/XY), ya que forman un clado monofilético derivado dentro del clado con mayor soporte estadístico y que agrupa a las secuencias de las especies con determinación sexual cromosómica de tipo XX/XY₁Y₂ (*R. acetosa*, *R. papillaris* and *R. intermedius*) (Figura 2).

Estos resultados se explican por la acción combinada de dos modelos evolutivos para el ADN satélite, uno saltatorio (amplificación diferencial de variantes según la hipótesis de la biblioteca de secuencias repetidas) y otro gradual (*molecular drive*) así como por las historias diferentes de esta familia de ADN satélite en diferentes linajes de acuerdo a su localización, número de copias y organización en los genomas analizados, como veremos en las secciones siguientes.

IV.2.3. Análisis de la evolución de las secuencias RAE180 de acuerdo a su localización cromosómica

Para llevar a cabo este análisis se han comparado las tasas de cambio y la de evolución concertada de las secuencias RAE180 en dos linajes diferentes de las especies dioicas del género *Rumex*. Por un lado, especies con sistema de cromosomas sexuales XX/XY en las que este ADN satélite se localiza en autosomas (*R. acetosella* y *R. suffruticosus*) y por otro, especies con sistema de cromosomas sexuales XX/XY₁Y₂, en las que este ADN satélite se localiza en cromosomas sexuales (*R. acetosa* y *R. papillaris*). En este segundo caso, se ha analizado también de forma diferenciada la evolución de RAE180 según que estas secuencias se localicen en autosomas o que lo hagan en cromosomas sexuales.

En la Tabla 3 se muestran el número de clones secuenciados y de repeticiones analizadas, la composición en pares de bases de las secuencias, la longitud de la unidad monomérica, la variabilidad intraespecífica (medida como diversidad nucleotídica, π , o número medio de diferencias nucleotídicas por sitio entre dos secuencias; Nei, 1987), así como la cuantificación del número de repeticiones del monómero RAE180 en los genomas de las cuatro especies objeto de estudio.

La variabilidad intraespecífica resultó ser baja en *R. acetosella* y *R. suffruticosus* (5,4% y 5,1% respectivamente), siendo similar para las repeticiones aisladas a partir de plantas de sexo

diferente. La familia de ADN satélite RAE180 está muy poco representada en el genoma de la especie *R. acetosella* (7×10^{-4} % del genoma), mientras que está presente en mayor porcentaje (0,04% del genoma) en el genoma de *R. suffruticosus*.

En el caso de *R. acetosa* y *R. papillaris*, la variabilidad intraespecífica fue del 20,9% en la primera y del 21,9% en la segunda, no existiendo grandes diferencias debidas al sexo (son similares entre los machos y las hembras de las dos especies). Las repeticiones RAE180 representan un 0,01% del genoma de las hembras de las dos especies, mientras que este porcentaje asciende al 5,6% en el caso de los machos de ambas especies.

Tabla 3. ADN satélite RAE180 en el genoma de cuatro especies dioicas del género *Rumex*

Especie	Nº de clones			Nº de repeticiones			%AT	Variabilidad			Longitud Monómero	Porcentaje	
	♀	♂	Total	♀	♂	Total		♀	♂	Total		♀	♂
<i>R. acetosella</i>	5	8	13	15	24	39	64,4	0,058	0,054	0,054	163-183	0,0007	0,0007
<i>R. suffruticosus</i>	9	7	16	16	12	28	65,4	0,053	0,040	0,051	181-184	0,04	0,04
<i>R. acetosa</i>	18	32 ^a	51	18	42 ^a	60	68,4	0,199	0,199 ^a	0,209	181-186	0,01	5,60
<i>R. papillaris</i>	15	23 ^a	38	15	31 ^a	46	67,9	0,153	0,211 ^a	0,219	182-184	0,01	5,60

^a Datos tomados de Navajas-Pérez *et al.*, 2007b

La distancia genética media de las secuencias RAE180 entre las especies *R. suffruticosus* y *R. acetosella* fue de 0,226 (0,247 tras excluir del análisis los sitios polimórficos compartidos). Considerando un tiempo de divergencia entre estas dos especies de unos 4 millones de años (Navajas-Pérez *et al.*, 2005a), se estima que la tasa de cambio para esta familia de ADN satélite es de 28×10^{-9} sustituciones por sitio y por año (Tabla 4), casi tres veces más que la estimada para las secuencias RAE180 de machos pertenecientes a las especies *R. acetosa* y *R. papillaris* (Navajas-Pérez *et al.*, 2007b; Tabla 4). La divergencia entre las secuencias de ambas especies también se ha estudiado analizando el patrón de variación en cada posición nucleotídica considerada de modo independiente entre todas las repeticiones RAE180 de las dos especies. De este modo, se ha clasificado cada posición nucleotídica como monomórfica o polimórfica, diferenciando en el último caso entre sitios polimórficos compartidos y no compartidos entre las dos especies, y clasificando cada sitio polimórfico no compartido de acuerdo al modelo descrito en el apartado 7 de Material y Métodos. La Tabla 4 recoge los resultados de este análisis. Al comparar los monómeros RAE180 de las especies *R. acetosella* y *R. suffruticosus*, un 14% de los sitios correspondieron a las categorías IV, V y VI, mientras que el 24% de los sitios fueron estados de transición (categorías II y III). Se identificaron además 8 sitios polimórficos compartidos entre las dos especies (Tabla 4).

Por otro lado, al comparar los monómeros RAE180 aislados a partir de hembras (sólo el locus autosómico) de las especies *R. acetosa* y *R. papillaris*, la distancia genética entre las dos especies fue de 0,197. Sin embargo, casi la mitad de los sitios polimórficos de cada especie se correspondían a sitios polimórficos compartidos entre las hembras de las dos especies (62 sitios polimórficos compartidos). Un resultado similar se obtuvo cuando se compararon los monómeros RAE180 de los machos de estas dos especies (Navajas-Pérez *et al.*, 2005b, Tabla 4). Al excluir los sitios polimórficos compartidos, la distancia genética media entre las secuencias de las hembras de ambas especies fue de 0.045, por lo que la tasa evolutiva en este caso sería de 9×10^{-9} por sitio y por año. En machos, estos valores fueron de 0,033y $8,25 \times 10^{-9}$ respectivamente. No se identificó ningún sitio de tipo IV, V o VI, mientras que el 34% de los sitios correspondió a estados de transición II y III (Tabla 4). Al igual que ocurre con las secuencias procedentes de los machos (Navajas-Pérez *et al.* 2005b), los monómeros RAE180 de las hembras de las dos especies aparecen entremezclados en el correspondiente árbol de secuencias obtenido mediante el método neighbor-joining (Figura 4).

En el género *Rumex*, la amplificación del ADN satélite ha acompañado a los reordenamientos cromosómicos acaecidos durante la evolución de las especies que lo constituyen. La amplificación de las familias de ADN satélite RAYSI y RAE730 se produjo en el linaje que condujo a las especies con determinación compleja del sexo XX/XY₁Y₂ (Navajas-Pérez *et al.*, 2005b, 2006). Ambas familias de ADN satélite comparten un origen común, si bien la familia RAE730 es autosómica y la familia RAYSI está localizada en el cromosoma Y (Navajas-Pérez *et al.*, 2005b, 2006, 2009). A su vez, la familia RAE180 se originó antes de que las especies XX/XY y XX/XY₁Y₂ divergieran, y no tiene relación con las dos familias de ADN satélite anteriores (Navajas-Pérez *et al.*, 2005b). En las especies XX/XY₁Y₂ las repeticiones están localizadas en los cromosomas Y, además de en un pequeño locus autosómico (Cuñado *et al.* 2007; Navajas-Pérez *et al.* 2009). De este modo, en un lapso de tiempo evolutivo relativamente corto desde el origen del clado XX/XY₁Y₂ (12-13 m.a.; Navajas-Pérez *et al.* 2005a), los cromosomas Y han sido sometidos a un proceso acelerado de degeneración genética mediada por la acumulación de las secuencias repetidas RAE180 y RAYSI. De hecho, el 5,6% del genoma de los individuos de sexo masculino de *R. acetosa* y *R. papillaris* está constituido por repeticiones de la familia RAE180, mientras que el porcentaje en hembras es del 0,01%. Asumiendo un tamaño del genoma de *R. acetosa* de 3234 Mb (Bennett y Leitch, 2004), este porcentaje supondría que, en las especies XX/XY₁Y₂, hay cerca de un millón de copias de secuencias RAE180 en los machos (localizadas en los cromosomas Y, y en el pequeño locus autosómico), mientras que en hembras las copias serían alrededor de unas 2000 (sólo en el locus autosómico). *R. acetosa* y *R. papillaris* son dos especies pertenecientes a un grupo homogéneo de especies caracterizadas por características morfológicas y cariológicas similares, incluidas su

sistema de determinación sexual cromosómico complejo (Wilby y Parker 1988; Ainsworth *et al.* 1999). Sin embargo, a pesar de estas similitudes cariotípicas y de la cantidad similar de secuencias RAE180 en sus cromosomas Y, la heterocromatina de los Ys está sujeta a mecanismos evolutivos de reordenamiento rápidos que conducen a diferencias en la distribución del ADN satélite en ambas especies (Navajas-Pérez *et al.*, 2009).

El caso de las especies con determinación sexual XX/XY es completamente diferente. En contraposición al avanzado estado de diferenciación genética entre los cromosomas X e Y en *R. acetosa* y *R. papillaris*, en las especies *R. acetosella* y *R. suffruticosus* la diferenciación genética entre los cromosomas sexuales se encuentra en un estadio temprano. Además, el cromosoma Y en estas especies es eucromático y aparentemente carece de familia alguna de ADN satélite (Cuñado *et al.*, 2007). De hecho, los monómeros RAE180 constituyen las únicas secuencias de ADN satélite identificadas en el genoma de las especies con determinación sexual de tipo XX/XY. No obstante, la familia RAE180 está poco representada en el genoma de estas especies. En *R. suffruticosus*, las repeticiones RAE180 se encuentran acumuladas en un locus pequeño de un par de autosomas (Cuñado *et al.*, 2007). En el caso de *R. acetosella*, la familia de ADN satélite RAE180 se ha aislado e identificado mediante PCR y Southern-blot, pero no se ha podido determinar su localización cromosómica mediante hibridación *in situ* fluorescente (FISH), lo que podría deberse a su presencia en un número bajo de copias o a una organización dispersa (no en tándem) que queda por debajo del nivel de resolución de la técnica (Cuñado *et al.*, 2007). Estos datos corroboran los datos de cuantificación obtenidos en este trabajo. En el caso de *R. suffruticosus*, los monómeros RAE180 suponen un 0,04% del genoma. Asumiendo un tamaño de genoma similar al de *R. acetosa* se podría hacer un cálculo aproximado, que resultaría en una estimación de unas 7000 copias de la repetición RAE180 en *R. suffruticosus*. En el caso de *R. acetosella*, la familia de ADN satélite RAE180 representa el $7,33 \times 10^{-4}$ % de su genoma. El tamaño estimado del genoma de esta especie tetraploide es de 6566 Mb (Bennett y Leitch, 2004), por lo que las repeticiones RAE180 ascenderían a unas 200. Esta diferencia en el número de copias (35 veces menos en *R. acetosella* que en *R. suffruticosus*) explicarían las diferencias en los resultados de la hibridación FISH (Cuñado *et al.*, 2007).

La familia de ADN satélite RAE180 está presente en el genoma de las especies dioicas euroasiáticas y americanas de *Rumex*, mientras que está ausente en el resto de especies del género. Este dato apoya la hipótesis de que RAE180 estaba presente en el genoma del ancestro común a todas las especies dioicas de *Rumex*, de modo que el origen de las secuencias RAE180 puede datarse en al menos 15-16 millones de años (Navajas-Pérez *et al.* 2005a). Las diferencias en la cantidad de repeticiones RAE180 entre las especies analizadas en este trabajo podrían reflejar eventos de amplificación diferencial dependientes del tiempo evolutivo y del lugar físico

de acumulación en el genoma. Para el ADN satélite RAE180 hay tres escenarios: i) el mantenimiento de un bajo número de copias RAE180 en *R. acetosella*, ii) una moderada amplificación en un locus autosómico en *R. suffruticosus*, y iii) una amplificación masiva en los cromosomas Y de las especies con determinación sexual XX/Y₁Y₂ manteniéndose una moderada amplificación en un locus autosómico. A pesar del origen común de todas las secuencias RAE180, y del corto lapso de tiempo evolutivo desde la separación entre los dos grupos de especies -XX/XY vs XX/XY₁Y₂- (12-13 millones de años; Navajas-Pérez *et al.* 2005a), nos encontramos con dos situaciones:

1. Las secuencias RAE180 no se han amplificado dentro de los genomas de especies con determinación sexual tipo XX/XY, excepto en un único locus autosómico detectable mediante FISH solo en la especie *R. suffruticosus*. La situación en el caso de la especie *R. acetosella*, en la que las pocas copias de las repeticiones RAE180 no se detectan mediante la técnica de FISH, podría ser considerada una situación ancestral que se ha mantenido en la línea XX/XY dioica de *Rumex*. Un estado derivado dentro de esta línea sería el proceso de amplificación detectado en la especie *R. suffruticosus*. Este proceso de amplificación podría haber ocurrido después de la separación entre estas dos especies, en los últimos cuatro millones de años según los datos de Navajas-Pérez *et al.* (2005a) referidos a la divergencia en la secuencia ITS entre *R. acetosella* y *R. suffruticosus*. Los cambios cuantitativos inducidos por la amplificación de las repeticiones monoméricas RAE180 podrían ocurrir en el curso del proceso de especiación, dando lugar a un perfil específico en cada especie. Así, la divergencia en la secuencia resulta evidente entre *R. suffruticosus* y *R. acetosella*. De hecho, el 14% de las posiciones nucleotídicas en cada repetición son sitios diferenciados (mutaciones diagnósticas de especie o estados Strachan tipo IV-VI) entre las dos especies (Tabla 4). Solo el 4% de los sitios nucleotídicos son sitios polimórficos ancestrales compartidos por ambas especies. Estos datos indican la eficacia del proceso de *molecular drive* en la expansión de nuevas variantes que conducen a homogeneidad intraespecífica (5,1% y 5,4% de variación en la secuencia respectivamente, Tabla 3) y divergencia interespecífica (22,6% de divergencia entre las secuencias de las dos especies, Tabla 4) tras la separación entre las dos especies, un patrón conocido como evolución concertada. Los estados de transición en el proceso de diferenciación (estados II y III) representan el 24% de las posiciones nucleotídicas (Tabla 4). Este dato apoya el proceso gradual de diferenciación entre las secuencias de *R. suffruticosus* y *R. acetosella* después de la amplificación de las secuencias RAE180, y confirman el doble origen del patrón observado: el efecto homogeneizador de una amplificación relativamente recientemente en *R. suffruticosus*, y el efecto de los mecanismos moleculares de homogeneización una vez amplificado el locus en esta especie. Comparando lo que ocurre con las secuencias RAE180 que están localizadas en los cromosomas Y (Navajas-Pérez *et al.* 2005b), las repeticiones en tándem

RAE180 muestran valores de divergencia elevados cuando se comparan estas dos especies. De hecho, nosotros calculamos una tasa de sustitución de 28×10^{-9} sustituciones por sitio y por año, una tasa 3 veces más alta que la estimada para las secuencias RAE180 ligadas a los cromosomas Y (Navajas-Pérez *et al.* 2007b; Tabla 4) pero similar a la tasa evolutiva determinada para el ADN satélite en otras especies como es el caso del complejo *Drosophila obscura* (Bachmann y Sperlich, 1993). En el género *Rumex*, otra familia de ADN satélite autosómica, la familia RAE730, presenta una tasa de sustitución de 22×10^{-9} sustituciones por sitio y por año. La comparación entre *R. acetosa* y *R. papillaris*, que divergieron hace unos 2,5 millones de años (Navajas-Pérez *et al.*, 2005a), mostró un 4% de estadios II y III, y un 30% de estadios IV a VI (Navajas-Pérez *et al.*, 2007b, Tabla 4). La variabilidad intraespecífica fue del 4,6% para *R. acetosa* y del 2,8% para *R. papillaris* (Navajas-Pérez *et al.*, 2007b, Tabla 4). Mientras que hemos encontrado niveles similares de variación intraespecífica en *R. suffruticosus* y *R. acetosella*, el patrón de distribución de esa variación dentro de las repeticiones en tándem de *R. acetosella* es específico, con la aparición de HORs (Navajas-Pérez, 2005), lo que podría indicar ciertas limitaciones del proceso de *molecular drive* en esta especie (Pons y Gillespie, 2003), quizá como consecuencia de la presencia de secuencias RAE180 solo en cantidades basales (Cuñaado *et al.*, 2007).

2. Las secuencias RAE180 se han amplificado enormemente dentro del genoma de las especies con determinación sexual de tipo XX/Y₁Y₂. Este proceso de amplificación ha ocurrido en un evento independiente del que ocurrió en la línea de especies tipo XX/XY. La amplificación del número de repeticiones de RAE180 en el cromosoma Y de las especies XX/Y₁Y₂ podría haber ocurrido recurrente e independientemente varias veces en cada nueva línea, originándose así cada especie XX/Y₁Y₂ actual. Si este fuera el caso, podría esperarse un patrón de amplificación diferencial con divergencia genética entre especies. Asumiendo la naturaleza no recombinante de los cromosomas Y, se esperaría también un elevado número de sitios polimórficos con un perfil específico de especie. Sin embargo, hemos encontrado niveles similares de amplificación (Tabla 3), y un número elevado de sitios polimórficos compartidos (59% de las posiciones nucleotídicas por repetición, Tabla 4) entre *R. acetosa* y *R. papillaris*. Estos datos apoyan la idea de que:

a) el principal evento de amplificación de la familia de ADN satélite RAE180 precedió a la divergencia entre estas dos especies, a pesar de la existencia de nuevos eventos de amplificación y reorganización subsiguientes (Navajas-Pérez *et al.*, 2009).

b) la mayoría de la variabilidad intraespecífica encontrada en cada especie es ancestral, originada antes de la divergencia entre las dos especies.

c) después de la separación entre estas dos especies, el proceso de *molecular drive* conducente a la evolución concertada ha fallado. De hecho, hay elevados niveles de variabilidad en las dos especies (19.9% en *R. acetosa* y 21.1% in *R. papillaris*; Tabla 3), y solo un sitio diagnóstico específico de una especie (Tabla 4). Estos resultados se explican sobre la base de la ausencia de recombinación en los cromosomas Y, un factor clave en el proceso de *molecular drive* (Navajas-Pérez *et al.*, 2005b). Pero también el patrón del locus autosómico de *R. acetosa* y *R. papillaris* se asemeja al comportamiento de la secuencia RAE180 ligada al cromosoma Y, un resultado inesperado ya que *a priori* cabría esperar un escenario evolutivo similar en los loci autosómicos en ambos tipos de especies, aquellas con determinación sexual XX/XY y aquellas con determinación XX/XY₁Y₂. Por contra, los loci autosómicos de *R. acetosa* y *R. papillaris* muestran elevados niveles de variabilidad intraespecífica (23,2% y 15,3% respectivamente, Tabla 3), sin diferencias fijadas entre *R. acetosa* y *R. papillaris* (Tabla 4). Estos niveles de amplificación similares (Tabla 3) junto con el elevado número de sitios polimórficos compartidos entre *R. acetosa* y *R. papillaris* (34% de las posiciones nucleotídicas por repetición, Tabla 3) de nuevo traen a colación que la variabilidad intraespecífica encontrada en cada una de las especies en el locus autosómico es ancestral, originada antes de la separación entre ambas especies, después de la cual, el proceso conducente a la evolución concertada también ha fallado en este caso. Esto podría explicarse también debido a las reducidas tasas de recombinación del locus autosómico de *R. acetosa* y *R. papillaris*. Podríamos asumir para explicarlo que el pequeño locus autosómico encontrado en estas dos especies tiene un comportamiento particular respecto a los procesos de homogeneización por expansión-contracción que deben conducir a la evolución concertada por el que la recombinación estaría reducida. La variabilidad puede permanecer durante largos periodos evolutivos en un genoma debido a la acción reducida de los mecanismos de intercambio no recíproco con lo que las diferentes variantes persistirían como una “biblioteca” (Mravinac *et al.* 2002; Meštrović *et al.* 2006) como ocurriría en el locus autosómico de especies XX/XY₁Y₂. A partir de una biblioteca de secuencias, la amplificación de distintas variantes podrían dar lugar más tarde a diferentes subfamilias de un ADN satélite particular como en el caso de RAYSI en *Rumex acetosa* (Navajas-Pérez *et al.*, 2006) o a un tándem amplificado de una de las variantes como hemos encontrado para el ADN satélite RAE180 en *R. suffruticosus*. Una explicación alternativa implicaría que parte de la variabilidad genética encontrada para RAE180 en autosomas de especies XX/XY₁Y₂ procede de los cromosomas Y por algún tipo de mecanismo de transferencia genética.

En conclusión, el análisis que hemos realizado demuestra el efecto de la localización, de la organización y del número de repeticiones en la evolución de una familia de ADN satélite. El proceso de *molecular drive* conduce a la evolución concertada de las repeticiones de una familia de ADN satélite. Los mecanismos moleculares de intercambio no recíproco son factores clave en el proceso de *molecular drive* que conducen a la homogeneización de los tándemes de secuencias repetidas dentro del genoma de una especie. Las familias de ADN satélite en los cromosomas Y no recombinantes muestran menores niveles de evolución concertada, de modo que aumenta la variabilidad intraespecífica sin que se incremente la divergencia interespecífica. Por contra, en los loci que se recombinan libremente ocurre evolución concertada de las secuencias, con diferenciación genética entre especies, como ocurre en los loci autosómicos de *R. acetosella* y *R. suffruticosus*. Sin embargo, parece que la evolución satélite de ADN también podría tener períodos evolutivos de estasis (enlentecidos) con variabilidad remanente en las secuencias por acción reducida de los mecanismos moleculares de intercambio no recíproco, como parece haber ocurrido en el locus autosómico de *R. acetosa* y *R. papillaris*, algo que podría depender del número de repeticiones y de los procesos que intervienen en su amplificación.

IV.2.4. Caracterización de las secuencias RAE180 y diferenciación genética entre las dos razas cromosómicas de *Rumex hastatulus*

La familia de ADN satélite RAE180 representa el 2×10^{-4} % del genoma de *R. hastatulus*, sin diferencias cuantitativas entre individuos de diferente sexo o raza, lo cual coincide con lo esperado según una localización autosómica de estas repeticiones. Suponiendo que el genoma de *R. hastatulus* (cuyo tamaño se desconoce) tiene un tamaño similar, aunque debe ser menor, al de *R. acetosa* (3234 Mb, Bennett y Leitch, 2004), este porcentaje supondría unas decenas de copias de repeticiones RAE180. Este es el menor número de copias RAE180 identificado en el genoma de las especies del género *Rumex* que, como hemos visto en la sección anterior, varían entre unas 200 y 7000 copias en los genomas de *R. acetosella* y *R. suffruticosus* (XX/XY) respectivamente, y alrededor de un millón de copias en la especie *R. acetosa* (XX/XY₁Y₂). Por tanto, como se ha descrito en otras especies del género *Rumex* con sistema cromosómico XX/XY (Cuñado *et al.* 2007), en el genoma de *R. hastatulus* las repeticiones RAE180 son autosómicas y están poco representadas. Este dato lo confirmamos mediante hibridación *in situ* fluorescente (FISH) con una sonda RAE180 específica de esta especie. Esta técnica no nos permitió identificar señales de hibridación en los cromosomas de *R. hastatulus*. Tampoco se observaron regiones heterocromáticas (mediante tinción con DAPI) en los cromosomas sexuales o en los autosomas de *R. hastatulus*.

Estos datos confirman que aunque *R. hastatulus* comparte la configuración cromosómica con *R. acetosa* y las otras especies con sistema XX/XY₁Y₂, ésta configuración debió surgir de forma independiente en ambos linajes (Smith 1969; Navajas-Pérez *et al.* 2005a). Además, a pesar del origen convergente del sistema XX/XY₁Y₂ en ambas líneas, los cromosomas sexuales de la raza de Carolina del Norte de *R. hastatulus* son todavía eucromáticos y no presentan acumulación de secuencias de ADN satélite, uno de los principales componentes de la heterocromatina. Una posible explicación a este hecho sería que el origen del sistema de determinación sexual XX/XY₁Y₂ de la raza de Carolina del Norte de *R. hastatulus* es reciente. De hecho, asumiendo una tasa media de cambio en el ADN nuclear de plantas es de 0,6% por sitio y por millón de años (Gaut, 1996), y utilizando los datos de distancia media corregida del ITS estimados por Navajas-Pérez *et al.* (2005a), la divergencia entre estas dos razas debió ocurrir hace aproximadamente unos 600.000 años. Esto contrasta con el origen del clado de *R. acetosa* y especies relacionadas, que se ha estimado en unos 12-13 millones de años (Navajas-Pérez *et al.* 2005a).

Se han aislado y secuenciado un total de 76 unidades repetitivas del genoma de diferentes individuos de la raza de Texas (43 de machos y 33 de hembras) y 58 repeticiones de la raza de Carolina del Norte (24 de machos y 34 de hembras). Los números de acceso de la EMBL para estas secuencias son FN561567 a FN561625 (Carolina del Norte) y FN563834 a FN563909 (Texas). El análisis de estas secuencias RAE180 demuestra la existencia de tres subfamilias de ADN satélite en ambas razas de *R. hastatulus*: TX1, TX2 y TX3 en la raza de Texas, y NC1, NC2 y NC3 en la raza de Carolina del Norte. Cada subfamilia RAE180 se caracteriza por la presencia de sitios diagnósticos, representado cada uno por una mutación particular compartida por todas las secuencias de la subfamilia, mientras que en el mismo sitio todas las secuencias de las otras subfamilias presentan un nucleótido diferente (Figura 5). En ambas razas, la divergencia entre subfamilias fue mayor que la variación dentro de cada subfamilia (Tablas 5 y 6), y los análisis filogenéticos agrupan a las secuencias según la subfamilia a la que pertenecen (Figuras 6 y 7).

```

#TX_M41 TATTGAATTATATCGATCATA-TGTGAAAAA-ATCGGCCCGTTCC-GCCAC--CGATACTCGAGAACGGTCAATATCA
#TX_M14 .....G.....G.....-T.....-ATT.GATCG.T.C.A.-AT.....A.....C.....
#TX_F65 .....G.....C.T.....TT.TCT.....ACAAT.GTTTG.TTCAAA.AAT.....T..A.....C.....

#TX_M41 TCGAA-TAGTT---AGCAATAGTGAAG---GAAGTTCAATGAAAAATAGTCGATATTCCACCTATGTCGATGACCCGGAA
#TX_M14 .T...-C.A.-CT...G.....ACGTGAAA.....G.....G.....C.....C....A.G.
#TX_F65 -T...A.....-CT.T.G...A...T---GTG.....ATA.T.GA...G..T...A.A..C....AAG.

#TX_M41 CTCGTTTTTGGTTAAATCTTTTTGTATATTTTGATCAA
#TX_M14 ...A.....CC.....G.....
#TX_F65 ...A.....CT..C...A....GT.....

A

#NC_F22 TAATGGATTTTCTCGATCATA-TT-GAAGAA-ATCGGCCTGTT-C-GGCCATCGATTCTCGAAACCAA-CAAATATCA
#NC_F14 ..TG....A.G.....AT-A.T.G.A.-.ATT.GA.CG.T.C.T.....A.....GG-.C.....
#NC_M41 ..T.....A..T.....TG..TTT.A.-CAAT.GG.TG.TTCAAAA.....A..T...-..G-.C.....

#NC_F22 CTTAA-CAGTT-CTATCGATAARTGAAG---GAAGTTTGGTGAAAAATAGTCGATATTGACCTATATCAATGACCCGAGA
#NC_F14 T-G...-A..C...G.....G.ACGTGAAA.....C.A.....T...GC.G.C....G..
#NC_M41 -.G..AT....-.....T---...C..G.....GA.T.GAT....T....AG.C....A-.

#NC_F22 CTCGCTTTTTCGTTAAATCTTTTTGTATATTTTGATCAA
#NC_F14 ...AT.....C.....C.....
#NC_M41 ...AT.....T..CC.ACC...G..TA....CCCC..

B

```

Figura 5. Alineamiento múltiple de secuencias representativas RAE180 de las dos razas de *Rumex hastatulus*. Subfamilia TX1/NC1: TX_M41 y NC_F22; Subfamilia TX2/NC2: TX_M14 y NC_F14; Subfamilia TX3/NC3: TX_F65 y NC_M41. Se muestra completa la primera secuencia del alineamiento, mientras que en el resto se muestran las posiciones nucleotídicas variables que corresponden en este caso a posiciones diagnósticas específicas de cada subfamilia. Los guiones representan deleciones con respecto a la secuencia de referencia.

Por tanto, teniendo en cuenta los sitios diagnóstico, las distancias genéticas y la filogenia, podemos adscribir las secuencias RAE180 de cada especie a subfamilias diferentes que, en un sentido amplio, si nos tomamos la licencia de denominar repeticiones “ortólogos” a todas aquellas derivadas de la amplificación de un mismo tipo de repetición en un único o varios eventos de amplificación, podríamos tratar a las diferentes subfamilias como grupos “parálogos” compuestos por repeticiones ortólogos de la familia RAE180. La relación ortóloga entre las subfamilias RAE180 de las dos razas de *R. hastatulus* se constata por la agrupación en clados de estas repeticiones en el árbol filogenético en el que se incluyen todas las secuencias estudiadas en ambas razas (Figura 8), así como en la comparación de éste árbol con los árboles de las secuencias RAE180 de cada raza (Figuras 6 y 7) y el árbol filogenético de las repeticiones RAE180 de todas las especies dioicas de *Rumex* analizadas (Figura 2). En la Figura 8 pueden observarse 3 grupos de repeticiones ortólogos: TX1/NC1, TX2/NC2 y TX3/NC3 (Figura 8). De hecho, las secuencias de las diferentes razas que pertenecen a distintas subfamilias comparten sitios diagnósticos (Figura 9). Las distancias genéticas entre las razas de Texas y Carolina del Norte (Tabla 7) fueron más altas para la subfamilia 1 (0,252) y más bajas para la subfamilia 2 (0,050), mostrando la subfamilia 3 un valor intermedio (0,179).

Tabla 5. Análisis de la variación entre cada subfamilia RAE180 y de la divergencia entre subfamilias en la raza de Texas de *R. hastatulus*

Raza	Variación intrafamiliar		
	TX1	TX2	TX3
Texas	0,019	0,004	0,036
	Divergencia interfamiliar		
TX1/TX2	0,159		
TX1/TX3	0,353		
TX2/TX3	0,246		

Tabla 6. Análisis de la variación entre cada subfamilia RAE180 y de la divergencia entre subfamilias en la raza de Carolina del Norte de *R. hastatulus*

Raza	Variación intrafamiliar		
	NC1	NC2	NC3
Carolina del Norte	0,181	0,069	0,200
	Divergencia interfamiliar		
NC1/NC2	0,256		
NC1/NC3	0,343		
NC2/NC3	0,339		

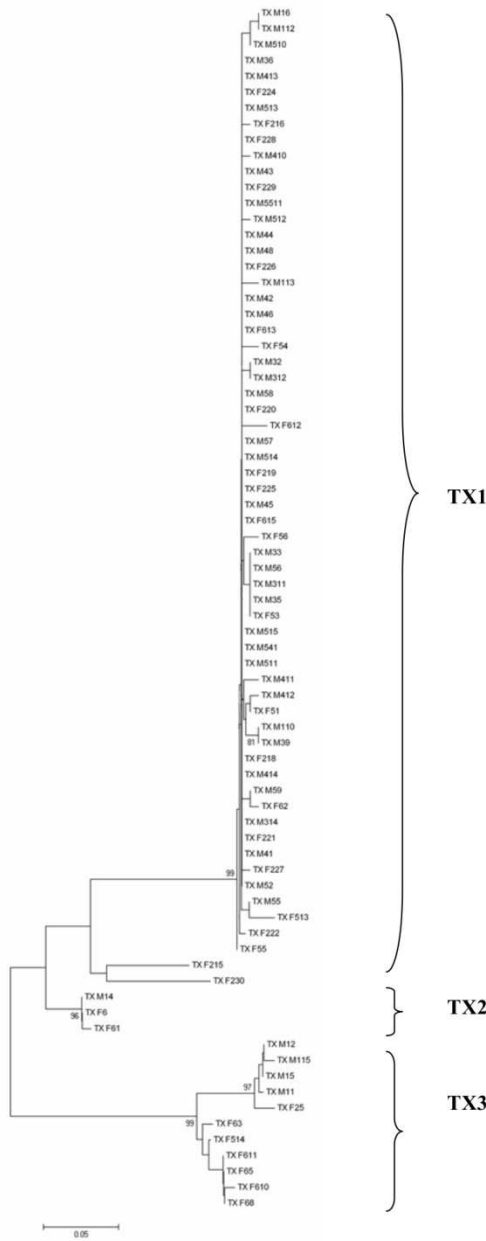


Figura 6. Árbol de secuencias de las subfamilias RAE180 en la raza de Texas (TX) de *R. hastatulus* (M) macho, (F) hembra. El primer número de la secuencia corresponde al número del individuo analizado y los siguientes al número de repetición según un código establecido durante la clonación. Los números sobre los nodos representan valores de *bootstrap*.

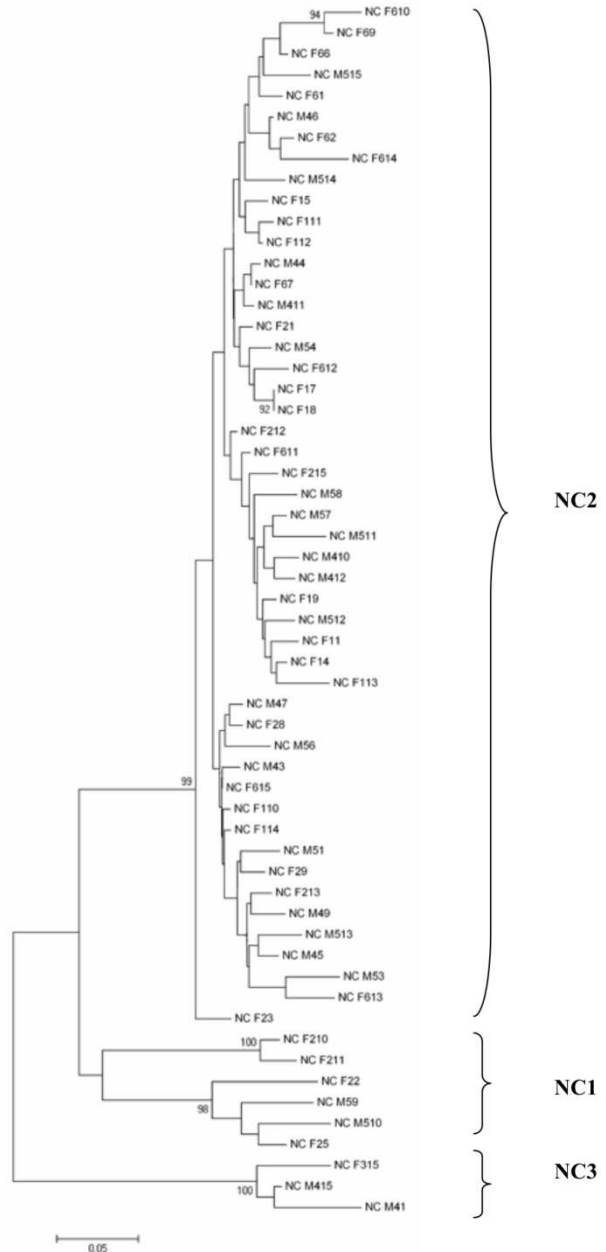


Figura 7. Árbol de secuencias de las subfamilias RAE180 en la raza de Carolina del Norte (NC) de *R. hastatulus*. (M) macho, (F) hembra. El primer número de la secuencia corresponde al número del individuo analizado y los siguientes al número de repetición según un código establecido durante la clonación. Los números sobre los nodos representan valores de *bootstrap*.

Tabla 7. Distancias genéticas entre las razas de Texas y Carolina del Norte de *R. hastatulus*

	Divergencia
Entre secuencias ortólogas	
TX1/NC1	0,252
TX2/NC2	0,050
TX3/NC3	0,179
Entre secuencias más comunes para cada raza	
TX1/NC2	0,190

No obstante, la abundancia de cada subfamilia RAE180 es diferente entre razas. En la raza de Texas, la subfamilia TX1 es la más abundante (82% de las secuencias repetidas analizadas), mientras que solo el 4% de las secuencias RAE180 pertenecen a la subfamilia TX2, y el 14% pertenecen a la subfamilia TX3. En la raza de Carolina del Norte, la subfamilia NC2 es la más abundante ya que supone el 85% de las secuencias repetidas, alrededor del 10% de las secuencias pertenecen a la subfamilia NC1 y el 5% a la subfamilia NC3. También los análisis filogenéticos de las secuencias RAE180 de las dos razas de *R. hastatulus* reflejan la existencia de dos clados principales (Figura 8) correspondiendo cada uno de ellos a secuencias de la raza de Texas y de la raza de Carolina del Norte. Este resultado es debido al hecho de que más del 80% de las repeticiones de cada raza pertenecen a dos subfamilias diferentes: la subfamilia 1 (TX1) en el caso de la raza de Texas y la subfamilia 2 (NC2) en el caso de la raza de Carolina del Norte, con un valor de diferenciación genética entre TX1 y NC2 del 19% (Tabla 7). El reemplazamiento de una variante monomérica por otra en diferentes especies (o poblaciones como en las razas de *R. hastatulus*) es una característica común del ADN satélite como consecuencia de la evolución dinámica de este tipo de secuencias. Además, la comparación de las secuencias ortólogas de las dos razas cromosómicas muestran un nivel importante de diferenciación genética (Tabla 7). Esta diferenciación genética del ADN satélite RAE180 entre las dos razas cromosómicas de *R. hastatulus* es comparable a la encontrada para el ADN satélite RAE730 de otras especies de *Rumex* (Navajas-Pérez *et al.* 2005b), o la familia de ADN satélite EcoRI de espáridos (De la Herrán *et al.* 2001). Así, el ADN satélite RAE180 resulta ser un marcador molecular útil para la identificación genética de individuos de diferentes razas.


```

#TX_M41 TATTGAATTATATCGATCATA-TGTGAAAAA-ATCGGGCCGTTCC-GCCAC--CGATACTCGAGAACGGTCAATATCAT
#NC_F22 ..A..G...T.C.....-T...G...-.....T...-..G.C-AT...T...A.C.AACA.....C
#TX_M14 .....G.....G.....-T.....-ATT.GATCG.T.C.A..-AT.....A.....C.....
#NC_F14 .....G.....G.....AT-AT..G.....-ATT.GATCG.T.C.T.C-AT.....A.C.....C.....
#TX_F65 .....G.....C.T.....TT.TCT.....ACAAT.GTTTG.TTCAAA.AAT.....T..A.....C.....-
#NC_M41 .....G.....C.T.....TG.T.TT.....-CAAT.GGTTG.TTCAAA.-AT.....T..A..-A.C.....-

#TX_M41 CGAA-TAGTT---AGCAATAGTGAAG----GAAGTTC AATGAAAATAGTCGATATTCACCTATGTCGATGACCGGAACT
#NC_F22 TT..-C....-CT.T.G...A.....-...TGG.....G.....A..A.....AG...
#TX_M14 T...-C.A..-CT..G....ACGTGAAA.....G.....G.....C...C...A.G...
#NC_F14 -...-C.A..CCT..G....ACGTGAAA.....G.....G.T...C...C...G...
#TX_F65 T...A.....-CT.T.G...A...T----...GTG.....ATA.T.GA...G..T...A.A..C...AAG...
#NC_M41 T...A.....-CT.T.G...A...T----...GGG.....GA.T.GAT..G..T...A.A..C...A-G...

#TX_M41 CGTTTTGGTTAAATCTTTTTGTATATTTGATCAA
#NC_F22 ..C....C.....TA.....
#TX_M14 .A.....CC.....G.....
#NC_F14 .A.....CC.....TC.....
#TX_F65 .A.....CT..C...A...GT.....
#NC_M41 .A.....CT..CC.A.C...G..TA.....CCCC...

```

Figura 9. Secuencia consenso de las tres subfamilias RAE180 en las razas de Texas y Carolina del Norte de *R. hastatulus*. Se muestra completa la primera secuencia del alineamiento, correspondiente a la subfamilia 1 de la raza de Texas (TX1), mientras que en el resto se muestran las posiciones nucleotídicas variables que corresponden en este caso a posiciones diagnósticas específicas de cada subfamilia. Los guiones representan deleciones con respecto a la secuencia de referencia.

De acuerdo con la hipótesis de la biblioteca (Fry and Salser 1977; Meštrović *et al.* 1998, 2006; Mravinac *et al.* 2002; Pons *et al.*, 2004), táxones relacionados pueden compartir una biblioteca de diferentes secuencias de ADN satélite (diferentes familias de ADN satélite pero también diferentes variantes monoméricas o subfamilias de una familia de ADN satélite), que pueden amplificarse diferencialmente en cada taxón, con el subsiguiente reemplazamiento de una secuencia por otra variante en diferentes especies (o poblaciones). Los resultados descritos en *R. hastatulus* apoyan esta hipótesis. La presencia de tres subfamilias RAE180 en las dos poblaciones analizadas sugiere que su existencia precedió al aislamiento y la diferenciación de las dos razas cromosómicas de *R. hastatulus*. Incluso presentando un número de secuencias repetidas aún bajo (unas pocas decenas según nuestros datos), se ha producido un proceso de amplificación diferencial en cada una de las razas de estas especies para las diferentes subfamilias RAE180, lo que concuerda con la hipótesis de la biblioteca.

Por otro lado, la dinámica evolutiva de las secuencias RAE180 en *R. hastatulus* permite aclarar el resultado obtenido tras el análisis comparativo de las repeticiones RAE180 aisladas del total de especies dioicas del género *Rumex* analizadas (apartado IV. 2.2 de esta memoria). El árbol filogenético de las secuencias RAE180 (Figura 2) no se ajusta a la filogenia de las especies dioicas de *Rumex* (Figura 1 Introducción, y Figura 10). Por el contrario, el árbol apoya de nuevo la hipótesis de la biblioteca para la evolución de las secuencias del ADN satélite RAE180 en el género *Rumex*. Según esto, podría asumirse la existencia de un conjunto de variantes RAE180 divergentes en el genoma del ancestral común de las especies dioicas de *Rumex*, a partir del cual diferentes variantes se amplificaron en diferentes linajes. La variante TX1/NC1, predominante en la raza de Texas de *R. hastatulus* está estrechamente relacionada con las secuencias RAE180 de *R. acetosella*, apareciendo como el tipo más antiguo de

secuencias con respecto a la variante TX2/NC2. La variante derivada, TX2/NC2, está reemplazando a la variante TX1/NC1 en la raza de Carolina del Norte de *R. hastatulus*. En cualquier caso, la familia RAE180 tiene localización autosómica y se encuentra pobremente representada en estas dos especies. Por otro lado, habría una variante específica amplificada en *R. suffruticosus* (Cuñado *et al.* 2007; Navajas-Pérez *et al.* 2009b). En esta especie, las repeticiones RAE180 están muy homogeneizadas. Finalmente, hay una variante relacionada con la de *R. suffruticosus* (Figura 2) que comparte con las tres especies XX/XY₁Y₂ (*R. acetosa*, *R. papillaris* and *R. intermedius*). La evolución de las repeticiones RAE180 en la línea XX/XY₁Y₂ es particular ya que, como hemos visto en la sección anterior, existen niveles altos de variabilidad intraespecífica en estas especies y bajas tasas de evolución y homogeneización de las secuencias. La mayor parte de la variabilidad intraespecífica encontrada en cada una de estas especies sería variabilidad ancestral, y la naturaleza no recombinante de los cromosomas Y explicaría cómo los procesos que conducen a la evolución concertada han fallado antes y después de la diversificación de las especies XX/XY₁Y₂ (Navajas-Pérez *et al.* 2005b) tal como se ha discutido en la sección anterior.

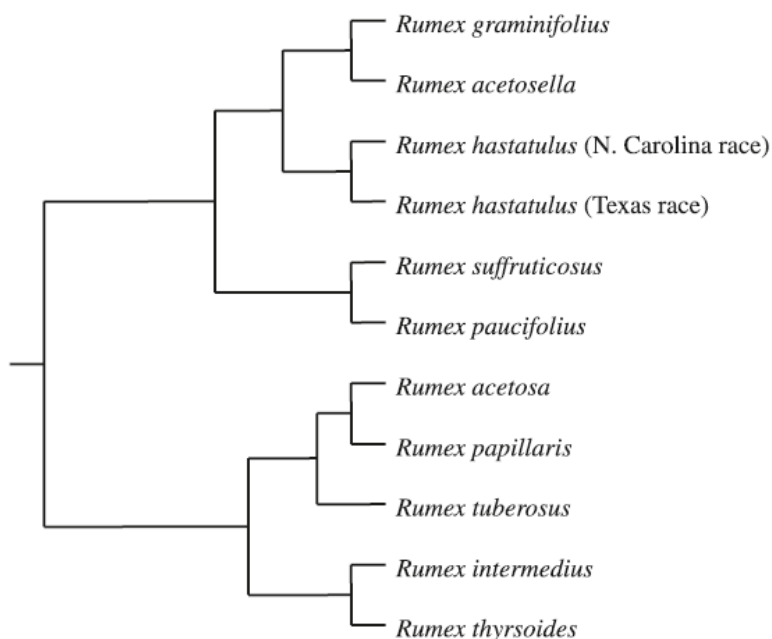


Figura 10. Relaciones filogenéticas entre especies dioicas del género *Rumex*. De acuerdo a los datos de Navajas-Pérez *et al.*, 2005a.

*Análisis del ADN satélite HinfI en la subfamilia
Carduoideae (Asteraceae)*

V.1. Introducción

V.1.1. Familia Asteraceae

La familia de plantas vasculares Asteraceae Bercht y J. Presl (1820) (también conocida como Compositae) se encuentra enmarcada en el Orden Asterales. Es la familia de fanerógamas con mayor diversidad, representando aproximadamente el 10% de todas las plantas con flores (Funk *et al.*, 2005) y siendo la familia de angiospermas más grande en términos de número de especies (Anderberg *et al.*, 2007). La última revisión de Steven (2012, www.mobot.org/MOBOT/research/APweb/) señala la existencia de 1.620 géneros y 23.600 especies. Sorprendentemente, aunque es la familia más grande de plantas con flores su origen es relativamente reciente, habiéndose diversificado no hace más de 50 millones de años (Funk *et al.*, 2005). Por esta razón, las Asteráceas ofrecen una excelente oportunidad para comprender el proceso adaptativo en grupos de plantas con una radiación relativamente reciente a nivel global.

Su distribución es subcosmopolita, pues está representada en todos los continentes excepto en la Antártida y la isla de Groenlandia (Funk *et al.*, 2005; Stevens, 2012). El éxito de la expansión por el mundo de esta familia de angiospermas, cultivadas desde la Antigüedad (Dempewolf *et al.*, 2008), se debe a la presencia de metabolitos secundarios, la característica morfología de su inflorescencia y a la plasticidad en sus hábitats (Carlquist, 1976; Hendry, 1996; Stuessy y Garver, 1996). La amplia diversidad de hábitats colonizados por las compuestas comprende desde el nivel del mar hasta el límite altitudinal de la vegetación, cuya mayor diversidad se halla en las zonas tropicales, subtropicales y templadas, abundando en regiones de montaña y en áreas áridas y semiáridas. Respecto a su morfología, la principal característica que las define es la disposición de sus flores reunidas en inflorescencias racemosas denominadas capítulos. Cada flor está rodeada de una a varias hileras de brácteas involucrales, sobre un receptáculo común de la inflorescencia que termina en el escapo o rama florífera (Weberling, 1989).

La clasificación más reciente del grupo ha sido propuesta por Panero y Funk (2002, 2008), y está basada en el análisis de secuencias cloroplastidiales. En este estudio se identifican 12 subfamilias estrictamente monofiléticas y 40 tribus (Panero y Funk, 2008) (Figura 11). No obstante, las relaciones entre los linajes principales no están totalmente resueltas. Parece ser que el origen de la familia tuvo lugar en Sudamérica a principios del Cenozoico, en el periodo Terciario, desde el final del Paleoceno a la mitad del Eoceno, habiendo tenido un gran éxito en términos de diversificación de especies mediante la dispersión fuera de la frontera sudamericana, logrando expandirse al resto del mundo.

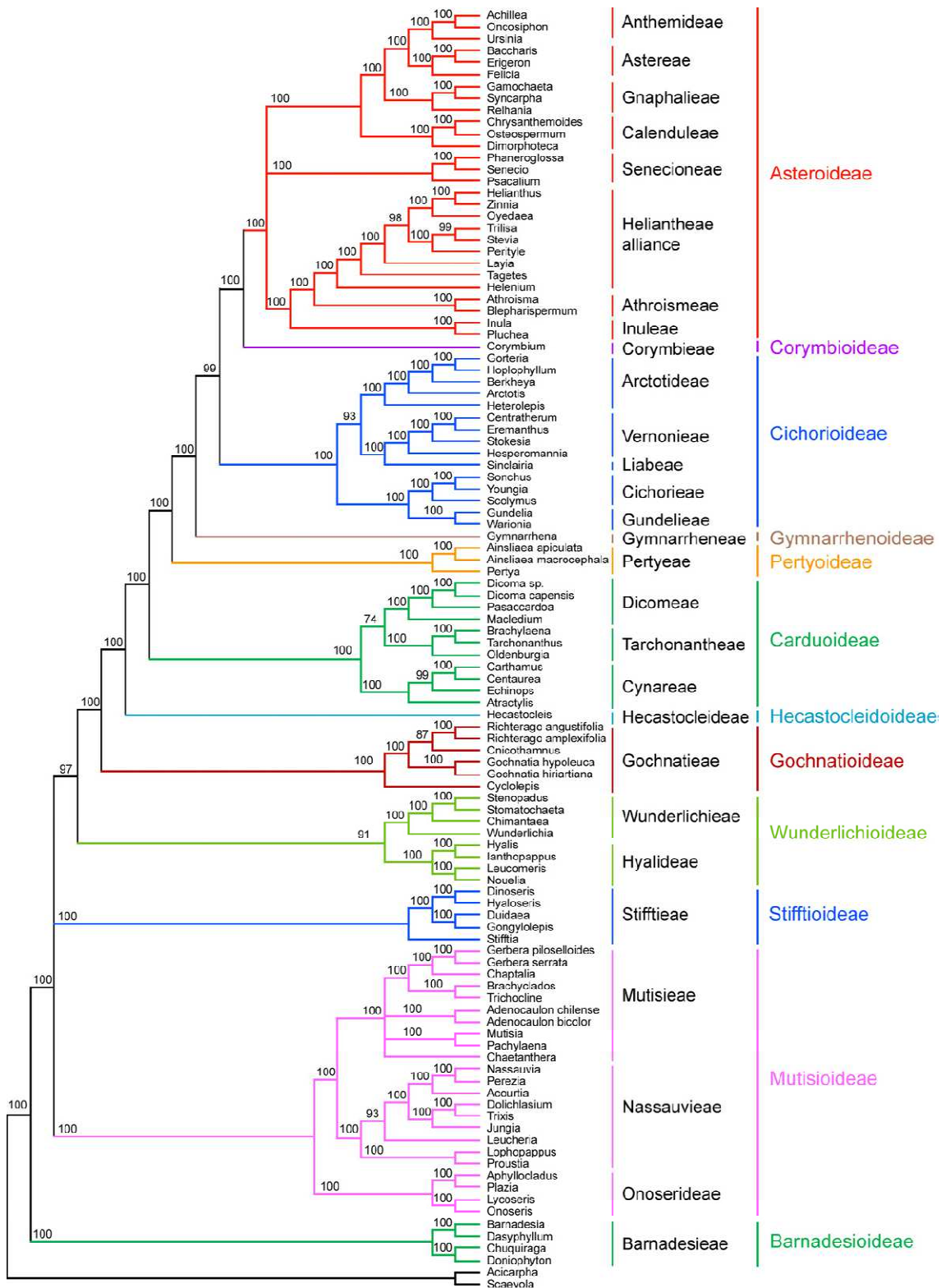
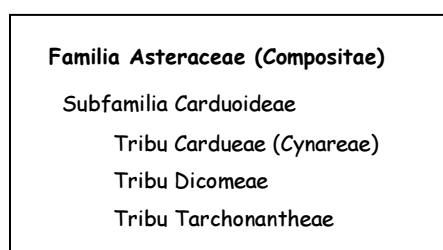


Figura 11. Árbol filogenético de la familia Asteraceae construido mediante inferencia bayesiana empleando 10 loci cloroplastidiales como marcadores. Tomado de Panero y Funk (2008).

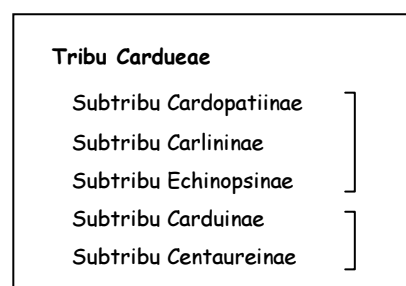
V.1.2. Tribu *Cardueae* Cass.

Dentro de las *Asteráceas*, en este trabajo hemos analizado las especies de la Tribu *Cardueae* que pertenece a la subfamilia *Carduoideae*. A nivel morfológico, esta subfamilia, no presenta ninguna sinapomorfía que la defina pero casi todos sus miembros tienen estilos papilosos con papilas mayoritariamente confinadas en un anillo bajo el estigma. No obstante, la subfamilia constituye un grupo monofilético (Panero y Funk 2002; García-Jacas *et al.*, 2002; Susanna *et al.*, 2006).

Actualmente, están reconocidas tres tribus en *Carduoideae*: *Cynareae* o *Cardueae*, *Dicomeae* y *Tarchonantheae*. La tribu *Cardueae* contiene más del 90% de la diversidad de especies del grupo. Se han identificado 2.400 especies repartidas en 72 géneros (Susanna y García-Jacas, 2007, 2009), destacando *Centaurea* (250 especies), *Cousinia* (600), *Saussurea* (300), *Cirsium* (250), *Jurinea* (200), *Echinops* (120), *Carduus* (90), *Serratula* (70) y *Onopordum* (60) (Una revisión completa se puede encontrar en <http://www.mobot.org/MOBOT/research/APweb/>). Esta tribu está ampliamente distribuida pero la mayor diversidad de géneros y especies se centran en el Hemisferio Norte, especialmente en Europa y Asia central existiendo baja representación en América y Australia (Hellwig, 2004). Las filogenias basadas en ADN nuclear (Susanna *et al.*, 1995) así como las basadas en ADN nuclear y cloroplastidial (García-Jacas *et al.*, 2002; Susanna *et al.*, 2006) confirman su monofilia.



En la actualidad, la tribu *Cardueae* se divide en cinco grupos taxonómicos con categoría de subtribu: *Cardopatiinae*, *Carlininae*, *Echinopsinae*, *Carduinae* y *Centaureinae* (Susanna *et al.*, 1995; García-Jacas *et al.*, 2002; Susanna *et al.*, 2006, Susanna and García-Jacas, 2009). La clasificación de la tribu es complicada debido a la enorme diversidad morfológica y al gran número de especies, incluyendo la controversia dificultades en la delimitación de la tribu, en la delimitación de los límites entre las subtribus, y en la delimitación de los géneros (revisado en Susanna y García-Jacas, 2009). Los análisis moleculares más recientes identifican a las subtribus *Cardopatiinae*, *Carlininae* y *Echinopsinae* como grupos basales que están menos representados en cuanto a número de especies, mientras que *Carduinae* y *Centaureinae* constituyen grupos derivados más diversificados, representando



más del 90% de las especies de esta tribu (Susanna y García-Jacas, 2009). Las subtribus *Cardopatiinae*, *Carlininae*, *Echinopsinae* y *Centaureinae* son grupos naturales mientras que *Carduinae* es un grupo parafilético con respecto a *Centaureinae* (García-Jacas *et al.* 2002; Susana *et al.*, 2006 Andenberg *et al.*, 2007; Susanna and García-Jacas, 2009). En particular, el grupo *Arctium-Cousinia-Saussurea-Jurinea* de *Carduinae* es un grupo hermano de *Centaureinae* (Hellwig 1996, Häffner y Hellwig 2000; García-Jacas *et al.* 2002; Susana *et al.*, 2006). Nuestros estudios se han centrado en primer lugar en la subtribu *Centaureinae* que se ha estudiado en mayor profundidad al menos en cuanto al número de géneros y especies analizadas y representa el grupo mejor estudiado desde el punto de vista filogenético y evolutivo. Posteriormente, se han extendido nuestros estudios al análisis de *Carlininae*, *Echinopsinae* y *Carduinae*, por subtribus y globalmente.

V.1.3. Subtribu *Centaureinae*

La familia *Asteraceae*, en general, y la subtribu *Centaureinae*, en particular, se consideran un buen ejemplo para estudiar la adaptación en las radiaciones recientes a escala global. Se encuentra distribuida en la región florística Mediterráneo-Turquestánica (Dittrich 1977) aunque existen representantes en África tropical (*Plectocephalus*, *Ochrocephala*), en el Norte y Sur de América (*Plectocephalus*), Este de Asia (*Serratula*) y Australia (*Rhaponticum*). La Cuenca Mediterránea es el centro de mayor diversidad de especies (Wagenitz 1986; Meusel y Jäger 1992). Esta subtribu comprende 31 géneros con aproximadamente 800 especies siendo *Centaurea*, *Colymbada*, *Psephellus*, *Klasea*, *Rhaponticoides*, *Cyanus*, *Rhaponticum*, *Carthamus*, *Tricholepis* y *Volutaria* los géneros más representados. Existe una gran variedad de formas de vida, desde arbustos anuales de mayor o menor tamaño a hierbas perennes que son las más comunes. A nivel morfológico, los géneros de esta subtribu tienden a desarrollar cabezuelas con las flores marginales estériles que, en los más primitivos, persisten como estaminoides (Wagenitz y Hellwig, 1996). Los caracteres filogenéticos que diferencian mejor los principales grupos de la subtribu han sido el tipo grano de polen descrito por Wagenitz (1955) y el número básico cromosómico, caracterizado por disploidía, estando ambos relacionados.

Respecto al número cromosómico, se observa una tendencia a la reducción del mismo (García-Jacas *et al.*, 1996; Hellwig *et al.*, 1996). La diversidad en el número básico de cromosomas oscila entre $n=16$ en *Volutaria* y $n=7$ en *Centaurea*. Los táxones con números cromosómicos superiores a $2n=24$ se consideran los más primitivos de la subtribu e, igualmente, presentan polen de alguno de los tipos más primitivos. Por otro lado, los táxones con número cromosómico $2n=24$ o inferior son considerados más evolucionados, con granos de polen de tipos más evolucionados (Susanna *et al.*, 1995; Wagenitz y Hellwig, 1996).

Los géneros que integran la subtribu Centaureinae y habitan en el Mediterráneo tienen un origen relativamente reciente, destacando los más recientes por un mayor número de especies. Es decir, el grupo más antiguo es menos diverso que el de origen más reciente. Se desconoce si se debe a que el grupo más antiguo nunca llegó a ser tan diverso como el de origen más reciente o si quizás en algún momento se produjo una disminución en el número de especies (Hellwig, 2004). Durante la historia geológica relativa a los últimos treinta

millones de años, en el continente europeo se han sucedido procesos cladogenéticos, en concreto dos radiaciones adaptativas, claves en la aparición y expansión de la subtribu Centaureinae. Estas dos radiaciones han generado los dos grandes grupos de la subtribu, en momentos geológicos distintos. Uno de ellos incluye los géneros basales y el otro al grupo de especies derivadas (García-Jacas *et al.*, 2001; Hellwig, 2004). Parece ser que la primera radiación ocurrió en el periodo Terciario, entre el final del Oligoceno y el Mioceno, dando lugar a los grupos basales, mientras que los grupos derivados se originaron en una segunda radiación en el periodo Cuaternario entre el Plioceno y el Pleistoceno (Hellwig, 2004).

El clado derivado, segunda radiación, engloba a la mayoría de las especies actualmente clasificadas dentro del género *Centaurea* y del complejo *Carthamus*. El género *Centaurea* incluye tres subgéneros: *Acrocentron*, *Centaurea* (anteriormente denominado como grupo *Jacea*), y *Cyanus* (Susanna y García-Jacas, 2009). El complejo *Carthamus* incluye a los géneros *Carduncellus*, *Carthamus*, *Femeniasia* y *Phonus* (García-Jacas *et al.*, 2001). El resto de los géneros constituyen los grupos basales (primera radiación): *Cheirolophus*, *Crupina*, *Klasea*, *Psephellus*, *Rhaponticoides* (anteriormente *Centaurea sensu stricto*), *Rhaponticum*, *Serratula*, o *Volutaria*, entre otros (García-Jacas *et al.*, 2001).

Aunque ambos grupos, el basal y el derivado, están presentes en la región Mediterránea, sólo el grupo derivado tiene un número elevado de especies (Hellwig, 2004). En los diferentes linajes de la subtribu, se observan diferentes modos de evolución que se ven reflejados en los caracteres morfológicos y en las secuencias de ADN. Hellwig (2004) describió la historia evolutiva de los géneros de la subtribu distribuidos en la cuenca Mediterránea relacionándola con su historia geológica. Según Hellwig (2004), los representantes de cada uno de los

Subtribu Centaureinae	
1ª Radiación (Grupos basales)	2ª Radiación (Grupos derivados)
Gr. <i>Rhaponticum</i>	<i>Centaurea</i>
<i>Plectocephalus</i>	<i>Centaurea (Jacea)</i>
Gr. <i>Amberboa</i>	<i>Acrocentron</i>
Gr. <i>Klasea</i>	<i>Cyanus</i>
<i>Psephellus</i>	Complejo <i>Carthamus</i>
<i>Crupina</i>	<i>Carthamus</i>
Gr. <i>Rhaponticoi-</i>	<i>Phonus</i>
<i>des/Cheirolophus</i>	<i>Carduncellus</i>
	<i>Femeniasia</i>
	Gr. <i>Colymbada</i>

diferentes clados, por lo general, crecieron simpátricamente pero tomando finalmente diferentes rutas evolutivas. El elevado número de nichos ecológicos han permitido la coexistencia de múltiples especies que han evolucionado dentro de la subtribu *Centaureinae*, por lo tanto, los diferentes clados simpátricos han seguido distintos procesos evolutivos (Hellwig, 2004). Además, los procesos de expansión y reducción de hábitats apropiados han favorecido la divergencia de distintos grupos originando especies alopátricas vicariantes. Los grupos que contienen especies perennes presentan un menor nivel de especiación por aislamiento geográfico, a menudo precedida por variación morfológica clinal y reticulación. En el caso de los grupos que mayoritariamente contienen especies anuales por el contrario, la especiación por lo general es alopátrica pero en algunos casos la especiación ha sido simpátrica (Hellwig, 2004).

Teniendo todo esto en cuenta, la subtribu *Centaurinae* parece ser un grupo modelo excelente para analizar el ADN satélite y la influencia que la biología y/o la evolución de sus diferentes linajes han tenido en la evolución de esta clase de ADN repetitivo.

V.1.4. Géneros estudiados de la subtribu *Centaureinae*

El género *Centaurea* es el más diversificado, con un total de unas 250 especies (Susanna y García-Jacas, 2007) de las cuáles sobre un centenar de ellas se localizan en la Península Ibérica. Se trata de un género cuya extraordinaria diversidad y complejidad taxonómica ha llevado a constantes revisiones sobre su taxonomía y sistemática (Bremer, 1994; Susanna *et al.*, 1995). De hecho, se ha demostrado el carácter parafilético del género en su concepción clásica (Susanna y García-Jacas 1995, 2000). La división infragenérica de *Centaurea* incluye distintos grupos que actualmente son reconocidos como subgéneros (Susanna y García-Jacas, 2009): *Centaurea*, *Acrocentron* y *Cyanus*. Dentro del género *Centaurea* las hibridaciones interespecíficas y la evolución reticulada están ampliamente documentadas especialmente en el subgénero *Centaurea* (Suárez-Santiago *et al.*, 2005, 2007), pero también en el subgénero *Acrocentron* (Font *et al.* 2002, 2009).

Una clasificación clásica de *Centaurea* en sentido amplio, *Centaurea sensu lato*, normalmente incluía lo que en ese momento se consideraba un subgénero denominado *Centaurea* que se dividía en varias secciones entre las que se encontraba la sección *Centaurea* que, a su vez, se subdividía en cinco subsecciones (Agababian, 1997). Una de ellas, *Centaurea* subsección *Ruthenicae*, integraba las especies *Centaurea alpina* y *C. linaresii* (Agababian, 1997). Sin embargo, los análisis filogenéticos realizados por García-Jacas *et al.* (2001), confirman que *Centaurea* sección *Centaurea* está compuesta en realidad por especies muy alejadas del resto de especies del género ocupando una posición basal. Los datos morfológicos también coinciden con esta separación puesto que estas especies comparten diversas

sinapomorfias, no compartidas por el resto de las especies del género, como el tipo de grano de polen, el tipo de vilano doble y el número básico cromosómico de $x=15$ (Wagenitz 1955; Agababian y Goukasian 1994; Wagenitz y Hellwig 1996; Eren, 2007). En la actualidad, esta sección se ha separado del género *Centaurea* y ha adoptado la categoría de género denominándose *Rhaponticoides* (Greuter, 2003; Greuter *et al.*, 2005). Está compuesto de unas 32 especies distribuidas desde Portugal y Marruecos hasta Mongolia (Hellwig, 2004). La mayoría de las especies son endémicas de áreas pequeñas o se distribuyen en áreas disyuntas (Agababian, 1997; Hellwig, 2004). Además, el género *Rhaponticoides* se diversificó en la primera radiación. Por tanto, constituye un género basal.

El complejo *Carthamus* L. está compuesto por 50 especies circunmediterráneas. La delimitación subtribal del complejo *Carthamus* ha experimentado frecuentes modificaciones según los caracteres morfológicos empleados en las diversas clasificaciones. La delimitación genérica del complejo también ha ido cambiando en función de los diferentes autores. Actualmente, el complejo incluye los géneros *Carduncellus*, *Carthamus*, *Femeniasia*, y *Phonus* (Vilatersana *et al.*, 2000; García-Jacas *et al.*, 2001). Según Vilatersana *et al.* (2000) existe una aparente segregación entre las especies anuales del este del Mediterráneo pertenecientes al género *Carthamus* y las perennes del oeste de los géneros *Carduncellus*, *Phonus* y *Femeniasia*. Es decir, hay una relación directa entre la biogeografía de los géneros y la clasificación basada en ITS.

Géneros analizados subtribu Centaureinae

Familia Asteraceae (Compositae)

Subfamilia Carduoideae

Tribu Cardueae

Subtribu Centaureinae

Complejo Carthamus

Género *Carduncellus*

Género *Carthamus*

Sección *Carthamus*

Sección *Atractylis*

Género *Femeniasia*

Género *Phonus*

Género *Centaurea*

Subgénero *Centaurea* (Jacea)

Sección *Willkommia*

Sección *Acrolophus/Phalolepis*

Subgénero *Acrocentron*

Subgénero *Cyanus*

Género *Cheirilopus*

Género *Crupina*

Género *Klasea*

Género *Plectocephalus*

Género *Rhaponticoides*

Género *Stizolophus*

Género *Serratula*

Grupo *Rhaponticum*

Género *Acroptilon*

Género *Callicephalus*

Género *Centaurothamus*

Género *Myopordon*

Género *Ochrocephala*

Género *Oligochaeta*

Género *Rhaponticum*

Grupo *Volutaria* (Amberboa)

Género *Volutaria*

El género *Carthamus* compuesto por 25 especies (Keil, 2006; Vilatersana *et al.*, 2007; Sassanuma *et al.*, 2008) se subdivide en dos secciones: *Carthamus* y *Atractylis* (Vilatersana *et al.*, 2000a, 2005; Bowles *et al.*, 2010). Por un lado, la sección *Carthamus*, que coincide con la descrita por Hanelt (1963) con la misma denominación, integra la especie de mayor relevancia, *Carthamus tinctorius* comúnmente conocida como cártamo, con un gran interés económico (Mündel *et al.*, 2004). Por otro lado, la sección *Atractylis* que consta, entre otras de la especie *Carthamus lanatus*. Se plantea la posibilidad de que esta especie tenga un origen híbrido y, actualmente, se haya naturalizado como hierba nociva.

Curiosamente, *Phonus*, cuyas especies han sido segregadas taxonómicamente del género *Carthamus* está más próximo a *Carduncellus* que a *Carthamus*, a pesar de su diferente biología, hábitats e historia biogeográfica (Vilatersana *et al.*, 2000).

Respecto al género *Carduncellus*, Vilatersana *et al.* (2000) acepta la incorporación propuesta por Hanelt (1963) y Rivas Goday y Rivas Martínez (1967) de *C. caeruleus*, *C. dianius* y *C. calvus* en el grupo *Caeruleus* de este género con *C. cuatrecasasii*, *C. hispanicus* y *C. mitissimus*, desapareciendo el género *Lamottea* como grupo independiente.

Los principales clados basales de la subtribu *Centaureinae* han sido establecidos, pero algunos géneros no han podido ser claramente clasificados en ningún grupo. Los análisis filogenéticos moleculares indican que estos géneros constituyen un linaje antiguo dentro de la subtribu, lo que sugiere que la divergencia entre todos estos géneros es antigua (García-Jacas *et al.*, 2001; Susanna and García-Jacas, 2009). Sin embargo, análisis filogenéticos recientes sugieren considerar las entidades informales como grupos naturales, lo que permite resolver parcialmente las relaciones filogenéticas (Figura 12).

El grupo *Volutaria* se encuentra en la base de los árboles filogenéticos e incluye a los géneros *Volutaria* (estudiado en este trabajo), *Amberboa* y *Mantisalca*, entre otros (Susanna *et al.*, 2011). El género *Volutaria* posiblemente se originó entre el este de África y el sur Arabia (Hellwig, 2004). El rango de distribución de sus especies está comprendido desde el Norte de África, sureste de España, Islas Canarias y el medio este. Las especies analizadas en esta memoria presentan diferente localización. Así, *V. crupinoides* es una especie endémica del Sáhara, *V. muricata* es Ibero-magrebí y *V. lippii* tiene una distribución Mediterránea-Saharosíndica.

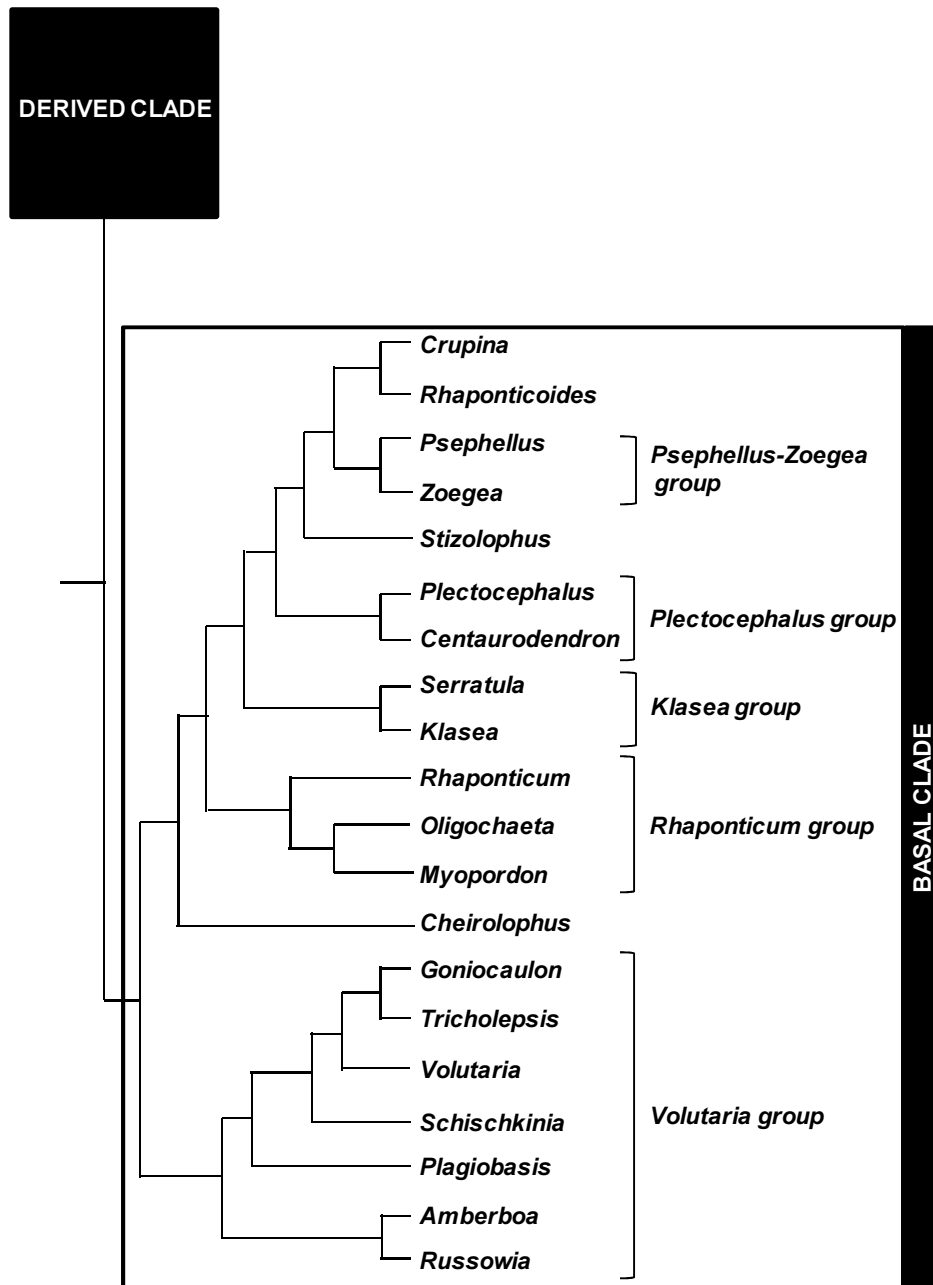


Figura 12. Reconstrucción tentativa de la filogenia del clado basal que incluye a los principales géneros de la subtribu *Centaureinae*. Basado en los trabajos de García-Jacas *et al.* (2001) y Susanna *et al.* (2011).

El resto del clado basal de la subtribu *Centaureinae* forma una tricotomía (Susanna *et al.*, 2011). El primer clado de esta tricotomía, está constituido por especies de *Cheirolophus*. Éste es uno de los géneros de la subtribu con representación continental e insular, en el archipiélago balear y canario. Este género incluye 20 especies distribuidas desde el oeste de la Cuenca Mediterránea hasta la Macaronesia. Susanna *et al.* (1999), mediante secuencias ITS y ETS, elaboraron una filogenia del género agrupándolo en dos clados bien definidos (Figura 13): el clado Mediterráneo y el Macaronésico.

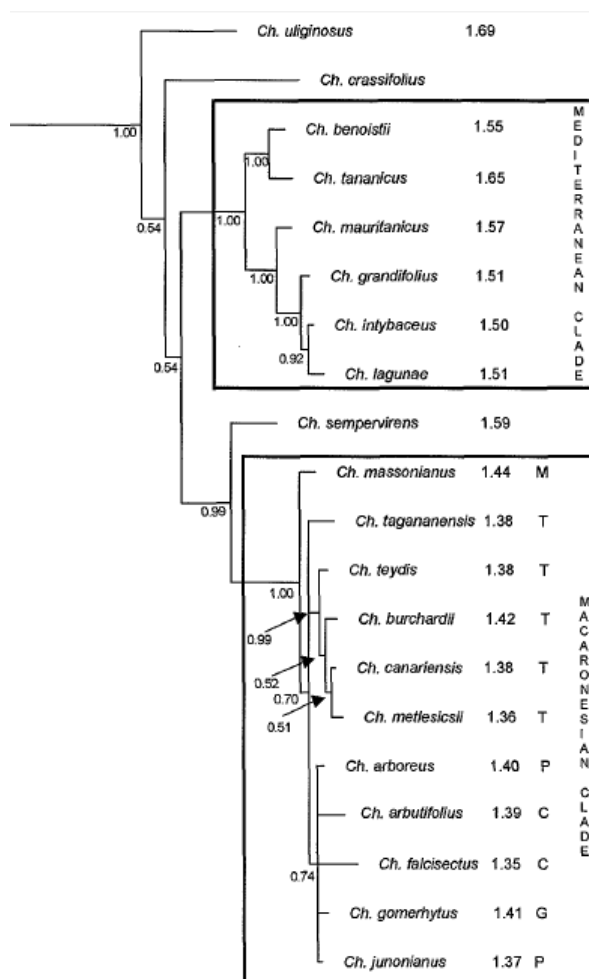


Figura 13. Representación esquemática del árbol filogenético del género *Cheirolophus*. Tomado de Susanna *et al.* (1999).

El segundo clado de esta tricotomía, incluye *Rhaponticum* (incluidas las especies inicialmente identificadas como *Acroptilon* y *Leuzea*), *Myopordon* y *Oligochaeta*. El grupo *Rhaponticum* comprende unas 40 especies en 8 géneros: *Rhaponticum* Vaill. (= *Stemmacantha*), *Acroptilon*, *Callicephalus*, *Leuzea*, *Myopodon*, *Ochrocephala*, *Oligochaeta* y *Centaurothamnus* (Hidalgo *et al.*, 2006). El grupo *Rhaponticum* es monofilético (Hidalgo *et al.*, 2006). Se subdivide en dos clados, uno oriental compuesto mayoritariamente por especies del centro, mitad y este de Asia, Australia y este de Europa que corresponden a *Rhaponticella* y otro occidental que incluye especies del Norte de África, Islas Canarias y Europa que corresponden a *Leuzea* (Hidalgo *et al.*, 2006), lo que implica una parafilia del género *Rhaponticum* (Hidalgo *et al.*, 2006) (Figura 12). Hidalgo *et al.* (2006) confirmaron por primera vez mediante análisis moleculares que las especies de *Leuzea* y *Rhaponticum* están íntimamente relacionadas y propusieron que las especies de *Leuzea* se integraran en el género *Rhaponticum* (Figura 14).

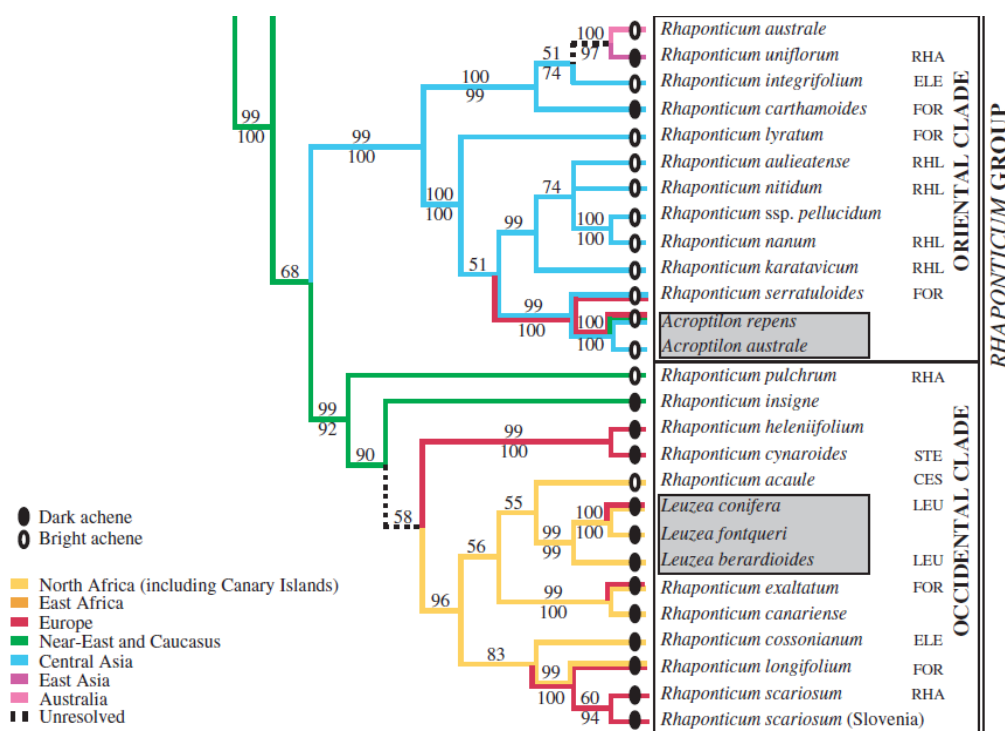


Figura 14. Filogenia del género *Rhaponticum* y *Leuzea*. Tomada de Hidalgo *et al.* 2006

Un tercer clado contiene los restantes táxones. Este último clado comprende a su vez un clado y una politomía con tres clados. El primer clado incluye, entre otros, a los géneros *Klasea* y *Serratula*. Se han descrito aproximadamente 70 táxones del género *Klasea*. Análisis moleculares sobre la relación entre *Klasea* y *Rhaponticum* confirman la proximidad entre ambos grupos (Hidalgo *et al.*, 2006). Esta proximidad también fue sugerida por Lessing (1832) quien consideró a *Klasea* como un subgénero de *Rhaponticum*. La delimitación de *Klasea* desde su descripción, ha experimentado diversas interpretaciones en torno a otro género, *Serratula* s.l. En un principio, algunas de las especies de *Klasea* se consideraron como especies pertenecientes a *Carduus* L. y *Centaurea* aunque, posteriormente, se incorporaron al género *Serratula* otorgándole el rango de sección dentro de *Serratula* L. (de Candolle 1838; Borisova, 1963). Sin embargo, habría que considerarlos como géneros independientes con *Klasea* ocupando una posición basal en la filogenia y siendo *Serratula* un grupo derivado de la subtribu Centaureinae (Martins y Hellwig, 2005; Hidalgo *et al.*, 2006). Así, la mayor parte de las especies de *Serratula* se han transferido a *Klasea* (Martins y Hellwig, 2005).

Por último, una politomía incluiría tres clados (Susanna *et al.*, 2011): el primero contiene el género *Plectocephalus*, el segundo el género *Stizolophus*, mientras que el tercero sería un clado moderadamente apoyado que contiene a los géneros *Crupina* y *Rhaponticoides* entre otros.

V.1.5. Subtribu *Carduinae*

La subtribu *Carduinae* es parafilética (Bremer 1994; Susana *et al.*, 1995; Häffner y Hellwig, 1999; García-Jacas *et al.*, 2002; Susanna *et al.*, 2006; Andenberg *et al.*, 2007; Susanna y García-Jacas, 2009). La mejor solución cladística para el problema parafilético de *Carduinae*, es considerar las subtribus *Carduinae* y *Centaureinae* como un grupo natural. Probablemente lo sean aunque no exista apoyo estadístico (García-Jacas *et al.*, 2002). Las filogenias moleculares obtenidas por García-Jacas *et al.* (2002) y Susanna *et al.* (2006) consideran a *Centaureinae* el grupo hermano de la subtribu *Carduinae*.

Dentro de la subtribu *Carduinae* pueden identificarse diferentes grupos informales (Andenberg *et al.*, 2007; Susanna and García-Jacas, 2009) (Figura 15):

a) El grupo basal *Xeranthemum*: *Amphoricarpos*, *Chardinia*, *Siebera* y *Xeranthemum*.

b) Los grupos *Berardia* y *Staelina*: constituidos por especies aisladas, sin afinidad con ningún otro género de la subtribu *Carduinae*.

c) El grupo *Onopordum*: al grupo *Onopordum* pertenecen los géneros *Onopordum* (el más diversificado), *Alfredia*, *Ancathia*, *Lamyropappus*, *Olgaea*, *Synurus*, *Syreitschikovia* y *Xanthopappus* (García-Jacas *et al.*, 2008). La delimitación genérica del grupo, con la excepción de *Onopordum*, no está clara (Susanna *et al.*, 2006; García-Jacas *et al.*, 2008). El género *Onopordum* junto con *Alfredia* comparten una característica morfológica poco común en la tribu y es la presencia de un receptáculo desnudo, sin escamas o cerdas. El género *Onopordum* y los que habitan en el centro y este de Asia como son *Ancathia*, *Synurus*, *Olgaea* y *Alfredia* constituyen un grupo monofilético (Häffner y Hellwig, 1999), con un fuerte carácter relictivo sugiriendo que posiblemente es de los grupos más antiguos de *Carduinae* (Häffner y Hellwig, 1999).

d) El grupo *Cynara*: *Cynara*, *Lamyropsis* y *Ptilostemon*, estrechamente relacionadas con los denominados comúnmente cardos (grupo siguiente).

e) El grupo *Carduus-Cirsium*, los llamados cardos (*Carduus*, *Cirsium*, *Notobasis*, *Picnomon*, *Silybum* and *Tyrimmus* y probablemente *Galactites* formando un grupo natural con el grupo *Cynara*. El grupo *Carduus-Cirsium* es uno de los grupos naturales más complejo, con flores muy espinosas. En cuanto a diversidad se refiere *Carduus* y *Cirsium* son los que más representantes tienen mientras que *Silybum* y *Galactites* son los que comprenden un menor número de especies. El género *Carduus* incluye aproximadamente unas 100 especies (Chaudhary, 2000) y está ampliamente distribuido por la región Mediterránea y Norte y Este de África (Bremer, 1994) siendo un género invasor en América del Norte (Olivieri *et al.*, 1985). El

género *Cirsium* se distribuye en el Hemisferio Norte con presencia también de poblaciones no nativas en el Hemisferio Sur. La mayor diversidad de especies se concentra en las montañas del Sur de Europa, el Cáucaso, Asia Central, oeste de China, Japón y noroeste de América (Jäger, 1977; Meusel y Jäger, 1992). El género *Cirsium* está compuesto por aproximadamente 300 especies. La mayoría de las especies forman híbridos interespecíficos siendo los procesos de hibridación especialmente intensos en Centroeuropa (Burès, 2004), en Norte América (Keil, 2006) y Japón (Kadota, 1995). *Silybum* se encuentra filogenéticamente próximo al género *Cirsium*. Se han estudiado las especies *S. marianum*, nativa de la Cuenca Mediterránea y originalmente distribuida por el sur de Europa y Norte de África y cultivada en Egipto, China, Irán y Argentina y la especie *S. eburneum* que es de distribución ibero-norteafricana. El género *Galactites*, distribuido en España y Norte de África, ocupa una posición basal en el grupo *Carduus* (Häffner y Hellwig 1999; Susanna *et al.*, 2006) aunque morfológicamente presenta más similitudes con *Carduus* y *Cirsium* que con los géneros más próximos a él (Petit, 1997; Susanna *et al.*, 2006). Se distribuye en España y Norte de África. *G. duriaei*, especie analizada en esta memoria, forma híbridos interespecíficos con otras especies del mismo género (Figuerola *et al.*, 1990).

f) El grupo *Arctium-Cousinia*: *Arctium*, *Cousinia*, *Hypacanthium* y *Schmalhausenia*.

g) Y el grupo *Jurinea-Saussurea*: *Jurinea*, *Saussurea* y *Dolomiaea*. El grupo *Jurinea-Saussurea*, es mayoritariamente endémico de la meseta Tibetana pero también se localiza en otras regiones del Hemisferio Norte. Es un grupo polifilético (Whang *et al.*, 2009). La delimitación entre el género *Jurinea* y el género *Saussurea* no está bien resuelta. *Jurinea*, habita en áreas montañosas y consta de bastante endemismos. En este trabajo, hemos analizado dos especies del género *Jurinea*, *J. fontqueri* y *J. humilis*. *J. fontqueri* es un endemismo raro cuya distribución se limita al extremo oriental del área mediterránea en Sierra Mágina (Jaén). *J. humilis* (Desf.) DC presenta una distribución un poco más amplia en la Península Ibérica. El grupo *Jurinea-Saussurea* y el grupo *Arctium-Cousinia* son grupos hermanos y muy próximos entre sí (Susanna *et al.*, 2006). A su vez, constituyen un grupo muy relacionado con la subtribu Centaureinae (Hellwig 1996, Häffner y Hellwig 2000; García-Jacas *et al.* 2002; Susana *et al.*, 2006).

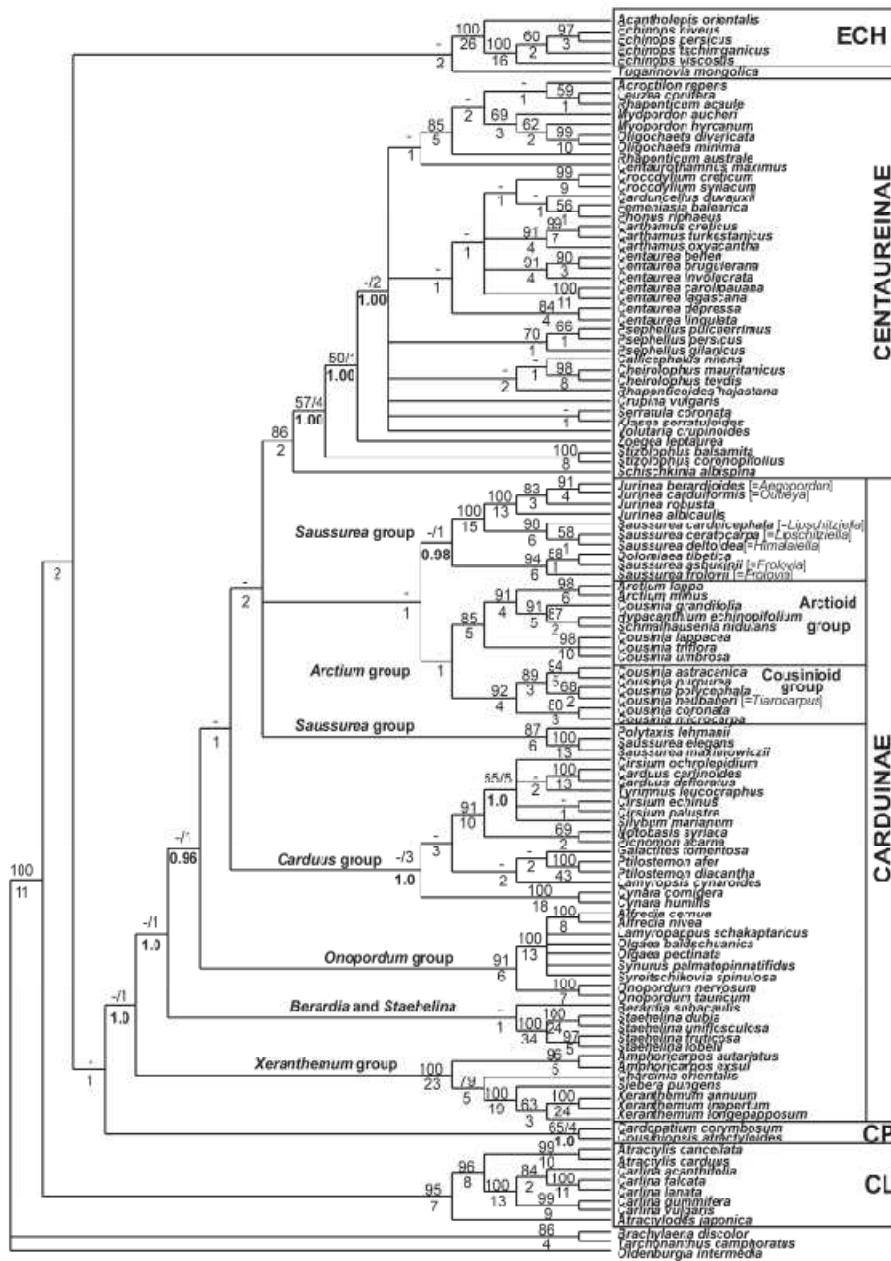


Figura 15. Filogenia de la tribu Cardueae basada en las secuencias ITS, trnL-trnF, and matK. Tomada de Susanna *et al.*, 2006. Abreviaturas de las subtribus: CL, Carlininae; CP, Cardopatinae; ECH, Echinopsinae.

V.1.6. Subtribu Echinopsinae

Clásicamente, se ha considerado que la subtribu Echinopsinae estaba compuesta por el género monotípico *Acantholepis* (*A. orientalis*) y por el género *Echinops*, considerando a *Acantholepis* próximo a *Echinops* pero como un género independiente (Bentham 1873, Hoffmann 1894, Dittrich 1977, Bremer 1994, Susanna y García-Jacas 2007). Sin embargo, en la actualidad, el género *Echinops sensu lato* (*Echinops* y *Acantholepis*) se considera el único género de la subtribu Echinopsinae (Cass.) Dumort (Susanna *et al.*, 2006). La característica morfológica

diferencial del género *Echinops* es la presencia de una inflorescencia compuesta, sincefalia, un rasgo sin paralelismo en el resto de los grupos de la Tribu *Cardueae* (Wagenitz 1976, Dittrich 1977, Petit 1988, 1997). García-Jacas *et al.* (2002) consideran que la inflorescencia compuesta de *Echinops* es una adaptación extrema y carente de relevancia sistemática.

El género *Echinops* s.l. se distribuye en África tropical, Cuenca Mediterránea, regiones templadas de Eurasia hasta Asia Central, Mongolia y noreste de China, con un elevado número de especies en el Cáucaso y en el Medio Este (Jäger, 1987). Posiblemente, tenga un origen poliploide mientras que *Acantholepis* sería diploide (Garnatje *et al.*, 2004a, b). Comprende unas 120 especies (Bobrov, 1997; Susanna y García-Jacas, 2007).

La uniformidad en su morfología permite que la delimitación taxonómica sea casi incuestionable pero esto también dificulta los intentos de establecer grupos naturales y una clasificación infragenérica. Está organizada en diferentes secciones, según diferentes autores: cinco (Rechinger, 1979), ocho (Shishkin y Bobrov 1997) o doce (Bunge, 1863). La filogenia elaborada por Sánchez-Jiménez *et al.* (2010) basada en marcadores moleculares, establece que *Echinops* es un grupo monofilético que se divide en nueve secciones, *Chamaechinops*, *Acantholepis*, *Psectra*, *Hololeuce*, *Hamolepis*, *Oligolepis*, *Phaeochaete*, *Ritropsis* y *Echinops*, existiendo dos linajes, uno de los cuáles incluye a las dos secciones analizadas en esta Tesis Doctoral, *Psectra* y *Echinops*. *Psectra*, incluye la especie *E. strigosus* (Sánchez-Jiménez *et al.*, 2010) que pertenece a un linaje basal y se distribuye en el Norte de África (Mauritania) y el sur de la Península Ibérica (Valdés, 2002). Otra de las especies analizadas en este estudio es la especie perenne *E. sphaerocephalus* que se incluye en la sección *Echinops* (Sánchez-Jiménez *et al.*, 2010) y se distribuye en Europa meridional y central.

Géneros analizados subtribu Echinopsinae

Familia Asteraceae (Compositae)

Subfamilia Carduoideae

Tribu Cardueae

Subtribu Echinopsinae

Género *Echinops sensu lato*

Sección *Chamaechinops*

Sección *Acantholepis*

Sección *Psectra*

E. strigosus

Sección *Hololeuce*

Sección *Hamolepis*

Sección *Oligolepis*

Sección *Phaeochaete*

Sección *Ritropsis*

Sección *Echinops*

E. sphaerocephalus

V.1.7. Subtribu Cardopatiinae

Aunque ha existido controversia sobre si los géneros que lo integran constituirían una subtribu independiente al resto o si estarían incluidos en las otras subtribus (García-Jacas *et al.*, 2002; Susanna *et al.*, 2006), en la actualidad se reconoce como una subtribu diferente (Susanna y

García-Jacas 2009). Esta quinta subtribu consta de dos géneros: *Cardopatium* y *Cousiniopsis*. A nivel filogenético estos géneros se encuentran próximos a la subtribu *Carlininae* (García-Jacas *et al.*, 2002; Susanna *et al.*, 2006). *Cardopatium* está compuesto por hierbas perennes que habitan desde este del Mediterráneo hasta Italia y Norte de África. *Cousiniopsis* está integrado por hierbas anuales de las estepas de Asia Central (Anderberg *et al.*, 2007).

V.1.8. Subtribu *Carlininae*

Probablemente sea el grupo más antiguo de la tribu (García-Jacas *et al.*, 2002). Un importante carácter que refuerza esta posición basal de *Carlininae* es la presencia de verdaderas lígulas en *Atractylis* (Dittrich, 1977), carácter compartido por muchos de los miembros más antiguos de otras tribus de la subfamilia *Carduoideae* como la tribu *Mutisieae*. Esta subtribu está compuesta por cinco géneros: *Atractylis*, *Atractylodes*, *Carlina* y *Thevenotia* y *Tugarinovia* (Susanna y García-Jacas, 2009). El género *Atractylis* contiene unas 30 especies principalmente Mediterráneas habitando algunas especies también en Eurasia e Islas Canarias. Análisis filogenéticos basados en ITS y en la región trnL-F indican que *Atractylis* y *Carlina* se encuentran en un mismo clado, mientras que *Atractylodes*, localizado en otro clado, es un grupo hermano de ambos (Peng *et al.*, 2012). En este trabajo, se han analizado las especie *Atractylis humilis* y *Atractylis leyseroides*. Del género *Carlina* se han descrito unas 28 especies distribuidas por el centro de Europa, la región Mediterránea y en el archipiélago Canario. Especies canarias de este género han formado híbridos interespecíficos (Marrero-Rodríguez, 1989; Francisco-Ortega *et al.*, 2000). En este trabajo se han analizado dos especies, *C. acaulis* y *C. salicifolia*. En relación a *C. acaulis* crece en hábitats rupícolas secos en el Centro y Sureste de Europa, desde los Pirineos a los Alpes, Península Ibérica y Balcánica siendo especialmente frecuente en el sureste y noreste de Polonia (Zajac y Zajac, 2001). *C. salicifolia* es endémica de la Macaronesia.

Géneros analizados subtribu *Carlininae*

Familia *Asteraceae* (*Compositae*)

Subfamilia *Carduoideae*

Tribu *Cardueae*

Subtribu *Carlininae*

Género *Atractylis*

A. humilis

A. leyseroides

Género *Atractylodes*

Género *Carlina*

C. acaulis

C. salicifolia

Género *Thevenotia*

Género *Tugarinovia*

V.2. Resultados

V.2.1. Subtribu *Centaureinae*

Se ha analizado el patrón evolutivo de la familia de ADN satélite *Hinfl* en 38 especies pertenecientes a 10 géneros de la subtribu *Centaureinae* (Tribu *Cardueae*, Subfamilia *Carduoideae*, Familia *Asteraceae*): *Centaurea*, *Rhaponticoides*, *Carduncellus*, *Carthamus*, *Phonus*, *Rhaponticum*, *Crupina*, *Klasea*, *Cheirolophus* y *Volutaria* (Ver esquema de los géneros analizados en la página 65). En todos los casos se obtuvieron amplificadores con un tamaño aproximado de 350 pb, que fueron clonados y secuenciados. En total, se han analizado 502 unidades monoméricas para esta subtribu.

Hemos incluido en el análisis 55 secuencias representativas, previamente analizadas en nuestro grupo (Suárez-Santiago *et al.*, 2007a), pertenecientes a seis especies del género *Centaurea* (*Centaurea granatensis*, *C. alba* subsp. *alba*, *C. corymbosa*, *C. boissieri* subsp. *willkommii*, *C. delicatula*, y *C. jaennensis*) y a *Phonus arborescens*. La media de secuencias seleccionadas (representativas de cada una de las tres subfamilias *Hinfl* caracterizadas) fue de ocho secuencias. El resto de secuencias repetidas (447) se han aislado y caracterizado para este trabajo y pertenecen a otras especies del género *Centaurea* y a especies de otros nueve géneros. La media de secuencias estudiadas para cada especie es de 14, no habiendo analizado en ningún caso menos de 10 secuencias por especie. Una excepción ha sido la especie *Crupina crupinastrum*, de la que analizamos 29 repeticiones porque analizamos dos poblaciones diferentes, una de las cuales estaba considerada inicialmente *C. matae*, que es en realidad sinónimo de *C. crupinastrum*. Las 55 secuencias seleccionadas del estudio previo (Suárez-Santiago *et al.* (2007a) corresponden a los siguientes números de acceso a la base de datos EMBL: AM712738-AM712751, AM712761-AM712774, AM712803-AM712809, AM712815-AM712821, y AM712828-AM712840. Los números de acceso a la base de datos EMBL para las 447 secuencias restantes analizadas en este estudio por primera vez fueron HF571538-HF571985.

Las 502 unidades monoméricas se clasificaron en ocho tipos monoméricos o subfamilias. Cada una de estas subfamilias se estableció de acuerdo a un conjunto de posiciones diagnósticas determinadas por mutaciones específicas compartidas por todas las secuencias de un tipo en concreto (Figura 16). Esta clasificación se confirmó mediante el análisis de variación intraespecífica y de la divergencia interespecífica de las secuencias *Hinfl* (Anexo I). En la mayor parte de los casos pueden encontrarse niveles altos de variación intraespecífica (Anexo I), que es consecuencia principalmente (*Centaurea*, *Rhaponticoides*, *Rhaponticum*, *Crupina*, *Cheirolophus* o *Volutaria*, por ejemplo) de la existencia de diferentes tipos de secuencias o

subfamilias compartidas entre los genomas de diferentes especies (Anexo II). De hecho, analizando las secuencias *HinfI* por tipos, la variación suele ser menor entre repeticiones del mismo tipo, aumentando cuando comparamos monómeros de tipos diferentes (Anexo II). En otros géneros, sólo hemos encontrado un tipo de secuencia común a todas las especies. No obstante, en algún caso, también la variación intraespecífica ha sido similar a la interespecífica. Es el caso, por ejemplo, del género *Carduncellus*.

Las subfamilias fueron nombradas con números romanos, del I al VIII, siguiendo la nomenclatura utilizada previamente (Suárez-Santiago *et al.*, 2007a) para tres de ellas (las subfamilias I, II y III). La mayor parte de la variación se encuentra concentrada en estas regiones diagnósticas, mientras que el resto de la secuencia monomérica está más conservada, con la excepción de sitios polimórficos distribuidos aleatoriamente, en especial en la región a partir de la cual se han diseñado los cebadores (Figura 16).

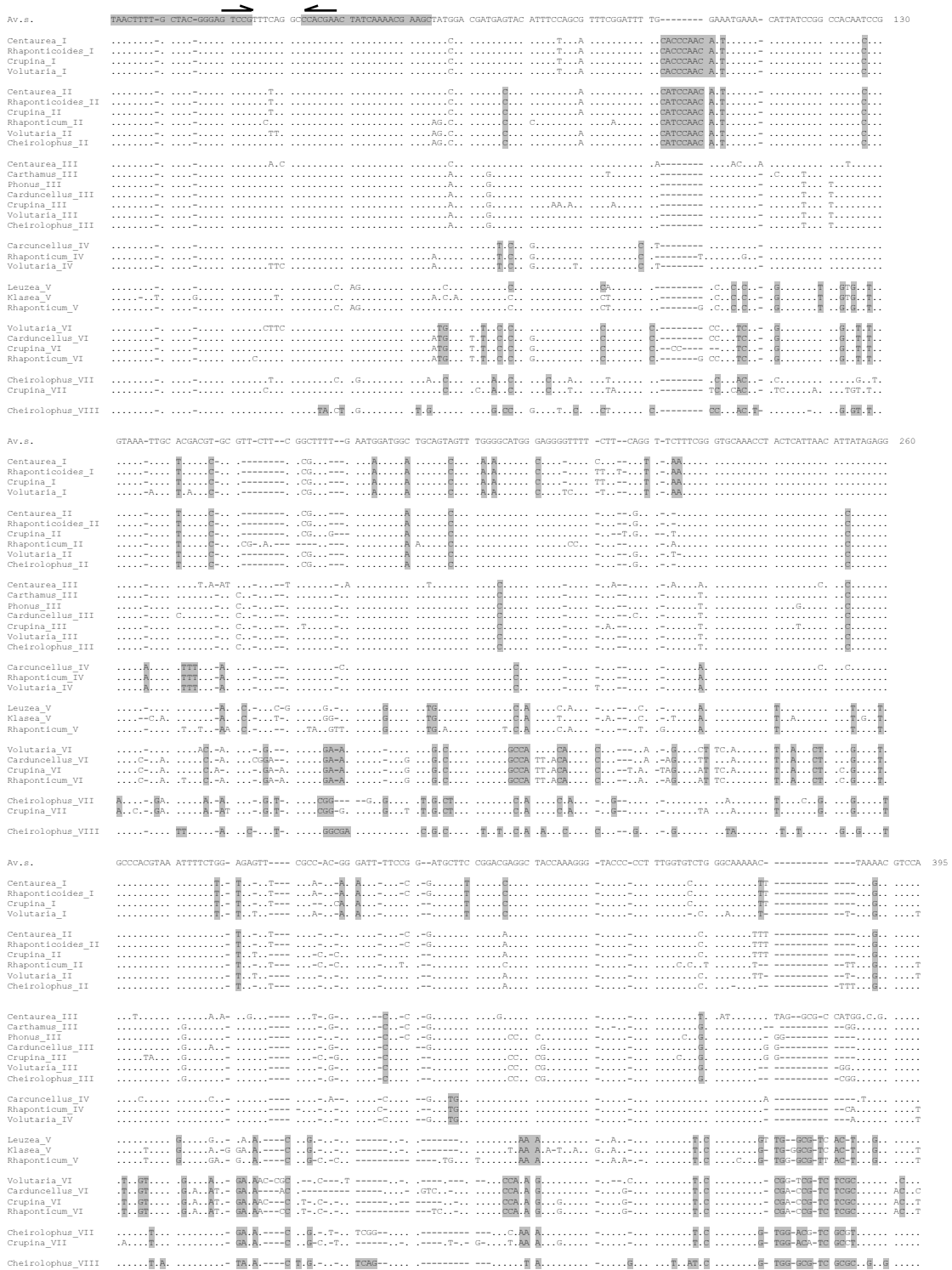


Figura 16. Alineamiento de las secuencias consenso de las distintas subfamilias del ADN satélite *HinfI* en los diferentes géneros de la subtribu *Centaureinae*. La primera línea corresponde a la secuencia consenso de todas las secuencias *HinfI*. En el resto de secuencias se muestran las posiciones nucleotídicas que varían con respecto a la secuencia consenso. Los puntos indican la misma base en las diferentes secuencias. Las posiciones diagnósticas para cada subfamilia están sombreadas en gris. La región de unión de los cebadores utilizados para la amplificación se muestra sombreada en gris en la secuencia consenso, y las flechas indican la orientación de los cebadores.

Tabla 8. Distribución por géneros de las subfamilias Hinfl en la subtribu Centaureinae. Los números representan el número de secuencias de la subfamilia identificadas en algunas de las especies. La línea gris separa el clado derivado y los grupos basales.

Género	Especie	Subfamilia Hinfl							
		I	II	III	IV	V	VI	VII	VIII
<i>Centaurea</i> *	<i>alba</i>	4	1	2					
	<i>corymbosa</i>	4	3						
	<i>boissieri</i>	6	8						
	<i>delicatula</i>	1	2	3					
	<i>jaennensis</i>	1	7						
<i>Acrocentron</i> *	<i>clementei</i>	6	5						
	<i>granatensis</i>		1	8					
	<i>saxifraga</i>	1	8	1					
<i>Cyanus</i> *	<i>cyanus</i>	4	8	1				1	
<i>Carthamus</i>	<i>lanatus</i>			15					
	<i>tinctorius</i>			18					
<i>Phonus</i>	<i>arborescens</i>			19					
<i>Carduncellus</i>	<i>caeruleus</i>			1	11				
	<i>mitissimus</i>			2	10		1		
	<i>cuatrecasasii</i>				10				
	<i>dianius</i>				14				
	<i>hispanicus</i>				12				
	<i>calvus</i>			4	10		2		
<i>Rhaponticoides</i>	<i>alpina</i>	9	6						
	<i>linaresii</i>	6	6	1					
<i>Crupina</i>	<i>crupinastrum</i>	9	15	2			1	2	
	<i>vulgaris</i>		3	4				6	
<i>Rhaponticum</i>	<i>acaule</i>		3		6	3	1		
	<i>berardioides</i>					13			
	<i>centauroides</i>					14			
	<i>coniferum</i>					12			
	<i>exaltatum</i>					15			
<i>Klasea</i>	<i>flavescens</i>					11			
	<i>leucantha</i>					13			
<i>Volutaria</i>	<i>muricata</i>		1	2			11		
	<i>tubuliflora</i>						14		
	<i>crupinoides</i>		2	3	4		3		
	<i>lippi</i>	1	3	9		1			
<i>Cheirolophus</i>	<i>falcisectus</i>							16	
	<i>intybaceus</i>							11	4
	<i>sempervirens</i>							5	7
	<i>sventenii</i>		1					11	2
	<i>teydis</i>		3	12					

La Tabla 8 muestra la distribución de las subfamilias Hinfl entre especies. Las especies de los géneros *Centaurea* y *Rhaponticoides* están caracterizadas por la presencia en sus genomas de secuencias Hinfl pertenecientes a las subfamilias I, II, y III, encontrándose secuencias de dos de las tres subfamilias y coexistiendo dos o más tipos en el genoma de la misma especie. En el caso de las dos especies analizadas del género *Rhaponticoides*, las secuencias pertenecen bien a la subfamilia I o bien a la II. Además, una secuencia en *R. linariseii* pertenece a la subfamilia III. En el caso de *Centaurea*, se han encontrado las subfamilias I y II en todas las especies del subgénero *Centaurea*, y la subfamilia III se encontró en dos especies. En la especie analizada del género *Cyanus* (*C. cyanus*), ocho de las catorce

repeticiones pertenecían a la subfamilia II, encontrándose también secuencias de la subfamilia I (cuatro repeticiones) y III (una repetición). Además, una de las secuencias aisladas en esta especie pertenece a la subfamilia VII. En el caso del subgénero *Acrocentron*, no se encontraron secuencias *HinfI* tipo III en *C. clementei*, mientras que la subfamilia I parece estar ausente en *C. gramatensis*. Los genomas de las dos especies analizadas del género *Crupina* mostraron secuencias de las subfamilias II y III, y adicionalmente, encontramos hasta seis (de 13) repeticiones de la subfamilia VII en *C. vulgaris*, y tres repeticiones (de 29) de las subfamilias VI (una) y VII (dos) en *C. crupinastrum* (Tabla 8). En las especies de los géneros *Carthamus* y *Phonus*, encontramos solo secuencias *HinfI* tipo III. La subfamilia IV es característica del género *Carduncellus*. No obstante, algunas especies de este género tienen también en sus genomas un bajo número de repeticiones de las subfamilias III y VI. La subfamilia V se ha encontrado en *Rhaponticum* y *Klasea*. Estas especies presentan únicamente este tipo de repetición *HinfI*, excepto en el caso de *R. acaule*, que porta también secuencias de las subfamilias II, IV, y VI en su genoma. De las cuatro especies del género *Volutaria* analizadas, las secuencias *HinfI* de *V. lippi* pertenecieron a las subfamilias I, II, III, o V, aunque principalmente a las de tipo III (nueve secuencias de catorce). Además, la mayoría de secuencias *HinfI* de *V. crupinoides* fueron de las subfamilias II, III, y IV. Sin embargo, la mayoría de secuencias de *V. muricata* y *V. tubuliflora*, junto a tres secuencias de *V. crupinoides* pertenecían a la subfamilia VI. Tres especies del género *Cheirolophus* (*C. intybaceus*, *C. sempervirens* y *C. sventenii*) comparten dos tipos de monómeros *HinfI* en sus genomas, los de las subfamilias VII y VIII. *C. sventenii* presenta también una repetición tipo II. La especie *C. falcisectus* presentó solo secuencias de la subfamilia VII. Por el contrario, entre las secuencias aisladas de *C. teydis*, doce pertenecieron a la subfamilia III y las otras tres a la subfamilia II (Tabla 8).

La reconstrucción filogenética mediante Máxima verosimilitud (ML) e inferencia Bayesiana muestra árboles de topología similar. En la Figura 17 se muestra el árbol consenso obtenido mediante análisis bayesiano a modo de esquema simplificado en el que se han eliminado los nombres de las secuencias (especie y número de repetición *HinfI*) en los nodos terminales. Las secuencias no se agrupan en los árboles en función de la especie ni del género al que pertenecen. Lo más habitual es que las secuencias se agrupen según la subfamilia a la que pertenecen (Figura 17 y Tabla 8). De hecho, en el árbol de la Figura 17 se han identificado los clados obtenidos con la correspondiente subfamilia *HinfI*, incluyendo recuadros que relacionan los clados con las subfamilias. Pueden distinguirse dos clados principales. El primero de ellos incluye cinco subclados, correspondiendo cada uno de ellos a las subfamilias I a V. El segundo clado incluye tres subclados, correspondiendo a las subfamilias VI a VIII.

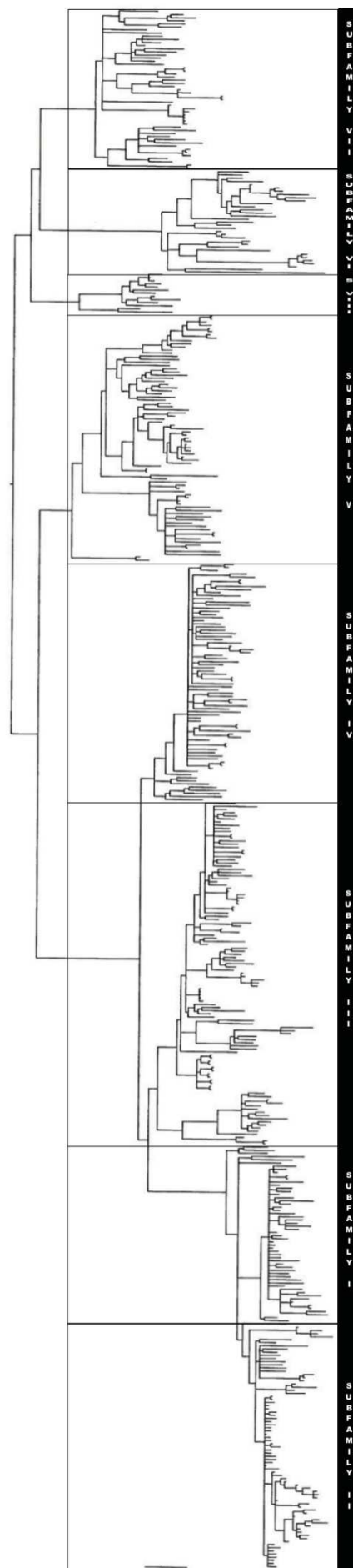


Figura 17. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias *HinfI* de la subtribu *Centaureinae*. Se muestra un árbol simplificado de secuencias *HinfI* pertenecientes a las diferentes subfamilias identificadas en la subtribu *Centaureinae* en el que se han eliminado los nombres de las secuencias en los nodos terminales. Los cuadros relacionan los clados con las subfamilias. Ver la versión extendida en el ANEXO IV.

Dentro de cada clado, las secuencias de las subfamilias I, II, o III de las diferentes especies aparecen entremezcladas de modo independiente de la especie a la que pertenecen, Sin embargo, vale la pena señalar que la mayoría de las secuencias de la subfamilia III pertenecientes a especies de *Phonus* y de *Carthamus* tienden a agruparse por afinidad taxonómica, por un lado, las secuencias de *P. arborescens* y, por otro lado, las secuencias de *C. tinctorius* y de *C. lanatus*, aunque aparecen algunas secuencias entremezcladas (Figura 18).

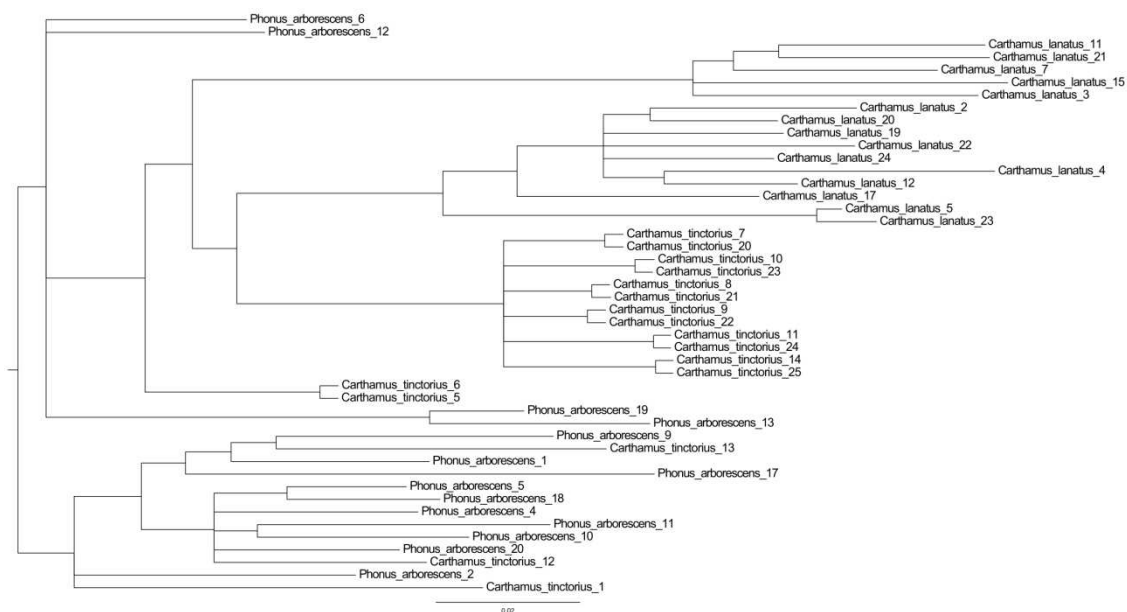


Figura 18. Secuencias *HinfI* pertenecientes a la subfamilia III de especies de los géneros *Carthamus* y *Phonus*.

Por el contrario, la comparación de las secuencias de la subfamilia III de estas especies y sus homólogas presentes en bajo número de copias en otras especies analizadas, muestra una elevada conservación interespecífica de las secuencias y la ausencia de mutaciones diagnósticas específicas de especie, por lo que aparecen entremezcladas en el clado de la subfamilia III en el árbol filogenético de todas las secuencias (Figura 17 y Anexo IV). Las secuencias *HinfI* de *Carduncellus* (subfamilia IV) aparecen entremezcladas sin separación por afinidad según la especie a la que pertenecen (árbol filogenético de todas las secuencias, Figura 17 y Anexo IV). La filogenia de las secuencias de la subfamilia V revela que las secuencias tienden a agruparse por géneros en dos subclados, uno compuesto por las secuencias de *Rhaponticum*, y el otro compuesto por las secuencias de *Klasea*. Además, las secuencias en estos casos tienden a estar agrupadas según la especie a la que pertenecen (Figura 19).



Figura 19. Secuencias *HinfI* pertenecientes a la subfamilia V de especies de los géneros *Rhaponticum* y *Klasea*.

Las secuencias de la subfamilia VI del género *Volutaria* se agrupan también por especies (Figura 20). Las secuencias de las dos subfamilias identificadas en género *Cheirolophus* (subfamilias VII y VIII), no se agrupan por especies en los árboles filogenéticos sino que aparecen entremezcladas en árbol filogenético de todas las secuencias (Figura 17 y Anexo IV).

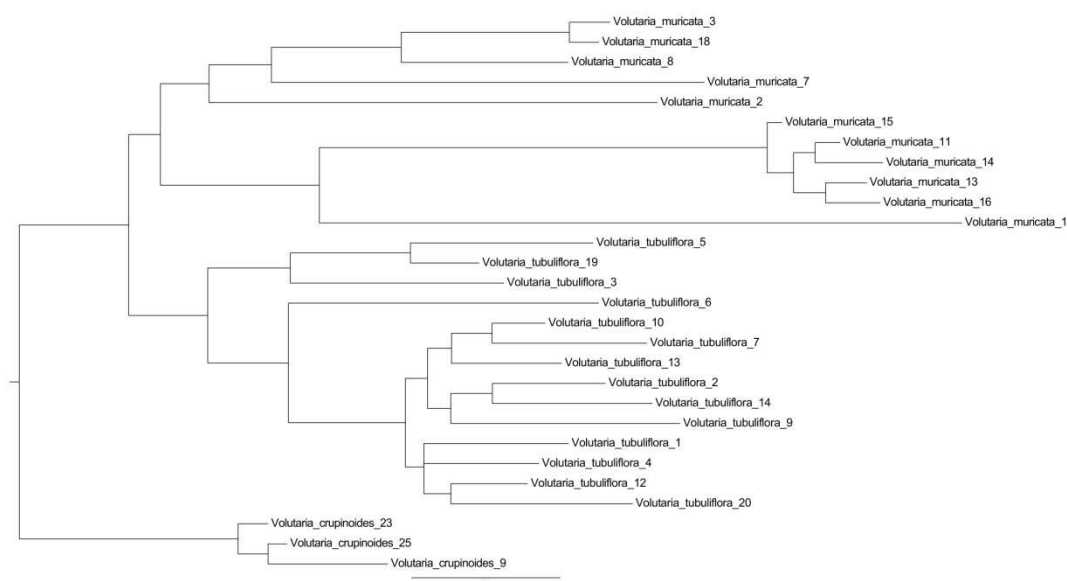


Figura 20. Secuencias *HinfI* pertenecientes a la subfamilia VI de especies del género *Volutaria*.

V.2.2. Subtribus *Carlininae*, *Echinopsinae*, y *Carduinae*

Además de en *Centaureinae*, hemos amplificado el ADN satélite *HinfI* en 16 especies pertenecientes a 9 géneros de las subtribus *Carduinae*, *Echinopsinae* y *Carlininae*, habiéndose obtenido un total de 234 secuencias procedentes de productos amplificados con tamaños aproximados de 350 pb en todos los casos. El análisis filogenético lo llevamos a cabo incluyendo 142 secuencias representativas de especies pertenecientes a la subtribu *Centaureinae*. Por tanto, el análisis incluyó un total de 376 repeticiones *HinfI* correspondientes a 55 especies de 20 géneros distintos.

La Figura 21 muestra el árbol filogenético inferido a partir de las secuencias de las especies de la Subtribu *Carlininae*, utilizando como secuencias de referencia externa (outgroups) las de las especies de *Centaureinae* pertenecientes a diferentes subfamilias. En este caso también, por calidad de la imagen y por espacio disponible, hemos incluido un esquema simplificado. Se trata del árbol obtenido en el que se han eliminado los nombres de las secuencias (especie y número de repetición *HinfI*) en los nodos terminales. La mayoría de repeticiones de *Carlininae*, las de *Atractylis humilis* y las de dos especies del género *Carlina* (*C. salicifolia* y *C. acaulis*), aparecen agrupadas en un subclado dentro del clado principal que contiene también a las secuencias de *Centaureinae* pertenecientes a la subfamilia II. Las distancias entre las secuencias de *Carlininae* y las de la subfamilia II de *Centaureinae* fueron siempre menores que cuando se compararon secuencias pertenecientes a diferentes subfamilias dentro de la subtribu *Centaureinae* (Anexo III). No obstante, las secuencias pertenecientes a la subtribu *Centaureinae* se diferencian de las pertenecientes a la subtribu *Carlininae* (Anexo III).

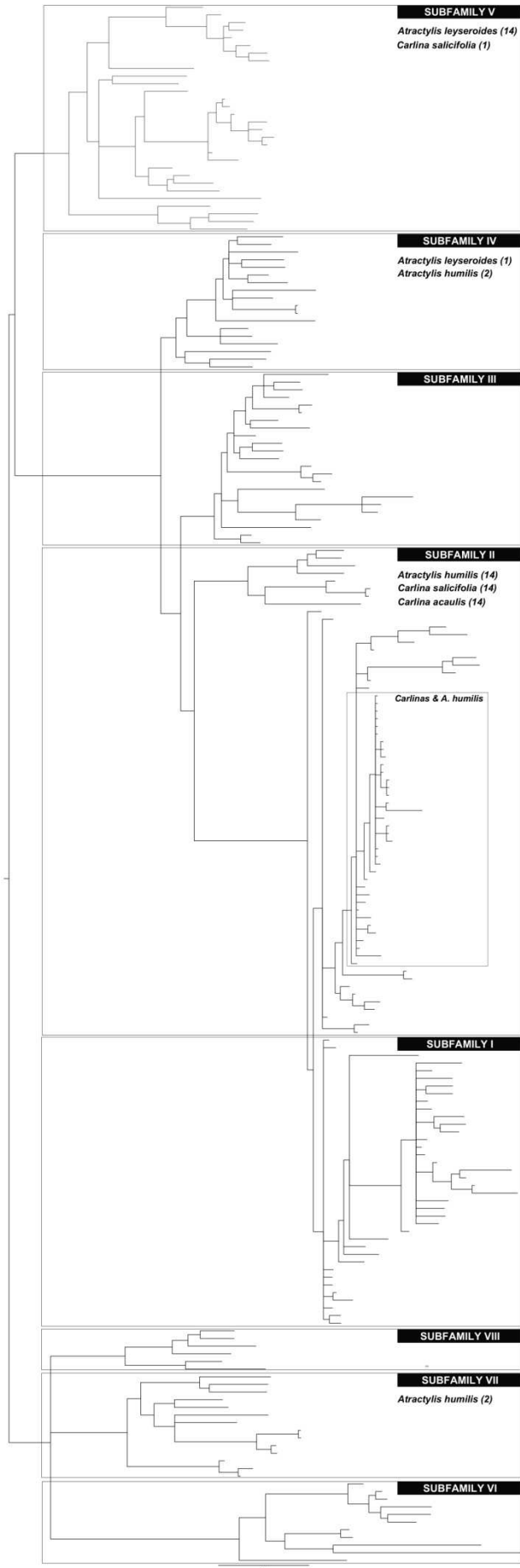


Figura 21. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias *HinfI* de las subtribus *Carlininae* y *Centaureinae*. Se muestra un árbol simplificado de secuencias *HinfI* aisladas de las diferentes especies de las subtribus *Carlininae* y *Centaureinae* en el que se han eliminado los nombres de las secuencias en los nodos terminales, y se han marcado en recuadros las especies de *Carlininae*. Los cuadros relacionan los clados con las subfamilias. Entre paréntesis se indica el número de secuencias identificadas en cada especie para la subfamilia correspondiente.

La distancia genética media calculada (Jukes-Cantor) entre las secuencias de la subfamilia II de las dos subtribus fue 0,09, mientras que las distancias dentro de *Centaureinae* para la subfamilia II oscilaron entre 0,05 y 0,08, y entre 0,02 y 0,05 dentro de *Carlininae*. Las secuencias *HinfI* de las diferentes especies de *Carlininae* aparecen entremezcladas en el árbol filogenético independientemente de la especie a la que pertenecen sin agrupación alguna por afinidad taxonómica (Figura 21). Por el contrario, las secuencias aisladas de la especie *Atractylis leyseroides* se agruparon en dos subclados distintos dentro del clado que contiene a las secuencias de la subfamilia V de *Centaureinae* (Figura 21). La distancia genética media entre las secuencias de *A. leyseroides* fue de 0,09, mientras que entre *A. leyseroides* y las de la subfamilia V de *Centaureinae* fue de 0,18 (Anexo III). Además, un pequeño número de secuencias (en concreto tres) de especies del género *Atractylis* aparecen agrupadas junto a las secuencias de las subfamilias IV (2 secuencias de *A. humilis* y 1 secuencia de *A. leyseroides*) y VII (2 secuencias de *A. humilis*) de *Centaureinae*, al igual que una secuencia de *Carlina salicifolia* se encuentra dentro del clado que contiene a las repeticiones de la subfamilia V (Figura 21).

Las secuencias aisladas de las especies del género *Echinops* se localizan en un subclado dentro del clado principal de repeticiones de la subfamilia II de *Centaureinae* (Figura 22). Las secuencias de *Echinops* se diferencian claramente de las secuencias de las especies de *Centaureinae* (distancia genética media Jukes-Cantor igual a 0,25; distancias de la subfamilia II dentro de *Centaureinae* entre 0,05 y 0,08, y dentro de *Echinopsinae* fue de 0,07; Anexo III). Al contrario que en el caso de *Carlininae*, las secuencias aisladas de dos especies diferentes de *Echinops* (*E. sphaerocephalus* y *E. strigosus*) se agrupan diferencialmente entre ellas.

El árbol de la Figura 23 muestra las relaciones filogenéticas entre las repeticiones *HinfI* de *Carduinae* y *Centaureinae*. Seis secuencias obtenidas de *Galactites duriaei*, todas las secuencias de *Onopordum nervosum*, así como todas las secuencias de las dos especies de *Cirsium* (*C. echinatum* y *C. pyrenaicum*) formaron un clado bien soportado estadísticamente. Las secuencias *HinfI* divergen entre los géneros analizados, y las repeticiones se agrupan en tres clados que coinciden con los tres géneros comparados (*Galactites*, *Cirsium* y *Onopordum*) (Figura 23), aunque las secuencias de *Cirsium* aparecen entremezcladas en el clado independientemente de la especie de la que provienen. Por otro lado, el resto de secuencias de *Galactites duriaei* y las secuencias aisladas a partir de especies de los géneros *Jurinea*, *Silybum* y *Carduus* aparecen entremezcladas en el árbol con las secuencias *HinfI* de las especies de *Centaureinae*. En concreto, la mayoría de las secuencias de *Jurinea* se agrupan junto a las secuencias de la subfamilia II, aunque dos secuencias aparecen agrupadas con secuencias de la subfamilia III, y otras dos con repeticiones de la subfamilia IV de *Centaureinae*. Las secuencias

de *Sylibum* se agrupan junto a repeticiones de las subfamilias II, III, V, o VII de *Centaureinae*. Las secuencias de *Carduus* aparecen entremezcladas con las de las subfamilias II, III, IV, V, VII y VIII de *Centaureinae* (Figura 23). Las especies de estos tres géneros se caracterizan por una particularidad que comparten con algunas especies de *Centaureinae*: la presencia en un mismo genoma de varios tipos de secuencias *HinfI*. Las distancias genéticas entre las especies de *Carduinae* y las de *Centaureinae* oscilaron entre 0,05 y 0,25, mientras que las distancias medias entre las secuencias de cada subfamilia dentro de cada subtribu variaron entre 0,05 y 0,09 para *Carduinae* y entre 0,05 y 0,11 para *Centaureinae* (Anexo III).

El árbol general inferido para las 376 secuencias *HinfI* confirma el agrupamiento y las relaciones filogenéticas entre las repeticiones *HinfI* descritas anteriormente cuando el análisis se lleva a cabo para cada subtribu independientemente (Figura 24).

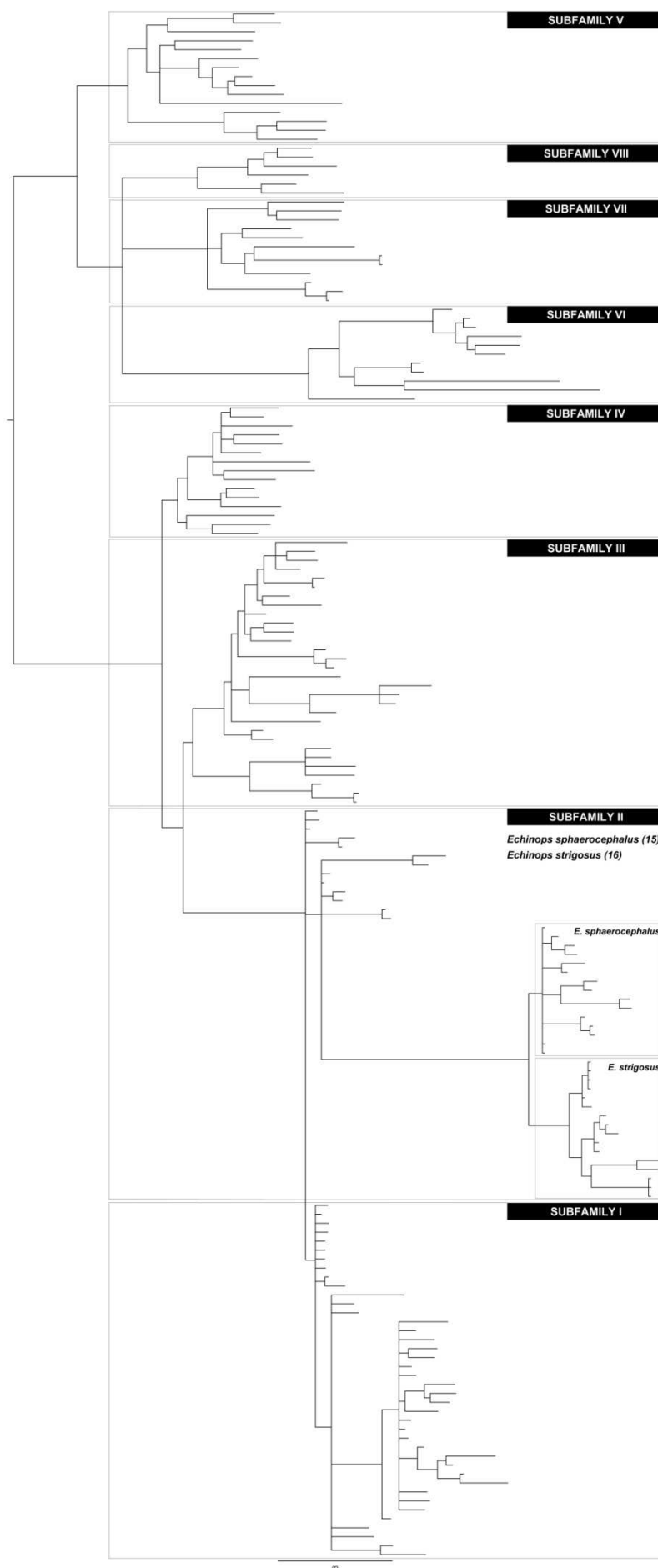


Figura 22. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias *HinfI* de las subtribus *Echinopsinae* y *Centaureinae*. Se muestra un árbol simplificado de secuencias *HinfI* aisladas de las diferentes especies de las subtribus *Echinopsinae* y *Centaureinae* en el que se han eliminado los nombres de las secuencias en los nodos terminales, y se han marcado en recuadros las especies de *Echinopsinae*. Los cuadros relacionan los clados con las subfamilias. Entre paréntesis se indica el número de secuencias identificadas en cada especie para la subfamilia correspondiente.

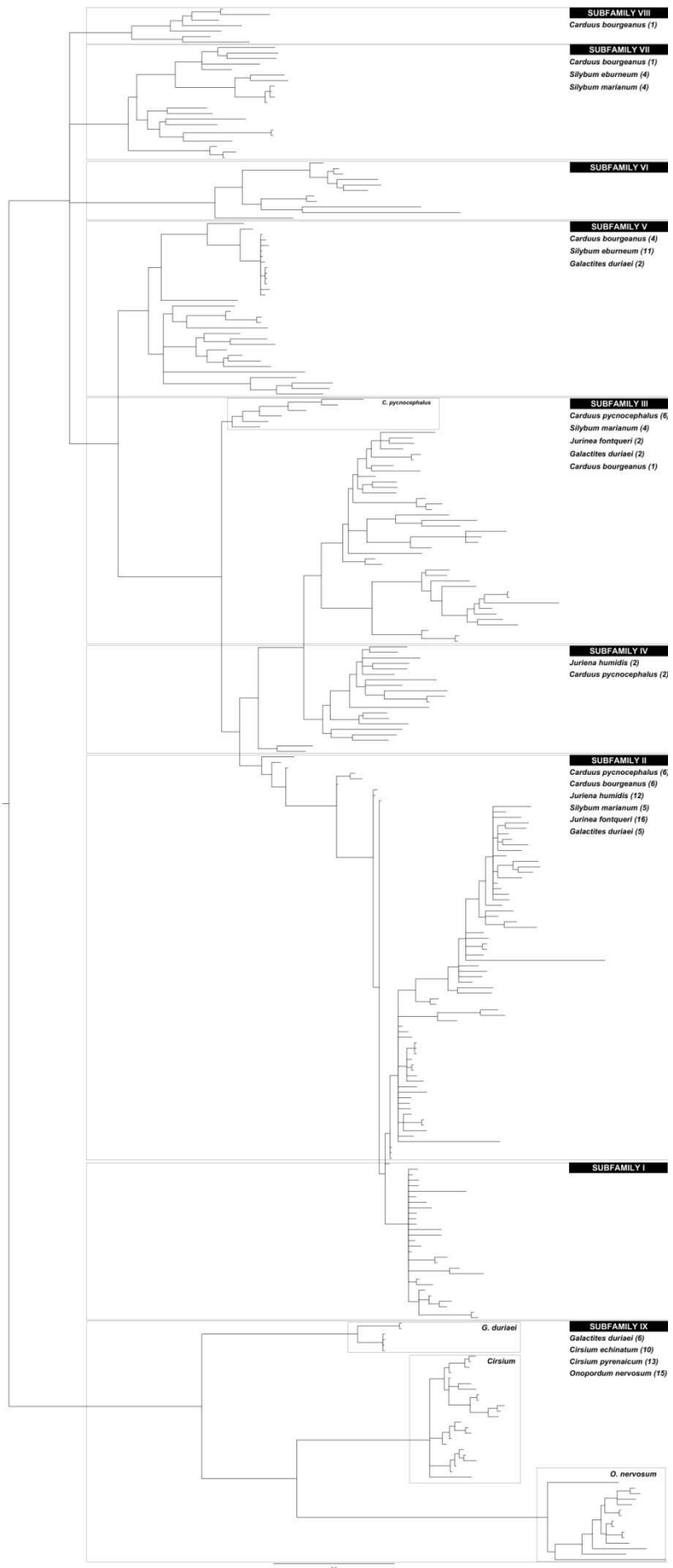


Figura 23. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias *Hinf*I de las subtribus *Carduinae* y *Centaureinae*. Se muestra un árbol simplificado de secuencias *Hinf*I aisladas de las diferentes especies de las subtribus *Carduinae* y *Centaureinae* en el que se han eliminado los nombres de las secuencias en los nodos terminales, y se han marcado en recuadros las especies de *Carduinae*. Los cuadros relacionan los clados con las subfamilias. Entre paréntesis se indica el número de secuencias identificadas en cada especie para la subfamilia correspondiente.

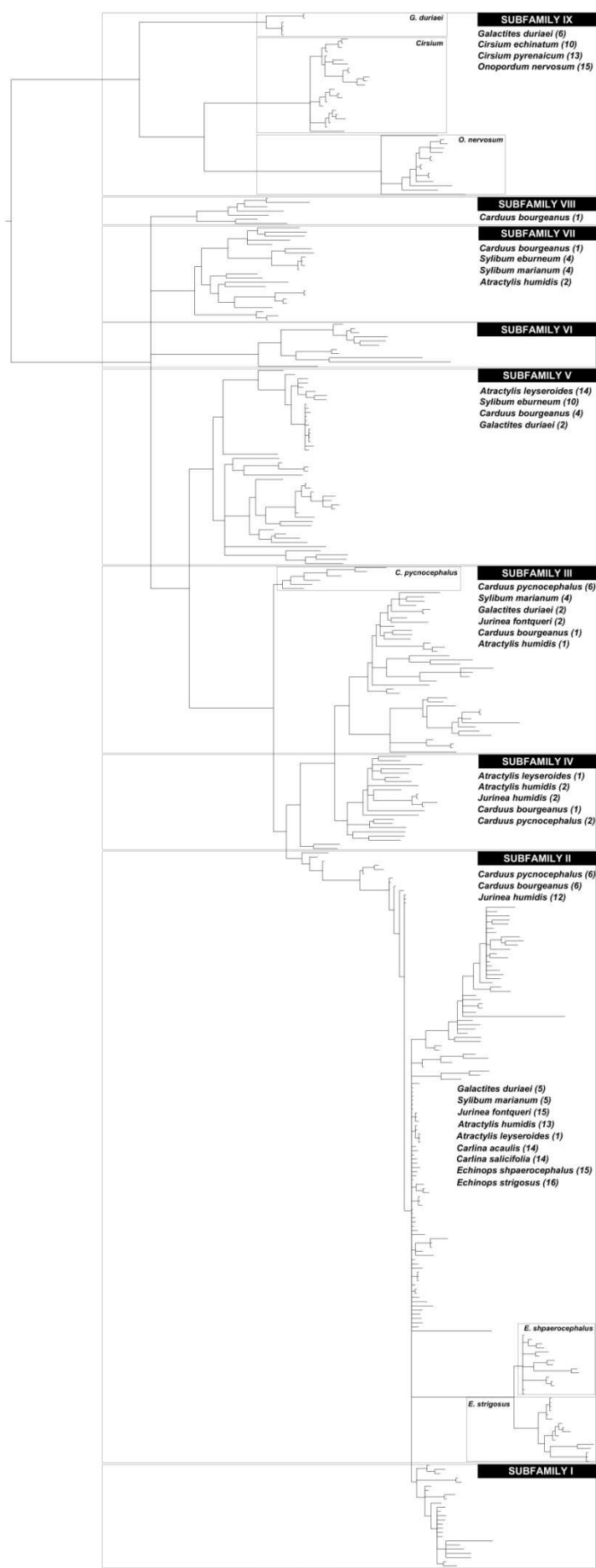


Figura 24. Filogenia obtenida mediante inferencia bayesiana basada en las secuencias *HinfI* de las subtribus de la tribu Cardueae. Se muestra un árbol simplificado de secuencias *HinfI* aisladas de las diferentes especies de las distintas subtribus (*Carlininae*, *Echinopsinae*, *Carduinae* y *Centaureinae*) de la tribu *Cardueae* en el que se han eliminado los nombres de las secuencias en los nodos terminales. Entre paréntesis se indica el número de secuencias identificadas en cada especie para la subfamilia correspondiente. Ver la versión extendida en el ANEXO V.

V.3. Discusión

Las secuencias HinfI se encuentran presentes en los genomas de todas las especies analizadas de las subtribus Carlininae, Echinopsinae y Carduinae, además de en las especies de la subtribu Centaureinae, por lo que está presente en prácticamente todas las especies de la Tribu Cardueae, la mayor de la familia de las compuestas (Susanna y García-Jacas, 2009), con lo que podemos estimar que esta familia de ADN satélite HinfI data de hace unos 29-24 m.a. (Kim *et al.*, 2005; Wang *et al.*, 2007). Que una familia de ADN satélite se conserve en un grupo tan amplio de especies y que tenga una antigüedad como la mencionada no es muy común entre las familias de ADN satélite, especialmente en plantas, en las que la familia de ADN satélite más antigua fue caracterizada en cicas (Cafasso *et al.*, 2003).

V.3.1. Subtribu Centaureinae

Entre las secuencias monoméricas HinfI, se han identificado hasta ocho subfamilias. Los cebadores utilizados para la amplificación por PCR fueron diseñados en regiones comunes a las secuencias de ADN de las subfamilias I y II identificadas mediante clonación convencional en *Centaurea boissieri* subsp. *funkii* (Suárez-Santiago *et al.*, 2007a). La pareja de cebadores amplificó el ADN satélite HinfI no sólo en las diferentes especies de *Centaurea*, sino también en especies de diferentes géneros de la subtribu Centaureinae. Este resultado no resulta inesperado si se considera que la región en la que se diseñaron los cebadores se encuentra muy conservada tanto entre las especies como entre las subfamilias descritas. Además, los cebadores han amplificado no solo las subfamilias I y II en las especies analizadas, sino también otras subfamilias (hasta ocho subfamilias diferentes). De hecho, en las especies más alejadas filogenéticamente de *Centaurea*, las secuencias amplificadas por los cebadores fueron aquellas propias de estas especies y que son diferentes a las subfamilias I y II. De este modo, se puede afirmar que no ha habido un sesgo en la amplificación de las diferentes subfamilias dentro de cada especie en particular debido a los cebadores, sino que, por el contrario, los cebadores amplifican los diferentes tipos de repeticiones HinfI de cada especie. Esta amplificación de secuencias monoméricas al azar de los cebadores debe estar además relacionada con la abundancia relativa de cada tipo de repetición en las distintas especies.

En general, estas subfamilias se han amplificado diferencialmente de acuerdo a las dos fases de radiación acaecidas en la evolución de la subtribu. Según García-Jacas *et al.* (2001) y Hellwig (2004), ha habido dos periodos de radiación que han generado dos grupos en la subtribu: el clado derivado y los grupos basales. El clado derivado, generado durante la segunda radiación, engloba la mayoría de las especies actualmente clasificadas como pertenecientes al

género *Centaurea* (subgéneros *Centaurea*, *Cyanus* y *Acrocentron*), y el complejo *Carthamus* (incluidos los géneros *Carthamus*, *Phonus* y *Carduncellus*). El resto de géneros forman parte de los grupos basales (primera radiación): *Psephellus*, *Rhaponticoides*, *Klasea*, *Serratula*, *Volutaria*, *Crupina* y *Cheirolophus*, entre otros (García-Jacas *et al.*, 2001).

Las subfamilias V a VIII se encuentran en los géneros más antiguos (primera radiación) (Tabla 8). La subfamilia V es el único tipo de secuencias presentes en los genomas de las especies pertenecientes a los géneros *Klasea* y *Rhaponticum*, identificándose unas pocas repeticiones pertenecientes a esta subfamilia en especies pertenecientes a los géneros *Rhaponticum* (tres repeticiones en *R. acaule*) y *Volutaria* (una repetición en *V. lippi*). La subfamilia VI se ha expandido predominantemente en dos de las cuatro especies de género *Volutaria* analizadas (*V. muricata* y *V. tubuliflora*). No obstante, se han identificado algunas secuencias de la subfamilia VI en dos especies pertenecientes al género *Carduncellus* (segunda radiación), y su presencia es también testimonial en los géneros *Rhaponticum* y *Crupina* (primera radiación). Las subfamilias VII y VIII son casi exclusivas del género *Cheirolophus*, aunque también están presente en especies de *Crupina* y en el género *Centaurea* (una repetición en *C. cyanus*).

Estos datos sugieren que todas estas variantes de las secuencias *HinfI*, clasificadas como distintas subfamilias, estaban probablemente presentes en el ancestro común a todas las especies de la subtribu *Centaureinae*, y que cada una de ellas se ha amplificado en diferentes géneros, quedando en el genoma de algunos de estos géneros algunas copias de alguna de las otras subfamilias. Esta hipótesis es también compatible con la distribución de las subfamilias I a IV. Estas subfamilias se han expandido predominantemente en el genoma de las especies pertenecientes al clado derivado de la subtribu *Centaureinae*. Así, los genomas de las especies del género *Centaurea* presentan las subfamilias *HinfI* I, II y III, las de los géneros *Carthamus* y *Phonus* solo la subfamilia III, y las del género *Carduncellus* las de la subfamilia IV. Sin embargo, algunas especies de este último género presentan también unas pocas copias de las subfamilias III y VI. Cabe destacar que varias de las especies que constituyen el grupo basal portan las subfamilias I a III como el tipo de secuencias mayoritarias en sus genomas. Por ejemplo, esto ocurre en las especies de los géneros *Rhaponticum* y *Crupina* (presentan también secuencias de las subfamilias VI y VII), así como *Volutaria crupinoides*, *V. lippii* y *Cheirolophus teydis*. En estas dos últimas especies, la subfamilia III predomina entre las unidades reptadas secuenciadas, pero también presentan repeticiones tipo I y/o II. *V. crupinoides* cuenta con repeticiones de las subfamilias II y III, pero también de las subfamilias IV y VI. *Rhaponticum acaule* tiene secuencias tipo II, IV, V y VI. Por tanto, todos estos datos sugieren que estas subfamilias se han expandido recientemente reemplazando a otras subfamilias en *Centaurea* (I-III), *Phonus* y *Carthamus* (III) y *Carduncellus* (IV) pero también

en géneros antiguos como *Crupina* (I y II), *Rhaponticoides* (I y II), *Volutaria* (*V. crupinoides*, II, III y IV y *V. lippi*, mayoritariamente de la III) o *Cheirolophus*, *C. teydis* (mayoritariamente III). Teniendo en cuenta todos estos datos, consideramos que las subfamilias I a IV estaban también presentes en el ancestro común de todas las especies integrantes de la subtribu Centaureinae, amplificándose diferencialmente cada una en diferentes géneros (Tabla 8).

El reemplazamiento de una secuencia por otra variante en diferentes especies es un rasgo común a las familias de ADN satélite que puede ser consecuencia del propio dinamismo de la evolución de este tipo de secuencias repetidas (Plohl *et al.*, 2010, 2012). Los mecanismos moleculares de intercambio no recíproco (entrecruzamiento desigual, conversión génica, replicación mediante círculo rodante y re inserción, e intercambio mediado por transposones) pueden expandir nuevas variantes que aparecen en las unidades de repetición individuales de una familia de secuencias, y los cambios se fijan en una población en la que los individuos se aparean aleatoriamente mediante reproducción sexual a través del proceso de dos etapas denominado *molecular drive*, que conduce a la evolución concertada (Plohl *et al.*, 2010, 2012; Pérez-Gutiérrez *et al.*, 2012). Los cambios en el ADN satélite debido a la acumulación gradual de variabilidad en las secuencias dan lugar a divergencia de las secuencias repetidas en grupos de organismos aislados reproductivamente (Plohl *et al.*, 2012). No obstante, dada la dispersión de las subfamilias en cada linaje de la tribu Centaureinae, otra hipótesis alternativa (pero no excluyente con la anterior) puede explicar esta distribución diferencial de las secuencias. De acuerdo con la hipótesis de la biblioteca, distintos taxa relacionados comparten una biblioteca de secuencias conservadas de ADN satélite diferentes: distintas familias de ADN satélite, pero también pueden ser diferentes variantes monoméricas o subfamilias de una misma familia de ADN satélite como en el caso que nos ocupa, de modo que cualquiera de ellas puede ser amplificada diferencialmente en cada especie, con el subsiguiente reemplazamiento de una secuencia por otra variante en diferentes especies (Mravinac *et al.*, 2002; Meštrović *et al.*, 2006) Cuando esto ocurre, puede darse el caso de que exista una variante que es mayoritaria en una especie y, a su vez, que existan unas pocas copias conservadas de dicha variante en otras especies. De la misma forma, la primera especie también conservaría algunas copias de otras variantes en su genoma. Por supuesto, la comparación de secuencias del mismo tipo entre especies diferentes reflejará (como es el caso que nos ocupa) una elevada conservación interespecífica y una completa ausencia de mutaciones diagnósticas específicas de especie, tal y como ha sido descrito previamente en escarabajos del género *Palorus* (Meštrović *et al.*, 1998). Esta hipótesis ha sido probada en diferentes especies de insectos (Meštrović *et al.*, 1998; Mravinac *et al.*, 2002; Cesari *et al.*, 2003; Pons *et al.*, 2004), y podría explicar los resultados de la dispersión de las repeticiones HinfI obtenidos en la subtribu Centaureinae.

De cualquier modo, la variabilidad global de los monómeros de un ADN satélite en un genoma es un rasgo complejo que depende de diferentes factores como su localización, organización y el número de repeticiones (Navajas-Pérez *et al.*, 2005), el tiempo de divergencia (Pérez-Gutiérrez *et al.*, 2012), factores biológicos (Luchetti *et al.*, 2003; Robles *et al.*, 2004; Suárez-Santiago *et al.*, 2007b) o limitaciones funcionales (Mravinač *et al.*, 2005). De hecho, algunos patrones de evolución de las repeticiones *HinfI* en linajes específicos de *Centaureinae* podrían haber resultado de la influencia de algunos de estos factores.

Así, cabe destacar que, como discutiremos en las siguientes secciones, la variabilidad observada en las especies de *Centaureinae* para la familia de ADN satélite *HinfI* tiene al menos tres fuentes que hacen que tengamos observaciones como la de que en una misma especie convivan diferentes unidades repetitivas de diferentes subfamilias *HinfI*. Por un lado, como hemos mencionado más arriba, los efectos de la amplificación diferencial de secuencias a partir de una biblioteca ancestral de variantes diferentes. Por otro lado, la escasa diferenciación existente entre especies, la cual puede deberse bien a un escaso tiempo de divergencia entre las especies comparadas desde que se separaron de un ancestral común, bien a una historia evolutiva de hibridación recurrente entre especies.

Clado derivado

Según la filogenia de la subtribu *Centaureinae* (García-Jacas *et al.*, 2001), dentro del clado derivado, el complejo *Carthamus* ocupa la posición basal, mientras que los subgéneros *Centaurea* (*Jacea*) y *Cyanus* del género *Centaurea* ocupan la posición más derivada, cuya estrecha relación como clados hermanos está firmemente establecida (Susanna y García-Jacas, 2009). El subgénero *Acrocentron* del género *Centaurea* ocupa una posición intermedia, aunque no quedan totalmente claras las conexiones entre este subgénero y los subgéneros *Jacea* y *Cyanus* (García-Jacas *et al.*, 2001; Susanna and García-Jacas, 2009).

Con los resultados obtenidos en este trabajo, podemos poner de relieve dos observaciones con respecto a la dispersión y evolución de las secuencias *HinfI* en este grupo.

La primera es que se han identificado cuatro subfamilias *HinfI* (denominadas I, II, III y IV) que se han expandido a través de los genomas de estas especies durante el proceso de especiación. No obstante, las cuatro subfamilias no están igualmente distribuidas en todos los taxa analizados (Tabla 8). Las subfamilias I y II se han expandido en segundo lugar como subfamilias específicas del género *Centaurea*, reemplazando casi completamente a la subfamilia III y completamente a la subfamilia IV, mientras que la subfamilia III se ha amplificado en *Phonus* y *Carthamus*, y la subfamilia IV lo ha hecho en *Carduncellus*. Esta hipótesis se ve corroborada por el árbol filogenético mostrado en la Figura 17, en el que las subfamilias I y II

aparecen en clados hermanos, mientras que la subfamilia IV es basal a las otras tres subfamilias. Sin embargo, las cuatro subfamilias han sido también identificadas en especies pertenecientes al grupo de especies basales (primera radiación). Además de la presencia de algunas copias de estas cuatro subfamilias en algunas de las especies de los grupos basales, que apoyan esta hipótesis, en algunas de estas especies (*Cheirolophus teydis*, *Volutaria lippi* y especies del género *Crupina*), estas subfamilias constituyen los principales componentes de las secuencias *HinfI* dentro de sus genomas. Este hecho apoya la hipótesis de la biblioteca tal y como se ha discutido previamente: estas subfamilias serían tan antiguas como el resto de subfamilias identificadas en este trabajo, que se han expandido o amplificado diferencialmente en los distintos linajes. Independientemente de la filogenia, de acuerdo con la hipótesis de la biblioteca, existe una convergencia en la expansión de algunas variantes en diferentes linajes. Esto nos recuerda que la hipótesis no predice ni qué variante de la biblioteca será amplificada, ni si alguna estará favorecida por presión selectiva alguna (Fry y Salser, 1977).

En segundo lugar, queremos destacar los aspectos relacionados con la evolución concertada de estas secuencias. Patrones de especiación diferencial han dado lugar a diferentes patrones de evolución de las secuencias *HinfI* en diferentes linajes. Diferentes subfamilias del ADN satélite *HinfI* coexisten dentro de la mayoría de los taxa analizados en el género *Centaurea*. La presencia de diferentes subfamilias en los genomas, igualmente representadas en la mayoría de los casos, puede explicarse como resultado de una evolución reticulada dentro de una parte de este género (Suárez-Santiago *et al.*, 2007a, b). La hibridación que ocurre entre diferentes especies de este género debe estar implicada en el mantenimiento de las diferentes subfamilias del ADN satélite *HinfI* dentro de un genoma en particular. Además, el flujo génico entre especies reduciría la cantidad de diferencias genéticas entre esas especies y aumentaría la variación intraespecífica. Así, considerando el escenario evolutivo propuesto, en contra de lo que se esperaría según el modelo de evolución concertada, se encontrarían valores similares o incluso más altos de variación intraespecífica que de divergencia interespecífica. Además, la hipótesis de la biblioteca podría explicar la existencia de un número bajo de copias de algunas subfamilias en algunas especies. Las especies de los géneros *Carthamus* y *Phonus* presentan solo secuencias de tipo III, y en ellas, la variación intraespecífica es menor que la divergencia interespecífica, un indicio de evolución concertada. *Carthamus* y *Phonus*, anteriormente considerados sinónimos, son dos géneros separados (Vilatersana *et al.*, 2000). De hecho, se ha demostrado que el género *Phonus* se encuentra más cerca del género *Carduncellus* que de *Carthamus*, y presentan rasgos biológicos y hábitats diferenciados, así como distinta historia biogeográfica y evolutiva (Vilatersana *et al.*, 2000). El género *Carthamus* contiene dos grupos bastante diferentes: *Carthamus sensu stricto*, que incluye solo la sección *Carthamus* y del que

C. tinctorius es la especie tipo, y la sección *Atractylis*, con la especie tipo *C. lanatus* (Vilatersana *et al.*, 2005). Las secuencias tipo III de ambas especies aparecen diferenciadas en el árbol de secuencias de esta subfamilia (Figura 18). Por otro lado, no se aprecia evolución concertada en las secuencias de *Carduncellus*, la especie portadora de la subfamilia IV. La variabilidad intraespecífica es mayor que la divergencia interespecífica en este género, y las secuencias *HinfI* aparecen entremezcladas en el árbol filogenético de la Figura 15 (ver también Anexo IV) y en los árboles inferidos a partir de las secuencias de *Carduncellus* únicamente (Figura 15 y Anexo IV). La especiación en *Carduncellus* es muy reciente en relación con la de otros géneros de Centaureinae, y los valores de divergencia entre las secuencias *HinfI* son muy bajos, al igual que se ha visto que ocurre con los valores de divergencia para la región ITS en un estudio llevado a cabo por Vilatersana *et al.*, 2000. La imposibilidad de resolver las relaciones filogenéticas dentro del género *Carduncellus* pone de manifiesto el bajo nivel de variación genética en muchos géneros de origen reciente (Neporkroeff y Sytsma, 1996; Susanna *et al.*, 1999). De hecho, en ausencia de restricciones selectivas y biológicas, el ritmo de la evolución concertada de una familia de secuencias de ADN satélite debe depender básicamente del tiempo de divergencia entre especies (Pérez-Gutiérrez *et al.*, 2012).

Grupos basales

De nuevo, resaltaremos dos observaciones en relación a la distribución y evolución de las secuencias *HinfI* en estas especies.

Primero, hay cuatro subfamilias *HinfI* (V, VI, VII y VIII) que se han expandido a través de los genomas de estas especies acompañando a los procesos de especiación más importantes. Al contrario que las subfamilias I a IV, en este caso, cada subfamilia se ha expandido casi exclusivamente dentro de uno o dos géneros específicos: la subfamilia V es exclusiva de los géneros *Rhaponticum* y *Klasea*, la subfamilia VI se ha amplificado en algunas especies de *Volutaria* y las subfamilias VII y VIII se amplificaron en la mayoría de especies analizadas de *Cheirolophus*. Podemos concluir en este caso que estas amplificaciones son antiguas y restringidas a linajes específicos. La presencia de algunas copias conservadas de estos tipos de repeticiones *HinfI* en un par de especies adicionales, de nuevo apoyan la hipótesis de la biblioteca.

Segundo, en referencia a la evolución concertada, una vez más podemos afirmar que patrones de especiación diferencial han dado lugar a diferentes patrones de evolución de las secuencias *HinfI* en diferentes linajes. Las secuencias de la subfamilia V evolucionan de manera concertada dentro del clado *Klasea-Rhaponticum* (especies que comparten la subfamilia V). Todas estas secuencias son, por lo general más similares dentro de una misma especie que entre

especies, y tienden a agruparse según la especie a la que pertenecen, confirmando así que la expansión de esta subfamilia precedió a la diversificación del género y que esta especiación es antigua dentro de la tribu *Centaureinae*. Nuestros resultados confirman la estrecha relación entre las especies de los géneros *Klasea* y *Rhaponticum* (Martins y Hellwig, 2005; Hidalgo *et al.*, 2006). La subfamilia VI de secuencias *HinfI* se encuentra en *Volutaria muricata* y *V. tubuliflora* (así como de manera residual en un par de especies más). Las secuencias de estas especies se agrupan por separado, un signo de homogeneización dentro de las especies y de diferenciación entre las dos especies (Figura 20), es decir, un signo de evolución concertada. Sin embargo, las secuencias de las subfamilias VII y VIII están distribuidas igualmente por todas las especies de *Cheirolophus* analizadas en este estudio (excepto in *C. teydis* como se ha discutido previamente). Dentro de este género, las secuencias *HinfI* de diferentes especies son más similares entre sí que cuando se comparan con secuencias homólogas de otras especies, y aparecen entremezcladas en los árboles filogenéticos. Como en el caso de *Carduncellus*, Susanna *et al.* (1999) determinaron que la especiación en *Cheirolophus* es muy reciente en relación a otros taxa de la subtribu *Centaureinae*, lo que podría ser un factor que influye en la ausencia de evolución concertada (Pérez-Gutiérrez *et al.*, 2012).

V.3.2. Subtribus *Carlininae*, *Echinopsinae*, y *Carduinae*

Tal y como ocurre para las secuencias *HinfI* en *Centaureinae*, hemos podido identificar diferentes tipos de secuencias o subfamilias en los distintos genomas de las especies analizadas pertenecientes a estas subtribus. Teniendo en cuenta la filogenia y las distancias genéticas, podemos adscribir las secuencias *HinfI* de cada especie a subfamilias diferentes. En este contexto, en un sentido amplio, si como ya hemos hecho anteriormente en esta memoria, nos tomamos la licencia de denominar repeticiones “ortólogas” a todas aquellas derivadas de la amplificación de un mismo tipo de repetición en un único o varios eventos de amplificación, podríamos tratar a las diferentes subfamilias como grupos “parálogos” compuestos por repeticiones ortólogas de la familia *HinfI* (Tabla 9). Como consecuencia de la comparación de subfamilias ortólogas con parálogas, este trabajo revela que las comparaciones entre secuencias pertenecientes a especies filogenéticamente distantes da lugar, en la mayoría de los casos, a distancias genéticas menores y relaciones filogenéticas más cercanas que cuando se comparan secuencias de especies menos distantes, más cercanas filogenéticamente, dependiendo del tipo de subfamilia que presentan en su genoma. Éste es el caso de la subtribu *Echinopsinae*, en la que las secuencias *HinfI* son ortólogas a las secuencias de la subfamilia II de *Centaureinae* (Tabla 9). Las secuencias *HinfI* de tres de las especies analizadas de la subtribu *Carlininae* (*Carlina salicifolia* y *C. acaulis*, y *Atractylis humilis*) son ortólogas también a las secuencias de

la familia II de *Centaureinae* (Tabla 9), mientras que las secuencias de la cuarta especie analizadas de esta subtribu, *Atractilis leyseroides*, pertenecen a la subfamilia V (Tabla 9). Así, en la subtribu *Carduinae* podemos diferenciar dos grupos de secuencias (Figuras 23 y 24 y Tabla 9). Por un lado, están el grupo de secuencias compuestas por las aisladas de los géneros *Jurinea* (grupo *Jurinea-Saussurea*) y *Silybum* y *Carduus* (grupo *Carduus-Cirsium*), cuyos genomas portan varios tipos de repeticiones *HinfI* pertenecientes a las subfamilias II a V, dependiendo de la especie, y también pertenecientes a las subfamilias VII y VIII en un par de casos. El árbol de las Figuras 23 y 24 refleja este hecho, con secuencias de estas especies diseminadas en varios clados de acuerdo con su ortología más que de acuerdo con la filogenia de las especies. Además, se puede identificar una nueva subfamilia muy diferenciada del resto, específica de los géneros *Cirsium* y *Onopordum*, que se encuentra presente también en el genoma de *Galactites duriaei* (seis de las quince secuencias analizadas en esta especie pertenecen a esta nueva subfamilia mientras que el resto pertenecen a las subfamilias II, III y V). Siguiendo la nomenclatura anterior, se ha denominado a esta nueva variante como subfamilia IX (Tabla 9).

Tabla 9. Distribución por géneros de las subfamilias *HinfI* en la subtribus *Carlininae*, *Echinopsinae* y *Carduinae*. Los números entre paréntesis representan el número de secuencias de la subfamilia identificadas en algunas de las especies.

Género	Especie	Subfamilia <i>HinfI</i>								
		I	II	III	IV	V	VI	VII	VIII	IX
<i>Carlina</i>	<i>salicifolia</i>		14			1				
	<i>acaulis</i>		14							
<i>Atractylis</i>	<i>leyseroides</i>				1	14				
	<i>humilis</i>		14		2			2		
<i>Echinops</i>	<i>sphaerocephalus</i>		16							
	<i>strigosus</i>		16							
<i>Jurinea</i>	<i>fontqueri</i>		16	2						
	<i>humilis</i>		12		2					
<i>Silybum</i>	<i>marianus</i>		5	4				4		
	<i>eburneum</i>					11		4		
<i>Carduus</i>	<i>bourgeanus</i>		6	1	1	4		1	1	
	<i>pycnocephalus</i>		6	8						
<i>Galactites</i>	<i>duriaei</i>		5	2		2				6
<i>Cirsium</i>	<i>echinatum</i>									10
	<i>pyrenaicum</i>									13
<i>Onopordum</i>	<i>nervosum</i>									15

Estos resultados confirman los descritos anteriormente en la subtribu *Centaureinae* (ver más arriba) Así, una biblioteca de variantes de las secuencias *Hinfl* debió estar presente en el ancestral común de, no solo las especies de la subtribu *Centaureinae*, sino de las especies de toda la Tribu *Cardueae*. A partir de esta biblioteca, cualquier variante se habría amplificado diferencialmente en los distintos linajes con el consiguiente reemplazamiento de una variante por otra en las diferentes especies de la Tribu *Cardueae* (Plohl *et al.*, 2012). Además, como se ha mencionado más arriba, la hipótesis de la biblioteca predice que las repeticiones de una subfamilia dominante en una especie podrían estar presentes en un bajo número de copias en otras especies y las comparaciones entre ellas mostrarían una elevada conservación interespecífica de la secuencia, así como la total ausencia de mutaciones diagnósticas específicas de especie (Plohl *et al.*, 2012;) tal y como hemos identificado en la tribu *Cardueae*. La dispersión de las subfamilias *Hinfl* en diferentes especies de la tribu *Cardueae* y el hecho de que se compartan distintas subfamilias amplificadas diferencialmente dentro de la misma especie, apoyan esta hipótesis y es una prueba de la evolución saltatoria del ADN satélite. Este modo de evolución ha sido descrito anteriormente en insectos (Meštrović *et al.*, 1998; Mravinac *et al.*, 2002; Cesari *et al.*, 2003) y en anfibios (Felicciello *et al.*, 2005, 2006). No obstante, en este trabajo se analiza en profundidad por primera vez una familia de ADN satélite en un amplio grupo de especies de plantas donde se demuestra la hipótesis de la biblioteca. Queda aún una cuestión importante por aclarar, que es la coexistencia de repeticiones de diferentes subfamilias con abundancia similar dentro del mismo genoma. En estos casos, podrían estar implicados otros factores como la evolución reticulada tal y como ha sido descrito en el género *Centaurea* (Suárez-Santiago *et al.*, 2007). Una explicación que podría valer también para las observaciones hechas en *Galactites duriaei* en cuyo genoma conviven varias subfamilias y se ha documentado que es una especie que suele hibridar con especies relacionadas. Curiosamente, las hibridaciones también están documentadas en el género *Cirsium* y en los genomas de sus especies sólo se conserva un tipo de variante *Hinfl*, y además las secuencias de las dos especies comparadas están diferenciadas. No obstante, la amplificación aleatoria o la conversión génica (Drouin and Dover, 1990; Plohl y Ugarković, 1994 a y b) de diferentes variantes *Hinfl* en el mismo linaje o el escaso tiempo de divergencia desde la separación entre las especies comparadas (Pérez-Gutiérrez *et al.*, 2012) podrían también explicar este tipo de variabilidad observada en algunas especies de la Tribu *Cardueae*.

Desde un punto de vista filogenético, los resultados descritos para estas subtribus podrían parecer desconcertantes, ya que las secuencias de especies de diferentes subtribus se agrupan juntas en un mismo clado en los árboles filogenéticos y las distancias genéticas entre ellas son menores que cuando se comparan secuencias aisladas a partir de especies de la misma subtribu si pertenecen a diferentes subfamilias, incluso cuando se comparan secuencias del

mismo género o de la misma especie. Por este motivo, es coherente la licencia de hablar de subfamilias parálogas. De hecho, cuando las secuencias del mismo tipo (repeticiones ortólogas de la misma subfamilia) se comparan, las secuencias se agrupan juntas de acuerdo con la delimitación actualmente aceptada de las subtribus. Esto ocurre en el caso de especies de las subtribus *Carlininae* y *Echinopsinae*. Cuando se tiene esto en cuenta, lo que el análisis llevado a cabo en este estudio encuentra para cada subfamilia cuando se comparan secuencias de diferentes subtribus, sería el resultado de la diferenciación gradual en las secuencias de *HinfI* por *molecular drive* (Plohl *et al.*, 2012; Pérez-Gutiérrez *et al.*, 2012). Sin embargo, la tasa de evolución concertada depende de la localización del ADN satélite, de su organización y del número de repeticiones (Navajas-Pérez *et al.*, 2005, 2009), del tiempo (Pérez-Gutiérrez *et al.*, 2012), de limitaciones funcionales (Mravinac *et al.*, 2005) o de factores biológicos (Luchetti *et al.*, 2003, 2006, Robles *et al.*, 2004; Suárez-Santiago *et al.*, 2007a). En ausencia de restricciones selectivas y biológicas, la tasa de evolución concertada de una familia de secuencias de ADN satélite debe depender básicamente del tiempo de divergencia entre especies (Pérez-Gutiérrez *et al.*, 2012). Sin embargo, no hemos encontrado diferenciación de las secuencias dentro *Carlininae*. Las secuencias *HinfI* de la subfamilia II aisladas a partir de estas especies aparecen entremezcladas en el árbol filogenético, una indicación de la ausencia de evolución concertada, que podría ser explicada como consecuencia del tiempo transcurrido (Pérez-Gutiérrez *et al.*, 2012) asumiendo una separación reciente entre las especies comparadas, a pesar de que pertenecen a dos géneros diferentes. Por contra, hemos encontrado también en este estudio que las secuencias de dos especies diferentes se agrupan por separado en dos subclados diferentes, una indicación de evolución concertada. Alternativamente, se podría asumir que existieran diferentes tasas de cambio en la secuencia de estos dos linajes diferentes (Adegoke *et al.*, 1993).

La situación en la subtribu *Carduinae* no es tan simple, ya que podemos distinguir entre dos grupos de secuencias, tal y como se ha mencionado anteriormente. Por un lado, el conjunto compuesto por las repeticiones *HinfI* de los géneros *Jurinea* (grupo *Jurinea-Saussurea*) y *Carduus* y *Silybum* (grupo *Carduus-Cirsium*). En este caso, no hay una diferenciación entre las secuencias de ambas subtribus, *Centaureinae* y *Carduinae*, sino que sus secuencias aparecen entremezcladas sin agrupación por afinidad taxonómica. Por otro lado, las secuencias de *Onopordum* (grupo *Onopordum*) y *Cirsium* (grupo *Carduus-Cirsium*), así como *Galactites duriaei* (*Carduus-Cirsium* grupo) pertenecen a una nueva subfamilia (subfamilia IX) muy diferenciada del resto de subfamilias, no sólo de aquellas aisladas de especies de la misma subtribu, sino también de las presentes en cualquier otra subtribu. Estos datos apoyan la parafilia de la subtribu *Carduinae* (García-Jacas *et al.*, 2002; Susanna *et al.*, 2006; Susanna y García-Jacas, 2007, 2009). Los resultados encontrados para las secuencias de *HinfI* de *Jurinea*, *Carduus* y *Silybum* reflejan el mismo comportamiento que estas secuencias tienen en

Centaureinae, sin diferenciación genética entre sus especies. Por otra parte, las distancias genéticas entre secuencias HinfI de ambas subtribus para las repeticiones de la subfamilia II son menores que entre las especies de la subtribu Centaureinae en algunos casos. Sin embargo, *Cirsium*, *Onopordum* y *Galactites* se diferencian claramente de las especies anteriores. Estos datos no se basan únicamente en la presencia de la subfamilia IX en sus genomas, una consecuencia de la amplificación de esta subfamilia específica en estos géneros. Además de subfamilia IX, *Galactites duriaei* tiene también en su genoma secuencias de las subfamilias II, III y V, y, en todos los casos, las distancias genéticas entre las secuencias de HinfI de esta especie y las de Centaureinae fueron mayores (distancias genéticas medias que oscilan entre 0,07 y 0,23, dependiendo de la subfamilia) que la variabilidad intraespecífica (0,01 en todos los casos), un grado similar de divergencia a la que se encuentra en las comparaciones entre otras subtribus. Además, las secuencias de HinfI de la subfamilia IX han divergido entre los diferentes géneros analizados, y las repeticiones se agrupan en tres clados que coinciden con los tres géneros comparados (*Galactites*, *Cirsium* y *Onopordum*), lo que demuestra la divergencia progresiva presente en las secuencias de ADN satélite compartidas por un determinado grupo de especies relacionadas.

En este trabajo se ha pretendido profundizar en el estudio de la evolución del ADN satélite en el genoma de plantas. Para ello, hemos estudiado dos sistemas que nos permitían analizar la influencia de factores tanto intrínsecos al propio genoma, como extrínsecos. Los resultados obtenidos se corresponden con los cuatro análisis principales realizados y que nos llevan a las cuatro conclusiones siguientes:

PRIMERA:

Hemos analizado el efecto que tienen factores como la localización, la organización y el número de copias en el proceso de *molecular drive* que conduce al patrón de evolución concertada del ADN satélite. Para ello, hemos estudiado el ADN satélite RAE180 de las especies dioicas del género *Rumex*. En las especies XX/XY, el ADN satélite RAE180 se localiza en un único locus autosómico pequeño, mientras que en las especies XX/XY₁Y₂ además de estar presente en un locus autosómico pequeño se ha amplificado masivamente en los cromosomas Y. En este trabajo hemos encontrado que las repeticiones RAE180 del locus autosómico en las especies XX/XY siguen un patrón general de evolución concertada. Por el contrario, en las especies XX/XY₁Y₂ tanto en el locus autosómico como en el ligado al cromosoma Y, la variabilidad ancestral se ha mantenido, y se observan tasas bajas de homogeneización y de evolución concertada de las secuencias. Así, este estudio demuestra que los mecanismos moleculares de intercambio no recíproco son factores claves en el proceso de *molecular drive* de tal manera que una misma familia de ADN satélite tiene un comportamiento diferente dependiendo de que se encuentre en la región diferencial del cromosoma Y, una región con baja tasa de recombinación, o en autosomas. No obstante, incluso en regiones autosómicas, puede darse una alternancia entre periodos evolutivos de cambio rápido en las secuencias, y periodos evolutivos de estancamiento en los que la variabilidad se mantiene debido a la acción reducida de los mecanismos moleculares de intercambio no recíproco, como ha ocurrido en el locus autosómico de las especies XX/XY₁Y₂ lo cual podría depender a su vez del número de copias y de los procesos implicados en la amplificación de estas secuencias.

SEGUNDA:

En la especie *Rumex hastatulus*, la raza de Texas se caracteriza por un sistema cromosómico sexual XX/XY, mientras que la raza de Carolina del Norte ha evolucionado hasta un sistema cromosómico sexual complejo XX/XY₁Y₂. Sin embargo, los cromosomas sexuales en los individuos de la raza de Carolina del Norte son aún eucromáticos, y no han acumulado secuencias de ADN satélite, lo que contrasta con el resto de especies dioicas del género *Rumex* con cromosomas XX/XY₁Y₂. En *R. hastatulus*, hemos detectado la existencia de tres subfamilias RAE180. No obstante, mientras que en la raza de Texas la subfamilia TX1/NC1 es la más abundante, en la raza de Carolina del Norte la subfamilia más abundante es la TX2/NC2, por lo que en ésta última parece haberse producido un reemplazamiento de una por otra. El análisis filogenético de las secuencias RAE180 de *R. hastatulus* más las de otras cinco especies dioicas del género *Rumex* apoya la hipótesis de la biblioteca, según la cual asumimos que en el genoma ancestral de las especies dioicas de *Rumex* se encontraban presentes un conjunto de variantes RAE180 a partir del cual cada una de ellas se amplificó diferencialmente en diferentes linajes. La amplificación diferencial del ADN satélite RAE180 en cada linaje, en diferentes momentos evolutivos, y en diferentes posiciones cromosómicas, dio lugar a distintos patrones evolutivos de las secuencias.

TERCERA:

Hemos encontrado hasta ocho subfamilias diferentes de la familia *Hinfl* de ADN satélite en la subtribu *Centaureinae*. Las subfamilias V a VIII se han encontrado predominantemente en los géneros más antiguos (primera fase de radiación dentro de la subtribu, Oligoceno-Mioceno tardíos), aunque algunas copias de estas secuencias se encontraron también en algunas especies de los géneros derivados. Las subfamilias I a IV se amplificaron mayoritariamente en las especies derivadas (segunda fase de radiación, Plioceno a Pleistoceno), aunque las repeticiones de estas subfamilias están presentes también en las especies antiguas. Los árboles filogenéticos no agrupan las repeticiones por afinidad taxonómica, sino dependiendo de la subfamilia a la que pertenecen. Mientras que en los géneros más antiguos se observa evolución concertada, ni en los grupos de radiación reciente, ni en los grupos con una historia de hibridación recurrente, se aprecia diferenciación genética entre especies, llegando incluso a coexistir varias subfamilias dentro de una misma especie. Nuestros resultados sugieren que las ocho subfamilias estaban presentes en el ancestro común de todas las especies de *Centaureinae*, y que cada una de ellas se amplificó diferencialmente en los distintos géneros durante las dos fases de radiación según el modelo evolutivo de la biblioteca. Además, diferentes vías de especiación han dado lugar a diferentes patrones de evolución de las secuencias en distintos linajes. Así, la historia evolutiva de cada grupo de *Centaureinae* se ve reflejada en la evolución del ADN satélite y nuestros datos refuerzan el valor de las secuencias de ADN satélite como marcadores de los procesos evolutivos que ocurren en los grupos biológicos.

CUARTA:

La familia de ADN satélite *Hinfl* se encuentra conservada además en el genoma de las especies del resto de subtribus de la tribu *Cardueae*, que data de hace 29-24 m.a. Analizando toda la tribu, además de las ocho subfamilias anteriores, hemos encontrado una adicional, sumando un total de nueve subfamilias diferencialmente distribuidas en diferentes especies de cada subtribu. Como en *Centaureinae*, los árboles filogenéticos agrupan a las secuencias según la subfamilia a la que pertenecen. No obstante, cuando se comparan secuencias *Hinfl* del mismo tipo, las distancias genéticas y los árboles filogenéticos muestran diferenciación de los géneros y las subtribus en todos los casos excepto cuando se comparan secuencias *Hinfl* de algunas especies de *Centaureinae* entre sí, y cuando se comparan secuencias de algunas especies de la subtribu *Carduinae* con especies de la subtribu *Centaureinae*. Por tanto, asumimos que la biblioteca de las nueve subfamilias existía ya en el ancestral común de toda la tribu *Cardueae* y que cada una de ellas se habría amplificado diferencialmente en los distintos linajes con el consiguiente reemplazamiento de una variante por otra en las distintas especies. Este proceso continuaría con la divergencia gradual entre repeticiones de una subfamilia compartida por un grupo de especies, lo que conduciría a evolución concertada a una tasa que variaría entre los diferentes linajes. Estas diferentes tasas serían las responsables de la escasa diferenciación entre especies del mismo género en algunas comparaciones. Desde un punto de vista filogenético, estos resultados apoyan la actual clasificación en subtribus y la delimitación de los géneros de *Cardueae*, y apunta a la parafilia de la subtribu *Carduinae*.

- Abdelaziz M, Teruel M, Chobanov D, Camacho JPM & Cabrero J (2007). Physical mapping of rDNA and satDNA in A and B chromosomes of the grasshopper *Eyprepocnemis plorans* from a Greek population. *Cytogenet Genome Res* 119: 143-146.
- Adegoke JA, Árnason U & Widegren B (1993). Sequence organization and evolution, in all extant whalebone whales, of a DNA satellite with terminal chromosome localization. *Chromosoma* 102: 382-388.
- Agababian M (1997). *Centaurea* subg. *Centaurea* (Compositae): Delimitation and distribution of sections and subsections. *Lagascalía* 19: 889-902.
- Agababian M & Goukasian AV (1994). On the karyology of Armenian representatives of the *Centaurea* subgenus *Centaurea* (Asteraceae). *Thaiszia* 4: 171-173.
- Ainsworth CC, Lu J, Winfield M & Parker J (1999). Sex determination by X:autosome dosage: *Rumex acetosa* (sorrel). In: Ainsworth CC (ed) *Sex Determination in Plants*. Bios Scientific Publishers Limited, pp 124-136.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25: 3389-3402.
- Anderberg AA, Baldwin BG, Bayer RG, Breitwieser J, Jeffrey C, Dillon MO, Eldenäs P, Funk V, García-Jacas N, Hind DJN, Karis PO, Lack HW, Nesom G, Nordenstam B, Oberprieler CH, Panero JL, Puttock C, Robinson H, Stuessy TF, Susanna A, Urtubey E, Vogt R, Ward J & Watson LE (2007). Compositae. In: Kadereit JW & Jeffrey C (eds) *The families and genera of vascular plants*, 8: 61-558.
- Bachmann L & Sperlich D (1993). Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. *Mol Biol Evol* 10: 647-659.
- Barragán MJ, Martínez S, Marchal JA, Fernández R, Bullejos M, Díaz de la Guardia R & Sánchez A (2003). Pericentric satellite DNA sequences in *Pipistrellus pipistrellus* (Vespertilionidae; Chiroptera). *Heredity* 91:232-238.
- Bennett MD & Leitch IJ (2004). *Plant DNA C-values Database (Release 3.0)*.
- Benslimane AA, Dron M, Hartmann C & Rode A (1986). Small tandemly repeated DNA sequences of higher plants likely originate from a tRNA gene ancestor. *Nucl Acids Res* 14: 8111-8119.
- Bentham G (1873). Compositae. In: G. Bentham & JD Hooker (eds), *Genera Plantarum*. Lovell Reeve, London, 2: 162-533.
- Bobrov EG (1997). *Echinops* L. In: Shishkin BK & Bobrov EG (eds) *Flora of the USSR*. Dehra Dun: Bishen Singh, Mahendra Pal Singh; Koenigsten: Koeltz Scientific Books, 27: 1-70.
- Borisova AG (1993). Rod serpucha-*Serratula* L. In: Bobrov EG & Čerepanov SK (eds) *Flora SSSR*. Izdatel'stvo Akademii Nauk SSSR, Moskva Leningrad, 28: 259-301, 606-607.
- Bowles VG, Mayerhofer R, Davis C, Good AG & Hall JC (2010). A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data. *Plant Syst Evol* 287: 85-97.
- Bremer K (1994). *Asteraceae: Cladistics and classification*. Timber Press, Portland, Oregon pp. 725.

- Britten RJ & Kohne DE (1968). Repeated sequences in DNA. *Science* 161: 529–540.
- Bruvo B, Pons J, Ugarković D, Juan C, Petitpierre E & Plohl M (2003). Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (Coleoptera). *Gene* 312: 85-94.
- Bunge A (1863). Über die Gattung *Echinops*. *Bull Acad Imp Sci Saint-Pétersbourg* 6: 390-412.
- Burės P, Wang YF, Horová L & Suda J (2004). Genome size variation in central European species of *Cirsium* (Compositae) and their natural hybrids. *Ann Bot* 94: 353-363.
- Buscaino A, Allshire R & Pidoux A (2010). Building centromeres: home sweet home or a nomadic existence? *Curr Opin Gen Dev* 20: 1-9.
- Cafasso D, Cozzolino S, De Luca P & Chinali G (2003). An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organisations of the repetitive unit in the plant genome. *Gene* 311: 71-79.
- Carlquist S (1976). Tribal interrelationships and phylogeny of the Asteraceae. *Aliso* 8: 465-492.
- Cesari M, Luchetti A, Passamonti M, Scali V & Mantovani B (2003). Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* 312: 289-295.
- Chaudhary SA (2000). Flora of the Kingdom of Saudi Arabia. Ministry of Agriculture and Water, National Herbarium, National Agriculture and Water Research Center, Riyadh, KSA, vol II part 3: 117-202.
- Charlesworth D (2002). Plant sex determination and sex chromosomes. *Heredity* 88: 94-101.
- Connor McCarthy (1996-1998). Chromas v 1.45. School of Health Science, Griffith University, Gold Coast Campus. Southport, Queensland, Australia.
- Cuñado N, Navajas-Pérez R, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Santos JL & Garrido-Ramos MA (2007). The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): Identification of a new species with heteromorphic sex chromosomes. *Chromosome Research* 15: 825-832.
- De Candolle AP (1838). *Prodromus systematis naturalis regni vegetabilis* 6. París.
- De la Herrán R, Ruiz Rejón C, Ruiz Rejón M & Garrido-Ramos MA (2001a). The molecular phylogeny of the Sparidae (Pisces, Perciformes) based on two satellite DNA families. *Heredity* 87: 691-697.
- De la Herrán R, Robles F, Cuñado N, Santos JL, Ruiz Rejón M, Garrido-Ramos MA & Ruiz Rejón C (2001b). A heterochromatic satellite DNA is highly amplified in a single chromosome of *Muscari* (Hyacinthaceae). *Chromosoma* 110: 197-202.
- Degraeve N (1976). Contribution a l'étude cytotaxonomique des *Rumex*. IV Le genre *Acetosa* Mill. *Cellule* 71: 235-250.
- Dempewolf H, Rieseberg LH & Cronk QC (2008). Crop domestication in the Compositae: a family-wide trait assessment. *Genet Resour and Crop Evol* 55: 1141-1157

- Dittrich M (1977). Cynareae systematic review. In VH Heywood, JB Harborne & BL Turner (eds) *The biology and chemistry of the Compositae*. Academic Press, London and New York, 2: 999-1015.
- Dover GA (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.
- Dover GA (2002). Molecular drive. *Trends Genet* 18: 587-589.
- Drouin G & Dover GA (1990). Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. *J Mol Evol* 31: 132-150.
- Eren Ö (2007). The genus *Rhaponticoides* Vaill. (Asteraceae) in Turkey: a new species and first key. *Pl Syst Evol* 267: 13-23.
- Feliciello I, Picariello O & Chinali G (2005). The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. *Gene* 349: 153-164.
- Feliciello I, Picariello O & Chinali G (2006). Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. *Gene* 383: 81-92.
- Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.
- Figuerola R, Stübing G & Peris JB (1990). Notas breves: dos híbridos nuevos para la flora Valenciana. *Anales Jardín Botánico de Madrid* 47: 252.
- Font M, García-Jacas N, Vilatersana R, Roquet C & Susanna A (2009). Evolution and biogeography of *Centaurea* section *Acrocentron* inferred from nuclear and plastid DNA sequence analyses. *Ann Bot* 103: 985-997.
- Font M, Garnatje T, García-Jacas N & Susanna A (2002). Delineation and phylogeny of *Centaurea* sect. *Acrocentron* based on DNA sequences: a restoration of the genus *Crocodylium* and indirect evidence of introgression. *Plant Syst Evol* 234: 15-26.
- Francisco-Ortega J, Santos-Guerra A, Kim S-C & Crawford DJ (2000). Plant genetic diversity in the Canary Islands: a conservation perspective. *Am J Bot* 87: 909-919.
- Fry K & Salser W (1977). Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordi* and characterization of similar sequences in other rodents. *Cell* 12: 1069-1084.
- Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, García-Jacas N, Susanna A & Jansen RK (2005). Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *Biol Skr* 55: 343-374.
- García-Jacas N, Garbany-Casals M, Romashchenko K & Susanna A (2008). On the conflicting generic delineation in the *Onopordum* group (Compositae, Cardueae-Carduinae): a combined nuclear and plastid molecular approach. *Aust Syst Bot* 21: 301-311.

- García-Jacas N, Garnatje T, Susanna A & Vilatersana R (2002). Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Molec Phyl Evol* 22: 51-64.
- García-Jacas N, Susanna A & Ilarslan R (1996) Aneuploidy in the Centaureinae (Compositae): is $n=7$ the end of the series?. *Taxon* 45: 39-42.
- García-Jacas N, Susanna A, Garnatje T & Vilatersana R (2001). Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Ann Bot* 87: 503-515.
- Garnatje T, Vallès J, García S, Hidalgo O, Sanz M, Canela MA & Siljak-Yakovlev S (2004a). Genome size in *Echinops* L. and related genera (Asteraceae, Cardueae): karyological, ecological and phylogenetic implications. *Biol Cell* 96: 117-124.
- Garnatje T, Vilatersana R, Susanna A, Vallès J & Siljak-Yakovlev S (2004b). Contribution to the karyological knowledge of *Echinops* L. (Asteraceae, Cardueae) and related genera. *Bot J Linn Soc* 145: 337-344.
- Garrido-Ramos MA, de la Herran R, Ruiz Rejón C & Ruiz Rejón M (1998). A satellite DNA of the Sparidae family (Pisces, Perciformes) associated with telomeric sequences. *Cytogenet Cell Genet* 83: 3-9.
- Garrido-Ramos MA, de la Herrán R, Ruiz Rejón M & Ruiz Rejón C (1999). A subtelomeric satellite DNA family isolated from the genome of the dioecious plant *Silene latifolia*. *Genome* 42: 442-446.
- Garrido-Ramos MA, Jamilena M, Lozano R, Ruiz Rejón C & Ruiz Rejón M (1995). The *EcoRI* centromeric satellite DNA of the Sparidae family (Pisces, Perciformes) contains a sequence motive common to other vertebrate centromeric satellite DNAs. *Cytogenet Cell Genet* 71: 345-351.
- Gaut BS (1996). Molecular clocks and nucleotide substitution rates in higher plants. *Evol Biol* 30: 93-120.
- Greuter W (2003). The Euro+Med treatment of Cardueae (Compositae) – generic concepts and required new names. *Willdenowia* 33: 49-61.
- Greuter W, Agababian M & Wagenitz G (2005). Vaillant on Compositae-systematic concepts and nomenclatural impact. *Taxon* 54: 149-174.
- Gutknecht J, Sperlich D & Bachmann L (1995). A species specific satellite DNA family of *Drosophila subsilvestris* appearing predominantly in B chromosomes. *Chromosoma* 103: 539-544.
- Häffner E & Hellwig FH (1999). Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: An analysis based on ITS sequence data. *Willdenowia* 29: 27-39.
- Häffner E & Hellwig FH (2000). Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: an analysis based on ITS sequence data. *Willdenowia* 29: 27-39.
- Hellwig FH (1996). Untersuchungen zur phylogenie der Cardueae-Centaureinae (Compositae) unter verwendung molekularer und morphologisch-anatomischer merkmale. Habilitationsschrift Univ. Göttingen, Göttingen.

- Hellwig FH (2004). Centaureinae (Asteraceae) in the Mediterranean – history of ecogeographical radiation. *Plant Syst Evol* 246: 137-162.
- Hanelt P (1963). Monographische Übersicht der Gattung *Carthamus* L. (Compositae). *Feddes Repert Spec Nov Regni Veg* 67: 41-180.
- Hendry GAF (1996). Fructan and the ecology and evolution of the Compositae. In: Caligari PDS & Hind DJN (eds), *Compositae: Biology and Utilization*. Proceedings of the International Compositae Conference, Kew. Royal Botanic Gardens Kew, 2: 121-128.
- Hidalgo O, García-Jacas N, Garnatje T & Susanna A (2006). Phylogeny of *Rhaponticum* (Asteraceae, Cardueae-Centaureinae) and related genera inferred from nuclear and chloroplast DNA sequence data: Taxonomic and biogeographic implications. *Ann Bot* 97: 705-714.
- Higgins DG & Sharp PM (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73: 237-244.
- Hoffmann O (1894). Compositae. In: Engler A. & Prantl K. (eds), *Die natürlichen Pflanzenfamilien*, Leipzig 4: 87-387.
- Jäger EJ (1987). Arealkarten der Asteraceen-tribus als Grundlage der örogeographischen Sippencharakteristik. *Bot Jahrb Syst* 108: 481-497.
- Jäger EJ (1977). Wuchsform und Verbreitung der *Cirsium acaule*-Verwandtschaft in Eurasien. *Flora* 166: 75-92.
- Kadota Y (1995). Tribe 9. Cardueae. In: Iwatsuki K, Yamazaki T, Boufford DE & Ohba H (eds) *Flora of Japan, Angiospermae Dicotyledonae Sympetalae* 117-164.
- Keil DJ (2006). *Carthamus*. In: Barkley TM, Brouillet L, Jeude H, Strother JL, Gandhi K, Kiger RW, Yatskievych K & Sarucchi JL (eds) *Flora of North America*, New York 19: 178-181.
- Kim KJ, Choi KS & Jansen RK (2005). Two chloroplast DNA inversions originated simultaneously during the early evolution on the sunflower family (Asteraceae). *Mol Biol Evol* 22: 1783-1792.
- Langdon T, Seago C, Jones RN, Ougham H, Thomas H, Forster JW & Jenkins G (2000). De novo evolution of satellite DNA on the rye B chromosome. *Genetics* 154: 869-884.
- Lessing CF (1832). *Synopsis generum Compositarum*. Berlin: Dunker and Humblot.
- López-Flores I, de la Herrán R, Garrido-Ramos MA, Boudry P, Ruiz Rejón C & Ruiz Rejón M (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene* 339: 181-188.
- López-Flores I & Garrido-Ramos MA (2012). The repetitive DNA content of eukaryotic genomes. In: Garrido-Ramos MA (ed) *Repetitive DNA*. *Genome Dyn*. Basel, Karger, 7: 1-28.
- Lorite P, Carrillo JA, Tinaut A & Palomeque T (2002). Comparative study of satellite DNA in ants of the *Messor* genus (Hymenoptera, Formicidae). *Gene* 297: 113-122.

- Lorite P, Torres MI, & Palomeque T (2013) Characterization of two unrelated satellite DNA families in the Colorado potato beetle *Leptinotarsa decemlineata* (Coleoptera, Chrysomelidae). Bull Entomolog Res, en prensa.
- Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V & Mantovani B (2003). Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). J Mol Evol 56: 587-596.
- Luchetti A, Marini M & Mantovani B (2006). Non-concerted evolution of the RET76 satellite DNA family in Reticulitermes taxa (Insecta, Isoptera). Genetica 128: 123–132.
- Macas J, Neumann P, Novák P & Jiang J (2010). Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. Bioinformatics 26: 2101-2108.
- Marrero-Rodríguez A (1989). *Carlina texedae* (Asteraceae), nueva especie de Gran Canaria. Botánica Macaronésica 18: 69-84.
- Martins L & Hellwig FH (2005). Systematic position of the genera *Serratula* and *Klasea* within Centaureinae (Cardueae, Asteraceae) inferred from ETS and ITS sequence data and new combinations in *Klasea*. Taxon 54: 632-638.
- Meštrović N, Plohl M, Mravinac B & Ugarković D (1998). Evolution of satellite DNAs from the genus *Palorus*-experimental evidence for the “library” hypothesis. Mol Biol Evol 15: 1062-1068.
- Meštrović N, Castagnone-Sereno P & Plohl M (2006). Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. Mol Biol Evol 23: 2316-2325.
- Meusel H & Jäger EJ (1992). Vergleichende Chorologie der zentraleuropäischen Flora. Flora III, Fischer G, Jena Stuttgart New York.
- Mravinac B, Plohl M, Meštrović N & Ugarković D (2002). Sequence of PRAT satellite DNA “frozen” in some Coleopteran species. J Mol Evol 54: 774-783.
- Mravinac B, Plohl M & Ugarković D (2005). Preservation and high sequence conservation of satellite DNAs suggest functional constraints. J Mol Evol 61: 542-550.
- Mravinac B, Ugarković D, Franjević D, Plohl M (2005). Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. J Mol Evol 60: 513-525.
- Mravinac B & Plohl M (2007). Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. Gene 394: 45–52.
- Mündel HH, Blacksaw RE, Byers JR, Huang HC, Johnson DL, Keon R, Kubik J, McKenzie R, Otto B, Roth B & Standford K (2004). Safflower production on the Canadian prairies. Agriculture and Agri-Food Canada, Lethbridge Research Center, Alberta, pp 37.
- Navajas-Pérez R, de la Herrán R, López-González G, JAMILENA M, Lozano R, Ruiz Rejón C, Ruiz Rejón M & Garrido-Ramos MA (2005a). The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. Mol Biol Evol 22: 1929-1939.

- Navajas-Pérez R, de la Herrán R, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M & Garrido-Ramos MA (2005b). Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). *J Mol Evol* 60: 391-399.
- Navajas-Pérez R, Ruiz Rejón M, Garrido-Ramos MA, Aznarte JL & Rubio-Escudero C (2007a). SatDNA Analyzer: a computing tool for satellite-DNA evolutionary analysis. *Bioinformatics* 23: 767-768.
- Navajas-Pérez R, Rubio-Escudero C, Aznarte JL, Ruiz Rejón M & Garrido-Ramos MA (2007b). SatDNA Analyzer 1.2 as a valuable computing tool for evolutionary analysis of satellite-DNA families: revisiting Y-linked satellite-DNA sequences of *Rumex* (Polygonaceae). In: Hochreiter S, Wagner R (eds) *Lecture notes in computer sciences, bioinformatics research and development*, Springer-Verlag, Berlin, 4414: 131-139.
- Navajas-Pérez R, Schwarzacher T, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M & Garrido-Ramos MA (2006). The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene* 368: 61-71.
- Navajas-Pérez R, Schwarzacher T, Ruiz Rejón M & Garrido-Ramos MA (2009). Molecular cytogenetic characterization of *Rumex papillaris*, a dioecious plant with an XX/XY₁Y₂ sex chromosome system. *Genetica* 135: 87-93.
- Nei M (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nepokroeff M & Sytsma KJ (1996). Systematics and patterns of speciation and colonization in Hawaiian *Psychotria* and relatives based on phylogenetic analysis of ITS sequence data. *Am J Bot* 83: 181-182.
- Nijman IJ & Lenstra JA (2001). Mutation and recombination in cattle satellite DNA: A feedback model for the evolution of satellite DNA repeats. *J Mol Evol* 52: 361-371.
- Olivieri I (1985). Comparative electrophoretic studies of *Carduus pycnocephalus* L., *Carduus tenuiflorus* Curt. (Asteraceae), and their hybrids. *Am J Bot* 72: 715-718.
- Palomeque T & Lorite P (2008). Satellite DNA in insects: a review. *Heredity* 100: 564-573.
- Palomeque T, Carrillo JA, Muñoz-López M & Lorite P (2006). Detection of a *mariner*-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus *Messor* and their possible involvement for satellite-DNA evolution. *Gene* 371: 194-205.
- Panero JL & Funk VA (2002). Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proc Biol Soc Washington* 115: 760-773.
- Panero JL & Funk VA (2008). The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Mol Phyl Evol* 47:757-782.
- Peng HS, Yuan QJ, Li QQ & Huang LQ (2012). Molecular systematic of genus *Atractylodes* (Compositae, Cardueae): Evidence from internal transcribed spacer (ITS) and *trnL-F* sequences. *Int J Mol Sci* 13: 14623-14633.
- Pérez-Gutiérrez MA, Suárez-Santiago VN, López-Flores I, Romero AT & Garrido-Ramos MA (2012). Concerted evolution of satellite DNA in *Sarcocapnos*: a matter of time. *Plant Mol Biol* 78:19-29.

- Petit DP (1988). Le genre *Echinops* L. (Compositae: Cardueae) 1. Position phylétique et interprétation de l'incapitulescence. *Candollea* 43: 467-481.
- Petit DP (1997). Generic interrelationships of the Cardueae (Compositae): A cladistic analysis of morphological data. *Plant Syst Evol* 207: 173-203.
- Petrovič V, Pérez-García C, Pasantes JJ, Šatović E, Prats E & Plohl M (2009). A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* 124: 63-71.
- Pezer Ž & Ugarković D (2012). Satellite DNA-associated siRNAs as mediators of heat shock response in insects. *RNA Biology* 9: 587-595.
- Plohl M, Luchetti A, Meštrović N & Mantovani B (2008). Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. *Gene* 409: 72-82.
- Plohl M & Ugarković D (1994a). Analysis of divergence of *Alphitobius diaperinus* satellite DNA-roles of recombination, replication slippage and gene conversion. *Mol Gen Genet* 242: 297-304.
- Plohl M & Ugarković D (1994b). Characterization of two abundant satellite DNAs from the mealworm *Tenebrio obscurus*. *J Mol Evol* 39: 489-495
- Plohl M, Meštrović N & Mravinac B (2012). Satellite DNA evolution. In: Garrido-Ramos M.A. (ed), Repetitive DNA. *Genome Dyn Basel*, Karger, 7: 126-152.
- Plohl M, Petrovič V, Luchetti A, Ricci A, Šatović E, Passamonti M & Mantovani B (2010). Long-term conservation versus high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* 104: 543-551.
- Pons J & Gillespie RG (2004). Evolution of satellite DNA in a radiation of endemic Hawaiian spiders: does concerted evolution of highly repetitive sequences reflect evolutionary history? *J Mol Evol* 59: 632-641.
- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarković D & Juan C (2004). Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common 'satellite DNA library'. *Heredity* 92: 418-427.
- Pons J, Petitpierre E & Juan C (2002). Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Mol Biol Evol* 19:1329-40.
- Posada D & Crandall KA (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Rechinger KH (ed.) (1979). Flora des Iranischen Hochlandes und der umrahmenden Gebirge. Compositae III-Cynareae. Akademische Druck-u Verlagsanstalt, Graz.
- Renner SS & Ricklefs RE (1995). Dioecy and its correlates in the flowering plants. *Am J Bot* 82: 596-606.
- Rivas Goday S & Rivas Martínez S (1967). Acerca de los *Carthamo-Carduncellus* de la *Ononido-Rosmarinetea* peninsular. *Anales del Instituto Botánico de Cavanilles* 25: 188-197.

- Robles F, de la Herrán R, Ludwig A, Ruiz Rejón C, Ruiz Rejón M & Garrido-Ramos MA (2004). Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* 338: 133-142.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA & Huelsenbeck JP (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61: 539-542.
- Ruiz Rejón M (2004). Sex chromosomes in plants. In: *Encyclopedia of plant and crop science.*, Dekker Agropedia. Marcel Dekker, New York, Vol IV pp 1148–1151.
- Ruiz Rejón C, Jamilena M, Garrido-Ramos MA, Parker JS & Ruiz Rejón M (1994). Cytogenetic and molecular analysis of the multiple sex chromosome system of *Rumex acetosa*. *Heredity* 72: 209-215.
- Sánchez-Jiménez I, Lazkov GA, Hidalgo O & Garnatje T (2010). Molecular systematics of *Echinops* L. (Asteraceae, Cynareae): A phylogeny based on ITS and *trnL-trnF* sequences with emphasis on sectional delimitation. *Taxon* 59 (3):698-708.
- Sanger F, Nicklen S & Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci USA*. 74: 5463-5467.
- Sassanuma T, Sehgal D, Sasakuma T, Raina SN (2008). Phylogenetic analysis of *Carthamus* species based on the nucleotide sequence of the nuclear SACPD gene and chloroplast *trnL-trnF* IGS region. *Genome* 51: 721-727.
- Sharma S & Raina SN (2005). Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenet Genom Res* 109:15-26.
- Shibata F, Hizume M & Kuroki Y (1999). Chromosome painting of Y chromosomes and isolation of a Y chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. *Chromosoma* 108: 266-270.
- Shibata F, Hizume M & Kuroki Y (2000a). Differentiation and the polymorphic nature of the Y chromosomes revealed by repetitive sequences in the dioecious plant *Rumex acetosa*. *Chromosome Res* 8: 229-236.
- Shibata F, Hizume M & Kuroki Y (2000b). Molecular cytogenetic analysis of supernumerary heterochromatic segments in *Rumex acetosa*. *Genome* 43: 391-397.
- Shishkin BK & Bobrov EG (eds.) (1997). *Flora of the USSR* 27. Bishen Singh, Mahendra Pal Singh and Koeltz Scientific Books, Dehra Dun.
- Smith BW (1969). Evolution of sex-determining mechanisms in *Rumex*. *Chromosomes Today* 2: 172-182.
- Stevens PF (2001 onwards). Angiosperm Phylogeny Website. Disponible en: <http://www.mobot.org/MOBOT/research/APweb/>.
- Stimpson KM & Sullivan BA (2010). Epigenomics of centromere assembly and function. *Curr Opin Cell Biol* 22: 772-80.
- Strachan T, Webb D & Dover GA (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J* 4: 1701-1708.

- Stuessy TF & Garver D (1996). The defensive role of pappus in heads of Compositae. In: Caligari PDS & Hind DJN (eds), *Compositae: Biology and Utilization*. Proceedings of the International Compositae Conference, Kew 1994. Royal Botanic Gardens, Kew 2: 81-91.
- Suárez-Santiago VN (2005). Utilización de marcadores moleculares en la resolución de problemas sistemáticos y evolutivos: la sección *Willkommia* Blanca (Género *Centaurea* L., Compositae) y el género *Muscari* Mill. (Hyacinthaceae). PhD Thesis, University of Granada, Spain.
- Suárez-Santiago VN, Blanca G, Ruiz-Rejón M & Garrido-Ramos MA (2007a). Satellite-DNA evolutionary patterns under a complex evolutionary scenario: the case of *Acrolophus* subgroup (*Centaurea* L., Compositae) from the western Mediterranean. *Gene* 404: 80-92.
- Suárez-Santiago VN, Salinas MJ, García-Jacas N, Soltis PS, Douglas ES & Blanca G (2007b). Reticulate evolution in the *Acrolophus* subgroup (*Centaurea* L., Compositae) from the western Mediterranean: origin and diversification of section *Willkommia* Blanca. *Mol Phyl Evol* 43: 156-172.
- Susanna A, Galbany-Casals M, Romaschenko K, Barres L, Martín J & García-Jacas N (2011). Lessons from *Plectocephalus* (Compositae, Cardueae-Centaureinae). ITS disorientation in annuals and Beringian dispersal as revealed by molecular analyses. *Ann Bot* 108: 263-277.
- Susanna A, García-Jacas N, Soltis DE & Soltis PS (1995). Phylogenetic relationships in tribe Cardueae (Asteraceae) based on ITS sequences. *Amer J Bot* 82: 1056-1068.
- Susanna A, Garnatje T & García-Jacas N (1999). Molecular phylogeny of *Cheirolophus* (Asteraceae: Cardueae-Centaureinae) based on ITS sequences of nuclear ribosomal DNA. *Plant Syst Evol* 214: 147-160.
- Susanna A, García-Jacas N, Hidalgo O, Vilatersana R & Garnatje T (2006). The Cardueae (Compositae) revisited: Insights from ITS, *trnL-trnF*, and *matK* nuclear and chloroplast DNA analysis. *Ann Missouri Bot Gard* 93: 150-171.
- Susanna A & García-Jacas N (2007). Tribe Cardueae. In: Kadereit JW & Jeffrey C (eds), *Flowering Plants. Eudicots. Asterales. The families and general of Vascular Plants*, Springer, Berlín 8: 123-146.
- Susanna A & García-Jacas N (2009). Cardueae (Carduoideae) In: Funk VA, Susanna A, Stuessy TF & Bayer RJ (eds) *Systematics, Evolution, and Biogeography of the Compositae*. IAPT, Vienna, pp 293–313.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
- Ugarković D (2005). Functional elements residing within satellite DNAs. *EMBO Reports* 6: 1035-1039.
- Valdés B (2002). *Echinops* L. In: Valdés B, Rejdali M, Achhal el Kadmiri A, Jury LL & Montserrat JM (eds) *Checklist of vascular plants of N Morocco with identification keys*, Madrid, CSIC 2: 723-724.

- Vicari MR, Nogaroto V, Noleto RB, Cestari MM, Cioffi MB, Almeida MC, Moreira-Filho O, Bertollo LAC & Artoni RF (2010). Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. *J Fish Biol* 76: 1094-1116.
- Vilatersana R, Brysting AK & Brochmann C (2007). Molecular evidence for hybrid origins of the invasive polyploids *Carthamus creticus* and *C. turkestanicus* (Cardueae, Asteraceae). *Mol Phyl Evol* 44:610-621.
- Vilatersana R, Garnatje T, Susanna A & García-Jacas N (2005). Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. *Bot J Linn Soc* 147: 375-383.
- Vilatersana R, Susanna A, García-Jacas N & Garnatje T (2000). Generic delimitation and phylogeny of the *Carduncellus-Carthamus* complex (Asteraceae) based on ITS sequences. *Plant Syst Evol* 221: 89-105.
- Wagenitz G (1955). Pollenmorphologie und Systematik in der gattung *Centaurea* L. s.l. *Flora* 142: 213-279.
- Wagenitz G (1976). Systematics and phylogeny of the Compositae (Asteraceae). *Pl Syst Evol* 125: 29-46.
- Wagenitz G (1986). *Centaurea* in South-West Asia: patterns of distribution and diversity. *Proc Roy Soc Edinburgh* 89B: 11-21.
- Wagenitz G & Hellwig FH (1996). Evolution of characters and phylogeny of the Centaureinae. In DJN Hind & HJ Beentje (eds.). *Compositae: Systematics Proceedings of the International Compositae Conference*, Royal Botanic Gardens, Kew, 1: 491-510.
- Wang YJ, Liu JQ & Miede G (2007). Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Ann Bot* 99: 311-322.
- Wang YJ, Susanna A, Von Raab-Straube E, Milne R & Liu JQ (2009). Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai-Tibetan Plateau. *Biol Jour Lin Soc* 97: 893-903.
- Weberling F (1989). *Morphology of flowers and inflorescences*. Cambridge, Univ. Press 1-405.
- Wilby AS & Parker JS (1988). Recurrent patterns of chromosome variation in a species group. *Heredity* 61: 55-62.
- Zajac A & Zajac M (2001). *Distribution Atlas of Vascular Plants in Poland*. Institute of Botany of Jagiellonian University, Krakow.

VIII

Anexos

ANEXO I

Tabla 10. Variabilidad y divergencia medias de las secuencias *HinfI* en las especies de *Centaurea* analizadas. *C. boiss* (*C. boissieri*), *C. clem* (*C. clementei*), *C. cory* (*C. corymbosa*), *C. cyan* (*C. cyanus*), *C. deli* (*C. delicatula*), *C. jaen* (*C. jaennensis*), *C. saxi* (*C. saxifraga*)

ESPECIES	<i>C. alba</i>	<i>C. boiss</i>	<i>C. clem</i>	<i>C. cory</i>	<i>C. cyan</i>	<i>C. deli</i>	<i>C. grana</i>	<i>C. jaen</i>	<i>C. saxi</i>
<i>C. alba</i>	0.171	0.136	0.140	0.137	0.167	0.167	0.231	0.156	0.260
<i>C. boiss</i>		0.065	0.072	0.074	0.098	0.117	0.180	0.083	0.225
<i>C. clem</i>			0.080	0.078	0.108	0.126	0.188	0.095	0.235
<i>C. cory</i>				0.080	0.112	0.130	0.193	0.094	0.242
<i>C. cyan</i>					0.133	0.144	0.200	0.115	0.242
<i>C. deli</i>						0.158	0.199	0.132	0.229
<i>C. grana</i>							0.165	0.187	0.166
<i>C. jaen</i>								0.096	0.234
<i>C. saxi</i>									0.155

Tabla 11. Variabilidad y divergencia medias entre los grupos de secuencias *HinfI* en las especies de *Carduncellus* analizadas. Las tres primeras letras indican el género y las cuatro siguientes la especie. *Cardcaer* (*C. caeruleus*), *Cardcalv* (*C. calvus*), *Cardcuat* (*C. cuatrecasii*), *Carddian* (*C. dianius*), *Cardhisp* (*C. hispanicus*), *Cardmiti* (*C. mitissimus*)

ESPECIES	<i>C. caer</i>	<i>C. calv</i>	<i>C. cuat</i>	<i>C. dian</i>	<i>C. hisp</i>	<i>C. miti</i>
<i>C. caer</i>	0.115	0.167	0.116	0.112	0.099	0.266
<i>C. calv</i>		0.207	0.176	0.166	0.151	0.254
<i>C. cuat</i>			0.107	0.108	0.095	0.226
<i>C. dian</i>				0.101	0.090	0.226
<i>C. hisp</i>					0.075	0.206
<i>C. miti</i>						0.297

Tabla 12. Variabilidad y divergencia medias de las secuencias *HinfI* en las especies del grupo *Carthamus-Phonus* analizadas

ESPECIES	<i>C. lanatus</i>	<i>C. tinctorius</i>	<i>P. arborescens</i>
<i>C. lanatus</i>	0.099	0.108	0.118
<i>C. tinctorius</i>		0.025	0.091
<i>P. arborescens</i>			0.063

Tabla 13. Variabilidad y divergencia medias las secuencias *HinfI* en las especies de *Crupina* analizadas

ESPECIES	<i>C. crupinastrum</i>	<i>C. vulgaris</i>
<i>C. crupinastrum</i>	0.112	0.310
<i>C. vulgaris</i>		0.262

Tabla 14. Variabilidad y divergencia medias de las secuencias *HinfI* en las especies de *Rhaponticoides* analizadas

ESPECIES	<i>R. alpina</i>	<i>R. linaresii</i>
<i>R. alpina</i>	0.072	0.085
<i>R. linaresii</i>		0.100

Tabla 15. Variabilidad y divergencia medias las secuencias *HinfI* en las especies de *Klasea* analizadas

ESPECIES	<i>K. flav. subsp. flav.</i>	<i>K. flav. subsp. leuc.</i>
<i>K. flav. subsp. flav.</i>	0.073	0.104
<i>K. flav. subsp. leuc.</i>		0.108

Tabla 16. Variabilidad y divergencia medias las secuencias *HinfI* en las especies de *Rhaponticum* analizadas

ESPECIES	<i>R. acaule</i>	<i>R. berardioides</i>	<i>R. centauroides</i>	<i>R. coniferum</i>	<i>R. exaltatum</i>
<i>R. acaule</i>	0.223	0.351	0.336	0.318	0.358
<i>R. berardioides</i>		0.112	0.176	0.105	0.184
<i>R. centauroides</i>			0.151	0.167	0.167
<i>R. coniferum</i>				0.029	0.170
<i>R. exaltatum</i>					0.109

Tabla 17. Variabilidad y divergencia medias las secuencias *HinfI* en las especies de *Cheirolophus* analizadas

ESPECIES	<i>C. intybaceus</i>	<i>C. falcisectus</i>	<i>C. sempervirens</i>	<i>C. sventenii</i>	<i>C. teydis</i>
<i>C. intybaceus</i>	0.223	0.253	0.260	0.261	0.459
<i>C. falcisectus</i>		0.217	0.311	0.260	0.524
<i>C. sempervirens</i>			0.242	0.296	0.473
<i>C. sventenii</i>				0.227	0.452
<i>C. teydis</i>					0.119

Tabla 18. Variabilidad y divergencia medias las secuencias *HinfI* en las especies de *Volutaria* analizadas

ESPECIES	<i>V. crupinoides</i>	<i>V. lippi</i>	<i>V. muricata</i>	<i>V. tubuliflora</i>
<i>V. crupinoides</i>	0.275	0.277	0.443	0.413
<i>V. lippi</i>		0.213	0.503	0.517
<i>V. muricata</i>			0.363	0.330
<i>V. tubuliflora</i>				0.127

ANEXO II

Tabla 19. Variabilidad y divergencia medias entre las tres subfamilias *HinfI* en las especies de *Centaurea* analizadas

SUBFAMILIAS	Subfamilia I	Subfamilia II	Subfamilia III
Subfamilia I	0.072	0.213	0.276
Subfamilia II		0.040	0.260
Subfamilia III			0.216

Tabla 20. Variabilidad y divergencia medias entre las subfamilias *HinfI* en las especies de *Carduncellus* analizadas

SUBFAMILIAS	Subfamilia III	Subfamilia IV	Subfamilia VI
Subfamilia III	0.130	0.191	0.495
Subfamilia IV		0.104	0.422
Subfamilia VI			0.039

Tabla 21. Variabilidad y divergencia medias entre los grupos de secuencias *HinfI* en las especies de *Crupina* analizadas

SUBFAMILIAS	Subfamilia I	Subfamilia II	Subfamilia III	Subfamilia VI	Subfamilia VII
Subfamilia I	0.059	0.119	0.254	0.558	0.492
Subfamilia II		0.058	0.208	0.506	0.442
Subfamilia III			0.113	0.489	0.412
Subfamilia VI				-	0.404
Subfamilia VII					0.100

Tabla 22. Variabilidad y divergencia medias entre las subfamilias I, II y III *HinfI* en las especies de *Rhaponticoides* analizadas

SUBFAMILIAS	Subfamilia I	Subfamilia II	Subfamilia III
Subfamilia I	0.047	0.102	0.240
Subfamilia II		0.039	0.205
Subfamilia III			-

Tabla 23. Variabilidad y divergencia medias entre los grupos de secuencias *HinfI* en las especies del género *Rhaponticum* analizadas

SUBFAMILIAS	Subfamilia II	Subfamilia IV	Subfamilia V	Subfamilia VI
Subfamilia II	0.012	0.251	0.355	0.530
Subfamilia IV		0.091	0.323	0.437
Subfamilia V			0.100	0.409
Subfamilia VI				-

Tabla 24. Variabilidad y divergencia medias entre los grupos de secuencias HinfI en las especies de *Cheirolophus* analizadas

SUBFAMILIAS	Subfamilia II	Subfamilia III	Subfamilia VII	Subfamilia VIII
Subfamilia II	0.035	0.183	0.503	0.461
Subfamilia III		0.085	0.485	0.468
Subfamilia VII			0.192	0.380
Subfamilia VIII				0.113

Tabla 25. Variabilidad y divergencia medias entre las subfamilias HinfI en las especies de *Volutaria* analizadas

ESPECIES	Subfam. I	Subfam. II	Subfam. III	Subfam. IV	Subfam. V	Subfam. VI
Subfamilia I	-	0.108	0.257	0.271	0.456	0.667
Subfamilia II		0.048	0.203	0.214	0.423	0.560
Subfamilia III			0.111	-	0.495	0.533
Subfamilia IV				0.093	0.435	0.507
Subfamilia V					-	0.495
Subfamilia VI						0.220

ANEXO III

Tabla 26. Variabilidad y divergencia medias entre las secuencias de la subfamilia II de *HinfI* entre los géneros de las subtribus *Centaureinae*, *Carduinae*, *Carlinae* y *Echinopsinae* analizados

ESPECIES	<i>Centaurea</i>	<i>Crupina</i>	<i>Rhaponticoides</i>	<i>Rhaponticum</i>	<i>Cheirolophus</i>	<i>Volutaria</i>	<i>Jurinea</i>	<i>Carduus</i>	<i>Silybum</i>	<i>Galactites</i>	<i>Echinops</i>	<i>Atractylis</i>	<i>Carlina</i>
<i>Centaurea</i>	0.043	0.085	0.029	0.084	0.055	0.063	0.070	0.069	0.117	0.067	0.258	0.075	0.063
<i>Crupina</i>		0.069	0.069	0.092	0.076	0.083	0.073	0.070	0.122	0.066	0.260	0.070	0.062
<i>Rhaponticoides</i>			0.012	0.077	0.043	0.046	0.056	0.051	0.100	0.047	0.248	0.058	0.046
<i>Rhaponticum</i>				0.006	0.083	0.078	0.092	0.090	0.150	0.091	0.267	0.078	0.071
<i>Cheirolophus</i>					0.041	0.054	0.063	0.059	0.119	0.057	0.236	0.064	0.056
<i>Volutaria</i>						0.049	0.063	0.055	0.115	0.054	0.237	0.069	0.062
<i>Jurinea</i>							0.036	0.033	0.085	0.028	0.226	0.059	0.046
<i>Carduus</i>								0.028	0.086	0.022	0.229	0.053	0.039
<i>Silybum</i>									0.137	0.079	0.281	0.112	0.098
<i>Galactites</i>										0.012	0.218	0.049	0.033
<i>Echinops</i>											0.071	0.229	0.214
<i>Atractylis</i>												0.047	0.035
<i>Carlina</i>													0.016

Tabla 27. Variabilidad y divergencia medias entre las secuencias de la subfamilia III de *HinfI* entre los géneros de las subtribus *Centaureinae* y *Carduinae* analizados

ESPECIES	<i>Centaurea</i>	<i>Carduncellus</i>	<i>Phonus</i>	<i>Carthamus</i>	<i>Crupina</i>	<i>Rhaponticoides</i>	<i>Cheirolophus</i>	<i>Volutaria</i>	<i>Jurinea</i>	<i>Carduus</i>	<i>Silybum</i>	<i>Galactites</i>
<i>Centaurea</i>	0.180	0.205	0.181	0.184	0.218	0.224	0.189	0.210	0.267	0.245	0.227	0.253
<i>Carduncellus</i>		0.143	0.133	0.132	0.166	0.169	0.128	0.153	0.231	0.213	0.183	0.227
<i>Phonus</i>			0.066	0.102	0.119	0.088	0.114	0.116	0.239	0.221	0.179	0.227
<i>Carthamus</i>				0.103	0.147	0.148	0.130	0.126	0.246	0.228	0.188	0.241
<i>Crupina</i>					0.157	0.141	0.148	0.161	0.206	0.199	0.186	0.197
<i>Rhaponticoides</i>						-	0.141	0.157	0.271	0.265	0.209	0.262
<i>Cheirolophus</i>							0.158	0.142	0.214	0.193	0.176	0.209
<i>Volutaria</i>								0.145	0.229	0.204	0.198	0.218
<i>Jurinea</i>									0.069	0.062	0.183	0.054
<i>Carduus</i>										-	0.180	0.068
<i>Silybum</i>											0.175	0.177
<i>Galactites</i>												0.010

Tabla 28. Variabilidad y divergencia medias de las secuencias de la subfamilia IV de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados

ESPECIES	<i>Carduncellus</i>	<i>Rhaponticum</i>	<i>Volutaria</i>	<i>Jurinea</i>	<i>Carduus</i>	<i>Atractylis</i>
<i>Carduncellus</i>	0.127	0.122	0.114	0.144	0.148	0.125
<i>Rhaponticum</i>		0.115	0.105	0.133	0.130	0.107
<i>Volutaria</i>			0.085	0.129	0.143	0.117
<i>Jurinea</i>				0.018	0.121	0.063
<i>Carduus</i>					-	0.130
<i>Atractylis</i>						0.056

Tabla 29. Variabilidad y divergencia medias de las secuencias de la subfamilia V de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados

ESPECIES	<i>Klasea</i>	<i>Rhaponticum</i>	<i>Volutaria</i>	<i>Carduus</i>	<i>Silybum</i>	<i>Galactites</i>	<i>Carlina</i>	<i>Atractylis</i>
<i>Klasea</i>	0.112	0.252	0.303	0.323	0.340	0.299	0.289	0.302
<i>Rhaponticum</i>		0.162	0.201	0.316	0.227	0.179	0.165	0.189
<i>Volutaria</i>			-	0.182	0.033	0.207	0.225	0.126
<i>Carduus</i>				0.172	0.145	0.160	0.155	0.168
<i>Silybum</i>					0.004	0.193	0.230	0.131
<i>Galactites</i>						0.006	0.150	0.176
<i>Carlina</i>							0.095	0.090
<i>Atractylis</i>								-

Tabla 30. Variabilidad y divergencia medias de las secuencias de la subfamilia VII de HinfI entre los géneros de las subtribus Centaureinae, Carduinae y Carlininae analizados

ESPECIES	<i>Crupina</i>	<i>Cheirolophus</i>	<i>Carduus</i>	<i>Silybum</i>	<i>Atractylis</i>
<i>Crupina</i>	0.152	0.204	0.222	0.246	0.118
<i>Cheirolophus</i>		0.179	0.168	0.228	0.184
<i>Carduus</i>			-	0.142	0.213
<i>Silybum</i>				0.110	0.221
<i>Atractylis</i>					0.006

Tabla 31. Variabilidad y divergencia medias de las secuencias de la subfamilia VIII de HinfI entre los géneros de las subtribus Centaureinae y Carduinae analizados

ESPECIES	<i>Cheirolophus</i>	<i>Carduus</i>
<i>Cheirolophus</i>	0.167	0.192
<i>Carduus</i>		-

Tabla 32. Variabilidad y divergencia medias de las secuencias de la subfamilia IX de HinfI entre los géneros de la subtribu Carduinae analizados

ESPECIES	<i>Cirsium</i>	<i>Galactites</i>	<i>Onopordum</i>
<i>Cirsium</i>	0.051	0.527	0.633
<i>Galactites</i>		0.027	0.778
<i>Onopordum</i>			0.088

ANEXO IV

Árbol filogenético obtenido mediante inferencia bayesiana basado en las secuencias HinfI de la subtribu Centaureinae

ANEXO V

Árbol filogenético obtenido mediante inferencia bayesiana basado en las secuencias *Hinf*I de las subtribus (*Carlininae*, *Echinopsinae*, *Carduinae* y *Centaureinae*) de la tribu *Cardueae*.

Navajas-Pérez R, Quesada del Bosque ME, Garrido-Ramos MA. Effect of location, organization, and repeat-copy number in satellite-DNA evolution. *Mol Genet Genomics*, 2009; 282(4): 395-406. doi: 10.1007/s00438-009-0472-4.

Effect of location, organization, and repeat-copy number in satellite-DNA evolution

R. Navajas-Pérez · M. E. Quesada del Bosque ·
M. A. Garrido-Ramos

Received: 3 April 2009 / Accepted: 11 July 2009 / Published online: 4 August 2009
© Springer-Verlag 2009

Abstract Here, we analyze the evolutionary dynamics of a satellite-DNA family in an attempt to understand the effect of factors such as location, organization, and repeat-copy number in the molecular drive process leading to the concerted-evolution pattern found in this type of repetitive sequences. The presence of RAE180 satellite-DNA in the dioecious species of the plant genus *Rumex* is a noteworthy feature at this respect, as RAE180 satellite repeats have accumulated differentially, showing a distinct distribution pattern in different species. The evolution of dioecious *Rumex* gave rise to two phylogenetic clades: one clade composed of species with an ancestral XX/XY sex chromosome system and a second, derived clade of species with a multiple sex–chromosome system XX/XY₁Y₂. While in the XX/XY dioecious species, the RAE180 satellite-DNA is located only in a small autosomal locus, the RAE180 repeats are present also in a small autosomal locus and additionally have been massively amplified in the Y chromosomes of XX/XY₁Y₂ species. Here, we have found that the RAE180 repeats of the autosomal locus of XX/XY species are characterized by intra-specific sequence homogeneity

and inter-specific divergence and that the comparison of individual nucleotide positions between related species shows a general pattern of concerted evolution. On the contrary, both in the autosomal and the Y-linked loci of XX/XY₁Y₂ species, ancestral variability has remained with reduced rates of sequence homogenization and of evolution. Thus, this study demonstrates that molecular mechanisms of non-reciprocal exchange are key factors in the molecular drive process; the satellite DNAs in the non-recombining Y chromosomes show low rates of concerted evolution and intra-specific variability increase with no inter-specific divergence. By contrast, freely recombining loci undergo concerted evolution with genetic differentiation between species as occurred in the autosomal locus of XX/XY species. However, evolutionary periods of rapid sequence change might alternate with evolutionary periods of stasis with variability remaining by the reduced action of molecular mechanisms of non-reciprocal exchange as occurred in XX/XY₁Y₂ species, which could depend on repeat-copy number and the processes involved in their amplification.

Keywords Satellite-DNA · *Rumex* · Sex chromosomes · Concerted evolution · Homogenization

Communicated by S. Hohmann.

R. Navajas-Pérez and M. E. Quesada del Bosque contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-009-0472-4) contains supplementary material, which is available to authorized users.

R. Navajas-Pérez · M. E. Quesada del Bosque ·
M. A. Garrido-Ramos (✉)
Departamento de Genética, Facultad de Ciencias,
Universidad de Granada, 18071 Granada, Spain
e-mail: mgarrido@ugr.es

Introduction

Satellite-DNA evolution is a key subject in molecular evolutionary genetics. Understanding modes, paths and consequences of satellite-DNA evolution is an open question, which requires a profound analysis using appropriate study systems. In this context, the presence of RAE180 satellite-DNA in the dioecious species of the plant genus *Rumex* makes it a noteworthy trait as RAE180 satellite repeats

have accumulated differentially, showing a distinct distribution pattern in different species. The comparative analysis of the effect of these different patterns of location, organization, and repeat-copy number within different *Rumex* species should enhance the understanding of satellite-DNA evolution, especially in this case, as well as in relation to sex-chromosome evolution.

Satellite-DNA families are comprised tandem non-coding short repeated sequences distributed through the eukaryotic genomes at heterochromatin basically in centromeric and subtelomeric regions as well as chromosome-specific amplified regions (Ugarkovic and Plohl 2002). A satellite-DNA family could arise in a phylogenetically short period by explosive amplification (Bachmann and Sperlich 1993). Then, after its formation, the repeats of the family could follow a gradual mode of sequence evolution during a long evolutionary period (Bachmann and Sperlich 1993). The processes by which satellite-DNA families arise are not well known. A set of molecular-exchange mechanisms has been proposed to be involved in its origin by amplification of a tandem array of multi-copy sequences. Those mechanisms include unequal crossing-over (Smith 1976), transposition (Miller et al. 2000), or extrachromosomal rolling-circle replication and reintegration of tandem arrays into the genome (Felicciello et al. 2005, 2006). A recently originated tandem array is initially homogeneous in sequence because of the multi-copy amplification of the same repeat type. In the course of time, random mutations would accumulate and the repeats would diverge. However, the non-allelic repeats of a satellite-DNA family do not evolve independently, but concertedly.

Mechanisms such as unequal crossing-over or gene conversion (Ohta and Dover 1984) can gradually spread a sequence variant throughout a family within a sexual population in a population-genetic process called molecular drive (Dover 2002). Molecular drive leads to high homogeneity levels in a repetitive DNA family for species-diagnostic mutations, which subsequently gives rise to inter-specific genetic divergence, an evolutionary pattern of repetitive sequences known as concerted evolution (Dover 2002). Empirical observations indicate that the production rate of new sequence variants is a slower process than their rate of spread, while the general paucity of transition stages indicates also that the replacement is relatively fast (Ugarkovic and Plohl 2002). However, the rates of homogenization and fixation of sequence variants, i.e., the rates of sequence change, vary for each satellite-DNA family or even for the same satellite-DNA family within different lineages. Levels of sequence identity between repeats would depend on many factors in each species, such as the rates and biases of transfer between homologous and non-homologous chromosomes, number and distribution of repeat units, physical constraints within the genome, generation time,

and effective population size as well as selective constraints (Ohta and Dover 1984). We have analyzed the effects of some of these factors (location, rate of genetic transfer, and repeat number, organization, and distribution) on satellite-DNA evolution in dioecious species of the genus *Rumex*.

Molecular phylogeny classifies *Rumex* in three clades (Navajas-Pérez et al. 2005a). The basal clade is composed by hermaphroditic species. A second clade, derived, is composed of polygamous and gynodioecious species. Finally, a third clade is composed of all Eurasian and American dioecious species. There is an evolutionary trend within the genus *Rumex* from hermaphroditism to dioecy, with polygamy and gynodioecy as intermediate stages (Navajas-Pérez et al. 2005a). We also confirmed that, in the evolution of the genus *Rumex*, there is an evolutionary trend for a chromosome-number reduction from $x = 10$ (the basic ancestral chromosome number for the genus, which persisted in all hermaphroditic as well as in some polygamous *Rumex* species) to $x = 7$ (the basic number found in most dioecious species), with intermediate stages having $x = 9$ (gynodioecious and some polygamous species) and $x = 8$ (the Eurasian dioecious *Rumex suffruticosus*) (Degraeve 1975, 1976, 1980; Navajas-Pérez et al. 2005a). The dioecious clade is subdivided into two subclades (Navajas-Pérez et al. 2005a). One basal subclade is composed of species such as *R. suffruticosus* ($x = 8$) and *Rumex acetosella* ($x = 7$) having a XX/XY sex chromosome system, in which females are XX and males are XY. A second subclade is composed of species such as *Rumex acetosa* and *Rumex papillaris* and represents a derived state with a complex XX/XY₁Y₂ sex-chromosome system, in which females have a karyotype $2n = 12 + XX$ and the males are $2n = 12 + XY_1Y_2$. The origin of the double Y chromosome system might come from a chromosomal translocation between one member of a pair of autosomes and the X chromosome in an ancestral XX/XY species (Smith 1969) or from a process of misdivision of an ancestral Y chromosome (Navajas-Pérez et al. 2006). Taking a mean rate of change in plant nuclear DNA of 0.6% per site per million years (Gaut 1998), the ITS mean-distance corrected estimates between clades supported the origin of dioecy in *Rumex* between 15 and 16 mya, while the divergence time between the two dioecious subclades should be 12–13 mya (Navajas-Pérez et al. 2005a).

RAE180 satellite-DNA family is present within the genomes of all Eurasian and American dioecious *Rumex* species while being absent of the rest of *Rumex* species. This data support the idea, on the one hand, that the dioecious group is monophyletic and, on the other hand, that RAE180 repeats appeared in the common ancestor of these species, approximately 15–16 mya. However, whereas in the XX/XY dioecious species, RAE180 satellite-DNA is

located only in a small locus of a pair of autosomes (Cuñaado et al. 2007), these repeats are present also in a small autosomal locus and additionally have been massively amplified in the Y chromosomes of XX/X₁Y₂ species (Cuñaado et al. 2007; Navajas-Pérez et al. 2009). Comparing satellite-DNA sequences of two species with a XX/X₁Y₂ system such as *R. acetosa* and *R. papillaris*, we found that the repeats of the RAE180 satellite-DNA families located in the non-recombining Y chromosomes have lower rates of sequence evolution than the repeats of the autosomic RAE730 satellite-DNA family (Navajas-Pérez et al. 2005b). These data imply that patterns of satellite-DNA evolution depend on the chromosomal location of the repeats, as the absence of recombination between the sex chromosomes seriously influenced the rate of satellite-DNA sequence change. These conclusions were derived from a comparison between different satellite-DNA families. The present research opens the possibility of testing the behavior of repeats from the same satellite-DNA family when they are located in different loci. For this task, we isolated and sequenced RAE180 repeats of the autosomic loci of *R. acetosa* and *R. papillaris* (XX/X₁Y₂ species), and of the autosomic loci of *R. suffruticosus* and *R. acetosella* (XX/X₁Y₂ species) and we have studied their evolutionary patterns in comparison to those previously characterized for RAE180 repeats of the Y chromosomes of *R. acetosa* and *R. papillaris* (Navajas-Pérez et al. 2005b).

Materials and methods

Materials, DNA extraction, and RAE180 repeats isolation

Female and male plants of *R. suffruticosus* and *R. acetosella* (Puerto de Navacerrada, Segovia, Spain), *R. acetosa* (Sierra Nevada, Spain), and *R. papillaris* (Sierra de Baza, Spain) were used in this study. DNA isolation was performed using DNAeasy mini kit (Qiagen) according to the manufacturer's instructions.

For amplification of RAE180 sequences from female and male plants of *R. suffruticosus* and *R. acetosella* and from female plants of *R. acetosa* and *R. papillaris*, we used the pair of primers Rae180a 5'-TCATCGAACTTCATTCAT-3' and Rae180b 5'-TATAGTAATATCTCGATC-3' (Navajas-Pérez et al. 2005b). PCR amplifications were performed in 50- μ l reactions containing 10 ng of purified DNA, 2 mM of dNTPs, 2 mM of each primer, and 1.25 units of Taq-polymerase in 10 mM Tris-HCl at pH 8.3, 5 mM KCl, 2 mM MgCl reaction buffer. Thermal cycles consisted of 1 min at 94°, 1 min at 55°, and 1 min at 72°. The PCR products were electrophoresed in agarose gels, after which the bands were cut out of the gel, purified using GenElute PCRclean-Up Kit (SIGMA), ligated to the

pCR4-TOPO vector, and cloned using the TOPO TA Cloning Kit (Invitrogen), following the manufacturer's instructions. Recombinant clones were sequenced by the dideoxy-sequencing method using the automatic ABI-Prism 377 sequencer (Applied Biosystems). The EMBL accession numbers for all the sequences analyzed in this article are: AM398600 to AM398639, AJ580328 to AJ580351, AJ580453 to AJ580469, and AJ634478 to AJ634526.

Sequence analysis

For sequence analysis, multiple alignments were carried out using Clustal X (Thompson et al. 1997) followed by manual adjustments. For each species, we determined repeat length, nucleotide composition, and intra-specific variability (measured as nucleotide diversity, π , the average number of nucleotide differences per site between two sequences, Nei 1987) using the software satDNA Analyzer (Navajas-Pérez et al. 2007a). Intra-specific variability was estimated by comparison of all RAE180 repeats. The satDNA Analyzer program has been developed for the analysis of the patterns of variation at each nucleotide position among all units of a repetitive family within a species and for comparing a pair of species. Within each species, the program classifies each site accordingly as monomorphic (the same nucleotide in all the repeats of the species) or polymorphic (the site is variable between repeats).

In comparing the repeats of two species, the satDNA Analyzer program (Navajas-Pérez et al. 2007a) discriminates between shared and non-shared polymorphic sites. The program identifies polymorphic sites shared between two species when the same polymorphism is found in both species (see Supplementary material, Sm1, for a graphic representation). When this occurs, we assume that these are ancestral sites that appeared prior to the split between the two species (Navajas-Pérez et al. 2005b). By contrast, non-shared polymorphic sites are autapomorphies, representing different transitional stages in the process of intra-specific sequence homogenization and inter-specific divergence. Under the assumption that concerted evolution is a time-dependent process, the expected stages of transition during the spread of a variant repeat unit toward its fixation can be defined according to the model of Strachan et al. (1985). This is a method of partitioning the variation by analyzing the patterns of variation at each nucleotide site considered independently among all the repeats of a repetitive family when comparing a pair of species (Strachan et al. 1985; Navajas-Pérez et al. 2007a). This method classifies the sites in terms of six stages (Classes I–VI) in the spread of variant repeats through the family and the species (see Supplementary material, Sm1, for a graphic representation). Briefly, the Class I site represents complete homogeneity across all repeat units sampled from a pair of species, whereas

Classes II, III, and IV represent intermediate stages in which one of the species shows a polymorphism. The frequency of the new nucleotide variant at the site considered is low in Class II and intermediate in Class III, while Class IV represents sites in which a mutation has replaced the progenitor base in most members of the repetitive family in the other species. Class V represents diagnostic sites in which a new variant is fully homogenized and fixed in all the members of one of the species while the other species retains the progenitor nucleotide. A Class VI site represents an additional step over the stage of Class V (new variants appear in some of the members of the repetitive family at a site fully divergent between the two species).

Genetic distances between species were calculated according to the Jukes–Cantor method, comparing all the repeats from each species. Rates of RAE180 evolution were determined then according to the divergence times estimated in Navajas-Pérez et al. (2005a). Shared polymorphic sites were excluded from the alignments, and genetic distances and the rates of RAE180 sequence change were estimated from these alignments following Navajas-Pérez et al. (2005b). This procedure avoids the quantification of ancestral variability, prior to the split of two species from their common ancestor, in the determination of genetic distances between two species for a satellite-DNA family.

Phylogenetic analyses were conducted using MEGA version 4.1 (Tamura et al. 2007) and trees constructed by the neighbor-joining method (Saito and Nei 1987). Thousand bootstrap replicates (Felsenstein 1985) were performed to assess internal support for nodes.

Satellite-DNA quantification

To determine the relative percentage of RAE180 repeats within different genomes (males and females of the four species analyzed), a quantitative PCR (qPCR) approach was developed. PCR conditions were optimized for each assay with regard to annealing temperature and concentration of $MgCl_2$. Conventional PCR reactions for agarose gel-based detection were carried out in 25 μ l using 90 ng/ μ l genomic DNA template, 0.2 mM dNTPs, 200 mM each oligonucleotide primer (Rae180a and Rae180b), optimized $MgCl_2$, and 1 unit Taq DNA polymerase. Each sample was subjected to an initial denaturation of 10 min at 95°C, followed by 40 amplification cycles of denaturation at 95°C for 15 s, optimized annealing for 30 s, followed by extension at 72°C for 30 s. Quantitative PCR reactions were carried out using SensiMix/dT Kit, which includes the internal reference (ROX) (Quantace). Each qPCR reaction comprised 12.5 μ l SensiMix/dT, forward and reverse primers (Rae180a and Rae180b) at optimized concentrations of 10 pmol (final concentration), 1 μ l SYBR Green I solution 50 \times , 90 ng/ μ l genomic DNA template, and sterile water up

to a final volume of 25 μ l. Each sample was subjected to an initial denaturation of 10 min at 95°C as a pre-cycle heat activation of DNA polymerase, followed by 40 cycles of denaturation at 95°C for 15 s, annealing temperature 56.7°C for 30 s, followed by extension 30 s at 72°C. Quantitative PCR experiments were performed using a Real-Time PCR Cromo 4 thermocycler (BioRad).

Real-time kinetic quantitative PCR determines, for each sample well, the C_t , the fractional cycle number at which the well's accumulating fluorescence crosses a set threshold that is several standard deviations above baseline fluorescence. A plot of C_t versus $\log(\text{amount of input target DNA})$ is linear, allowing simple relative quantification of unknowns by comparisons to a standard curve derived from amplification, in the same plate, of serial dilutions of a reference DNA sample. The standard curve was generated from a 7-fold \log_{10} dilution series of a solution of 50 ng/ μ l of recombinant plasmid DNA containing one RAE180 repeat using Rae180a and Rae180b as PCR primers. All samples, including the external standards and non-template control, were run in triplicate. The amplification efficiency for each primer was determined from the linear slope of standard curve; only primers with a standard curve slope of around -3.3 were used for quantification. When analyzed by qPCR, the dilution series produced a set of standard curves, which were used to calculate the slope value using the Opticon3 software (BioRad).

Results

Isolation and analysis of RAE180 sequences from *R. suffruticosus* and *R. acetosella*

The main PCR product obtained for *R. suffruticosus* with Rae180a and Rae180b primers had an approximate length of 330 bp, although some PCR products of 520 bp were also obtained. Each amplified product was comprised one (330-bp product) or two (520-bp product) RAE180 repeat units plus a fragment of each 5' and 3' surrounding repeats. In the case of *R. acetosella*, the amplified PCR product had about 700 bp, composed of three repeat units plus a fragment of each 5' and 3' surrounding repeats. PCR products were cloned and sequenced. The number of clones and repeats sequenced and the results on base-pair composition, repeat length, intra-specific variability (measured as nucleotide diversity, π , the average number of nucleotide differences per site between two sequences, Nei 1987) and quantification of RAE180 repeats in the genomes of *R. acetosella* and *R. suffruticosus* are summarized in Table 1. Specifically, intra-specific variability was low for both species (5.1 and 5.4%, respectively), being similar for repeats isolated from different sexes. RAE180 sequences

Table 1 Number of clones and repeats sequenced and results on base pair composition, variability, repeat length, and quantification of RAE180 repeats in the genomes of four dioecious *Rumex* species

Species	Number of clones			Number of repeats			%AT	Variability			Repeat length	Percentage	
	Female	Male	Total	Female	Male	Total		Female	Male	Total		Female	Male
<i>R. acetosella</i>	5	8	13	15	24	39	64.4	0.058	0.054	0.054	167–183	0.0007	0.0007
<i>R. suffruticosus</i>	9	7	16	16	12	28	65.4	0.053	0.040	0.051	181–184	0.04	0.04
<i>R. acetosa</i>	18	32 ^a	51	18	42 ^a	60	68.4	0.199	0.199 ^a	0.209	181–186	0.01	5.60
<i>R. papillaris</i>	15	23 ^a	38	15	31 ^a	46	67.9	0.153	0.211 ^a	0.219	182–184	0.01	5.60

^a Taken from Navajas-Pérez et al. (2007b)

were poorly represented in the genome of *R. acetosella* ($7 \times 10^{-4}\%$ of the genome), while this satellite-DNA family represents the 0.04% of the genome of *R. suffruticosus*.

RAE180 sequences of these two species were phylogenetically analyzed (Figs. 1, 2; see also supplementary material, Sm2, for alignments). RAE180 repeats of *R. suffruticosus* were not specifically clustered by sequence type or source gender (Fig. 1). However, in the case of *R. acetosella*, the grouping of sequences resembled the tandem arrangement of each repeat within each sequenced clone in the order A–B–C (Fig. 2). This indicates the organization of RAE180 monomers in a higher-ordered repeated (HOR) supra-organization constituted by three or more repeats of the RAE180 satellite-DNA. It is worth mentioning here that the third repetition of the HOR underwent a 3' terminal deletion of 16 bp in all the clones analyzed, giving rise to 167-bp monomers in all cases (Fig. 2b). Notably also, in all clones analyzed, the next repeat in the tandem, which was sequenced partially because of the PCR procedure, also had a 25-bp fixed deletion at the 5' end (Fig. 2b). In this species, sequence differentiation between the three types of sequences was between 6 and 7%, while the variation between sequences of the same type ranged between 2 and 3%.

Isolation and analysis of autosomic RAE180 sequences from *R. acetosa* and *R. papillaris*

Females of *R. acetosa* and *R. papillaris* have only the minor autosomic locus in contrast to males, which have two Y chromosomes enriched in RAE180 repeats besides the autosomic locus (Cuñado et al. 2007). To analyze the evolutionary behavior of RAE180 in the autosomic location in such species, we then isolated by PCR amplification the RAE180 repeats from the genomes of females of both species. In both cases, the only PCR product obtained with RAE180a and RAE180b primers had an approximate length of 330 bp, consisting of one RAE180 repeat unit plus a fragment of each 5' and 3' surrounding repeats. The products were cloned and sequenced. Results, including quantification of RAE180 repeats in the genome of both females

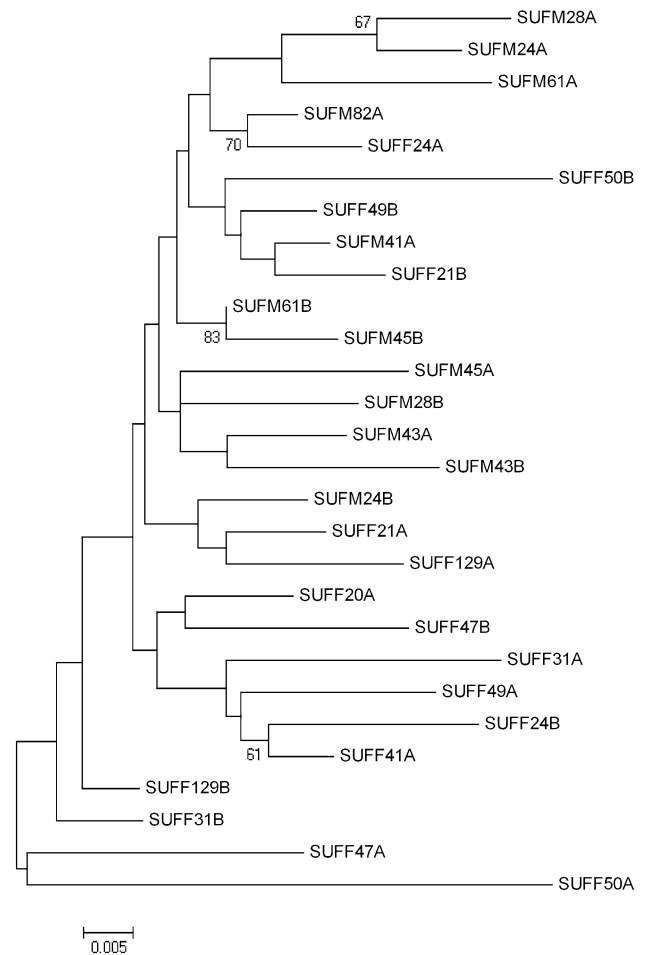


Fig. 1 Neighbor-joining tree of RAE180 repeats isolated from *R. suffruticosus*. RAE180 repeats were not specifically clustered by sequence type or source gender. Sequence code: first three letters (SUF) for species name (*R. suffruticosus*), next letter indicates gender (F, female; M, male) and the number indicates the number of the clone from which the repeat was sequenced. In some cases, within a clone, there were two adjacent RAE180 repeats (see text). In these cases, a final letter in the sequence code indicates the order of the repeat in the clone (A, first and B, second). 1,000 bootstrap replicates were performed to assess internal support for nodes. Only bootstrap values above 60% are indicated

and males, are summarized in Table 1 (data for males were taken from Navajas-Pérez et al. 2007b). In this case, intra-specific variability was 20.9% for *R. acetosa* and 21.9% for

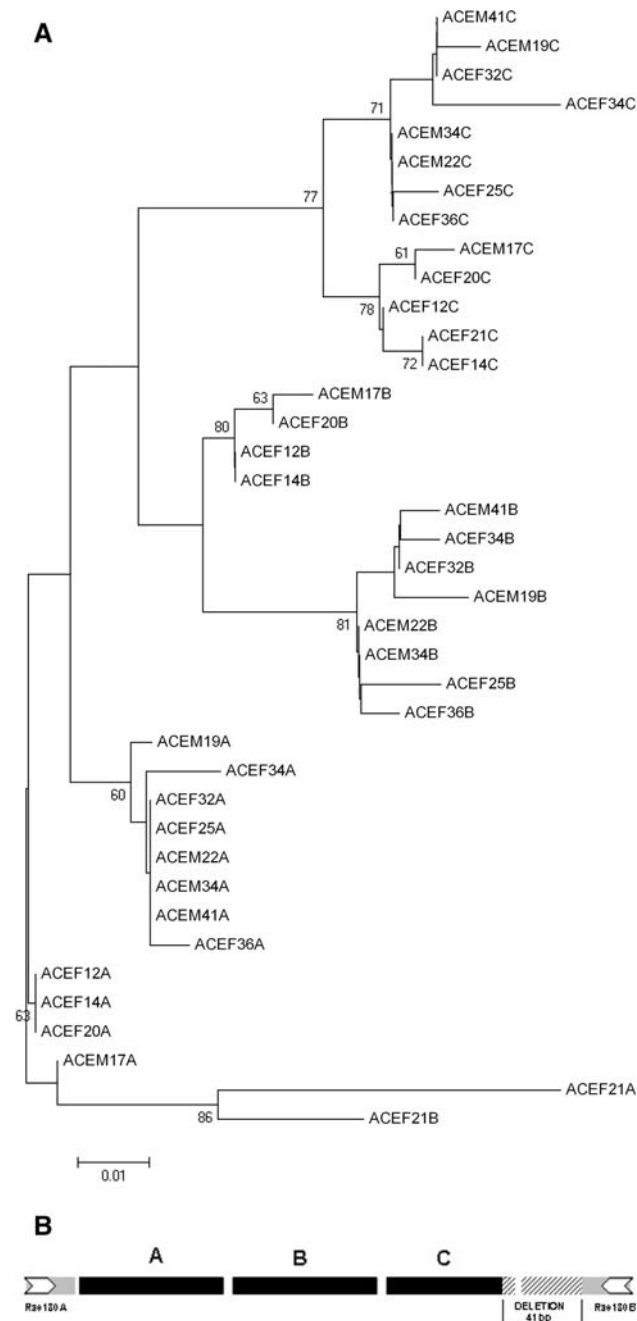


Fig. 2 a Neighbor-joining tree of RAE180 repeats isolated from *R. acetosella*. The grouping of sequences resembles the tandem arrangement of each repeat within each sequenced clone in the order A–B–C, which suggest the organization of RAE180 repeats in a higher-ordered repeated (HOR) supra-organization constituted by three or more repeats of the RAE180 satellite-DNA. Sequence code: first three letters (ACE) for species name (*R. acetosella*), next letter indicates gender (F, female; M, male), the number indicates the number of the clone, and the last letter indicates the order of the repeat in the clone (A, B, or C), as explained in (b). Only bootstrap values above 60% are indicated. **b** Schematic representation of the repeat disposition in clones of the *R. acetosella*-RAE180 library. The clone insets comprised three repeat units (A, B, and C, *black boxes*) plus a fragment of each 5' and 3' surrounding repeats (*grey boxes*; *arrows* indicate PCR primer localizations). The repeat C underwent a 3' terminal deletion of 16 bp in all the clones analyzed, giving rise to 167-bp monomers in all cases. Notably also, in all clones analyzed, the next repeat in the tandem, which was sequenced partially because of the PCR procedure, also had a fixed deletion of 25 bp in the 5' end

divergence time of 4 million years between the two species (Navajas-Pérez et al. 2005a), we can estimate the evolutionary rate for this satellite-DNA family at 28×10^{-9} substitutions per site per year, almost three-fold higher than that estimated for RAE180 sequences when comparing males of *R. acetosa* and *R. papillaris* (Navajas-Pérez et al. 2007b). Also, we have studied the sequence divergence by analyzing the patterns of variation at each nucleotide position considered independently among all RAE180 repeats when comparing the two species. Thus, we have classified each site within each species as monomorphic or polymorphic, discriminated between shared and non-shared polymorphic sites between the two species, and classified each non-shared polymorphic site according to the model proposed by Strachan et al. (1985) (see “Materials and methods” section). Table 2 summarizes all these data. A high percentage of Strachan sites of the categories IV, V, and VI were found (14% of the sites per repeat), while 24% of the sites per repeat were Strachan transition stages (II + III). There were eight shared polymorphic sites between the two species. Figure 3 shows the neighbor-joining tree of RAE180 sequences from these two species, which are clustered together according to taxonomic affinity into two main clades supported by 100% of bootstrap replicates.

Comparing the autosomic RAE180 repeats of *R. acetosa* and *R. papillaris* females, we found the genetic distance between the two species to be 0.197. However, almost half of the polymorphic sites within each species were shared between the females of the two species (62 shared polymorphic sites). A similar situation was found for RAE180 sequences of *R. acetosa* and *R. papillaris* males (Navajas-Pérez et al. 2005b). Excluding the shared polymorphic sites, the mean genetic distance between females of both species was 0.045 and thus the evolutionary rate in this case should be 9×10^{-9} substitutions per site per year. There are no Strachan sites of the classes IV–VI, while 34% of the

R. papillaris, being similar in both females and males. While 0.01% of the genome of females is composed of RAE180 repeats, these sequences represent the 5.6% of the genome of males in both species.

Concerted-evolution analysis of autosomic RAE180 repeats

The mean genetic distance for RAE180 sequences between *R. suffruticosus* and *R. acetosella* proved to be 0.226 (0.247 after excluding shared polymorphic sites). Considering a

Table 2 Interspecific comparative analysis of RAE180 repeats

Species comparison	SP (%)	Strachan sites II–III (%)	Strachan sites IV–VI (%)	Genetic distance	Evolutionary rate
<i>R. acetosa</i> versus <i>R. papillaris</i> males ^a	108 (59%)	33 (17%)	1 (0.5%)	0.033	8.25×10^{-9}
<i>R. acetosa</i> versus <i>R. papillaris</i> females	62 (34%)	62 (34%)	0 (0%)	0.045	9×10^{-9}
<i>R. acetosella</i> versus <i>R. suffruticosus</i>	8 (4%)	45 (24%)	27 (14%)	0.226	28×10^{-9}

The table shows the number and percentage of shared polymorphic sites (SP) between two genomes, the genetic distances and the evolutionary rates as well as the results on the classification of each non-shared polymorphic site according to the model proposed by Strachan et al. (1985) in six different stages of transition during the spread of a variant repeat unit toward its fixation. Genetics distances estimates were after excluding shared polymorphic sites

^a Taken from Navajas-Pérez et al. (2007b)

sites represented Strachan stages II–III (Table 2). Similar to the repeats from males (Navajas-Pérez et al. 2005b), RAE180 repeats from females of both species appear intermixed in the corresponding neighbor-joining tree (Fig. 4; see also Sm3 complementary material for sequence alignments).

Discussion

Satellite-DNA amplification has accompanied the karyotype rearrangements found during species evolution in *Rumex*. Amplification of RAYSI and RAE730 satellite-DNA families occurred in the lineage leading to XX/X₁Y₂ species (Navajas-Pérez et al. 2005b, 2006). Both satellite-DNA families share a common origin but one is autosomic (RAE730 family) and the other is located in the Y chromosomes (RAYSI) (Navajas-Pérez et al. 2005b, 2006). In turn, the RAE180 family analyzed in this article originated before the split between XX/X₁Y₂ species and has no relation with the other satellite-DNAs (Navajas-Pérez et al. 2005b). In XX/X₁Y₂ species, RAE180 repeats are located in the Y chromosomes with an additional small RAE180 autosomic site (Cuñado et al. 2007; Navajas-Pérez et al. 2009). Thus, in the relatively short evolutionary time span from the origin of the XX/X₁Y₂ clade (12–13 mya; Navajas-Pérez et al. 2005a), the Y chromosomes have undergone an accelerated process of genetic degeneration mediated by the accumulation of RAE180 and RAYSI repetitive sequences (Fig. 5). In fact, we demonstrated here that the 5.6% of the genome of *R. acetosa* and *R. papillaris* males is made up of the RAE180 repeats while the percentage in females is 0.01%. With the assumption of a genome size of 3,234 Mb (Bennett and Leitch 2004) for *R. acetosa*, this would indicate that there are about 1 million copies of RAE180 sequences in the males (repeats located in the Y chromosomes and in the minor locus of a pair of autosomes) and about 2,000 copies of RAE180 repeats in females (only the autosomic locus) of XX/X₁Y₂ species. *R. acetosa* and *R. papillaris* are two

species belonging to a homogeneous group of species characterized by similar morphological and karyological features, including their complex sex-chromosome system plus a sex-determination mechanism based on the X/A balance (Wilby and Parker 1988; Ainsworth et al. 1999). Nevertheless, although both species appear to have karyotype similarities and similar amounts of RAE180 sequences in their Y chromosomes, heterochromatin of the Ys is the subject of rapid evolutionary rearrangement mechanisms that led to differences in satellite-DNA distribution in both species (Navajas-Pérez et al. 2009).

On the other hand, the case for XX/X₁Y₂ species is different. Contrasting with the advanced state of genetic differentiation between the X and the Y chromosomes in *R. acetosa* and *R. papillaris*, we have found that *R. acetosella* and *R. suffruticosus* represent an earlier stage of genetic differentiation between sex chromosomes and also that the Y chromosome in these species is euchromatic and apparently lacks any type of satellite-DNA family (Cuñado et al. 2007). In fact, RAE180 was the only satellite-DNA family detected in the genomes of XX/X₁Y₂ species (Navajas-Pérez et al. 2006; Cuñado et al. 2007). However, RAE180 repeats are poorly represented within the genome of these species. RAE180 was found to be slightly amplified in *R. suffruticosus* in which the repeats are accumulated in a minor locus in a pair of autosomes (Cuñado et al. 2007). In the case of *R. acetosella*, even when PCR and Southern-blot hybridization techniques revealed the presence of RAE180 sequences within its genome, a FISH assay was unable to detect these repeats in their chromosomes, a situation that might be explained by the fact that RAE180 sequences in *R. acetosella* are underrepresented or non-tandemly organized at a level below the resolution of the FISH technique (Cuñado et al. 2007). These data corroborate the quantification developed in this research. In the case of *R. suffruticosus*, RAE180 repeats amount to 0.04% of the genome (for a rough estimate, if we assume a similar genome size to *R. acetosa* for this species, we could estimate some 7,000 copies of RAE180 sequences in *R. suffruticosus*). In contrast, we found that RAE180 satellite DNA represents

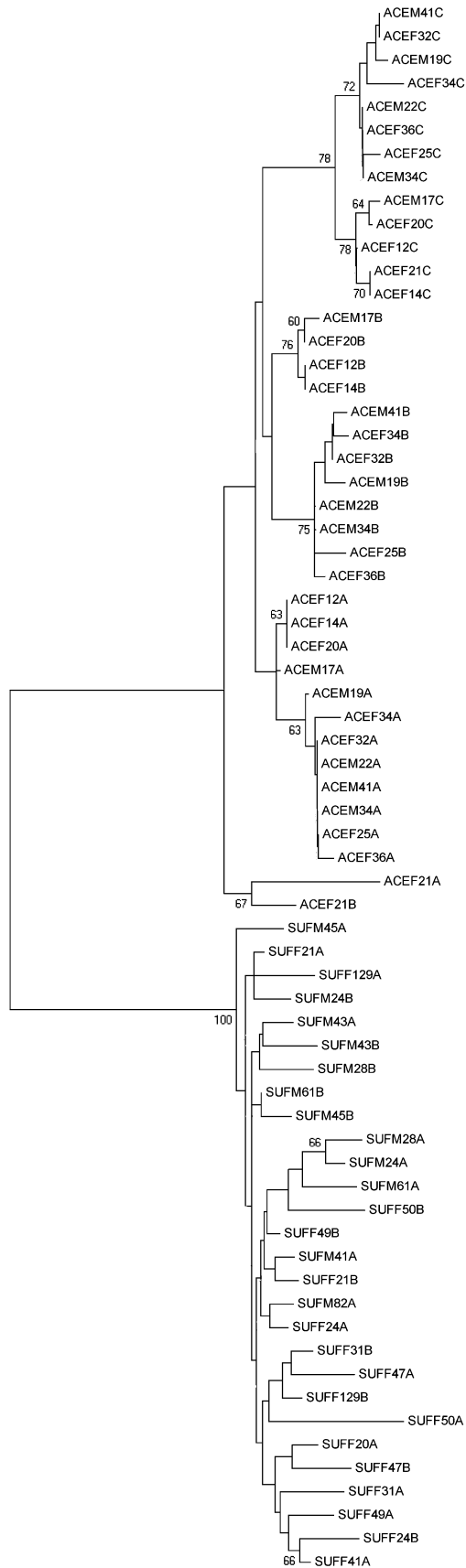


Fig. 3 Neighbor-joining tree of RAE180 repeats from *R. suffruticosus* and *R. acetosella*. The repeats were clustered according to taxonomic affinity into two main clades supported by 100% of bootstrap replicates, one composed of *R. suffruticosus* repeats and the other composed by the *R. acetosella* repeats. Only bootstrap values above 60% are indicated

$7.33 \times 10^{-4}\%$ of the *R. acetosella* genome. Genome size estimated for this tetraploid species is 6,566 Mb (Bennett and Leitch 2004), which could indicate an absolute number of about 200 RAE180 repeats in its genome. The differences estimated here for the percentage of RAE180 satellite DNA between the genomes of these two species, *R. suffruticosus* and *R. acetosella*, could explain the difference in FISH results (Cuñaado et al. 2007).

The RAE180 satellite-DNA family is present within the genomes of the Eurasian and American dioecious *Rumex* species, while it is absent from the rest of *Rumex* species. This data support the contention that the RAE180 satellite-DNA family was present in the genome of the common ancestor of all dioecious *Rumex* species thus dating the origin of RAE180 sequences at least in 15–16 mya (Navajas-Pérez et al. 2005a). The differences in the amount of RAE180 repeats through the different species analyzed here may indicate differential events of amplification, depending on the evolutionary time and the physical place of accumulation. For RAE180 satellite DNA, there are three different observable cases (Fig. 5): (i) maintenance of RAE180 repeats in low copy in *R. acetosella*, (ii) moderate amplification in one autosomal locus in *R. suffruticosus*, and (iii) massive amplification in Y chromosomes of XX/XY₁Y₂ species and a moderate amplification in one autosomal locus. Despite the common origin of all RAE180 sequences and the short evolutionary time lapse from the split of the two groups of species—XX/XY versus XX/XY₁Y₂—(12–13 mya—Navajas-Pérez et al. 2005a), there are two situations.

1. RAE180 sequences have not amplified within the genomes of XX/XY species except for a single autosomal locus detectable by FISH only in *R. suffruticosus*. The situation in *R. acetosella* in which there are a few copies of RAE180 repeats not detectable by the FISH technique might be considered an ancestral situation maintained in the XX/XY lineage of dioecious *Rumex*. A derived state within this lineage is the amplification process detected in *R. suffruticosus*. Thus, this amplification process might have occurred after the split of these two species and hence might have occurred in the last 4 million years, according to data gathered by Navajas-Pérez et al. (2005a) for ITS sequence divergence between *R. acetosella* and *R. suffruticosus*. Quantitative changes induced by the amplification of RAE180 satellite repeats could occur in the course of

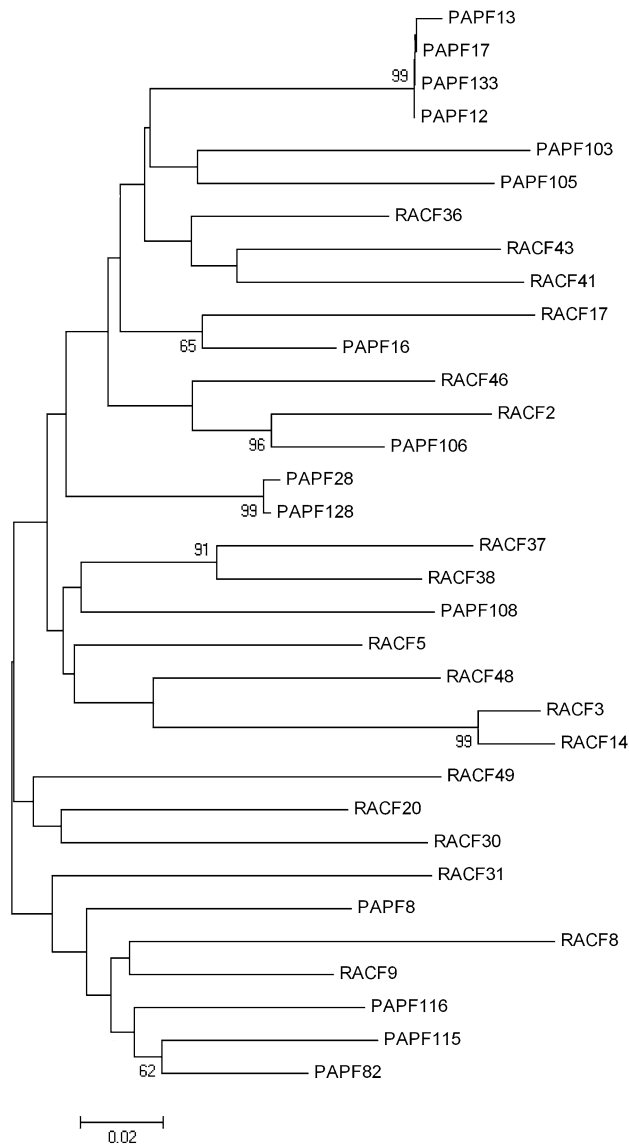


Fig. 4 Neighbor-joining tree of RAE180 repeats isolated from females of *R. acetosa* and *R. papillaris*. RAE180 repeats from females of both species appear intermixed in the tree indicating no differentiation between the two species for these types of sequences

the speciation process, forming a species-specific profile. Thus, sequence divergence is evident between *R. suffruticosus* and *R. acetosella*. In fact, 14% of the nucleotide positions per repeat are differentiated sites (species-diagnostic mutations or Strachan stages IV–VI) between the two species (Table 2). Only 4% of the nucleotide sites per repeat are ancestral polymorphic sites shared between both species. These data indicate effectiveness of the molecular drive process after species split in the spreading of new sequence variants leading to intra-specific homogeneity (5.1 and 5.4% of sequence variation, respectively; Table 1) and inter-specific divergence (22.6% of sequence divergence between the two species; Table 2), a pattern known as

concerted evolution. Transitional stages of differentiation (Strachan stages II and III) represent 24% of the nucleotide positions (Table 2). This data support a gradual process of sequence differentiation between *R. suffruticosus* and *R. acetosella* after RAE180 amplification and confirm the double source of the observed pattern: the effect in the molecular drive process of the relatively recent RAE180 amplification in *R. suffruticosus* and the effect of the molecular-exchange homogenizing mechanisms once amplified the locus in this species. Contrasting with that occurring for RAE180 sequences when they are located in the Y chromosomes (Navajas-Pérez et al. 2005b), the autosomal tandem arrays of RAE180 sequences show high degrees of sequence divergence when comparing these two related species. In fact, we detected a rate of sequence substitution of 28×10^{-9} substitutions per site per year, a rate three-fold higher than that estimated for Y-linked RAE180 sequences (Navajas-Pérez et al. 2007b) but similar to the evolutionary rate found for satellite DNA in other species such as in the *Drosophila obscura* complex (Bachmann and Sperlich 1993). In *Rumex*, other autosomic satellite-DNA family, the RAE730 family had a sequence substitution rate of 22×10^{-9} . The comparison between *R. acetosa* and *R. papillaris* that splits 2.5 mya (Navajas-Pérez et al. 2005a) demonstrated 4% of Strachan stages II–III and 30% of Strachan stages IV–VI (Navajas-Pérez et al. 2007b). Intra-specific variability was 4.6% for *R. acetosa* and 2.8% for *R. papillaris* (Navajas-Pérez et al. 2007b). While we have found similar levels of intra-specific variation in *R. suffruticosus* and *R. acetosella*, the distribution pattern of that variation within tandem arrays of *R. acetosella* is specific with the appearance of HORs, which could indicate certain limitations for molecular drive in this species (Pons and Gillespie 2003), perhaps as a consequence of the presence of RAE180 sequences only in basal amounts (Cuñado et al. 2007; this paper).

2. RAE180 repeats were dramatically amplified within the XX/X₁Y₁Y₂ species clade. This amplification process was an independent event to that occurring in the lineage of the XX/X₁Y species. The amplification of repeat-copy number of RAE180 in the Y chromosomes of the XX/X₁Y₁Y₂ lineage might have occurred recurrently and independently several times in each new lineage, leading to each current XX/X₁Y₁Y₂ species. If this was the case, we would expect differential amplification patterns with genetic divergence between species. Assuming the non-recombining nature of the Y chromosomes, we would also expect a high number of polymorphic sites with a species-specific profile. However, we have found similar amplification levels

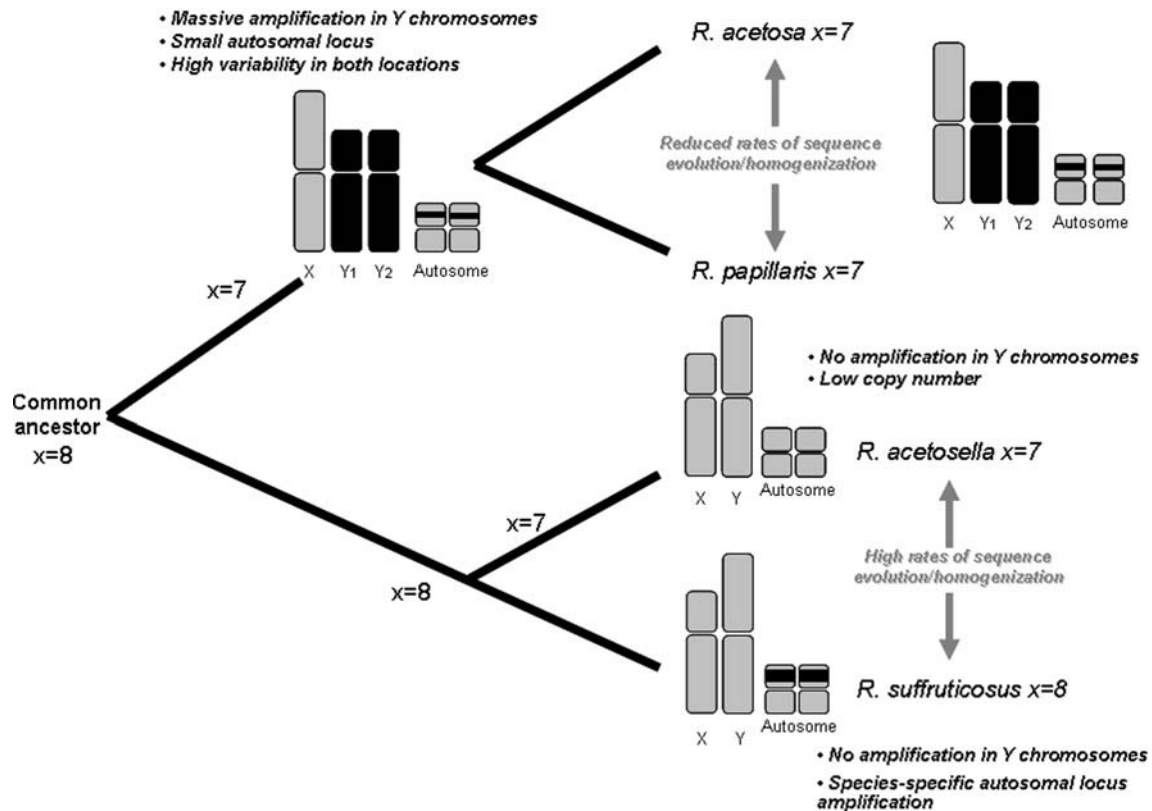


Fig. 5 Schematic representation of the evolutionary dynamics of RAE180 sequences in dioecious *Rumex* species. Black color in chromosomes indicates presence of RAE180 sequences

(Table 1) and a high number of shared polymorphic sites (59% of the nucleotide positions per repeat, Table 2) between *R. acetosa* and *R. papillaris*. These two data support the idea: (a) that the main amplification event of the RAE180 satellite-DNA family precedes the split between these two species, despite subsequent new events of amplification and reorganizations (Navajas-Pérez et al. 2009); (b) that most of the intra-specific variability found in each species is ancestral, originated prior to the split of these two species; and (c) that after the split of these two species, processes leading to concerted evolution has failed. In fact, there are high levels of variability in both species (19.9% in *R. acetosa* and 21.1% in *R. papillaris*; Table 1) and only one species-specific diagnostic site (Table 2). These results were explained on the basis of the absence of recombination in the Y chromosomes, a key factor in the molecular drive process (Navajas-Pérez et al. 2005b). The pattern of the autosomal locus of *R. acetosa* and *R. papillaris* resembles Y-linked RAE180 repeat behavior, an unexpected conclusion as a priori we assumed a similar evolutionary scenario for the autosomal loci in both XX/XY and XX/XY₁Y₂ species. In contrast, the autosomal loci of *R. acetosa* and *R. papillaris* showed high levels of intra-specific vari-

ability (23.2 and 15.3%, respectively; Table 1) with no fixed differences between the *R. acetosa* and *R. papillaris* (Table 2). Similar amplification levels (Table 1) and the high number of shared polymorphic sites (34% of the nucleotide positions per repeat, Table 2) between *R. acetosa* and *R. papillaris* again supports the contention that intra-specific variability found in each species at the autosomal locus is ancestral, originated prior to the split of these two species, and, after the split of these two species, processes leading to concerted evolution has failed. This situation might be explained in this case also by reduced rates of recombination for the autosomal locus of *R. acetosa* and *R. papillaris*. It appears that the small autosomal locus found in these two species was singled out of expansion–contraction homogenizing processes that should have lead to concerted evolution. Variability can remain for long evolutionary periods by reduced action of molecular mechanisms of non-reciprocal exchange and sequence variants persist as a library (Mravinac et al. 2002; Meštrović et al. 2006) from which the amplification of a particular divergent variant could give rise later to a new satellite-DNA subfamily (Navajas-Pérez et al. 2006) or to a new enlarged array of the sequence variant (as in *R. suffruticosus*). An alternative explanation

would imply that part of the genetic variability found in autosomes comes from the Y chromosomes by some sort of genetic-transfer mechanism.

In conclusion, this study demonstrates the effect of differential location, organization, and repeat-copy number in the evolution of a satellite-DNA family. Molecular drive leads to the concerted-evolution pattern of satellite-DNA repeats. Molecular mechanisms of non-reciprocal exchange are key factors in the molecular drive process leading to the homogenization of tandem arrays within the genome of a species. Satellite-DNAs in the non-recombining Y chromosomes show low rates of concerted evolution and intra-specific variability increase with no inter-specific divergence. By contrast, freely recombining loci undergo concerted evolution with genetic differentiation between species, as occurs between the autosomal loci of *R. acetosella* and *R. suffruticosus*. However, it appears that satellite-DNA evolution might also have evolutionary periods of stasis with variability remaining by reduced action of molecular mechanisms of non-reciprocal exchange, as occurred between the autosomal loci of *R. acetosa* and *R. papillaris*, something that could depend on repeat-copy number and the processes involved in its amplification.

Acknowledgments Our thanks to Dr. Ginés López González (Real Jardín Botánico de Madrid, Spain), and to the Parque Nacional de Sierra Nevada and to the Parque Natural de la Sierra de Baza, for kindly providing the material analyzed in this article. We also are deeply indebted to Dr. Miguel Burgos of the University of Granada for advice and help with the satellite-DNA quantification procedure. This research was supported by grant CGL2006-00444/BOS of the Ministerio de Ciencia e Innovación (Spain) and Fondo Europeo de Desarrollo Regional (FEDER, EU).

References

- Ainsworth CC, Lu J, Winfield M, Parker JS (1999) Sex determination by X:autosome dosage: *Rumex acetosa* (sorrel). In: Ainsworth CC (ed) Sex determination in plants. BIOS Scientific Publishers Limited, Oxford, pp 124–136
- Bachmann L, Sperlich D (1993) Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. Mol Biol Evol 10:647–659
- Bennett MD, Leitch IJ (2004) Plant DNA C-values database. Release 3.0
- Cuñado N, Navajas-Pérez R, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Santos JL, Garrido-Ramos MA (2007) The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. Chromosome Res 15:825–832
- Degraeve N (1975) Contribution à l'étude cytotaxonomique des *Rumex*. I Le genre *Rumex*. Caryologia 28:187–251
- Degraeve N (1976) Contribution à l'étude cytotaxonomique des *Rumex*. IV Le genre *Acetosa* Mill. Cellule 71:231–240
- Degraeve N (1980) Contribution à l'étude cytotaxonomique des *Rumex*. III Le genre *Acetosella* Fourr. Genetica 54:29–34
- Dover G (2002) Molecular drive. Trends Genet 18:587–589
- Feliciello I, Picariello O, Chinali G (2005) The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. Gene 349:153–164
- Feliciello I, Picariello O, Chinali G (2006) Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene 383:81–92
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
- Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. Evol Biol 30:93–120
- Meštrović N, Castagnone-Sereno P, Plohl M (2006) Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. Mol Biol Evol 23:2316–2325
- Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. Mol Biol Evol 17:1597–1609
- Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002) Sequence of PRAT satellite DNA “frozen” in some Coleopteran species. J Mol Evol 54:774–783
- Navajas-Pérez R, de la Herrán R, López González G, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2005a) The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. Mol Biol Evol 22:1929–1939
- Navajas-Pérez R, de la Herrán R, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2005b) Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). J Mol Evol 60:391–399
- Navajas-Pérez R, Schwarzacher T, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2006) The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. Gene 368:61–71
- Navajas-Pérez R, Ruiz Rejón M, Garrido-Ramos MA, Aznarte JL, Rubio-Escudero C (2007a) SatDNA Analyzer: a computing tool for satellite-DNA evolutionary analysis. Bioinformatics 23:767–768
- Navajas-Pérez R, Rubio-Escudero C, Aznarte JL, Ruiz Rejón M, Garrido-Ramos MA (2007b) SatDNA Analyzer 1.2 as a valuable computing tool for evolutionary analysis of satellite-dna families: revisiting Y-linked satellite-DNA sequences of *Rumex* (Polygonaceae). In: Hochreiter S, Wagner R (eds) Lecture notes in computer sciences, bioinformatics research and development, vol 4414. Springer-Verlag, Berlin, pp 131–139
- Navajas-Pérez R, Schwarzacher T, Ruiz Rejón M, Garrido-Ramos MA (2009) Molecular cytogenetic characterization of *Rumex papillaris*, a dioecious plant with an XX/XY1Y2 sex chromosome system. Genetica 135:87–93
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Ohta T, Dover G (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521
- Pons J, Gillespie RG (2003) Common origin of the satellite DNAs of the Hawaiian spiders of the genus *Tetragnatha*: evolutionary constraints on the length and nucleotide composition of the repeats. Gene 313:169–177
- Pons J, Petitpierre E, Juan C (2002) Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). Mol Biol Evol 19:1329–1340
- Saito N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
- Smith BW (1969) Evolution of sex-determining mechanisms in *Rumex*. Chromosomes Today 2:172–182
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

- Strachan T, Webb D, Dover G (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. EMBO J 4:1701–1708
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882
- Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles—causes and effects. EMBO J 21:5955–5959
- Wilby AS, Parker JS (1988) Recurrent patterns of chromosome variation in a species group. Heredity 61:55–62

Quesada del Bosque ME, Navajas-Pérez R, Panero JL, Fernández-González A, Garrido-Ramos MA. A satellite DNA evolutionary analysis in the North American endemic dioecious plant *Rumex hastatulus* (Polygonaceae). *Genome*, 2011; 54(4): 253-60. doi: 10.1139/g10-115.

A satellite DNA evolutionary analysis in the North American endemic dioecious plant *Rumex hastatulus* (Polygonaceae)

M.E. Quesada del Bosque, R. Navajas-Pérez, J.L. Panero, A. Fernández-González, and M.A. Garrido-Ramos

Abstract: We studied the evolution of RAE180 satellite DNA family in the North American endemic dioecious plant *Rumex hastatulus*. In this species, the Texas race is characterized by a single XX/XY sex chromosome system, whereas the North Carolina race has evolved a derived complex XX/XY₁Y₂ sex chromosome system. RAE180 repeats were autosomic and poorly represented (2×10^{-4} % of the genome) with no differences between individuals of different genders or different races of *R. hastatulus*. In fact, the sex chromosomes of the North Carolina race are still euchromatic, and they have not accumulated satellite DNA sequences, which contrasts with that occurring in the rest of dioecious XX/XY₁Y₂ *Rumex* species. In *R. hastatulus*, we detected the existence of three RAE180 subfamilies. Notwithstanding, while in the Texas race the TX1/NC1 subfamily is the most frequent, the TX2/NC2 subfamily is the most abundant in the North Carolina race. Additionally, the third, less represented subfamily (TX3/NC3) appears currently as relict sequences in both genomes. A common feature of RAE180 satellite is the sudden replacement of one sequence variant by another in different species (or populations as in *R. hastatulus* races). Thus, the phylogenetic analysis of RAE180 repeats from six dioecious *Rumex* species supports the “library” hypothesis. According to this hypothesis, we assume that a set of divergent RAE180 variants were present in the ancestral genome of dioecious *Rumex* species, from which novel tandem arrays originated by the amplification of different variants in different lineages. Differential levels of RAE180 satellite DNA amplification in each lineage, at different evolutionary times, and in different chromosomal positions gave rise to differential patterns of sequence evolution.

Résumé : Les auteurs ont étudié l'évolution de l'ADN satellite RAE180 chez le *Rumex hastatulus*, une plante dioïque endémique en Amérique du Nord. Chez cette espèce, la race Texas est caractérisée par un système à un seul chromosome sexuel XX/XY, tandis que la race North Carolina possède un système complexe dérivé de type XX/XY₁Y₂. Les répétitions RAE180 sont autosomiques et peu abondantes (2×10^{-4} % du génome) et ne présentent pas de différences entre individus des deux sexes ou entre les différentes races du *R. hastatulus*. De fait, les chromosomes sexuels de la race North Carolina sont encore euchromatiques et n'ont pas encore accumulé d'ADN satellite, contrairement à ce qui est observé chez les autres espèces dioïques XX/XY₁Y₂ du genre *Rumex*. Chez le *R. hastatulus*, les auteurs ont décelé la présence de trois sous-familles de RAE180. Tandis que chez la race Texas la sous-famille TX1/NC1 est la plus abondante, la sous-famille TX2/NC2 est la plus abondante chez la race North Carolina. De plus, une troisième sous-famille (TX3/NC3), la moins abondante, semble constituer une relique chez ces deux génomes. Le remplacement soudain d'un variant par un autre chez les différentes espèces (ou populations comme les races du *R. hastatulus*) est une caractéristique commune chez le satellite RAE180. Ainsi, l'analyse phylogénétique des répétitions RAE180 identifiées chez six espèces dioïques du genre *Rumex* supporte l'hypothèse de la « bibliothèque ». Selon cette hypothèse, les auteurs postulent qu'une collection de variants distincts de RAE180 était présente dans le génome ancestral des espèces dioïques de *Rumex* et que de nouveaux amas dupliqués en tandem sont apparus suite à une amplification de différents variants au sein des différents lignages. Des degrés différents d'amplification de l'ADN satellite RAE180 au sein de chaque lignage, à différents moments au cours de l'évolution et à différentes positions chromosomiques auraient donné lieu aux différents résultats observés.

Mots-clés : *Rumex hastatulus*, chromosomes sexuels, ADN satellite, évolution concertée, hypothèse de la « bibliothèque ».

[Traduit par la Rédaction]

Received 15 July 2010. Accepted 25 November 2010. Published on the NRC Research Press Web site at genome.nrc.ca on 1 April 2011.

Paper handled by Associate Editor Perry Gustafson.

M.E. Quesada del Bosque, R. Navajas-Pérez, A. Fernández-González, and M.A. Garrido-Ramos. Departamento de Genética, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain.

J.L. Panero. Section of Integrative Biology, 1 University Station, C0930, University of Texas, Austin, TX 78712, USA.

Corresponding author: M.A. Garrido-Ramos (e-mail: mgarrido@ugr.es).

Introduction

American and Eurasian dioecious species of the plant genus *Rumex* form a clade divided in two sister groups, one composed of species with an XX/XY sex chromosome system (*Rumex acetosella* and closely related species such as *Rumex graminifolius* and *Rumex paucifolius*, and *Rumex suffruticosus* and *Rumex hastatulus*) and the other including species with an XX/XY₁Y₂ system (*Rumex acetosa* and its relatives *Rumex papillaris*, *Rumex intermedius*, or *Rumex thyrsoides*) (Navajas-Pérez et al. 2005a). Sex chromosomes of the XX/XY species are barely differentiated, whereas the XX/XY₁Y₂ species are in a further stage of differentiation with heteromorphy between sex chromosomes (Cuñado et al. 2007). In addition, the Y chromosomes of XX/XY₁Y₂ species, but not those of XX/XY species, are heterochromatic and have accumulated a set of diverse repetitive sequences (Shibata et al. 1999, 2000; Navajas-Pérez et al. 2006, 2009a; Cuñado et al. 2007; Mariotti et al. 2006, 2009).

There are several satellite DNA families in *Rumex*: RAYSI (Y-specific) and RAE730 (autosomic) appeared only in the XX/XY₁Y₂ species of the genus (Navajas-Pérez et al. 2006, 2009a), as well as the RAYSII and RAYSIII families (Mariotti et al. 2009), whereas the RAE180 family is conserved in all dioecious species of *Rumex* (Navajas-Pérez et al. 2005b, 2006, 2009a, 2009b; Cuñado et al. 2007). In the XX/XY species, the RAE180 satellite DNA is located only in a small locus of a pair of autosomes, being poorly represented within their genomes (Cuñado et al. 2007; Navajas-Pérez et al. 2009a, 2009b). RAE180 repeats are characterized by intraspecific sequence homogeneity and interspecific divergence showing a general pattern of concerted evolution (Navajas-Pérez et al. 2009b). In contrast, in the case of XX/XY₁Y₂ species, the RAE180 sequences are present also in a small autosomal locus, but additionally have been massively amplified in the Y chromosomes (Cuñado et al. 2007; Navajas-Pérez et al. 2009a, 2009b). RAE180 sequences show very high levels of intraspecific variability, mostly ancestral because processes leading to concerted evolution have failed both before and after the diversification of the XX/XY₁Y₂ group. These results were explained on the basis of the absence of recombination in the Y chromosomes, a key factor in the concerted evolution process (Navajas-Pérez et al. 2005b, 2009b).

Interestingly, the North American endemic *R. hastatulus* resembles the global situation of the genus with two genetically — but not morphologically — differentiated chromosomal races that differ in their sex chromosome systems (Smith 1963, 1969; Navajas-Pérez et al. 2005a). The so-called Texas race is characterized by a simple XX/XY system, whereas the North Carolina race has a complex XX/XY₁Y₂ one. Despite its multiple sex chromosome system, according to molecular and cytogenetical data (Navajas-Pérez et al. 2005a), the North Carolina race of *R. hastatulus* is classified into the XX/XY group, implying secondary evolution from the XX/XY to the XX/XY₁Y₂ sex chromosome system (Smith 1969). More importantly, the same evolutionary change from one to the other sex chromosome system seems to have occurred independently in the two lineages of *Rumex*, one in the ancestor of the XX/XY₁Y₂ group and the other in *R. hastatulus*, as proposed earlier by Smith (1969) and Degraeve (1976).

In this context, given the parallel origin of the XX/XY₁Y₂ in two different lineages of dioecious species of *Rumex*, we sought to compare the events occurring in sex chromosomes evolution of both lineages and analyze (i) the genetic differentiation of the RAE180 satellite DNA between the Texas and North Carolina chromosomal races of *R. hastatulus* and (ii) the origin and evolution of this family of repeats, taking into account the existence of differential events of RAE180 satellite DNA amplification in different *Rumex* species. For this, we have compared RAE180 sequences from both *R. hastatulus* races as well as between these *R. hastatulus* repeats and other RAE180 repeats of the rest of dioecious *Rumex* species.

Materials and methods

Materials, DNA extraction, and RAE180 repeats isolation

Female and male plants were collected in the wild in Mason County, Texas (Texas race) and in Cumberland County, North Carolina (North Carolina race). The DNA from leaves of four males and three females from Texas and the DNA from leaves of two males and three females from North Carolina were isolated using DNAeasy mini kit (Qiagen) according to the manufacturer's instructions.

The characterization of RAE180 sequences was carried out separately in the two chromosomal races of *R. hastatulus*. Additionally, RAE180 sequences were also isolated from the genome of *R. intermedius* males (an XX/XY₁Y₂ species closely related to *R. acetosa*). For PCR amplification of RAE180 sequences from each selected individual, we used the primer pair Rae180a and Rae180b, 5'-TCATCGAAGCTT-CATTCAT-3' and 5'-TATAGTAATATCTCGATC-3', and PCR procedure as well as purification, cloning, and sequencing of the PCR products were as described in Navajas-Pérez et al. (2009b).

A total of 76 repeats from the Texas race (43 from males and 33 from females) and 58 repeats from the North Carolina race (24 from males and 34 from females) were used for sequence analyses. The EMBL accession numbers are FN561567 to FN561625 (North Carolina race) and FN563834 to FN563909 (Texas race). Additionally, eight RAE180 repeats were selected among recombinant clones obtained from *R. intermedius* males (EMBL accession numbers FN564362 to FN564369).

Sequence analysis

Multiple alignments were performed using ClustalX (Thompson et al. 1997) followed by manual adjustments. Alignments were constructed for the 76 Texas samples, for the 58 North Carolina samples, and for the Texas and North Carolina samples combined. Additionally, multiple alignments were made for a representative set of RAE180 sequences of six *Rumex* species: 16 sequences from *R. hastatulus* (eight from each race) and eight sequences from each of three XX/XY₁Y₂ species (*R. acetosa* and *R. papillaris* — Navajas-Pérez et al. 2005b; and *R. intermedius* — this paper) and two XX/XY species (*R. acetosella* and *R. suffruticosus* — Navajas-Pérez et al. 2009b). For the characterization of RAE180 sequences from *R. hastatulus*, we determined variation within race (measured as nucleotide diversity, π , the average number of nucleotide differences per site between

Fig. 1. Bootstrap consensus neighbor-joining tree of RAE180 repeats isolated from both *Rumex hastatulus* chromosomal races. Sequence code: TX, Texas; NC, North Carolina; F,M: female, male; numbers represent the number of the individual followed by the number of the repeat analyzed. To assess internal support for nodes, 1000 bootstrap replicates were performed. Only bootstrap values above 60% are indicated.

two sequences; Nei 1987) and differences between races (measured as the average number of nucleotide substitution per site between races, D_{xy} ; Nei 1987) using the software satDNA Analyzer (Navajas-Pérez et al. 2007).

For phylogenetic analyses of RAE180 sequences, we used the aligned data as distance matrix sources. Distances were calculated according to the Kimura gamma distance and trees constructed by the neighbor-joining method in MEGA v4.1 (Tamura et al. 2007). For selection of the distance measuring method, the aligned sequences were subjected to analysis using Modeltest v.3.6 (Posada and Crandall 1998), which performs a hierarchical test of likelihood fits under 56 different models of DNA substitution. To assess internal support for nodes, 1000 bootstrap replicates (Felsenstein 1985) were performed.

Satellite DNA quantification

A quantitative PCR (qPCR) approach, following the procedure described in Navajas-Pérez et al. (2009b), was used to quantify relative percentage of RAE180 repeats in different genomes (males and females of the two *R. hastatulus* races analyzed). After relative percent quantification of RAE180 repeats, a rough estimate of the number of RAE180 copies present in *R. hastatulus* males and females for each race was made by assuming a total genome size similar to *R. acetosa* (3234 Mb; Bennett and Leitch 2004), as there are no data for *R. hastatulus* genome size.

Chromosomal analyses

Root tips were obtained from seeds germinated in Petri dishes and then pretreated with 2 mmol/L 8-hydroxyquinoline for 4 h for metaphase accumulation. They were then fixed in ethanol-glacial acetic acid (3:1) until required. For DAPI staining and for fluorescence in situ hybridization (FISH) experiments, we followed the procedure described in Cuñado et al. (2007). As probes for FISH, we used RAE180 repeats directly amplified by PCR from the genomic DNAs of individuals of both *R. hastatulus* races.

Results and discussion

Characterization of RAE180 sequences and genetic differentiation between *Rumex hastatulus* chromosomal races

RAE180 satellite DNA represents about $2 \times 10^{-4}\%$ of the genome of *R. hastatulus*, with no differences in amounts between individuals of different genders and different races, which is consistent with an autosomic provenance of these repeats. Assuming that *R. hastatulus* has a similar genome size to *R. acetosa* (3234 Mb; Bennett and Leitch 2004) (there are no genome size data for *R. hastatulus*), this would equate to approximately a few tens of copies of RAE180 repeats.



Genome Downloaded from www.nrcresearchpress.com by 77.229.98.32 on 04/08/11 For personal use only.

Fig. 2. Multiple sequence alignment of representative sequences of RAE180 sequences from the two races of *Rumex hastatulus*. Subfamily TX1/NC1: TX_M41 and NC_F22; Subfamily TX2/NC2: TX_M14 and NC_F14; Subfamily TX3/NC3: TX_F65 and NC_M41.

```

TX_M41  TATTGAATTATATCGATCATA-TGTGAAAAA-ATCGGCCCGTTCC-GCCAC--CGATACTCGAGAACGGTCAATATCAT
NC_F22  ..A..G...T.C.....-T...G...-...T...-...G.C-AT...T...A.C.AACA.....C
TX_M14  ....G.....G.....-T.....-ATT.GATCG.T.C.A.-AT.....A.....C.....
NC_F14  ...G.G.....G.....AT-AT..G...-ATT.GATCG.T.C.T.C-AT.....A.C...C.....
TX_F65  ....G.....C.T.....TT.TCT...ACAAT.GTTTG.TTCAAA.AAT.....T..A.....C.....-
NC_M41  ....G.....C.T.....TG.T.TT...-CAAT.GGTTG.TTCAAA.-AT.....T..A.-.A.C.....-

TX_M41  CGAA-TAGTT---AGCAATAGTGAAG----GAAGTTCAATGAAAAATAGTCGATATTCACCTATGTCGATGACCCGAACT
NC_F22  TT.-C....-CT.T.G..A.....----.....TGG.....G.....A..A.....AG...
TX_M14  T...-C.A.-CT...G....ACGTGAAA.....G.....G.....G.....C...C...A.G...
NC_F14  -...-C.A..CCT...G....ACGTGAAA.....G.....G.....G.....G.T...C...C...G...
TX_F65  T...A.....-CT.T.G..A...T----.....GTG...ATA.T.GA..G..T...A.A..C...AAG...
NC_M41  T...A.....-CT.T.G..A...T----.....GGG.....GA.T.GAT..G..T...A.A..C...A-G...

TX_M41  CGTTTTTGGTTAAATCTTTTGTATATTTTGTATCAA
NC_F22  ..C....C.....TA.....
TX_M14  .A....CC.....G.....
NC_F14  .A....CC.....TC.....
TX_F65  .A....CT..C...A...GT.....
NC_M41  .A....CT..CC.A.C...G..TA...CCCC...

```

This is the smallest number of copies found for RAE180 sequences within *Rumex* species since we estimated about 200 and 7000 copies, respectively, in *R. acetosella* and *R. suffruticosus* (XX/XY species) and about one million copies of RAE180 repeats in the XX/XY₁Y₂ species *R. acetosa* (Navajas-Pérez et al. 2009b). Therefore, as reported for other *Rumex* species of the XX/XY clade (Cuñado et al. 2007; Navajas-Pérez et al. 2009b), RAE180 repeats are autosomic and poorly represented in the genome of *R. hastatulus*. Thus, after analysis with FISH using species-specific RAE180 satellite DNA probes, no hybridization signals were observed in *R. hastatulus* chromosomes, confirming that RAE180 sequences are scarcely represented in this species. No DAPI positive heterochromatic regions were observed in the sex chromosomes or the autosomes of *R. hastatulus* (not shown).

Our data reinforce the hypothesis that although the sex chromosome configurations (XX/XY₁Y₂) of *R. acetosa* and their relatives and those of *R. hastatulus* are similar, they arose independently (Smith 1969; Navajas-Pérez et al. 2005a). In addition, despite the convergent origin of the XX/XY₁Y₂ in both lineages, the sex chromosomes of the North Carolina race of *R. hastatulus* are still euchromatic, and they have not accumulated satellite DNA sequences, the main component of heterochromatic regions. A possible explanation could be that the origin of the XX/XY₁Y₂ sex chromosome system of the North Carolina race of *R. hastatulus* is quite recent. In fact, assuming a mean rate of change in plant nuclear DNA of 0.6% per site per million years (Gaut 1998) and using nuclear intergenic transcribed spacer (ITS) mean distance corrected estimates data of Navajas-Pérez et al. (2005a), the split between these two races occurred about 600 000 years ago. This contrasts with the origin of the *R. acetosa* clade, which is estimated to be 12–13 million years ago (mya) (Navajas-Pérez et al. 2005a).

Sequence analysis of RAE180 repeats demonstrated the existence of three satellite DNA subfamilies in both races of *R. hastatulus*: TX1, TX2, and TX3 in the Texas race and NC1, NC2, and NC3 in the North Carolina race. Each RAE180 subfamily was characterized by a set of diagnostic sites, each one representing a particular mutation shared by all the sequences of the subfamily, while at the same sites all

Table 1. Divergence between RAE180 sequences from both chromosomal races of *Rumex hastatulus*.

	Divergence
Between orthologous sequences	
TX1/NC1	0.252
TX2/NC2	0.050
TX3/NC3	0.179
Between most common sequences for each race	
TX1/NC2	0.190

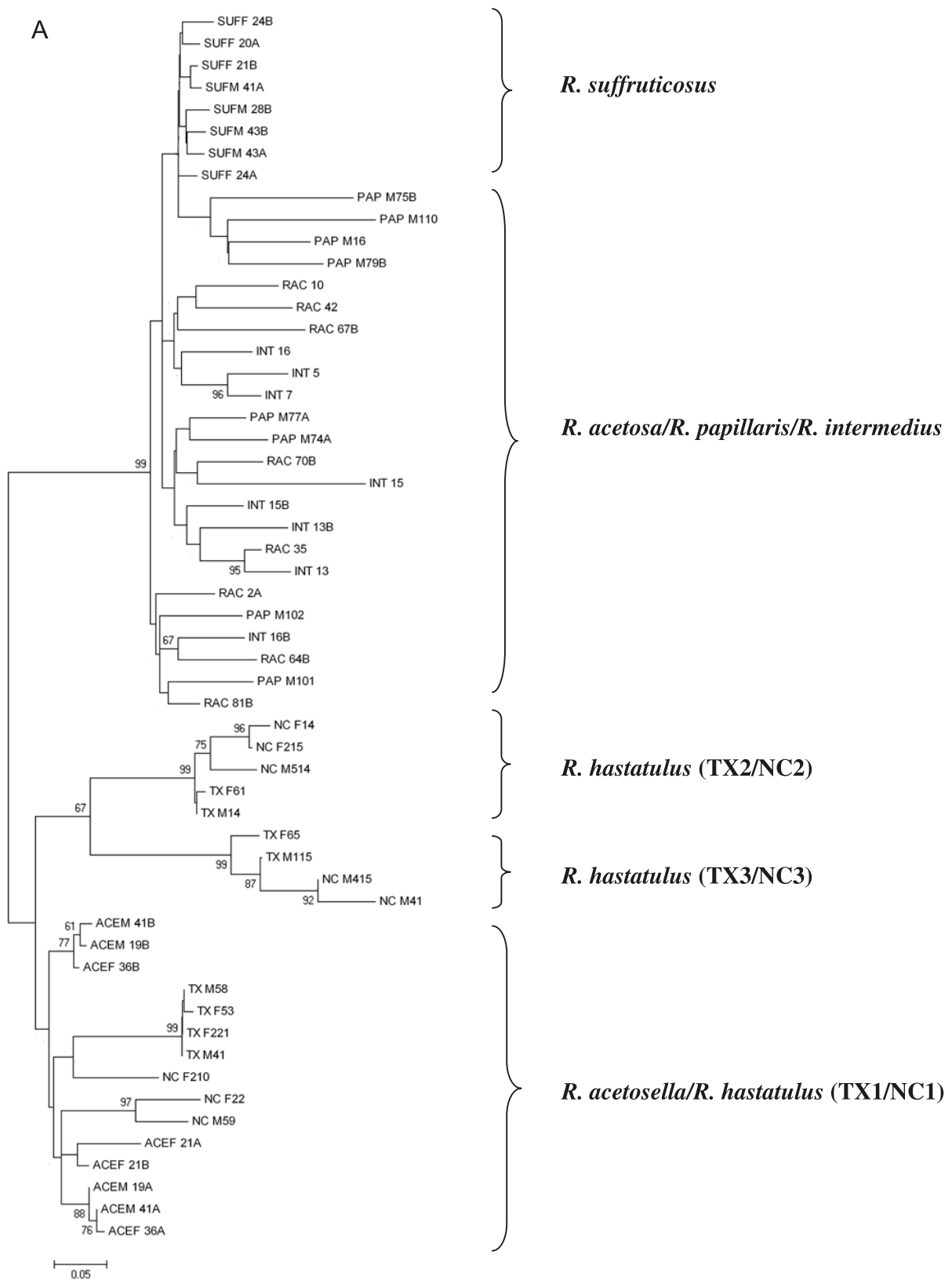
the sequences of the other subfamilies had a different nucleotide (see Supplementary data,¹ Fig. S1). In both races, divergence between subfamilies was higher than variation within subfamilies (Tables S1 and S2), and phylogenetic analysis grouped together sequences by subfamily provenance (Figs. S2 and S3).

Orthology of RAE180 subfamily types was determined based on shared clade membership on the combined-race phylogenetic tree (Fig. 1) and the comparison between individual race trees (Figs. S2 and S3) and the combined species tree. In the combined phylogenetic tree of RAE180 sequences, three groups of orthologous repeats were evident: TX1/NC1, TX2/NC2, and TX3/NC3 (Fig. 1). Sequences from different races shared diagnostic sites when they belonged to the same orthologous subfamily (Fig. 2). Genetic distances between the Texas and North Carolina races (Table 1) were highest for subfamily 1 (0.252) and lowest for subfamily 2 (0.050), with that of subfamily 3 well above the median (0.179).

However, the abundance of each RAE180 subfamily is different between races. In the Texas race, the TX1 subfamily was the most abundant (82% of the sequenced repeats), whereas only 4% of the RAE180 sequences belonged to the TX2 subfamily, and 14% belonged to the TX3 subfamily. In the North Carolina race, the NC2 subfamily was found to be the most abundant with 85% of the sequenced repeats. An average of 10% of the RAE180 sequences belonged to the NC1 subfamily, whereas 5% belonged to the NC3 subfamily.

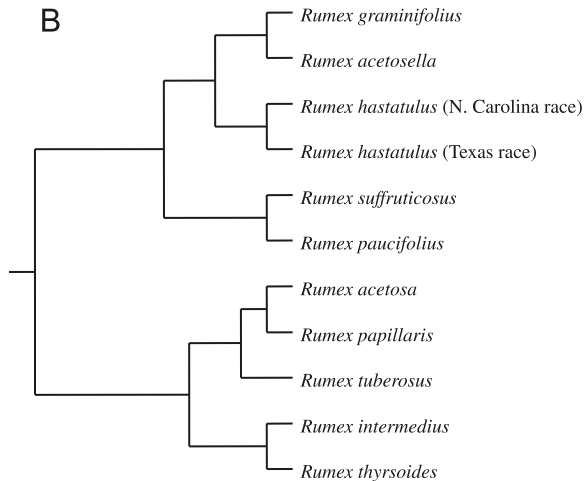
¹Supplementary data are available with the article at www.nrcresearchpress.com/gen

Fig. 3. (A) Bootstrap consensus neighbor-joining tree of RAE180 sequences from different dioecious species of the genus *Rumex*: RAC, *R. acetosa*; PAP, *R. papillaris*; INT, *R. intermedius*; ACE, *R. acetosella*; TX, *R. hastatulus*, Texas race; NC, *R. hastatulus*, North Carolina race. Only bootstrap values above 60% are indicated. (B) Phylogenetic relationships between dioecious *Rumex* species, according to Navajas-Pérez et al. (2005a).



Genome Downloaded from www.nrcresearchpress.com by 77.229.98.32 on 04/08/11
For personal use only.

Fig. 3 (concluded).



Concomitantly, phylogenetic analysis between RAE180 sequences of both *R. hastatulus* races show two major clades (Fig. 1), each corresponding to the Texas race and the North Carolina race. This result is due to the fact that more than 80% of the repeats for each race belong to two different subfamilies: subfamily 1 (TX1) in the case of Texas race and subfamily 2 (NC2) in the case of North Carolina race, with a value of genetic differentiation between TX1 and NC2 of 19% (Table 1). The replacement of one sequence variant by another in different species (or populations as in *R. hastatulus* races) is a common feature of satellite DNA as a consequence of the high dynamism of satellite DNA evolution, as discussed below. In addition, comparisons of orthologous sequences of both chromosomal races show a remarkable level of genetic differentiation (Table 1). RAE180 satellite DNA genetic differentiation between the two chromosomal races of *R. hastatulus* is comparable to that found for the RAE730 satellite DNA of other *Rumex* species (Navajas-Pérez et al. 2005b) and the *EcoRI* satellite DNA of sparids (de la Herrán et al. 2001). The RAE180 satellite DNA is thus a valuable taxonomic marker for the genetic identification of individuals from different races.

RAE180 origin and evolution

According to the “library” hypothesis (Fry and Salser 1977; Meštrović et al. 1998, 2006; Mravinac et al. 2002), related taxa share a library of different conserved satellite DNA sequences (different satellite DNA families but also monomer variants or subfamilies of a satellite DNA family), which may be differentially amplified in each taxon with the subsequent replacement of one sequence variant by another in different species (or populations). In *R. hastatulus*, we have found support for the library hypothesis. The presence of the three RAE180 subfamilies in the two populations analyzed suggests that the existence of this variation preceded the isolation and differentiation of the two *R. hastatulus* chromosomal races. Even though the number of repeated sequences is still low (less than 100 according to our data), there is a differential amplification process in each race of this species for the different RAE180 subfamilies, in agreement with the library hypothesis. In addition, our understand-

ing of the molecular evolution of RAE180 sequences in *R. hastatulus* has helped us understand the evolution of these repeats in other dioecious species of the genus *Rumex*.

The phylogenetic tree of RAE180 sequences isolated from dioecious species of the genus *Rumex* (Fig. 3A) is not congruent with the phylogeny of these species based on other markers (Fig. 3B). Using nuclear ITSs and chloroplast intergenic sequences, we have recently demonstrated that the genus *Rumex* contains three major clades (Navajas-Pérez et al. 2005a): a basal clade composed of strictly hermaphroditic species; a second clade composed of hermaphroditic, polygamous, and gynodioecious species of the genus; and a third clade including all the American and Eurasian dioecious species of *Rumex*. The third clade contains two subclades: (Fig. 3B) one containing *R. hastatulus*, labeled therein as the XX/XY group (*R. hastatulus*, *R. acetosella*, *R. graminifolius*, *R. paucifolius*, and *R. suffruticosus*) and the other composed of *R. acetosa* and relatives (*R. papillaris*, *R. intermedius*, or *R. thyrsoides*) and characterized by having an XX/XY₁Y₂ system (Degraeve 1976; Wilby and Parker 1988). According to Navajas-Pérez et al. (2005a), the dioecious clade of the genus *Rumex* split from the rest of other *Rumex* species 15–16 mya, whereas the two dioecious groups split approximately 12–13 mya.

However, most RAE180 sequences did not group by taxonomic affinity but rather by repeat type (Fig. 3). Most, but not all, sequences grouped according to the sex chromosome system. The exception in both cases was the eight sequences of *R. suffruticosus* that formed a monophyletic clade derived within a statistically well-supported clade containing all XX/XY₁Y₂ species (*R. acetosa*, *R. papillaris*, and *R. intermedius*) sequences. TX2/NC2 sequences and TX3/NC3 sequences each formed well-supported monophyletic clades also. However, TX1/NC1 sequences did not constitute a monophyletic clade and appear more closely related to *R. acetosella* sequences than to either TX2/NC2 or TX3/NC3, but without significant bootstrap support. Therefore, this tree supports again the library hypothesis for RAE180 satellite DNA evolution in *Rumex*.

Based on our phylogenetic analyses, we assume that the ancestral genome of dioecious *Rumex* species contained a set of divergent RAE180 variants from which novel tandem arrays originated by the amplification of different variants in different lineages. The TX1/NC1 variant, predominant in the Texas race of *R. hastatulus*, is closely related to the RAE180 sequences of *R. acetosella*, appearing then as an older type with respect to the TX2/NC2 variant. The derived TX2/NC2 variant is replacing the TX1/NC1 type in the North Carolina race of *R. hastatulus*. In any case, the RAE180 family is autosomic and poorly represented in these two species (Navajas-Pérez et al. 2009b; this paper). In addition, there is a species-specific variant slightly amplified in one pair of autosomes of *R. suffruticosus* (Cuñado et al. 2007; Navajas-Pérez et al. 2009b). In this species, RAE180 repeats are highly homogenized (Navajas-Pérez et al. 2009b; this paper). Finally, there is a sequence variant related to that of *R. suffruticosus* (Fig. 3) and shared by the three XX/XY₁Y₂ species (*R. acetosa*, *R. papillaris*, and *R. intermedius*). The evolution of RAE180 repeats in the XX/XY₁Y₂ lineage is unusual because we have found high levels of intraspecific variability in these species and slowed rates of sequence evolution and

homogenization (Navajas-Pérez et al. 2005b, 2009b; see also Fig. 3). Most of the high intraspecific variability found in each of these species would be ancestral variability, and the nonrecombining nature of the Y chromosomes would explain how processes leading to concerted evolution have failed both before and after the diversification of XX/XY₁Y₂ species (Navajas-Pérez et al. 2005b, 2009b; this paper).

In summary, the study of the RAE180 satellite DNA in *Rumex* reveals a complex evolutionary pattern where variability can remain by reduced action of molecular mechanisms of genomic turnover, thus leading to sequence variants persisting as a “library” (Meštrović et al. 1998, 2006; Mravinac et al. 2002). From this library, the amplification of particular variants could give rise later to new enlarged arrays in different lineages. Differential levels of RAE180 satellite DNA amplification in different lineages, at different evolutionary times, and in different chromosomal places gave rise to differential patterns of sequence evolution in which the onset of concerted evolution depended on factors such as location, organization, and copy number of the repeats (Navajas-Pérez et al. 2009b and references herein; this paper).

Acknowledgements

This research was supported by grant CGL2006-00444/BOS of the Ministerio of Ciencia e Innovación (Spain) and Fondo Europeo de Desarrollo Regional (FEDER, EU).

References

- Bennett, M.D., and Leitch, I.J. 2004. Plant DNA C-values database Release 3.0.
- Cuñado, N., Navajas-Pérez, R., de la Herrán, R., Ruiz Rejón, C., Ruiz Rejón, M., Santos, J.L., and Garrido-Ramos, M.A. 2007. The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. *Chromosome Res.* **15**(7): 825–833. doi:10.1007/s10577-007-1166-6. PMID:17899410.
- de la Herrán, R., Rejon, C.R., Rejon, M.R., and Garrido-Ramos, M.A. 2001. The molecular phylogeny of the Sparidae (Pisces, Perciformes) based on two satellite DNA families. *Heredity*, **87**(6): 691–697. doi:10.1046/j.1365-2540.2001.00967.x. PMID:11903564.
- Degraeve, N. 1976. Contribution à l'étude cytotonomique des *Rumex*. IV Le genre *Acetosa* Mill. *Cellule*, **71**: 231–240.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**(4): 783–791. doi:10.2307/2408678.
- Fry, K., and Salser, W. 1977. Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*, **12**(4): 1069–1084. doi:10.1016/0092-8674(77)90170-2. PMID:597857.
- Gaut, B.S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evol. Biol.* **30**: 93–120.
- Mariotti, B., Navajas-Pérez, R., Lozano, R., Parker, J.S., de la Herrán, R., Rejón, C.R., 2006. Cloning and characterization of dispersed repetitive DNA derived from microdissected sex chromosomes of *Rumex acetosa*. *Genome*, **49**(2): 114–121. doi:10.1139/G05-089. PMID:16498461.
- Mariotti, B., Manzano, S., Kejnovský, E., Vyskot, B., and Jamilena, M. 2009. Accumulation of Y-specific satellite DNAs during the evolution of *Rumex acetosa* sex chromosomes. *Mol. Genet. Genomics*, **281**(3): 249–259. doi:10.1007/s00438-008-0405-7. PMID:19085011.
- Meštrović, N., Plohl, M., Mravinac, B., and Ugarković, Đ. 1998. Evolution of satellite DNAs from the genus *Palorus* — experimental evidence for the “library” hypothesis. *Mol. Biol. Evol.* **15** (8): 1062–1068. PMID:9718733.
- Meštrović, N., Castagnone-Sereno, P., and Plohl, M. 2006. Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. *Mol. Biol. Evol.* **23**(12): 2316–2325. doi:10.1093/molbev/msl119. PMID:16982821.
- Mravinac, B., Plohl, M., Meštrović, N., and Ugarković, Đ. 2002. Sequence of PRAT satellite DNA “frozen” in some Coleopteran species. *J. Mol. Evol.* **54**(6): 774–783. doi:10.1007/s0023901-0079-9. PMID:12029359.
- Navajas-Pérez, R., de la Herrán, R., López González, G., Jamilena, M., Lozano, R., Ruiz Rejón, C., 2005a. The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. *Mol. Biol. Evol.* **22**(9): 1929–1939. doi:10.1093/molbev/msi186. PMID:15944442.
- Navajas-Pérez, R., de la Herrán, R., Jamilena, M., Lozano, R., Rejón, C.R., Rejón, M.R., and Garrido-Ramos, M.A. 2005b. Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). *J. Mol. Evol.* **60**(3): 391–399. doi:10.1007/s00239-004-0199-0. PMID:15871049.
- Navajas-Pérez, R., Schwarzacher, T., de la Herrán, R., Ruiz Rejón, C., Ruiz Rejón, M., and Garrido-Ramos, M.A. 2006. The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene*, **368**: 61–71. doi:10.1016/j.gene.2005.10.013. PMID:16324803.
- Navajas-Pérez, R., Rubio-Escudero, C., Aznarte, J.L., Rejón, M.R., and Garrido-Ramos, M.A. 2007. SatDNA Analyzer: a computing tool for satellite-DNA evolutionary analysis. *Bioinformatics*, **23** (6): 767–768. doi:10.1093/bioinformatics/btm005. PMID:17242030.
- Navajas-Pérez, R., Schwarzacher, T., Rejón, M.R., and Garrido-Ramos, M.A. 2009a. Molecular cytogenetic characterization of *Rumex papillaris*, a dioecious plant with an XX/XY₁Y₂ sex chromosome system. *Genetica*, **135**(1): 87–93. doi:10.1007/s10709-008-9261-y. PMID:18373205.
- Navajas-Pérez, R., Quesada del Bosque, M.E., and Garrido-Ramos, M.A. 2009b. Effect of location, organization, and repeat-copy number in satellite-DNA evolution. *Mol. Genet. Genomics*, **282** (4): 395–406. doi:10.1007/s00438-009-0472-4. PMID:19653004.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Posada, D., and Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**(9): 817–818. doi:10.1093/bioinformatics/14.9.817. PMID:9918953.
- Shibata, F., Hizume, M., and Kuroki, Y. 1999. Chromosome painting of Y chromosomes and isolation of a Y chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. *Chromosoma*, **108**(4): 266–270. doi:10.1007/s004120050377. PMID:10460415.
- Shibata, F., Hizume, M., and Kuroki, Y. 2000. Differentiation and the polymorphic nature of the Y chromosomes revealed by repetitive sequences in the dioecious plant, *Rumex acetosa*. *Chromosome Res.* **8**(3): 229–236. doi:10.1023/A:1009252913344. PMID:10841050.
- Smith, B.W. 1963. The mechanism of sex determination in *Rumex hastatulus*. *Genetics*, **48**(10): 1265–1288. PMID:17248184.
- Smith, B.W. 1969. Evolution of sex-determining mechanisms in *Rumex*. *Chromosomes Today*, **2**: 172–182.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software

- version 4.0. *Mol. Biol. Evol.* **24**(8): 1596–1599. doi:10.1093/molbev/msm092. PMID:17488738.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**(24): 4876–4882. doi:10.1093/nar/25.24.4876.
- Wilby, A.S., and Parker, J.S. 1988. Recurrent patterns of chromosome variation in a species group. *Heredity*, **61**(1): 55–62. doi:10.1038/hdy.1988.90.

Quesada del Bosque ME, López-Flores I, Suárez-Santiago VN, Garrido-Ramos MA. Differential spreading of HinfI satellite-DNA variants during radiation in Centaureinae. *Annals of Botany*, 2013; revisión enviada.

Differential spreading of HinfI satellite-DNA variants during radiation in Centaureinae

María Ester Quesada del Bosque, Inmaculada López-Flores, Víctor N. Suárez-Santiago*

and Manuel A. Garrido-Ramos

Departamentos de Genética y de Botánica*, Facultad de Ciencias, Universidad de Granada,
18071 Granada, Spain.

Correspondence:

Dr. Manuel A. Garrido-Ramos
Departamento de Genética
Facultad de Ciencias
Universidad de Granada
Avda. Fuentenueva s/n, 18071
Granada, Spain
Phone number: 958249710
Fax number: 958244073
e-mail: mgarrido@ugr.es

Abstract

- *Background and Aims:* The subtribe Centaureinae appears to be an excellent model group to analyze satellite DNA and assess the influence that the biology and/or the evolution of its different lineages have had in the evolution of this class of repetitive DNA. Phylogenetic analyses of Centaureinae support two main phases of radiation, leading to two major groups of genera of different ages. Furthermore, different modes of evolution are observed in different lineages reflected by morphology as well as DNA sequences.

- *Methods:* We have isolated and compared the sequences of 502 repeat units of the HinfI satellite-DNA family which were isolated from 38 species belonging to 10 genera of Centaureinae. A phylogenetic reconstruction was made by Maximum likelihood and Bayesian inference.

- *Key Results:* We have found up to eight different HinfI subfamilies based on the presence of a set of diagnostic positions given by a specific mutation shared by all the sequences of one group. Subfamilies V to VIII were found prevalingly in older genera (first phase of radiation within the subtribe, late Oligocene-Miocene), although some copies of these type of repeats were also found in some species of the derived genera. Subfamilies I to IV spread mostly in species of the derived clade (second phase of radiation, Pliocene to Pleistocene), although repeats of these subfamilies exist in older species. Phylogenetic trees did not group the repeats by taxonomic affinity but sequences were grouped by subfamily provenance. Concerted evolution was observed in HinfI subfamilies spread in older genera whereas no genetic differentiation was found between species, and several subfamilies even coexist within the same species, in groups of recent radiation or in groups with a history of recurrent hybridization of lineages.

- *Conclusions:* Our results suggest that the eight HinfI subfamilies were present in the common ancestor of Centaureinae and each one did spread differentially in different genera during the two main phases of radiation following the library model of satellite DNA evolution. Additionally, differential speciation pathways gave rise to differential patterns of sequence evolution in different lineages. Thus, the evolutionary history of each Centaureinae group is reflected in HinfI satellite DNA evolution. Our data reinforce the value of satellite-DNA sequences as markers of evolutionary processes occurring in biological groups.

Keywords: satellite DNA, concerted evolution, molecular drive, library hypothesis, Centaureinae, radiation.

Introduction

There are four tribes within the subfamily Carduoideae (Family Asteraceae), the most abundant and widespread being the tribe Cardueae. This tribe represents 90% of species diversity of the group, with 2400 species belonging to 72 genera (Susanna and Garcia-Jacas, 2009). Current classifications of the tribe accept five subtribes: Cardopatiinae, Carlininae, Echinopsinae, Carduinae, and Centaureinae (Susanna and Garcia-Jacas, 2009). Cardopatiinae, Carlininae and Echinopsinae form basal groups and are less represented in species number, while Carduinae and Centaureinae are more diversified derived groups (Susanna and Garcia-Jacas, 2009). The family Asteraceae in general, and the subtribe Centaureinae in particular, provides an excellent opportunity to understand adaptation in the recent radiation of a plant group at a global scale (Panero and Funk, 2008). The subtribe Centaureinae comprises 32 genera with over 800 species distributed in the Mediterranean-Oriental-Turkestanian region with representatives in tropical regions of Africa, North and South America, East Asia, and Australia (Susanna and Garcia-Jacas, 2007). Notwithstanding, the Mediterranean basin is the main center of diversity of the subtribe (Wagenitz, 1986; Meusel and Jäger, 1992). Phylogenetic analyses of Centaureinae support two main phases of radiation, leading to two major groups of genera of different ages, one which includes the basal genera of the subtribe and the other includes the derived ones (Garcia-Jacas *et al.*, 2001; Hellwig, 2004). The first phase may date to the late Oligocene and Miocene, while the major modern or derived clades appear to have differentiated around the transition from Pliocene to Pleistocene (Hellwig, 2004). The derived clade, the second radiation, encompasses most of the species currently classified in the genus *Centaurea* and the *Carthamus* complex. The genus *Centaurea* comprises three subgenera: *Acrocentron*, *Centaurea* (formerly the *Jacea* group), and *Cyanus* (Susanna and Garcia-Jacas, 2009). The *Carthamus* complex includes the genera *Carduncellus*, *Carthamus*, *Femeniasia*, and *Phonus* (Garcia-Jacas *et al.*, 2001). The rest of genera constitute the basal groups (first radiation): *Cheirolophus*, *Crupina*, *Klasea*, *Psephellus*, *Rhaponticoides* (formerly *Centaurea sensu stricto*), *Rhaponticum*, *Serratula*, or *Volutaria*, among others (Garcia-Jacas *et al.*, 2001). Both groups (basal and derived) have recent members in the Mediterranean but only the derived younger genera are present with many species (Hellwig, 2004). Different modes of evolution are observed in different lineages reflected by morphology as well as DNA sequences. Hellwig (2004) described the ecogeographical history of Centaureinae (Asteraceae) in the Mediterranean on the background of the geological history of the region. According to Hellwig (2004), the representatives of the different clades often grew sympatrically but taking

different evolutionary routes. The high numbers of different ecological niches, the fine mosaic-like habitats of different qualities are the basis for the coexistence of many species which have evolved within the Centaureinae (Hellwig, 2004). Expansion and reduction of suitable habitats led to the evolution of groups consisting of many now largely allopatric vicariant species. While in several groups of perennial species slow geographic speciation, often preceded by clinal morphological variation, and extended reticulation, predominate, other groups, containing mostly annuals, produced many reproductively isolated but sometimes sympatric species (Hellwig, 2004).

With this taken into account, the subtribe Centaureinae appears to be an excellent group model to analyze satellite DNA and the influence that the biology and/or the evolution of its different lineages have had in the evolution of this class of repetitive DNA.

Satellite-DNA sequences are families of short, highly repetitive sequences found at heterochromatin mainly in pericentromeric and subtelomeric regions of the eukaryote chromosomes (López-Flores and Garrido-Ramos, 2012; Plohl *et al.*, 2012). Monomer sequences of a satellite DNA evolve concertedly through a process of molecular drive in which mutations are homogenized in a genome and fixed in a population at a higher rate than the one at which they arise. This process results in rapid divergence of satellite sequences in reproductively isolated groups of organisms (Plohl *et al.*, 2012). Nevertheless, the overall variability profile of satellite-DNA monomers in a genome is a complex feature that depends on genomic conservation and divergence of satellite DNAs, distribution, and homogenization patterns among variants, putative selective constraints imposed on them, as well as on the reproduction mode and population factors (Plohl *et al.*, 2010, 2012). Therefore, concerted evolution might be slowed due to satellite-DNA location, organization, and repeat-copy number (Navajas-Pérez *et al.*, 2005, 2009), to functional constraints (Mravinac *et al.*, 2005) or to biological factors (Luchetti *et al.*, 2003, 2006; Robles *et al.*, 2004; Suárez-Santiago *et al.*, 2007a). In the absence of selective and biological constraints, the rate of concerted evolution of a family of satellite-DNA sequences should thus depend basically on the divergence time between species (Pérez-Gutiérrez *et al.*, 2012). However, the high rate of sequence changes in the course of concerted evolution is not the only possible cause that may explain diversity among dominant satellite DNAs in related species (Plohl *et al.*, 2012). In addition to the large potential of the sequence change, satellite DNAs are permanently altered in copy number by expanding and contracting arrays of satellite monomers. Since several satellite-DNA families or subfamily variants of one family may coexist in a genome, the evolution of species-specific

satellite-DNA composition can be directed by copy-number changes within a library of satellite sequences common for a group of species (Plohl *et al.*, 2012). According to the library hypothesis of satellite-DNA evolution, the occurrence of a species-specific profile of satellite DNAs result from differential amplifications and/or contractions within a pool of sequences shared by related genomes (Fry and Salser, 1977). The library of satellite DNAs represents a permanent source of sequences that can be independently amplified in each species into a dominant satellite DNA rapidly changing any profile of genomic satellite DNA (Plohl *et al.*, 2012).

The HinfI satellite-DNA family was isolated from the genomes of several species of the genera *Centaurea*, *Phonus*, and *Carthamus* (Suárez-Santiago *et al.*, 2007a). In *Centaurea*, the absence of concerted evolution was visualized by similar levels of intraspecific variation and interspecific divergence and by the lack of fixed species-diagnostic nucleotide sites as well as by the coexistence of several subfamilies within the same species, all of these features as a reflecting of the reticulate mode of evolution in this genus. Taking this into account and having found that this satellite DNA is conserved throughout the entire subtribe, we extend our previous study to analyze here the HinfI repeats in the rest of Centaureinae in order not only to check the different hypotheses on satellite-DNA evolution within and between the two main radiation rounds but also to look for the value of these types of sequences as markers of the different evolutionary processes in this highly diversified group.

Materials and Methods

We analyzed a total of 502 HinfI repeat units isolated from 38 species belonging to 10 genera of the subtribe Centaureinae: *Centaureae*, *Carthamus*, *Phonus*, *Carduncellus*, *Rhaponticoides*, *Rhaponticum*, *Klasea*, *Crupina*, *Cheirolophus*, and *Volutaria* (Table 1). We have included a few representative sequences (55) previously analyzed by us in Suárez-Santiago *et al.* (2007a) belonging to six species of *Centaurea* (*Centaurea granatensis*, *C. alba subsp. alba*, *C. corymbosa*, *C. boissieri subsp. willkommii*, *C. delicatula*, and *C. jaennensis*) and to *Phonus arborescens*. The mean number of selected sequences (representative of each of three HinfI satellite subfamilies characterized) was eight sequences. The rest of repeat sequences (447) were determined for the first time for this work and belong to additional species of *Centaurea* and to species of the other nine genera. The mean number of HinfI sequences studied from each new species analyzed was 14 and we never analyzed less than 10 repeats per species in

these cases (Table 1). An exception was *Crupina crupinastrum*, for which we analyzed 29 repeats because we studied two populations of this species, one formerly considered *C. matae*, a synonymy of *C. crupinastrum*. The EMBL accession numbers for the 55 sequences selected from our previous study in Suárez-Santiago *et al.* (2007a) were AM712738-AM712751, AM712761-AM712774, AM712803-AM712809, AM712815-AM712821, and AM712828-AM712840. The EMBL accession numbers for the rest of 447 sequences analyzed in this study for the first time were HF571538-HF571985.

Total genomic DNA was extracted from dried leaves from the Herbarium of the University of Granada (Spain) using the DNAeasy Plant Mini Kit (Qiagen). To amplify the monomeric units of the HinfI satellite DNA, we used the primer pairs CenHinf1 and CenHinf2: 5'-GCTTCGTTTTGATAGTTCGTGG-3' and 5'-TAACTTTTGCTACGGGAGTCCG-3'. These primers were internal primers, oriented oppositely with respect to each other, and designed from the most conserved region of HinfI satellite DNA sequences (see below). The location of these primers is indicated in Fig. 1. As satellite-DNA sequences are tandemly arranged, we expected to obtain complete monomer and multimer sequences flanked by partial monomer sequences that are excluded from the analysis. This procedure enables us to avoid the ends of the amplified product. PCR amplifications were performed in a GeneAmp 2700 Applied Biosystems Thermocycler. PCR reactions were performed in a volume of 50 µl of mix containing 10 mM 10xPCR buffer, 2.5 mM MgCl₂, 200 µM of a mix of each dNTP, 1 µM of each primer, 2.5 units of Taq DNA polymerase (Amersham Biosciences) and 10 ng of DNA. The PCR conditions were 40 cycles of 94°C for 60 sec, 50°C for 60 sec, and 72°C for 60 sec. After these 40 cycles, there was a 10-min extension period at 72 °C. Amplification was detected on 1.5% agarose gels stained with SYBR safe®. The PCR products were excised from the gels and the DNA purified using the GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences). The purified PCR products were ligated into the pSC-A-amp/kan vector of the StrataClone PCR Cloning Kit (StrataGene) and cloned in StrataClone SoloPack competent cells (StrataGene) following the manufacturer's recommendations. The transformed colonies were transferred onto a fresh plate and in a 10 µl of a PCR reaction mixture to verify their transformation. Positive clones were purified using a commercial kit (Dominion, Spain).

Automated sequencing of the plasmid clones was performed in both directions using the generic primers M13F and M13R. Thermal-cycling reactions were performed using the ABI Prism® Big Dye™ Terminator Cycle Sequencing Kit (Applied Biosystems). Sequencing was carried out on an ABI PRISM® 3100-Avant Genetic Analyzer (Applied Biosystems).

Multiple-sequence alignments were performed using ClustalX (Higgins and Sharp, 1988). Maximum likelihood (ML) and Bayesian analysis were used to infer phylogeny. For selection of the DNA substitution model we used MEGA v.5 (Tamura *et al.*, 2011). The model with the lowest AIC (Akaike Information Criterion) score was chosen (GTR+G). The ML analysis was conducted in MEGA v.5 (Tamura *et al.*, 2011). Bootstrap support values were calculated on 1,000 replicates (Felsenstein, 1985). Bayesian posterior probabilities (PPs) were estimated using MrBayes v3.2 (Ronquist *et al.*, 2012). The extent of rate variation across sites for individual data partitions was estimated by the shape parameter of the gamma distribution (s). Each Bayesian analysis was initiated with random starting trees and was run for one million generations with sampling frequency of trees set at the 500th iteration. For all analyses, the variance of split sequences was 0.01. The potential scale reduction factor (PSRF) was close to 1.0 for all of the substitution model parameters. The fraction of sampled values discarded as burn-in was set at 0.25. Posterior probabilities of 0.95-1.00 were considered statistically significant. Trees were displayed with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

The distribution of nucleotide variability throughout the monomer sequences was determined using the sliding window command (window length = 1, step size = 1) implemented in DNA polymorphism and DNA divergence between populations options of DnaSP v.5 (Librado and Rozas, 2009). The conserved and variable regions were considered significant when P_i exceeded the average $P_i \pm 2 \times SD$ (standard deviation), as described in Mravinac *et al.* (2004). In addition, conserved regions along the HinfI repeats were determined by the Conserved DNA Region option (Minimum window length = 19, Conservation threshold = 0,11) of the DnaSP v.5 program.

Sequence analysis revealed the existence of up to eight different HinfI satellite-DNA subfamilies amplified by the primers CenHinf1 and CenHinf2. These subfamilies were differentially represented in different species. In order to re-check the distribution of these subfamilies, we designed four additional primer pairs for the specific amplification of repeats of subfamilies I, III, V and VI, and probed them in a set of representative species. The sequences of these primers were chosen from specific regions defining the subfamily type and their location is indicated in Fig. 1. To amplify the monomeric units of the subfamily I of the HinfI satellite DNA, we used the primer pairs CentOne-F and CentOne-R (5'-GATTTTGCACCCAACAATAT-3' and 5'-TCCATCTATTCAAACGCGC-3'), for those of the subfamily III, we used the primer pairs CentThree-F and CentThree-R (5'-GATTTTGGAAATTAACAATT-3' and 5'-AGCCGAAGAAGGCAATAACT-3'), for those of subfamily V, the primers CentFive-F and

CentFive-R (5'-CAACGCAACGTTATCCGTCCGC-3' and 5'-TTTCTCGTCCGGACCGACCGTG-3') were used, and for the subfamily VI, the primers CentSix-F and CentSix-R (5'-CCGTGTAAAGTATCATGGAA-3' and 5'-TAGCGAGACGGTCGGTTT-3'). Presence/absence analysis of each subfamily in each analyzed species was achieved by gel electrophoresis on 1.5% agarose gels stained with SYBR safe®. Amplified products were sequenced to verify their subfamily provenance.

Results

The primer pairs CenHinf1 and CenHinf2 were used for the amplification of HinfI repeats from the genomes of a total of 38 species, the PCR products were cloned and a total of 502 HinfI cloned repeats were sequenced. These repeats were ascribed to eight monomer types or subfamilies. These subfamilies were established according to a set of diagnostic positions given by a specific mutation shared by all the sequences of one group (Fig. 1). They were designated with Roman numerals from I to VIII following the nomenclature previously used in Suárez-Santiago *et al.* (2007a) for three of them (subfamilies I, II and III). Additionally, the different types of sequences have diagnostic deletions found at different positions in the HinfI repeats. In order to study diversity distribution along the repeat sequences, we performed a sliding window analysis using a window length of 10 and step size 1 (see Supplementary Material Sm_1). Windows that exhibit diversity \leq (average +2 SD) were defined as variable, and those with diversity \geq (average -2 SD) were considered as conserved. The analysis reveals one conserved segment from positions 1 to 50 resulting from the overlapping of the neighboring windows. Conserved DNA Region option (Minimum window length = 19, Conservation threshold = 0,11) of the DnaSP v.5 program identified the same region, from sites 1 to 48, as a significant conserved region. This region coincides with that from which the primers were designed (Fig. 1). The rest of the repeat sequences was variable and included the diagnostic positions defining each subfamily.

Table 2 shows the distribution of HinfI subfamilies among species. The species of *Centaurea* and *Rhaponticoides* were characterized by the presence in their genomes of the HinfI sequences belonging to subfamilies I, II, and III, even with sequences of two or the three subfamilies coexisting within the same species. In the case of *Rhaponticoides*, in both species analyzed, sequences belonged either to subfamily I or to subfamily II, with one sequence of *R.*

linaresii belonging to subfamily III. In the case of *Centaurea*, the subfamilies I and II were within all species of the subgenus *Centaurea*, with the presence of subfamily III in two species. In the species analyzed of the subgenus *Cyanus* (*C. cyanus*), we found eight out of 13 repeats belonging to subfamily II although subfamilies I (four repeats) and III (one repeat) were also found. In addition, one of the sequences isolated from this species belonged to subfamily VII. In the case of the subgenus *Acrocentron*, *C. clementei*, we did not find HinfI type III sequences, while subfamily I appeared absent from *C. granatensis*. The genomes of the two species analyzed from the genus *Crupina* had sequences of subfamilies I, II, and III and, additionally, we found up to two 6 (out of 13) repeats of subfamily VII in *C. vulgaris* and 3 repeats (out of 29) of subfamilies VI (one) and VII (two) in *C. crupinastrum* (Tables 1 and 2). In the species of the genera *Carthamus* and *Phonus*, we found only HinfI type-III sequences. Subfamily IV is characteristic of the genus *Carduncellus*. Nevertheless, some species of this genus also have low-copy repeats of subfamilies III and VI in their genomes. Subfamily V was found in *Rhaponticum* and *Klasea*. These species had only this type of HinfI repeat, except in the case of *R. acaule* which also contained sequences of subfamilies II, IV, and VI in its genome. We analyzed four species of the genus *Volutaria* (Table 1). The HinfI sequences of *V. lippi* belonged to the subfamilies I, II, III, or V, mainly to type III (nine sequences out of 14). In addition, most HinfI sequences of *V. crupinoides* were of subfamilies II, III, and IV. However, most of the sequences of *V. muricata* and *V. tubuliflora* together three of *V. crupinoides* belonged to subfamily VI. Three species of the genus *Cheirolophus* (*C. intybaceus*, *C. sempervirens*, and *C. sventenii*) shared two types of HinfI monomers in their genomes, those of the subfamilies VII and VIII. *C. sventenii* also had a type-II repeat. The species *C. falcisectus* had only sequences of subfamily VII. In contrast, among the sequences isolated from *C. teydis*, 12 belonged to subfamily III and the other three to subfamily II (Table 2). We tested the reliability of CenHinf1 and CenHinf2 primers to detect the eight subfamilies in these species. For this, we designed four additional primer pairs for the specific amplification of repeats of subfamilies I, III, V and VI, and probed them in a set of representative species. We obtained similar results to those obtained with the general primers. Notwithstanding, we found some differences (Table 2 and Supplementary Material Sm_2). For example, the subfamily I is present in *Centaurea granatensis* and *Crupina vulgaris* (the only species of *Centaurea* and *Crupina* for which we did not obtain repeats of subfamily I after sequencing of amplified product using the primers CenHinf1 and CenHinf2). The subfamily III is found in species of *Volutaria* and *Cheirolophus* other than those previously detected by the "general" primers. However, the primers used were not able to amplify repeats of the subfamily III in *Carduncellus* (even in the event that we

detected a few repeats of this type using the CenHinf1 and CenHinf2 primers). Subfamily V is found not only in *Klasea* and *Rhaponticum* but also in *Volutaria*, like we previously found by sequencing CenHinf1/CenHinf2 amplified product. The subfamily VI was found only in *Volutaria* as detected by CenHinf1 and CenHinf2 primers.

Phylogenetic reconstruction by Maximum likelihood and Bayesian inference resulted in similar trees. A Bayesian majority-rule consensus tree with Bayesian PPs is shown in Supplementary Material Sm_3. A simplified tree layout without tip labels is shown in Fig. 2. The trees did not group sequences by specific or generic affinity. Although the tree grouped together the HinfI monomers of some particular genera or subtribe group, the most common view is that the repeats are grouped by HinfI subfamilies, which are conserved among different species, as classified in Fig. 1 and Table 2. The tree represented in Fig. 2 has indications of the correspondence between clades and HinfI subfamilies. We can distinguish two main clades. The first clade includes five subclades, each one corresponding to each of the subfamilies I to V. The second clade includes three subclades, one for each of the subfamilies VI to VIII.

Within each clade, sequences of subfamilies I, II or III of the different species are intermingled in a species-independent manner (Sm_3). However, it bears noting that most sequences of subfamily III belonging to *Phonus* and *Carthamus* species tend to be grouped by taxonomic affinity, on one hand the sequences of *Phonus arborescens* and, on the other, the sequences of *Carthamus tinctorius* and those of *Carthamus lanatus*, although there are some intermixed sequences from each one (Sm_3). By contrast, comparisons of subfamily III sequences of these species and low-copy counterparts of subfamily III in other species examined show high interspecific sequence conservation and the complete lack of any species-diagnostic mutations and therefore they appear to be intermixed within subfamily III clade (Sm_3). HinfI sequences of *Carduncellus* (subfamily IV) appear intermingled without separation by specific affinity (Sm_3). In the case of *Rhaponticum* and *Klasea* sequences tend to be grouped by specific affinity (Sm_3). Sequences of subfamily VI of *Volutaria* are separated according to specific precedence (Sm_3). However, the sequences of the two different subfamilies found in *Cheirolophus* (VII and VIII) are not grouped in phylogenetic trees by specific affinity but they appear intermixed (Sm_3).

Discussion

The HinfI sequences have been found to be present within the genomes of all the species analyzed of the subtribe Centaureinae. These species are representative of the entire range of groups defined within this subtribe (Garcia-Jacas *et al.*, 2001; Hellwig, 2004). The first phase of radiation of the subtribe might date to the late Oligocene and Miocene. Therefore, the HinfI satellite DNA would date to at least 28-23 Myr ago (Garcia-Jacas *et al.*, 2001; Hellwig, 2004). This is not very common among satellite-DNA families, specifically in plants, the most ancient found exceptionally in cycads (Cafasso *et al.*, 2003).

We identified up to eight HinfI subfamilies. Our results suggest that the eight HinfI subfamilies were present in the common ancestor of Centaureinae, each one spreading differentially in different genera. The differential spreading accompanied to the two main phases of radiation leading to two major groups in the Centaureinae (Garcia-Jacas *et al.*, 2001; Hellwig, 2004). Thus, subfamilies V to VIII are found prevalingly in older genera (first phase of radiation within the subtribe, late Oligocene-Miocene). Notwithstanding, a few repeats of subfamilies VI and VII were isolated from *Carduncellus* and *Centaurea* (derived clade). Subfamilies I to IV have expanded predominantly in the genomes of species belonging to the derived clade of Centaureinae (second phase of radiation, Pliocene to Pleistocene). Notably, there are several species of the basal groups having subfamilies I-III as the major representatives of HinfI sequences within their genomes. These data suggest that subfamilies I to IV have expanded recently, replacing other subfamilies in derived genera but also in older genera.

The replacement of one sequence variant by another in different species is a common feature of satellite-DNA that could be a consequence of the dynamics of satellite-DNA evolution (Plohl *et al.*, 2010, 2012). Molecular mechanisms of non-reciprocal exchange (unequal crossing-over, gene conversion, rolling-circle replication and re-insertion, and transposon-mediated exchange) would spread new sequence variants appearing in individual repeat units of a family of sequences and the changes are fixed in a population of randomly mating individuals by sexual reproduction according to a time-dependent two-step process called molecular drive, which leads to concerted evolution (Plohl *et al.*, 2010, 2012; Pérez-Gutiérrez *et al.*, 2012). Satellite DNA changes due to gradual accumulation of sequence divergence which results in divergence of satellite sequences in reproductively isolated groups of organisms (Plohl *et al.*, 2012). However, given the scattering of subfamilies in each Centaureinae lineage, an alternative but not mutually exclusive hypothesis could explain this differential distribution.

According to the “library” hypothesis, related taxa share a library of different conserved satellite-DNA sequences (different satellite-DNA families but also monomer variants or subfamilies of a satellite-DNA family), which may be differentially amplified in each taxa. Variability can remain for long evolutionary periods by reduced action of molecular mechanisms of non-reciprocal exchange and sequence variants persist as a library (Mravinac *et al.*, 2002; Mestrovic *et al.*, 2006) from which any of them may be differentially amplified in each taxon with the subsequent replacement of one sequence variant by another in different species. When this occurs, the study of unrelated species-specific dominant satellite-DNA repeats reveals the presence of low-copy counterparts of each of them in other examined species and comparisons of high-copy and low-copy monomer variants of these satellites show high interspecific sequence conservation and the complete lack of any species-diagnostic mutations, as found in *Palorus* (Meštrović *et al.*, 1998). This hypothesis has been proved in insects (Meštrović *et al.*, 1998; Mravinac *et al.*, 2002; Cesari *et al.*, 2003; Pons *et al.*, 2004) and plants (Navajas-Pérez *et al.*, 2009; Quesada del Bosque *et al.*, 2011), and could explain the main observation made in Centaureinae concerning the scattering of HinfI types. Variation in satellite profiles is found in this case by changes in copy number (Plohl *et al.*, 2012).

However, the overall variability profile of satellite-DNA monomers in a genome is a complex feature that depends on several factors such as location, organization, and repeat-copy number (Navajas-Pérez *et al.*, 2005, 2009), time (Pérez-Gutiérrez *et al.*, 2012), biological factors (Luchetti *et al.*, 2003, 2006; Robles *et al.*, 2004; Suárez-Santiago *et al.*, 2007a) or functional constraints (Mravinac *et al.*, 2005). Some patterns of HinfI repeat evolution in specific lineages of Centaureinae might result from the influence of some of these factors, discussed below.

Derived clade

According to the phylogeny of the subtribe Centaureinae (Garcia-Jacas *et al.*, 2001), within the derived clade, the *Carthamus* complex occupies the most basal position, while the subgenera *Jacea* and *Cyanus* of *Centaurea* show the most derived position, for which a sister relationship is firmly established (Susanna and Garcia-Jacas, 2009). The subgenus *Acrocentron* of *Centaurea* occupies an intermediate position, although connections between this subgenus and *Jacea* and *Cyanus* remain unclear (Garcia-Jacas *et al.*, 2001; Susanna and Garcia-Jacas, 2009).

We can emphasize two main observations with respect to HinfI sequences spreading and evolution in this group. First, there are four HinfI subfamilies (I to IV) that have spread through the genomes of these species accompanying the major speciation processes. However, the four repeat subfamilies are not equally distributed in all the taxa analyzed (Table 2). Subfamilies I and II should have spread secondarily as specific subfamilies of the genus *Centaurea*, almost completely replacing subfamily III and completely replacing subfamily IV, while subfamily III expanded in *Phonus* and *Carthamus*, and subfamily IV expanded in *Carduncellus*. Nevertheless, the four subfamilies are found in species belonging to the basal clade (first radiation). In addition to the presence of low-copy repeats of these four subfamilies in some species of the basal groups, in some of these species (*Cheirolophus teydis*, *Volutaria lippi*, species of the genus *Crupina*) these types are the main components of HinfI sequences within their genomes. These findings support the library hypothesis, as discussed above. In this sense, these subfamilies might be as old as the rest of the subfamilies studied in this paper, differentially expanding in different lineages. Regardless of phylogeny, in agreement with the library hypothesis, there is a convergence in the spreading of some subfamily variants in different lineages. It should be remembered that the hypothesis predicts neither whether any of the sequences of the library can be amplified into a major satellite family/subfamily nor whether there is selective pressure favoring some sequences or the amplification mechanism involved (Fry and Salser, 1977).

Second, several comments on concerted evolution should be emphasized. Differential speciation pathways gave rise to differential patterns of sequence evolution in different lineages. Different subfamilies coexist within most of the taxa analyzed in the genus *Centaurea*. The presence of different subfamilies within their genomes was explained as a result of reticulate evolution within a part of this genus (Suárez-Santiago *et al.*, 2007a,b). Hybridization should be a process that maintains different HinfI satellite-DNA subfamilies within one particular genome. In addition, gene flow between taxa should reduce the amount of genetic differences between those taxa but should increase the amount of intraspecific variation. Thus, under this evolutionary scenario, contrary to the expectations on the concerted-evolution model, we should find similar or even higher levels of intraspecific variation than the values of interspecific divergence. Additionally, the library hypothesis might explain the existence of additional copies of some subfamilies in some species. Species of *Carthamus* and *Phonus* have type III sequences and they are almost differentiated in such a way that intraspecific variation is lower than interspecific divergence, a sign of concerted

evolution. *Phonus* and *Carthamus*, formerly considered synonymous, are two separate genera (Vilatersana *et al.*, 2000). In fact, *Phonus* is closer to *Carduncellus* than to *Carthamus* and has differentiated biological traits and habitats, as well as biogeographical and evolutionary stories (Vilatersana *et al.*, 2000). *Carthamus* contains two rather different groups: *Carthamus sensu stricto*, which includes only sect. *Carthamus* and has *C. tinctorius* as the type species, and the section *Atractylis*, which has as type species to *C. lanatus* (Vilatersana *et al.*, 2005). Type III sequences of both species are differentiated in the tree of Sm_3. Concerted evolution is not found in *Carduncellus*, the species of which bear subfamily IV. Intraspecific variability is greater than interspecific divergence in this genus and HinfI sequences appear intermixed in the phylogenetic tree. Speciation in *Carduncellus* is very recent relative to other Centaureinae, and the HinfI divergence values such as those found in a different study for the ITS divergence values were very low (Vilatersana *et al.*, 2000). Failure to resolve relationships within the genus *Carduncellus* highlights the extremely low level of variation in many genera of recent origin (Nepokroeff and Sytsma, 1996, Susanna *et al.*, 1999). In fact, in the absence of selective and biological constraints, the rate of concerted evolution of a family of satellite-DNA sequences should depend basically on the divergence time between species (Pérez-Gutiérrez *et al.*, 2012).

Basal groups

Major clades of basal Centaureinae have been established but some genera could not be clearly classified in any group. Molecular phylogeny indicated that these genera constitute an old stock in the subtribe, suggesting that divergence between all these genera is old (García-Jacas *et al.*, 2001; Susanna and García-Jacas, 2009). However, recent phylogenetic analyses (Susanna *et al.*, 2011) suggested informal entities as natural groups and phylogenetic relationships have been partially resolved. The *Volutaria* group is located close to the base of the phylogenetic trees and includes the genera *Volutaria* and *Amberboa* among others (Susanna *et al.*, 2011). The crown node for the clade of basal Centaureinae received good support but relationships among the main groups in that clade remain poorly resolved, forming a general trichotomy (Susanna *et al.*, 2011). It includes a strongly supported clade of *Cheirolophus*, another equally supported clade which encompasses *Rhaponticum* (including former *Acroptilon* and *Leuzea*), *Myopordon* and *Oligochaeta*, and a third clade containing the remaining taxa. This latter clade comprises a clade which includes among others, *Klasea* and

Serratula and, finally, an unsupported polytomy should include three clades (Susanna *et al.*, 2011): the first contains the genus *Plectocephalus*, and the second contains the genus *Stizolophus*, whereas the third is a moderately supported clade, containing the genera *Crupina* and *Rhaponticoides*, among others.

Here, we also highlight two main observations with respect to HinfI sequence distribution and evolution in these species. First, there are four HinfI subfamilies (V to VIII) that have spread through the genomes of these species, accompanying the major speciation processes. On the contrary to the subfamilies I to IV, in this case, each subfamily has spread almost exclusively within one or two specific genera: subfamily V is almost exclusive of the genera *Rhaponticum* and *Klasea* (although we have detected it also in some species of *Volutaria*), subfamily VI spread in some species of *Volutaria* and subfamilies VII and VIII spread in most *Cheirolophus* species analyzed. We can conclude in this case that these spreading are old and restricted to specific lineages. The presence of some copies of these types of HinfI repeats in a few additional species again supports the library hypothesis.

Second, referring to concerted evolution, we can conclude in this case also that differential speciation pathways gave rise to differential patterns of sequence evolution in different lineages. Sequences of subfamily V evolve in concert within the *Klasea-Rhaponticum* clade (Sm_3). Sequences are in general more similar within species than between species and they tend to be grouped according species affinity, thus confirming both that the spreading of this subfamily precedes the diversification of the genus and that this speciation is old within Centaureinae. Subfamily VI of HinfI sequences is found in *Volutaria muricata* and *Volutaria tubuliflora* (as well as low-copy repeats in a few more species). The sequences of these species cluster separately, a sign of homogenization within species and differentiation between these two species (Sm_3), i.e. concerted evolution. However, sequences of subfamilies VII and VIII are distributed equally throughout all *Cheirolophus* species analyzed in this study (except in *C. teydis*, as discussed above). Within this genus, HinfI sequences of different species are more similar than when compared within species and they appear intermingled in the phylogenetic trees. As with *Carduncellus*, Susanna *et al.* (1999) found speciation in *Cheirolophus* to be very recent relative to other Centaureinae, which might be a factor influencing the absence of concerted evolution (Pérez-Gutiérrez *et al.*, 2012).

References

- Cafasso D, Cozzolino S, De Luca P, Chinali G. 2003. An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organisations of the repetitive unit in the plant genome. *Gene* **311**: 71-79.
- Cesari M, Luchetti A, Passamonti M, Scali V, Mantovani B. 2003. Polymerase chain reaction amplification of the *Bag320* satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* **312**: 289-295.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.
- Fry K, Salser W. 1977. Nucleotide sequence of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **12**: 1069-1084.
- Garcia-Jacas N, Susanna A, Garnatje T, Vilatersana R. 2001. Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Annals of Botany* **87**: 503–515.
- Hellwig FH. 2004. Centaureinae (Asteraceae) in the Mediterranean-history of ecogeographical radiation. *Plant Systematics and Evolution* **246**:137-162.
- Hidalgo O, Garcia-Jacas N, Garnatje T, Susanna A. 2006. Phylogeny of *Rhaponticum* (Asteraceae, Cardueae-Centaureinae) and related genera inferred from nuclear and chloroplast DNA sequence data: taxonomic and biogeographic implications. *Annals of Botany* **97**: 705-714.
- Higgins DG, Sharp PM. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.
- López-Flores I, Garrido-Ramos MA. 2012. The repetitive DNA content of eukaryotic genomes. In: Garrido-Ramos MA, ed. *Repetitive DNA. Genome Dynamics vol 7*. Basel: Karger, 1-28.

- Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V, Mantovani B. 2003. Unisexuality and molecular drive: *Bag320* sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *Journal of Molecular Evolution* **56**: 587-596.
- Luchetti A, Marini M, Mantovani B. 2006. Non-concerted evolution of *RET76* satellite DNA family in *Reticulitermes* taxa (Insecta, Isoptera). *Genetica* **128**: 123–132.
- Martins L, Hellwig FH. 2005. Systematic position of the genera *Serratula* and *Klasea* within Centaureinae (Cardueae, Asteraceae) inferred from ETS and ITS sequence data and new combinations in *Klasea*. *Taxon* **54**: 632-638.
- Meusel H, Jäger EJ. 1992. *Vergleichende Chorologieder zentraleuropäischen Flora* 3. Band III (Text- und Kartenteil). Jena, Stuttgart, New York: Fischer Verlag.
- Meštrović N, Castagnone-Sereno P, Plohl M. 2006. Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. *Molecular Biology and Evolution* **23**: 2316-2325.
- Meštrović N, Plohl M, Mravinac B, Ugarkovic D. 1998. Evolution of satellite DNAs from the genus *Palorus*-experimental evidence for the library hypothesis. *Molecular Biology and Evolution* **15**: 1062-1068.
- Mravinac B, Plohl M, Meštrovi N, Ugarkovic D. 2002. Sequence of PRAT satellite DNA “frozen” in some Coleopteran species. *Journal of Molecular Evolution* **54**: 774-783.
- Mravinac B, Plohl M, Ugarkovic D. 2004. Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene* **332**: 169–177.
- Mravinac B, Plohl M, Ugarkovic D. 2005. Preservation and high sequence conservation of satellite DNAs indicate functional constraints. *Journal of Molecular Evolution* **61**: 542-550.
- Navajas-Pérez R, de la Herrán R, Jamilena M, Lozano R, Rejón CR, Ruiz Rejón MR, Garrido-Ramos MA. 2005. Reduced Rates of Sequence Evolution of Y-Linked Satellite DNA in *Rumex* (Polygonaceae). *Journal of Molecular Evolution* **60**: 391-399.
- Navajas-Pérez R, Quesada del Bosque ME, Garrido-Ramos MA. 2009. Effect of location, organization and repeat-copy number in satellite-DNA evolution. *Molecular Genetics and Genomics* **282**: 395-406.

Nepokroeff M, Sytsma KJ. 1996. Systematics and patterns of speciation and colonization in Hawaiian *Psychotria* and relatives based on phylogenetic analysis of ITS sequence data. *American Journal of Botany* **83**:181-182.

Panero JL, Funk VA. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* **47**: 757-782.

Pérez-Gutiérrez MA, Suárez-Santiago VN, López-Flores I, Romero AT, Garrido-Ramos MA. 2012. Concerted evolution of satellite DNA in *Sarcocapnos*: a matter of time. *Plant Molecular Biology* **78**: 19-29.

Plohl M, Meštrović N, Mravinac B. 2012. Satellite DNA evolution. In: Garrido-Ramos MA, ed. *Repetitive DNA. Genome Dynamics vol 7*. Basel: Karger, 126-152.

Plohl M, Petrović V, Luchetti A, Ricci A, Satović E, Passamonti M, Mantovani B. 2010. Long-term conservation vs high sequence divergent: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* **104**:543–551.

Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarkovic D, Juan C. 2004. Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common 'satellite DNA library'. *Heredity* **92**: 418-427

Quesada del Bosque ME, Navajas-Pérez R, Panero JL, Fernández-González A, Garrido-Ramos MA. 2011. A satellite DNA evolutionary analysis in the North American endemic dioecious plant *Rumex hastatulus* (Polygonaceae). *Genome* **54**: 253-260.

Robles F, de la Herrán R, Ludwig A, Ruiz-Rejón C, Ruiz-Rejón M, Garrido-Ramos MA. 2004. Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* **338**:133-142.

Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höna SH, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

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

Suárez-Santiago VN, Salinas MJ, Garcia-Jacas N, Soltis PS, Douglas ES, Blanca G. 2007b. Reticulate evolution in the *Acrolophus* subgroup (*Centaurea* L., Compositae) from the western

Mediterranean: Origin and diversification of section *Willkommia* Blanca. *Molecular Phylogenetics and Evolution* **43**: 156-172.

Susanna A, Galbany-Casals M, Romaschenko K, Barres L, Martin J, Garcia-Jacas N. 2011. Lessons from *Plectocephalus* (Compositae, Cardueae-Centaureinae): ITS disorientation in annuals and Beringian dispersal as revealed by molecular analyses. *Annals of Botany* **108**: 263–277.

Susanna A, Garcia-Jacas N. 2007. Tribu Cardueae. In: Kadereit JW, Jeffrey C, eds. *Flowering plants. Eudictos. Asterales (vol. 8 in series The families and genera of vascular plants, Kadereit JW. ed.)*. Berlin, Heidelberg: Springer Verlag, 123-146.

Susanna A, Garcia-Jacas N. 2009. Cardueae (Carduoideae) In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, eds. *Systematics, Evolution, and Biogeography of the Compositae*. IAPT: Vienna, 293–313.

Susanna A, Garnatje T, Garcia-Jacas N. 1999. Molecular phylogeny of *Cheirolophus* (Asteraceae: Cardueae-Centaureinae) based on ITS sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution* **214**: 147-160.

Tamura K, Peterson D, Peterson N, Stecher G, Masatoshi N, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N. 2005. Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. *Botanical Journal of the Linnean Society* **147**: 375-383.

Vilatersana R, Susanna A, Garcia-Jacas N, Garnatje T. 2000. Generic delimitation and phylogeny of the *Carduncellus-Carthamus* complex (Asteraceae) based on ITS sequences. *Plant Systematics and Evolution* **221**: 89-105.

Wagenitz G. 1986. *Centaurea* in South-West Asia: patterns of distribution and diversity. *Proceeding of the Royal Society of Edinburgh* **89B**: 11–21.

Figure Legends

Figure 1. Sequence alignment of average sequences of the different HinfI satellite-DNA subfamilies from different genera. The first line is an average sequence (Av.s.) for the whole HinfI sequences. Only base changes to the average sequence are shown in individual genera and subfamilies. Diagnostic positions for each subfamily are shaded black. The annealing regions of the primer pairs used for the amplification of the repeats are indicated (grey).

Figure 2. Majority-rule consensus tree based on Bayesian MCMC analysis of the HinfI repeat sequences. This figure represents a simplified tree layout without tip labels indicating the correspondence between major clades and HinfI subfamilies. Bayesian posterior probability values (PP) for the main nodes are indicated. For a detailed version of this tree refer to the figure of Supplementary material (Sm_3).

Sm_1. Identification of conserved (shaded) and variable segments in HinfI satellite DNA using sliding window of 10 bp. Average diversity (Pi) is indicated with dashed line, while the average diversity ± 2 S.D. is indicated with dotted line.

Sm_2. PCR amplification of HinfI variant repeats using subfamily specific primers. a) Subfamily I and Subfamily III. From left to right: *Centaurea jaennensis*, *C. granatensis*, *C. cyanus*, *Carthamus lanatus*, *Phonus arborescens*, *Carduncellus caeruleus*, *C. cuatrecasasii*, *Crupina vulgaris*, *Rhaponticum acaule*, *R. coniferum*, *Klasea flavescens*, *Volutaria tubuliflora*, *V. lippi*, *Cheirolophus sempervirens*, *C. teydis*. b) Subfamilies V and VI. From left to right: *Centaurea granatensis*, *Carthamus lanatus*, *Phonus arborescens*, *Carduncellus caeruleus*, *Crupina vulgaris*, *Rhaponticum coniferum*, *Klasea flavescens*, *Volutaria tubuliflora*, *Cheirolophus sempervirens*. In the case of subfamily VI, a PCR fragment of about 1 Kb was amplified from the genome of *Klasea flavescens* but was an unspecific amplification.

Sm_3. Majority-rule consensus tree based on Bayesian MCMC analysis of the HinfI repeat sequences. Numbers above branches are Bayesian posterior probability values (PP).

Table 1. List of species analyzed in this paper.

Taxa	Locality/Voucher/Reference ^a	Number of repeats analyzed
<i>Centaurea</i> L.		
<i>Centaurea</i> subg. <i>Acrocentron</i>		
<i>C. clementei</i> Boiss.	SP: Cádiz, Grazalema/GDAC 5102	11
<i>C. granatensis</i> Boiss.	SP: Granada, Huétor Santillán/ GDA 46119 ^b	9
<i>C. saxifraga</i> Coincy	SP: Granada, Zújar/GDA 499332	10
<i>Centaurea</i> subg. <i>Cyanus</i>		
<i>C. cyanus</i> L.	SP: Ctra. Ávila-Salamanca/ GDAC 23553	14
<i>Centaurea</i> subg. <i>Centaurea</i>		
<i>Centaurea</i> sect. <i>Acrolophus-Phalolepis</i> (Cass.) Dostál		
<i>C. alba</i> L. subsp. <i>alba</i>	SP: Ávila, Sierra del Águila/GDAC 24948 ^b	7
<i>C. corymbosa</i> Pourret	FR: Narbonne, La Clape/BC ^b	7
<i>Centaurea</i> sect. <i>Willkommia</i> Blanca		
<i>C. boissieri</i> D.C. subsp. <i>willkommii</i> (Schultz Bip. ex Willk.) Dostál	SP: Murcia, Sierra Espuña/GDAC 6649 ^b	14
<i>C. delicatula</i> Breittw and Podlech	TN: Djebel Chambi/MPU ^b	6
<i>C. jaemensis</i> Degen and Debeaux in Degen	SP: Jaén, Pozo Alcón/GDAC 6724 ^b	8
<i>Carthamus</i> L.		
<i>C. tinctorius</i> L.	SP: Sevilla/ GDAC 13709	18
<i>C. lanatus</i> L.	SP: Córdoba, Sierra Morena, Los Patalos/GDAC 39253	15
<i>Phonus</i> Hill		
<i>P. arborescens</i> (L.) G. López	SP: Granada, Sierra de Lújar/GDA 18987 ^c	19
<i>Carduncellus</i> Adans.		
<i>C. caeruleus</i> (L.) C. Presl	SP: Málaga/ GDA 49305	12
<i>C. calvus</i> Boiss. & Reut.	MO: Beni Hosmar, Montis Borgues/GDA 3998	16
<i>C. cuatrecasasi</i> G. López	SP: Granada, Pedro Martínez/ GDA 25041	10
<i>C. dianthus</i> Webb	SP: Alicante, Mongó/GDAC 4193	14
<i>C. hispanicus</i> Boiss. ex DC.	SP: Granada, Sierra Sagra/GDAC 20480	12
<i>C. mitissimus</i> (L.) DC.	SP: León, Solana del Puerto de Ventana/ GDAC 13047	13
<i>Rhaponticoides</i> Vaill.		
<i>R. alpina</i> (L.) M.V. Agab. & Greuter*	SP: Jaén, S ^a de las Villas/ GDAC 55133	15
<i>R. linaresii</i> (Lázaro Ibiza) M.V. Agab & Greuter*	SP: Palencia, Dueñas/ GDAC 13180	13
<i>Rhaponticum</i> Vaill.		
<i>R. acaule</i> (L.) DC.	MO: Región de Monte Arruit/ GDA 5265	13
<i>R. berardioides</i> Batt.	MAR: Talasentant / GDA 4998	13
<i>R. confertum</i> (L.) DC.	SP: Granada, Pantano de Cubillas/GDA 42407	12
<i>R. centauroides</i> (L.) Holub	SP: Lérida, Valle de Arán/GDAC 24874	14
<i>R. exaltatum</i> (Willk.) Greuter	SP: Ávila, Hoyocasero/ GDAC 34999	15
<i>Klasea</i> Cass.		
<i>K. flavescens</i> (L.) Holub subsp. <i>flavescens</i>	SP: Granada, Lanjarón/ GDA 54612	11
<i>K. flavescens</i> subsp. <i>leucantha</i> (Cav.) Cantó & Rivas Mart.	SP: Granada, Cerro Gordo/ GDAC 10730	13
<i>Crapina</i> (Pers.) DC.		
<i>C. crupinastrum</i> (Moris) Vis.	SP: Granada, Alhama de Granada/GDA 54194	29
<i>C. vulgaris</i> Cass.	SP: Córdoba, S ^a Morena/ GDAC 42205	13
<i>Cheirolophus</i> Cass.		
<i>C. falcisectus</i> Montelongo & Moraleda	SP: Gran Canaria, Jardín Botánico Viera y Clavijo/GDAC 44225	16
<i>C. intybaceus</i> (Lam) Dostál	SP: Almería, Níjar/ GDA 25217	15
<i>C. sempervirens</i> (L.) Pomel	SP: Granada, S ^a de Cázulas / GDA 52606	12
<i>C. sventenii</i> (A. Santos) G. Kunkel	SP: Gran Canaria, Jardín Botánico Viera y Clavijo/ GDAC 44224	14
<i>C. teydis</i> (Buch) G. López	SP: Santa Cruz de Tenerife, Circo de Las Cañadas del Teide/ GDA 41753	15
<i>Volutaria</i> Cass.		
<i>V. crupinoides</i> (Desf.) Maire	MO: Driouch/GDAC 39797	12
<i>V. lippi</i> (L.) Maire	SP: Almería/ GDA 52154	14
<i>V. muricata</i> (L.) Maire	SP: Málaga / GDAC 42882	14
<i>V. tubuliflora</i> (Murb.) Sennen	SP: Almería, Tabernas / GDAC 43411	14

^aFR, France. MO, Morocco. SP, Spain. TN, Tunisia. GDA/GDAC, University of Granada Herbarium. BC, Barcelona herbarium. MPU, University of Montpellier Herbarium.

^bSequences previously described in Suárez-Santiago *et al.* (2007a)

^cFour repeats previously described in Suárez-Santiago *et al.* (2007a) plus 15 new repeats

*These species have been considered synonyms

Table 2. Survey of HinfI subfamilies in Centaureinae. Numbers indicate the number of sequences isolated of each subfamily from each species. The symbol + indicates that the presence of the subfamily was detected in that species by PCR amplification using subfamily specific primers. Color codes indicate the subfamily type. A gray line separates the derived clade and the basal groups.

Genus	Species	HinfI subfamily							
		I	II	III	IV	V	VI	VII	VIII
<i>Centaurea</i> *	<i>alba</i>	4	1	2					
	<i>corymbosa</i>	4	3						
	<i>boissieri</i>	6	8						
	<i>delicatula</i>	1	2	3					
	<i>jaennensis</i>	1	7						
<i>Acrocentron</i> *	<i>clementei</i>	6	5						
	<i>granatensis</i>	+	1	8					
	<i>saxifraga</i>	1	8	1					
<i>Cyanus</i> *	<i>cyanus</i>	4	8	1				1	
<i>Carthamus</i>	<i>lanatus</i>			15					
	<i>tinctorius</i>			18					
<i>Phonus</i>	<i>arborescens</i>			19					
<i>Carduncellus</i>	<i>caeruleus</i>			1	11				
	<i>mitissimus</i>			2	10		1		
	<i>cuatrecasasii</i>				10				
	<i>dianius</i>				14				
	<i>hispanicus</i>				12				
	<i>calvus</i>			4	10		2		
<i>Rhaponticoides</i>	<i>alpina</i>	9	6						
	<i>linaresii</i>	6	6	1					
<i>Crupina</i>	<i>crupinastrum</i>	9	15	2			1	2	
	<i>vulgaris</i>	+	3	4				6	
<i>Rhaponticum</i>	<i>acaule</i>		3		6	3	1		
	<i>berardioides</i>					13			
	<i>centauroides</i>					14			
	<i>coniferum</i>					12			
	<i>exaltatum</i>					15			
<i>Klasea</i>	<i>flavescens</i>					11			
	<i>leucantha</i>					13			
<i>Volutaria</i>	<i>muricata</i>		1	2			11		
	<i>tubuliflora</i>			+		+	14		
	<i>crupinoides</i>		2	3	4		3		
	<i>lippi</i>	1	3	9		1			
<i>Cheirolophus</i>	<i>falcisectus</i>							16	
	<i>intybaceus</i>							11	4
	<i>sempervirens</i>			+				5	7
	<i>sventenii</i>		1					11	2
	<i>teydis</i>		3	12					

* *Centaurea*, *Acrocentron* and *Cyanus* are subgenera of the genus *Centaurea*

Av. s. TAACTTTT-G CTAC-GGGAG TCCGTTTCAG GCCCAGCAAC TATCAAACG AAGGTATGGA CGATGAGTAC ATTTCCAGCG TTTCGGATTT TG----- GAAATGAAA- CATTATCCGG CCACAATCCG 130

Centaurea_IC.....T..A.....CACCCAAC A.T.....
Rhaponticoides_IC.....T..A.....CACCCAAC A.T.....
Crupina_IC.....T..A.....CACCCAAC A.T.....
Volutaria_IC.....T..A.....CACCCAAC A.T.....

Centaurea_IIT.....C.....C.....A.....CATCCAAC A.T.....
Rhaponticoides_IIT.....C.....C.....A.....CATCCAAC A.T.....
Crupina_IIT.....C.....C.....A.....CATCCAAC A.T.....
Rhaponticum_IIC.....AG.C.....C.....C.....A.....CATCCAAC A.T.....
Volutaria_IITT.....AG.C.....C.....A.....CATCCAAC A.T.....
Cheirolophus_IIAG.C.....C.....A.....CATCCAAC A.T.....

Centaurea_IIIA.C.....C.....A-----AC..A.....T.....
Carthamus_IIIA.....G.....T.....-----C.....T.....
Phonus_IIIA.....G.....-----T.....T.....
Carduncellus_IIIA.....G.....-----T.....T.....
Crupina_IIIA.....G.....AA.A.....A.....T.....
Volutaria_IIIA.....G.....-----T.....T.....
Cheirolophus_IIIA.....G.....-----T.....T.....

Carcuncellus_IVH.C.....G.....G.....T-----
Rhaponticum_IVA.....H.C.....G.....T-----T.....G.....
Volutaria_IVTTC.....A.....H.C.....G.....T-----

Leuzea_VC. AG.....C.....C.....CA.....-----C.C.C.....G.....T.....
Klasea_VT.....G.....T.....A.C.A.....C.....C.....C.....G.....T.....GT.....
Rhaponticum_VC. AG.....C.....C.....C.....G.....-----G.....C.....G.....T.....G.....

Volutaria_VICTTC.....TC.....C.....C.....C.....C.....CC.....TC.....
Carduncellus_VIAGC.....T.....C.....C.....G.....C.....C.....CC.....TC.....
Crupina_VIAGC.....T.....C.....C.....G.....C.....C.....CC.....TC.....
Rhaponticum_VIC.....AGC.....T.....H.....C.....G.....C.....C.....G.....CC.....TC.....

Cheirolophus_VIIT.....C.....G.....A.C.....H.C.....C.....A.....T.....-----G.....AC.....C.....G.T.....
Crupina_VIIC.....C.....C.....C.....C.....A.....C.....T.....TA.....TC.....CAC.....TC.....A.....TGT.T.....

Cheirolophus_VIIITA.C.....G.....H.G.....G.CC.....G.....T.....G.....G.....-----CC.....AC.....H.....S.GT.....

Av. s. GCCCACGTAA ATTTCTGG- AGAGTT---- CGCC-AC-GG GATT-TTCCG G--ATGCTTC CGGACGAGGC TACCAAAGGG -TACCC-CCT TTGGTGTCTG GGCAAAAAC- ----- ----TAAAAC GTCCA 395

Centaurea_I A...T...A...A...-C-G... C... ..C... ..T... ..S...
 Rhaponticoides_I T...A...A...-C-G... C... ..C... ..T... ..S...
 Crupina_I T...C...A...A...-C-G... C... ..C... ..T... ..S...
 Volutaria_I T...A...A...-C-G... C... ..C... ..A... ..T... ..T

Centaurea_II T...-C-G... A... ..C... ..TTT... ..S...
 Rhaponticoides_II T...-C-G... A... ..C... ..TTT... ..S...
 Crupina_II T...C-C... A... ..C... ..TTT... ..S...
 Rhaponticum_II T...C-C... T... ..C... ..C.C.T... ..T... ..TT... ..S... ..T
 Volutaria_II T...C... A... ..C... ..TT... ..S... ..T
 Cheirolophus_II T...-C-G... A... ..C... ..TTT... ..S...

Centaurea_IIIT... ..A.A...G... ..T-G... -C-C-G... ..G... ..AT... ..TAG--GCG-C CATGG.C.G.
 Carthamus_IIIG... ..G... ..C-C-G... ..G... ..G... ..GG... ..GG... ..GG...
 Phonus_IIIG... ..G... ..C-C-G... ..CC... ..C... ..C... ..G... ..G... ..G...
 Carduncellus_IIIG... ..A... ..G... ..C... ..G... ..G... ..G... ..G... ..G... ..G...
 Crupina_IIITA... ..G... ..C-G... ..CC... ..CG... ..C... ..G... ..G... ..G... ..G...
 Volutaria_IIIG... ..G... ..C... ..CC... ..CG... ..G... ..G... ..G... ..G... ..G...
 Cheirolophus_IIIG... ..G... ..C... ..CC... ..CG... ..G... ..G... ..G... ..G... ..G...

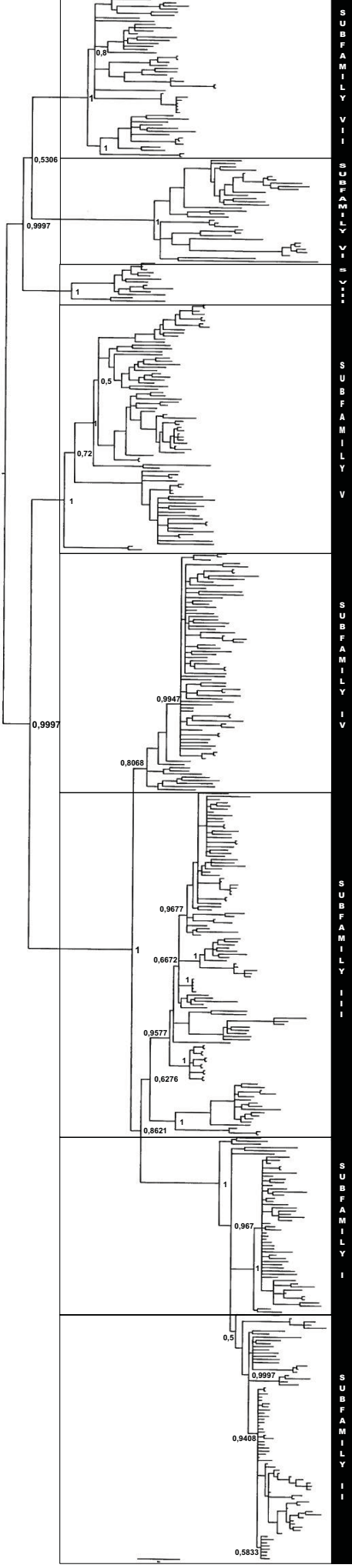
Carcuncellus_IVC... ..C... ..A... ..C... ..G... ..TC... ..A... ..T... ..T... ..T
 Rhaponticum_IVC... ..C... ..C... ..TC... ..CA... ..CA... ..CA... ..CA... ..CA... ..T
 Volutaria_IVC... ..C... ..C... ..TC... ..A... ..A... ..A... ..A... ..A... ..T

Leuzea_VG... ..A.A...C... ..S... ..AA A... ..A... ..T... ..C... ..G... ..TC-GCG-TC AC-T... ..S... ..T
 Klasea_VT... ..G... ..A.G...GA... ..C... ..S... ..T... ..A... ..A... ..G... ..A... ..T... ..C... ..G... ..TC-GCG-TC AC-T... ..S... ..T
 Rhaponticum_VT... ..G... ..GA... ..C... ..S-C... ..TG... ..T... ..AA A... ..A.A... ..A... ..C... ..G... ..TC-GCG-TC AC-T... ..S... ..T

Volutaria_VIGT... ..G... ..A... ..GA... ..AAC-CGC... ..-C... ..T... ..GTC... ..CCA... ..A... ..G... ..-C... ..T... ..C... ..G... ..TC... ..G... ..TCGG-TC TCGC... ..S... ..T
 Carduncellus_VIGT... ..G... ..A... ..AT... ..GA... ..AAC-AC... ..-C... ..GTC... ..CCA... ..A... ..G... ..-G... ..T... ..C... ..G... ..TC... ..G... ..TCGG-TC TCGC... ..S... ..T
 Crupina_VIGT... ..G... ..A... ..AT... ..GA... ..AAC-C... ..T... ..C... ..GTC... ..CCA... ..A... ..G... ..-G... ..T... ..C... ..G... ..TC... ..G... ..TCGG-TC TCGC... ..S... ..T
 Rhaponticum_VIGT... ..G... ..A... ..AT... ..GA... ..AAC-CC... ..T... ..C... ..TC... ..CCA... ..A... ..G... ..-G... ..T... ..C... ..G... ..TC... ..G... ..TCGG-TC TCGC... ..S... ..T

Cheirolophus_VIIA... ..A... ..G... ..S... ..T... ..TCGG... ..C... ..AA... ..A... ..G... ..-C... ..T... ..C... ..G... ..TCGG-ACG-TC GCG... ..S... ..T
 Crupina_VIIA... ..A... ..G... ..S... ..C... ..T... ..TCGG... ..C... ..AA... ..A... ..G... ..-C... ..T... ..C... ..G... ..TCGG-ACA-TC GCG... ..S... ..T

Cheirolophus_VIIIA... ..A... ..TA... ..A... ..C... ..T... ..TCAC... ..A... ..G... ..-C... ..T... ..AT... ..G... ..TCGG-TC GCG... ..S... ..T



Quesada del Bosque ME *et al.* Analysis of HinfI satellite DNA variants in the tribe Cardueae ,en preparación.

Analysis of HinfI satellite DNA variants in the tribe Cardueae

Abstract

- *Background and Aims:* Controversies including difficulties in the delimitation of the tribe, in the delimitation of the boundaries between the subtribes and in the generic delimitation complicate the classification of the tribe Cardueae. Current subtribal classification of Cardueae recognizes the subtribes Carlininae, Cardopatiinae and Echinopsinae as basal subtribes to Carduinae and Centaureinae, the core Cardueae representing more than 90% of species of the tribe. This classification is not fully satisfactory taking into account the very probable paraphyly of Carduinae. We analyze in this paper whether the HinfI satellite DNA family is conserved in the tribe Cardueae and whether the analysis of its evolution might shed light on the subtribal and generic delimitations of this tribe and its current classification.

- *Methods:* We have isolated and compared HinfI repeats from 55 species belonging to 20 genera of Cardueae selected among the subtribes Carlininae, Echinopsinae, Carduinae and Centaureinae. Phylogenetic reconstructions were made by Maximum likelihood and Bayesian inference.

- *Key Results:* We have found up to nine different HinfI subfamilies differentially scattered in different species of each subtribe. Phylogenetic trees grouped together sequences by subfamily affinity instead of species provenance. Notwithstanding, when comparing HinfI repeats of the same type, genetics distances and phylogenetic trees reflected genus and subtribal differentiation in all cases except in comparisons between HinfI sequences of some Centaureinae and between HinfI sequences of some Carduinae species and those of Centaureinae.

- *Conclusions:* The HinfI satellite-DNA family is conserved in the genome of the species of the tribe Cardueae, the largest of Compositae, dating back to 29-24 mya. A "library" of nine HinfI subfamilies would exist in the common ancestor of all Cardueae, which have been differentially amplified in different lineages with the subsequent replacement of one sequence variant by another in different Cardueae species. This process would have been followed of gradual divergence between repeats of a subfamily shared by a particular group of species leading to concerted evolution to a rate that differ between lineages. These different rates might be the responsible of the scarce differentiation found between species of the same

genus in some comparisons. From a phylogenetic standpoint, the results support the current subtribal and generic delimitation of Cardueae but point to the paraphyly of Carduinae.

Keywords: satellite DNA, library hypothesis, concerted evolution, Cardueae, Carlininae, Echinopsinae, Carduinae, Centaureinae.

Introduction

The tribe Cardueae is one of the largest of Compositae, with some 2400 species in 72 genera (Susanna and Garcia-Jacas, 2009). This tribe is one of the five in which the subfamily Cardioideae is divided. The classification of the tribe is complicated due to great morphology diversity and to the large number of species, controversies including difficulties in the delimitation of the tribe, in the delimitation of the boundaries between the subtribes and in the generic delimitation (revised in Susanna and Garcia-Jacas, 2009). Latest molecular analyses confirm that Cardueae is monophyletic and includes the subtribes Carlininae and Echinopsinae, most times segregated of the tribe because are rather different from the subtribes Carduinae and Centaureinae, the core Cardueae, which represents more than 90% of species of the tribe (Susanna et al., 2006; Susanna and Garcia-Jacas, 2009). Current subtribal classification of Cardueae recognizes these four subtribes and the subtribe Cardopatiinae with Carlininae, Cardopatiinae and Echinopsinae as basal subtribes (Susanna and Garcia-Jacas, 2009). Notwithstanding, this classification is a conservative approach and not fully satisfactory taking into account the very probable paraphyly of Carduinae (Garcia-Jacas *et al.*, 2002; Andenberg et al., 2007; Susanna and Garcia-Jacas, 2009).

Subtribe Carlininae includes the genera *Atractylis*, *Atractylodes*, *Carlina*, *Thevenotia* and *Tugarinovia* (Susanna et al., 2006). Subtribe Cardopatiinae includes two monotypic genera, *Cardopatium* and *Cousiniopsis* while subtribe Echinopsinae includes only the genus *Echinops* (Susanna et al., 2006; Susanna and Garcia-Jacas, 2009). Contrasting with the other three subtribes, Centaureinae and Carduinae encompass 90% of the species of Cardueae. Majority of recent studies have confirmed the paraphyly of Carduinae when the monophyletic Centaureinae alone, or Centaureinae plus their sister group (*Arctium-Cousinia-Jurinea-Saussurea*; Hellwig 1996, Häffner and Hellwig 2000; Garcia-Jacas et al. 2002; Susanna et al., 2006) are moved to a different subtribe (Andenberg et al. , 2007; Susanna and Garcia-Jacas, 2009). There are some informal groups within the subtribe Carduinae (Andenberg et al. , 2007; Susanna and Garcia-Jacas, 2009): the basal *Xeranthemum* group (*Amphoricarpos*, *Chardinia*,

Sieberia and *Xeranthemum*); the *Berardia* and *Staezelina* groups, extreme isolates with no affinities to any other genus of Carduinae; the *Onopordum* group (*Onopordum* s.str. and related genera); the *Cynara* group (*Cynara*, *Lamyropsis* and *Ptilostemon*), closely related to the "thistles"; *Carduus-Cirsium* group, the "thistles" (*Carduus*, *Cirsium*, *Notobasis*, *Picnomon*, *Silybum* and *Tyrimnus* and probably *Galactites*); the *Arctium-Cousinia* group (*Arctium*, *Cousinia*, *Hypacanthium* and *Schmalhausenia*); and the *Jurinea-Saussurea* group (*Jurinea*, *Saussurea* and *Dolomiaea*). Subtribe Centaureinae comprises 31 genera with over 800 species. Phylogenetic analyses of Centaureinae support two main phases of radiation, leading to two major groups of genera of different ages, one includes the basal genera of the subtribe and the other includes the derived ones (Garcia-Jacas et al., 2001; Hellwig, 2004). The derived clade, the second radiation, encompasses most of the species currently classified in the genus *Centaurea* and the *Carthamus* complex. The genus *Centaurea* comprises three subgenera: *Centaurea* (formerly the *Jacea* group), *Acrocentron* and *Cyanus* (Susanna and Garcia-Jacas, 2009). The *Carthamus* complex includes the genera *Carduncellus*, *Carthamus*, *Femeniasia* and *Phonus* (Garcia-Jacas et al., 2001). The rest of genera constitute the basal groups (first radiation): *Psephellus*, *Rhaponticoides* (*Centaurea sensu stricto*), *Leuzea*, *Klasea*, *Serratula*, *Volutaria*, *Crupina* or *Cheirolophus*, among others (Garcia-Jacas et al., 2001).

The HinfI satellite-DNA family is conserved throughout the entire subtribe Centaureinae. We have described up to eight HinfI different subfamilies differentially amplified in different species (Quesada del Bosque et al., in press). Notwithstanding, the phylogenetic scattering of HinfI sequence variants through genomes of the Centaureinae species resembled the main phases of radiation of the subtribe (Quesada del Bosque et al., in press). Thus, subfamilies V to VIII were found in older genera (first phase of radiation within the subtribe) while subfamilies I to IV expanded predominantly in the genomes of species belonging to the derived clade of Centaureinae (second phase of radiation). Furthermore, the evolution of HinfI sequences within genomes explained the evolution of major lineages in Centaureinae (Quesada del Bosque et al., in press). Additionally, however, some low-copy counterparts of each HinfI subfamily were found in several species other than those in which they predominate, as it is predicted by the "library" hypothesis of satellite-DNA evolution (Plohl et al., 2012; Quesada del Bosque et al., in press).

We were prompted to analyze in this paper whether the HinfI satellite DNA is conserved in the tribe Cardueae and whether the analysis of its evolution might shed light on the subtribal and generic delimitations of this tribe and its current classification. In this paper, we have analyzed

a total of 376 repeat sequences from 55 species belonging to 20 genera of Cardueae. Of these, a selection of 142 repeats was taken from previous analyzed species of Centaureinae from which we choose representative sequences of each subfamily from each species. The rest of sequences were isolated for the first time for this study. A total of 234 new HinfI repeats were obtained from the genome of species of two genera of Carlininae (*Carlina* and *Atractylis*), the only genus of Echinopsinae (*Echinops*) and six genera of Carduinae (*Jurinea*, *Silybum*, *Carduus*, *Galactites*, *Cirsium* and *Onopordum*).

Materials and Methods

We analyzed a total of 376 HinfI repeat units isolated from 55 species belonging to 20 genera of the tribe *Cardueae* (Table 1). We have included 142 representative sequences previously analyzed by us in Quesada del Bosque *et al.* (in press) belonging to all 39 the species of *Centaureinae* previously studied. The rest of repeat sequences (234) were determined for the first time for this work and belong to additional species of the rest of Cardueae subtribes (Table 1). The mean number of HinfI sequences studied from each new species analyzed was 15 and we never analyzed less than 10 repeats per species in these cases (Table 1). An exception was *Crupina crupinastrum*, for which we analyzed 29 repeats because we studied two populations of this species, one formerly considered *C. matae*, a synonymy of *C. crupinastrum*. The sequences selected from our previous study in Quesada del Bosque *et al.* (in press) were the sequences with EMBL accession numbers The EMBL accession numbers for the rest of 234 new sequences analyzed in this study for the first time were

In addition, the presence of HinfI sequences was analyzed in subfamilies other than Carduoideae, such as the basal Mutisioideae (genus *Gerbera*) or the derived Cichorioideae (genera *Scorzonera* and *Tragopogon*).

Total genomic DNA was extracted from dried leaves from the Herbarium of the University of Granada (Spain) using the DNAeasy Plant Mini Kit (Qiagen). To amplify the monomeric units of the HinfI satellite DNA, we used the primer pairs CenHinf1 and CenHinf2: 5'-GCTTCGTTTTGATAGTTCGTGG-3' and 5'-TAACTTTTGCTACGGGAGTCCG-3'. These primers were internal primers, oriented oppositely with respect to each other. As satellite-DNA sequences are tandemly arranged, we expected to obtain complete monomer and multimer sequences flanked by partial monomer sequences that are excluded from the analysis. This procedure

enables us to avoid the ends of the amplified product. PCR amplifications were performed in a GeneAmp 2700 Applied Biosystems Thermocycler. PCR reactions were performed in a volume of 50 µl of mix containing 10 mM 10xPCR buffer, 2.5 mM MgCl₂, 200 µM of a mix of each dNTP, 1 µM of each primer, 2.5 units of Taq DNA polymerase (Amersham Biosciences) and 10 ng of DNA. The PCR conditions were 40 cycles of 94°C for 60 sec, 50°C for 60 sec, and 72°C for 60 sec. After these 40 cycles, there was a 10-min extension period at 72 °C. Amplification was detected on 1.5% agarose gels stained with SYBR safe[®]. The PCR products were excised from the gels and the DNA purified using the GFX[™] PCR DNA and Gel Band Purification Kit (Amersham Biosciences). The purified PCR products were ligated into the pSC-A-amp/kan vector of the StrataClone PCR Cloning Kit (StrataGene) and cloned in StrataClone SoloPack competent cells (StrataGene) following the manufacturer's recommendations. The transformed colonies were transferred onto a fresh plate and in a 10 µl of a PCR reaction mixture to verify their transformation. Positive clones were purified using a commercial kit (Dominion, Spain).

Automated sequencing of the plasmid clones was performed in both directions using the generic primers M13F and M13R. Thermal-cycling reactions were performed using the ABI Prism[®] Big Dye[™] Terminator Cycle Sequencing Kit (Applied Biosystems). Sequencing was carried out on an ABI PRISM[®] 3100-Avant Genetic Analyzer (Applied Biosystems).

Multiple-sequence alignments were performed using ClustalX (Higgins and Sharp, 1988). Maximum likelihood (ML) and Bayesian analysis were used to infer phylogeny. For selection of the DNA substitution model we used MEGA v.5 (Tamura *et al.*, 2011). The model with the lowest AIC (Akaike Information Criterion) score was chosen (GTR+G). The ML analysis was conducted in MEGA v.5 (Tamura *et al.*, 2011). Bootstrap support values were calculated on 1,000 replicates (Felsenstein, 1985). Bayesian posterior probabilities (PPs) were estimated using MrBayes v3.2 (Ronquist *et al.*, 2012). The extent of rate variation across sites for individual data partitions was estimated by the shape parameter of the gamma distribution (s). Each Bayesian analysis was initiated with random starting trees and was run for one million generations with sampling frequency of trees set at the 500th iteration. For all analyses, the variance of split sequences was 0.01. The potential scale reduction factor (PSRF) was close to 1.0 for all of the substitution model parameters. The fraction of sampled values discarded as burn-in was set at 0.25. Posterior probabilities of 0.95-1.00 were considered statistically significant. Trees were displayed with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Results

PCR resulted in the amplification of HinfI repeats from the genomes of all studied species of the family Cardueae confirming that the four subtribes analyzed shared the HinfI satellite-DNA family (Centaureinae, Carduinae, Echinopsinae and Carlininae). However, the PCR assay failed to amplify this type of repeats in species belonging to subfamilies such as Mutisioideae (genus *Gerbera*) or Cichorioideae (genera *Scorzonera* and *Tragopogon*).

Figure 1 shows a bayesian phylogenetic tree inferred for the HinfI sequences isolated from Carlininae species using as reference outgroups a set of representative HinfI repeats from species of Centaureinae belonging to different HinfI subfamilies. Most repeats of Carlininae, those of *Atractylis humilis* and of the two species of the genus *Carlina*, *C. salicifolia* and *C. acaulis*, appeared clustered in a subclade within a major clade containing also the Centaureinae HinfI sequences belonging to subfamily II (see also extended figure in Supplementary material, Sm_1). Distances between sequences of Carlininae and those of subfamily II of Centaureinae were always lowest than when comparing sequences belonging to different subfamilies within Centaureinae. Notwithstanding, subfamily II sequences of Centaureinae are differentiated from those of Carlininae. The mean Jukes-Cantor genetic distance between subfamily II HinfI sequences of these two subtribes was 0.09 while distances within Centaureinae for subfamily II ranged from 0.05 to 0.08 and within Carlininae ranged between 0.02 and 0.05. HinfI sequences of different species of Carlininae are intermixed in the phylogenetic tree independently of the species provenance and do not group by taxonomic affinity. In contrast, the sequences isolated from *A. leyseroides*, formed two distinct subclades within a clade containing sequences of subfamily V of Centaureinae. Mean genetic distance between sequences of *A. leyseroides* was 0.09 and between *A. leyseroides* and those of subfamily V of Centaureinae was 0.18. In addition, a few sequences (three) of *Atractylis* species appeared clustered together subfamily IV sequences, the same that one sequence of *C. salicifolia* was framed within clade containing subfamily V repeats.

The sequences isolated from *Echinops* species appeared in a distinct subclade within a major clade containing the repeats of subfamily II of Centaureinae (Figures 2 and Sm_2). The sequences of *Echinops* are clearly differentiated from the sequences of species of Centaureinae (mean Jukes-Cantor genetic distance of 0.25; distances within Centaureinae for subfamily II ranged between 0.05 and 0.08 while within Echinopsinae, the mean genetic

distance was 0.07). In contrast to Carlininae, the sequences isolated from two different species of *Echinops* (*E. sphaerocephalus* and *E. strigosus*) grouped differentially by taxonomic affinity. The tree of the Figures 3 and Sm_3 represents the phylogenetic relationships between HinfI repeats of Carduinae and Centaureinae. Six sequences from *Galactites duriaei*, all the sequences of *Onopordum nervosum* as well as all the sequences of the two species of *Cirsium* (*C. echinatum* and *C. pyrenaicum*) formed a clade, statistically highly supported. HinfI sequences diverged accordingly between the different genera analyzed and repeats are grouped in three clades that coincide with the three compared genera (*Galactites*, *Cirsium* and *Onopordum*), although the sequences of *Cirsium* were grouped intermixed together in the same clade independently of specific provenance. However, the rest of sequences of *G. duriaei* and the sequences isolated from species of the genera *Jurinea*, *Sylibum* and *Carduus* appeared in the tree intermixed with HinfI sequences of Centaureinae species. Specifically, most sequences of *Jurinea* grouped together with those of subfamily II, although two sequences clustered with subfamily III repeats and two other together with repeats of subfamily IV. *Sylibum* sequences grouped together with repeats belonging to subfamilies II, III, V or VII. Sequences of *Carduus* appeared intermixed with those of subfamilies II, III, IV, VI, VII and VIII. The species of these three genera are characterized by a feature shared with some Centaureinae species, that is, the presence in the same genome of several types of HinfI sequences. Genetic distances between sequences of these Carduinae species and those of their Centaureinae counterparts ranged between 0.05 and 0.25, while mean distances between sequences of each subfamily within each subtribe ranged between 0.05 and 0.09 for Carduinae and between 0.05 and 0.11 for Centaureinae.

The inferred tree for all HinfI sequences is shown in a large figure in Supplementary material (Sm_4). This tree confirm the same clustering and phylogenetic relationships of HinfI repeats found in independent subtribal analyses (Figures 1, 2 and 3).

Discussion

The HinfI satellite-DNA family is conserved in the genome of the species of the tribe Cardueae, the largest of Compositae (Susanna and Garcia-Jacas, 2009). We have tried to isolate this type of sequences from the genome of related species of different Compositae subfamilies such as the basal Mutisioideae (genus *Gerbera*) or the derived Cichorioideae (genera *Scorzonera* and *Tragopogon*) by PCR amplification. However, the set of primers used in our studies failed to

amplify HinfI sequences from these species circumscribing this family of repeats to the subfamily Carduoideae, dating back to 29-24 mya (Kim *et al.*, 2006; Wang *et al.*, 2007).

As occurred for the HinfI sequences in Centaureinae (Quesada del Bosque *et al.*, in press), different HinfI subfamilies or sequence types amplified depending of the lineage. According to the phylogeny and genetic distances, we can ascribe HinfI sequences from each species to a different subfamily type. In this context, in a broad sense, if we take the license to call "orthologous" repeats to all repeats derived from the amplification of a same type of repeat in a single or several events of amplification, we can treat different subfamilies as "paralogous" groups composed of "orthologous" repeats of the HinfI family (Table 2). As a consequence of comparing orthologous versus paralogous subfamilies, our study reveals that comparisons between sequences of phylogenetically distant species result most times in shorter genetic distances and closest phylogenetic relationships than when comparing sequences from closer species, depending on the subfamily type that they have. That is the case for Echinopsinae, in which the HinfI sequences are orthologous to sequences of Centaureinae subfamily II (Table 2). HinfI sequences of three of the analyzed species of Carlininae (*Carlina salicifolia* and *C. acaulis* and *Atractylis humilis*) are orthologous also to sequences of Centaureinae subfamily II (Table 2) while sequences of the fourth species analyzed of this subtribe, *Atractylis leyseroides*, belong to the subfamily V (Table 2). Notably, we find two set of sequences in Cardueinae (Figures 1 and 2; Table 2). On one hand, there are a set composed by those isolated from the genera *Jurinea* (*Jurinea-Saussurea* group) and *Silybum* and *Carduus* (*Carduus-Cirsium* group), which genomes bear several HinfI types of repeats belonging to subfamilies II to V, depending on the species, but also belonging to subfamilies VII and VIII in a few cases. The trees of Figures 3, Sm_3 and Sm-4 reflect this point, with sequences of these species spread in several clades according to their "orthology" more than according to species phylogeny. In addition, we have found in this study a new subfamily type, highly differentiated, specific of the genera *Cirsium* and *Onopordum*, which was also found in the genome of *Galactites duriaei* (six out fifteen of its sequences belonged to this new subfamily while the others belonged to subfamilies II, III and V. By analogy to previous studies, we have called the new HinfI type of sequences, subfamily IX (Table 2).

This study supports our previous findings in Centaureinae (Quesada del Bosque *et al.*, in press). We can assume that a "library" of HinfI sequence variants or subfamilies existed in the common ancestor of, not only Centaureinae but, all Cardueae. From this library any variant would have been differentially amplified in different lineages with the subsequent

replacement of one sequence variant by another in different Cardueae species (Plohl *et al.*, 2012; Quesada del Bosque *et al.*, in press). In addition, the "library" hypothesis predicts that unrelated species-specific dominant satellite-DNA subfamily repeats would have low-copy counterparts of each of them in other species and comparisons of high-copy and low-copy monomer variants of these subfamilies would show high interspecific sequence conservation and the complete lack of any species-diagnostic mutations (Plohl *et al.*, 2012; Quesada del Bosque *et al.*, in press) as we have found in Cardueae. The scattering of HinfI subfamilies in different species of Cardueae and the sharing of different subfamily types differentially amplified in the same species, supports this hypothesis and it is a probe of saltatory evolution of satellite DNA. This mode of evolution has been found previously in insects (Meštrović *et al.*, 1998; Mravinac *et al.*, 2002; Cesari *et al.*, 2003), in amphibians (Felicciello *et al.*, 2005, 2006) and in plants (Navajas-Pérez *et al.*, 2009; Quesada del Bosque *et al.*, 2011). Notwithstanding, this is the first report in which a large group of plant species is deeply analyzed for a satellite DNA family demonstrating the "library" hypothesis. There is still an important question to elucidate as it is the coexistence of repeats of different subfamilies similarly abundant in the same genome. In these cases, other factors such as reticulate evolution would be involved, as might occur in the genus *Centaurea* (Suárez-Santiago *et al.*, 2007). But random amplification of different HinfI variants in the same lineage or gene conversion (Drouin and Dover, 1990; Plohl and Ugarkovic, 1994a,b) could also explain this type of variability observed in some Cardueae species.

From a phylogenetic standpoint, the results obtained could appear puzzling because sequences from species of different subtribes are clustered together in the same clades of the trees and genetic distances between them are lowest than when comparing sequences isolated from species of the same subtribe when belonging to different subfamilies, even when comparing sequences of the same genus or species. By this, it is coherent to talk about "paralogous" subfamilies. In fact, when sequences of the same type ("orthologous" repeats of the same subfamily) are compared, the sequences group together according to the currently accepted subtribe delimitation. This is so in the case of species of Carlininae and Echinopsinae. With this taken into account, what we have found in this study for each subfamily when comparing sequences of different subtribes would be the result of the gradual differentiation in HinfI sequences by molecular drive (Plohl *et al.*, 2012; Pérez-Gutiérrez *et al.*, 2012). Monomer sequences of a satellite DNA evolve concertedly through a process of molecular drive in which mutations are homogenized in a genome and fixed in a population at a higher rate than the one at which they arise. This process results in rapid divergence of

satellite sequences in reproductively isolated groups of organisms (Plohl *et al.*, 2012). Nevertheless, the rate of concerted evolution depends on satellite-DNA location, organization, and repeat-copy number (Navajas-Pérez *et al.*, 2005, 2009), functional constraints (Mravinac *et al.*, 2005) or biological factors (Luchetti *et al.*, 2003, 2006; Robles *et al.*, 2004; Suárez-Santiago *et al.*, 2007a). In the absence of selective and biological constraints, the rate of concerted evolution of a family of satellite-DNA sequences should depend basically on the divergence time between species (Pérez-Gutiérrez *et al.*, 2012). However, we have not found sequence differentiation within Carlininae. The HinfI sequences of subfamily II isolated from these species appeared intermixed in the phylogenetic tree, an indication of absence of concerted evolution, which might be explained as a consequence of time (Pérez-Gutiérrez *et al.*, 2012) assuming a recent split between the compared species even though they belong to two different genera. In contrast, we have found that sequences of two different species cluster separately in two different subclades, an indication of concerted evolution. Alternatively, we could assume different rates of sequence change in these two different lineages (Adegoke *et al.*, 1993).

The situation in Carduinae is not so direct since we can distinguish between two groups of species as mentioned above. On one hand, the set composed by HinfI repeats of the genera *Jurinea* (*Jurinea-Saussurea* group) and *Carduus* and *Silybum* (*Carduus-Cirsium* group). In this case, there is not a differentiation between sequences of both subtribes, Centaureinae and Carduinae, but their sequences appear intermixed without cluster by taxonomic affinity. On the other hand, *Onopordum* (*Onopordum* group) and *Cirsium* (*Carduus-Cirsium* group), as well as *Galactites duriaei* (*Carduus-Cirsium* group) have their own subfamily highly differentiated from the rest of sequences, not only of those isolated from species of the same subtribe but also from any other subtribe. These data support the paraphyly of Carduinae (García-Jacas *et al.*, 2002; Susanna *et al.*, 2006; Susanna and García-Jacas, 2007, 2009). The results found for the HinfI sequences of *Jurinea*, *Carduus* and *Silybum* resemble the behavior of these sequences in Centaureinae with no genetic differentiation between their species. Furthermore, the genetic distances between HinfI sequences of both subtribes for subfamily II repeats are lower than between species of the tribe Centaureinae in some comparisons. However, *Cirsium*, *Onopordum* and *Galactites* would be clearly differentiated from the rest of “thistles”. This data is not based only in the presence of subfamily IX in their genomes, a consequence of the amplification of this specific subfamily in these genera. In addition to subfamily IX, *Galactites duriaei* also have in its genome sequences of the subfamilies II, III and V, and, in all the cases, the genetic distances between HinfI sequences of this species and those of Centaureinae was

higher (mean genetic distances ranging between 0.07 and 0.23, depending of the subfamily) than intraspecific variability (0.01 in all cases), a similar degree of divergence to that found in comparisons between other subtribes. In addition, HinfI sequences of the subfamily IX have diverged accordingly between the different genera analyzed and repeats are grouped in three clades that coincide with the three compared genera (*Galactites*, *Cirsium* and *Onopordum*) demonstrating the gradual divergence accounted for satellite DNA sequences shared by an specific group of related species.

