UNIVERSIDAD DE LEÓN

DEPARTAMENTO DE BIOLOGÍA MOLECULAR (ÁREA DE GENÉTICA)

TESIS DOCTORAL

DESARROLLO DE MARCADORES GENÉTICOS A PARTIR DE SECUENCIAS DE RETROTRANSPOSONES DEL GENOMA DE LENTEJA, *Lens culinaris* Medik

Mª Rita Rey Baños

2015

Esta tesis ha sido realizada gracias a una beca predoctoral de la Universidad de León, a las Ayudas a Grupos de Excelencia GR113 de la Junta de Castilla y León y a los proyectos AGL2009-07853, AGL2013-44714-R y LEGRESIST.

A mi familia

Agradecimientos

A lo largo de estos años he tenido la suerte de conocer buenos amigos y de pasar momentos inolvidables que me han ayudado a susperar los malos momentos y a disfrutar de los buenos.

A mis directores, el Dr. D. Marcelino Pérez de la Vega y el Dr. D. Luís Saenz de Miera Carnicer por su paciencia y por todo el apoyo que me han brindado durante todos estos años de trabajo.

A los profesores y doctores del área de Genética: Francisca Vaquero, Francisco Javier Vences, Pedro García, Maribel Peláez, María Luisa Ruiz, Ana Isabel González, Carlos Polanco y Richard Fratini por su ayuda y apoyo a lo largo de estos años.

A Ana y Juan que siempre han estado ahí para todo lo que he necesitado y sin los cuales esta tesis no se podría haber llevado a cabo.

A la mesa del café, en la que debo añadir a Marta y Luís Carlos, por esos momentos de camaradería con nuestros crucigramas, santorales y desayunos panocheros.

A mis compañeros doctorandos Nati, Carlos David, Carlos y Ari, con los que he compartido laboratorio y confidencias y a los que me une una gran amistad.

A toda mi familia, debo decir que muy numerosa, que siempre ha estado apoyándome y animándome en todo momento desde todos los puntos cardinales de la península, las islas y incluso desde otros países. Gracias sobre todo a mis padres y mis hermanos y, en especial, a mi abuela Domi que estuvo a mi lado casi hasta el final y a mis pequeñajos Juan, Carmen, David, Rodrigo, Pilar y Elena que han ido llegando y creciendo durante estos años de trabajo.

A mis grandes amigas Eri, Rebe, Sandra y mi hermana, Elena, con las que he pasado los mejores momentos de mi vida y al resto de mis amigos y compañeros de "viaje" que siempre están dispuestos a echar una mano y a estar donde y cuando se les necesita.

En fin, a todos aquellos sin los que el trabajo a veces no puede salir adelante y sin los que éste no merecería la pena. A todos vosotros muchas gracias por estar ahí.

Índice

1. Introducción ... 3

1.1. La lenteja ... 3
1.1.1. Características morfológicas de la lenteja ... 3
1.1.2. Origen y expansión de la lenteja ... 4
1.1.3. Taxonomía de *Lens* Miller ... 6
1.1.4. Variabilidad de la lenteja cultivada ... 9
1.1.5. Importancia nutricional y agronómica ... 11

1.2. Los elementos transponibles ... 12
1.2.1. Elementos de clase I o retroelementos ... 13
1.2.2. Elementos de clase II o transposones ... 17

1.3. Los retrotransposones LTR ... 19
1.3.1. Características estructurales ... 19
1.3.2. Ciclo de movilización ... 25
1.3.3. Efecto de los retrotranposones en el genoma de plantas ... 30

1.4. Marcadores moleculares basados en retrotransposones ... 31

1.5. Mapas genéticos ... 40

2. Objetivos ... 49

3. Materiales y métodos ... 53

3.1. Material Vegetal ... 53

3.2. Extracción de ADN ... 53

3.3. Electroforesis en geles de agarosa ... 54

3.4. Amplificación de retrotransposones ... 55

3.5. Amplificación de repeticiones terminales largas LTRs ... 57
3.5.1. Amplificación de LTRs a partir de secuencias internas conservadas del retrotransposón ... 57
3.5.2. iPBS (inter-Primer Binding Site) ... 60

3.6. Extracción de bandas específicas ... 61

3.7. Clonación y transformación ... 62

3.8. Secuenciación de ADN ... 65

3.9. Análisis de secuencias ... 65

3.10. Diseño de cebadores ... 66

3.11. Técnicas para la búsqueda de marcadores moleculares ... 67
3.11.1. SSAP ... 67
3.11.2. REMAP ... 71
3.11.3. iPBS ... 72

3.12. Mapeo genético ... 73
3.12.1. Cálculo de la fracción de recombinación ... 73
3.12.2. Construcción del mapa genético ... 74

3.13. Otros programas informáticos utilizados ... 76

4. Resultados y discusión..... 79

4.1. Búsqueda de marcadores derivados de retrotransposones en el genoma de lenteja 79

4.1.1. Cebadores derivados del gen de la Retrotranscriptasa 79

4.1.1.1. Retrotranscriptasa de retrotransposones *Ty1-copia* 80

4.1.1.2. Retrotranscriptasa de retrotransposones *Ty3-gypsy* 88

4.1.2. Cebadores derivados de un retrotransposón de lenteja 97

4.1.3. Retrotransposnes de plantas. Corolario..... .105

4.2. Identificación de repeticiones largas terminales o LTRs..... 107

4.2.1. Identificación de LTRs a partir de las secuencias obtenidas con los cebadores diseñados a partir de la retrotranscriptasa de *copia* y *gypsy*..... 107

4.2.2. Identificación de LTRs a partir de la secuencia de la RNasaH..... 112

4.2.3. Identificacón de LTRs a partir de secuencias PBS 117

4.2.4. Identificación de un LTR completo descrito en *Lens culinaris* 124

4.3. Obtención de marcadores moleculares..... 125

4.3.1. SSAP (Sequence-Specific Amplification Polymorphism)..... 126

4.3.2. REMAP (REtrotransposon-Microsatellite Amplified Polymorphisms) 127

4.3.3. iPBS (inter-Primer Binding Site) 128

4.4. Localización de los marcadores en un mapa genético..... 131

4.4.1. Mapa genético de marcadores derivados de secuencias de retrotransposones..... 131

4.4.2. Mapa genético ampliado con otros marcadores..... 134

5. Conclusiones..... 151

6. Bibliografía..... 155

Abreviaturas

A: adenina.

ADN: ácido desoxirribonucleico.

ADNc: ADN copia.

ARNm: ARN mensajero.

AFLP: polimorfismo en longitud de fragmentos amplificados.

ALP: polimorfismo de longitud amplificado.

BLAST: herramienta para la búsqueda de alineamientos.

C: citosina.

CAP: secuencias de restricción amplificadas y polimórficas.

CCD: dominio central catalítico.

CDD: base de datos de dominios conservados.

cM: centimorgan.

CRF: Centro de Recursos Fitogenéticos

D: distancia.

EDTA: ácido etilendiaminotetraacético.

env: gen que codifica para proteínas de la envuelta de los virus.

ET: elemento transponible.

FAOSTAT: organización de las Naciones Unidas para la alimentación y la agricultura.

G: guanina.

gag: gen que codifica para proteínas de maduración y empaquetamiento.

HRM: fusión de alta resolución.

ICARDA: Centro Internacional de Investigación Agrícola en Áreas de Secano (Alepo, Siria).

INIA: Instituto Nacional de Investigación y Tecnología Agraria y Alimentación.

int: gen que codifica para la integrasa (int).

IPTG: isopropil-β-D-tiogalactósido.

IR: repeticiones invertidas.

IRAP: polimorfismo amplificado entre retrotransposones.

ISSR: repeticiones entre microsatélites.

ITAP: polimorfismos dirigidos a intrones de genes.

FAD: Flavín adenín dinucleótido.

LARD: derivados largos de retrotransposones.
LB: medio de cultivo Luria Bertani.
LINE: elemento nuclear disperso largo.
LOD: logaritmo de las probabilidades.
LTR: repeticiones largas terminales.
M: molaridad.
MAS: selección asistida por marcadores.
MITE: elementos transponibles con repeticiones invertidas en miniatura
NADH: Nicotinamida adenina dinucleótido reducida.
NCBI: centro nacional de información biotecnológica.
PAGE: electroforesis en gel de poliacrilamida.
PBS: sitio de unión de proteínas.
PCR: reacción en cadena de la polimerasa.
pol: gen que codifica para la retrotranscriptasa y para la RNasaH.
PPT: región rica y purinas.
P/A: presencia/ausencia.
QTL: locus de un carácter cuantitativo.
RAPD: ADN polimórfico amplificado al azar.
RBIP: polimorfismo de amplificación basado en retrotransposones.
REMAP: polimorfismos amplificado entre retrotransposón y microsatélite.
RFLP: polimorfismo en longitud de fragmentos de restricción.
RIL: línea consanguínea recombinante.
RIVP: polimorfismo de variación interna de retrotransposones.
RNasaH: ribonucleasa H.
RNT-LTR: retrotransposones con LTR.
RUP: polimorfismo del UTR de un retrotransposón.
SDS: dodecilsulfato sódico.
SINE: elemento nuclear disperso corto.
SNP: polimorfismo de un único nucleótido.
SRAP: polimorfismo amplificado relacionado con secuencias.
SSAP: polimorfismo de amplificación de secuencia específica.
SSR: microsatélite.
T: timina.
TAE: tampón de Tris-acetato y EDTA.

TAM: marcador basado en microarrays.

TIR: repeticiones terminales repetidas.

TRIM: retortransposnes en miniatura con repeticiones terminales.

TSD: diana de duplicación.

U: unidades.

udm: unidades de mapa.

UTR: región líder no traducible.

UV: ultravioleta.

VLP: partícula similar a virus.

VNTR: repeticiones en tándem de número variable.

X-gal: 5-bromo-4-cloro-3-indolil-β-D-galactopiranósido

Relación de figuras

Figura1.1: Ilustración de *Lens culinaris*..... 3

Figura1.2: Datos arqueológicos. Zona de distribución y domesticación de *L. orientalis* 5

Figura1.3: Tipos de transposición: replicativa y no replicativa 13

Figura1.4: Organización estructural general de los diferentes tipos de ET en plantas 18

Figura1.5: Organización genómica de los retrotransposones tipo *Ty1-copia* y *Ty3-gypsy* 19

Figura1.6: Estructura típica de la LTR de un retrotransposón 20

Figura1.7: Ciclo de los retrotransposones LTR 27

Figura1.8: Esquema de la retrotransposición 28

Figura1.9: Esquema de una técnica de marcadores moleculares tipo “southern”.... 33

Figura1.10: Esquema general de la PCR..... 34

Figura1.11: Esquema de los marcadores moleculares basados en retrotransposones más utilizados 40

Figura 3.1 : Vector pGEM®-T diseñado para la clonación de productos de PCR..... 64

Figura 3.2: Analizador de fragmentos de ADN, ABI Prism 310 70

Figura 4.1: Bandas obtenidas con los cebadores diseñados para *Ty1-copia* y *Ty3-gypsy*. 80

Figura 4.2: Minipreparación ultrarrápida de las colonias positivas 81

Figura 4.3: Secuencias de aminoácidos obtenidas con los cebadores diseñados a partir del gen de la retrotranscriptasa de *Ty1-copia* 83

Figura 4.4: Árbol filogenético basado de la retrotranscriptasa de retrotransposones *Ty1-copia*..... 86

Figura 4.5: Secuencias de aminoácidos obtenidas con los cebadores diseñados a partir del gen de la retrotranscriptasa de *Ty3-gypsy*..... 90

Figura 4.6: Ampliación de las secuencias *Ty3-gypsy*..... 94

Figura 4.7: Árbol filogenético de la retrotranscriptasa de retrotransposones *Ty3-gypsy*..... 96

Figura 4.8: Bandas obtenidas con los cebadores diseñados para un retrotransposón de tipo desconocido de *Lens culinaris*. 98

Figura 4.9: Secuencias obtenidas con los cebadores diseñados a partir del retrotransposón de tipo desconocido de *Lens culinaris*..... 100

Figura 4.10A: Árbol filogenético de las secuencias obtenidas diseñados a partir de una secuencia de un retrotransposón de tipo desconocido de *Lens culinaris* correspondiente a *Ty1-copia*..... 104

Figura 4.10B: Árbol filogenético de las secuencias obtenidas diseñados a partir de una secuencia de un retrotransposón de tipo desconocido de *Lens culinaris* correspondiente a *Ty3-gypsy* 105

Figura 4.11: Localización de las diferentes secuencias obtenidas en el tipo de retrotransposón correspondiente a *Lens*..... 105

Figura 4.12: Bandas obtenidas con los cebadores NLc..... 108

Figura 4.13: Secuencias nucleotídicas y aminoacídicas, obtenidas con los cebadores NLc1, NLc3 y NLc4 111

Figura 4.14: Esquema de la técnica “nested PCR” o PCR anidada..... 113

Figura 4.15: Bandas obtenidas con los cebadores Rnasah1 y Rnasah2 114

Figura 4.16: Secuencias de proteínas obtenidas con los cebadores Rnasah1 y Rnasah2..... 116

Figura 4.17: Secuencias de proteínas, PPT y LTR obtenidas con los cebadores Rnasah1 y Rnasah2..... 116

Figura 4.18: Patrones de bandas obtenidos con diferentes cebadores PBS 119

Figura 4.19: Secuencias obtenidas a partir de los cebadores PBS 122

Figura 4.20: Secuencia obtenida tras la amplificación con los cebadores LTR-Ps y LTR-1F..... 125

Figura 4.21: Mapa genético obtenido con los marcadores basados en retrotransposones 133

Figura 4.22: Mapa genético obtenido con todos los marcadores analizados en una población RIL F_7..... 136

Figura 4.23A: Relación entre los grupos GL_R3 y GL_R4 con $GL_{RC}3$..... 138

Figura 4.23B: Relación entre los grupos GL_R1 y GL_R5 con $GL_{RC}1$ 139

Figura 4.24A: Reordenamiento de los marcadores de los grupos GL_R1 y GL_R5 al unirse para formar $GL_{RC}1$ 140

Figura 4.24B: Reordenamiento de los marcadores de los grupos GL_R2 y $GL_{RC}2$..... 141

Figura 4.24C: Reordenamiento de los marcadores de los grupos GL_R3 y GL_R4 al unirse para formar $GL_{RC}3$ 142

Figura 4.25: Relación entre el grupo $GL_{RC}3$ del mapa completo obtenido en este trabajo con los grupos GL2, GL4 y GL9 descritos por Neyra (2014)..... 144

Relación de tablas

Tabla1.1: Agrupaciones de las especies de *Lens* 8

Tabla1.2: Relación de los mayores exportadores e importadores de lenteja del mundo en el año 2011 12

Tabla 3.1: Secuencias de los cebadores utilizados y nombre con el que se denominaron a lo largo del estudio 55

Tabla 3.2: Cebadores diseñados a partir de zonas flanqueantes de microsatélites (SSRs) de lenteja 72

Tabla 4.1: Relación de las secuencias localizadas en la base de datos del NCBI que presentan un nivel de identidad mínimo del 90% con los amplificados de cebadores Copia, Gypsy y Tnana. 106

Tabla 4.2: Cebadores diseñados a partir de algunas de las secuencias obtenidas con los cebadores diseñados para *Ty1-copia* 108

Tabla 4.3: Relación de las secuencias localizadas en la base de datos del NCBI que presentan un nivel de identidad mayor del 90% con los amplificados de los cebadores NLc 109

Tabla 4.4: Cebadores utilizados con la técnica iPBS 118

Tabla 4.5: Secuencia de los cebadores diseñados a partir de los LTRs descritos en este trabajo, en el genoma de *Lens culinaris* cv. Lupa 126

Tabla 4.6: Relación de polimorfismos en función de cada uno de los marcadores moleculares utilizados 129

Tabla 4.7: Polimorfismos obtenidos con los diferentes cebadores 130

Tabla 4.8: Resumen del mapa genético basado en retrotransposones 132

Tabla 4.9: Resumen del mapa genético completo 135

Tabla 4.10: Relación de los grupos GL_{RC} con los grupos descrito por Sharpe *et al*. (2013) y su posible equivalencia con los cromosomas de *Medicago truncatula* 146

INTRODUCCIÓN

1. Introducción

1.1. La lenteja

1.1.1. Características morfológicas de la lenteja

La lenteja (*Lens culinaris* Medik. ssp. *culinaris*) es una especie herbácea anual de porte erecto o semierecto, adaptada a climas secos, que presenta una altura media de entre 30 y 35 cm, es fundamentalmente autógama, diploide con 2n = 14 cromosomas y con un tamaño de genoma de, aproximadamente, 4.063 Mpb (Arumuganathan y Earle, 1991).

Los tallos, cuadrados y estriados, son delgados al igual que las ramas, mientras que el sistema radicular está formado por una raíz pivotante delgada y raices laterales fibrosas. La superficie de las raices está cubierta de numerosos nódulos de formas variadas. Las hojas son pinnadas o paripinnadas, con zarcillos en las hojas superiores y foliolos ovalados y lanceolados. Las flores son papilionadas blancas, violetas o azules y pueden aparecer de forma individual o en grupos, unidas por pedúnculos florales que aparecen en los nodos superiores de la planta. Los frutos son ablongos, comprimidos lateralmente, con una longitud de 6 a 20 mm y una anchura de entre 3,5 y 11 mm, y contienen una, dos o raramente tres semillas de forma lenticular y ricas en proteínas y fibra (Pérez de la Vega *et al*., 2011).

Figura1.1: Ilustración de *Lens culinaris*. Fuente original: ''Flora von Deutschland, Österreich und der Schweiz'' 1885, Gera, Alemania.

1.1.2. Origen y expansión de la lenteja

Según Pérez de la Vega *et al*. (2011), la lenteja es una de las primeras especies domesticadas en el creciente fértil, y se considera tan antigua como el trigo, la cebada o los guisantes (Harlan, 1992). Se ha cultivado a lo largo de los últimos 10.000 años en los ambientes más extremos.

Según los datos arqueológicos dados por Cubero (1981), los restos más antiguos de lentejas silvestres (10.500 años a.C.) se encontraron en Mureybit (Siria) mientras que los más antiguos de lenteja cultivada (aproximadamente 9000 años a.C) aparecieron en Hacilar (Turquía), Ramad (Siria), Jarmo (Iraq), Jericó (Palestina), Beidha (Jordania) y Ali Kosh (Iran). Las lentejas más antiguas que aparecieron en Grecia se han datado hacia el año 8000 a.C., mientras que en Europa Central no aparecen hasta el 5.000-7.000 a.C. Más tarde llegaron a la India (3.000-4.000 a.C.) y Europa occidental (3.000-3.500 a.C.). La lenteja debió llegar a Egipto antes de lo que indican los restos arqueológicos por su proximidad geográfica y cultural con el creciente fértil, sin embargo, los problemas de preservación que sufren los materiales orgánicos presentes en el delta del Nilo, dificultan la datación real. Los restos más antiguos en la Península Ibérica datan de aproximadamente el año 5.500 a.C. (Buxó, 1997).

Ladizinsky (1999) describió la lenteja cultivada como monomórfica para varios caracteres, mientras que se encontraron polimorfismos en las accesiones de la forma silvestre *Lens orientalis*. Sin embargo, tres accesiones del este de Turquía y norte de Siria, compartían las mismas características con el cultigen y podrían considerarse, por tanto, como miembros del acervo del que se derivó la lenteja cultivada. Zohary (1999) llegó a la conclusión de que la lenteja se domesticó sólo una vez o muy pocas veces, basándose en polimorfismos cromosómicosy del ADN.

Con todos los datos disponibles, Lev-Yadun *et al*. (2000) sugirieron un lugar que se superpone al área del Creciente Fértil (Figura 1.2), la región del sur de Turquía y norte de Siria, como el lugar más probable de domesticación de la lenteja, donde

algunas poblaciones de *L. orientalis* fueron sometidas a selección para dar lugar a un nuevo cultivo, *L. culinaris*.

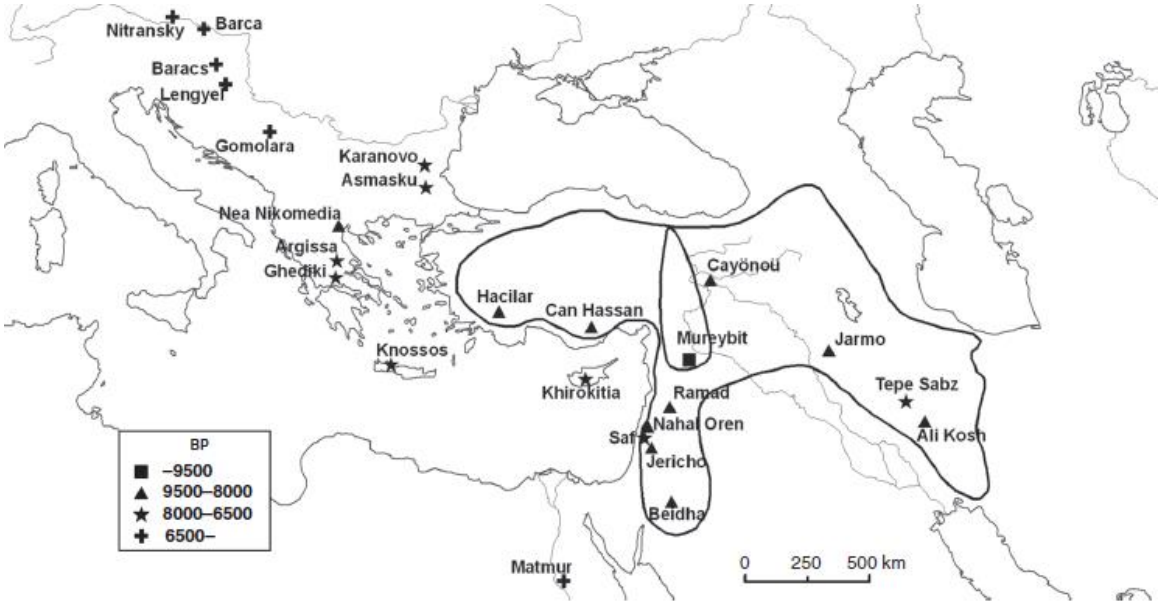


Figura1.2: Datos arqueológicos. Zona de distribución y domesticación de *L. orientalis*. Las líneas indican la zona principal de distribución de *L. orientalis* y el área de domesticación propuesta dentro de ella. (Tomado de Cubero *et al*., 2009).

Según Cubero *et al*. (2009), es muy probable que la expansión de la lenteja ocurriera simultánemente con la de otras especies domesticadas en el creciente fértil. A partir de la zona de origen, el cultivo se fue extendiendo por Grecia, centro y oeste de Europa a lo largo del Danubio, el delta del Nilo y el este de la India. La introducción de la lenteja en la India debió de ser con una muestra de semillas muy pequeña ya que la variabilidad genética encontrada en la zona es muy limitada, a pesar de ser la región de cultivo más extensa del mundo. Los primeros restos arqueológicos encontrados en la península Ibérica pertenecen al Neolítico temprano y, en base al tamaño de la semilla, se trata de restos de la variedad microsperma (Buxó, 1997). El cultivo fue introducido en América por los españoles en el siglo XVI.

La variabilidad genética en el centro de origen se ha mantenido hasta nuestros días. Ferguson *et al*. (1998), mediante el uso de marcadores moleculares, localizaron áreas que presentaban gran diversidad genética en diferentes especies: *L. culinaris* ssp.

orientalis (sureste de Turquía, noreste de Siria, y Jordania), *L. odemensis* (sur de Siria), *L. ervoides* (región entre Siria y Turquía) y *L. nigricans* (este de Turquía).

1.1.3. Taxonomía de *Lens* Miller

El nombre científico de la lenteja, *Lens culinaris*, lo estableció Friedrich Kasimir Medikus (1736-1808) en 1787 (Cubero *et. al*., 2009). A lo largo de la historia se han usado diferentes sinónimos como *Lens esculenta* Moench, e, incluso, se ha incluido como especie de otros géneros como *Cicer lens*, *Vicia lens* o *Ervum lens*. Sin embargo, en el Congreso Internacional de Botánica celebrado en Edimburgo en 1964, se decidió establecer como nombre del género *Lens* Miller (Cubero *et. al*., 2009). Dicho género pertenece a la tribu Vicieae D. C., junto con los géneros *Lathyrus*, *Pisum* y *Vicia* (Gunn, 1969). Otros géneros como *Cicer* y *Vavilovia* han sido incluidos en ocasiones (Gunn y Kluve, 1976) aunque el primero se ha excluido en base a nuevos estudios filogenéticos y la inclusión del segundo es muy discutida (Clarke y Kupicha, 1976, Sinjushin y Akopian, 2011).

La clasificación del género *Lens* realizada en base a características morfológicas ha ocasionado problemas en la identificación de las diferentes especies. De esta manera, Boissier describió la especie *Ervum cyaneum* como una mezcla de las características de *ervoides*, *nigricans* y *orientalis*. También han existido dudas taxonómicas en cuanto a la clasificación de *L. orientalis* como especie o subespecie y en cuanto a la posición filogenética de *L. nigricans*.

Ya que *L. orientalis* es el antecesor silvestre de la especie cultivada debe ser considerada como subespecie, es decir, *L. culinaris* ssp. *orientalis*. (Harlan and De Wet, 1971). Sin embargo, la clasificación clásica del género incluye cuatro especies: *L. culinaris*, *L. orientalis*, *L. nigricans* y *L. ervoides*. (Ladizinsky, 1979). En 1984, algunas accesiones de *L. nigricans* se reclasificaron como *L. odemensis* y se propuso una nueva clasificación taxonómica del género: *L. culinaris* con las subespecie cultivada *L. c. culinaris* y las subespecies silvestres *L. c. orientalis* y *L. c. odemensis* y *L. nigricans* con las subespecies *L. n. nigricans* y *L. n. ervoides* (Ladizinsky *et al*., 1984). Aunque, más

tarde, el propio Ladizinsky estableció una nueva taxonomía que incluía *L. culinaris*, con las subespecies *L. c. culinaris* y *L. c. orientalis*, *L. odemensis*, *L. ervoides* y *L. nigricans*. (Ladizinsky, 1993). Posteriormente, se incluyeron dos nuevas especies: *L. tomentosus* (Ladizinsky, 1997) y *L. lamottei* (Van Oss *et al*., 1997).

También se han utilizado métodos moleculares para la clasificación del género. De esta forma, Ferguson *et al*. (2000), utilizando marcadores moleculares y morfológicos, y Zimniak-Przybylska *et al*. (2001), cuya nomenclatura filogenética se basa en SDS-PAGE (electroforesis en gel de poliacrilamida con dodecilsulfato sódico) de proteínas totales de semilla, llegaron a la conclusión de que tanto *L. odemensis* como *L. tomentosus* eran subespecies de *L. culinaris*. La clasificación del género *Lens* propuesta entonces fue, de nuevo, de cuatro especies: *L. culinaris*, con cuatro subespecies, *culinaris*, *orientalis*, *tomentosus* y *odemensis*, junto a *L. ervoides*, *L. nigricans* y *L. lamottei*.

Estudios más recientes basados en marcadores moleculares en el espaciador interno transcrito del ADNr (ITS) (Mayer y Bagga, 2002; Sonnante *et al*., 2003; Durán *et al*., 2004) (Tabla 1.1) sugieren:

- Características divergentes entre *L. nigricans* y el resto de las especies.
- Relación subespecífica entre *L. culinaris* y *L. orientalis*.
- *L. tomentosus* como taxón más cercano a *culinaris-orientalis*.
- *L. lamottei* especie más cercana a *L. ervoides*.
- *L. odemensis* suele aparecer unida a *tomentosus-culinaris-orientalis*.

Tabla1.1: Agrupaciones de las especies de *Lens*. Agrupaciones obtenidas tras la realización de diferentes análisis tanto moleculares como morfológicos. (Tomada de Cubero *et al*. 2009).

Método[1]	Clusters						Referencias
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6	
SDS-PAGE	*culinaris* *orientalis* *tomentosus* *odemensis*						**Zimniak-Przybylska *et al.* (2001)**
ITS	*culinaris* *orientalis*	*tomentosus* *odemensis* *ervoides* *lamottei*	*nigricans*				**Mayer y Bagga (2002)**
ITS + cpDNA	*culinaris* *orientalis*	*odemensis*	*ervoides*	*nigricans*			**Mayer y Bagga (2002)**
ITS	*culinaris* *orientalis*	*tomentosus* *odemensis* *lamottei* *ervoides*	*nigricans*				**Sonnanate *et al.* (2003)**
FISH	*culinaris* *orientalis*	*tomentosus*	*odemensis*	*lamottei*	*ervoides*	*nigricans*	**Galasso (2003)**
RAPD + ISSR	*culinaris* *orientalis*	*tomentosus*	*odemensis*	*lamottei*	*ervoides*	*nigricans*	**Durán *et al.* (2004)**
Cruzabilidad	*culinaris* *orientalis*	*odemensis* *nigricans*	*ervoides*				**Fratini *et al.* (2004)** **Fratini y Ruiz (2006)**
Morfología	*culinaris*$^{(m)}$ *orientalis* *odemensis* *nigricans*	*ervoides*	*culinaris*$^{(M)}$				**Fratini *et al.* (2006)**

[1] Abreviaturas utilizadas: cpDNA: ADN de cloroplasto; FISH: hibridación in situ con fluorescencia; ISSR: secuencias entre microsatélites; ITS: espaciador interno transcrito del rDNA; RAPD: ADN polimórfico amplificado al azar; SDS-PAGE: electroforesis en gel de poliacrilamida con dodecilsulfato sódico de proteínas totales de semilla. (m): microsperma; (M): macrosperma.

De esta manera, en la actualidad se habla de una clasificación consensuada de seis especies (Pérez de la Vega *et al*., 2011):

L. culinaris Medik.

L. culinaris. subsp *culinaris*; forma cultivada.

L. culinaris. subsp *orientalis* (Boiss.) Ponert; ancestro silvestre.

L. odemensis (Godr.) Ladiz.

L. ervoides (Bring.) Grande.

L. nigricans (Bieb.) Godr.

L. tomentosus Ladiz.

L. lamottei Czfr.

Los diferentes resultados obtenidos en los estudios indicados se deben al proceso evolutivo del género. Todas las especies del género comparten características morfológicas y bioquímicas y, además, los resultados dependen de las accesiones analizadas así como de los marcadores y caracteres estudiados (Cubero *et al*., 2009).

1.1.4. Variabilidad de la lenteja cultivada

Alefeld, en 1866, definió ocho subespecies de lenteja cultivada: *L. c. schniffspahni* (syn. *orientalis*), *L. c. himalayensis* (syn. *nigricans*), *L.c. punctata* (syn. *culinaris*), *L. c. hypochloris*, *L. c. nigra* (syn. *culinaris*), *L. c. vulgaris*, *L. c. nummularia* y *L. c. abyssinica.* Sin embargo, hasta 1930 no se realizaron estudios importantes sobre la especie. Es en ese año cuando Barulina publica el estudio más detallado y completo de la lenteja cultivada hecho hasta ese momento. Ella consideraba dos subespecies en función del tamaño de la semilla: *macrosperma* y *microsperma*, aunque existen evidencias moleculares que indican que dichas subespecies son, en realidad, variedades de la subespecie *L. c. culinaris*. También consideraba la distribución de una serie de caracteres que definen grupos regionales. Los caracteres que definen las subespecies en su estudio son: vainas, semillas, diferencias de tamaño en las flores y, secundariamente, tamaño de los foliolos y longitud y frondosidad de la planta. La identificación varietal se definía por otra serie de caracteres como son la dehiscencia, la longitud del cáliz y el número de flores por pedúnculo.

Lens culinaris presenta, frente a las especies silvestres, mayor tamaño de la planta, raquis más largo, mayor número de foliolos por hoja, mayor área de foliolo, más flores por pedúnculo, pedúnculos más cortos o iguales que el raquis, mayor frecuencia de flores blancas y semillas y vainas más grandes. (Cubero *et al*., 2009).

Existe una gran discusión sobre el origen de estas características ya que *L. odemensis* podría ser la responsable del mayor tamaño de la semilla y de otros caracteres relacionados característicos de las lentejas cultivadas localizadas en las zonas más occidentales, mientras que las variedades del este son más pequeñas lo que correspondería con una evolución del cultivo a partir de *L. orientalis*. El origen de la variedad pilosae (endemismo indio) podría haber tenido su origen en el sur de Turquía y el norte de Siria donde se han localizado las formas pilosas de *orientalis* y *culinaris*. El pequeño cáliz de *aethiopicae* podría tener su origen en *L. ervoides*. El origen de estas diferentes características también se puede atribuir a procesos clásicos de evolución de cultivos como son: mutación, selección, deriva genética y cruzamientos entre plantas.

Erksine *et al*. (1989) realizaron un estudio de la variación morfológica de las lentejas de los países con mayor producción y encontraron tres grupos regionales. El gran parecido que presentan las variedades de regiones adyacentes indica que las condiciones ambientales han influido de forma muy importante en la evolución de la lenteja cultivada así como la migración y el intercambio de semillas entre diferentes pueblos.

En resumen, la lenteja se domesticó en el Oriente Próximo y, más concretamente, en las montañas del sureste de Turquía y norte de Siria. El material de origen serían poblaciones de *L. orientalis* aunque los agricultores pudieron utilizar otras especies del género dando lugar a poblaciones mixtas. Sin embargo, el análisis de marcadores moleculares ha mostrado una variabilidad genética muy baja entre las diferentes variedades de lenteja cultivada (Sonnante *et al*., 2003; Durán *et al*, 2004) lo que indicaría que los morfotipos *macrosperma* y *microsperma* podrían ser, únicamente, variantes de caracteres cuantitativos resultantes de selección disruptiva (Sonnante *et al*.,

2003). Es difícil establecer el peso genético que han podido tener el resto de especies silvestres en el acervo genético del cultigen.

1.1.5. Importancia nutricional y agronómica

La lenteja se consume tradicionalmente como fuente de energía, proteínas y hierro además de por su contenido en fibra, minerales, vitaminas y antioxidantes aunque, además de estos componentes, también contiene sustancias antinutritivas como inhibidores de la proteasa, taninos, α-galactosidosacáridos o ácido fítico. Presenta un bajo contenido en grasas (0,7-4,3%), siendo más del 80% de las mismas insaturadas (Petterson *et al*., 1997; Urbano *et al*., 2007). Además, debido a su alto contenido en lisina y triptófano, su consumo (junto con arroz y trigo) proporciona un buen balance nutricional de aminoácidos para una correcta nutrición. También se utiliza en menor grado para la alimentación animal (Erksine *et al*., 1990).

En el año 2013, según la FAOSTAT (Organización de las Naciones Unidas para la Alimentación y la Agricultura), la lenteja se cultivó en 51 países en, aproximadamente, 4,9 millones de hectáreas a lo largo de los cinco continentes, este cultivo se realizó en prácticamente todas las regiones de la tierra exceptuando la Antártida, las regiones árticas, las zonas subtropicales y las grandes altitudes tropicales. Los mayores productores de lenteja fueron Canadá (1.880.500 Tm), India (1.134.000 Tm), Turquía (417.000 Tm), Australia (324.100 Tm), Estados Unidos (227.658 Tm), Nepal (226.931 Tm), China (150.000), Etiopía (129.833 Tm) y Siria (125.000 Tm). Sin embargo, los mayores productores no son, en muchos casos, los mayores consumidores lo que provoca un gran movimiento de producto entre diversos países. A continuación se muestra una tabla con los datos de los mayores exportadores e importadores a nivel mundial en el año 2012 ya que los datos comerciales sólo están publicados hasta dicha fecha.

Tabla1.2: Relación de los mayores exportadores e importadores de lenteja del mundo en el año 2012.

País	Exportación (Tm)	Importación (Tm)
Australia	**337.657**	2.004
Bangladesh	0	**171.000**
Canadá	**1.125.301**	11.083
Emiratos Árabes Unidos	**23.314**	**93.902**
Estados Unidos	**181.441**	34.394
India	779	**441.264**
Nepal	**35.823**	6.948
Sri Lanka	8.331	**114.713**
Turquía	**197.413**	**168.806**
España	2.100	49.906

España es el país desarrollado en el que, posiblemente, se consume una mayor cantidad de lentejas y de otras leguminosas grano por habitante dado que las leguminosas están muy bien valoradas y se incluyen en la típica dieta mediterránea. A pesar de ser uno de los mayores productores de lentejas en Europa, sus necesidades superan con creces esta producción, de manera que es necesario importar unas 50.000 Tm al año.

1.2. Los elementos transponibles

Los elementos transponibles (ET) se encuentran distribuidos en todos los genomas y tienen la capacidad de moverse de unos sitios a otros ya sea por un mecanismo de cortar y pegar (transposones) o por un mecanismo indirecto mediado por

un ARN intermediario (retroelementos). Este tipo de elementos fueron descubiertos por Bárbara McClintock (1950) en maíz, sus estudios se basaron en análisis citogenéticos y en la observación de los fenotipos de las plantas, principalmente, de las mazorcas en diferentes cruces.

Los ET se dividen en dos clases en función del tipo de transposición (Figura 1.3):

- Clase I o retroelementos: El ADN se transcribe a ARN que a su vez se retrotranscribe generando una copia de ADN, esta copia de ADN es la que se inserta de nuevo en el genoma huésped (transposición replicativa), de manera que el número de copias puede sufrir un gran incremento en el genoma a lo largo del tiempo.

- Clase II o transposones: en estos elementos es el propio ADN el que se moviliza e inserta. La transposición puede ser tanto replicativa como no replicativa en función de si el transposón se duplica o se mueve directamente de un lugar a otro (Lewin, 1997).

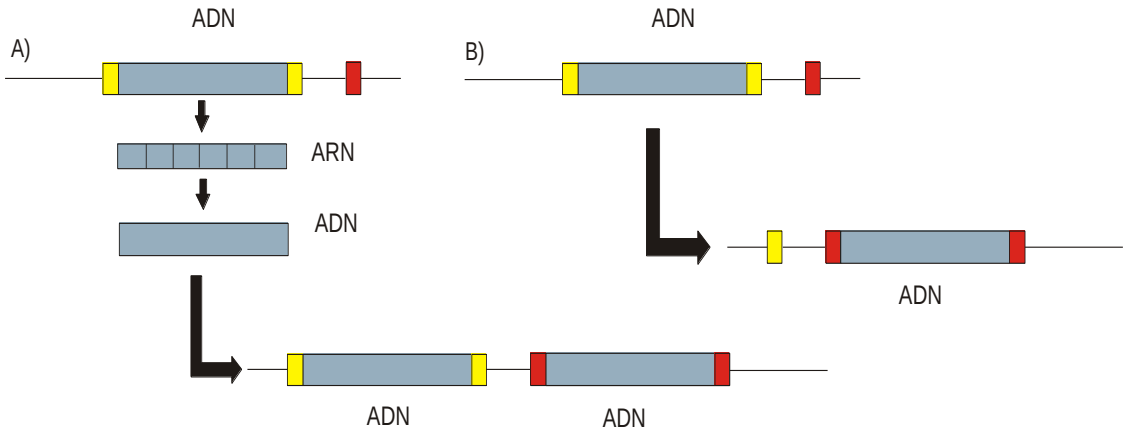


Figura1.3: Tipos de transposición: replicativa y no replicativa. A) Transposición replicativa mediada por un ARN intermediario propia de los retroelementos (clase I). B) Transposición no replicativa o conservativa sin productos intermediarios, transposones (clase II).

1.2.1. Elementos de clase I o retroelementos

Este grupo está formado por retrovirus, retrosecuencias (pseudovirus) y retrotransposones, todos ellos relacionados evolutivamente. A continuación se muestra

la clasificación para este tipo de elementos según Hansen y Heslop-Harrison (2004), dicha clasificación sigue el esquema de los seres vivos celulares.

Clase: Retroelementos
- Orden: Retrovirales
  - Suborden: Orthoretrovirineae (presentan función integrasa)
  - Suborden: Pararetrovirineae (no presentan función integrasa)
    - Familia: Hepadnaviridae (infecta animales; p.e. HBV)
    - Familia: Caulimoviridae (infecta plantas, p.e. CaMV)
  - Suborden: Retrotransposineae: Retrotransposones LTR
    - Familia: Pseudoviridae (*Ty1-copia*)
    - Familia: Metaviridae (*Ty3-gypsy*)
- Orden: Retrales
  - Suborden: Retroposineae: Retrotransposones no-LTR
  - Suborden: Retronineae: intrones mitocondriales de grupo II

El orden retrovirales está formado por elementos con una capacidad potencial para infectar como son los ortoretrovirus (suborden Orthoretrovirineae) cuyo genoma está formado por ARN que, una vez replicado, se inserta en el genoma de la célula infectada, y los pararetrovirus (suborden Pararetrovirineae) con un genoma de ADN, que necesitan la ayuda de un ARN intermediario para poder replicarse y que carecen de integrasas; así como por otros elementos sin capacidad infectiva pero derivados de los retrovirus, de manera que presentan un origen parásito, denominados retrotransposones (suborden Retrotransposineae) formados a su vez por dos familias: Pseudoviridae (*Ty1-copia*) y Metaviridae (*Ty3-gypsy*). Por otro lado, el orden Retrales lo forman elementos sin capacidad infectiva que se dividen en el suborden Retroposineae, formado por retrotransposones sin LTR (no-LTR) y el suborden Retronineae que incluye los intones mitocondriales de grupo II (Hansen y Heslop-Harrison, 2004).

Los retrotransposones son elementos genéticos móviles que, además de en otros muchos organismos, se pueden encontrar en todo el reino vegetal (Kumar y Bennetzen, 1999; Fedoroff, 2000). Los retrotransposones se mueven de un lugar del genoma a otro a través de un ARN intermediario que se convierte en ADN extracromosómico por la

acción de la enzima transcriptasa-inversa/RNasaH codificada, en la mayoría de los casos, por el propio retrotransposón (Bingham y Zachar, 1989; Boeke y Corces, 1989), para, posteriormente, insertarse en otro punto del genoma (Figura 1.3A). Este tipo de transposición puede dar lugar a mutaciones en el genoma del organismo celular si el elemento móvil se inserta cerca o dentro de un gen.

Los retrotransposones son la clase más abundante y extendida de elementos móviles en eucariotas (Kumar y Bennetzen, 1999). Existen dos tipos de retrotransposones en los genomas de eucariotas, los retrotransposones LTR, que presentan una secuencia repetida en ambos extremos (Long Terminal Repeat) y que aparecen en un gran número de copias en plantas con genomas grandes, más de un millón de copias totales por núcleo haploide, y los no LTR, que carecen de dicha secuencia repetida y que se encuentran en un número de copias menor, aproximadamente 250.000 (Kumar y Bennetzen, 1999).

- **Retrotransposones LTR**

En este grupo se encuentran los tipos *Ty3-gypsy* y *Ty1-copia* que contienen todas las funicones necesarias para su replicación autónoma (proteasa, integrasa, retrotranscriptasa y RNasaH). La principal diferencia entre los dos grupos es el orden de las secuencias codificantes en sus genomas. Se trata de elementos estructuralmente similares a los retrovirus pero sin capacidad infectiva. Además de éstos, también existen otros retrotransposones no autónomos como son los LARD (Large Retrotransposon Derivate) y los TRIM (Terminal Repeat Retrotransposon In Miniature) que no se encuentran incluidos en la clasificación de Hansen y Heslop-Harrison.

Grupo *Ty1-copia*: este tipo de retrotransposones pertenece a la familia Pseudoviridae y se encuentra presente en algas y en todo el reino Plantae, desde briófitas hasta gimnospermas y angiospermas (Flavell *et al*., 1974; Voytas *el al*., 1992). Estos elementos tienen una longitud que puede variar en un amplio rango, entre las 100 pb y las 5 kpb y están presentes en un número muy elevado de copias en el genoma, por ejemplo, *BARE-1* está presente en más de 50.000 copias en el genoma de cebada

(Suoniemi *et al.*, 1996A) y *Opie1* en más de 30.000 copias en el genoma de maíz (SanMiguel *et al*., 1996). Las secuencias terminales no codifican para ninguna proteína pero en ellas se encuentran los promotores y terminadores asociados con la transcripción. En el dominio interno se encuentra la región codificante formada por los genes *gag* y *pol.* El gen *pol* codifica para la retrotranscriptasa, la proteasa, la RNasaH y la integrasa, enzimas necesarias para la replicación, transposición e inserción de la forma ADN del retrotransposón en su nueva localización. El gen *gag* da lugar a proteínas relacionadas con la maduración y el empaquetamiento del ARN del retrotransposón para permitir su integración en el genoma del hospedador. (Kumar y Bennetzen, 1999).

Grupo *Ty3-gypsy*: este grupo, de la familia Metaviridae, también se encuentra ampliamente distribuido en el reino Plantae y también existen ejemplos de gran número de copias en los genomas huésped como es el caso de *Cinful-1* con 20.000 copias en maíz (SanMiguel y Bennetzen, 1998) y *Del1* con 13.000 en *Lilium henryi* o más de 40.000 en *L. longiflorum* (Leeton y Smith, 1993). Los retrotransposones de este tipo codifican las mismas funciones que los de tipo *Ty1-copia* y presentan una organización genómica muy similar con algunas modificaciones en el orden de los componentes del gen *pol*, como puede verse en las figuras 1.4 y 1.5.

LARD (Large Retrotransposon Derivate): se trata de elementos largos de entre 5,5 y 8,5 Kpb descritos por primera vez en cebada (Kalendar *et al*., 2004). Su dominio interno contiene ADN conservado no codificante que da lugar a una estructura secundaria importante para el ARN mensajero aunque aún no se sabe cuál es su función en el ciclo del retrotransposón. Parece ser que este tipo de retrotransposones necesita la ayuda de otro retrotransposón relacionado con los metavirus para su movilización (Havecker *et al*., 2004).

TRIM (Terminal Repeat Retrotransposon In Miniature): este tipo de elemento no autónomo se identificó por primera vez un intrón del gen de la uresasa de la patata (Witte *et al*., 2001). Son secuencias muy pequeñas, de unas 540 pb. Están

formados por unas LTR de entre 100 y 250 pb, sitios de unión para la transcriptasa inversa y un pequeño fragmento interno (Kalendar *et al*., 2008).

- **Retrotransposones no-LTR**

En este grupo se encuentran los LINE (Long Interspersed Nuclear Elements), los SINE (Short Interspersed Nuclear Elements) y los retropseudogenes.

LINE: estos elementos presentan uno o dos marcos de lectura abierto. Todos ellos codifican para una transcriptasa inversa y algunos, aunque no todos, para una endonucleasa y para una ribonucleasa H. Su longitud puede variar entre 100 y 1.000 pb (Kazazian, 2004).

SINE: se trata de elementos no autónomos. Las secuencias del extremo 5' derivan de un ARNt, ARN 5S o de ARN 7SL y presentan actividad promotora de la ARN polimerasa III mientras que el extremo 3' deriva de los LINE (Okada, 1991; Ohshima y Okada, 2005).

Retropseugenes: surgen por la retrotranscripción de ARNm procesados y la inserción del ADNc obtenido (Vanin, 1985). Estos elementos se distinguen del resto de pseudogenes por la pérdida de sus intrones, la presencia de una cola de poliA en el extremo 3' y por estar flanqueados por dianas de duplicación del punto de inserción (Gonçalves *et al*., 2000).

1.2.2. Elementos de clase II o transposones

Los transposones aparecen tanto en los genomas de plantas como en los de animales. Se trata de elementos móviles que utlizan un mecanismo de corta-pega para autopropagarse. Este mecanismo lo lleva a cabo la transposasa que reconoce las regiones TIR (Terminal Inverted Repeats) que contienen en ambos extremos. Existen elementos no autónomos que dependen de transposones activos para llevar a cabo su transposición, dos ejemplos de familias de transposones en plantas son *Ac/Ds* donde *Ac (Activator)* es el elemento autónomo capaz de activar a *Ds (Dissociation),* o *pSpm/dSpm*

donde *pSpm (Supresor-mutator)* activa el elemento defectivo *dSpm (defective Spm)* (McClintock, 1950; Huang *et al*., 2012).

Otro tipo importante de transposones no autónomos son los MITE (Miniature Inverted-repeat Transposable Elements). Se trata de elementos descritos por primera vez en plantas (Bureau y Wessler, 1992) que presentan un alto número de copias y cuyo origen parece estar relacionado con deleciones en transposones autónomos (Yang *et al*., 2009).

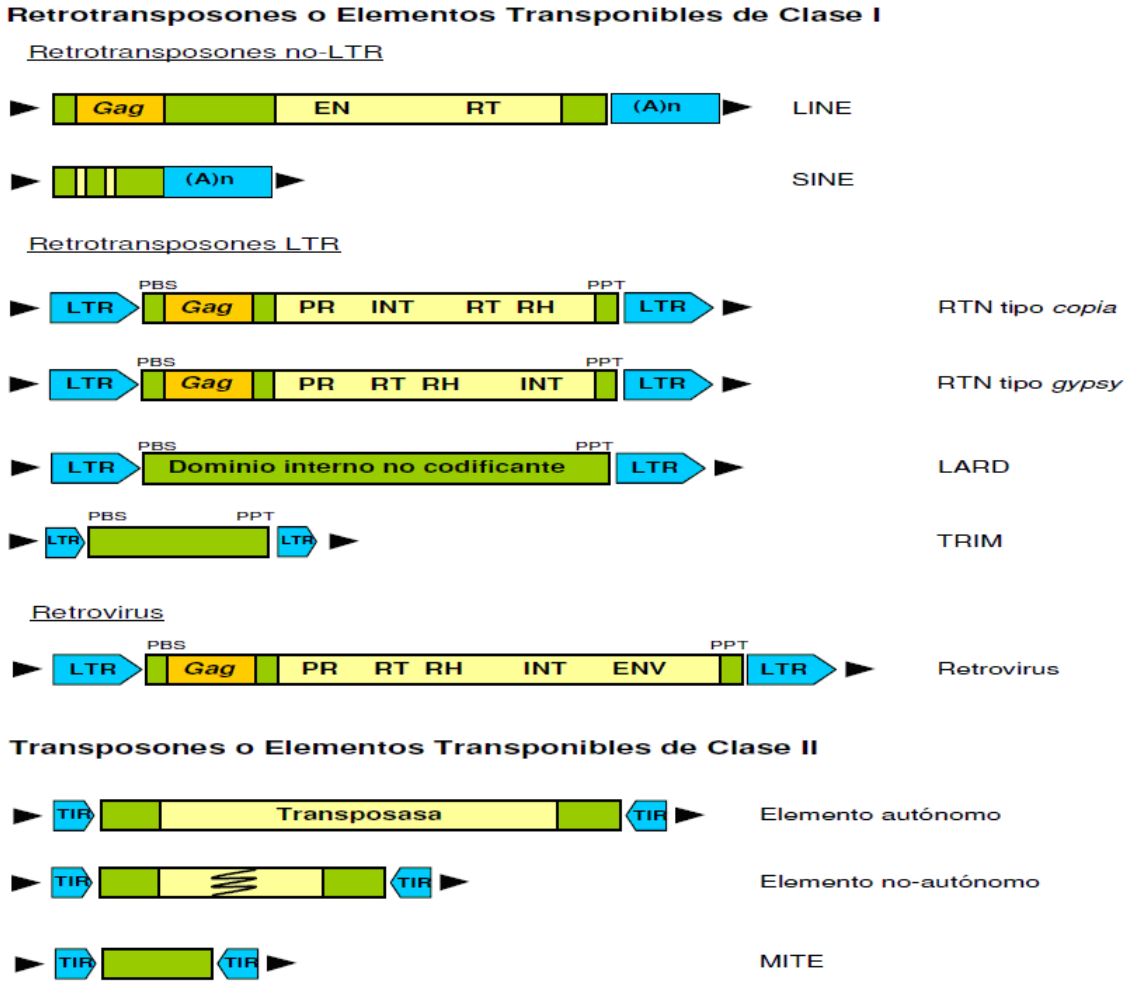


Figura1.4: Organización estructural general de los diferentes tipos de ET en plantas. EN: endonucleasa; RT: transcriptasa inversa; (A)n: cola de poliA; LTR: Long Terminal Repeat; PR: proteasa; INT: integrasa; RH: RNasaH; PBS: *Primer Binding Site*; PPT: *Poly Purine Track*; TIR: repeticiones invertidas terminales. (Tomado de Rico, 2005).

1.3. Los retrotransposones LTR

Este trabajo está centrado en los retrotransposones tipo LTR (RTN-LTR) que permiten el desarrollo de diferentes tipos de marcadores moleculares basados, principalmente, en la secuencia de dichas regiones terminales.

1.3.1. Características estructurales

Los RTN-LTR se caracterizan por la presencia de repeticiones largas terminales (LTR) en sus extremos, que flanquean el dominio interno que codifica las proteínas necesarias para su replicación y transposición (genes *gag* y *pol*). Junto al LTR5´ se encuentra la región PBS (Primer Binding Site) y la LTR3´ se sitúa junto a una región rica en purinas (PPT, Poly Purine Track). (Figura 1.5).

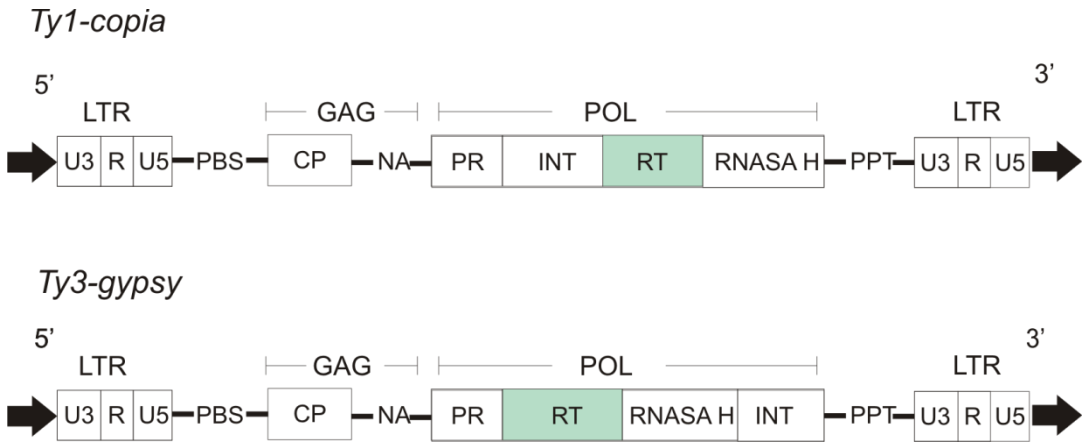


Figura1.5: Organización genómica de los retrotransposones tipo *Ty1-copia* y *Ty3-gypsy* (Tomado de Kumar y Bennetzen, 1999).

- **LTR**

Las LTR se encuentran situadas en los extremos, con la misma orientación y pueden tener una longitud de entre 100 pb (*Tos17*, Hirochika *et al*., 1992) y 5,2 Kpb (*Ogre*, Neumann *et al*., 2003).

Las LTR de cada familia de retrotransposones son diferentes y muy útiles para poder datar las inserciones en los genomas, ya que la secuencia de ambas LTR es idéntica en el momento de inserción pero, con el paso del tiempo, se producen diferentes tipos de mutaciones que hacen que difieran una de la otra (SanMiguel *et al*., 1998).

Ambas LTR presentan secuencias cortas invertidas en sus extremos (IR, Inverted Repeats). La secuencia consenso de las LTR es 5´ TG…CA 3´ aunque varía la longitud de la secuencia de unos retrotransposones a otros (Neuvéglise *et al*., 2002).

Además, las LTR de la mayoría de retrotransposones están flanqueadas por una duplicación de la diana de inserción denominada TSD (Target Site Duplication). Se trata de una secuencia de 5 pb de ADN del genoma huésped (Novikov *et al*., 2012).

Las LTR comprenden tres dominios: U3, R y U5 (Figuras 1.5 y 1.6). Estos componentes son iguales en ambas LTR aunque actúan de forma diferente ya que la LTR5´ es la que contiene la región reguladora de la transcripción (Flabet *et al.*, 2007).

- U3: esta zona se encuentra entre el extremo 5´ y el sitio de inicio de la transcripción. Contiene una secuencia promotora de la ARN polimerasa II, una caja TATA y un elemento iniciador que permiten su transcripción. Se han observado algunos retrotransposones en los que aparece un sitio alternativo de unión a una ARN polimerasa III.

- R: se encuentra entre U3 y U5. Generalmente presenta repeticiones en tándem

- U5: se extiende desde el sitio de poliadenilación, que indica en final de la transcripción, hasta el extremo 3´ de la LTR.

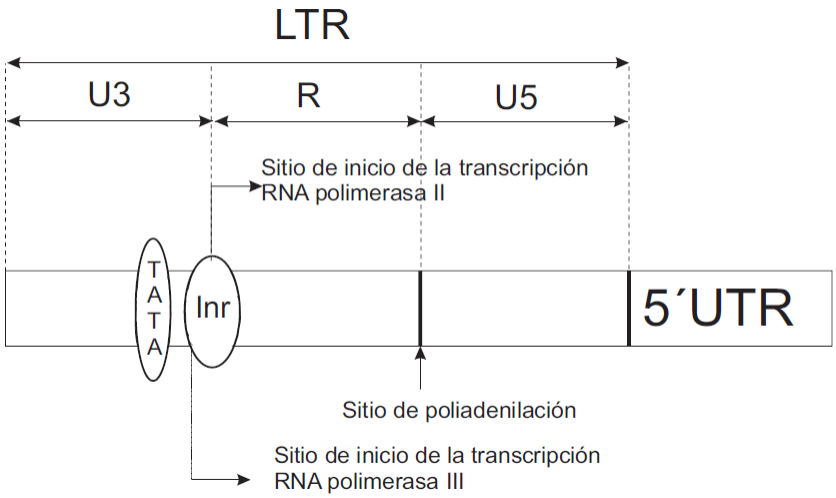


Figura1.6: Estructura típica de la LTR de un retrotransposón. Inr: iniciador. TATA: caja TATA (Tomado de Flabet *et al.*, 2007).

- **UTR (UnTraslated Region)**

También conocido como ULR (Untranslated Leader Region), se trata de una región no traducible que se encuentra entre las LTR y la región codificante del retrotransposón (Figura 1.6). La región 5´UTR de los retrotransposones es una región multifuncional que participa en la transcripción (Flabet *et al*., 2007).

- **PBS (Primer Binding Site)**

Se encuentra a continuación de la LTR5´ (Figura 1.5) y es el sitio de unión al cebador para llevar a cabo la síntesis de la cadena de ADNc (-).

Se trata de una secuencia de entre 8 y 49 pb complementaria a un ARNt específico del genoma huésped. De forma general y dependiendo de la familia del retrotransposón, esta secuencia puede ser complementaria al extremo 3´ del brazo aceptor del ARNt o a 13 nucleótidos internos del ARNt que incluyen el anticodón, aunque existen excepciones a esta norma como son los retrotransposones *Tca1* y *Tca2* de *C. albicans* que presentan una región PBS complementaria a una secuencia interna diferente al anticodón del ARNt (Neuvèglise, 2002).

El ARNt, que participa en la retrotranscripción, más común en plantas es el de metionina, cuya secuencia complementaria al PBS es TGGTATCAGAGC, aunque se han descrito retrotransposones complementarios a otros transferentes distintos, como el ARNt de ácido glutámico en *Bagy2* de cebada y *Cyclops2* de guisante o el ARNt de arginina en *Grande* de maíz y *RIRE2* de arroz (Rico, 2005).

- **PPT (PolyPurine Track)**

Esta secuencia rica en purinas se encuentra junto al dominio U3 de la LTR3´ (Figura 1.5). Es el sitio de unión del cebador para la síntesis de la cadena (+). Se forma durante la síntesis de la cadena (-) de ADN y se mantiene tras la degradación del ARN debido a su resistencia frente a la RNasaH por lo que sirve de molde para la retrotrancriptasa. (Nair *et al*., 2012).

- **Dominio interno**

Entre las dos LTR se encuentra la zona codificante formada por ORFs (Open Reading Frame) o marcos de lectura abierta que se corresponden con los genes *pol* y *gag*.

La región codificante del gen *gag* está muy poco conservada, aunque suele presentar una zona que codifica para una región rica en cisteínas (CCHC) en todos los retrotransposones, mientras que la del gen *pol* presenta importantes similitudes tanto en los retrotransposones como en los retrovirus (Rashkova *et al*., 2002).

Existe un primer codón de terminación entre los dos genes y otro a continuación del gen *pol*. Según el mecanismo utilizado para superar la primera señal de parada se han identificado cinco tipos de organizaciones *gag/pol* en los retrotransposones (Lewin, 1997; Gao *et al*., 2003):

1) Obtención de una única poliproteína que se procesará para generar las proteínas funcionales. Esto ocurre por la presencia de un glutamil-tRNA que reconoce el primer codón de fin permitiendo que continúe la traducción. Casi la mitad de los retrotransposones presenta esta organización.
2) Obtención de dos poliproteínas diferentes a partir de los genes *gag* y *pol*. En este caso el primer codón de fin es reconocido normalmente. Esta organización se ha descrito en el retrotransposón *RIRE2* de arroz (Ohtsubo *et al*., 1999) y en el elemento *BELKamikaze* de *Bombyx mori* (Abe *et al*., 2001).
3) *gag* y *pol* separados en tres o más marcos de lectura abiertos. Parece ser que este tipo de organización se debe a mutaciones. Tanto en este tipo de organización como en los dos siguientes, el primer codón de fin ha desaparecido, la síntesis de dos poliproteínas diferentes depende de un cambio en el marco de lectura de ambos genes.
4) *pol* en el marco de lectura -1 con relación a *gag*. Esta estructura sólo aparece en los retrotransposones tipo *Ty3-gypsy*y *DIRS* de animales.

5) *pol* en el marco de lectura +1 con relación a *gag*. Esta organización se encuentra en muy pocos retrotransposones tanto de animales como de plantas y hongos.

Todos los retrotransposones tienen en común un exceso de expresión de *gag* frente a *pol* ya que necesitan mucha más cantidad de proteínas estructurales (proporción 20:1) que permiten la viabilidad del elemento, de manera que desarrollan diferentes estrategias de regulación post-transcripcional que determinan su organización estructural (Forbes *et al*., 2007).

Estos genes codifican para todas las proteínas necesarias para llevar a cabo el ciclo de los retrotransposones. El gen *gag* codifica para proteínas estructurales de la matriz, nucleocápsida y de la cápsida necesarias para formar una partícula similar a la viral (VLP, Virus Like Particle), mientras que el gen *pol* codifica para las enzimas proteasa, integrasa, transcriptasa inversa y ribonucleasaH.

- **Proteasa:** esta enzima es la encargada de procesar la poliproteína para obtener proteínas funcionales. Es un proteasa de tipo aspártico que, en los retroelementos, se caracteriza por presentar un motivo D(S/T)G en el centro activo (Peterson-Burch y Voytas, 2002). El primer paso del procesamiento es el corte de los enlaces peptídicos de los extremos de la propia proteasa de manera que ésta se escinde de la poliproteína y, a continuación, rompe el resto de los enlaces dando lugar primeramente a la transcriptasa inversa y la integrasa y después al resto de proteínas de forma secuencial (Goodenow *et al*., 2002).

- **Transcriptasa inversa:** esta enzima es la encargada de formar el ADNc a partir del ARN del retrotransposón para lo cual presenta actividad ADN-polimerasa dependiente de ARN (copia el ARN a ADNc, cadena -) y actividad ADN-polimerasa dependiente de ADN (sintetiza la segunda cadena de ADNc, cadena +) (Rico, 2005). La secuencia completa de la proteína se mantiene conservada entre retrotransposones de organismos muy distantes filogenéticamente (Flavell *et al*., 1992).

- **RibonucleasaH:** se encarga de degradar el ARN del heterodúplex ADN-ARN, que se forma durante la retrotranscripción de manera no específica en presencia de cationes divalentes, para poder sintetizar la segunda cadena de ADN. Esta actividad se encuentra en la misma retrotranscriptasa, más concretamente en el extremo C-terminal y tiene actividad tanto endo como exonucleasa 3´ y 5´ (Rico, 2005). Se han descrito dos familias (tipo1 y tipo2) en función de sus secuencias aminoacídicas y sus propiedades bioquímicas. Ambas familias se caracterizan por la presencia de cuatro sitios activos, donde se localiza la actividad catalítica, que son tres residuos de ácido aspártico y uno de ácido glutámico (denominado motivo DEDD) (Goedken y Marqusee, 2001; Nowotny *et al*, 2005). Además, en los retrotransposones *Ty1-copia*, se han descrito dos motivos conservados que se han utilizado en estudios filogenéticos y en el diseño de cebadores para la búsqueda de retrotransposones en el genoma de diversos organismos (Pearce *et al*., 1999).

- **Integrasa:** permite la integración del ADN complementario del retrotransposón en el genoma huésped. Presenta tres elementos conservados: un motivo de dedo de Zinc en la región N-terminal implicada en la actividad de la enzima, una región central en la que se encuentra la actividad catalítica de la enzima (CCD, Catalytic Core Domain), muy conservada en retrovirus y retrotransposones y caracterizada por la presencia del motivo DD(35)E, formado por dos residuos de ácido aspártico y uno de ácido glutámico, separado del residuo anterior por 35 aminoácidos, y la región C-terminal que presenta un dominio de unión inespecífica al ADN (Jaskolski *et al*, 2009; Qi *et al*., 2013).

- **Gen *env***

La diferencia entre los retrovirus y los retrotransposones es la presencia o ausencia del gen *env* que codifica para un grupo de glicoproteínas que reconocen receptores proteicos en la superficie de las células hospedadoras permitiendo su unión y

entrada a través de la membrana celular del huésped para su infección (Vicient *et al*., 2001).

La presencia de la pared celular en las plantas hace que las proteínas codificadas por el gen *env* no sean tan efectivas como en las células animales, lo que explicaría el escaso número de retrovirus en plantas, sin embargo, existen algunos retrotransposones en los que aparece un elemento extra en el extremo 3´ similar al gen *env* de los retrovirus, como es el caso de algunos retrotransposones de la familia *Ty3-gypsy* como *Athila/Tat1* de *Arabidopsis thaliana* (Wright y Voytas, 1998), *Cyclops* de guisante (Chavanne *et al*., 1998), *Calypso* de soja (Peterson-Burch *et al*., 2000), *Bagy2* de cebada o *Rigy2* de arroz (Vicient *et al*., 2001). En la familia *Ty1-copia* sólo se ha encontrado un grupo que presente esta proteína, el retrotransposón *SIRE1* de soja considerado un proretrovirus (Laten *et al*., 1998).

1.3.2. Ciclo de movilización

El primer paso en el ciclo de los retrotransposones (Figura 1.7), es la transcripción del ADN mediante la acción de una polimerasa II obteniéndose un ARNm que se traduce a continuación y da lugar a las poliproteínas GAG y POL. La proteasa codificada en el propio elemento es la encargada de procesar los polipéptidos para obtener las proteínas funcionales. La proteína GAG contiene tres dominios: cápsida, nucleocápsida y matriz. La cápsida tiene función de polimerización que permite la formación de la partícula similar a un virus (VLP) en cuyo interior tendrá lugar la transcripción inversa. El mecanismo por el cual el ARN se introduce en la VLP se denomina empaquetamiento y se produce tras el reconocimiento de una determinada secuencia del ARN (PSI: packaging signal o señal de empaquetamiento) por la nucleocápsida.

El ARN de los retrovirus, generalmente, aparece en forma de dímero en el interior del VLP, esta dimerización se produce antes o simultáneamente con el empaquetamiento. La dimerización se ha observado en algunos retrotransposones como *Ty1* de *Saccharomyces cereviseae*.

Una vez empaquetado (Figura 1.8), la región PBS es reconocida por un ARNt y la retrotranscriptasa inicia la síntesis de una cadena (-) corta complementaria a los dominios R-U5 de la LTR5´ donde se encuentra con una señal de parada. La RNasaH degrada la región transcrita y la cadena de ADNc sintetizada es transferida al extremo 3´ por complementariedad con el dominio R de la LTR3´ desde donde continúa la síntesis de la cadena (-) hasta llegar a la señal de parada. A medida que avanza la síntesis, la RNasaH va degradando el ARN del híbrido ARN-ADN formado. La síntesis de la cadena (+) comienza con la unión del cebador al PPT, se sintetizan las regiones U3-R-U5 y el PBS y se produce un nuevo salto al extremo 5´ donde presenta complementariedad con el PBS y finaliza la síntesis con la obtención de una doble cadena de ADNc con dos LTR idénticas en ambos extremos.

A continuación (Figura 1.7), se produce la entrada del VLP en el núcleo para proceder a la integración. La integrasa se une a la LTR de una forma concreta en función de determinadas secuencias localizadas en sus extremos. Esta enzima corta la doble cadena del ADN huésped de forma asimétrica dando lugar a extremos cohesivos que serán reparados pon un sistema endógeno (Sabot y Schulman, 2006).

El mecanismo de transposición replicativa de los retrotransposones los convierte en elementos muy invasivos, de manera que es imprescindible el control de distribución y número de copias para disminuir al máximo posible los efectos adversos en el genoma huésped.

Ya que los retrotransposones no se pueden transponer sin la presencia de un ARN molde, la forma más fácil de controlar su actividad es mediante una regulación a nivel transcripcional bloqueando su transcripción. En plantas, y en otros organismos, se conocen muchos ejemplos de retrotransposones silenciados en los tejidos somáticos durante el desarrollo, o que son activados por hormonas o factores de desarrollo y asociados a modificaciones epigenéticas (Boeke y Corces, 1989; Ding y Lipshitz, 1994; Law y Jacobsen, 2010).

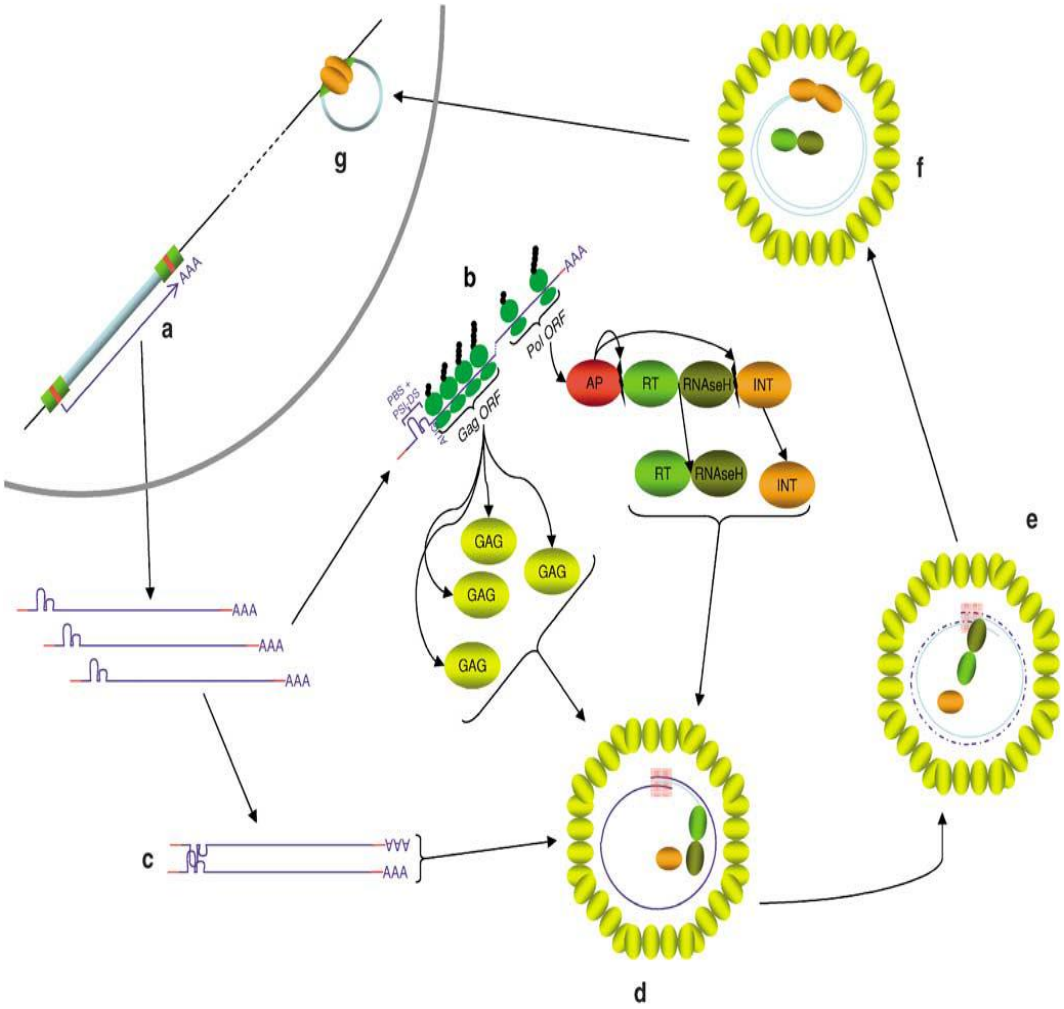


Figura1.7: Ciclo de los retrotransposones LTR. a) Transcripción del ARNm. b) Traducción y síntesis proteica. c) Dimerización del ARN. d) Empaquetamiento del ARN e inicio de la retrotranscripción. e) Degradación del ARN e inicio de la síntesis de la segunda cadena de ADNc. f) Fin de la síntesis de la doble cadena de ADNc y unión de la integrasa a las LTR. g) Rotura de la doble cadena de ADN nuclear e integración de la nueva copia en una zona diferente del genoma. (Tomado de Sabot y Schulman, 2006).

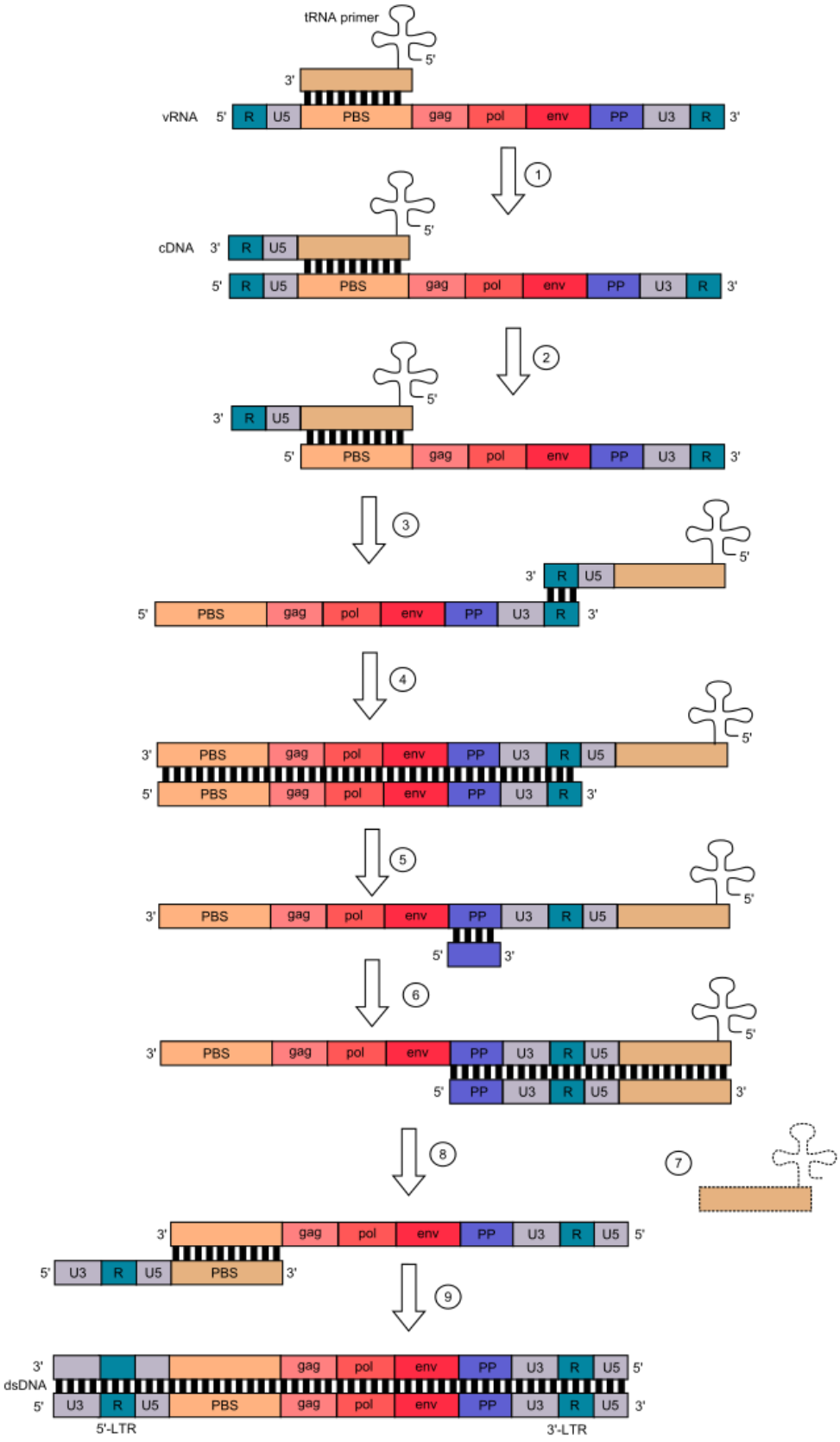


Figura1.8: Esquema de la retrotransposición. Los diferentes pasos del proceso están explicados en el texto.

Hasta finales de los años 90 la actividad transcripcional de los retrotransposones sólo se había podido demostrar en una pequeña cantidad de elementos, tanto del tipo *Ty1-copia* como *Ty3-gypsy*, en diversas especies como el tabaco (Grandbastien *et al*., 1989; Pouteau *et al*., 1991), maíz (Varagona *et al*., 1992), o el trigo (Moore *et al*., 1991), todos ellos aislados tras su transposición. Sin embargo, más adelante se pudo demostrar la presencia de transcritos que comenzaban con elementos de LTRs en especies como maíz (Avramova *et al*., 1995; San Miguel *et al*., 1996; Turcich *et al*., 1996), cebada (Suoniemi *et al*., 1996B), patata (Pearce *et al*., 1996), soja (Bi y Laten, 1996), tabaco (Grandbastien *et al*., 1997; Takeda *et al*., 1998) y arroz (Hirochika *et al*., 1996A).

Los análisis de secuencia realizados en retrotransposones de diversas especies de plantas han mostrado una gran cantidad de elementos que contienen errores en los marcos de lectura, de manera que dichos elementos se encuentran inactivos (Flavell *et al*., 1992; Voytas *et al*., 1992; Hirochika e Hirochika, 1993). Sin embargo, también se han localizado otros elementos que sí se transcriben dentro de la planta (Pouteau *et al*., 1991; Hirochika *et al*., 1996A; Pearce *et al*., 1997; Suoniemi *et al*., 1996B). La transcripción se encuentra inactiva durante el desarrollo normal de la planta pero puede ser inducida por estreses tanto bióticos como abióticos, incluyendo cultivos celulares, heridas o ataques de patógenos (Wessler, 1996; Grandbastien, 1998).

La metilación del ADN del huésped también ha demostrado ser una forma de control de la expresión reduciendo la capacidad de multiplicación de retrotransposones y retrovirus ya que la metilación inhibe la transcripción y además es heredable epigenéticamente (Yoder *et al*., 1997; Law y Jocabsen, 2010).

Algunos retrotransposones pueden ser activados tras el aislamiento de protoplastos o mediante cultivos celulares o de callos. Ambos procesos inducen modificaciones en el metabolismo de las células y en la expresión génica mediante la activación de genes relacionados con el estrés y el crecimiento, debido a la hidrólisis de la pared celular en el caso del aislamiento de protoplastos o a que la respuesta observada

en el cultivo celular o de callo es similar que ante la presencia de una herida (Todorovska, 2007).

La primera evidencia de la activación de un retrotransposón debido a estrés se observó al detectar la transposición de *Tnt1A* en el gen estructural de la nitrato reductasa en plantas de tabaco regeneradas a partir de protoplastos derivados de cultivos celulares (Grandbastien *et al*., 1989), también se observó que la expresión de este mismo elemento se inducía tras el aislamiento de protoplastos de una hoja de tabaco (Pouteau *et al*., 1991).

También se han observado casos de regulación postranscripcional en hongos, como es el caso de las proteínas FUS3P y TFIIH presentes en el genoma huésped. FUS3P es capaz de inhibir la acumulación de las proteínas expresadas por *Ty1* y TFIIH degrada los ADNc. (Kumar y Bennetzen, 1999).

1.3.3. Efecto de los retrotransposones en el genoma de plantas

Los retrotransposones han invadido con éxito prácticamente todas las partes del genoma de plantas, llegando a ocupar hasta el 80% del ADN nuclear como en el caso del genoma de maíz (SanMiguel *et al*., 1998). Los retrotransposones se encuentran dispersos por todo el genoma aunque se han observado casos en los que su distribución está determinada por la familia de retrotransposones y la especie estudiada. Por ejemplo, los retrotransposones tipo *Ty3-gypsy* aparecen agrupados en regiones centroméricas en maíz, arroz o trigo mientras que los *Ty1-copia* de *Allium* aparecen en la zona heterocromática de los extremos de los cromosomas. (Kumar y Bennetzen, 1999).

El análisis de secuencias de retrotransposones obtenidas en diversos estudios muestra una gran heterogeneidad de las mismas en los genomas de plantas, mucho más alta que en *Drosophila* y en levaduras (Flavell *et al*., 1992). Esta heterogeneidad sugiere que dichos elementos ya aparecían en las primeras etapas de la evolución de las plantas y que han ido evolucionando de forma diferencial hasta llegar a las especies actuales (Gribbon *et al*., 1999).

Una de las razones más importantes de la variabilidad en el tamaño de los genomas de plantas es la presencia de un número mayor o menor de retrotransposones. Por ejemplo, se ha demostrado que la presencia de retrotransposones de una única familia es la responsable de un aumento del tamaño del genoma de *Vicia* de hasta un 50% entre diferentes especies. Así, *Ogre* está presente en todas las especies de *Vicia* en proporciones que varían entre el 38% de *Vicia pannonica* que presenta un genoma haploide de 6615 Mpb y el 0,1% en *Vicia lathyroides* cuyo genoma haploide contiene 2573 Mpb (Neumann *et al*., 2006).

Los retrotransposones parecen haber tenido un importante papel en la evolución de la estructura y regulación de genes en plantas. Se han localizado genes de plantas que presentan, en sus regiones flanqueantes, secuencias de retrotransposones con función reguladora (White *et al*., 1994).

1.4. Marcadores moleculares basados en retrotransposones

Un marcador genético es cualquier diferencia fenotípica controlada genéticamente y utilizada en el análisis genético (Aranzana *et al*., 2000). Existen unos criterios que deben cumplir los marcadores moleculares para poder ser utilizados (Agarwal *et al*., 2008):

- Ser polimórfico y distribuirse uniformemente por todo el genoma.
- Generar múltiples marcadores independientes y fiables.
- Proporcionar una adecuada resolución de las diferencias genéticas.
- Permitir un análisis simple, rápido y barato.
- Requerir pequeñas cantidades de tejido u otro tipo de muestras.
- Vincularse con fenotipos diferentes.
- Ser neutro y no epistático.
- Ser codominante.
- No requerir información previa acerca del genoma de un organismo.

El uso de marcadores está muy extendido en la actualidad en los análisis genéticos y en la mejora vegetal debido a sus múltiples utilidades como pueden ser, la obtención de "huellas genéticas" (fingerprinting) de individuos, variedades y poblaciones, el análisis de la estructura y diversidad genética en poblaciones, el establecimiento de relaciones filogenéticas entre diferentes individuos y especies, la construcción de mapas genéticos de alta densidad y la localización de genes de interés económico, el mapeo de QTL (Quantitative Trait Loci) y la mejora asistida por marcadores (MAS, Marker Assisted Selection).

Los marcadores moleculares se pueden dividir en dos grupos: marcadores bioquímicos y marcadores de ADN. Los primeros se basan en el uso de proteínas sin actividad enzimática o de isoenzimas o aloenzimas mientras que, los segundos, utilizan fragmentos de ADN y han servido para solucionar muchos de los problemas no resueltos por los marcadores bioquímicos.

Las proteínas derivan directamente de los genes mediante los procesos de transcripción, procesamiento y traducción. Las isoenzimas, descubiertas en 1957 por Hunter y Markert, son diferentes variantes de una misma enzima, presentes en una especie y que pueden presentar diferentes propiedades. Las técnicas para su estudio son sencillas y no requieren equipos de laboratorio complejos. El uso de estos marcadores consiste, simplemente, en una separación electroforética y en una tinción específica de manera que se obtendrán bandas a diferentes alturas del gel cuando existan diferencias en la composición de aminoácidos o en la estructura de la proteína (Cubero, 2003).

Las principales ventajas que presentan las isoenzimas son la capacidad de relacionar los resultados con los genes que las codifican, que suelen tener una base genética codominante y que, generalmente, son neutros en cuanto a la selección.

Por otro lado, el uso de este tipo de marcadores ha ido disminuyendo debido a los problemas técnicos que presentan en muchas ocasiones. Además, no permiten analizar genomas completos y sólo detectan genes que se expresen en algún lugar o momento determinado. Los resultados pueden variar en función del tejido analizado y,

además, existe polimorfismo ontogénico, de manera que los resultados pueden variar entre individuos jóvenes y adultos.

Los marcadores de ADN se desarrollaron como una nueva generación de marcadores moleculares que ha servido para solucionar muchos de los problemas no resueltos por las isoenzimas. Los marcadores genéticos se pueden dividir en dos grupos en función de las técnicas empleadas para su análisis:

- Técnicas que emplean hibridación tipo "Southern".
- Técnicas basadas en PCR.

Las técnicas de hibridación tipo "Southern" permiten analizar variaciones en la longitud o tamaño de los fragmentos de ADN producidos por la acción de enzimas de restricción específicas. Los fragmentos obtenidos tras la acción de dichas enzimas se separan por electroforesis en geles de agarosa y, a continuación se transfieren, por capilaridad o mediante vacío, a una membrana de nitrocelulosa o de nylon (esto es lo que se denomina transferencia tipo "southern"). Por último, se detectan las bandas mediante la hibridación de una sonda marcada que permite la visualización de las bandas hibridadas.

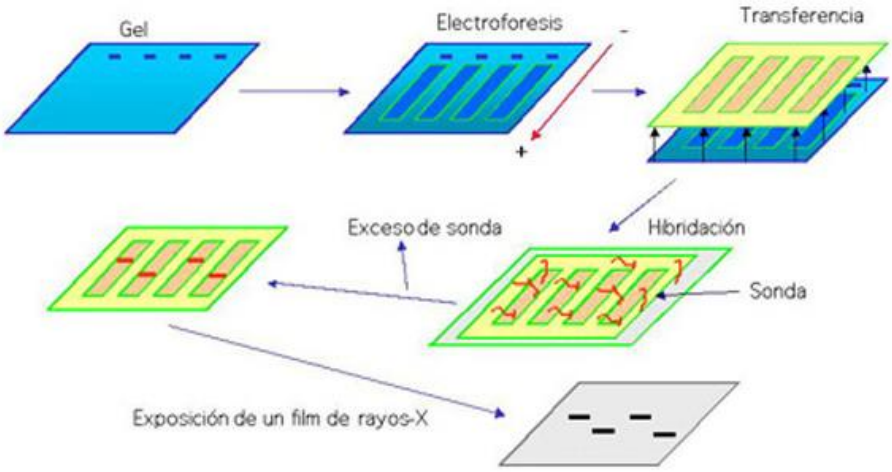


Figura1.9: Esquema de una técnica de marcadores moleculares tipo "southern".

Dentro de esta técnica se encuentran los marcadores RFLP (Restriction Fragment Length Polymorphism), que son los primeros marcadores de ADN utilizados y que se basan en las diferencias en el tamaño de los fragmentos de ADN obtenidos tras la acción de enzimas de restricción (Jeffreys, 1979) y los VNTR (Variable Number Tandem Repeat), en los que las diferencias se deben al número de repeticiones de secuencias repetidas en tándem de entre nueve y cien pares de bases (minisatélites) que varía entre unos individuos y otros (Jeffreys *et al*., 1985).

La técnica de PCR, o reacción en cadena de la polimerasa, fue desarrolla por Kary B. Mullis en 1987 y permite la multiplicación in vitro de fragmentos específicos de ADN de manera que permite la detección de secuencias o genes de interés en el genoma de un individuo.

Dicha técnica se basa en la amplificación de fragmentos de ADN a partir de secuencias de nucleótidos denominadas cebadores, que son capaces de reconocer secuencias blanco de las que son complementarias. Cada ciclo de una reacción de PCR incluye tres pasos:

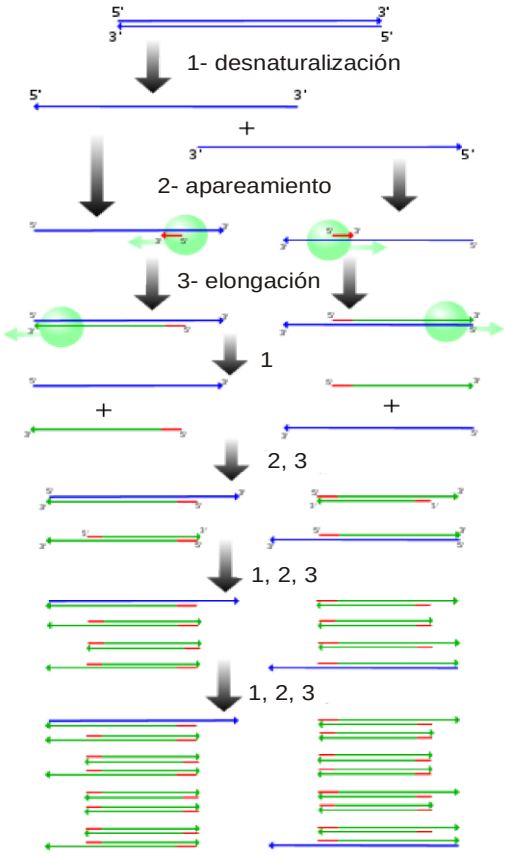


Figura1.10: Esquema general de una PCR.

- Desnaturalización: consiste en la separación de las dos hebras que componen la doble hélice del ADN. Se consigue calentando la muestra hasta alcanzar 95^{o}C.

- Apareamiento: unión de los cebadores a las zonas complementarias.

- Elongación: en este paso, una ADN polimerasa termoestable sintetiza la cadena complementaria a la cadena molde utilizando el cebador como iniciador para la reacción.

Este proceso se lleva a cabo en un termociclador que permite realizar los cambios de temperatura necesarios para llevar a cabo el proceso descrito. Los ciclos se repiten un número de veces determinado para conseguir la cantidad de ADN deseada en cada caso.

En este grupo se incluye una gran variedad de marcadores:

- RAPD (Random Amplified Polimorphic DNA): esta técnica permite amplificar regiones arbitrarias del genoma usando un único cebador de secuencia aleatoria de diez pares de bases (Welsh y McClelland, 1990; Williams *et al*., 1991):
- SSR (Simple Sequence Repeats): los microsatélites son secuencias de entre dos y seis pares de bases que se repiten de forma consecutiva y la variación en el número de repeticiones entre individuos produce diferentes alelos (Tautz y Renz 1984).
- ISSR (Inter-Simple Sequence Repeats): se trata de marcadores basados en la amplificación de regiones que se encuentran entre microsatélites lo que implica el conocimiento previo de la secuencia del genoma (Godwin *et al*., 1997).
- AFLP (Amplified Fragment Length Polymorphism): esta técnica se basa en la restricción del genoma con enzimas de restricción, seguida de una ligación de las secuencias adaptadoras de dichas enzimas y la posterior amplificación de los fragmentos obtenidos utilizando cebadores complementarios a los adaptadores de las enzimas utilizadas (Vos *et al*., 1995)
- Restricción con enzimas tipo CEL: esta técnica se basa en el uso de endonucleasas de restricción de tipo CEL, que se caracterizan por reconocer fragmentos de cadena única, para detectar mutaciones en el genoma, esta enzima es capaz de reconocer desapareamientos producidos por cambios en una única base por lo que se puede utilizar en estudios de SNP (Simple Nucleotide Polymorphism) (Oleykowski *et al*., 1998; Sobrino *et al*, 2005; Galeano *et al*., 2009).
- CAP (Cleaved Amplified Polimorphic Sequences): técnica derivada de los RFLP en la que se sustituye el "southern blott" por una PCR. Se identifican polimorfismos en longitud de fragmentos de restricción utilizando las endonucleasas

para digerir fragmentos de ADN amplificados mediante PCR (Konieczny y Ausubel, 1993).

- SRAP (Sequences Related Amplified Polymorphism): se trata de la amplificación de regiones correspondientes a marcos de lectura abiertos (ORFs) (Li y Quirós, 2001).

- HRM (High Resolution Melting): se trata de una técnica que permite detectar SNP que no se encuentran en dianas de enzimas de restricción. Consiste en realizar una PCR a tiempo real con un colorante fluorescente de manera que, tras la amplificación, se aumenta la temperatura para que las dos hebras de ADN se desnaturalicen y liberen la fluorescencia de manera que se puedan detectar diferencias en las temperaturas de fusión de los amplicones (Wittwer *et al*., 2003)

Los marcadores de ADN presentan muchas ventajas ya que no se ven afectados por el ambiente, están presentes en cualquier momento del desarrollo del individuo, permiten una detección temprana, son universales, muy abundantes, requieren poca cantidad de ADN para los análisis y el ADN es muy estable y específico para cada individuo. Sin embargo, son relativamente costosos y algunos requieren equipos relativamente sofisticados para su análisis.

La presencia de retrotransposones en todos los grupos de plantas estudiados así como por tratarse de elementos repetitivos, permite que se puedan desarrollar marcadores genéticos a partir de sus secuencias, más concretamente, a partir de los extremos LTR. El análisis de este tipo de marcadores permite la obtención de un gran número de polimorfismos entre individuos (Waugh *et al*., 1997; Ellis *et al*. 1998; Porceddu *et al*., 2002; Tam *et al*., 2005). La posición de las inserciones de los retrotransposones son irreversibles y permanecen estables durante millones de años (Schulman *et al*., 2004; Vitte *et al*., 2004; Jing *et al*., 2005) lo que permite hacer estudios filogenéticos partiendo de un estado ancestral carente de inserciones.

En la actualidad se utilizan diversos tipos de marcadores genéticos basados en retrotransposones (Kalendar *et al*., 2011). Estos marcadores se basan en la secuencia de

las regiones LTR de los retrotransposones (Schulman, 2007; Agarwal *et al*., 2008) y de otras regiones internas conservadas (Kalendar *el al.*, 2010). Las más extendidas son:

- SSAP (Sequence Specific Amplified Polymorphism) (Figura 1.11A): esta técnica, también conocida como "transposon display" si se utiliza con transposones (van den Broeck *et al*., 1998), es una modificación de los AFLPs y es la técnica de marcadores derivados de retrotransposones más utilizada hasta el momento. Consiste en la amplificación de secuencias que se encuentren entre la región LTR de un retrotransposón y una diana de restricción a la que se liga un adaptador complementario al cebador. Este método combina la alta resolución de los AFLPs con la especificidad del cebador obtenido a partir de la secuencia dela LTR (Waugh *et al*., 1997; Tam *et al*., 2005). Esta técnica se ha utilizado hasta el momento con una gran variedad de especies como son cebada (Waugh *et al*., 1997), guisante (Ellis *et al*., 1998; Flavell *et al*., 1998; Pearce *et al*., 2000), trigo (Gribbon *et al*., 1999; Queen *et al*., 2004), alfalfa (Porceddu *et al*., 2002), alubia (Galindo *et al*., 2004), anacardo (Syed *et al*., 2005), tomate y pimiento (Tam *et al*., 2005), lechuga (Syed *et al*., 2006), haba (Martín Sanz *et al*., 2007) y vid (Castro *et al*., 2012). Todos estos estudios muestran que el desarrollo de un sistema de marcadores SSAP depende, directamente, del tipo de retrotransposones encontrados en cada caso. El principal inconveniente que presenta esta técnica es la necesidad de analizar los datos mediante electroforesis capilar, más compleja y cara que la electroforesis en gel de agarosa que se utiliza en el resto de técnicas basadas en retrotransposones.

- IRAP (Inter-Retrotransposon Amplified Polymorphism) (Figura 1.11B): con esta técnica se pretende amplificar secuencias que se encuentren entre dos retrotransposones, para lo cual se utilizan dos cebadores diseñados a partir de la secuencia de las LTR. Se pueden utilizar dos cebadores diferentes o el mismo, y se pueden diseñar a partir de retrotransposones de la misma familia o de dos familias diferentes. El principal inconveniente de este método es la distancia entre los dos elementos móviles ya que, si ésta es muy grande, los segmentos amplificados serían demasiado largos, lo que dificulta su análisis por PCR y disminuye la resolución en los

geles de agarosa (Kalendar y Schulman, 2006). Esta técnica se ha probado en diversidad de especies de plantas como cebada (Brik *et al*., 2006), guisante (Smýkal, 2006), arroz (Branco *et al*., 2007) o vid (Pereira *et al*. 2005; Castro *et al*., 2012).

- REMAP (Retrotransposon-Microsatellite Amplified Polymorphism) (Figura 1.11C): se trata de una variante de la técnica IRAP que consiste en la combinación de los cebadores diseñados a partir de las LTRs con cebadores para microsatélites o SSR (Simple Sequence Repeats) para identificar posibles polimorfismos en el segmento amplificado entre ambos cebadores (Kalendar *et al*., 1999; Branco *et al*., 2007). En esta técnica se pueden utilizar como cebadores secuencias de microsatélites de dos, tres o cuatro repeticiones o cebadores complementarios de zonas flanqueantes de SSRs conocidos en la especie de estudio (Branco *et al*., 2007; Peredo *et al*., 2009). Hasta el momento se ha desarrollo esta técnica para el estudio de diferentes especies de plantas como arroz (Chadha y Gopalakrishna, 2005; Branco *et al*. 2007), vid (Pereira *et al*., 2005; Castro *et al*., 2012) o cebada (Brik *et al*., 2006).

- RBIP (Retrotransposón Based Amplified Polymorphism) (Figura 1.11D): esta técnica permite detectar polimorfismos en la integración de un elemento en un locus concreto y consiste en una amplificación en la que se utilizan dos cebadores diseñados a partir del ADN que flanquea la zona de inserción y un tercero obtenido a partir de la secuencia de una LTR del retrotransposón o de otra región interna de éste. Al utilizar los dos primeros cebadores se amplificará una región carente del retrotransposón, mientras que, si se amplifica un cebador flanqueante y el del retrotransposón, se amplificará el extremo del elemento móvil analizado (Schulman, 2007). Si el retrotransposón que estamos estudiando fuera muy corto, se obtendrían dos bandas cuando este esté presente ya que se amplificaría tanto el extremo como el elemento completo, sin embargo, en el caso de retrotransposones largos, los cebadores externos estarían demasiado separados y no se obtendría el amplicón de la zona ocupada. El principal inconveniente de esta técnica es la necesidad de conocer la secuencia de las zonas flanqueantes de la zona de inserción. En los últimos años se ha desarrollado esta técnica para el estudio de diferentes plantas como guisante (Smýkal *et al*., 2008A y B;

Jing *et al*., 2010), arroz (Petit *et al*., 2009), café (Hamon *et al*., 2011) y peral (Kim *et al*., 2012).

- TAM (Tagged microarray marker): se trata de una variante de la técnica RBIP en la que se sustituyen los geles de agarosa por micromatrices o "microarrays" (Flavell *et al*., 2003; Jing *et al*., 2007). Las PCR en las que se detectan las regiones ocupadas o no por los retrotransposones, se visualizan, todas juntas, en una membrana con una sonda específica de cada locus. Esta técnica de ha utilizado en guisante para analizar marcadores RBIP codominantes en una colección de germoplasma con miles de muestras (Jing *et al*., 2010).

- RUP (Retrotransposon UTR polymorphism fingerprinting): es un marcador molecular basado en la variabilidad en longitud de la región interna no traducible (UTR) que aparece en algunos retrotransposones (Pelsy, 2007). Se realiza una PCR utilizando cebadores derivados de una LTR y de la región *gag*. Las diferencias de tamaño de la región UTR son los suficientemente estables como para permitir realizar estudios filogenéticos. Pelsy (2007) utilizó esta técnica para definir el genotipo específico de 94 accesiones de *Vitis* (fingerprinting) mientras que, anteriormente, se había necesitado un mínimo de seis marcadores SSR con un alto grado de polimorfismo, para poder identificar las diferentes variedades (This *et al*., 2004).

- RIVP (Retrotransposon Internal Variation Polymorphism): esta técnica no utiliza la secuencia de las LTR sino que utiliza como cebadores secuencias internas del retrotransposón que revelan variaciones en la estructura interna (Vershinin *et al*, 1999).

- iPBS (inter- Primer Binding Site) (Figura 1.11E): esta técnica se desarrolló para la búsqueda de regiones LTR en cualquier genoma tanto animal como vegetal (Kalendar el al., 2010). Consiste en una reacción de PCR en la que se puede utilizar un único cebador o una combinación de dos cebadores diseñados a partir de las regiones PBS (Primer Binding Site) de diferentes retrotransposones. Esta región se encuentra junto a una de las LTR por lo que, durante la amplificación, se obtendrían, teóricamente, los extremos de dos retrotransposones próximos en el genoma así como la

región que se encuentra entre ellos. Con esta técnica se han conseguido secuenciar las LTR de retrotransposones *Ty1-copia*, *Ty3-gipsy*, TRIM y LARD. Kalendar *et al* (2010) han diseñado 83 cebadores de 12, 13 y 18 pb que han demostrado su eficacia en especies animales como vaca, yak, oveja o gallina y en plantas tanto angiospermas como gimnospermas (manzana, cebada, maíz, lino, *Medicago truncatula*, sandía o patata entre otras). En los geles de agarosa se pueden observar entre 15 y 50 bandas de 100 a 5000 pb y presentan un alto grado de polimorfismo dominante.

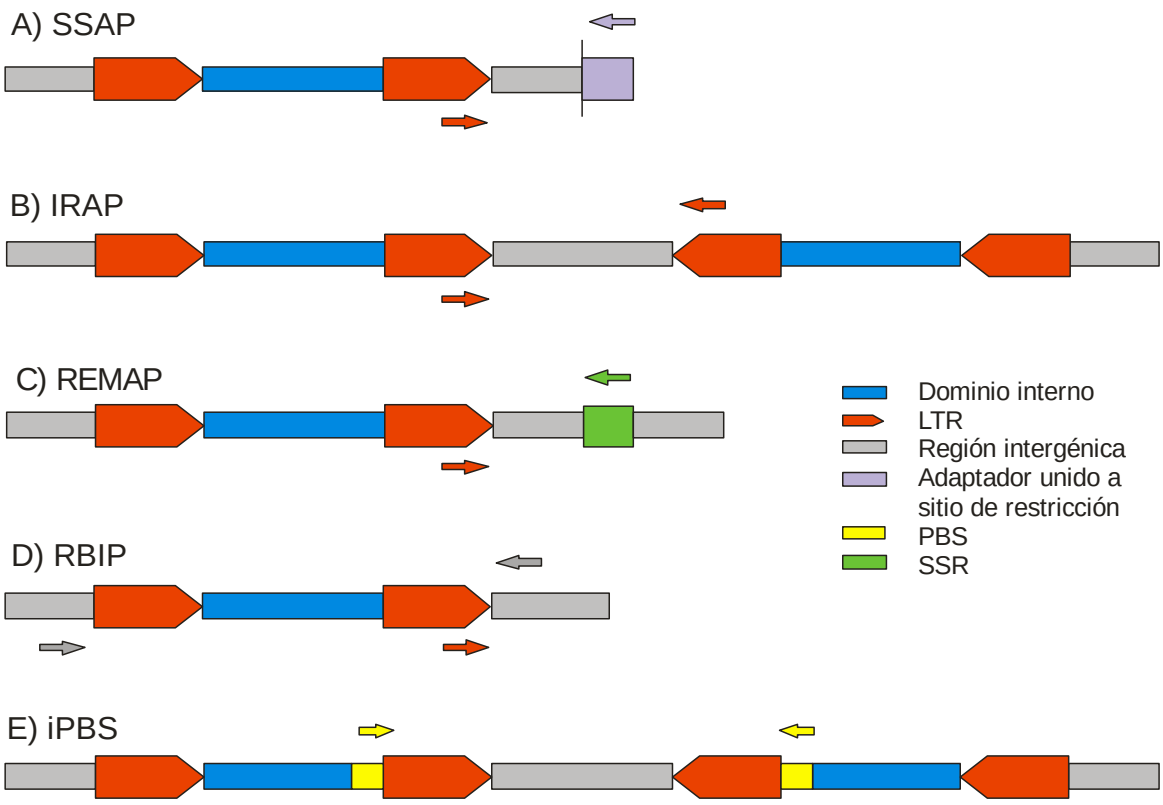


Figura1.11: Esquema de los marcadores moleculares basados en retrotransposones más utilizados.

1.5. Mapas genéticos

Un mapa genético es una representación gráfica de la disposición lineal de los loci que se están estudiando y la distancia relativa entre ellos expresada en unidades de mapa (udm) o centiMorgan (cM). El número de grupos de ligamiento ideal de una determinada especie debe coincidir con el número cromosómico haploide.

Dos o más genes están ligados cuando se encuentran situados en un mismo cromosoma y no segregan de forma independiente. Cuanto menor sea la probabilidad de que ocurra una recombinación, más cerca se encontrarán dos genes en el mapa genético.

Los primeros análisis de ligamiento en lenteja los realizaron Zamir y Ladizinsky (1984) y se basaron en marcadores morfológicos e isoenzimas. Mapearon un marcador morfológico y ocho isoenzimas en dos grupos de ligamiento utilizando un cruzamiento entre *L. c. culinaris* y *L. c. orientalis*. Más tarde, analizando un cruzamiento entre *L. c. culinaris* y *L. ervoides*, Tadmor *et al*. (1987) obtuvieron cinco grupos de ligamiento tras el mapeo de dos marcadores morfológicos, nueve isoenzimas, cuatro loci de proteínas de semilla y un punto de translocación. Muehlbauer *et al*. (1989) describieron los polimorfismos de aloenzimas para 18 loci ligados, demostrando su herencia monogénica y el ligamiento entre ellos y con otros cuatro genes que controlan caracteres morfológicos. En este estudio se obtuvieron seis pequeños grupos de ligamiento que contenían 14 de los loci analizados. Este trabajo demostró la existencia de sintenia entre *Lens* y *Pisum* ya que varios de los grupos de ligamiento estaban conservados en ambas especies. Tahir y Muehlbauer (1994) fueron los primeros en utilizar líneas RIL (Recombinant Inbred Lines) para realizar mapas de lenteja, identificaron seis grupos de ligamiento que incluían 17 loci isoenzimáticos y cuatro morfológicos.

El primer mapa de lenteja realizado a partir de marcadores basados en ADN fue descrito por Havey y Muehlbauer en 1989. Utilizaron para el estudio una población F_2 de un cruzamiento entre *L. c. culinaris* y *L. c. orientalis* y obtuvieron nueve grupos de ligamiento que incluían seis caracteres morfológicos, ocho isoenzimas y veinte marcadores RFLP. En 1992, Weeden *et al*., obtuvieron once grupos de ligamiento con el análisis de tres marcadores morfológicos, 37 isoenzimas y 26 RFLP. Los resultados de este mapa se compararon con mapas de *Pisum* de manera que se pudieron confirmar los resultados obtenidos anteriormente en el estudio de sintenia realizado por Muehlbauer *et al*. (1989), identificándose ocho regiones sinténicas entre las dos especies.

Más tarde, se obtuvieron nuevos mapas estudiando poblaciones segregantes de cruzamientos subespecíficos e interespecíficos. Tahir *et al*. (1993), unificando los resultados de diferentes estudios, propusieron diez grupos de ligamiento basados en el análisis de siete marcadores morfológicos, 25 isoenzimas, 38 RFLP, cuatro proteínas de semilla, un marcador de cloroplasto y un marcador ribosomal. Vaillancourt y Slinkard (1993) obtuvieron cuatro grupos de ligamiento a partir de una población F_2 analizando 17 isoenzimas y 11 marcadores morfológicos y, un año más tarde, Tahir y Muehlbauer (1994) identificaron seis grupos de ligamiento que incluían 17 isoenzimas y cuatro caracteres morfológicos en poblaciones RIL de lenteja.

A medida que las técnicas basadas en PCR se fueron introduciendo en los laboratorios de forma rutinaria, el número de marcadores genéticos mapeados en el genoma de lenteja fue aumentando considerablemente. Así, Eujayl *et al*. (1997) analizaron 76 loci segregantes en una población F_2 y, utilizando el programa MAPMAKER (Lander *et al*., 1987) con un "LOD score" de ligamiento de 3,0, obtuvieron nueve grupos de ligamiento que incluían 28 RAPD, un RFLP, un marcador morfológico y otros tres marcadores.

Un año más tarde, Eujayl *et al*. (1998), publicaron el primer mapa genético extenso en lenteja utilizando 177 marcadores (89 RAPD, 79 AFLP, seis RFLP y tres de tipo morfológico), lo desarrollaron a partir del análisis de 86 líneas RIL de un cruzamiento intersubespecífico de *L. c. orientalis*. La segregación fue analizada mediante el MAPMAKER con un "LOD score" de 4,0 y redujeron el mapa a siete grupos de ligamiento, este fue el primer mapa en el que se conseguía obtener un número de grupos de ligamiento igual al número cromosómico haploide de lenteja.

Rubeena *et al*. (2003) publicaron el primer mapa de lenteja utilizando variedades de una misma especie, para ello analizaron una población F_2 de un cruce entre ILL5588 e ILL7537 que se diferencian en la resistencia a *Ascochyta* que presenta el primero frente a la susceptibilidad del segundo, es el primer mapa en el que se analizan genes

análogos de resistencia (RGA). En este estudio se analizaron 114 marcadores (100 RAPD, 11 ISSR y tres RGA) y se obtuvieron nueve grupos de ligamiento.

El primer mapa genético de lenteja que incluía marcadores SSR fue publicado por Durán *et al*. (2004). Se estudió una población F_2 de 113 individuos del cruzamiento entre *L. c. culinaris* (Lupa) y *L. c. orientalis* (BG16880) y se obtuvieron diez grupos de ligamiento con 161 de los 298 marcadores analizados (71 RAPD, 39 ISSR, 83 AFLP, dos SSR y cinco caracteres morfológicos).

En 2005, Hamwieh *et al*., añadieron 39 SSR y 49 AFLP a los marcadores analizados por Eujayl *et al*. (1998). Utilizando un nuevo programa, JOINMAP (Van Ooijen y Voorrips, 2001), y niveles de LOD que variaban entre 3,0 y 6,0, obtuvieron 14 grupos de ligamiento con 286 marcadores. Aunque aumentaron el número de grupos de ligamiento alcanzado en el mapa previo se consiguió disminuir, considerablemente, la longitud del mismo (751 cM frente a los 1073 cM anteriores) y la distancia media entre marcadores (2,6 cM frente a 6 cM).

Phan *et al*. (2007), publicaron el primer mapa de ligamiento basado en genes conocidos de lenteja. En este caso se utilizó una población F_5 intraespecífica y se analizaron 79 ITAP (Intron-Targeted Amplified Polymorphic) y 18 SSR. Se obtuvieron siete grupos de ligamiento aplicando el programa MULTIPOINT y un LOD de 9,0, que ocupaban 928,4 cM de longitud. Este estudio evidenció las relaciones de macrosintenia simple y directa entre *Lens* y *M. truncatula*.

Tullu *et al*. (2008) desarrollaron un mapa constituido por 240 marcadores (177 AFLP, 54 RAPD y nueve SSR) que presentaba 12 grupos de ligamiento. Además de estos marcadores, mapearon QTL relacionados con resistencia a *Ascochyta* y a antracnosis así como otros relacionados con la precocidad y altura de la planta.

En 2010, Tanyolac *et al*. publicaron un mapa desarrollado con 94 individuos de una población RIL del cruzamiento Precoz x WA8949041. Se estudiaron 162

marcadores (64 AFLP, 23 ISSR, 75 RAPD y cuatro morfológicos). Este mapa está formado por 11 grupos de ligamiento.

En ese mismo año, Saha *et al*. (2010) publicaron un nuevo mapa con 14 grupos de ligamiento, desarrollado a partir de 206 plantas de una población RIL del cruzamiento ILL6002 x ILL5588 tras el análisis de 145 marcadores (27 RAPD, 27 SSR, 89 SRAP y dos morfológicos). Además, en el mapa obtenido se identificaron doce QTL relacionados con el diámetro y peso de la semilla, el tamaño de la planta y el tiempo de floración.

En 2012, de la Puente, describió dos nuevo mapas genéticos desarrollados a partir del estudio de 73 marcadores (10 SNP-CEL, 11 CAPS, 26 SSR, dos RAPD, cinco SRAP, cinco debido a diferencias en el tamaño de amplicones (ALP), 14 marcadores definidos por la presencia o ausencia de amplificación (P/A)) en 113 individuos de una población F_2 y de 85 marcadores (82 ISSR y tres morfológicos) en la población RIL del cruzamiento entre la especie cultivada *L. c. culinaris* (Lupa) y la silvestre *L. c. orientalis*. El mapa de la población RIL consta de 10 grupos de ligamiento mientras que, con los marcadores analizados en la F_2, junto con los descritos por Durán *et al*. (2004), se obtuvo un mapa con nueve grupos de ligamiento.

Neyra (2014) amplió el mapa anterior obtenido con la población RIL añadiendo 16 marcadores (seis de PCR específica, seis CAPs y cuatro SNPs) y consiguió reducir el número de grupos de ligamiento a nueve

El avance más importante de los últimos tiempos en el mapeo genético de lenteja es el mapa publicado por Sharpe *et al*. (2013). Este mapa se desarrolló en una población RIL intraespecífica de *Lens culinaris* (CDC Robin x 964a-46), contiene seis marcadores SSR y 537 SNP y presenta siete grupos de ligamiento. Los cebadores SNP se desarrollaron a partir de la secuenciación del transcriptoma de lenteja mediante la técnica de secuenciación 454. En este mapa se han localizado, de nuevo, regiones sinténicas entre los genomas de *L. culinaris* y *M. truncatula*.

Fedoruk *et al*. (2013) realizaron una revisión y amplificación del mapa anterior de manera que consiguieron localizar 560 marcadores (550 SNP, seis SSR y cuatro morfológicos) en siete grupos de ligamiento y consiguieron disminuir la longitud del mapa pasando de 834,7 a 697 cM lo que implica una distancia media entre marcadores de 1,2 cM que indica un nivel de saturación muy alto.

Yilmaz Temel *et al*. (2015) desarrolló un mapa que incluye 388 marcadores (376 SNP, tres SSR y nueve ISSR) localizados en nueve grupos de ligamiento, siete de ellos más grandes que se podrían corresponder con los siete cromosomas de lenteja. La longitud que presenta es de 432,8 cM con una distancia media entre marcadores de 1,11 CM que es la más baja publicada hasta este momento.

El último mapa de lenteja publicado hasta la fecha es el de Andeden *et al*. (2015). Este mapa, basado en el uso de marcadores SSR analizados en una población F_2, incluye 47 marcadores localizados en siete grupos de ligamiento. Presenta una longitud de 303,9 cM y una distancia media entre marcadores de 7,06 cM.

OBJETIVOS

2. Objetivos

Los objetivos planteados para este trabajo han sido:

1- Identificar retrotransposones del tipo LTR, en el genoma de *Lens culinaris*.

2- Identificar y secuenciar diferentes regiones de retrotransposones, incluyendo LTR, para la obtención de marcadores genéticos útiles en la construcción de mapas genéticos.

3- Obtener otros marcadores moleculares desarrollados a partir de elementos móviles de lenteja.

4- Analizar la segregación de dichos marcadores en una población RIL del cruzamiento entre la especie cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880.

5- Obtener un mapa genético basado en marcadores derivados de retrotransposones e integrarlo en mapas previos generados a partir de estudios en la misma población RIL.

6- Comparar el mapa genético obtenido con otros mapas de lenteja y establecer las relaciones de sintenia con la especie de referencia *Medicago truncatula*.

MATERIALES Y MÉTODOS

3. Materiales y métodos

3.1. Material Vegetal

Para los estudios realizados se ha utilizado material biológico de la especie *Lens culinaris* ssp. *culinaris* cv. Lupa y una población RIL (línea consanguínea recombinante) F_7 de 94 líneas, pertenecientes al cruzamiento entre la forma cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880. Dicho cruzamiento fue obtenido en el laboratorio del Área de Genética de la Universidad de León por el doctor Frattini.

3.2. Extracción de ADN

Para extraer el ADN, se utilizó el juego de productos para aislamiento de ADN de Quiagen (Dneasy® Planta Mini Kit). El ADN se aisló a partir de harina procedente de semillas individuales, previo machacado por métodos físicos, y a partir de hojas de plantas crecidas en condiciones estériles, en cuyo caso se utilizó nitrógeno líquido para la homogeneización.

Para obtener las plantas en condiciones estériles se sumergen las semillas unos segundos en etanol, se dejan unos 20 segundos en hipoclorito sódico (lejía) diluida (1:3) y, por último, se hacen dos lavados seguidos con agua destilada y esterilizada, todo ello en cabina de flujo laminar. A continuación, las semillas se introducen, individualmente, en tubos con algodón estéril húmedo y se dejan germinar en una cámara de crecimiento. Las hojas se recogieron 10 días después de la siembra.

Para la extracción se siguió el protocolo descrito por el fabricante basado en la utilización de dos columnas, la primera de ellas retiene impurezas mientras que la segunda retiene el ADN que se recupera, tras varios lavados, mediante una elución con agua milli Q.

Tras la extracción, se procedió a la cuantificación del ADN utilizando el espectrofotómetro NanoDrop ND-1000 (Thermo Scientific). Se trata de un espectrofotómetro UV/Visible que permite conocer tanto la concentración del ADN

como su pureza midiendo la absorbancia de la muestra a 230, 260 y 280 nm. Las muestras se diluyeron a una concentración de 20 ng/µl para amplificaciones y de 40 ng/µl para digestiones con endonucleasas.

3.3. Electroforesis en geles de agarosa

Otro método para comprobar el estado y cantidad del ADN obtenido en las extracciones, así como el resultado de las amplificaciones, es la realización de electroforesis en gel de agarosa con una concentración que puede variar entre el 0,8 y el 2 % según el tamaño de las bandas esperadas en cada caso, en tampón TAE 1X (Tris acetato 40 mM y EDTA 1 mM, pH 8) a 85-100 voltios (las condiciones exactas se indican en cada uno de los apartados). Como marcador de peso molecular se utilizó una "escalera" comercial de fragmentos (1 kb plus DNA ladder de Invitrogen), y se cargaron entre 10 y 40 µl de muestra de ADN + 5 µl de tampón de carga "Blue juice" (azul de bromofenol 0,025% + glicerol 50%) que permite ver cómo se va desplazando el frente y mantiene las muestras en el fondo del pocillo.

Para observar las bandas en el gel se utilizaron dos tipos de tinciones diferentes:

- Bromuro de etidio: en este caso se sumergió el gel en una solución de bromuro de etidio (0,5 µg/ml) con agitación orbital suave durante 20 minutos y, posteriormente, se fotografió el gel.
- Red Safe (iNtRON): este compuesto se añadió directamente en el gel. Se utilizó 1 µl de Red Safe 1000X por cada ml de tampón TAE 1X, de manera que no fue necesaria la tinción tras la electroforesis.

En ambos casos, el gel se fotografía en un transiluminador Gene Genius Bio Imaging System (Syngene) de luz ultravioleta conectado mediante un ordenador con el programa GeneSnap (Syngene), que permite la visualización de los fragmentos de ADN obtenidos en la amplificación.

3.4. Amplificación de retrotransposones

Para buscar los posibles retrotransposones en el genoma de lenteja se llevó a cabo una amplificación mediante PCR (Polimerase Chain Reaction) utilizando como molde el ADN genómico de *Lens culinaris* cv. Lupa.

La PCR se realizó utilizando cebadores degenerados obtenidos a partir del gen de la retrotranscriptasa y diseñados para amplificar retrotransposones tipo *Ty1-copia* (Hirochika y Hirochika,1993) y tipo *Ty3-gypsy* (Friesen *et al*., 2001), que se han utilizados en guisante con buenos resultados (Dixit *et al*., 2006), y otros diseñados a partir de una secuencia de *Lens culinaris* (Tnana) (Llorente, 2001) (ver apartado 3.10), que presenta una buena homología con secuencias de retrotransposones conocidos, en este último caso se diseñaron tres cebadores "forward" o directo y uno "reverse" o inverso con los que se obtuvieron tres combinaciones diferentes (Tabla 3.1).

Tabla 3.1: Secuencias de los cebadores utilizados y nombre con el que se denominaron a lo largo del estudio.

Cebador	**Secuencia**	**Tm**
Copia F	5´ CARATGGAYGTNAARAC 3´	47,9 ^{o}C
Copia R	5´ CATRTCRTCNACRTA 3´	42,4 ^{o}C
Gypsy F	5´ MRNATGTGYGTNGAYTAYMG 3´	55,3 ^{o}C
Gypsy R	5´ RCAYTTNSWNARYTTNGCR 3´	52,4 ^{o}C
Tnana F1	5´ GAAAAATMGAKCTGARTTATTCTC 3´	55,0 ^{o}C
Tnana F2	5´ GTACACRATCTYTCWCCTGG 3´	57,3 ^{o}C
Tnana F3	5´ CAAAARRGYTATCGTTGCTA 3´	52,2 ^{o}C
Tnana R	5' GTTCACAKGTTRCYAKTGCC 3'	57,3 ^{o}C

Nomenclatura utilizada para las posiciones degeneradas: N = A, C, G o T; K = G o T; M = A o C; R = A o G; S = G o C; W = A o T; Y = C o T

Se realizaron diferentes reacciones en función del tipo de retrotransposón que se quería encontrar así como de los cebadores utilizados. En ambos casos se utilizó un tampón de amplificación avanzado con $MgCl_2$ (500 mM KCl, 100 mM Tris-HCl pH 8,3 (a 25°C), 15 mM Mg^{2+}).

- *Ty1-copia* y *Ty3-gypsy* (Dixit *el al*., 2006): se llevaron a cabo reacciones de amplificación de 50 µl de volumen total:

ADN genómico (20 ng/µl)	5 µl
Cebador directo (5 pmol/µl)	10 µl
Cebador inverso (5 pmol/µl)	10 µl
Tampón de amplificación 10X	5 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	5 µl
Taq polimerasa 5 U/µl (5Prime)	1 µl
H_2O milli Q	14 µl

Los programas de PCR fueron los siguientes: para los cebadores correspondientes a los retrotransposones *Ty1-copia*: 3 min a 94 ^oC, 40 ciclos (1 min 94^{o}C, 50 s 45^{o}C, 40 s 68^{o}C) y 10 min 68^{o}C, y para los de tipo *Ty3-gypsy* la reacción fue de 3 min a 94^{o}C, 40 ciclos (1 min 94^{o}C, 50 s 50^{o}C, 40 s 68^{o}C) y 10 min 68^{o}C.

- Secuencia Tnana de *Lens culinaris* (Llorente, 2001): en este caso se realizaron reacciones de 100 µl

ADN genómico (20 ng/µl)	5 µl
Cebador directo (5 pmol/µl)	10 µl
Cebador inverso (5 pmol/µl)	10 µl
Tampón de amplificación 10X	10 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	10 µl
Taq polimerasa 5 U/µl (5Prime)	0,5 µl
H_2O milli Q	54,5 µl

El programa de PCR, para estos cebadores, fue el siguiente: 3 min 94^{o}C, 30 ciclos (45 s 94^{o}C, 1:40 min 52^{o}C, 2 min 68^{o}C) y 10 min 68^{o}C.

Las reacciones se llevaron a cabo en termocicladores de Applied Biosystems (GeneAmp PCR System 9700 y 2720 Thermal Cycler)

Los resultados de las amplificaciones se observaron tras una electroforesis en gel de agarosa al 1% y 100 V que se dejó correr durante 1 hora. Cuando la amplificación ha tenía éxito se seguía el protocolo de extracción de bandas, transformación y clonación (ver apartados 3.6 y 3.7).

3.5. Amplificación de repeticiones terminales largas LTRs

Para localizar las LTRs de los retrotransposones en el genoma de lenteja, se utilizaron dos técnicas diferentes: localización de LTRs a partir de secuencias internas conservadas del retrotransposón e iPBS (inter-Primer Binding Site).

3.5.1. Amplificación de LTRs a partir de secuencias internas conservadas del retrotransposón

Esta técnica consta de varios pasos:

1- Diseño de cebadores: a partir de los extremos de las secuencias localizadas con los cebadores Copia, Gypsy y Tnana se diseñaron nuevos cebadores (ver apartado 3.10) que permitieran avanzar a lo largo de la secuencia para alcanzar la región LTR.

2- Digestión parcial del material genético de lenteja con una enzima de restricción. En este caso se utilizó la enzima *Mse*I (Biolabs). Para llevar a cabo esta digestión se realizó la siguiente reacción:

*Mse*I	2 U (0,2 µl)
Tampón 10x (NEB2)	10 µl
ADN	2,5 µg (60 µl)
H_2O milli Q	29,8 µl

La enzima actuó durante 20 minutos a 37^{o}C e inmediatamente después se incubó la muestra durante 10 minutos a 80^{o}C para detener la digestión y conseguir fragmentos de un tamaño aproximado de 1 kb.

3- Ligación o unión de los adaptadores complementarios a la zona de corte de la enzima utilizada en la digestión y complementarios entre sí: *Mse*I1 (5´-

GACGATGGATCCTGAG-3´) y *Mse*I2 (5´-TACTCAGGATCCAT-3´). El primer paso de la ligación es la mezcla de los adaptadores para conseguir una concentración final de 2 pm/µl. La mezcla se desnaturalizó incubándola 3 minutos a 95^{o}C y se dejó renaturalizar a temperatura ambiente durante 15 minutos.

4- Reacción de ligación:

ADN digerido	15 µl
Tampón de ligación 10X	2 µl
Ligasa T4 (1 U/µl)	0,1 µl
Adaptador	1 µl
H_2O milli Q	1,9 µl

La reacción se incuba durante 2 horas a 20^{o}C.

5- El siguiente paso consistió en una amplificación en dos pasos (Van der Linden *et al*., 2004) de manera que, en un primer paso, se realizó una PCR en la que sólo se incluyó el cebador específico del retrotransposón para aumentar el número de fragmentos que incluyan la secuencia específica, mientras que, en la segunda, se añadió también el cebador complementario al adaptador. Se utilizó este tipo de reacción para evitar la amplificación masiva de fragmentos que presenten el adaptador en ambos extremos.

La primera amplificación consiste en una reacción de 25 µl de volumen total:

ADN con el adaptador	4 µl
Cebador específico (5 pmol/µl)	3 µl
Tampón de amplificación 10X (Quiagen)	2,5 µl
dNTPmix (2 mM cada uno de los cuatro dNTPs)	1,3 µl
Taq polimerasa Hot Start 5 U/µl (Quiagen)	0,1 µl
H_2O milli Q	14,1µl

Una vez finalizada la primera reacción se añaden al mismo tubo:

Cebador *Mse*I (5 pmol/µl)	3 µl
Cebador específico (5 pmol/µl)	3 µl
Tampón de amplificación 10X	2,5 µl
dNTPmix (2 mM cada uno de los cuatro dNTPs)	2,5 µl
Taq polimerasa Hot Start 5 U/µl (Quiagen)	0,1 µl
H_2O milli Q	13,9µl

La reacción de PCR para las dos reacciones fue: 15 min a 95^{o}C, 30 ciclos (30 s 95^{o}C; 1:40 min 45-56^{o}C, 2 min 72^{o}C) y 7 min a 72^{o}C. La temperatura de apareamiento depende del cebador utilizado en cada reacción (Tabla 3.2).

6- Comprobación de la amplificación en un gel de agarosa al 1 % y 100 V, para lo cual se cargan 10 µl de muestra. Si la amplificación ha tenido éxito se sigue el protocolo de extracción de bandas, transformación y clonación (ver apartados 3.6 y 3.7).

Esta técnica también se utilizó con los cebadores degenerados Rnasah1: 5'-MGNACNAARCAYATHGA-3' y Rnasah2: 5'-GCNGAYATNYTNACNAA-3', diseñados a partir de la región RNasaH de un retrotransposón de guisante (Pearce *et al*., 1999).

Además, con estos últimos cebadores, se utilizó la técnica "nested PCR" que consiste en la digestión parcial del genoma con la enzima *MseI* seguida de una ligación de adaptadores similar a la descrita en la técnica anterior. A continuación, se realiza de una doble PCR de manera que, en la primera amplificación se utiliza el cebador más interno de la secuencia, en este caso Rnasah1 y en la segunda el cebador Rnasah2 ambos combinados con el cebador complementario al adaptador.

Para esta última técnica, las amplificaciones consisten en reacciones de 50 µl de volumen que incluye:

ADN con el adaptador	7,5 µl
Cebador RnasaHI/II (5 pmol/µl)	6 µl
Cebador MseI	6 µl
Tampón de amplificación 10X (Quiagen)	5 µl
dNTPmix (2 mM cada uno de los cuatro dNTPs)	5 µl
Taq polimerasa 5 U/µl (5Prime)	0,2 µl
H_2O milli Q	20,3 µl

La reacción de PCR para las dos reacciones fue: 2 min a 94^{o}C, 30 ciclos (1 min 94^{o}C; 1:30 min 45^{o}C, 1:30 min 72^{o}C) y 7 min a 72^{o}C.

Además de estos cebadores, también se llevó a cabo una amplificación sencilla utilizando cebadores diseñados a partir de la secuencia de una LTR del retrotransposón *Angela* de *Pisum sativum* por Smýkal *et al*. (2009). Dichos cebadores, denominados LTR-Ps: 5'- TGAAGGAGAATTGCGACCCAAAGCGC-3' y LTR-1F: 5'-TGTTGGTGTAAGCCCCTAGAGGCC-3', se utilizaron para comprobar la obtención de una LTR completa de *Lens culinaris* descrita por los mismos autores.

La PCR consistió en una reacción de 50 µl de volumen con:

ADN genómico (20 ng/µl)	5 µl
Cebador directo (5 pmol/µl)	10 µl
Cebador inverso (5 pmol/µl)	10 µl
Tampón de amplificación 10X	5 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	5 µl
Taq polimerasa 5 U/µl (5Prime)	1 µl
H_2O milli Q	14 µl

La reacción consistió en 3 min a 94 ^{o}C, 40 ciclos (1 min 94 ^{o}C; 50 s 50 ^{o}C ; 40 s a 68 ^{o}C) y 10 min a 68 ^{o}C.

3.5.2. iPBS (inter-Primer Binding Site)

Esta técnica consiste en una reacción de PCR en la que se puede utilizar un único cebador o una combinación de dos cebadores diseñados a partir de las regiones

PBS (Primer Binding Site) de diferentes retrotransposones (Kalendar *el al.*, 210) (ver Tabla 4.4, página 121).

Para llevar a cabo esta técnica, se realizó una reacción de PCR (Kalendar *el al.*, 2010) utilizando como sustrato el ADN genómico de *Lens culinaris* cv. Lupa. La amplificación consistió en una reacción de 20 µl de volumen que incluían:

ADN genómico (20 ng/µl)	2 µl
Cebador 1 (5 pmol/µl)	1,2 µl
Cebador 2 (5 pmol/µl)	1,2 µl
Tampón de amplificación 10X (Biotools)	2 µl
dNTPmix	2 µl
$MgCl_2$	1,2 µl
Taq polimerasa 5 U/µl (5Prime)	0,2 µl
Pfu polimerasa 1 U/µl (Biotools)	0,04 µl
H_2O milli Q	10,2 µl

3.6. Extracción de bandas específicas

Para poder extraer las secuencias amplificadas tras la PCR, se llevaron a cabo electroforesis con la mayor cantidad de muestra posible, en este caso se llenaron los pocillos con 40 µl de la reacción de PCR + 5 µl de tampón de carga. En esta electroforesis se utilizó un gel de agarosa al 2 % ya que es necesaria una mayor resolución y definición de las bandas para una mejor extracción de éstas.

Una vez teñido el gel, el corte de las bandas se llevó a cabo sobre un transiluminador de luz ultravioleta que permite su visualización. Este proceso se realizó extremando las medidas de protección debido a la larga exposición a la luz UV.

Tras la obtención de dichos fragmentos, se recuperó el ADN a partir de los geles de agarosa mediante un juego de productos de extracción. Se utilizaron dos diferentes siguiendo, en cada caso, las indicaciones del fabricante:

- QIAquick Gel Extraction Kit, de Quiagen
- UltraClean GelSpin DNA Purification Kit, de Mo Bio.

Ambos protocolos se basan en el uso de membranas de sílica-gel que retienen las impurezas para obtener un ADN totalmente purificado.

Una vez extraídas las bandas se comprobó el éxito de la extracción mediante electroforesis en gel de agarosa al 1 % y 100 V.

3.7. Clonación y transformación

Una vez obtenidos los fragmentos deseados, se llevó a cabo la clonación de los diferentes productos de amplificación en el plásmido pGEM®-T (Promega) este vector está diseñado para la clonación de productos de PCR, se suministra cortado y con extremos 3' T protuberantes, de modo que actúan como extremos cohesivos con los extremos 3' de los amplificados por PCR que contienen adeninas (Promega, 2007). Se utilizó para la ligación la ligasa del fago T4 siguiendo el protocolo recomendado por Promega:

ADN	3 µl
Tampón de ligación	5 µl
Ligasa de T4	1 µl
pGEM®-T	1 µl

La reacción se deja incubar una hora a temperatura ambiente o toda la noche a 4^{o}C. Se obtiene mayor número de colonias recombinantes dejando la reacción durante toda la noche.

Pasado este tiempo, se procedió a la transformación en células competentes de *E. coli* de la cepa DH5α obtenidas según el protocolo INOUE (Inoue *et al*., 1990). Una de las principales diferencias que presenta esta técnica frente a otras es la temperatura de incubación de las células, ya que se utilizan a 18^{o}C en lugar de los 37^{o}C utilizados generalmente, y que ha demostrado una mayor eficacia.

El procedimiento para la transformación se detalla a continuación:

1- Se descongelan las células competentes en un baño de hielo.
2- Se añaden 100µl de células competentes al tubo que contiene los 10µl de mezcla de reacción.
3- Se mantienen 20-30 minutos en hielo.
4- Se someten a un choque térmico a 42^{o}C de 45-50 segundos.
5- Se mantienen en hielo 2 minutos.
6- Se añade 1 ml de medio LB (Luria Bertani: 10 g bacto triptona, 5 g extracto de levadura, 10 g NaCl; pH 7,3 ajustado con NaOH) (Miller, 1972).
7- Se mezcla suavemente y se incuba 1 hora a 37^{o}C.
8- Se añaden 100 µl de X-Gal 5 % y 100 µl de IPTG 100 mM al tubo de la transformación, para poder distinguir fácilmente las colonias transformadas de las que no lo estén.
9- Se extienden 500 µl de la mezcla de transformación en una placa con medio LB suplementado con agar y ampicilina (100 µg/µl), 300 µl en otra y el resto en una tercera placa lo que permitirá elegir aquella placa en la que las colonias se encuentren más separadas.
10- Se Incuban a 37^{o}C durante 16 horas. No se debe incubar durante más tiempo ya que la ampicilina puede empezar a degradarse.

Al utilizar medio LB suplementado con ampicilina se permite sólo el crecimiento de bacterias con el plásmido ya que éste presenta un gen de resistencia a dicho antibiótico. Al crecer las bacterias en presencia de IPTG y X-Gal, las colonias recombinantes se verán de color blanco porque tienen inactivado el gen *Lac*Z que se encuentra en la zona de inserción del plásmido (Figura 3.1), mientras que las bacterias sin inserto aparecerán de color azul y de menor tamaño.

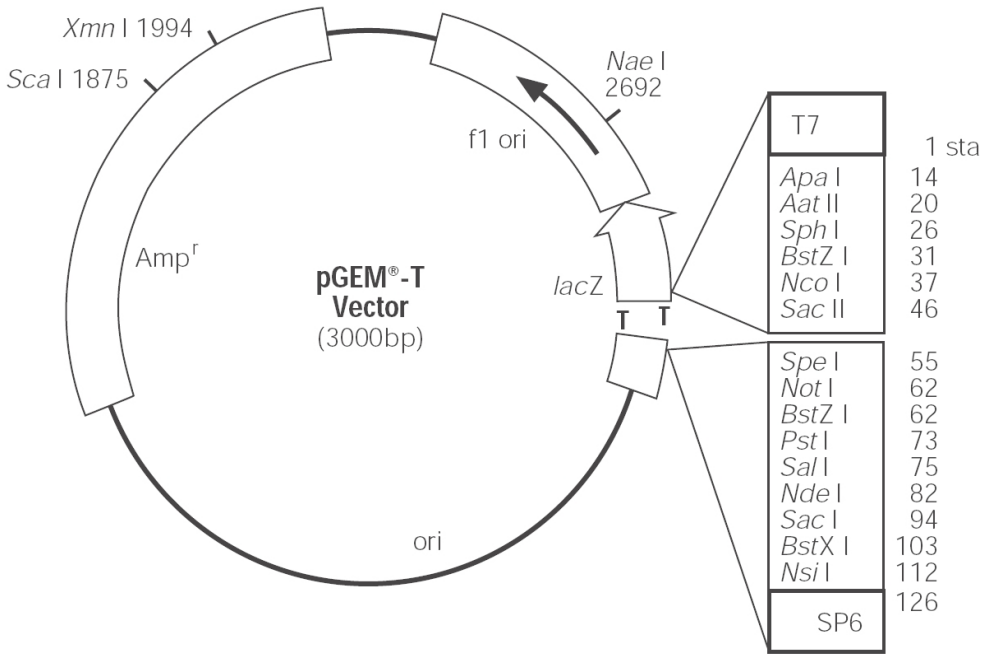


Figura 3.1: Vector pGEM®-T diseñado para la clonación de productos de PCR.

Una vez crecidas, se seleccionan algunas colonias transformantes recombinantes (blancas) y se pasan a una placa nueva. En este trabajo se pasaron 16 colonias blancas y 4 azules que se utilizaban como control. Para comprobar la presencia del inserto en las colonias transformantes positivas se realizó la técnica de "minipreparación" ultrarrápida de colonias positivas que se describe a continuación:

1- Se resuspende media colonia en 50 µl de EDTA 0,01 M, pH 8
2- Se añaden 50 µl de "cracking buffer" (0,2 g sacarosa; 40 µl NaOH 5 M; 50 µl SDS 10%; 900 µl H_2O) recién hecho y se mezclan en un agitador en vórtice.
3- Se incuba a 70^oC durante 5 minutos.
4- Se añaden 1,5 µl de KCl 4 M y 2 µl de bromofenol 0,4 %. Se mezclan con el agitador en vórtice.
5- Se mantienen en hielo 5 minutos.
6- Se centrifugan 4 minutos.
7- Se mantienen, al menos, 10 minutos en hielo.
8- Se carga un gel de agarosa al 1% con 15 µl del sobrenadante.

Las colonias que presentan un inserto del tamaño esperado se preparan para su secuenciación.

3.8. Secuenciación de ADN

Las colonias que presentan el inserto se resuspenden en 20 ml de medio LB suplementado con 20 μl de ampicilina (100 μg/μl) y se incuban a 37^{o}C y en agitación (300 r.p.m.) durante toda la noche. A continuación se centrifugan las células a 3800 r.p.m., durante 5 minutos y se elimina el sobrenadante de manera que se recupera únicamente el sedimento, la centrifugación debe ser suave para evitar que se rompan las células y se libere el ADN. Para la secuenciación se utilizaron los cebadores universales: T7 (5′-GTAATACGACTCACTATAGGGC-3′) y SP6 (5′-GCTATTTAGGTGACACTATAGAA-3′).

La secuenciación de ADN se realizó en el Servicio de Secuenciación del Laboratorio de Técnicas Instrumentales de la Universidad de León. Para la reacción de secuenciación se utilizó DYEnamic ET Dye Terminator kit (MegaBACE, Amersham), se trata de una modificación del método de los didesoxinucleótidos (Sanger *et al*., 1977) en el que cada didesoxinucleótido está marcado con un sistema de fluorocromos distinto. La amplificación se llevó a cabo en un termociclador MJ Research PTC-200. Los fragmentos de ADN se analizaron en un secuenciador MegaBACE 500 (Amersham Biosciences).

3.9. Análisis de secuencias

Una vez obtenidas las secuencias, éstas se analizaron con la ayuda del editor de alineamiento del programa MEGA5 (Tamura *el al*., 2011). El programa permite el alineamiento manual e incluye los métodos Clustal y Muscle para obtener alineamientos automáticos. Además, se usó la base de datos GeneBank del NCBI (“Nacional Center for Biotechnology Information”) utilizando los algoritmos BLASTn (permite la comparación entre secuencias de nucleótidos) y BLASTp (comparación entre secuencias de aminoácidos) (Altschul *et al*, 1990) para comparar las secuencias obtenidas con otras ya conocidas.

Para la representación de las secuencias en este trabajo se ha utilizado la siguiente nomenclatura:

- CLc para secuencias obtenidas con los cebadores de *Ty1-copia*.
- GLc: para secuencias tipo *Ty3-gypsy*.
- TLc: para secuencias obtenidas a partir de las tres combinaciones de cebadores de la secuencia Tnana de *Lens culinaris*.

A continuación se indica el número de la banda seguido de un guion y el número del clon correspondiente

Este programa también permite construir árboles filogenéticos para lo cual, una vez alineadas las secuencias, se convierten al formato "mega" que permite la generación de los árboles.

Para elegir el mejor modelo se utilizó una opción del mismo programa que permite realizar comparaciones entre 24 modelos diferentes. En todos los casos se escogió el modelo de Tamura con tres parámetros, que tiene en cuenta el efecto del contenido de G+C sobre las sustituciones de las secuencias, y una distribución gamma con cinco parámetros, que valora la heterogeneidad de los cambios en los diferentes sitios de la secuencia (Tamura, 1992). Con el modelo se obtuvieron matrices de distancia y árboles filogenéticos del tipo Neighbor Joining (Saitou y Nei, 1987), que permiten agrupar las secuencias según la similitud existente entre las secuencias de lenteja y otras secuencias obtenidas de las bases de datos. Para comprobar la fiabilidad de los nodos, se ha utilizado el método "bootstrap" con 1000 repeticiones que permite determinar los datos de distancia y la desviación estándar de cada uno de ellos.

3.10. Diseño de cebadores

A partir de los extremos de las secuencias localizadas con los cebadores Copia, Gypsy y Tnana se diseñaron nuevos cebadores para avanzar a lo largo de la secuencia y poder alcanzar la región LTR. También se diseñaron cebadores a partir de las secuencias de LTRs localizadas en el genoma de lenteja. Para ello se utilizó la aplicación Primer Quest de Integrated DNA Technologies (IDT).

Los cebadores se diseñaron intentando que tuvieran una temperatura de apareamiento de alrededor de 55-60^{o}C y un contenido en CG del 40-60 %. (ver Tabla 4.2, página 109 y Tabla 4.5, página 129).

3.11. Técnicas para la búsqueda de marcadores moleculares

Los análisis de marcadores se realizaron con 94 individuos y los parentales de una población RIL del cruzamiento intraespecífico entre *Lens culinaris* ssp. *culinaris* cv. Lupa y *Lens culinaris* ssp. *orientalis* BG16880. Se utilizaron tres técnicas diferentes para la búsqueda de marcadores moleculares: SSAP, REMAP y PBS.

3.11.1. SSAP

Para llevar a cabo la técnica SSAP (Sequence-Specific Amplification Polymorphism) se siguió el protocolo que se describe a continuación (Syed y Flavell, 2006):

1- Digestión del ADN genómico: se utilizó una combinación de dos enzimas de restricción *Mse*I (Biolabs) y *Pst*I (Biolabs). *Mse*I es una enzima con una diana de corte de 4 nucleótidos (T$^{\blacktriangledown}$TAA) mientras que la de *Pst*I es de 6 nucleótidos (CTGCA$^{\blacktriangledown}$G).

ADN (40 ng/µl)	20 µl
*Mse*I (10 U/µl)	0,5 µl
*Pst*I (20 U/µl)	0,25 µl
Tampón 10X (NEB2)	3 µl
H_2O milli Q	6,25 µl

Incubación a 37 ^{o}C de 3 horas a toda la noche. Se paró la reacción incubando a 80^{o}C durante 20 minutos.

Comprobación de la digestión mediante electroforesis en gel de agarosa al 1 % y 100 V.

2- Preparación de adaptadores y ligación: se utilizaron adaptadores complementarios a la zona de corte de ambas enzimas (*Mse*I1: 5´-GAGTCCTGAGTAGCAG, *Mse*I2: 5´-TACTCAGGACTCAT, *Pst*I1: 5´-CTCGTAGACTGCGTACATGCA y *Pst*I2 5´-TGTACGCAGTCTAC). Se trata de adaptadores que dejan extremos cohesivos 5' para *Mse*I y 5' y 3' para *Pst*I, en función del corte de cada una de las enzimas. El adaptador *Mse*I se utiliza a una concentración final de 50 pmol/µl, mientras que el adaptador *Pst*I se necesita a una concentración final de 25 pmol/µl. Antes de utilizarlos en la reacción de ligación se tienen que desnaturalizar incubándolos a 95^{o}C durante 3 minutos y dejar que renaturalicen a temperatura ambiente durante 15 minutos. La reacción de ligación es la siguiente:

ADN digerido	25 µl
Adaptador *Mse*I (50 pmol/µl)	1 µl
Adaptador *Pst*I (25 pmol/µl)	0,5 µl
Tampón de ligación 10X (NEB)	3 µl
Ligasa de T4 400 U/µl (NEB)	0,5 µl

Se incuba a temperatura ambiente durante 3 horas o toda la noche a 4^{o}C. A continuación se inactiva la ligasa por choque término a 65^{o}C durante 10 minutos.

3- Pre-amplificación: se realizó una primera amplificación utilizando cebadores para los adaptadores *Mse*I (5´-GATGAGTCCTGAGTAA) y *Pst*I (5´-GACTGCGTACATGCAG). La reacción es la siguiente:

ADN ligado (dilución 1/10)	4 µl
Cebador *Mse*I (5 pmol/µl)	4 µl
Cebador *Pst*I (5 pmol/µl)	4 µl
Tampón de amplificación 10X (5prime)	2,5 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	2,5 µl
Taq polimerasa (5prime)	0,2 µl
H_2O milli Q	7,8 µl

El programa de PCR en este caso es el siguiente: 1min a 95^{o}C, 30 ciclos (1 min 94^{o}C, 1 min 60^{o}C, 1 min 72^{o}C) y 7 min a 72^{o}C.

Las amplificaciones se observan en geles de agarosa al 1,5% y 80 V.

En este paso se podrían añadir bases selectivas a los cebadores si se observan demasiadas bandas en los geles. En este experimento no se añadió ningún nucleótido selectivo a los cebadores.

4- Amplificación SSAP: en esta reacción se utilizó un cebador diseñado a partir de una LTR y el cebador complementario al adaptador de una de las enzimas. Se realizaron reacciones de 10 µl como se muestra a continuación:

ADN preamplificado	1,5 µl
Cebador enzima (5 pmol/µl)	0,6 µl
Cebador SSAP (5 pmol/µl)	1,5 µl
Tampón de amplificación 10X (Biotools)	1 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	1 µl
Taq polimerasa Hot Split (Biotools)	0,5 µl
$MgCl_2$	1 µl
H_2O milli Q	2,9 µl

El programa de PCR consistió en una reacción "touch-down" con las siguientes características: 5 min a 95^{o}C, 13 ciclos (1 min 94^{o}C, 1 min 65^{o}C disminuyendo 0,7^{o}C por ciclo, 1 min 72^{o}C), 22 ciclos (1 min 94^{o}C, 1 min 56^{o}C, 1 min 72^{o}C) y 7 min a 72^{o}C.

5- Electroforesis capilar: las muestras amplificadas se analizaron mediante electroforesis capilar llevada a cabo en un ABI PRISM 310 Genetic Analyzer (Applied Biosystems) cuyo esquema se muestra en la (figura 3.2). Los cebadores deben estar marcados con fluorocromos en el extremo 5´ para permitir su visualización mediante este tipo de electroforesis. Los fluorocromos utilizados fueron FAM para LTR1, LTR2 y LTR3 y JOE para LTR4. Los cebadores LTR1, 2 y 3 se pueden utilizar y analizar simultáneamente con LTR4 por tener fluorocromos diferentes. Además, se utiliza un

control de peso molecular marcado con un tercer fluorocromo, en este caso se utilizó ROX-500 (Applied Biosystems).

Las muestras se preparan previamente:

ADN amplificado	2 µl
GeneScan-500 Rox Size Standard	0,5 µl
Formamida	12 µl

Se desnaturalizan las muestras incubando 5 minutos a 95^{o}C y se introducen inmediatamente en un baño de hielo durante, al menos, cinco minutos antes del análisis.

La electroforesis se realizó con el polímero POP4 (Applied Biosystems) a una temperatura de 60^{o}C para evitar la renaturalización de las muestras y a 15 kV durante 27 minutos. La carga de muestras se realiza por ionización a 15 kV durante 18 segundos.

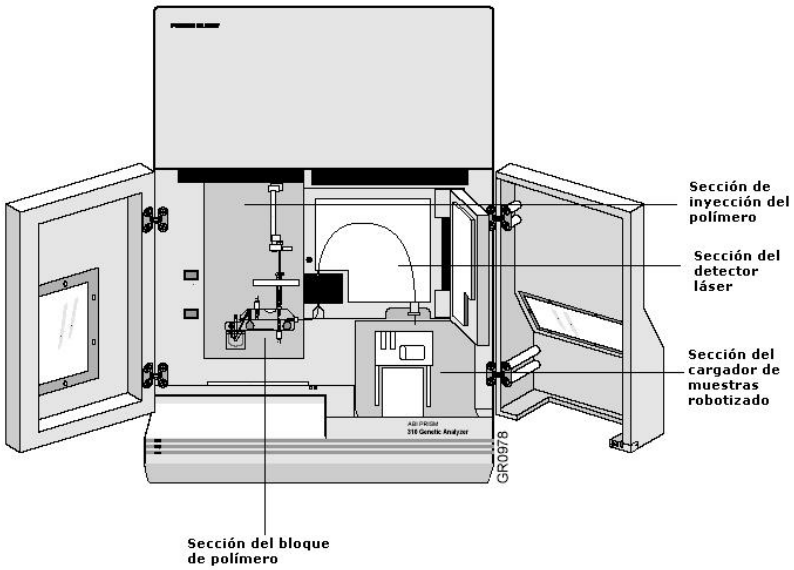


Figura 3.2: Analizador de fragmentos de ADN, ABI Prism 310.

El analizador ABI Prism está controlado mediante ordenador por el software específico Data Collection v 3.1. Los datos se analizaron con el programa Gene Mapper v 4.0 de Applied Biosystems.

Con este programa se analizan los electroferogramas obtenidos por electroforesis capilar y se obtienen matrices de genotipos para cada una de las combinaciones de cebadores.

Los polimorfismos detectados deben ser comprobados de forma manual comparando los electroferogramas de cada una de las muestras con los valores asignados en las matrices, para ello se exportan éstas a una hoja de cálculo donde, mediante una serie de "macros", se convierte la matriz de genotipos en una matriz de presencia (1) - ausencia (0) que permite comprobar fácilmente si los valores asignados en las matrices son los correctos.

3.11.2. REMAP

Esta técnica (REtrotransposon-Microsatellite Amplified Polymorphisms) consiste en una amplificación por PCR utilizando como cebadores oligonucleótidos diseñados a partir de las LTRs de los retrotransposones y microsatélites (SSR).

En este trabajo se han utilizado cebadores diseñados a partir de microsatélites de dos, tres y cuatro repeticiones: $(AC)_{10}$, $(GT)_{10}$, $(CT)_{10}$, $(GAA)_5$, $(GCA)_5$, $(GACA)_4$, así como secuencias diseñadas a partir de zonas flanqueantes de microsatélites localizados en el genoma de *L. culinaris* cv. *ILL5588* (Hamwieh et al., 2005) (Tabla 3.2).

La reacción de PCR es la siguiente:

ADN genómico (20 ng/µl)	4 µl
Cebador SSR (5 pmol/µl)	0,8 µl
Cebador LTR (5 pmol/µl)	0,8 µl
Tampón de amplificación 10X (5Prime)	2 µl
dNTPmix (2 mM de cada uno de los 4 dNTPs)	2 µl
Taq polimerasa (5Prime)	0,2 µl
H_2O milli Q	10,2 µl

El programa de PCR en el termociclador es de 4 min a 95^{o}C, 32 ciclos (40 s 94^{o}C, 40 s 60^{o}C, 2 min 72^{o}C) y 5 min a 72^{o}C. Debido a las diferentes temperaturas de

apareamiento de los cebadores (Tabla 3.5) fue necesario modificar alguno de los parámetros de la PCR y, en algunos casos se realizaron reacciones "touch-down".

Una vez terminadas las reacciones, se visualizaron en un gel de agasora al 1% y 80 V.

Tabla 3.2: Cebadores diseñados a partir de zonas flanqueantes de microsatélites (SSRs) de lenteja.

Cebador	Secuencia F	Secuencia R	Tª óptima
SSR13	5' GAAACAACACCGAAATACAC 3'	5' CGAAGTCAGATGAAGTTTG 3'	53 °C
SSR 19	5' GACTCATACTTTGTTCTTAGCAG 3'	5' GAACGGAGCGGTCACATTAG 3'	58 °C
SSR 33	5' CAAGCATGACGCCTATGAAG 3'	5' CTTTCACTCACTCAACTCTC 3'	56 °C
SSR 48	5' CATGGTGGAATAGTGATGGC 3'	5' CTCCATACACCACTCATTCAC 3'	57 °C
SSR 59-2	5' CCAAATACTGCAACACACCG 3'	5' GTTCCCATCAGGCAGAAGG 3'	58 °C
SSR 80	5' CCATGCATACGTGACTGC 3'	5' GTTGACTGTTGGTGTAAGTG 3'	56 °C
SSR 96	5' GTTATCTTCCAGCGTC 3'	5' GATATACAATCAGAGATG 3'	49 °C
SSR 99	5' GGGAATTTGTGGAGGGAAG 3'	5' CCTCAGAATGTCCCTGTC 3'	57 °C
SSR 107	5' GCGGCGAGCAAATAAAT 3'	5' GGAGAATAAGAGTGAAATG 3'	51 °C
SSR 113	5' CCGTAAGAATTAGGTGTC 3'	5' GGAAAATAGGGTGGAAAG 3'	51 °C
SSR 124	5' GTATGTGACTGTATGCTTC 3'	5' GCATTGCATTTCACAAACC 3'	52 °C
SSR 130	5' CCACGTATGTGACTGTATG 3'	5' GAAAGAGAGGCTGAAACTTG 3'	55 °C
SSR 151	5' GGTAGGTGAGATAGTTG 3'	5' GGAGCAAGAAGAAGCAG 3'	51 °C
SSR 154	5' GGAATTTATCACACTATCTC 3	5' GACTCCCAACTTGTATG 3'	51 °C
SSR 156	5' GTACATTGAACAGCATCATC 3	5' CAAATGGGCATGAAAGGAG 3'	53 °C
SSR 167	5' CACATATGAAGATTGGTCAC 3'	5' CATTTATGTCTCACACACAC 3'	54 °C
SSR 184	5' GTGTGTACCTAAAGCCTTG 3'	5' GTAAGTTGATCAAACGCCC 3'	55 °C
SSR 199	5' GTGTGCATGGTGTGTG 3'	5' CCATCCCCCTCTATC 3'	51 °C
SSR 204	5' CACGACTATCCCACTTG 3'	5' CTTACTTTCTTAGTGCTATTAC 3'	53 °C
SSR 212-1	5' GACTCATTGTTGTACCC 3'	5' GCGAGAAGAATGGTTG 3'	50 °C
SSR 213	5' CACTCGCACCTCTTATG 3'	5' GGCCGAGGTTGTAACAC 3'	51 °C
SSR 215	5' CATTAATATTTCTTTGGTGC 3'	5' CTTTTCTTCTCTTCCCC 3'	50 °C
SSR 233	5' CTTGGAGCTGTTGGTC 3'	5' GCCGCCTACATTATGG 3'	52 °C
SSR 302	5' CAAGCCACCCATACACC 3'	5' GGGCATTAAGTGTGCTGG 3'	56 °C
SSR 309-2	5' GTATGTCGTTAACTGTCGTG 3'	5' GAGGAAGGAAGTATTCGTC 3'	50 °C
SSR 317-1	5' AGTGACAACAAAATGTGAGT 3'	5' GTAGCAATAATTACACCCAC 3'	53 °C
SSR 317-2	5' CACGTAACATCTTGCTTATG 3'	5' GTATCAAACTTATGGTGAAATC 3'	53 °C
SSR 323	5' GTGGGTGTAATTATTGCTAC 3'	5' GTACCTAGTTTCATCATTG 3'	51 °C
SSR 336	5' GTGTAACCCAACTGTTCC 3'	5' GAAATTGTCTCTTAGCAAG 3'	54 °C

[1] Los nombres de los cebadores son los utilizados por Hamwieh *et al.*, 2005

3.11.3. iPBS

Se trata de las mismas reacciones de PCR descritas en el apartado 3.5.2. En este caso se analiza la presencia o ausencia de las bandas amplificadas en la población RIL del cruzamiento estudiado para generar marcadores.

Algunos de estos cebadores PBS se probaron también en todas los parentales de *Lens* que se han utilizado en este laboratorio para realizar cruzamientos.

3.12. Mapeo genético

Para realizar los mapas genéticos de ligamiento se tuvieron en cuenta los marcadores moleculares polimórficos que presentaban una segregación esperada para las líneas RIL, es decir, una segregación 1:1. Para comprobar el tipo de segregación se realizó una prueba χ^2, cuando el valor del χ^2 observado mostró una probabilidad $p > 0,05$ se aceptó la hipótesis nula y, por lo tanto, se incluyó en el estudio. Además de esta segregación, también se incluyeron los marcadores que presentaran una segregación 3:1 la cual se correspondería con dos bandas polimórficas del mismo tamaño con segregación 1:1.

Se utilizaron los programas informáticos MapMaker v 3.0b (Lander *et al*., 1987) y CarthaGene v 1.3 (de Givry *et al*., 2005) para determinar las fracciones de recombinación entre parejas de marcadores y obtener los diferentes grupos de ligamiento, así como establecer el orden de los marcadores en los mismos.

3.12.1. Cálculo de la fracción de recombinación

Para calcular la fracción de recombinación se calcula la proporción de gametos recombinantes que forma un heterocigoto, frente al número de gametos totales bajo criterios de máxima verosimilitud.

Se utilizó el método de puntuación LOD (Logarithm of odds ratio) para medir la probabilidad de ligamiento:

$$LOD = \log \frac{P_r}{P_{(0,5)}}$$

siendo P_r la probabilidad de obtener la población segregante analizada considerando una fracción de recombinación r y $P_{(0,5)}$ la misma probabilidad considerando r = 0,5, situación de independencia. Se admitirá el ligamiento cuando la relación sea de al

menos 10000:1 (LOD = 4) y se rechazará cundo sea de menos de 100:1 (LOD = 2). La fracción de recombinación se determina cuando el valor de LOD es máximo.

3.12.2. Construcción del mapa genético

Para obtener un mapa genético es necesario seguir los pasos que se detallan a continuación.

- **Determinación de los grupos de ligamiento**

Los grupos de ligamiento se establecieron bajo los criterios de LOD superior a 4 y fracción de recombinación menor de 0,50.

- **Ordenación de los marcadores**

La ordenación lineal de los marcadores en cada uno de los grupos de ligamiento debe ser la que permita obtener con mayor probabilidad los datos observados. Ya que el número de órdenes posibles para n marcadores es n!/2, no sería posible analizar todos los órdenes posibles para los grupos de ligamiento que presenten un número relativamente alto de marcadores. Por ello, se han desarrollado algoritmos que permiten determinar el orden mediante análisis más sencillos (Tan y Fu, 2006). En este tipo de algoritmos están basados los programas informáticos que permiten la construcción de mapas.

Para obtener los mejores mapas con el programa MapMaker es necesario seguir una serie de pasos diferentes en función del tamaño del grupo de ligamiento:

- Cuando se trabaja con grupos de menos de siete marcadores, se utiliza el comando COMPARE de manera que se obtienen los valores de máxima verosimilitud para cada uno de los órdenes posibles de los marcadores

- Cuando se tienen grupos de más de siete marcadores se realiza un análisis de tres puntos. Este análisis consiste en comparar cada uno de los tripletes posibles de loci ligados para establecer el orden más probable y, a continuación se utiliza el comando ORDER varias veces para obtener un esqueleto alrededor del cual se colocan, uno a uno, el resto de marcadores utilizando el comando TRY, que permite situar cada uno de

ellos en el lugar más probable dentro del mapa obtenido anteriormente, finalmente el comando COMPARE permite asegurar que se ha obtenido el orden de los marcadores que hace máxima la probabilidad de la descendencia.

En cuanto al programa CarthaGene, se utilizan diferentes algoritmos genéticos con los que se obtiene el mapa más probable incluyendo todos los marcadores de cada grupo de ligamiento:

- Annealing: partiendo de un mapa inicial se va alterando el orden de los marcadores hasta maximizar determinados parámetros del programa.

- Taboo: se basa en realizar cambios en el orden entre marcadores adyacentes de un mapa de partida evitando aquellos que el programa considera tabú, hasta alcanzar valores de máxima verosimilitud.

- Genetic: es un algoritmo basado en una simulación de un proceso de reproducción-mutación-selección.

El último paso en la construcción del mapa, es la comprobación del orden de los marcadores de cada grupo. Para ello se utiliza el comando RIPPLE de MapMaker o la opción FLIP de CarthaGene. En ambos casos, el programa comprueba todas las permutaciones posibles para cada grupo de cinco marcadores adyacentes e indica cuál de ellas es la más probable.

- **Cálculo de las distancias genéticas**

La fracción de recombinación no es aditiva por lo que se necesitan funciones de mapa que relacionen la distancia de mapa con las fracciones de recombinación que se han calculado.

Las dos funciones de mapa más utilizadas son:

- Función de Haldane (Haldane, 1919): esta función se basa en la inexistencia de interferencia y asume que los sobrecruzamientos suceden al azar a lo largo de todo el genoma.

$$D = -\frac{1}{2}\log(1-2r)$$

donde D es la distancia de mapa y r la fracción de recombinación.

- Función de Kosambi (Kosambi, 1944): esta función tiene en cuenta la presencia de cierto grado de interferencia, la probabilidad de que ocurra un sobrecruzamiento en una zona determinada es menor si ya se ha producido otro en la misma zona o en sus proximidades

$$D = \frac{1}{4}\log\frac{(1+2r)}{(1-2r)}$$

donde D es la distancia de mapa y r la fracción de recombinación.

En este trabajo se han realizado los mapas utilizando la función de Kosambi ya que los supuestos de esta función se acercan más a la realidad.

3.13. Otros programas informáticos utilizados

- **Total Lab**: (Phoretix): se trata de un programa de análisis de electroforesis. Permite detectar bandas en los geles para determinar de forma semiautomática, su presencia o ausencia y poder analizar la segregación de los marcadores. Se utilizó para analizar los geles en los que aparecían bandas con pequeñas diferencias de tamaño y que, simplemente por observación, eran difíciles de distinguir.

- **MapChart v 2.2**: (Voorrips, 2002): este programa permite obtener los gráficos correspondientes a los mapas de ligamiento obtenidos con los marcadores analizados.

RESULTADOS Y DISCUSIÓN

4. Resultados y discusión

4.1. Búsqueda de marcadores derivados de retrotransposones en el genoma de lenteja

Para la búsqueda de retrotransposones en el genoma de lenteja se han utilizado dos tipos de cebadores según la secuencia que se ha utilizado para su diseño. Así se estudiaron cebadores derivados del gen de la retrotranscriptasa y cebadores derivados de otras regiones internas de un retrotransposón de lenteja

La búsqueda se ha realizado sobre el ADN genómico procedente de diferentes individuos de la variedad cultivada Lupa de *Lens culinaris*.

4.1.1. Cebadores derivados del gen de la Retrotranscriptasa

El primer tipo de cebadores utilizado en este trabajo está basado en la secuencia del gen de la retrotranscriptasa o transcriptasa inversa. Este gen ha sido el elegido en numerosos estudios debido a que es la proteína más conservada tanto en retrovirus como en retrotransposones, dada la presencia de dominios altamente conservados con actividad enzimática (Doolittle *et al*., 1989).

Para la amplificación de retrotransposones del tipo *Ty1-copia* se han utilizado cebadores degenerados derivados de dos de sus dominios conservados, como son QMDVKT y YVDDM (Hirochika y Hirochika, 1993), y para los retrotransposones *Ty3-gypsy* los dominios conservados a partir de los cuales se obtuvieron los cebadores fueron RMCVDYR y YAKLSKC (Friedsen *et al*., 2001).

Las diferentes reacciones de PCR dieron como resultado un patrón de bandas característico para cada uno de los tipos de retrotransposones (Figura 4.1), de manera que cada una de las bandas obtenidas se pudo aislar para permitir su clonación en plásmidos, la transformación de células de *E. coli* y, por último, la secuenciación de los amplicones para su posterior análisis. Se observó que los patrones de bandas obtenidos con cada pareja de cebadores eran constantes entre los individuos de Lupa utilizados, como se esperaba de individuos de un cultivar genéticamente uniforme.

Para la representación de las secuencias se ha utilizado la siguiente nomenclatura:

- CLc para secuencias obtenidas con los cebadores de *Ty1-copia*.
- GLc para secuencias obtenidas con los cebadores de *Ty3*-gypsy.

A continuación se indica el número de la banda seguido de un guion y el número del clon correspondiente.

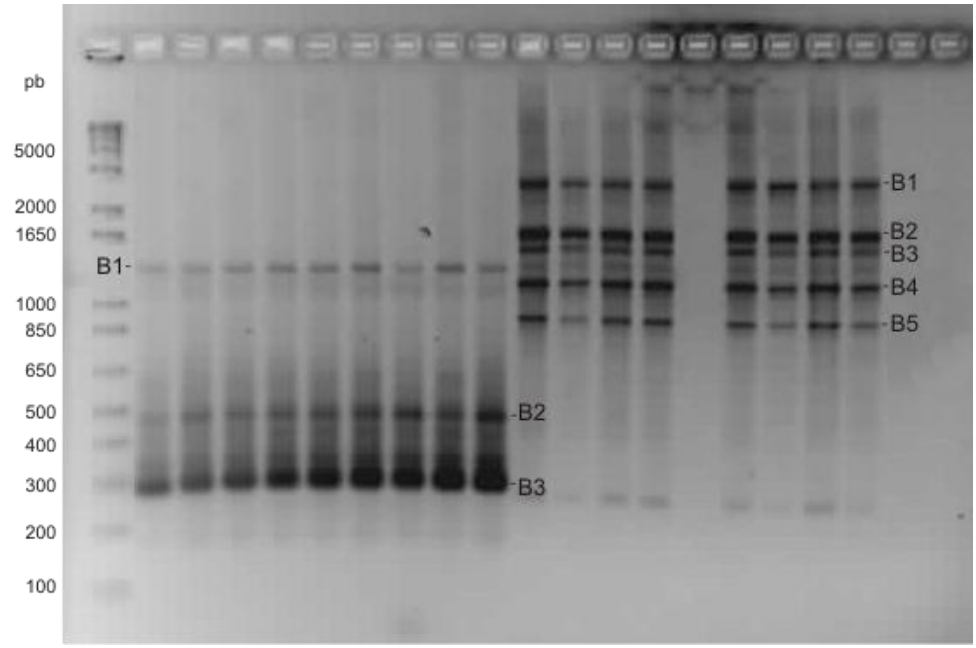


Figura 4.1: Bandas obtenidas con los cebadores diseñados para *Ty1-copia*, a la izquierda, y *Ty3-gypsy*, a la derecha. Se indican las bandas aisladas para su clonación y secuenciación. Cada uno de los carriles se corresponde con un individuo diferente de Lupa. A la izquierda se indica el tamaño de cada una de las bandas de un marcador de peso molecular utilizado como control.

4.1.1.1. Retrotranscriptasa de retrotransposones *Ty1-copia*

El fragmento de la retrotranscriptasa esperado tras la amplificación con los cebadores para el tipo de retrotransposón *Ty1-copia* tiene unas 270 pb, lo que correspondería con la banda de menor peso molecular (banda 3); sin embargo se procedió al aislamiento de las tres bandas visualizadas para comprobar el resultado de la amplificación.

Tras el aislamiento de las bandas se procedió a su transformación y clonación, de manera que, en placas de medio LB con ampicilina, se obtuvieron dos tipos de colonias: transformantes recombinantes (blancas), que incluyen la secuencia aislada, y

no recombinantes (azules), que carecen de la misma. De esta manera se puede observar la diferencia de tamaño de las bandas entre las colonias recombinantes y las no recombinantes que se utilizan como control (Figura 4.2).

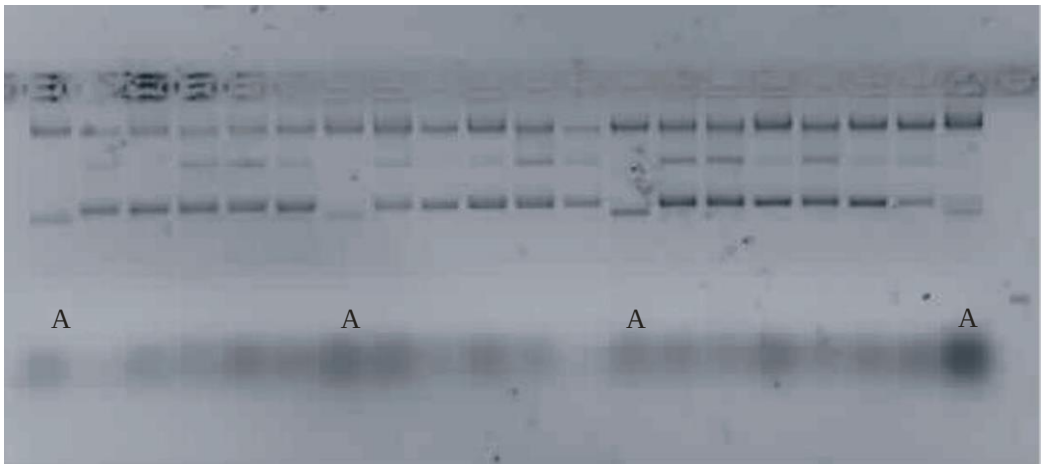


Figura 4.2: Minipreparación ultrarrápida de las colonias positivas. Con la letra A se indican las colonias azules que se utilizaron como control. Se puede observar el menor peso molecular de la banda sin inserto frente al resto de bandas correspondientes a colonias transformadas.

En el caso de la banda 1, tras la realización de la transformación y la clonación, se llevaron a cabo minipreparaciones de 48 de las colonias crecidas en las placas, sin embargo no se observaron diferencias de tamaño entre las colonias recombinantes y las control, a pesar de que debían presentar una diferencia de aproximadamente 1500 pares de bases, de manera que no se procedió a la secuenciación de dichos clones.

A partir de la banda 2 se obtuvieron clones que se pudieron secuenciar sin problemas, sin embargo, al comparar estas secuencias con otras publicadas en las bases de datos, se observó que presentaba similitudes con secuencias de retrotranscriptasas correspondientes a retrotransposones tipo *Ty3-gypsy* por lo que se estudiarán con las secuencias amplificadas para *Ty3-gypsy* (ver página 91). En este caso se secuenciaron 12 colonias, tras su comprobación mediante una minipreparación y, a partir de éstas, se obtuvieron siete secuencias diferentes.

Tras la realización de las minipreparaciones de las colonias obtenidas a partir del aislamiento de la banda 3, se escogieron 23 colonias al azar para su secuenciación. Una

vez obtenidas las secuencias, se alinearon utilizando el método ClustalW (Thompson *et al*., 1994) incluido en el programa Mega5 (Tamura *et al*., 2011), de esta manera, se observó que los fragmentos amplificados no estaban formados por una secuencia única si no que aparecían diferencias entre unas y otras de manera que se pueden distinguir 21 elementos diferentes.

Para obtener mayor información de las secuencias nucleotídicas de los diferentes clones, se dedujeron las secuencias aminoacídicas y todas ellas se compararon entre sí y con secuencias de otros retrotransposones tipo *copia* ya descritos en diferentes especies y presentes en la base de datos GeneBank del NCBI ("Nacional Center for Biotechnology Information") utilizando los algoritmos BLASTn (permite la comparación entre secuencias de nucleótidos) y BLASTp (comparación entre secuencias de aminoácidos) (Altschul *et al*, 1990), de manera que se obtuvieron homologías con fragmentos de retrotransposones de diversas especies vegetales (Tabla 4.1: ver apartado 4.1.3.) y se pudieron definir siete grupos en función del nivel de identidad, más del 90% de aminoácidos comunes entre las 21 secuencias analizadas.

Como se puede observar en la Figura 4.3, diez de las secuencias obtenidas (CLc3-20, CLc3-02, CLc3-22, CLc3-14, CLc3-04, CLc3-07, CLc3-16, CLc3-11, CLc3-6, CLc3-23) presentan codones de stop internos y además se pueden observar deleciones en las secuencias CLc3-22, CLc3-14 y CLc3-19, lo que implicaría cambios en los marcos de lectura. Estas modificaciones en las secuencias conllevan la pérdida de función de las proteínas codificadas.

El 50% de las secuencias observadas presenta un marco de lectura continuo lo que indicaría un alto grado de conservación de la retrotranscriptasa de estos elementos en el genoma de *Lens culinaris* y la posibilidad de que puedan ser activos de forma independiente.

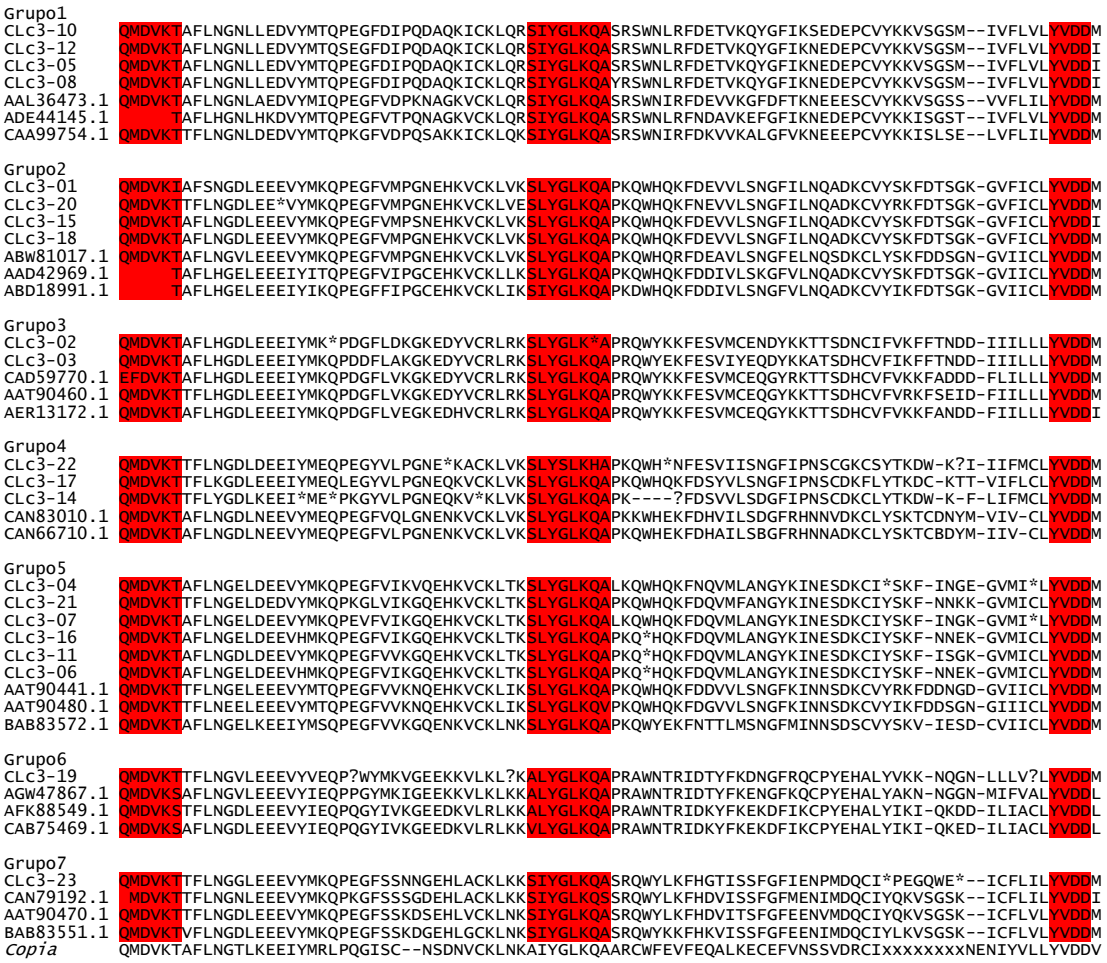


Figura 4.3: Secuencias de aminoácidos obtenidas con los cebadores diseñados a partir del gen de la retrotranscriptasa de retrotransposones *Ty1-copia.* Estas secuencias se comparan con otras secuencias publicadas en la base de datos del NCBI (AAL36473.1: *Setaria faberi*; ADE44145.1: *Corchorus olitorius*; CAA99754.1: *Oryza sativa*; ABW81017.1: *Arabodopsis lyrata*; AAD42969.1: *Orobanche ramosa*; ABD18991.1: *Orobanche cernua*; CAD59770.1: *Cicer arietinum*; AAT90460.1: *Vigna radiata*; AER13172.1: *Phaseolus vulgaris*; CAN83010.1: *Vitis vinifera*; CAN66710.1: *Vitis vinifera*; AAT90441.1: *Vigna radiata*; AAT90480.1: *Vigna radiata*; BAB83572.1: *Silene latifolia*; AGW47867.1: *Phaseolus vulgaris*; AFK88549.1: Arabidopsis halleri; CAB75469.1: Arabidopsis thaliana; CAN79192.1: *Vitis vinifera*; AAT90470.1: *Vigna radiata*; BAB83551.1: *Silena gallica*). Con el fondo rojo se indican los motivos conservados de la proteína. La última secuencia se corresponde con la secuencia consenso descrita para *copia* por Xiong y Eickbush (1990).

Estas polimerasas se descubrieron en retrovirus en 1970 (Baltimore; Temin y Mizutani) y desde entonces se han localizado en retrovirus tanto animales como vegetales (Toh *et al*., 1983), en retrotransposones LTR (Saigo *et al*., 1984) y no LTR (Fawcett *et al.*, 1986) y en diferentes grupos de mitocondrias (Michel y Lang, 1985) y, aunque se han localizado otras regiones codificantes conservadas entre algunos de estos elementos, la retrotranscriptasa es la única proteína que presenta dominios conservados

en todos ellos (Xiong y Eickbush, 1990). Por ello, los cebadores utilizados en el presente trabajo también se hayan usado para identificar retrotransposones tipo *Ty1-copia* en plantas como tomate (*Lycopersicum chilense*) (Yañez *et al*., 1998), cebada (*Hordeum vulgare*), trigo (*Triticum aestivum*) (Gribbon *et al*., 1999), en coníferas (Friedsen *et al*., 2001) o judía mungo (*Vigna radiata*) (Dixit *et al*., 2006).

A pesar de las numerosas diferencias presentes entre las distintas secuencias, todas ellas presentan motivos funcionales conservados, como son la secuencia SLYGLKQA propia de retrotransposones tipo *copia* (Flavell *et al*., 1992; Hirochika y Hirochica, 1993; Rico, 2005) y, lógicamente, las secuencias utilizadas para diseñar los cebadores.

La amplificación de secuencias del genoma de lenteja con los cebadores degenerados para *copia*, produce resultados que se corresponden con los obtenidos por otros autores. Así, Friedsen (2001), realizó el estudio de diversas especies de diferentes familias de coníferas y, al igual que en el presente estudio, obtuvo una banda de 270 pb correspondiente a dicha familia de retrotransposones (el 95% de las secuencias analizadas presentaba homología con retrotransposones *Ty1-copia*) y, en alguna de las especies estudiadas, otra de 420 pb que se correspondía con retrotransposones del tipo *Ty3-gypsy*, como ha ocurrido con algunas de las secuencias amplificadas en lenteja.

Además, el uso de estos cebadores para amplificar fragmentos del gen de la retrotranscriptasa ha permitido describir diversas familias de retrotransposones tipo *Ty1-copia* en muchas especies de plantas, lo que demuestra su naturaleza universal y la conservación de dichos elementos a lo largo del reino vegetal (Flavell *et al*., 1992; Voytas *et al*, 1992.).

Para poder analizar las relaciones evolutivas entre los diferentes tipos de secuencias se utilizó el programa Mega5 que permite realizar mediante un análisis de máxima verosimilitud comparaciones entre 24 modelos de evolución diferentes. De esta manera, se puede estimar la distancia entre parejas de secuencias según el modelo que se ajuste mejor a los datos y, por lo tanto, los explique mejor. Así, se eligió el modelo

de Tamura con tres parámetros, que considera la diferencia de probabilidad con la que se producen transiciones y transversiones así como el efecto del contenido de G+C sobre las sustituciones nucleotídicas, y una distribución gamma, que valora la heterogeneidad de los cambios en los diferentes sitios de la secuencia (Tamura, 1992). Con este modelo se obtuvo una matriz de distancias a partir de la cual se diseñó un árbol filogenético basado en el algoritmo Neighbor Joining (Saitou y Nei, 1987), que permite agrupar las secuencias sin asumir la hipótesis de reloj molecular en su evolución. Para comprobar la fiabilidad de los nodos del árbol se ha utilizado un análisis "bootstrap" realizando 1000 remuestreos.

Aunque el gen de la retrotranscriptasa está muy conservado y es semejante en todas las secuencias, en el árbol se pueden observar claramente los siete grupos determinados por identidad en las búsquedas BLAST los cuales se indican con diferentes colores en la Figura 4.4.

En el árbol, se puede ver como cada uno de los grupos de secuencias se une de forma monofilética con una fiabilidad en el "bootstrap", que varía del 85% en el caso del grupo 3 (amarillo) al 100% en el resto.

Asimismo, se puede observar que los grupos 2 (azul) y 5 (marrón) son clados hermanos unidos con una confianza del 82%. Estos dos clados presentan homología tanto con retrotransposones de leguminosas como con los de otras especies como *Arabidopsis lyrata* u *Orobanche ramosa.*

El grupo 4 (verde) presenta homología con retrotransposones de diferentes individuos de *Vitis vinífera* con una confianza del 84%.

El grupo 3 (amarillo) es característico de leguminosas, se une con diferentes especies de éstas como *Cicer arietinum* y *Phaseolus vulgaris* con una fiabilidad del 100%.

El grupo 7 (gris) está formado por una única secuencia que se une con diferentes especies (*Vitis vinifera* y *Silene gallica*) con una confianza del 100%.

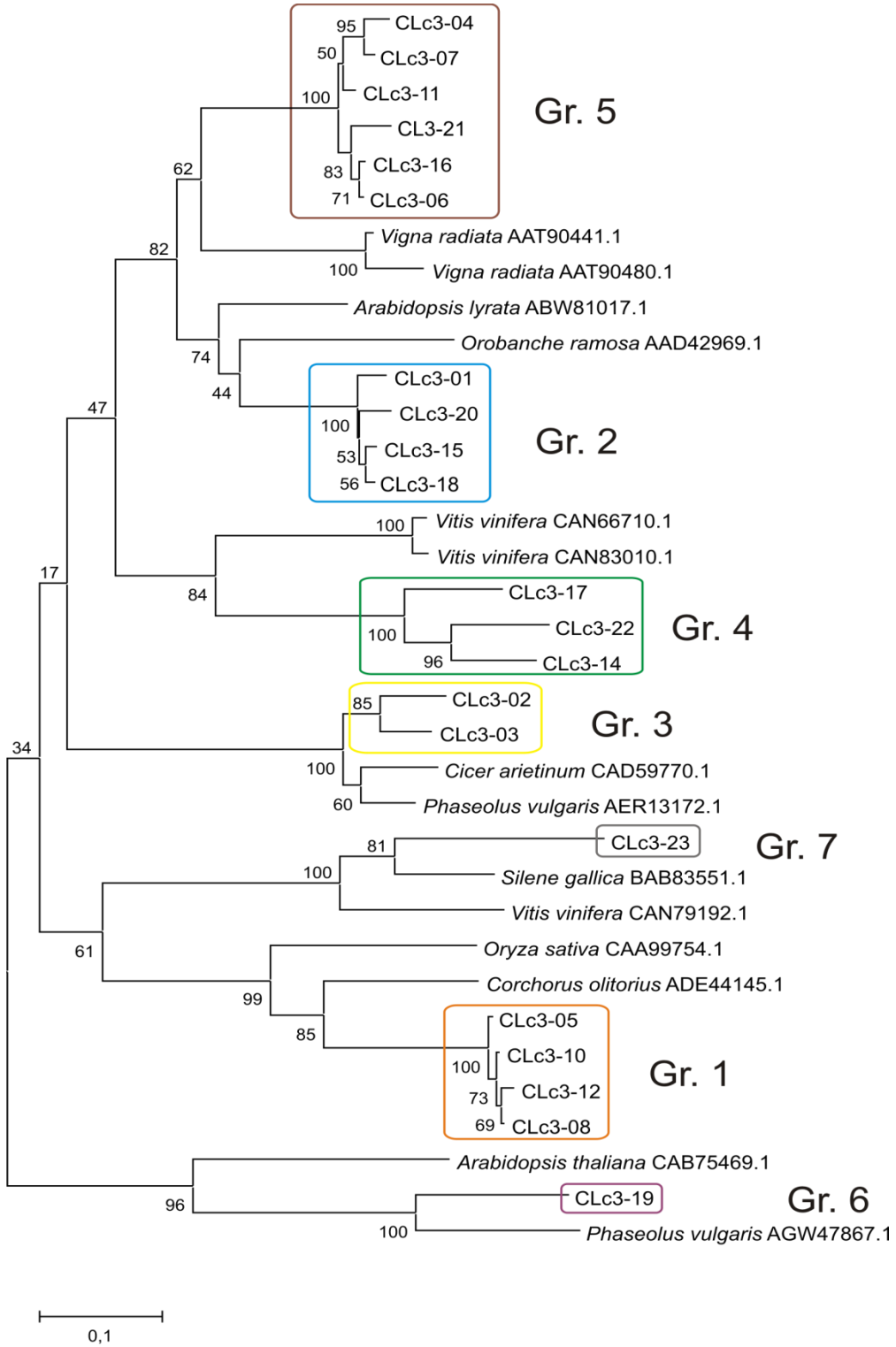


Figura 4.4: Árbol de la retrotranscriptasa de retrotransposones tipo *copia*. La divergencia entre las secuencias está definida por la longitud de las ramas y expresado en unidades de distancia genética. En cada nodo se indican los valores de "bootstrap". Se han recuadrado con diferentes colores los distintos grupos de secuencias obtenidos.

En cuanto al grupo 1 (naranja), se puede observar que apareció antes del origen de las monocotiledóneas ya que presenta homología con especies de este grupo (*Oryza sativa*) y con eudicotiledóneas (*Corchuros olitorius*) con las que se une con una fiabilidad del 99% y se une con el grupo 7 con un valor muy bajo de "bootstrap" (61%). El origen anterior de determinados transposones a la aparición de las monocotiledóneas, también se ha observado en otros trabajos que describen retrotransposones tipo *copia* en diferentes especies de angiospermas (Flavell *et al*., 1997; Woodrow *et al*., 2012).

Por último, el grupo 6 (morado), presenta homología tanto con retrotransposones de leguminosas como de *Arabidopsis*, con la que se une con una fiabilidad en el "bootstrap" del 96%.

Dentro de cada grupo, las diferentes secuencias de lenteja parecen copias parálogas de origen reciente como en los grupos uno, dos y seis, o más antiguas como en los grupos tres y cuatro.

Se han realizado numerosos estudios filogenéticos basados en la secuencia de la retrotranscriptasa de retrotransposones del tipo *Ty1-copia*. Uno de los trabajos más importantes es el realizado por Flavell *et al.* (1992) en el que se identificaron retrotransposones tipo *copia* en especies representativas de todos los grupos de plantas superiores y se encontraron secuencias similares en algunas de las especies estudiadas, lo que podría indicar un origen remoto de dichos elementos y su transmisión a las líneas germinales de las especies divergentes. Lo mismo ocurre en el trabajo de Dixit *et al*. (2006) el en que se obtuvieron seis grupos de secuencias que mostraban similitudes tanto con especies relacionadas como con otras muy alejadas evolutivamente. Por otra parte, el hecho de que en algunos grupos no haya secuencias de leguminosas similares a las de lenteja pero sí otras de especies más alejadas evolutivamente es probablemente el resultado de lo fragmentario, incompleto y desigual de la representación de secuencias de retrotransposones en las bases de datos.

4.1.1.2. Retrotranscriptasa de retrotransposones *Ty3-gypsy*

Las secuencias esperadas tras la amplificación con los cebadores para la retrotranscriptasa de los retrotransposones del tipo *Ty3-gypsy* deberían tener unos 420 pb; sin embargo, todas las bandas obtenidas son mayores de lo esperado, así que se procedió al aislamiento de todas ellas (Figura 4.1) para determinar las secuencias de los diferentes fragmentos amplificados.

Tras el aislamiento de las bandas se siguió con la transformación y clonación de la misma forma que con los amplificados de *copia*. A continuación se seleccionaron 16 colonias procedentes de cada una de las bandas y se comprobó la presencia de inserto mediante minipreparaciones de colonias transformantes y recombinantes. Algunas de las colonias seleccionadas carecían de inserto de tamaño apreciable, de manera que se secuenciaron 14 colonias de la banda 1, 10 de la banda 2, 11 de la banda 3, 9 de la banda 4 y 11 de la banda 5.

A continuación, se comprobó que muchas de los fragmentos analizados presentaban el cebador GypsyF en ambos extremos lo que podría deberse al alto nivel de degeneración que presentan los cebadores utilizados (casi el 50% de las posiciones de dichos cebadores están degeneradas), aunque también podría deberse a una deleción en la parte interna del retrotransposón, de manera que se hubiese perdido la zona de unión al cebador inverso o a una mutación en la secuencia de dicha región, lo que habría provocado la unión inespecífica del cebador directo por falta de la región complementaria al inverso. Esto ha complicado el análisis de las secuencias ya que, únicamente, uno de los extremos de la mayoría de las secuencias presenta homología con retrotransposones, mientras que el otro no coincide con dichos elementos, de hecho, en algunos casos, ni siquiera se encuentran homologías en las bases de datos. Aun así, la mayor parte de estas secuencias se corresponden con el gen de la retrotranscriptasa y, en algunos casos, se ha llegado a amplificar un fragmento de la RNasaH. Además, una vez alineadas, se observó que el nivel de identidad entre alguna de las secuencias obtenidas a partir de diferentes bandas, era muy alto.

Tras el alineamiento de todas las secuencias obtenidas se han definido tres grupos de secuencias, en función de su nivel de identidad (más del 90% de aminoácidos comunes). Posteriormente, se compararon con diversas especies publicadas en la base de datos GeneBank del NCBI mediante BLASTn y BLASTp (Tabla 4.1: ver apartado 4.1.3.) y se encontraron homologías de todas ellas con retrotranscriptasas de *Ty3-gypsy*.

Utilizando la base de datos CDD (Conserved Domain Database) del NCBI (Marchler-Bauer *et al*., 2010) se han podido localizar diferentes motivos conservados en la secuencia de aminoácidos de la retrotranscriptasa. Algunas de las secuencias eran exactamente iguales así que, de las 55 obtenidas a partir de las colonias que presentaban inserto, se han incluido en el estudio las 22 secuencias que incluían alguna diferencia entre ellas (Figura 4.5).

Entre las secuencias del primer grupo (Figura 4.5), tres (GLc4-16, GLc3-05 y GLc4-08) presentan los dos cebadores utilizados en esta amplificación mientras que el resto de las secuencias presentan el cebador GypsyF en ambos extremos, observándose que una de las secuencias obtenidas (GLc4-08) presenta un codón de stop interno. En el alineamiento de todas ellas se observa la presencia de los motivos PRID y FXDD (en este caso FIDD) que se encuentran conservados en retrotransposones del tipo *Ty3-gypsy* de diversas especies vegetales (Rajput y Upadhyaya, 2009). Además, presentan 11 de los 13 sitios activos descritos en las retrotranscriptasas de *Ty3-gypsy* (Huang *et al*, 1998), los dos sitios que faltan se encontrarían en posiciones no secuenciadas. La secuencia GLc2-01 es más larga pero se ha eliminado el extremo 3´ por no presentar homología con el gen de la retrotranscriptasa ni con otras secuencias conocidas.

Figura 4.5: Secuencias de aminoácidos obtenidas con los cebadores diseñados a partir del gen de la retrotranscriptasa de retrotransposones *Ty3-gypsy*. Los alineamientos de dichas secuencias se comparan con otras secuencias publicadas en la base de datos del NCBI (AAT85845.1: *Vigna radiata*; ABD28305.1: *Medicago truncatula*; XP006574052: *Glycine max*; AAQ82033.1: *Pisum sativum*; XP006584201: *Glycine max*; XP004506193: *Cicer arietinum*). Con el fondo rojo se indican los motivos conservados de la proteína y, en negrita, los aminoácidos que forman el centro activo de la misma. La secuencia denominada *gypsy* es la secuencia consenso descrita por Xiong y Eickbush (1990).

El segundo grupo está formado por las secuencias obtenidas a partir de la banda 2 (CLc2) recuperada tras la amplificación con los cebadores para *Ty1-copia* y por secuencias correspondientes a la amplificación de *Ty3-gypsy*, estas últimas presentan el cebador GypsyF en ambos extremos, (Figura 4.5). La traducción de estas secuencias contiene los 13 sitios activos descritos para estas retrotranscriptasas. Se ha eliminado la parte final de la mayoría de las secuencias por no presentar homología con secuencias descritas en las bases de datos, de manera que podrían afectar a los estudios de filogenia y evolución. Además, se ha eliminado la parte inicial de las secuencias obtenidas con los cebadores Copia que se correspondería con la región delimitada por los cebadores CopiaF y GypsyF, estas secuencias podrían ser el resultado de una quimera formada por un retrotransposón *Ty3-gypsy* incluido dentro de otro de tipo *Ty1-copia*.

En las secuencias de aminoácidos se puede observar el motivo PRID aunque, en este caso, la arginina (R) ha sido sustituida por histidina (H) en todas las secuencias. Además, presenta el motivo YVDD en el que se encuentra el centro activo de la retrotranscriptasa y, a continuación, el motivo SKCEF algo menos conservado aunque sí que presenta el residuo de lisina (K) imprescindible en las polimerasas ARN-dependientes (Suoniemi *et al*., 1998). Este último motivo se corresponde con parte de la secuencia del cebador GypsyR (F/YAKLSKC) sin embargo, la presencia de cambios en la secuencia del retrotransposón, ha impedido la unión de dicho cebador por falta de complementariedad de las bases produciéndose una unión inespecífica del cebador GypsyF en una posición más alejada. (Figura 4.5).

Los últimos 35 aminoácidos codificados por las secuencias GLc4-12 y GLc4-15 se corresponden con la RNasaH en la que se ha podido localizar el primer residuo del centro activo (DEDD) descrito para este gen (Goedken y Marqusee, 2001; Nowotny *et al*., 2005) aunque no está conservado, en todas las secuencias se ha sustituido el ácido aspártico (D) por una leucina (L) (ver descripción RNasaH en la página 107).

El tercer grupo está formado por una única secuencia (GLc4-09), con el cebador GypsyF duplicado, también debido a mutaciones en la zona de unión del cebador

GypsyR. Presenta los 13 sitios conservados del centro activo de la retrotranscriptasa de *Ty3-gypsy* además de los motivos conservados PRID, FIDD y SKCEF. Este último presenta diversos cambios de secuencia pero incluye la lisina (K) característica de este tipo de proteínas. Además, presenta el fragmento inicial de la RNasaH de *Ty3-gypsy* en la que se encuentran los dos primeros residuos conservados de los cuatro descritos (DE). En el siguiente apartado se describe con más detalle la estructura de la RNasaH.

También, se puede destacar la presencia de codones de fin internos en las secuencias GLc4-15, Clc2-08, GLc4-05 así como deleciones y/o mutaciones en la secuencia de la retrotranscriptasa en 17 de las 28 secuencias descritas de manera que sólo el 32% de los retrotransposones amplificados podrían ser funcionales.

Estos cebadores degenerados han sido utilizados por distintos autores para buscar retrotransposones tipo Ty3-*gypsy* en diferentes plantas, incluyendo coníferas (Friedsen *et al*., 2001), judía mungo (*Vigna radiata*) (Dixit *et al*., 2006) o quinoa (*Chenopodium quinoa*) (Kolano *et al*., 2013), entre otras.

Muchas de las secuencias de retrotransposones encontradas en genomas vegetales presentan mutaciones similares a las detectadas en el presente estudio de manera que es necesario incluir huecos que supongan cambios en el marco de lectura para poder obtener las secuencias aminoacídicas de los retrotransposones publicadas en las bases de datos (Flavell *et al*., 1992; Hirochika y Hirochika, 1993; Gribbon *et al.*, 1999).

Diferentes estudios sugieren que la gran mayoría de los retrotransposones de plantas se encuentran inactivos tanto por mutaciones como por silenciamiento epigenético (Hirochika *et al*., 2000; Devos *et al*., 2002; Wright y Voytas, 2002). En arroz, por ejemplo, sólo tres entre veinte familias de retrotransposones identificadas pudieron expresarse en cultivos celulares (Hirochika *et al*., 1996B).

Algunas de las secuencias amplificadas se extienden más allá de la región de la retrotranscriptasa esperada (Figura 4.5). Estas secuencias, pertenecientes a los grupos 2

y 3, se muestran en la Figura 4.6. En el grupo 2 se mantiene una alta identidad entre las diferentes secuencias y se puede observar que los fragmentos obtenidos con los cebadores de *copia* no llegan hasta estas posiciones. Las secuencias GLc3-04, GLc1-04, GLc4-15, GLc4-05 del grupo 2 y GLc4-09 (Grupo 3) incluyen el principio del gen de la RNasaH y presentan el primer aminoácido conservado del dominio (DEDD) aunque en el grupo 2 se ha sustituido por una leucina (L).

Los resultados obtenidos con los cebadores para *gypsy* demuestran que éstos son menos eficaces que los de *copia*, esto coincide con los resultados descritos por otros autores al utilizar estos mismos cebadores con diferentes especies vegetales. La variedad de retrotransposones identificados del tipo *Ty3-gypsy* es mucho menor que el obtenido con los cebadores para *Ty1-copia* lo que demostraría que la retrotranscriptasa está menos conservada en los retrotransposones *gypsy*, de manera que dificulta el diseño de cebadores degenerados y su uso generalizado en especies diferentes (Su y Brown, 1997; Friedsen *et al*, 2001).

Los fragmentos de retrotransposones obtenidos a partir del genoma de lenteja presentan los motivos estructurales conservados esperados en los retrotransposones LTR típicos y presentan elevadas similitudes con secuencias de dichos elementos descritos en otras especies e incluidos en las bases de datos tanto en el caso de los retrotransposones tipo *Ty1-copia* como en los *Ty3-gypsy*. En las secuencias tipo *gypsy* se observa una menor heterogeneidad en comparación con las secuencias de tipo *copia* ya que únicamente se han detectado tres grupos frente a los siete de *copia*, lo mismo que ocurre con *Vigna radiata* en la que se han descrito tres grupos de *gypsy* frente a seis de *copia* utilizando los mismos cebadores (Dixit *et al*., 2006). Esta menor heterogeneidad también ha sido descrita por otros autores utilizando cebadores diferentes (Muthukumar y Bennetzen, 2004).

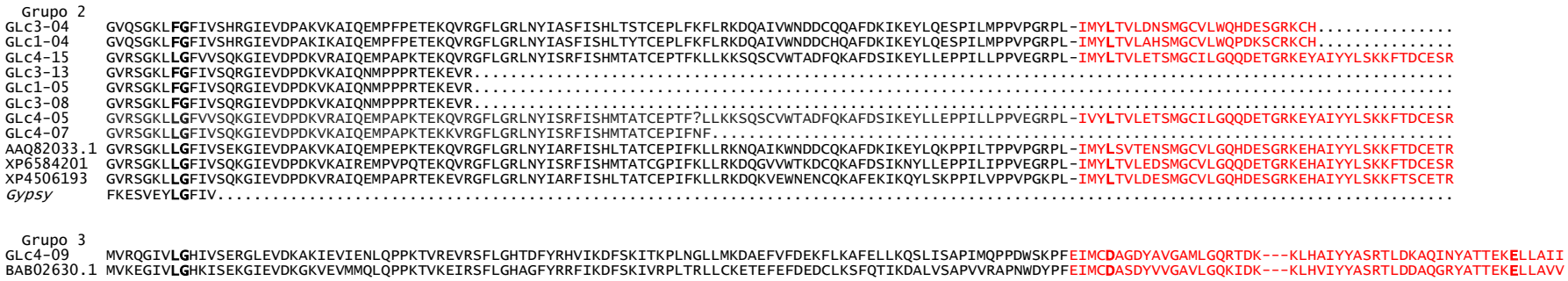


Figura 4.6: Ampliación de las secuencias de aminoácidos obtenidas con los cebadores diseñados para la retrotrancriptasa de retrotransposones del tipo *Ty3-gypsy*. En negrita se indican los aminoácidos que forman el centro activo de la retrotranscriptasa y en rojo el fragmento correspondiente a la RNasaH.

Al igual que con las secuencias obtenidas con los cebadores *copia*, se ha realizado un estudio filogenético con las secuencias nucleotídicas correspondientes a los cebadores tipo *Ty3-gypsy*. En este caso sólo se han analizado las secuencias que se corresponden con fragmentos del gen de la retrotranscriptasa.

El árbol se ha obtenido siguiendo el mismo procedimiento descrito para *copia*, de manera que se ha seleccionado como el modelo más adecuado, mediante un análisis de máxima verosimilitud utilizando el programa Mega5, el de Tamura con tres parámetros con una distribución gamma de cinco categorías, modelo que coincide con el utilizado con las secuencias de *Ty1-copia* (Tamura, 1992). Con este modelo se obtuvo una matriz de distancias a partir de la cual se diseñó un árbol filogenético basado en el algoritmo Neighbor Joining (Saitou y Nei, 1987). Para comprobar la fiabilidad de los nodos del árbol se ha utilizado un análisis "bootstrap" realizando 1000 remuestreos (Figura 4.7).

Se pueden observar claramente los tres grupos descritos anteriormente en función de su identidad, aunque uno de ellos presenta varios subgrupos o ramificaciones, indicados en el árbol mediante diferentes recuadros. Cada una de estas ramificaciones constituye un clado monofilético que presenta, en todos los casos, valores de "bootstrap" mayores del 97%. Todos estos clados hermanos se unen entre ellos con una fiabilidad del 80% y se puede observar que todas las secuencias obtenidas con los cebadores para *copia* pertenecen al mismo clado hermano, excepto CLc2-12. Este grupo, definido por el color azul, parece ser propio de leguminosas ya que sólo se han encontrado homologías con retrotransposones de especies de dicha familia (Figura 4.7).

El grupo 3, marcado en la figura con el color verde, está formado por una única secuencia de lenteja que se une con una fiabilidad del 100% en el "bootstrap" con una secuencia de *Arabidopsis thaliana*.

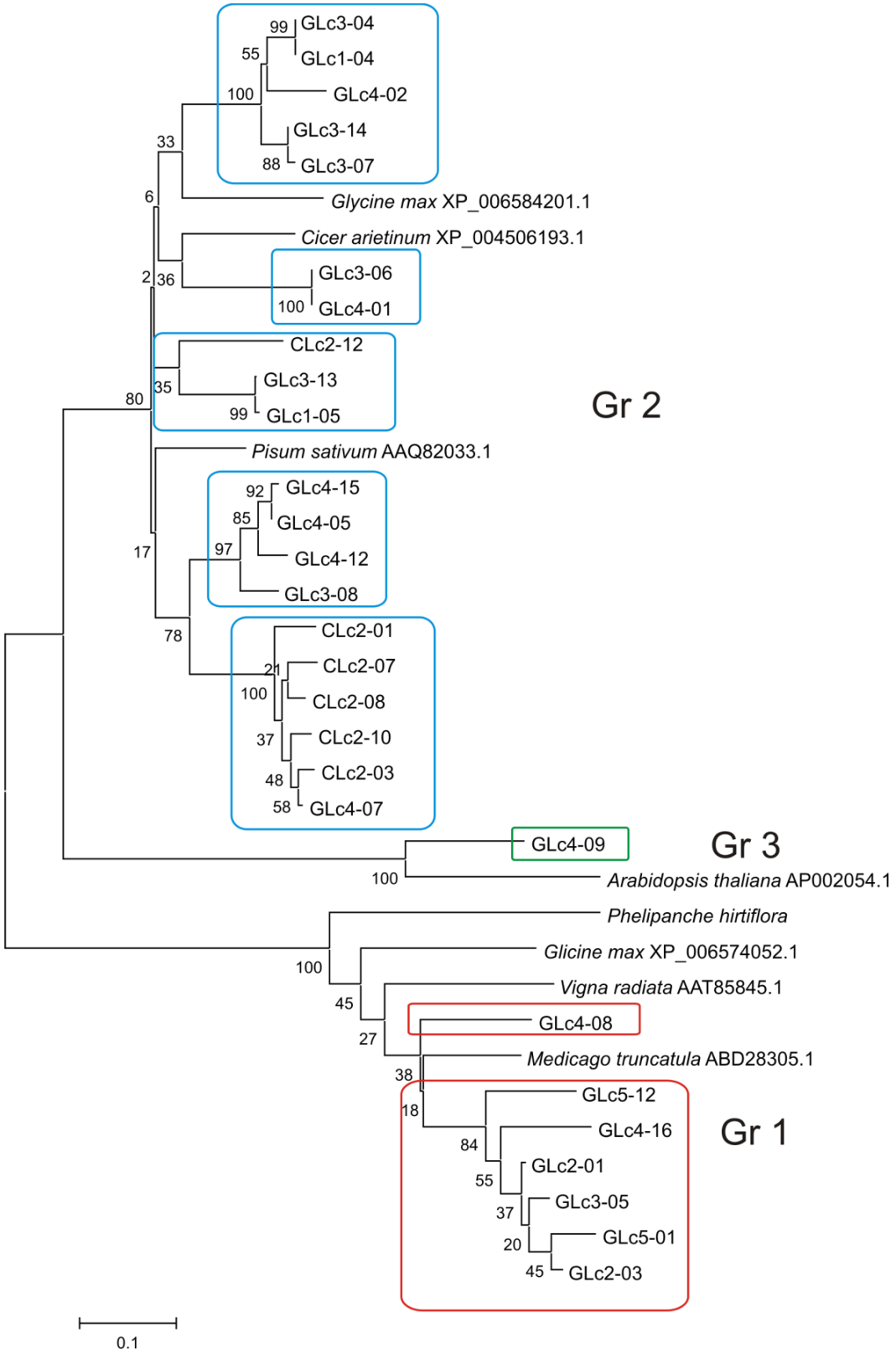


Figura 4.7: Árbol filogenético de retrotransposones tipo *gypsy*. La divergencia entre las secuencias está definida por la longitud de las ramas y expresado en unidades de distancia genética. En cada nodo se indican los valores de "bootstrap". Se han recuadrado con diferentes colores los distintos grupos de secuencias obtenidos.

En cuanto al último grupo 1, rojo, se puede observar que todas las secuencias de lenteja analizadas, excepto GLc4-08, se encuentran unidas formando un grupo monofilético con un valor de "bootstrap" del 84% y, a su vez, se unen con otras secuencias de lenteja u otras plantas formando un grupo mayor con una fiabilidad del 100%. En este grupo, las secuencias de lenteja presentan homología con retrotransposones de diferentes leguminosas y con *Phelipanche hirtiflora*. El conjunto de las leguminosas formaría un grupo monofilético pero con una fiabilidad "bootstrap" muy baja (45%).

A partir de los cebadores degenerados diseñados para *gypsy* y *copia*, se han amplificado miembros de distintos tipos de retrotransposones y sus secuencias presentan importantes similitudes entre distintas especies de plantas, esto sugeriría que los genes de la retrotranscriptasa de todos estos retrotransposones están relacionados por un origen monofilético común (Xiong y Eiskbush, 1990; Flavell *et al*., 1992; Friedsen, 2001).

Los estudios realizados por Xiong y Eiskbush (1990) sugieren, además, un origen común de todos los elementos que presentan esta polimerasa, desde retrovirus hasta retrotransposones y mitocondrias, debido a las similitudes que presentan en sus secuencias de aminoácidos.

4.1.2. Cebadores derivados de un retrotransposón de lenteja

A partir de una secuencia de *Lens culinaris* correspondiente a un retrotransposón de tipo desconocido secuenciado en el laboratorio por Llorente (2001), se han diseñado 3 cebadores en un sentido (TnanaF1, TnanaF2 y TnanaF3) y uno en sentido contrario (TnanaR) con los que se han obtenido secuencias correspondientes a diferentes zonas de distintos retrotransposones tanto del tipo *Ty3-gypsy* como del tipo *Ty1-copia*.

Estos cebadores han sido diseñados en el propio laboratorio por lo que no se habían utilizado hasta el momento, esto impedía conocer a priori el número y el tamaño de las bandas que se van a amplificar.

Los amplicones obtenidos con estos cebadores a partir del genoma de *Lens* mostraron el patrón de bandas característico que se muestra en la Figura 4.8.

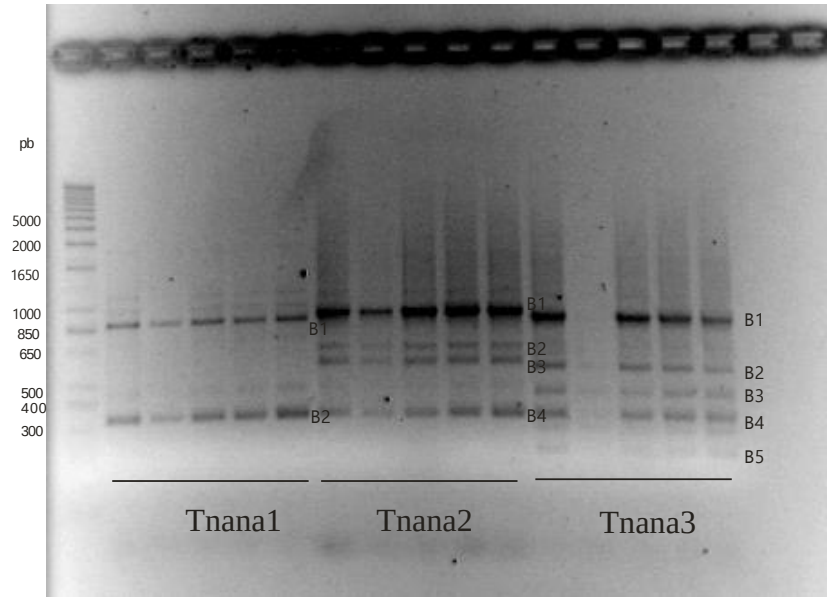


Figura 4.8: Bandas obtenidas con los cebadores diseñados para un retrotransposón de tipo desconocido de *Lens culinaris.* Se indican las bandas aisladas para su clonación y secuenciación. A la izquierda se indica el peso molecular de cada una de las bandas del marcador de peso molecular utilizado como control.

Para nombrar estas secuencias se ha utilizado la denominación TLc añadiendo el número del cebador correspondiente, el número de la banda y, precedido por un guion, el número de la colonia.

Tras el aislamiento de las bandas se realizó su transformación y clonación. A continuación se seleccionaron 16 colonias de cada una de las bandas y se comprobó la presencia del inserto mediante minipreparaciones de colonias recombinantes. En alguno de los casos no se obtuvo ninguna colonia recombinante o muy pocas de manera que de las 16 colonias estudiadas para cada una de las bandas se enviaron a secuenciar las siguientes colonias:

Cebador TnanaF1: 4 colonias de la banda 1 y 8 de la banda 2.

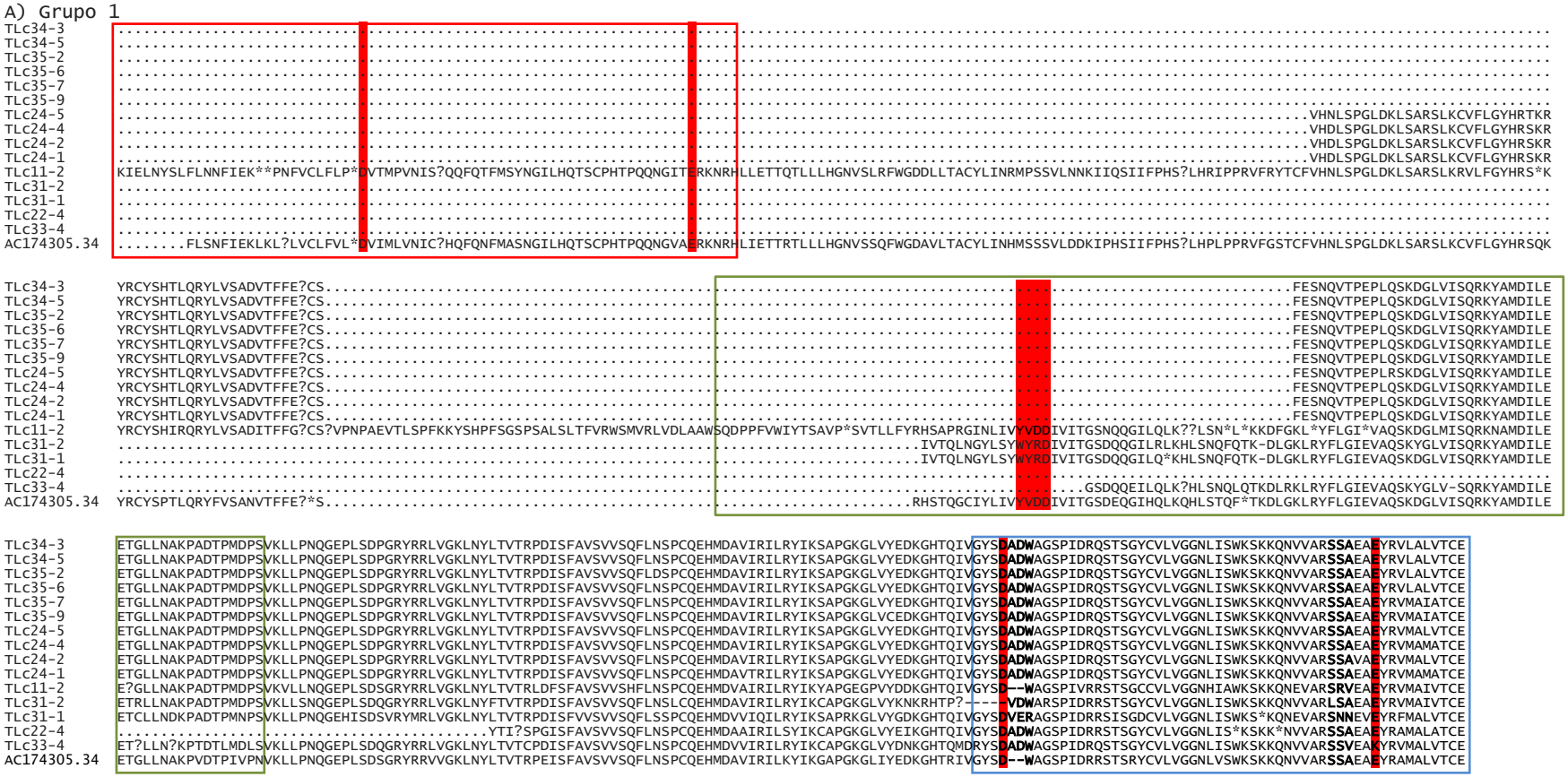
Cebador TnanaF2: 6 colonias de las bandas 1 y 2, 9 de la banda 3 y 5 de la 4.

Cebador TnanaF3: no se consiguieron colonias recombinantes de la banda 2 y del resto de las bandas se secuenciaron, 6 colonias de la banda 1, 4 de la banda 3, 5 de la banda 4, 15 de la banda 5 y 8 de la banda más pequeña.

Una vez obtenidas las secuencias, se analizaron mediante BLAST (BLASTn y BLASTp) para compararlas con secuencias de nucleótidos y aminoácidos publicadas en la bases de datos del NCBI (tabla 4.1: ver apartado 4.1.3.). Además, la base de datos CDD (Conserved Domain Database) del NCBI (Marchler-Bauer *et al*., 2010) ha permitido determinar la presencia de diferentes motivos conservados en las diferentes proteínas localizadas. En la Figura 4.9 se muestran, únicamente, las secuencias de aminoácidos que presentan diferencias entre sí divididas en tres grupos en función de sus identidades.

El primer grupo está formado por secuencias obtenidas a partir de los tres cebadores (TnanaF1, TnanaF2 y TnanaF3) aunque en tres de ellas (TLc33-4, TLc31-2 y TLc31-1) no se observa el cebador F ya que el electroferograma obtenido tras la secuenciación contenía interferencias en la zona inicial, todas las secuencias del grupo presentan el cebador TnanaR en el extremo contrario. Las secuencias de este grupo muestran una identidad superior al 90% con un retrotransposón *Ty1-copia*, más concretamente con una región interna que incluye la integrasa, la retrotranscriptasa y la RNasaH de manera que se ha podido secuenciar el gen *pol* prácticamente en su totalidad.

Se puede observar una secuencia mucho más larga que el resto (TLc11-2), obtenida con el cebador TnanaF1, mientras que el resto de las secuencias sólo presentan parte de ella, en alguno de los casos se trata de secuencias que han perdido la parte interna del retrotransposón lo que implicaría la pérdida de capacidad de retrotransposición autónoma del mismo. También, cabe destacar la secuencia TLc22-4 que presenta el cebador TnanaF2 en una situación más adelantada que en el resto de las secuencias amplificadas con dicho cebador.


A) Grupo 1
TLc34-3
TLc34-5
TLc35-2
TLc35-6
TLc35-7
TLc35-9
TLc24-5
TLc24-4
TLc24-2
TLc24-1
TLc11-2
TLc31-2
TLc31-1
TLc22-4
TLc33-4
AC174305.34

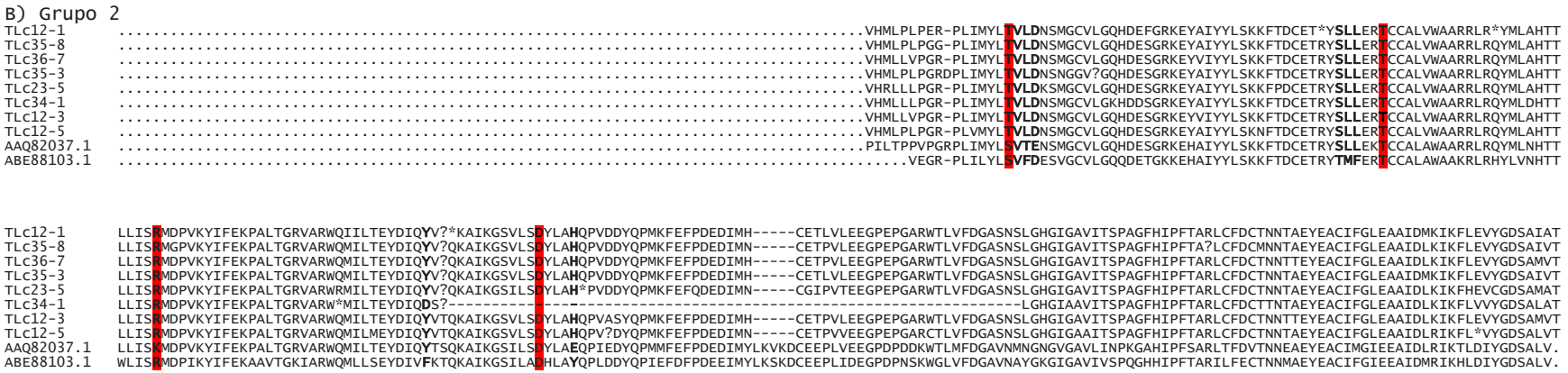


Figura 4.9: Secuencias obtenidas con los cebadores diseñados a partir de un retrotransposón de tipo desconocido de *Lens culinaris*. A) Secuencias de aminoácidos que presentan homologías con la integrasa, la retrotranscriptasa y la RNAsaH identificadas con un retrotransposón tipo *Ty1-copia* por comparación con la secuencia AC174305.34 de *Medicago truncatula* publicada en la base de datos del NCBI. Los recuadros indican cada uno de los fragmentos conservados; rojo: motivo CCD de la integrasa; verde: retrotranscriptasa; azul: RNAsaH. Con el fondo rojo se representan los residuos de los centros activos de las proteínas y en negrita los residuos de unión ARN-ADN conservados. B) secuencias de aminoácidos que presentan homología con la RNasaH de un retrotransposón del tipo *Ty3-gypsy* comparadas con otras secuencias publicadas en la base de datos del NCBI (AAQ82037.1: *Pisum sativum*; ABE88103.1: *Medicago truncatula*). Las secuencias están alineadas con la región de la RNasaH del grupo 1.

En todas las secuencias aparece un salto en el marco de lectura antes de comenzar el fragmento codificante para la retrotranscriptasa. Además, aparecen diferentes saltos en el marco de lectura y codones de fin internos en las secuencias TLc11-2, TLc31-2, TLc21-1, TLc22-4 y TLc33-4.

La primera parte de las secuencias se corresponde con la integrasa. Esta proteína es la que permite la integración del ADN del retrotransposón en el genoma del huésped. Consta de tres dominios conservados: el dominio N-terminal, un dominio central en el que se encuentra la actividad catalítica de la enzima (CCD, Catalytic Core Domain) caracterizado por la presencia del motivo DD(35)E y el dominio C-terminal. (Jaskolski *et al*, 2009). La secuencia TLc11-2 incluye dos de los residuos del motivo DD(35)E, el segundo ácido aspártico y el ácido glutámico. Los dominios N-terminal y C-terminal no están conservados por lo que sus secuencias aminoacídicas no se encuentran en las bases de datos CDD. Esto hace que no se pueda confirmar que el fragmento que se encuentra a continuación del CCD se corresponda con el dominio C-terminal de la proteína.

En su parte intermedia, las secuencias presentan similaridad con una retrotranscriptasa. Se puede observar el motivo conservado YVDD característico de las transcriptasas inversas de los retrotransposones *Ty1-copia*.

El final de las secuencias se corresponde con la RNasaH. Se trata de una RNasa de tipo1 de *Ty1-copia*. Este tipo de proteínas se caracterizan por la presencia de un centro activo formado por cuatro residuos: tres residuos de ácido aspártico (D) y uno de ácido glutámico (E). Estos cuatro sitios forman el denominado motivo DEDD (Goedken y Marqusee, 2001; Nowotny *et al*., 2005). Es este caso aparecen los dos primeros residuos del motivo DDED, aunque el segundo ácido aspártico (D) ha sido sustituido por un ácido glutámico (E), y 8 de los 11 sitios de unión del híbrido ARN-ADN mediante iones metálicos, necesarios para el correcto funcionamiento de la enzima. La ausencia de alguno de estos sitios de unión impediría la formación de la estructura secundaria necesaria para la hidrólisis del ARN (Nowotny *et al*., 2007). De estos sitios

cabe destacar la presencia de una histidina en el extremo C-terminal que se encuentra conservada en RNasasH de bacterias, virus y humanos (Tisdale *et al*., 1991; Oda *et al*., 1993).

El segundo grupo representado está formado por secuencias que presentan el cebador TnanaR en ambos extremos. Estas secuencias se corresponden con una RNasaH tipo1 de *Ty3-gypsy* de características similares a las descritas para la RNasaH tipo1 de *Ty1-copia*. Se trata de un retrotransposón tipo *Ogre* descrito en guisante por Neumann *et al*. (2003) que se caracteriza por su gran tamaño (22 Kb) y por la presencia de un marco de lectura abierto adicional que codifica para una proteína con función desconocida. Constituye al menos el 5% del genoma de *Pisum sativum* y se transcribe tanto en hojas como en raíces y flores (Neumann *et al*. 2003).

Todas las secuencias de este segundo grupo presentan el motivo correspondiente al centro activo de la RNasaH muy modificado (TTRD en lugar de DEDD), sin embargo, sólo el último aspártico (D) está conservado en todos los amplificados. El primer ácido aspártico (D) y el ácido glutámico (E), han sido sustituido por treonina (T) y el segundo ácido aspártico por arginina (R). Además, las secuencias presentan los 11 sitios de unión del híbrido ARN-ADN mediante iones metálicos, necesarios para el correcto funcionamiento de la enzima.

Las secuencias del grupo 2, presentan similitudes con secuencias definidas en diferentes especies como *Pisum sativum* o *Medicago truncatula* lo que indicaría la conservación de la secuencia del gen *pol* (codificante para proteasa, integrasa, transcriptasa inversa y ribonucleasaH) en las leguminosas estudiadas.

Se ha realizado un estudio filogenético con estas secuencias, para ello se ha utilizado el programa Mega5 para elegir el mejor modelo de evolución mediante métodos de máxima verosimilitud. Como en los caso anteriores, éste ha sido el modelo de Tamura con tres parámetros y distribución gamma (Tamura, 1992). Este modelo permitió obtener una matriz de distancias a partir de la cual se diseñó un árbol filogenético basado en el algoritmo Neighbor Joining (Saitou y Nei, 1987) y para

comprobar la fiabilidad de los nodos del árbol se utilizó un análisis "bootstrap" realizando 1000 remuestreos. De esta manera se han obtenido dos árboles correspondientes a los grupos 1 y 2 (Figura 4.10). Se han analizado los dos grupos de secuencias por separado debido a las diferencias existentes de la región amplificada entre uno y otro. El estudio conjunto reduciría la secuencia común a menos de 250 nucleótidos.

En el árbol correspondiente al grupo 1 se puede observar como casi todas las secuencias muestran una alta similitud exceptuando algunas de ellas como TLc22-4, TLc11-2, TLc31-1, TLc31-2 y TLc33-4. Todas parecen más próximas entre sí que con *Medicago*. El análisis "bootstrap" parece indicar la presencia de una politomía en el origen de estas secuencias y una amplificación reciente de la secuencia que originó el grupo mayoritario.

En cuanto al segundo árbol, se puede observar un grupo de secuencias de alta similitud (TLc12-1, TLc12-3, TLc35-3 y TLc23-4) y un origen monofilético de las secuencias de lenteja posterior a su divergencia con la secuencia de *Pisum sativum*.

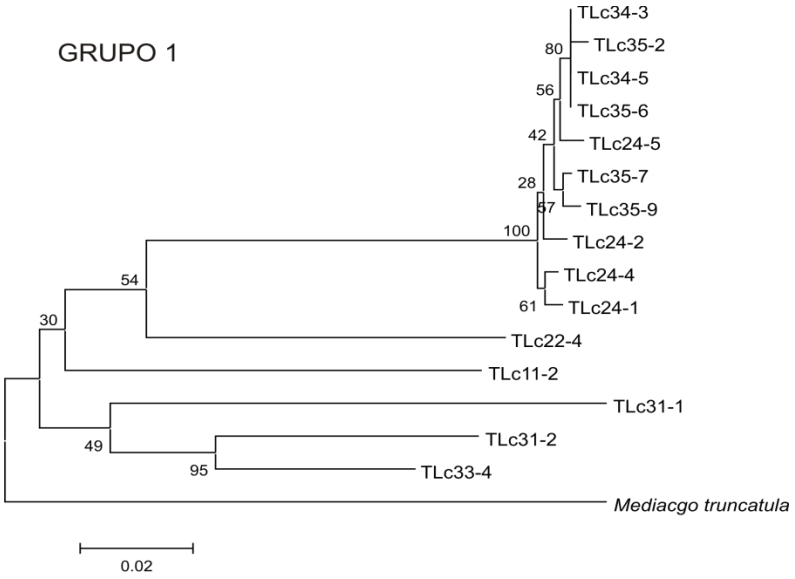


Figura 4.10A: Árbol filogenético de las secuencias obtenidas a partir de una secuencia de un retrotransposón de tipo desconocido de *Lens culinaris* correspondientes a *Ty1-copia*. La divergencia entre las secuencias está definida por la longitud de las ramas y expresado en unidades de distancia genética. En cada nodo se indican los valores de "bootstrap".

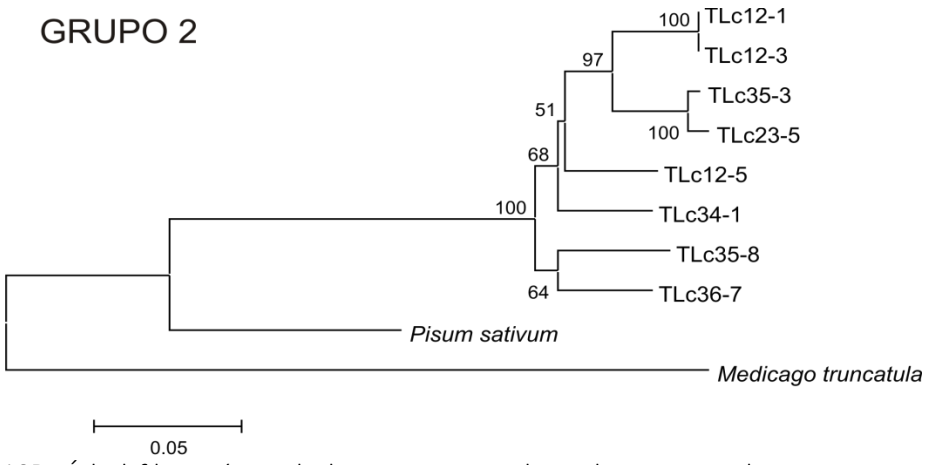


Figura 4.10B: Árbol filogenético de las secuencias obtenidas a partir de una secuencia de un retrotransposón de tipo desconocido de *Lens culinaris* correspondientes a *Ty3-gypsy*. La divergencia entre las secuencias está definida por la longitud de las ramas y expresado en unidades de distancia genética. En cada nodo se indican los valores de "bootstrap".

4.1.3. Retrotransposones en plantas. Corolario.

La tabla 4.1 muestra la variedad de especies de diferentes taxones con los que presentan similitudes las secuencias amplificadas con los cebadores descritos. La gran mayoría de las especies pertenecen a las leguminosas pero aparecen especies de otros grupos como *Setaria faberi*, *Corchorus olitorius*, *Oryza sativa* o *Silene gallica*.

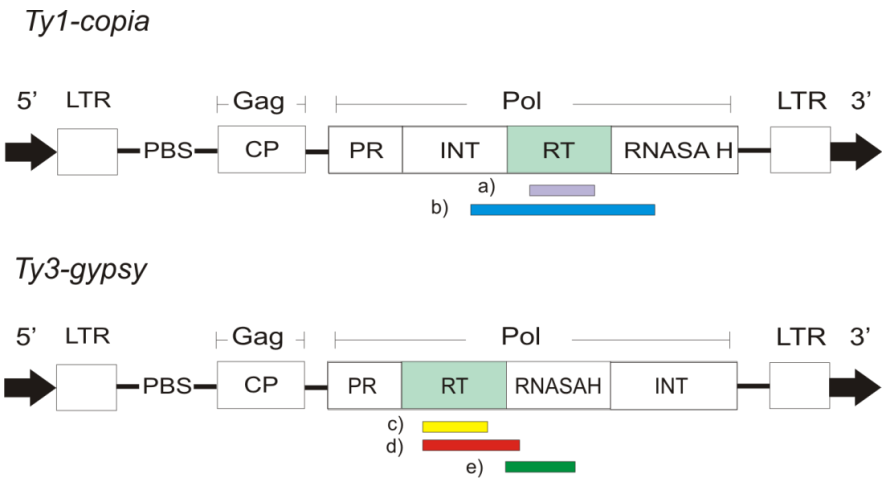


Figura 4.11: Localización de las diferentes secuencias obtenidas en el tipo de retrotransposón correspondiente. a) secuencias obtenidas con los cebadores diseñados a partir de regiones conservadas de la retrotranscriptasa de *Ty1-copia*, b) primer grupo de las secuencias obtenidas con los cebadores diseñados a partir de un retrotransposón de tipo desconocido de *Lens*, c) grupo 1 de las secuencias obtenidas con los cebadores diseñados a partir de regiones conservadas de la retrotranscriptasa de *Ty3-gypsy*, d) grupos 2 y 3 de las secuencias obtenidas con los cebadores diseñados a partir de regiones conservadas de la retrotranscriptasa de *Ty3-gypsy*, e) segundo y tercer grupo de secuencias obtenidas con los cebadores diseñados a partir de un retrotransposón de tipo desconocido de *Lens*.

Tabla 4.1: Relación de las secuencias localizadas en la base de datos del NCBI que presentan un nivel de identidad mínimo del 90% con los amplificados de los cebadores Copia, Gypsy y Tnana. En la tabla se indican los identificadores nucleotídicos y aminoacídicos (cuando está definido), así como las principales características de cada secuencia y el grupo de retrotransposones al que pertenecen.

Número de accesión	Identificador proteína	Especie	Características	Tipo de retrotransposón
AF227036.1	AAL36473.1	*Setaria faberi*	Transcriptasa reversa	*Ty1-Copia*
GU734724.1	ADE44145.1	*Corchorus olitorius*	Transcriptasa reversa	*Ty1-Copia*
Z75518.1	CAA99754.1	*Oryza sativa*	Transcriptasa reversa. Retrotransposón Rtr23	*Ty1-Copia*
AF155230.1	AAD42969.1	*Orobanche ramosa*	Transcriptasa reversa	*Ty1-Copia*
EU162608.1	ABW81017.1	*Arabidopsis lyrata*	Poliproteína gag-pol	*Ty1-Copia*
DQ376329.1	ABD18991.1	*Orobanche cernua*	Transcriptasa reversa	*Ty1-Copia*
AJ535751.1	CAD59770.1	*Cicer arietinum*	Transcriptasa reversa	*Ty1-Copia*
AY684660.1	AAT90460.1	*Vigna radiata*	Transcriptasa reversa	*Ty1-Copia*
HQ632856.1	AER13172.1	*Phaseolus vulgaris*	Poliproteína gag-pol	*Ty1-Copia*
AM451833.2	CAN83010.1	*Vitis vinifera*	Proteína hipotética VITISV_010102	*Ty1-Copia*
AM426055.2	CAN66710.1	*Vitis vinifera*	Proteína hipotética VITISV_006397	*Ty1-Copia*
AY684634.1	AAT90441.1	*Vigna radiata*	Transcriptasa reversa	*Ty1-Copia*
AY684684.1	AAT90480.1	*Vigna radiata*	Transcriptasa reversa	*Ty1-Copia*
AB065058.1	BAB83572.1	*Silene latifolia*	Transcriptasa reversa	*Ty1-Copia*
KF303287.1	AGW47867.1	*Phaseolus vulgaris*	Poliproteína	*Ty1-Copia*
JX035977.1	AFK88549.1	*Arabidopsis halleri*	Transcriptasa reversa	*Ty1-Copia*
AL138659.1	CAB75469.1	*Arabidopsis thaliana*	Transcriptasa reversa	*Ty1-Copia*
AM467485.2	CAN79192.1	*Vitis vinifera*	Proteína hipotética VITISV_000234	*Ty1-Copia*
AY684673.1	AAT90470.1	*Vigna radiata*	Transcriptasa reversa	*Ty1-Copia*
AB065032.1	BAB83551.1	*Silene gallica*	Transcriptasa reversa	*Ty1-Copia*
AY683021.1	AAT85845.1	*Vigna radiata*	Transcriptasa reversa	*Ty3-gypsy*
AC147774.3	ABD28305.1	*Medicago truncatula*	Transcriptasa reversa	*Ty3-gypsy*
XM_006573989.1	XP_006574052.1	*Glycine max*	DNA polimerasa	*Ty3-gypsy*
AY299395.1	AAQ82033.1	*Pisum sativum*	Poliproteína gag-pol	*Ty3-gypsy*
XM_006584138.1	XP_006584201.1	*Glicine max*	Proteína no caracterizada LOC100789592	*Ty3-gypsy*
XM_004506136.1	XP_004506193.1	*Cicer arietinum*	Proteína no caracterizada LOC1101494924	*Ty3-gypsy*
AP002054.1	BAB02630	*Arabidopsis thaliana*	Poliproteína pol	*Ty3-gypsy*
AY299395.1	AAQ82037.1	*Pisum sativum*	Poliproteína gag-pol	*Ty3-gypsy*
AC147431.19	ABE88103.1	*Medicago truncatula*	Ribonucleasa H	*Ty3-gypsy*
AC174305.34		*Medicago truncatula*	Clon mth2-108j24	*Ty1-copia*
XM_003537652	XP_003537700	*Glycine max*	Tioredoxina reductasa	

La diversidad de familias de retrotransposones encontrada, así como las similitudes que presentan con elementos de diversas especies vegetales tanto relacionadas con las leguminosos como pertenecientes a otros taxones, indican la abundancia en el genoma y su presencia y conservación a lo largo de todo el reino vegetal tanto de la familia *Ty1-copia* como de *Ty3-gypsy* (Tabla 4.1).

La Figura 4.11 muestra un esquema con la localización de las diferentes secuencias obtenidas a partir de zonas conservadas de retrotransposones *Ty1-copia* y *Ty3-gypsy*.

4.2. Identificación de repeticiones largas terminales o LTRs

4.2.1. Identificación de LTRs a partir de las secuencias obtenidas con los cebadores diseñados a partir de la retrotranscriptasa de *copia* y *gypsy*.

Una vez secuenciadas todas las bandas amplificadas con los cebadores específicos Gypsy, Copia y Tnana, se diseñaron nuevos cebadores complementarios a los extremos de las secuencias obtenidas en cada caso, para poder avanzar por ellas hasta alcanzar el extremo del retrotransposón donde se encuentran las LTRs ya que son estas las regiones necesarias para el diseño de nuevos marcadores moleculares.

Para poder avanzar a lo largo del retrotransposón, es necesario realizar una digestión del genoma mediante endonucleasas de restricción, de manera que se pueda ligar un adaptador al extremo digerido y realizar una amplificación utilizando el cebador específico y otro complementario al adaptador.

Aunque se diseñaron cebadores a partir de todas las secuencias obtenidas, solamente se probaron seis, todos ellos correspondientes a secuencias obtenidas con los cebadores de *Ty1-copia*. En la Tabla 4.2 se indica el nombre de cada una de ellas y su secuencia de procedencia así como el grupo correspondiente de los descritos para *Ty1-copia* y *Ty3-gypsy* en el apartado anterior.

A continuación, se llevó a cabo el protocolo de digestión, ligación y amplificación. La amplificación se realizó en dos pasos, un primer paso en el que sólo se utilizó el cebador específico de cada una de las secuencias y una segunda amplificación con el cebador específico y el correspondiente al adaptador de la enzima de restricción. De esta manera se intentó evitar una amplificación masiva de fragmentos que presentaran el cebador del adaptador en ambos extremos.

Tabla 4.2: Cebadores diseñados a partir de algunas de las secuencias obtenidas con los cebadores diseñados para *Ty1-copia*.

Cebador	Origen	Grupo	Secuencia	T_m
NLc-1	CLc3-10	Gr 1 *copia*	5' CTACAAGAAGGTTAGTGG 3'	46,2°C
NLc-2	CLc3-01	Gr 2 *copia*	5' GTGTAAGCTAGTTGAGTC 3'	46,5°C
NLc-3	CLc2-12	Gr 2 *gypsy*	5' CTTATCAGAGAGCTATGG 3'	45,5°C
NLc-4	CLc3-02	Gr 3 *copia*	5' GAGTCTGTTATGTGTGAG 3'	46,5°C
NLc-5	CLc3-19	Gr 6 *copia*	5' CTCTACGTCGATGATATG 3'	45,9°C
NLc-6	CLc2-10	Gr 2 *gypsy*	5' CCTTCATAACACCTTGGG 3'	49,8°C

Tras la realización de las amplificaciones se observaron nuevos patrones de bandas en función del cebador utilizado en cada caso. (Figura 4.12).

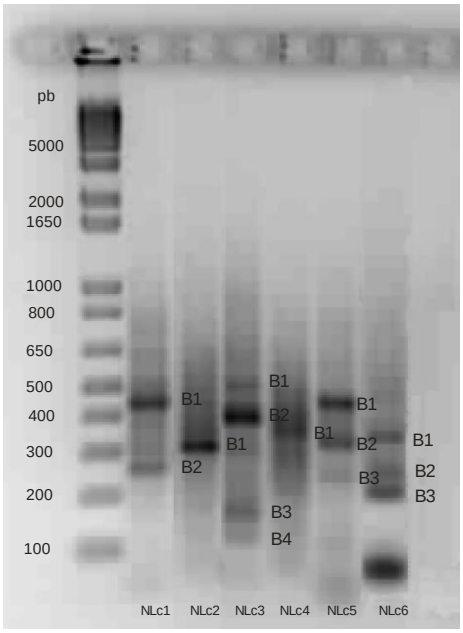


Figura 4.12: Bandas obtenidas con los cebadores NLc (NLc1, NLc2, NLc3, NLc4, NLc5 y NLc6). Se indican las bandas aisladas para su clonación y secuenciación. A la izquierda se indica el tamaño de cada una de las bandas del marcador de peso molecular utilizado como control.

El siguiente paso consistió en el aislamiento de todas las bandas amplificadas, su ligamiento en un vector de clonación y la posterior transformación y clonación de células competentes. De todas las colonias recombinantes obtenidas, se seleccionaron 16 de cada una de las bandas, se comprobó la presencia del inserto mediante minipreparaciones de colonias recombinantes. Con las bandas 1 y 4 de NLc3 y la banda

2 de NLc5 no se obtuvieron colonias con insertos de tamaño apreciable aunque sí que aparecían colonias blancas en las placas, del resto de las bandas se secuenciaron cinco colonias.

Las secuencias se alinearon con el algoritmo Clustal en el programa MEGA5 de forma que se pudo comprobar que todas ellas eran iguales en función del cebador utilizado y la banda de la que procedían, de manera que sólo se incluirá una secuencia de cada grupo en la Figura 4.13.

A continuación, se compararon dichas secuencias con otras incluidas en la base de datos GeneBank del NCBI mediante el uso de los algoritmos BLASTn y BLASTp de manera que se encontraron niveles de identidad mayores del 90% con fragmentos de retrotranscriptasas de otras especies vegetales (Tabla 4.3).

Tabla 4.3: Relación de las secuencias localizadas en la base de datos del NCBI, que presentan un nivel de identidad mayor del 90% con los amplificados de los cebadores NLc (NLc1, NLc3 y NLc4). En la tabla se indican las referencias nucleotídicas y aminoacídicas, así como las principales características de cada secuencia y el grupo de retrotransposones al que pertenecen.

Número de accesión	Identificador proteína	Especie	Características	Tipo de retrotransposón
AF077408.1	AAC26250.1	*Arabidopsis thaliana*	Transcriptasa reversa	*Ty1-Copia*
AC137921.2	AAR06328.1	*Oryza sativa*	Transcriptasa reversa	*Ty1-Copia*
AY299398.1	AAQ82037.1	*Pisum sativum*	Poliproteína gag-pol	*Ty3*-Gypsy
XM_004490430.1	XP_004490487.1	*Cicer aritinum*	Proteína no caracterizada LOC10149483	*Ty3*-Gypsy
HQ622856.1	AER13172.1	*Phaseolus vulgaris*	Poliproteína gag-pol	*Ty1-Copia*
AM447752.2	CAN65281.1	*Vitis vinifera*	Proteína hipotética VITISV_035564	*Ty1-Copia*

Para poder avanzar a lo largo de las secuencias hasta alcanzar las LTRs era necesario diseñar cebadores que estuviesen lo más cerca posible del extremo de la secuencia. Esto dificultó la obtención de buenos cebadores ya que no en todos los casos estos extremos presentan las características idóneas para su diseño, ya fuese por una proporción inadecuada de nucleótidos G y C o por la formación de estructuras secundarias. De esta manera no siempre se amplificaban secuencias únicas ni se obtenía la secuencia original, únicamente se ha conseguido amplificar la secuencia deseada con la banda 1 de NLc1 (NLc1B1), la banda 2 de NLc3 (NLc3B2) y la banda 1 de NLc4

(NLc4B1) (Figura 4.13). Con el resto de cebadores se han obtenido secuencias que no se corresponden con aquellas de las que proceden.

Con los cebadores NLc1 y NLc4 se han obtenido secuencias correspondientes a retrotransposones del tipo *Ty1-copia* mientras que, con el cebador Nlc3, el retrotransposón localizado pertenece a la familia *Ty3-gypsy*.

Todas estas secuencias excepto NLc4B1, mostradas en la Figura 4.13 como secuencias de aminoácidos, presentan el motivo conservado YVDD en el que se encuentra el centro activo de la retrotranscriptasa (Suoniemi *et al*., 1998) y, en la secuencia Nlc3B2, dos de los 13 residuos del centro activo descritos por Huang *et al* (1998).

Como se puede observar en la Figura 4.13, las secuencias NLc1B1 y NLc4B1 presentan codones de parada antes de la terminación de la proteína lo que las convierte en proteínas no funcionales.

Con esta técnica se han conseguido secuencias más largas que las obtenidas en las primeras amplificaciones, sin embargo no se ha conseguido alcanzar las regiones de las LTRs debido al pequeño tamaño de las bandas amplificadas. Se podrían haber obtenido secuencias más largas probando diferentes enzimas de restricción que presentasen dianas menos comunes en el genoma, sin embargo existen otros problemas como son la longitud de dichos elementos, en algunos caso de miles de pares de bases, y a interferencias en la unión de los cebadores debidos a las zonas de inserción de los RTs que convierten este proceso en un trabajo largo y tedioso. Por ello, se decidió utilizar nuevas técnicas basadas en cebadores diseñados a partir de regiones que se encuentren próximas a las LTRs como es el caso de la RNasaH en los retrotransposones tipo *copia* (Pearce *et al*., 1999) o de la región PBS adyacente a la LTR 5’ (Kalendar *et al*., 2010).

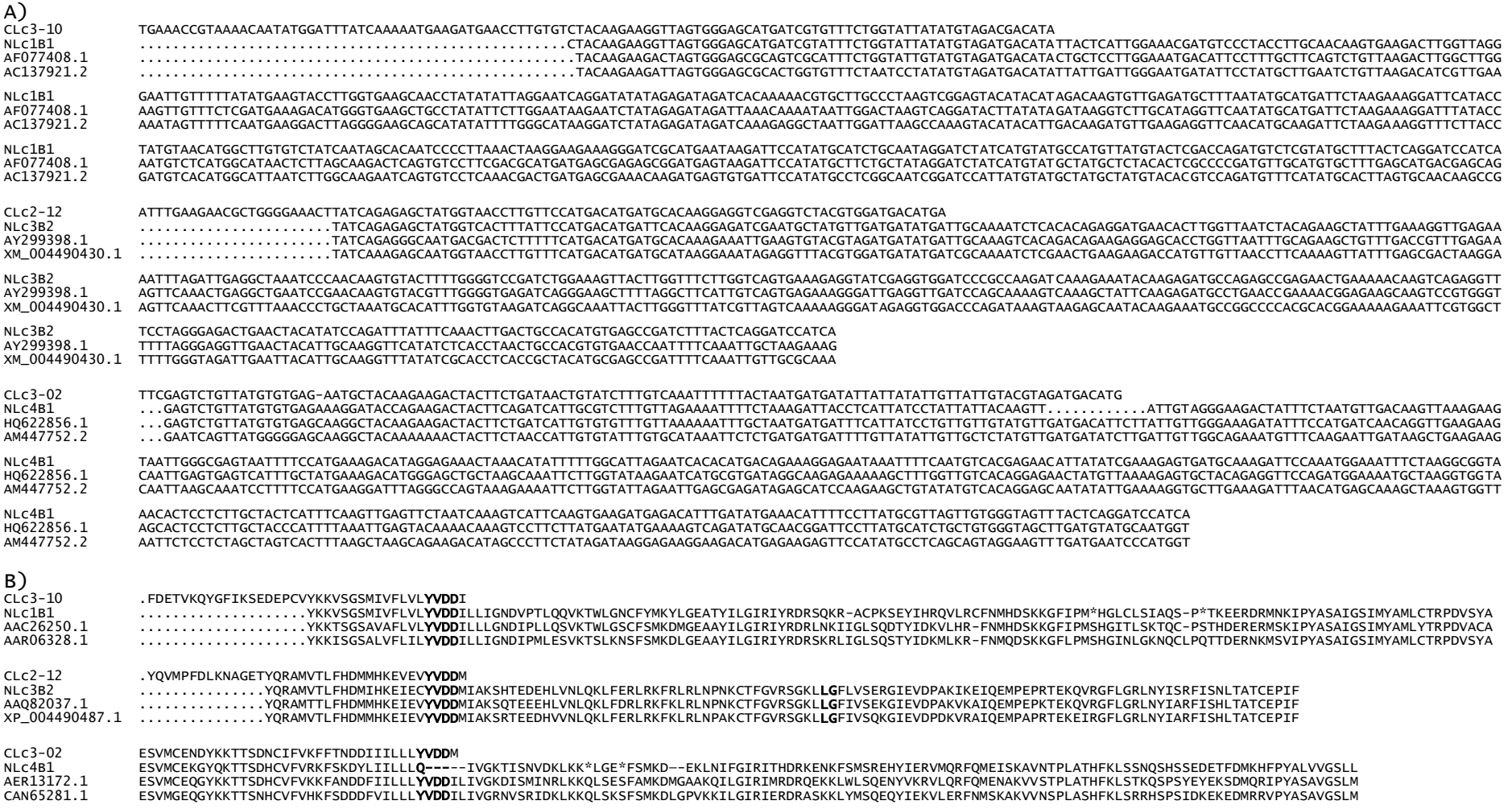


Figura 4.13: Secuencias nucleotídicas (A) y aminoacídicas (B), obtenidas con los cebadores NLc1, NLc3 y NLc4, diseñados a partir de las secuencias analizadas en el apartado anterior. En primer lugar se muestra la secuencia de partida y, a continuación, la secuencia localizada comparada con otras especies diferentes publicadas en la base de datos del NCBI. En negrita se indican las regiones conservadas del gen de la retrotranscriptasa localizadas en estas secuencias.

4.2.2. Identificación de LTRs a partir de la secuencia de la RNasaH

Esta proteína se encarga de degradar el ARN en el heterodúplex ARN-ADN que se forma durante la retrotranscripción y de liberar la PPT (polypurine tract), que presenta resistencia a la enzima debido a la pérdida de una de las histidinas catalíticas propia de las RNasasH del resto de elementos transponibles, y actúa como cebador para iniciar la amplificación de la segunda cadena de ADN. Esta actividad es imprescindible para los retrotransposones LTR ya que éstos, a diferencia de los no LTR, llevan a cabo su ciclo de vida en el citoplasma de las células eucariotas o en el interior de estructuras similares a virus por lo que no pueden acceder a las RNasas de la célula huésped que se encuentran en el núcleo celular, todo ello implica que las RNasas de los retrotransposones LTR presenten dominios propios y un mayor nivel de conservación en sus secuencias (Malik, 2005).

Se utilizaron técnicas basadas en el uso de dos cebadores diseñados a partir de una región interna conservada de dicha proteína (RNasah1 y RNasah2), de un retrotransposón de guisante (Pearce *et al*., 1999).

Con estos cebadores se utilizaron dos técnicas diferentes, ambas basadas en una digestión parcial del genoma seguida de una ligación de adaptadores a los extremos de los cortes y una amplificación doble. La diferencia se encuentra en este último paso de la reacción ya que una de ellas consiste en una PCR anidada o “nested PCR” (Figura 4.14), que consiste en una primera amplificación con el cebador Rnasah1 junto con el del adaptador, seguida de otra amplificación en la que se combina este último con el cebador Rnasah2, mientras que la otra técnica es similar a la utilizada con los cebados NLc del apartado anterior, es decir una PCR sencilla tanto con el cebador Rnasah1 como con Rnasah2.

Estos cebadores sólo se puede utilizar para amplificar retrotransposones del tipo *Ty1-copia* ya que sólo éstos tienen la RNasaH próxima a la LTR (Figura 1.5) y además, la mayor parte de estos retrotransposones, presentan unas 500 pb entre el primer motivo

conservado de la RNasaH y la LTR lo que permite alcanzar esta última región tras una única amplificación (Pearce *et al.*, 1999). Por ello se utilizaron unas condiciones específicas para la digestión del genoma que permitiesen obtener fragmentos de menos de 1000 pb, lo que facilitaría su extracción y clonación para obtener las secuencias deseadas.

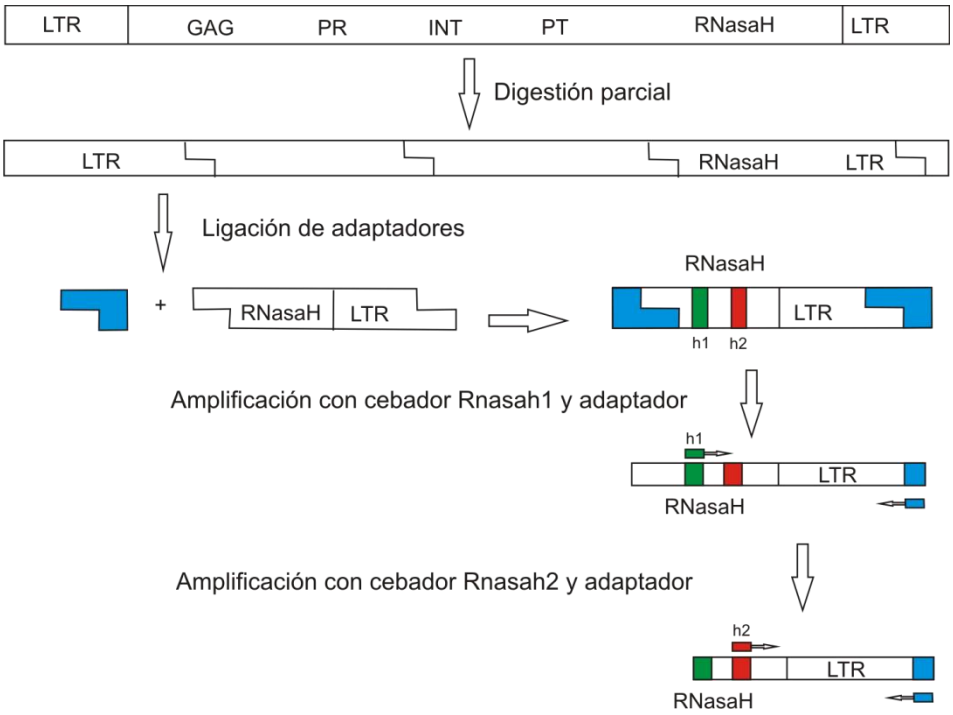


Figura 4.14: Esquema de la técnica "nested PCR" o PCR anidada. Técnica utilizada con los cebadores diseñados a partir de regiones conservadas de la RNasaH.

En la Figura 4.15 se pueden observar los patrones de bandas obtenidos tras las amplificaciones con ambos cebadores.

A partir de estos geles se aislaron las diferentes bandas amplificadas para proceder a la transformación de células competentes. Se seleccionaron 16 de todas las colonias recombinantes crecidas en las placas de LB con ampicilina correspondientes a cada una de las bandas, y se realizaron minipreparaciones de colonias positivas para comprobar la presencia del inserto esperado y, de éstas, se secuenciaron cinco colonias de cada una de las bandas aisladas.

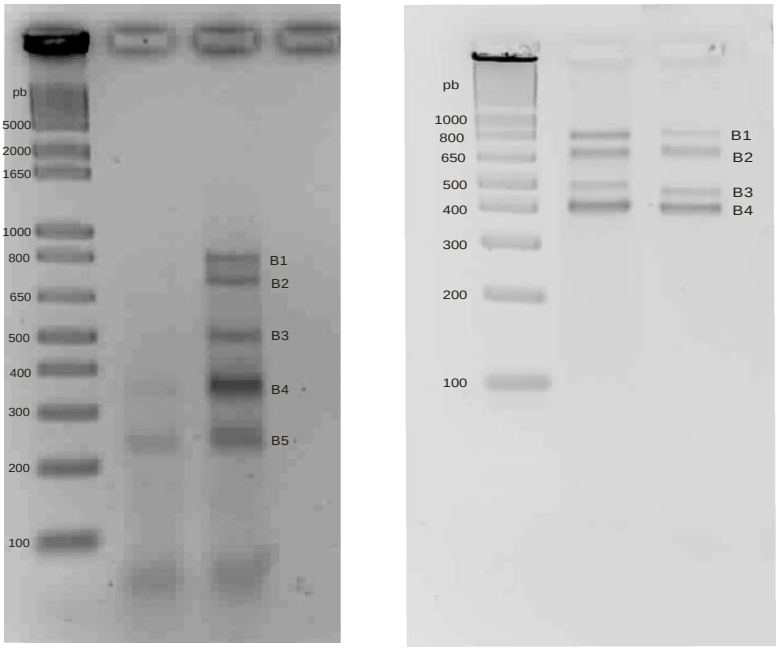


Figura 4.15: Bandas obtenidas con los cebadores RNasah1 (PCR sencilla) a la izquierda y RNasah2 (PCR nested) a la derecha. Se indican las bandas aisladas para su clonación y secuenciación. A la izquierda se indica el tamaño de cada una de las bandas del marcador de peso molecular utilizado como control.

Una vez obtenidas las secuencias, se alinearon mediante el programa MEGA5 de manera que se pudo comprobar que todas aquellas obtenidas a partir de una misma banda eran iguales. Además, se introdujeron en la base de datos del NCBI para compararlas con otras secuencias ya descritas, mediante los algoritmos BLAST.

No todas las bandas obtenidas mostraron homología con la RNasaH de manera que se localizaron retrotransposones similares a los descritos por Pearce (1999) para *Pisum sativum* (*Tps*) y *Vicia faba* (*Tvf*), además de otros retrotransposones diferentes, únicamente con tres de las secuencias amplificadas.

Los nombres de las secuencias indican el cebador con el que se ha realizado la amplificación (RNH1 o RNH2) y el clon del que proceden, precedido de un guion.

Como se puede observar en la Figura 4.16, la secuencia obtenida a partir de la banda 4 del cebador RNasah2 mediante PCR anidada (RNH2-4), presenta homología con alguno de los retrotransposones de *Pisum sativum* (*Tps*3 y *Tps*5) y *Vicia faba* (*Tvf*5 y *Tvf*6) descritos por Pearce *et al*. (1999). En dicho trabajo, se localizaron 9

retrotransposones de *Pisum sativum* (*Tps*), 6 elementos de *Vicia fava* (*Tvf*) y 8 de *Picea abies* (*Tpa*).

La secuencia obtenida a partir de la banda 3 amplificada con el cebador RNasaH1, en este caso se trata de una amplificación sencilla, (RNH1-3), presenta homología con un retrotransposón del tipo *SIRE-1* descrito en *Glicine max*. Este tipo de elementos pertenecen a la familia *Ty1-copia* y son los únicos de esta familia encontrados hasta el momento que presentan el gen *env* que codifica para una envoltura similar a la de los virus por lo que son considerados como proretrovirus (Laten *et al*., 1998). En este caso se ha localizado un retrotransposón que presenta una RNasaH similar a la presente en elementos del tipo *SIRE-1* pero que carece del gen *env*, lo mismo que ocurre con otros retrotransposones descritos en otras leguminosas como *Pisum sativum* (*Tp*3) y *Vicia fava* (Pearce, 2007), la gran similitud que existe entre la secuencia de la RNasaH de ambos elementos sugiere una historia evolutiva común aunque no está claro si el origen está en un virus ancestral que ha perdido el gen *env* en todas las especies de leguminosas excepto en soja o si se trata de una adquisición reciente del genoma de ésta leguminosa (Laten *et al*., 1998).

En la Figura 4.17 se muestran las secuencias de aminoácidos traducidas a partir de los amplificados con los cebadores Rnasah, de manera que se puede observar la secuencia correspondiente a la RNasaH, una vez alcanzado el triplete de fin de traducción muestran los nucleótidos que se corresponden con la región PPT (PolyPurine Track) y la LTR.

La secuencia RNH2-2, aislada a partir de la banda 2 del cebador Rnasah2, presenta homología con un fragmento de una RNasaH tipo 1 como la localizada con los cebadores específicos de *Ty3-gypsy*. En este caso correspondiente al retrotransposón *Ogre* de *Pisum sativum*, pero no se ha alcanzado el extremo de la LTR ya que se trata de un retrotransposón de gran tamaño y, en este caso, la distancia entre la RNasaH y la LTR es demasiado grande como para poder secuenciarla con una sola amplificación de manera que se ha obtenido una región interna de la proteína.

```
RNH2-4    ...............................................ADIFTKALDVNQFEVLRGKLGICIYEDL*
Tps5      ..................................................ADMFTKALDANQFECLRGKLGICVYENL*
Tvf6      ..................................................ADIFTKALDANQLENLRGKLGICQHEKL*
Tvf5      ...............................................ADIFIKALDANQFENLRGKLGICQHEEL*
Tps3      ...............................................ADIFTKALDTNPFECLRGKLRICIHENP*


RNH1-3    PIFCDNTSAINLTKNPVLHSRTKHIDIRHHFIRDLVEDKVITLEHV?TELKLADIFTKALDATQFENLKSKLGICLYEEI*
G. max    .LYCDNLSAINISKNPVQHSRTKHIDIRHHYIRDLVDDKVITLEHVDTEEQIADIFTKALDANQFEKLRGKLGICLLEDL*
Tp3       ........................LRHHFIRDLVEDKVITLEHVATELQLADIFTKALDATQFENLRGKLRICIHENP*


RNH2-2    DPSKCTFGVQSGKLLGFIVSHRGIEVDPVKVKAIQDMPVPETEKQVRGFLGRLNYIARFISHLTSTCEPLFKLLRKDQAIVWNDDCQKAFDKVKEYLQEPPILMPPVPGRPLIMYLTVLDNSMGCVLGQHDESG
Ogre      .PNKCTFGVRSGKLLGFIVSEKGIEVDPAKVKAIQEMPEPKTEKQVRGFLGRLNYIARFISHLTATCEPIFKLLRKNQAIKWNDDCQKAFDKIKEYLQKPPILIPPVPGRPLIMYLSVTENSMGCVLGRHDESG

RNH2-2    RREYAIYYLSKKFTDCETRYSLLERTCCALVWAARRLRQYMLAHTTLLISRMDPVKYIFEKPALTGRVSRWQMILTEYDIQYVTQDPS
Ogre      RKEHAIYYLSKKFTDCETRYSLLEKTCCALAWAARRLRQYMLNHTTLLISKMDPVKYIFEKPALTGRVARWQMILTEYDIQ.......
```

Figura 4.16: Secuencias de proteínas obtenidas con los cebadores RNasah1 y RNasah2, diseñados a partir de dos motivos conservados de la secuencia de la RNasaH de *Pisum sativum*. Solamente se han incluido las regiones codificantes de la proteína.

```
RNH2-4    ADIFTKALDVNQFEVLRGKLGICIYEDL....[41pb].........AGAAG-TGCAAGAATACTCAAGAACTCTGTAAAGGCTGGAAATGTATTCTTTGGTGATGATAAAGCGTTGGTTGAGGTCA (110pb)
Tps5      ADMFTKALDANQFECLRGKLGICVYENL....[70pb].........GAGAGA-TGTGATGGTGTAAAGATTCTATGAGATTTATTTACCTTGACTCAATTTTTGACTAAAAAAGGGGGGAGA
Tvf6      ADIFTKALDANQLENLRGKLGICQHEKL....[-9pb].........GAGGAATTGTAGCAATTAATTTTGCTTAAAACAAGGGTTGAAGGAGTTATTTTTGTTAAGTCTATTTTGATTGGGTTAATC (83pb)
Tvf5      ADIFIKALDANQFENLRGKLGICQHEEL....[-9pb].........GAGAAATTGTAGCAATTAATTTTGCTCAAAACAAGGGTTGAAGGAGTTATTTTTGTTAAGTCTATTTTAAGTGGGTTAATC (79pb)
Tps3      ADIFTKALDTNPFECLRGKLRICIHENP....[ 8pb].......AAAGGAG-TGTGTGGTTAATATCCCAAAGGATATTTCTAACAAAAATACGAAGGAGTTTTGCAAAGTGGTTGAA
RNH1-3    ADIFTKALDATQFENLKSKLGICLYEEI....[42pb]....GATAGAATGATGGTCCTGAGTAAGAACAAGTATTCACTTTCGAGGATTGATGATTTGATGGATCCTGAGTAATGTCGATATGTT (204pb)
Tp3       ADIFTKALDATQFENLRGKLRICIHENP....[29pb].....AAAGGAAGATGTTG
```

Figura 4.17: Secuencias de proteínas, PPT y LTR obtenidas con los cebadores RNasah1 y RNasah2. Estos cebadores se diseñaron a partir de dos motivos conservados de la secuencia de la RNasaH de *Pisum sativum*. En este caso se muestra la secuencia aminoacídica correspondiente a la RNasaH hasta el triplete de fin de la traducción seguida de las secuencias nucleotídicas de la región del PPT (azul) y del inicio de la LTR (verde) ya que no se trata de regiones codificantes. Entre corchetes se indica el número de nucleótidos presentes entre la proteína y la región PPT y, entre paréntesis, el número de nucleótidos omitidos de las secuencias.

4.2.3. Identificación de LTRs a parir de secuencias PBS

Kalendar *et al*. (2010), propusieron el uso de la combinación de cebadores, iguales o diferentes, obtenidos a partir de las regiones PBS (Primer Binding Site o sitio de unión de cebadores) de distintos retrotransposones para generar marcadores moleculares. Esta región se encuentra junto a uno de las LTRs por lo que, con ellos, se consiguió secuenciar los extremos de dos retrotransposones que se encuentren próximos en el genoma.

Kalendar *et al*. (2010) recopilaron las secuencias de retrotransposones LTR publicadas en las bases de datos NCBI, RetrOryza y TREP. En ellas localizaron la región correspondiente al PBS y, a partir de ésta, diseñaron cebadores que permitieran la amplificación de la LTR contigua. Este tipo de cebadores ha demostrado su efectividad tanto en animales como en plantas y es la única técnica descrita hasta el momento que permite obtener las secuencias de las LTR de cualquier tipo de retrotransposones incluidos TRIM (Terminal-Repeat Retrotransposons in Miniature) y LARD (Large Retrotransposns Derivatives), los cuales carecen de dominios codificantes. La única limitación del método es que solamente se amplifican retrotransposones que se encuentren cercanos en el genoma y en sentido inverso uno con respecto al otro (Kalendar *et al*., 2010).

En este trabajo se han utilizado doce de los estos cebadores en reacciones de amplificación en las que se añaden estos cebadores de forma individual o combinados de dos en dos (Tabla 4.4).

Tras la amplificación se realizaron geles de electroforesis para comprobar el número de bandas y el tamaño de cada una de ellas y, en función de estos parámetros, realizar una purificación del gel para integrar las secuencias en el vector de clonación y transformar células competentes directamente en el caso de que el aislamiento de las diferentes bandas fuese muy complicado o, por el contrario, aislarlas y realizar transformaciones de bandas individuales como se había estado haciendo con el resto de cebadores hasta el momento.

Tabla 4.4: Cebadores utilizados con la técnica iPBS.

Cebador	Secuencia	Identificación[1]	Tª óptima
PBS-1	5’ ACTTGGATGCTGATACCA 3’	2231	52ºC
PBS-2	5’ AACCGACCTCTGATACCA 3’	2249	51ºC
PBS-3	5’ GACCTAGCTCTAATACCA 3’	2256	51ºC
PBS-4	5’ GAACTTGCTCCGATGCCA 3’	2373	51ºC
PBS-5	5’ GAACCCTTGCCGATACCA 3’	2398	51ºC
PBS-6	5’ CCCCTCCTTCTAGCGCCA 3’	2400	51ºC
PBS-7	5’ CGACCTGTTCTGATACCA 3’	2229	52,5ºC
PBS-8	5’ TCTAGGCGTCTGATACCA 3’	2230	52,9ºC
PBS-9	5’ TCATGGCTCATGATACCA 3’	2252	53,2ºC
PBS-10	5’ TCGAGGCTCTAGATACCA 3’	2253	51,0ºC
PBS-11	5’ TCCCCAGCGGAGTCGCCA 3’	2395	52,8ºC
PBS-12	5’ AAACTGGCAACGGCGCCA 3’	2399	52,0ºC

[1] Estas cifras indican la numeración de los cebadores en el artículo de Kalendar *et al.*, 2010.

En la Figura 4.18 se puede observar que, por ejemplo, con el cebador PBS1 y con las combinaciones de los cebadores PBS1-PBS2 y PBS1-PBS3, se obtuvieron unos patrones de bandas que hacían casi imposible su aislamiento debido a la gran densidad de bandas obtenidas y a la presencia de bandas de tamaños muy próximos, de manera que se realizó la transformación del amplicón completo, sin embargo en otras amplificaciones se obtuvieron bandas que se aislaron fácilmente, este es el caso del cebador PBS2.

Una vez realizada la transformación y la clonación, se llevaron a cabo minipreparaciones de colonias recombinantes con 16 clones de cada una de las bandas extraídas y se secuenciaron cinco clones de cada una de ellas. En aquellas muestras en las que se transformó el amplicón completo, se realizaron las minipreparaciones con 50 colonias de cada muestra y se secuenciaron una o dos colonias de cada uno de los diferentes tamaños observados.

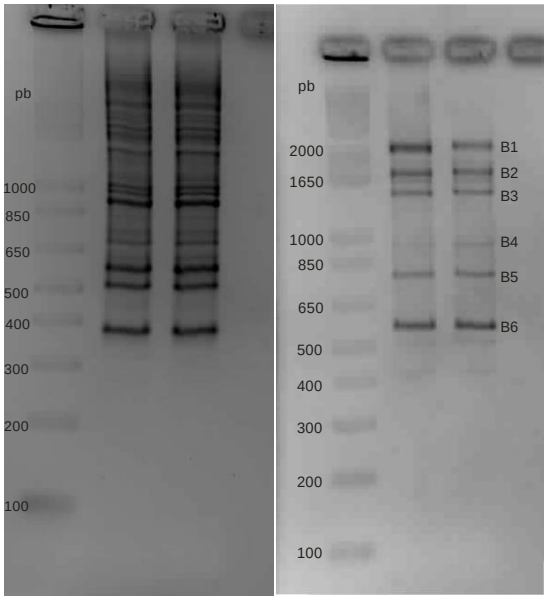


Figura 4.18: Patrones de bandas obtenidos con diferentes cebadores PBS. Patrón obtenido con el cebador PBS1 a la izquierda y bandas amplificadas con PBS2 a la derecha. Se indican las bandas aisladas para su clonación y secuenciación. A la izquierda se indica el tamaño de cada una de las bandas del marcador de peso molecular utilizado como control.

Tras la secuenciación se realizó un análisis de secuencias utilizando el programa MEGA5 de manera que se pudo comprobar la presencia de 135 secuencias diferentes de las cuales, tras su análisis BLAST, solamente 73 mostraron homología con fragmentos de retrotransposones y se han podido localizar seis LTRs semejantes a secuencias publicadas en la base de datos del NCBI:

- LTR completa del retrotransposón tipo *Angela* descrito con más detalle en el siguiente apartado (ver apartado 4.2.4). En la Figura 4.19 se muestra la secuencia para *Lens* obtenida a partir de la amplificación con el cebador PBS7.

- LTR del retrotransposón *Cassandra*. Se trata de un retrotransposón TRIM (Terminal-Repeat Retrotransposons in Miniature), éstos están compuestos por LTRs de entre 100 y 250 pb y un dominio interno formado por un PBS muy conservado complementario de una metionil tRNA y un PPT separados por una pequeña secuencia que puede llegar a contener únicamente 34 nucleótidos, de manera que carecen de regiones que codifiquen para proteínas, por lo que se trata de elementos no autónomos que dependen de otros retrotransposones para replicarse (Antonius-Klemola *et al*., 2006; Sabot y Schulman, 2006; Kalendar *et al*., 2008). El retrotransposón *Cassandra* ha

sido descrito en diversas especies de todo el reino vegetal y tiene como característica especial la presencia de un RNA 5S incluido en la LTR, este RNA presenta promotores, para la RNA polimerasa III en lugar de para la polimerasa II propia de los retrotransposones LTR, estos promotores (cajas A, EI y C) están muy conservados al igual que el terminador (Cloix *et al*., 2003; Kalendar *et al*. 2008). Todos ellos están señalados en la Figura 4.19. Esta secuencia se obtuvo con los cebadores PBS1, PBS2 y la combinación de PBS2 y PBS3.

- LTR completa de un retrotransposón de la familia *Ty1-copia* denominado gmw2-109b11-re-3 descrito en soja (*Glycine max*) (Wawrzynski *et al*., 2008). Esta secuencia se ha obtenido tras la amplificación con las combinaciones de cebadores PBS1-PBS3 y PBS3-PBS5.

- Extremo 3' de la LTR del retrotransposón tipo *SIRE* (*Ty1-copia*) denominado *SIRE1*-13 (Pearce, 2007). Esta secuencia se ha obtenido tras la amplificación con los cebadores PBS3, PBS7 y la combinación de PBS2 y PBS3. Se trata de retrotransposones descritos en *Glycine* con características poco habituales ya que presentan un gen que codifica para la envuelta propia de los retrovirus y es el único conocido de este tipo que pertenece a la familia *Ty1-copia* además de presentar una gran longitud (11Kb) (Laten *et al*. 1998). Se han realizado análisis de su secuencia que demuestran que se trata de elementos relativamente nuevos en el genoma de soja (Laten *et al*., 2003). La región descrita se corresponde con la LTR por lo que no se puede asegurar que este elemento presente el gen *env* en *Lens* ya que, hasta el momento, se han encontrado homologías entre secuencias de diferentes leguminosas pero sólo se ha localizado dicho gen en *Glycine*.

- Extremo 3' de una LTR similar al *Peabody* de la familia *Ty3-gypsy* descrito en guisante (*Pisum sativum*) (Volman *et al*., 1998). Este retrotransposón presenta 7,8 Kb aproximadamente y se estima que forma entre el 2 y el 3% del genoma de guisante siendo muy baja su variabilidad lo que sugeriría una aparición reciente (Macas *et al*., 2007). Esta secuencia se obtuvo con los cebadores PBS1, PBS3 y la combinación de PBS1-PBS2.

- Con la combinación de cebadores PBS1 y PBS7 se ha conseguido secuenciar un fragmento de LTR que coincide con la región PBS y una LTR parcial de un retrotransposón *Ogre* de *Lathyrus sativus*. (Ceccarelli *et al*., 2010).

En la Figura 4.19 sólo se indican las secuencias que se corresponden con las regiones LTRs de los retrotransposones amplificados con los cebadores PBS.

A)
PBS7 TGTTGGTGTAAGCCC-TAGAGGCCATTGTTTATGGTACTTGTATCGAATTATTTATTAATTATAAAAGAAATTTCTTTATTATGTTTGTTGTTCCAAAATAATAAAGTCCCTATAATAGCTAGTCTGTTTAATGAAATGTTAAGTATGACTTAATC
Psat TGTTGGTGTAAGCCCCTAGAGGCCATTGTTTATGGTACTTGTATCGAACTATTTATTAATAATAAAAGACAGTTCTTTATTAGGTTTGTCGTTCCAAAATTATAAAGTCCCAAGAATAGCTAGTCCGTTTAATGAAATATTAAGTGTGACTTAATC
PBS7 ATGAGATCCCATTAAACATAAGAACACTATTCTTAAAGTATCTGTAGTCAAGCTTTATTGTGAAATGGGATAACATTAAAGCATTAAGACTATAATGTAGATAGGCTGATAATCGCATCTCATGGATCATGGATAAGGAGTTATCAAGTCTTAA-C
Psat ATGAGATCTCATTAAACATAAGGACACTATTCTTAAAGTATTCGTAGTCAAGCTTTATTGTGAAATGGGATAACATTAAAGCATTGAGACTATTATGTAGATAGGCTGATGATCACATCTCATGGATCATGGATAATGAGGTATCAAGTCTTAAAC
PBS7 ATAAGTATGAATATTAGGAGTAATATTTATACTGGATTGACCCACTATGAGAGTACTACAAAAAATGTTATGCAAAGTGTCATAATTTATTCTCATGGTGATAGTGGTGTGTACCACCATTCGACCTGAAACCACTATGGACCTTAGATGTAGAGT
Psat ATAAGTATGAATATTAGGAGTAATATTTATACTGGATTGACACGCTATGAGAGTACTACATAAAATGTTATGCAAAGTGTCATAAGTTATTCTCATGGTGATAATGGTGTACTCCAC----------
PBS7 GAGTGCTTTATTGCTGATCAAACATTGTCCGTAGCTGGATGACCATAAAGACAATTGATGGGCACTCCACAAAGCATGCTGAGGGACACGAGTGACCTAGATGGATTTTTCCCATCCTGCGT-ACAGGATAA-TGTCTAACGGTCCAATATTGAAC
Psat ----------AAAGCATGCCGAGGGACATGAGTGACCTAGATGGAATTTGTCCATCCTGCGTAACAGGATAAATGTCTAAGGGCCCAATATTGAAC
PBS7 TAGACAAGGGGGTCAAATCTGTTACCTTATGTTTGATATAGACATAGGGGCAAAAGGGTAACTGTACATACAAGTAATTATCACGAAAAGGTTTTTGTCAGATCACAT----------GGGTA---GTCATG-ATTGCTATGCTT-CAG
Psat TAGACAAGGGAGACACGGATTATGCCTTGTGTTTGATATAGACATAGGGGCAAAAGGGTAACTGTACATACAAGTAATTATCACGAAAAGGTTTTTGTCAGATCACATGAAAATTTTGAGACTTTGGGTAGCAGTCATGTATTGCTAGATAAACCC
PBS7 -ACA-TTGCTCATGTTAAA-GCATGATTTAA-ATAAT-TGCCAACGTCGCGAGAACCTACCAGGGTCAC-ACACATAAGGACAAATTTATGAGAGATAGAGAAAATAAGGAACACCGTAAGGTACGTTGTTCGTAAGTGGAATACGAAATATGGTA
Psat TCACTATTATTATGTTAAATGCATGTTTTAAAATAATTTGTCAACGTCACGAGAACCT-ATAAGGTCACTACACATAAGGACGAATTGATGAGAGATAGAGAAAATAAGGAACACCGTAAGGTACGGTGTACTTAAGTAGAATACCAAATATGGTA
PBS7 ATATATCATATGCTTAAGTAGAATACGAAATATGGTAAAA-TACCATACGCTTAAGTGAAATATAGTATGGTACGCTTAAATGGGCTTTTTAGCTTAAAGCACACACAAGTGGTTCTGTAAATAGAACACTTGTGTATAAGCCA-----TCAGTTG
Psat AAATACCATAAGCTTCAATAAAATAAGAAATATGGTAAAAATATCATACGCTTAAGTGAAATACAGTATGGTACGCTTAAATGGGCCTTTTAGCTTAAAGCCCACACAAGTGGTTCTATAAATAGAACACTTGTGCGTAAGCTATTTTCCCAGTTG
PBS7 TTGAAACTCTGAATTACGTATTCCCTCTTCTCCCTCTCTCAAAGGATTTCATTAATATCAGCTAGCACTGAGACTGAAGGCGTTCTGTTTGTGTGGACTGAGTATAGGCGTTGTTCTTCAATCAACGCACATGGTCAATTCATCAATACTGGAGCA
Psat TTGAAACCCTGAATTTTGTATTCCCTCCTCTCCCTCTCTCAAAGGTCTTCATTCATATTAGCTAGCATTGAGCCTGAAGGCATTCCGTTCGTGTGGTCTGAGTAGAGGCATTATTCTTCGATCAACGCACATGATCAATCCTTCGATACTGCATCA
PBS7 AAGAAGTGAAATTTTTGTTTTTCCGCTGC----------GTTTT--GGATCGTGACTTCCCTTTAGT-GGTATCAC
Psat AGGGTTTTAA----------CCGCCACAAGAGGTAACTATTTATATCACAGATCATGCCCATTCGTAAGGATCAC
B)
PBS1 TGTAACACCCGAGA--CCGAAGAGGGCGGGGAGT----ACCCGAGGCCGATGTGGGCGG-AGAGTGGTTGTCGGTGCATGAGGCATGACAATGAGACACTCTTGGAACCTTCT-AGGCAAAAACCGGATTCACATCGGAATGAGAGGGATCTTGAG
PBS2 TGTAACACCCGAGA--CCGAAGAGGGCGGGGAGT----ACCTGAAACCGATGAGGGCAG-AGAGTGGT-GCCGGTGCATGAGGCATGACAATGAGACACTCTTGGAAGCTTCT-AGGAGAAAACCAGATTCACATCGGAATGAGAGGGATCCTAAG
PBS2/3 TGTAACACTCGAGGA-CGGA-GA----------GGAAGG-AGAGTGGTTATCGGTGCATAAGGCATGACAATTAGACACTCCTGGAAGCTGCT-AGGCGAAAAACCAATTCACGTTGGAATGAGAGAGATCCTGAG
TRIM TGTAACACCCGAGG--TCGGAGAGGCCGAAGAGTGGTTACATGTAACACCCGAGGGCGATGGAGTGGTTGTCGATGCATGAGGCATGACAATGAGACACTCCTAGCAGCTTCTAGGGGAAAAACCGAATTCACATCGGAATGAGAGGGATCCTGAG
PBS1 GCTGTGCAGGTATGGGACTGCATGGTTAAAGGAAGGCTTATAAGGAAAGATTGGTACTACCTATACAAACAAGATGCATCTTCTTTTCGGTAGCCTAACAGATAAGAACTCCACAGTTAAGCATGCTTGACTTGGAGTAGTATTTGGATGGGTGAC
PBS2 GCTGTGCATGTATGGGACTGCATGGTTGAAGGAACGCTTATAAGGATTAATTGGTACTACCTATAAAAACAAGATGCATCTTCTTTTCGGTAGCCTAACAGTTAAGAACTCCACGGTTAAGACTGCCTGGCTTGGAGTAGTATTTGGATGGGTGAC
PBS2/3 GCTGTGCAGGTATAAGACTACATGGTTAAAGGAAGGCTTATAAGGATAGATTGGTACTACCTATACAAACAAGATGCATCTTCTTTTCGGTAGCCTAACAGATAAGAACTCCACAGTTAAGTGTGCTTGACTTGGAGTAGTATTTAGACGGGTGAC
TRIM GCTATGCAGGTATGAGACTGCATGGTTGAATGAAGGCTTATAAGGATTAATTGGTACTACCTATACCAACAAGATGCATCTTCTTTTCGGTAGCCCAATAGATAAGAACTCCATAGTTAAGCGTGCTTGACTTGGAGTAGTATTGGGATGGGTGAC
PBS1 CTTCTGGGAAGTTTTCTGGAAAGTGTGTGAGTGAGGACAAAGCATGCCGAAAAGACTCGTGTTGGTTTGTGGGGTCAGTCGGTAAATCCTGAAAGTAGTATGGGGCGTTACA
PBS2 CTTCTGGGAAGTTTTTCGAAAAGTGTGTGAGTGAGGACAAAGCATGCCGAAAAGACTCGTGTTGGTTTATGGGGTCAGTCGGTAAATCCTGATAGCAGTATGAGGCGCTACA
PBS2/3 CTTCTGAGAAGTTTTCCAGAAAGCGTGTGAGTGAGGACAAAGCATGCTGAAAAGACCCGTGTTGGTTTGTGGGGTTATTCGGTAAATCCTAAAACTAGTATGAGGCATTACA
TRIM CTTCTGGGAAGTTTCCCGGAAAGTGTGCGAGTGAGGACAAAACACGCTGAAAAGACTCGTGTTGGTTTGTGGGGTCAGTCGGTAA-TCCTGAAAGCAGTTTGGGGTGTTACA
C)
PBS1/3 TGTTGGTAGTTTAATAGTTGATATAAA---------TTTTTTTATAAATTTGTTTAGAAAGAAG-ACTATGCCATTTATTTCTTGAATGATCATTTAAAAA-TTATCAATTGCCTATTTACAGGCGCAATGCA----------
PBS3/5 ----------TTG-TATATATGATATAATCCCTTTTATATATTATGTTTAAAAGAAG-ACTATGCCATTTATTTCTTGAATGATCATTTAAAAA-TTATCAATTGCCTATTTACAGGCGCAATGCA----------
Gmax TGTTGGTAGTTGATTAGATGATAGTTGATAGTTGATAGTTTTAGAGAA-TTGTTTTAGAAAGAAGGCTATGTCATTGATTTCTTGAATGATTAATTACAACATA-CCAATTGCCTATATATAGGCTCAAGTCACCAACCTCTCAATGATGGTGAAT
PBS1/3 ----------CTTTCTAG-GTGTTTCATGATAGATTAGTATCATGATAGTTCTAGGTTCTACTATCATTTCC-ATACACTTG----------
PBS3/5 ----------CTTTCTAG--TGTTTCATGATA-ATTA-TATCATGATAGTTCTAGGTTCTACTAT-ATTTCC-ATACACTTG----------
Gmax GTTTTTACATTACTTTAGGTCTTTCTAGAGT-TTTCATGATAGATTAGTATCATGATAGTTCTAGGTTCTTCTATC-TTATACATACACTTATAGTTCTAGATTGTTCTACCATATTCATACATATTATAGTTCTAGATAGTTCTACTATATTCAT
PBS1/3 --------TAGTTCTAGGTTCTTATATCATTTGCATACACATTATATTTAAGAATATTCTAGAAATTTGTAGCAATTTCAACA
PBS3/5 --------TA-TTCTAGGTTCTTATATCATTTGCATACACATTATATTTAAGAATATTCTAGAAATTTGTAGCAATTTCAACA
Gmax ACTCACTATAGTTCTAGGATATTCTACCATTTTCATACATATTATATTTAAGAATATTCTAGAAATTTGTAGCAATTTCAACA

```
D)
PBS3      GGGAAGTGAGAGGGG-TCTCATATCTAGGAGT---TCTTAGATAGAAATATAGCACGGGTAGTGATTAGGTGAGTAGATTGTAAATTGGTGATAGTTTGTTGTTGGTCTACGAACTGATATTATTTAGTGGATTTCCTTC-CTGGCTTGGTAGCCC
PBS2/3    GGGAAGTGAGAGGGG-TCTCATATCTAGGAGT----TTTAGATAGAAATATAGCACGGGTAGTGATTAGGTGAGTAGATTGTAAATTGGTGATAGTTTGTTGTTGGT-TACGAACTGATATTATTTAGTGGATTTC--TT-CTGGCTTGGTAG-CC
PBS7      GGGGAGTGAGTAGG-TTCTCATACCTAAGAG--GTTCTTAGGTAAAAATT-ACATTGGGTAGATATTAAGTGAGAAG-TAGTAAACCAGAGGTTGTTTGTACTTTG--AATTAATAT-TAC--TTTAGTGGATTTTCTTCCCTGGCTTGGTAGCCC
SIRE      GGG-AGTGAGAGGGGTTCTCATATCTAAGAGTGGCTCTTAGGTAGAGATT--GCACGGGTAGTGGTTAGGTGAGAAGGTTGTAAA-CAGTGACTGTTAGATCTTCG--AACTAACAC-TAT--TTTAGTGGATTTCCT-CCCTGGCTTGGTAGCCC

PBS3      CTAGATGTAGGTGACGTTGCACCGAACTGGGTTAACAACTCT-GTGTGTT-CTTTACTTACT---GTTGTACTTCTTACCTTGCCATTGTTTTGTGTTCAGAGATCA-GTGTCTCCACATCTTGTAGT-ACATCGAAGATCTGAC---TGC-----
PBS2/3    CTAGATGTAGGTGACGTTGCACCGAACTGGGTTAACAACTCT-GTGTGTT-CTTTACTTACT---GTTGTACTTCTTACCTTGCCATTGTTTTGTGTTCAGAGATCA-GTGTCTCCACATCTTGTAGT-ACATCGAAGATCTGAC---TGC-----
PBS7      TCAAACGTAGAT-ACGTTGCATCGAACTGGGTTAACAATTGAT-TGTGTTCATTTATTTACAATTGTTT--TCATTT
SIRE      CCAGATGTAGGTGACGTTGCACCGAACTGGGTTAACAATTCTCGTGTGTT-ATTTACTT------GTTTAATCTGTT-CATACTCTCAAATATAA--TCTG-CATGTTCTGAAGCGTGAT---GTCGTGACATCT--GGTACGACATCTGGCATTG

PBS3      ----CAGAATTTCA
PBS2/3    ----CAGAATTTCA
SIRE      GTATCAGAATTTCA


E)
PBS1      TGTAT-----TTCAA-T--ATTATACG-TGTGTTTGTGTTTGAGTTTATGTTTTTG-ATTAAGATGATATTGTTTTGCCTGAATATATGATCTAATTAGGGTGAATGA-TGTGTTTGGTTACTTAGCATTACATAATGTTTTATAATGCTTATTAT
PBS3      TGTAT-----TCCAA-T--ATTATACG-TGTGTTTGTGTTTGAGT--ATGATTATG-----AGTTGACATTGTTTTGCTTGAATATATGATCTAATTAGGATGATTGA-TGTGTTAATTTACTTAACATTACATAATATCCTATAATGCTTGTTAT
PBS1/2    TGTAT-----TTCAA-T--ATTATACG-TGTGTTTGTGTTTG-AGTTATGTTTTTG-ATTAAGATGATATTGTTTTGCCTGAATATATGATCTAATTAGGGTGAATGA-TGTGTTTGGTTACTTAGCATTACATAATGTTTTATAATGCTTATTAT
Peabody   TGTATGAATATTTAATTTAATTGAGGAGTGGAATTGTGTTGATATTAAGTATGATG-TTTAAGTTGATGTGCCGTTAC-TGAATGTATGTTGCGATTAGGGTGATGAAATGTGTAAATTTACTTAGCATTACATGATGATTTATAATGCTTATTAT

PBS1      ATTGAATGAAAAACTCACCCTTACAACTATTTTTCAGGTAATGATCAGTGAGGATTGAGTAGAAGCTAGTATACGAAGTCC---------TAGACTAGTGGGTCATGCTCTGGTAGATGTAACATCGGGATGGGATTTGTGACGTTTGGTTTTGTTG
PBS3      ATCGATTGAGGAACTCACCCTTACAACTATTTTTCAGGTAATGAGCAATGAGGACAGAGTAGGAGATAGTATGCGAAGTCC---------TAGACTAGTGGGTCATGCTCTGATAGGTGTAACATCGGGTTGGGATTTTGTGCTATCTATTTTGTTG
PBS1/2    ATTGAATGAAAAACTCACCCTTACAACTATTTTTCAGGTAATGATCAGTGAGGATTGAGTAGAAGCTAGTATACGAAGTCC---------TAGACTAGTGGGTCATGCTCTGGTAGATGTAACATCGGGATGGGATTTGTGACGTTTGGTTTTGTTG
Peabody   ATCGATTGAGGAACTCACCCTTACAACTATTTTTCAGGTAACGAGCAATGA--GTTGAGTAGAAGCTAATGCTTGGAGTCTAGTGTAGTCTCCTAAGTGGGTCATGCTCTGATAGATGTAACATCGGGACGGGAT--GTTTTTAATTGTTTTATTG

PBS1      GTT--TTGCTGAATAAATT-ATATGTATTTATTTTACATATTTTATAGGA----TTGAATTTTATTCCGTTGCGA-CGATGCAATAT---TTTGGTTTGTTTAATA-AATGAGCATGACAGGTTTATTAGTGATTTG-TGTGAAGTGTCAGTGTGA
PBS3      TTT---TGGTGAATAAATT-ACATGTAAT-ATGATACATGGTTTGCTGTTG--TTGGATTTTCTTTCTGTTGCGAATTGTGCAATA----TTTAGTTTGATTAATA-AATGAGCATGACAGGTTTCTTTGGTGAATGGTGTGAAGTGTCAGTGTGA
PBS1/2    GTT--TTGCTGAATAAATT-ATATGTATTTATTTTACATATTTTATAGGA----TTGAATTTTATTCCGTTGCGA-CGATGCAATAT---TTTGGTTTGTTTAATA-AATGAGCATGACAGGTTTATTAGTGATTTG-TGTGAAGTGTCAGTGTGA
Peabody   TTGG-TTGTTGAATCATTTTACATGTAAT-ATGTTATATGTTTTGAATGGTTGATTTAATTTCTATCCGCTGCGTGTGATGCAAGAAATGTTTATTTTGATTGA-ATAAAGAGCATGACAG--TTATTA------TGGTGTGAAATGTT--TGTGA

PBS1      CACCCTTGAATTGCATGTATACTCTGATT-TGTAAT--TTTATTGTTATATATTTAC-GGGGTATTTTAGAAGGATGTTA-CA
PBS3      CACCCTTGGATTGCATAT-TACTCTGATT-TATATTTTTTTGTTTTTAATTAATTATTGGGGTATTTTAGAAGGGTGTTA-CA
PBS1/2    CACCCTTGAATTGCATGTATACTCTGATT-TGTAAT--TTTATTGTTATATATTTAC-GGGGTATTTTAGAAGGATGTTA-CA
Peabody   CACCCTTGAA-TGCATTTAT-TTCTGATCGAAAAGTT-GTAACTTTAATTAAATAATTGGGGTATTTTAGGAGGGTGTTAACA


F)
PBS1/7    CCTGTGGAGTACCACAGACGTGAGGGGTGCTAAT-ACCTTCCCCTTGCGTAACC--AACATCCTTACCC-ATCTCTTGGTTGC-GA-GACC-AT-CAGATTCTCTT---TTTCGT-AGGTTTA-CTAG--CGTTTTCCTTTCCCTCTCTTTGGGAT
Lathyrus  CC-GTCGAGTACGACGGACGTGAGGGGTGCTAAT-ACCTTCCCCTTGCGTAACCGA---CTCCCGAATCCAGGTTTTGGTTGC-GACGACC-ATAC---CTTTCCCGTGTTGTGCGAGGTTTACTCTGAGCG-TTTCCTTTCCCTCCCGT-GGGA

PBS1/7    -AAATAA-TGATGGTGGCGCACTCCGCTGGGGA.....................................................
Lathyrus  TAAATAACGCACGGTGGCGACTCTTCTGTTCAGCCGGCGCCGATTTTCGCAGTAGCGACAGCTGGCGACTCTGCTGGGGATC
```

Figura 4.19: Secuencias obtenidas a partir de los cebadores PBS. El nombre de las secuencias indica el cebador con el que se obtuvo cada una de ellas. Cada uno de los grupos de secuencias se corresponde con cada una de las LTR obtenidas: A) retrotransposón tipo *Angela*, B) *Cassandra*, en rojo se indica la secuencia correspondiente al RNA 5S y recuadrados los promotores para la polimerasa III: caja A (verde), caja EI (morado) y caja C (marrón) y el terminador (azul), C) retrotransposón descrito en *Glycine max*, D) retrotransposón tipo *SIRE*, E) retrotransposón similar a *Peabody* de *Pisum sativum* y F) retrotransposón *Ogre* descrito en *Lathyrus sativum*.

4.2.4. Identificación de un LTR completo descrito en Lens culinaris

Además, se probaron los cebadores LTR-Ps y LTR-1F, éstos se utilizaron para comprobar la obtención de una LTR completa de *Lens culinaris* descrita por Smýkal *et al*. (2009). La amplificación produjo una banda única que pudo secuenciarse sin necesidad de clonar. La secuencia incluye un SSR que también se identifica en la secuencia descrita por Smýkal *et al*. (2009). El resultado de la amplificación se muestra en la Figura 4.20.

Existen teorías que explican la aparición de los microsatélites a partir de elementos repetitivos del genoma tanto en plantas (Koike *et al*., 2006; Tero *et al*., 2006), e insectos (Wilder y Hollocher, 2001; Meglecz *et al.* 2007) como en nematodos (Johnson *et al*., 2006), lo que sugiere una relación entre los SSR y los elementos transponibles. En estas teorías se basó el grupo de Smýkal *et al*. (2009) para obtener cebadores a partir de regiones de LTRs deducidas a partir de SSRs conocidas en el genoma de guisante y de secuencias completas de retrotransposones. De esta manera se localizó un nuevo grupo de retrotransposones de la familia *Ty1-copia*, denominados *Angela*, que presentan un microsatélite incluido en la secuencia de las LTRs. Este grupo fue localizado en 30 especies diferentes de leguminosas y en todas ellas presentaba microsatélites basados en la repetición del dinucleótido TC. En el caso de *Lens culinaris* este SSR se corresponde con $(TC)_3$ CC $(TC)_3$ (Figura 4.20).

Estas regiones repetitivas podrían estar relacionadas con la iniciación de la transcripción en determinadas condiciones mediante la formación de ADN-H o de triple hélice cuando se encuentran antes de la caja TATA (Lu *et al*., 2003) que es precisamente, lo que ocurre en los retrotransposones tipo *Angela*. En este LTR se puede observar la caja TATA a 15 nucleótidos del microsatélite aunque, en la secuencia obtenida en el presente trabajo, presenta una mutación de un nucleótido con respecto a la secuencia de Smýkal *et al*. (2009).

```
LTRL    TGTTGGTGTAAGCCCCTAGAGGCCATTGTTTATGGTACTTGTATCGAACTATTTATTAATAATAAAAGACAGTTCTTTATTAGGTTTGTCGTT
Lens    TGTTGGTGTAAGCCC-TAGAGGCCATTGCTTATGGTACTTGTATCGAATTATTTATTAATAATAAAAGGCAGTTCTTTATTATGTTTGTTGTA

LTRL    CCAAAATTATAAAGTCCCAAGAATAGCTAGTCCGTTTAATGAAATATTAAGTGTGACTTAATCATGAGATCTCATTAAACATAAGGACACTAT
Lens    CCAAAATAATAAAGTCCCTATAATAGATAGTCCGTTTAATGAAATGTTAAGTGTGACTTAATCATGAGATCCCATTAAACATAAGGACATTAT

LTRL    TCTTAAAGTATTCGTAGTCAAGCTTTATTGTGAAATGGGATAACATTAAAGCATTGAGACTATTATGTAGATAGGCTGATGATCACATCTCAT
Lens    TATTAAAGTATCAGTAGTCAAGCTTTATTATGAAATGGGATAACATTAAAGCATTGAGACTATTATGTGGATAGGCTGATGATCACATCTCAT

LTRL    GGATCATGGATAATGAGGTATCAAGTCTTAAACATAAGTATGAATATTAGGAGTAATATTTATACTGGATTGACACGCTATGAGAGTACTACA
Lens    GGATCATGGATAAGGAGTTATCAAGTCTTAA-CATAAGTATGAATATTATGAGTAATATTTATATTGGATTGACCCGCTATGAGTGTACTACA

LTRL    TAAAATGTTATGCAAAGTGTCATAAGTTATTCTCATGGTGATAATGGTGTACTCCAC----------------------------------
Lens    TAAAATATTATGCAAAGTGTCATAAGTTATTCTCAAGGTGATAGGGGTGTGTACCACCCTTCGACCTGAAACCATTATAGACCCTAGATGTAG

LTRL    ----------------------------------------------------------------------AAAGCA-TGCCGAGGG-ACA
Lens    AGTCAAGTGCTTTTTTGCTGATCGACCGTTGTCCCTTTTTGGCTGACCCTAAAGGCCGTGGGGGTACTCCCCCAAAGCCTTGCCTGGGGGACC

LTRL    -TGAGT-GACCTAG-ATGGAATTTGTCCATCCTGCGT-AACAGGATAAA-TGTCTAAGGGCCCAA--TATTGAACTAGACA-AGGGAGACACG
Lens    TTGAGGGGACCTAGCATGGGTTTT-CCCCTCCTGCGTTAA--GGATACAGTATT-AA--GCCCTTCATATTGAAC--GGGTCAGGGGGG-ACA

LTRL    GATTAT---GCCTTGTGTTTGATATAGACATAGGGGCAAAAGGGTAACTGTACATACAAGTAATTATCACGAAAAGGTTTTTGTCAGATCACA
Lens    AAGTATACGGCCTT-AGATCAAAAAAGACATAAAGGCAAAAGGGTAATTGTACATACAAGTAATT-TCATGGAAAGGTTTCTGTCATATCACA

LTRL    TGAAAATTTTGAGACTTTGGGTAGCAGTCATGTATTGCTAGATAAACGCTCACTATTATTATGTTAAATGCATGTTTTAAAATAATTTGTCAA
Lens    TGAAAATTTTGTGACTTTGGGTAGAAGTGATGTGTTGCTAGATAACCAATCACTTTTATTATGTTAAATGCATGATTTAATATAATT-GCCAA

LTRL    CGTCACGAGAACCT-ATAAGGTCACTACACATAAGGACGAATTGATGAGAGATAGAGAAAATAAGGAACACCGTAAGGTACGGTGTACTTAAG
Lens    TGTCGCGAGAAACTTACAGGGTCAC-ACACATAAGGATGGATTGACAAGAGATAGAGAAAATAAGGAACACTGTAAGATACGGTGTACTTAAG

LTRL    TAGAATACCAAATATGGTAAAATACCATAAGCTTCAATAAAATAAGAAATATGGTAAAAATATCATACGCTTAAGTGAAATACAGTATGGTAC
Lens    TAGAATAAGAAATATGGTAAAATACTATACGCTTAAGTAAAATATGAAATATGGTAAAA-TACCATACACTTAAGT-----------------

LTRL    GCTTAAATGGGCCTTTTAGCTTAAAGCCCACACAAGTGGTTCTATAAATAGAACACTTGTGCGTAAGCTATTTTCCCAGTTGTTGAAACCCT-
Lens    -CTTAAATGAGATTTTTAGCTTAAAGCTCACACAAGTGGTTGTATAAATTGAACACTTGTGCATAAGCTAAGTTC--AATTGTTGAAACCCTA

LTRL    --GAATTTTGTATTCCCTCTCTCCCTCTCTCAAAGGTCTTCATTCATATTAGCTAGCATTGAGCCTGAAGGCATTCCGTTCGTGTGGTCTGAG
Lens    CTGAATTTCGTATTCCCTCTCTCCCTCTCTCAAAGGCCTTCATTCATATCAGCTAGCACAGAGACTGAAGGCATTCTGTTCGTGTAGACTGAG

LTRL    TAGAGGCATTATTCTTCGATCAACGCACATGATCAATCCTTCGATACTGCATCAAGGGTTTTAACCGCCACAAGAGGTAACTATTTATATCAC
Lens    TGGAGGCGTTGTCCTTGGATCAACACACGTGATCAATCTATCGATACCAGAGTAAAGGTTTTAATCGCCATAAGAGGTAATTGTTTCTATCAC

LTRL    AGATCATGCCCATTCGTAAGGATCAC
Lens    TGATCATGCCCCTTCGTAAGGATCAC
```

Figura 4.20: Secuencia obtenida tras la amplificación con los cebadores LTR-Ps y LTR-1F (LTRL) comparada con la secuencia descrita para *Lens culinaris* por Smýcal *et al*. (2009) (*Lens*), en rojo se muestra la secuencia del microsatélite y en azul la caja TATA.

4.3. Obtención de marcadores moleculares

Para llevar a cabo el análisis de polimorfismos, se ha utilizado una población RIL (línea consanguínea recombinante) F_7 de 94 líneas, pertenecientes al cruzamiento entre la forma cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880.

Se han utilizado diferentes tipos de marcadores moleculares como son SSAP, REMAP (basados en el uso de la región LTR) y PBS, obteniéndose un total de 256 marcadores: 48 del tipo SSAP, 9 REMAP y 199 del tipo iPBS.

Los marcadores moleculares basados en retrotransposones LTR han demostrado ser mucho más informativos que otros tipos de marcadores moleculares, así Ellis *et al*. (1998), consiguieron obtener 281 marcadores mediante la técnica SSAP frente a 101 utilizando AFLPs, Kalendar *et al*. (1999) localizaron 6 marcadores a partir de dos ISSRs

frente a 36 marcadores con dos combinaciones de cebadores IRAP y 91 con cinco combinaciones de cebadores mediante la técnica REMAP, Tam *et al*. (2005) realizaron un estudio de diversidad genética entre el tomate y el pimiento utilizando SSAP, AFLP y SSR y obtuvieron un número de polimorfismos mucho más elevado en ambas especies mediante la técnica SSAP: entre 26 y 30 polimorfismos por pareja de cebadores frente a los 13 de los AFLPs y 2,4 de los SSRs; y finalmente Waugh *et al*. (1997) obtuvieron una proporción de productos polimórficos de alrededor del 0,1 con los AFLPs frente al 0,26 alcanzada con los SSAPs.

Estos marcadores también han demostrado ser más eficaces en lenteja ya que en este trabajo se han localizado 258 marcadores utilizando tres técnicas diferentes basadas en retrotransposones mientras que de la Puente García (2012) localizó 73 marcadores utilizando siete técnicas diferentes (SNP-CEL, CAPS, SSR, RAPD, SRAP, ALP y P/A) y Neyra (2014) localizó 16 marcadores utilizando CAPS, PCR y HRM.

4.3.1. SSAP (Sequence-Specific Amplification Polymorphism)

Se trata de una modificación de la técnica de AFLPs en la que se utilizan cebadores diseñados a partir de las LTRs de retrotransposones, combinados con cebadores complementarios a los adaptadores de enzimas de restricción.

A partir de cuatro de las LTRs secuenciadas, se han diseñado cebadores (Tabla 4.5) que cumplan las características adecuadas para llevar a cabo el protocolo de los SSAP (Syed y Flavell, 2006).

Tabla 4.5: Secuencia de los cebadores diseñados a partir de los LTRs descritos en este trabajo, en el genoma de *Lens culinaris* cv. Lupa.

Cebador	Secuencia
LTR1 (*Angela-like*)	5´-ATC ATG CCC CTT CGT AAG GAT CAC-3´
LTR2 (*Cassandra*)	5´-ATT TGG ATG GGT GAC CTT CTG GGA-3´
LTR3 (*Glycine*)	5´-GGC GCA ATG CAC TTT CTA GGT GTT-3´
LTR4 (*Peabody*)	5´-GTG TGA CAC CCT TGA ATT GCA TGT-3´

Para realizar la técnica SSAP, se digirió parcialmente el genoma utilizando dos enzimas de restricción: *MseI* y *PstI*. Tras la digestión, se procedió a la ligación de los adaptadores de ambas enzimas. A continuación se llevó a cabo una amplificación utilizando los cebadores complementarios a los adaptadores seguida de una segunda amplificación en la que se combinaron un cebador LTR con un cebador del adaptador. Por último, se realizó el análisis de los resultados mediante una electroforesis capilar.

Se analizaron los resultados obtenidos con los cuatro cebadores de las LTR combinadas con el cebador para el adaptador de *MseI* y con la combinación de cebadores LTR2-*PstI*. A partir de las cinco combinaciones analizadas se detectaron 100 polimorfismos de los cuales 46 presentaban una segregación 1:1, dos una segregación 3:1, que se corresponderían con dos bandas del mismo tamaño con segregación 1:1, y el resto no segregaba correctamente (Tabla 4.6: ver página 137).

Esta técnica se ha utilizado para desarrollar marcadores y realizar estudios de diversidad y evolución en gran cantidad de especies. En guisante, por ejemplo, se han desarrollado marcadores SSAP a partir del retrotransposón *PDR1* y se ha observado que este tipo de marcadores es más eficiente que los AFLPs a la hora de discriminar entre diferentes especies de *Pisum* (Ellis *et al*., 1998). Se han realizado estudios de diversidad basados en este tipo de marcadores entre especies del género *Vicia*, en las que se obtuvieron unos 25 marcadores por cada par de cebadores en las dos especies estudiadas (Martín Sanz *et al*., 2007) y se han realizado estudios de diversidad genética entre tomate y pimiento utilizando esta técnica junto con otros tipos de marcadores cuyos resultados se han comentado en el punto anterior (Tam *et al*, 2005).

4.3.2. REMAP (REtrotransposon-Microsatellite Amplified Polymorphisms)

Esta técnica consiste en una amplificación en la que se utiliza un cebador diseñado a partir de una LTR combinado con un microsatélite. Los cebadores de las LTRs utilizados son los mismos usados con los SSAP (Tabla 4.5).

Se probaron los cuatro cebadores diseñados con 35 cebadores obtenidos a partir de regiones flanqueantes de microsatélites de lenteja (Hamwieh *et al*., 2005) (Tabla 3.2: ver página 72) y con los microsatélites $(AC)_{10}$, $(GT)_{10}$, $(CT)_{10}$, $(GAA)_5$, $(GCA)_5$ y $(GACA)_4$ pero sólo se obtuvieron polimorfismos con tres de las combinaciones (LTR1-AC, LTR4-GT y LTR2-SR66R) el resto de cebadores no presentaron polimorfismos o presentaron problemas durante las amplificaciones que impidieron sus análisis. De estas tres combinaciones se obtuvieron diez bandas polimórficas de las cuales nueve presentaban una segregación 1:1, de manera que se pudieron incluir en el mapa genético (Tabla 4.6: ver página 137).

Se puede encontrar gran variedad de bibliografía basada en este tipo de marcadores. En ella se destaca el alto grado de polimorfismo observado con estos marcadores que alcanza el 97% en el caso del arroz (Branco *et al*., 2007), el 91% en trigo (Nasri *et al*., 2013) o el 90% en vid (Villano *et al*., 2014).

4.3.3. iPBS (inter-Primer Binding Site)

En este caso se utilizaron los cebadores diseñados para la amplificación y secuenciación de regiones LTR a partir de las secuencias PBS (Tabla 4.4) directamente como marcadores moleculares, ya que muchas de las bandas observadas en los geles presentaban polimorfismos entre los parentales.

Se realizaron amplificaciones con los doce cebadores y con todas las combinaciones posibles entre los mismos y, tras el análisis de los resultados, se comprobó que once de las combinaciones no presentaban diferencias entre los parentales mientras que en el resto se observaron 233 polimorfismos en la población RIL de los que 184 presentaban una segregación 1:1 y 15 se correspondían con segregaciones 3:1 (segregación 1:1 de dos bandas que presentan el mismo tamaño) (Tabla 4.6).

La técnica iPBS, desarrollada en un principio para secuenciar retrotransposones, ha demostrado su eficacia como una técnica para marcadores moleculares tanto en plantas como en animales ya que se obtienen patrones polimórficos que permiten la

discriminación tanto interespecífica (Kalendar *et al*., 2010) como intraespecífica, esta última demostrada entre diferentes especies del género *Lens* (Baloch *et al*., 2015).

Baloch *et al*. (2015) consiguieron 150 bandas polimórficas utilizando 10 marcadores PBS para estudiar 7 especies diferentes del género *Lens* lo que indica una media de 15 polimorfismos por pareja de cebadores, mientras que en el presente trabajo se ha obtenido una media de 3,5 estos resultados se pueden explicar por el hecho de que en este trabajo se han estudiado los descendientes de un único cruzamiento entre dos subespecies de una misma especie mientras que Baloch *et al*. compararon individuos de especies distintas por lo que las diferencias entre ellos deben ser mucho mayores.

A continuación se muestra una tabla que resume la cantidad de polimorfismos encontrados así como el tipo y la media de polimorfismos por pareja de cebadores. En ella se puede observar que la mayor cantidad de polimorfismos obtenidos se corresponden con la técnica iPBS, aunque la técnica SSAP es mucho más eficaz ya que se obtuvo un mayor número de marcadores por reacción de amplificación, 20 marcadores frente a los 3 obtenidos con REMAP y los 3,5 de los iPBS.

En cuanto a la segregación de los polimorfismos se puede ver que el 90% de los localizados mediante REMAP, el 85% de los iPBS y el 48% de los SSAP presentan segregación 1:1 o 3:1.

Tabla 4.6: Relación de polimorfismos en función de cada uno de los marcadores moleculares utilizados. También se muestra el tipo de segregación presente en cada uno de ellos.

Cebador	Número de polimorf.	Media de polimof. por reacción de ampl.	Segregación			% marcadores 1:1 o 3:1
			H_0(1:1)	H_0(3:1)	H_0 sig	
SSAP	100	20	46	2	52	48%
REMAP	10	3	9	0	1	90%
iPBS	233	3,5	184	15	34	85%

Las diferentes técnicas utilizadas presentan tanto ventajas como inconvenientes. Las dos técnicas no relacionadas con AFLPs (PBS y REMAP) son más sencillas y baratas, ya que el único equipamiento necesario para su realización y análisis es un termociclador y un sistema de electroforesis y detección para geles de agarosa, mientras que los SSAP precisan del uso de endonucleasas, adaptadores (lo que hace que los reactivos necesarios para esta técnica sean más costosos) y un sistema de electroforesis

capilar más un programa informático que permita el análisis de los resultados obtenidos. Otro de los factores a tener en cuenta es el tiempo empleado en la obtención de los marcadores, en el caso de los iPBS y los REMAP sólo es necesario realizar una PCR y analizar las bandas obtenidas en los geles mientras que en el caso de los SSAP es necesaria una reacción de restricción, una ligación de adaptadores y dos reacciones de PCR antes del análisis de la electroforesis y, además, de los polimorfismos obtenidos, sólo la mitad presenta una segregación apropiada para poder incluirlos como marcadores en un mapa genético.

Tabla 4.7: Polimorfismos obtenidos con los diferentes cebadores. También se muestra el tipo de segregación que presenta cada uno de ellos.

CEBADORES	BANDAS AMPLIF.	BANDAS POLIM.	SEGREGACÓN		
			H_o (1:1)	H_o (3:1)	H_o Sig
LTR1-MseI	25	12	**4**	**0**	8
LTR2-MseI	31	30	**17**	**2**	11
LTR3-MseI	13	12	**5**	**0**	7
LTR4-MseI	12	12	**6**	**0**	6
LTR2-PstI	50	34	**14**	**0**	20
LTR1-AC	5	3	**3**	**0**	0
LTR4-GT	5	4	**3**	**0**	1
LTR2-SR66R	4	3	**3**	**0**	0
PBS1 y 2	10*	2	**2**	**0**	0
PBS1 y 3	10*	4	**4**	**0**	0
PBS1 y 4	10*	3	**2**	**0**	1
PBS1 y 6	8*	2	**1**	**0**	1
PBS1 y 7	12*	2	**2**	**0**	0
PBS1 y 8	8	4	**3**	**0**	1
PBS1 y 9	12*	2	**1**	**0**	1
PBS1 y 10	13*	4	**4**	**0**	0
PBS1 y 11	7	1	**1**	**0**	0
PBS2	12	5	**5**	**0**	0
PBS2 y 3	16*	6	**6**	**0**	0
PBS2 y 4	7	1	**0**	**0**	1
PBS2 y 5	10	2	**1**	**1**	0
PBS2 y 6	12*	5	**4**	**1**	0
PBS2 y 7	9	2	**2**	**0**	0
PBS2 y 8	14	5	**4**	**1**	0
PBS2 y 9	8*	4	**3**	**0**	1
PBS2 y 10	15	8	**7**	**1**	0
PBS2 y 11	10	1	**1**	**0**	0
PBS2 y 12	9	2	**1**	**0**	1
PBS3	9	6	**6**	**0**	0
PBS3 y 4	9*	2	**1**	**0**	1
PBS3 y 5	15	6	**5**	**1**	0
PBS3 y 6	14	3	**3**	**0**	0
PBS3 y 7	18*	1	**1**	**0**	0
PBS3 y 8	16*	2	**2**	**0**	0
PBS3 y 11	6	4	**4**	**0**	0
PBS3 y 12	8*	2	**1**	**0**	1
PBS4 y 5	14*	3	**2**	**0**	1
PBS4 y 6	15	2	**2**	**0**	0

CEBADORES	BANDAS AMPLIF.	BANDAS POLIM.	SEGREGACÓN		
			H_o (1:1)	H_o (3:1)	H_o Sig
PBS4 y 7	16	2	**2**	**0**	0
PBS4 y 8	9	7	**5**	**0**	2
PBS4 y 9	10*	5	**3**	**1**	1
PBS4 y 10	9	4	**4**	**0**	0
PBS4 y 11	7*	4	**4**	**0**	0
PBS5	7	3	**2**	**0**	1
PBS5 y 6	14	3	**2**	**0**	1
PBS5 y 7	8*	3	**2**	**0**	1
PBS5 y 8	20*	8	**6**	**2**	0
PBS5 y 9	13	7	**3**	**1**	3
PBS5 y 10	13	3	**3**	**0**	0
PBS5 y 11	5*	1	**1**	**0**	0
PBS5 y 12	12*	1	**1**	**0**	0
PBS6	13*	4	**3**	**0**	1
PBS6 y 7	9	5	**3**	**1**	1
PBS6 y 8	19*	10	**3**	**2**	5
PBS6 y 9	14*	7	**6**	**1**	0
PBS6 y 10	10	3	**3**	**0**	0
PBS6 y 11	7	3	**2**	**0**	1
PBS6 y 12	15*	3	**3**	**0**	0
PBS7	6*	3	**3**	**0**	0
PBS7 y 8	15*	4	**3**	**0**	1
PBS7 y 9	7*	4	**3**	**0**	1
PBS7 y 10	10*	3	**1**	**1**	1
PBS8	13*	5	**4**	**0**	1
PBS8 y 9	13*	1	**1**	**0**	0
PBS8 y 11	13*	1	**1**	**0**	0
PBS8 y 12	9*	2	**2**	**0**	0
PBS9	7	3	**3**	**0**	0
PBS9 y 10	12*	4	**3**	**0**	1
PBS9 y 11	7	4	**4**	**0**	0
PBS9 y 12	13*	3	**3**	**0**	0
PBS10	8*	1	**1**	**0**	0
PBS10 y 11	12*	4	**3**	**0**	1
PBS10 y 12	12*	2	**2**	**0**	0
PBS11 y 12	8	2	**2**	**0**	0
PBS12	10	5	**3**	**1**	1
TOTAL	886	343	**239**	**17**	87

* En estos casos, el número de bandas amplificadas es aproximado ya que estos cebadores dan lugar a patrones muy complejos con muchas bandas y de tamaños similares, por lo que sólo se han tenido en cuenta aquéllas que se pudieran analizar claramente en toda la descendencia.

4.4. Localización de los marcadores en un mapa genético

El siguiente paso del estudio fue el desarrollo de un mapa genético de lenteja que incluyera todos los polimorfismos encontrados que segregaran correctamente. Para ello se utilizaron herramientas informáticas basadas en distintos métodos de análisis como MapMaker v 3.0b (Lander *et al*., 1987) y CarthaGene v 1.3 (de Givry *et al*., 2005) que permiten agrupar y ordenar los marcadores analizados en determinados grupos de ligamiento. Aunque los grupos que se obtienen con ambos programas son los mismos, el orden de los marcadores en los grupos grandes (> 50 marcadores) presenta menores valores de longitud y mayores valores de probabilidad al utilizar los algoritmos del programa CarthaGene, además, el programa MapMaker no permite analizar más de 100 marcadores en un mismo grupo. También se utilizó el programa MapChart v 2.2 (Voorrips, 2002) con el que se obtuvieron las representaciones gráficas de dichos grupos.

4.4.1. Mapa genético de marcadores derivados de secuenciación de retrotransposones

Se realizó un primer mapa genético que incluye los marcadores desarrollados en el presente estudio. Posteriormente, se construyó un segundo mapa en el que se añadieron otros marcadores analizados en el área de genética de la Universidad de León utilizando la misma población RIL.

Ambos mapas se obtuvieron con valores de LOD superiores a 4 entre parejas de marcadores, lo que indicaría que la probabilidad de que los marcadores contiguos en el mapa, estén ligados es 10000 veces superior a que se comporten como independientes.

En el primer estudio de ligamiento se incluyeron los 256 marcadores obtenidos mediante las técnicas SSAP, REMAP y PBS. De ellos, 86 se comportaron como independientes mientras que el resto, 170, pudieron localizarse en un mapa genético formado por 12 grupos de ligamiento que abarcaban 1970,5 cM lo que indica una distancia media entre marcadores adyacentes de 11,6 cM (Figura 4.21). Este mapa presenta tres grupos de ligamiento grandes, GL_R2 con 32 marcadores, GL_R3 (53

marcadores) y GL_R4 (34 marcadores), otros dos grupos intermedios GL_R1 y GL_R6 formados por 12 marcadores y otros siete grupos pequeños: GL_R5 que incluye seis marcadores, GL_R7, GL_R9, GL_R11 y GL_R12 formados por cuatro marcadores, GL_R8 con tres marcadores y, por último, GL_R10 que presenta únicamente dos marcadores (Tabla 4.8).

Tabla 4.8: Resumen del mapa genético basado en retrotransposones.

Grupo de ligamiento	Longitud (cM)	Número de marcadores (%)	Distancia media entre marcadores (cM)
GL_R1	86,8	12 (7,06)	7,23
GL_R2	332,0	32 (18,82)	10,38
GL_R3	744,3	53 (31,18)	14,04
GL_R4	456,9	34 (20,00)	13,44
GL_R5	57,8	6 (3,53)	6,63
GL_R6	125,1	12 (7,06)	10,43
GL_R7	37,3	4 (2,35)	9,32
GL_R8	19,4	3 (1,76)	6,47
GL_R9	47,3	4 (2,35)	11,83
GL_R10	9,2	2 (1,19)	4,60
GL_R11	45,0	4 (2,35)	11,25
GL_R12	48,2	4 (2,35)	12,05
Total	1970,5	170 (100)	11,60

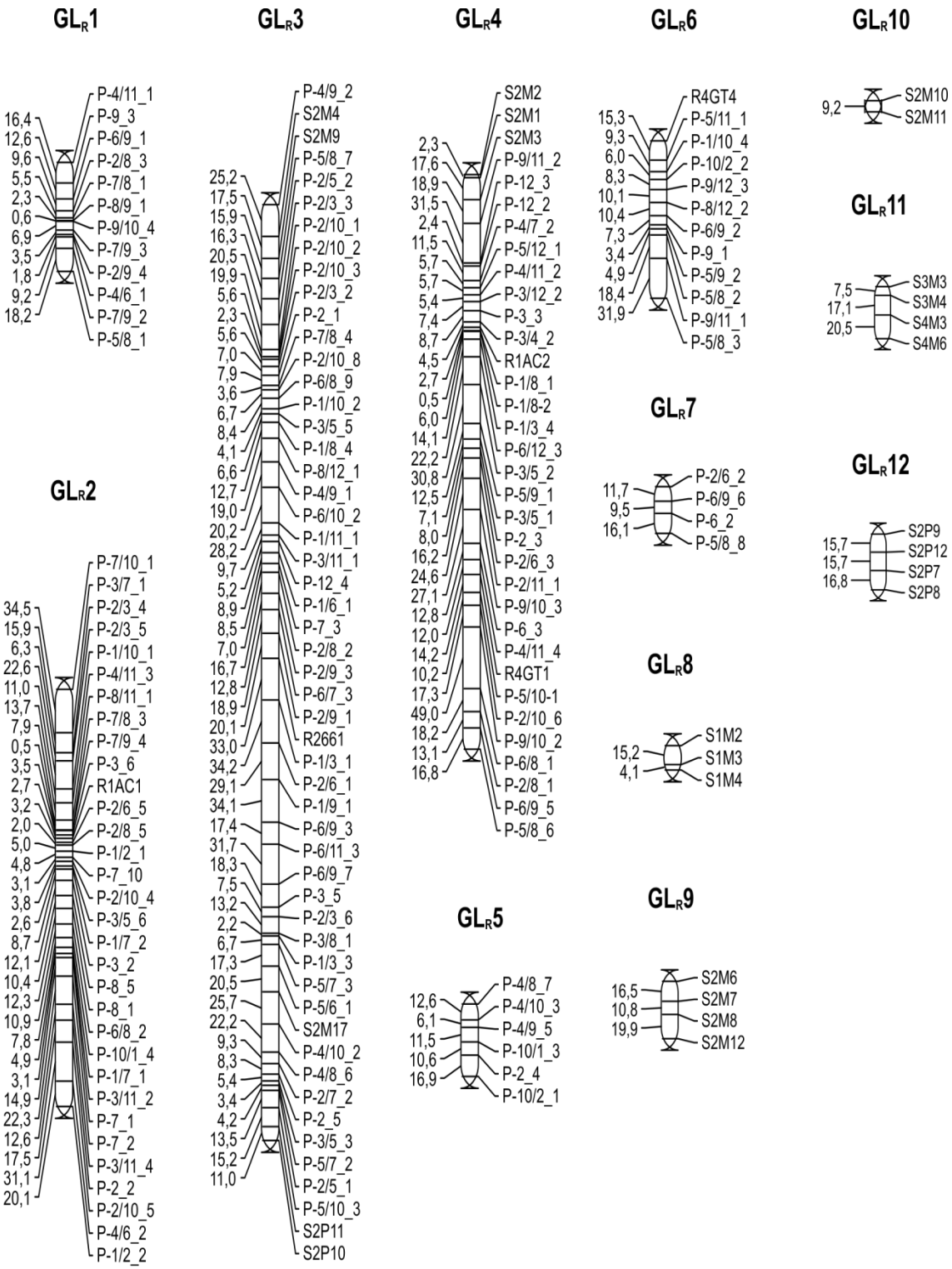


Figura 4.21: Mapa genético obtenido con los marcadores basados en retrotransposones. Estos marcadores se han desarrollado a partir de diferentes técnicas basadas en retrotransposones analizados en una población RIL F_7 pertenecientes al cruzamiento entre la forma cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880.

Como se puede observar, el mapa presenta cinco grupos (GL_R8-GL_R12) formados únicamente por marcadores del tipo SSAP. En la bibliografía consultada de otras especies vegetales no se han encontrado ejemplos de mapas obtenidos con diferentes tipos de marcadores en los que ocurra algo similar, aunque si se ha descrito un mapa genético en *Pennisetum esquamulatum* formado por 27 grupos de ligamiento utilizando 290 marcadores SSAP un SCAR y un AFLP, de manera que no se puede deducir si los SSAP se localizarían en pequeños grupos de ligamiento en presencia de otro tipo de marcadores (Huo *et al*., 2009). Sin embargo, la aparición de estos grupos de ligamiento se podría deber a la distribución de los retrotransposones en los cromosomas ya que lo hacen de una forma no aleatoria a lo largo del genoma (Peterson-Burch *et al*., 2004) y existen diferentes zonas intergénicas de preferencia para los elementos móviles (San Miguel *et al*., 1996; Walbot y Petrov, 2001). En los cromosomas se ha observado la ausencia, o la presencia pero en muy baja proporción, de retrotransposones en determinadas regiones como centrómeros, telómeros, heterocromatina o la región del organizador nucleolar (Schmidt *et al*., 1995; Brandes *et al*., 1997). Por esto pueden aparecer zonas en las que se encuentren agrupados varios retrotransposones lo suficientemente separadas del resto de marcadores como para no poder asociarlas con alguno de los grupos de ligamiento más grandes de manera que se formarían grupos independientes.

4.4.2. Mapa genético ampliado con otros marcadores

Para obtener un mapa más completo se han añadido otros 119 marcadores analizados en el laboratorio de Genética de la facultad de Biología de la Universidad de León y con la misma población RIL (de la Puente, 2012; Neyra, 2014). Este grupo de marcadores está formado por CAPs, ISSRs, SSRs, longitud de PCR específica, HMR (High Resolution Melt) y tres marcadores morfológicos.

De los 375 marcadores incluidos en este segundo mapa, 93 se comportan como independientes, incluyendo 80 de los 88 marcadores independientes del mapa anterior. En el nuevo mapa se incluyen 284 marcadores agrupados en 13 grupos de ligamiento (Figura 4.22 y Tabla 4.9). En este caso el mapa presenta una longitud de 2.657,7 cM y

una distancia media entre marcadores de 9,41 cM. Los grupos $GL_{RC}2$, $GL_{RC}6$, $GL_{RC}7$, $GL_{RC}8$, $GL_{RC}9$, $GL_{RC}10$; $GL_{RC}11$ y $GL_{RC}12$ se corresponden con los grupos homónimos del mapa anterior (GL_R), $GL_{RC}1$ (33 marcadores) está formado por la unión de los grupos GL_R1 y GL_R5, $GL_{RC}3$ (146 marcadores) está formado por GL_R3 y GL_R4 y aparecen tres grupos nuevos denominados $GL_{RC}13$ (cinco marcadores), $GL_{RC}14$ (tres marcadores) y $GL_{RC}15$ (dos marcadores) (Figura 4.23A y 4.23B).

Tabla 4.9: Resumen del mapa genético completo. Mapa obtenido al añadir otros tipos de marcadores.

Grupo de ligamiento	Longitud (cM)	Número de marcadores (%)	Distancia media entre marcadores (cM)
$GL_{RC}1$	245,5	33 (11,62)	7,44
$GL_{RC}2$	389,7	49 (17,37)	7,95
$GL_{RC}3$	1509,9	146 (51,77)	10,34
$GL_{RC}6$	177,9	18 (6,38)	7,95
$GL_{RC}7$	95,9	9(3,19)	10,65
$GL_{RC}8$	19,4	3 (1,06)	6,47
$GL_{RC}9$	47,3	4 (1,42)	11,83
$GL_{RC}10$	9,2	2 (0,71)	4,60
$GL_{RC}11$	45,0	4 (1,42)	11,25
$GL_{RC}12$	48,2	4 (1,42)	12,05
$GL_{RC}13$	33,2	5 (1,77)	6,64
$GL_{RC}14$	32,9	3 (1,06)	10,97
$GL_{RC}15$	0,6	2 (0,71)	0,30
Total	2654,7	282 (100)	9,41

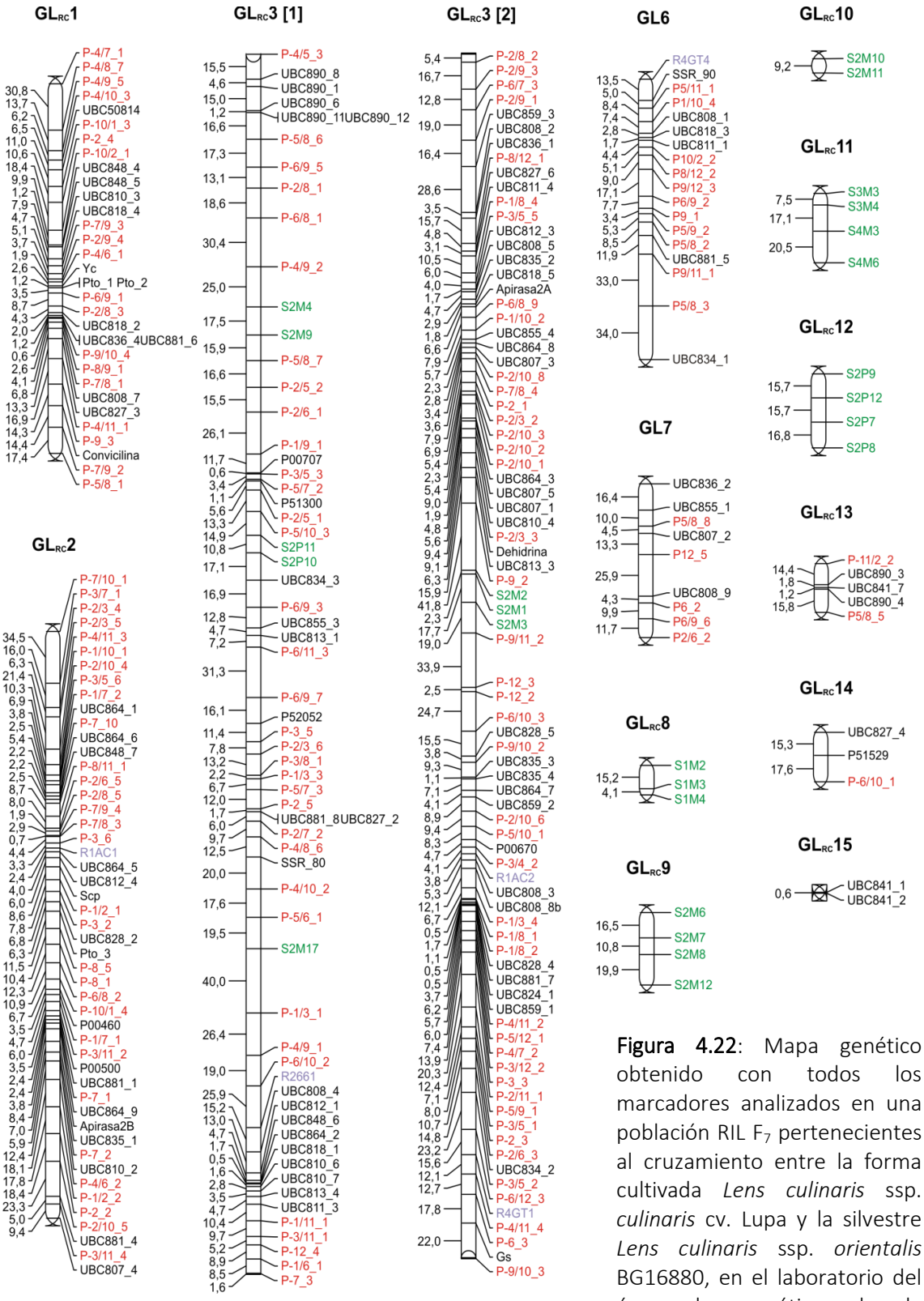


Figura 4.22: Mapa genético obtenido con todos los marcadores analizados en una población RIL F_7 pertenecientes al cruzamiento entre la forma cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880, en el laboratorio del área de genética de la Universidad de León. Los marcadores desarrollados en el presente trabajo aparecen con diferentes colores (SSAP, REMAP, PBS).

Al comparar los dos mapas, se puede destacar la reducción en la distancia media entre marcadores de 11,6 a 9,41 cM y la localización, en este segundo mapa, de ocho marcadores basados en retrotransposones que en el primer mapa se comportaban como independientes. Estos marcadores aparecen en $GL_{RC}1$ (P-4/7_1), en $GL_{RC}3$ (P4/5_3, P-9_2 y P-6/10_3), en $GL_{RC}7$ (P-12_5), en $GL_{RC}13$ (P-11/2_2 y P5/8_5) y el $GL_{RC}14$ (P6/10_1). En todos los casos, excepto en $GL_{RC}7$, se trata de grupos nuevos con respecto al mapa GL_R ya sea por la unión de diferentes grupos ($GL_{RC}1$ y $GL_{RC}3$) o porque sólo aparezcan en el segundo mapa ($GL_{RC}13$y $GL_{RC}14$).

En ambos mapas se puede observar que en algunos grupos aparecen regiones mucho más saturadas que otras, generalmente los marcadores aparecen más cercanos en las partes centrales y tienden a alejarse en los extremos. Esto puede tener un origen evolutivo ya que las diferencias de densidades de marcadores en diferentes zonas del genoma pueden deberse a distintos grados de conservación de las secuencias de ADN, los segmentos altamente saturados pueden coincidir con regiones del genoma con un alto grado de diferenciación entre los parentales y las zonas con menor densidad serían las más conservadas entre las especies del cruzamiento analizado (Vallejos *et al*., 1992). También puede ser debido a las diferencias en las frecuencias de recombinación de las diferentes zonas del cromosoma ya que en los centrómeros y en los telómeros la recombinación es 10 veces menor que en otras zonas del genoma por lo que en estas estructuras los marcadores estarían más cerca unos de otros (Tanksley *et al*., 1992).

En los grupos más grandes aparecen regiones con distancias entre marcadores adyacentes muy grandes llegando, en algunos casos, a superar los 30 cM, por lo que podrían resultar poco fiables (Figura 4.22).

En las siguientes Figuras (4.23A y 4.23B) se puede observar la formación de grupos nuevos en el segundo mapa (GL_{RC}) a partir de la unión de grupos del mapa anterior (GL_R).

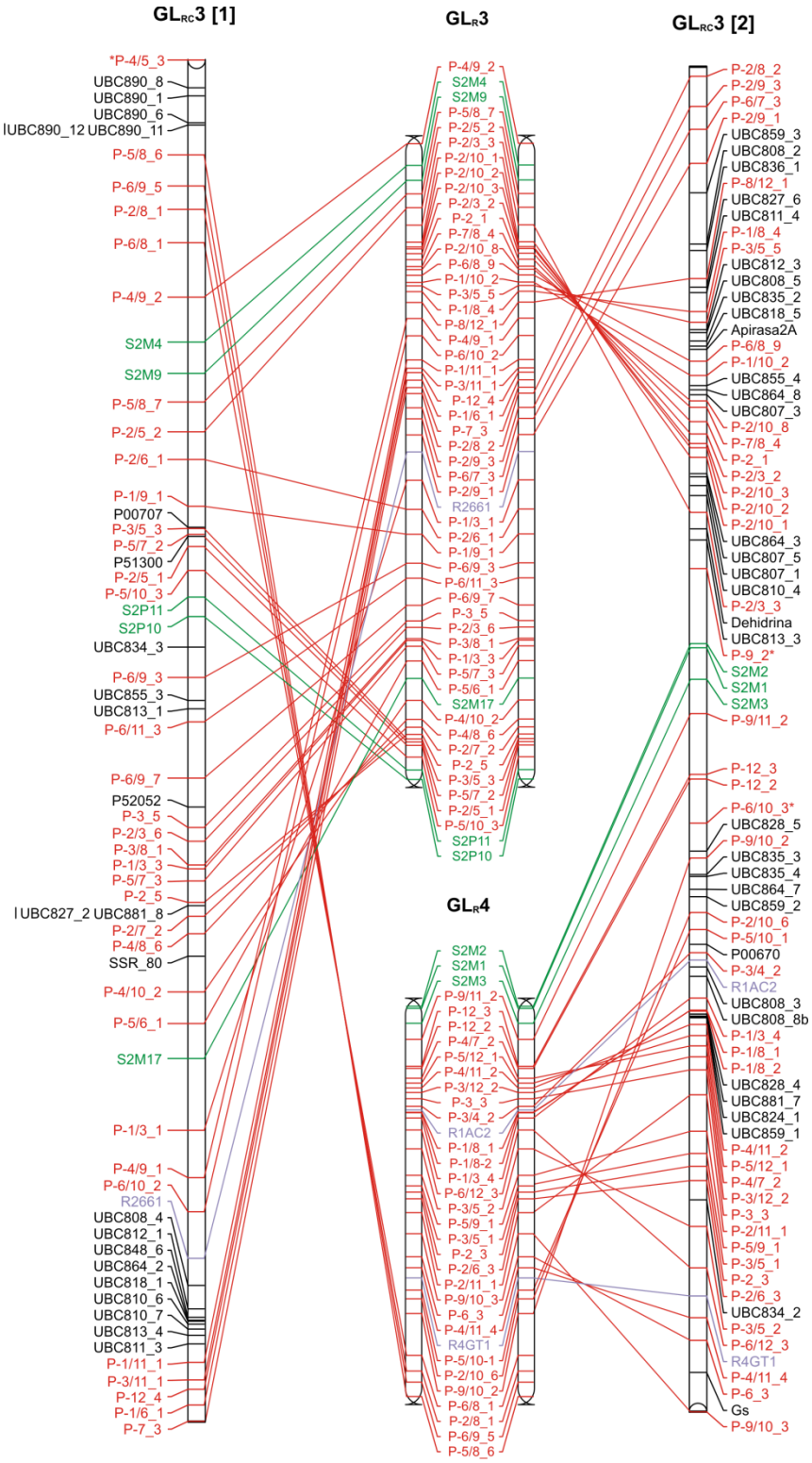


Figura 4.23A: Relación entre los grupos GL_R3 y GL_R4 con $GL_{RC}3$. En la figura se ve claramente como el segundo es el resultado de la unión de los dos primeros tras la incorporación de marcadores no basados en retrotransposones. Se han señalado con un asterisco los marcadores que aparecen como independientes en el primer mapa pero no en el segundo.

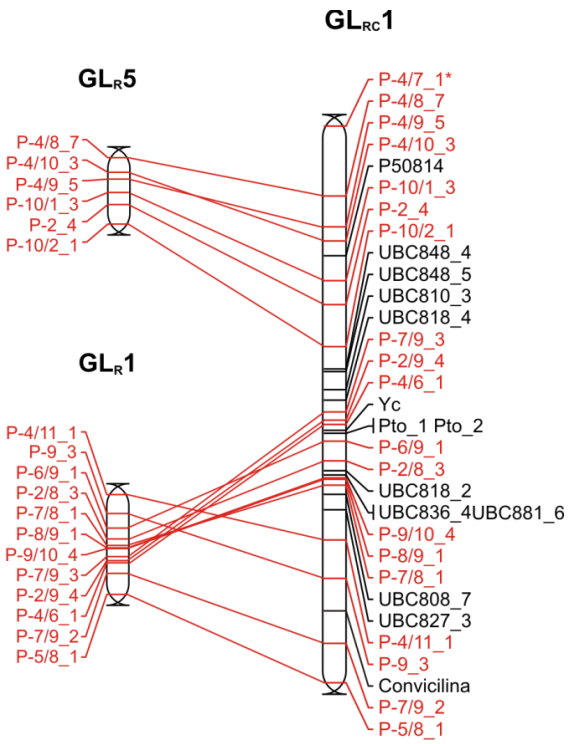


Figura 4.23B: Relación entre los grupos GL_R1 y GL_R5 con $GL_{RC}1$. En la figura se ve claramente como los grupos del segundo mapa (GL_{RC}) son el resultado de la unión de dos grupos del primero (GL_R) tras la incorporación de marcadores no basados en retrotransposones. Se han señalado con un asterisco los marcadores que aparecen como independientes en el primer mapa pero no en el segundo.

En las Figuras 4.24A, 4.24B y 4.24C se puede destacar la presencia de reordenamientos de los marcados dentro de los grupos de ligamiento $GL_{RC}1$, $GL_{RC}2$ y $GL_{RC}3$ del segundo mapa con respecto al orden de los mismos en el primer mapa. En ambos mapas aparecen agrupaciones internas determinadas de marcadores que varían de lugar entre un mapa y otro. Además, dentro de alguna de estas agrupaciones internas, los marcadores sufren cambios en la ordenación de los mismos.

En la Figura 4.24A se pueden ver los reordenamientos sufridos en las agrupaciones locales de marcadores que forman los grupos GL_R1y GL_R5 tras su unión para formar $GL_{RC}1$, la mayoría de estas agrupaciones locales sólo han sufrido cambios de posición manteniéndose el orden de los marcadores que las forman pero se puede

observar como en otras se han producido cambios internos en el orden de los marcadores.

En el caso GL_R2 y $GL_{RC}2$ (Figura 4.24B) no se ha producido ninguna unión de grupos de ligamiento sin embargo, los marcadores han cambiado de posición al añadir los no basados en retrotransposones y se puede destacar la desaparición de una posición del mapa con una distancia entre marcadores superior a 30 cM que aparece en el primer mapa pero no en el segundo.

En $GL_{RC}3$ (Figura 4.24C), cabe destacar la primera agrupación local de marcadores que proviene de GL_R4 mientras que el resto de agrupaciones locales que se encuentran a continuación provienen de GL_R3, esto se debe a que esta agrupación local de marcadores está separado del resto por distancias muy grandes (30,4 cM en el caso de $GL_{CR}3$ y 49 cM en GL_R4) por lo que el programa informático lo coloca de forma arbitraria en uno de los extremos del grupo de ligamiento correspondiente. La agrupación local de marcadores número 6 de $GL_{RC}3$ también se encuentra a una distancia grande del siguiente marcador (40 cM) mientras que en GL_C3 se encuentra muy cerca de los marcadores adyacentes.

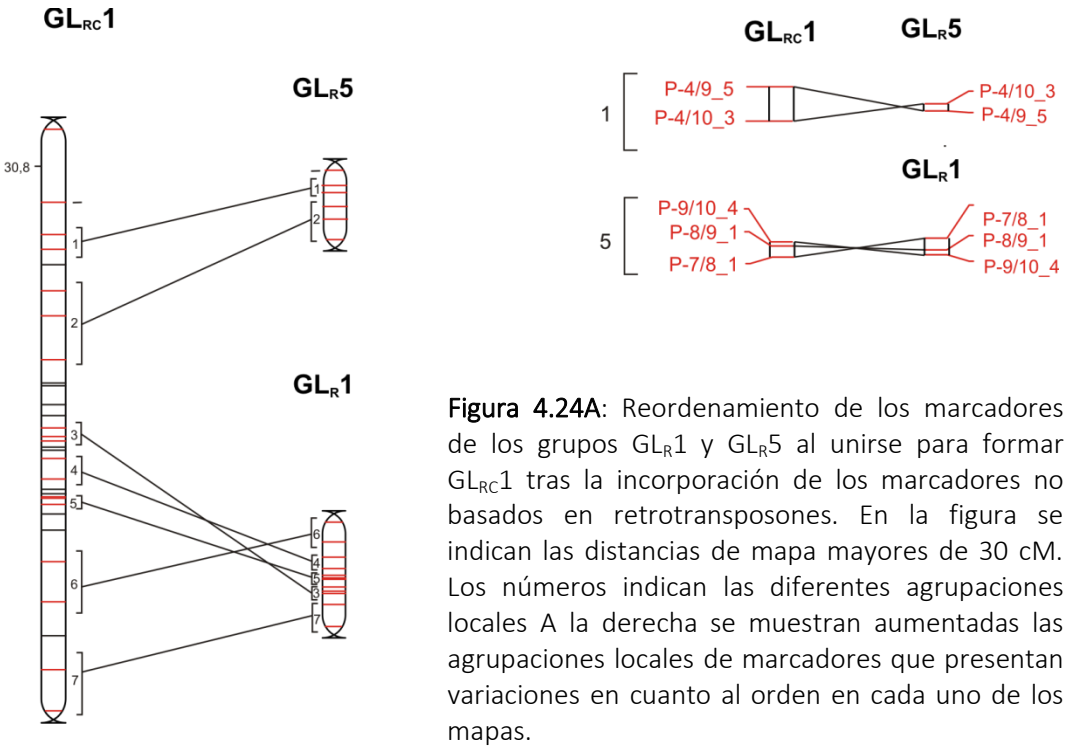


Figura 4.24A: Reordenamiento de los marcadores de los grupos GL_R1 y GL_R5 al unirse para formar $GL_{RC}1$ tras la incorporación de los marcadores no basados en retrotransposones. En la figura se indican las distancias de mapa mayores de 30 cM. Los números indican las diferentes agrupaciones locales A la derecha se muestran aumentadas las agrupaciones locales de marcadores que presentan variaciones en cuanto al orden en cada uno de los mapas.

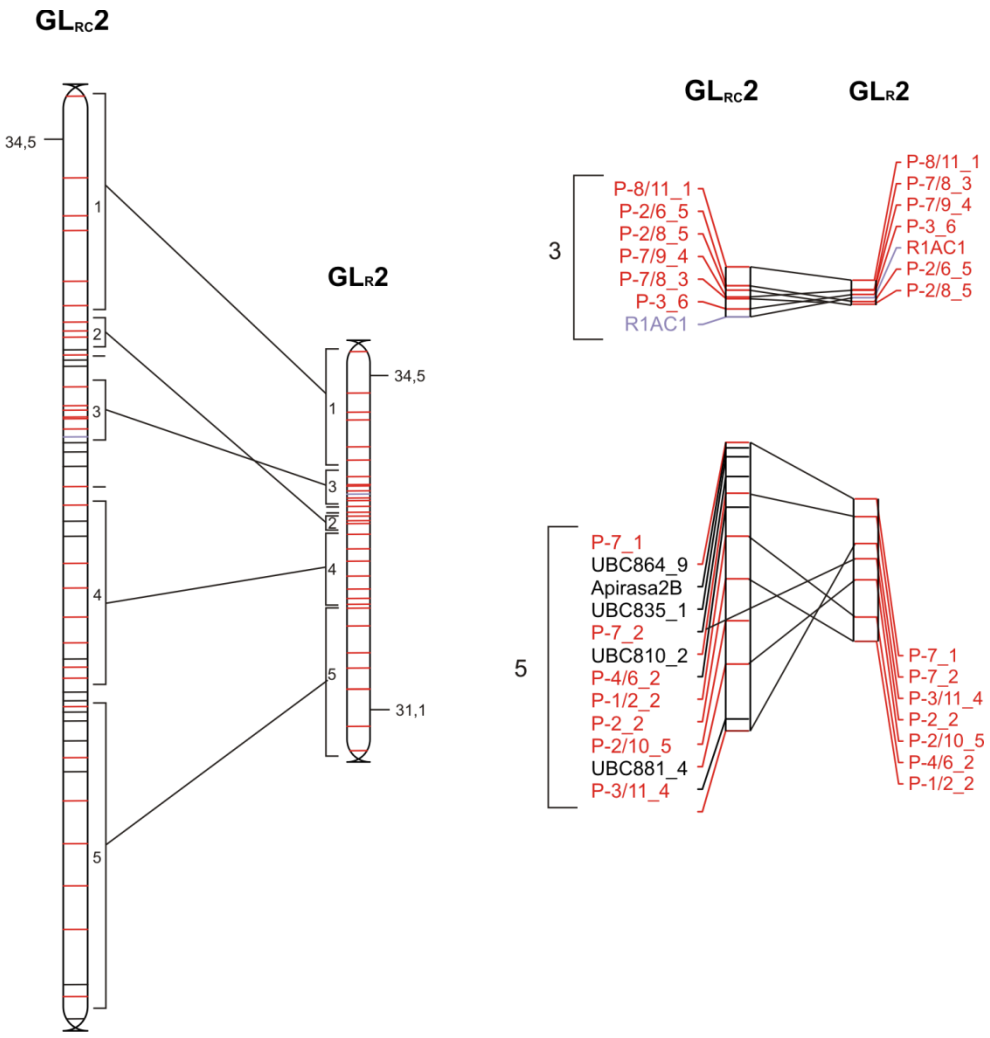


Figura 4.24B: Reordenamiento de los marcadores de los grupos GL_R2 y $GL_{RC}2$ tras la incorporación de los marcadores no basados en retrotransposones. En la figura se indican las distancias de mapa mayores de 30 cM. Los números indican las diferentes agrupaciones locales. A la derecha se muestran aumentadas las agrupaciones locales de marcadores que presentan variaciones en cuanto al orden en cada uno de los mapas.

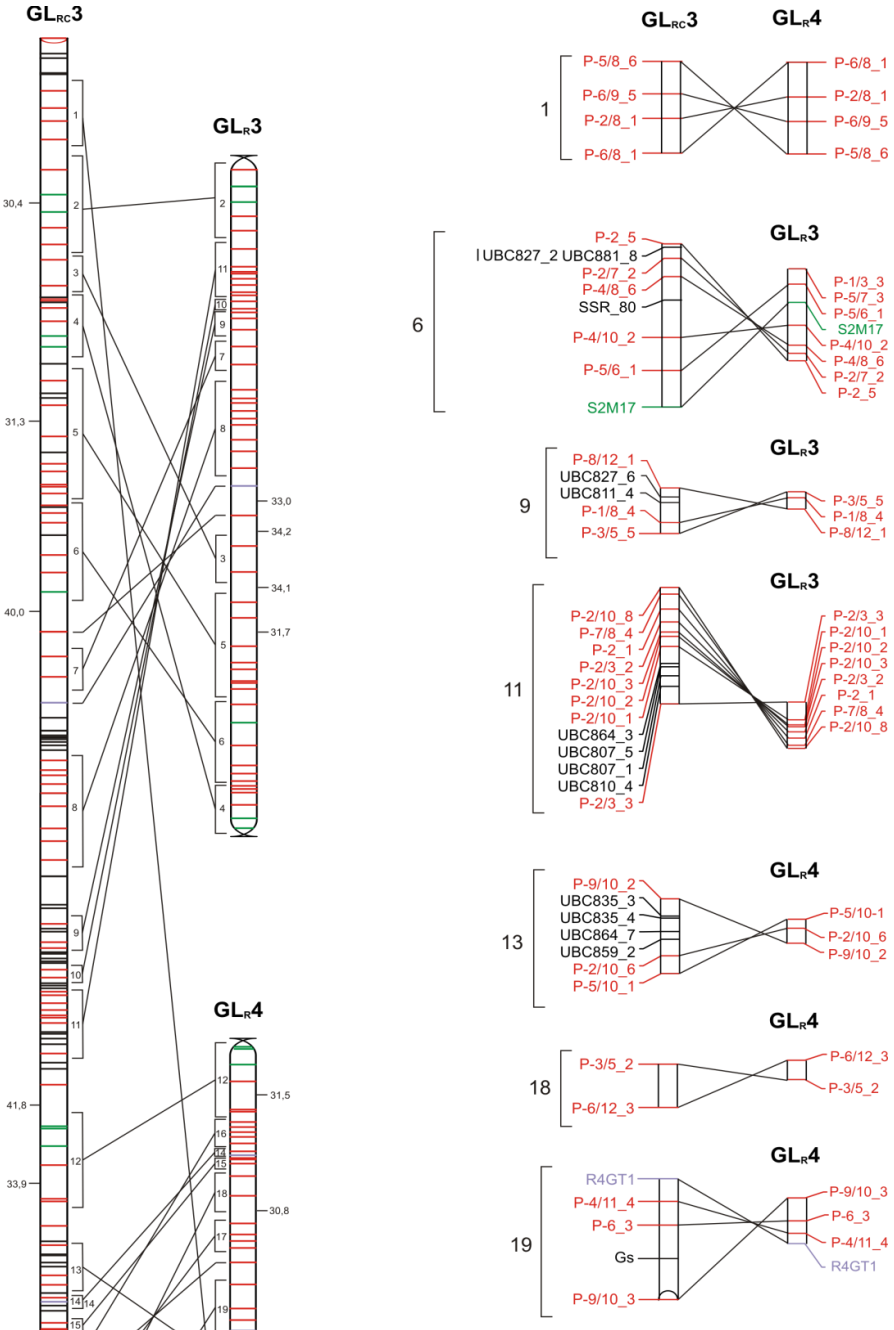


Figura 4.24C: Reordenamiento de los marcadores de los grupos GL_R3 y GL_R4 al unirse para formar $GL_{RC}3$ tras la incorporación de los marcadores no basados en retrotransposones. En la figura se indican las distancias de mapa mayores de 30 cM. Los números indican las diferentes agrupaciones locales. A la derecha se muestran aumentadas las agrupaciones locales de marcadores que presentan variaciones en cuanto al orden en cada uno de los mapas.

Desde el primer mapa realizado en lenteja en 1989 por Havey y Muehlbauer, se han obtenido numerosos mapas genéticos basados en diferentes tipos de marcadores tanto moleculares como morfológicos (Eujalyl *et al*., 1997 y 1998; Rubeena *et al*., 2003 y 2006; Durán *et al*., 2004; Kahraman *et al*., 2004; Phan *et al*., 2007; Tullu *et al*., 2008; Tanyolac *et al*., 2010; Saha *et al*., 2010; de la Puente, 2012; Sharpe *et al*., 2013; Fedoruk *et al*., 2013; Neyra, 2014; Yilmaz Temel *et al*., 2015; Andeden *et al*., 2015), pero nunca se habían utilizado marcadores basados en retrotransposones con este fin.

En este trabajo se han descrito trece grupos de ligamiento utilizando marcadores basados en retrotransposones, aunque cabe destacar la presencia de cinco grupos formados, únicamente, por marcadores SSAP (Figura 4.21) que no superan los cuatro marcadores por grupo. Esto ha hecho que no se pueda disminuir el número de grupos de ligamiento obtenidos en el último mapa descrito para la misma población RIL, nueve grupos (Neyra, 2014), aunque sí que se han producido uniones de grupos de dicho mapa en el actual (Figura 4.25). Por otro lado los grupos $GL_{RC}6$ y $GL_{RC}7$ están unidos en el mapa presentado por Neyra (2014) esto puede deberse a la diferencia en los parámetros utilizados a la hora de definir el mapa ya que Neyra (2014) utilizó un valor de LOD de tres y en el presente trabajo se ha utilizado el valor cuatro que es más restrictivo.

Figura 4.25: Relación entre el grupo GL$_{RC}$3 del mapa completo obtenido en este trabajo con los grupos GL2, GL3 y GL9 descritos por Neyra (2014). En la figura se ve claramente como GL$_{RC}$ es el resultado de la unión de los tres grupos de Neyra (2014).tras la incorporación de marcadores basados en retrotransposones.

El mapa obtenido en este trabajo no ha conseguido alcanzar los siete grupos de ligamiento correspondientes al número haploide de *Lens*, Sharpe *et al*. (2013) fue el primero en conseguirlo mediante el uso de marcadores SNP y SSR en una población RIL de *Lens culinaris*. Este mapa consta de 543 marcadores repartidos en siete grupos de ligamiento de entre 58 y 226 cM lo que indica un alto grado de saturación. La longitud total del mapa es de 834,7 cM. Se han encontrado relaciones de colinearidad en la posición de los marcadores genéticos entre *Medicago truncatula* y *Lens culinaris* aunque existen reordenamientos cromosómicos que son responsables de la diferencia en el número de cromosomas de ambas especies (Phan *et al*., 2007; Ellwood *et al*., 2008), esto permitió a Sharpe *et al*. (2013) numerar los grupos de ligamiento obtenido en lenteja de manera que se correspondieran, lo mejor posible, con los cromosomas descritos en *Medicago truncatula*.

Se ha realizado una revisión y ampliación del mapa de Sharpe (Fedoruk *et al*., 2013) en el que se añadieron al mapa otros 13 marcadores SNP, y cuatro morfológicos. De todos los marcadores incluidos en el mapa, tres han sido utilizados también para obtener el mapa más completo del presente trabajo (Yc, Scp y SSR80), de manera que permiten relacionar alguno de los grupos de ligamientos de dicho mapa con los obtenidos en los trabajos anteriormente citados y así situarlos en los cromosomas de lenteja. A continuación se muestra una tabla en la que se relacionan los grupos de ligamientos de los mapas descritos con los grupos del mapa de Fedoruk *et al*. (2013) y su equivalencia con los cromosomas de *Medicago truncatula*.

Yilmaz Temel *et al*. (2015) describió un mapa en lenteja en el que se incluyen 388 marcadores (376 SNP, tres SSR y nueve ISSR) localizados en nueve grupos de ligamiento, siete de ellos de gran tamaño que se podrían corresponder con los siete cromosomas de lenteja. La longitud que presenta es de 432,8 cM con una distancia media entre marcadores de 1,11 cM que es la más baja publicada hasta este momento. Presenta un marcador en común con el mapa GL_{RC} (UBC808_1) que se encuentra en el grupo de ligamiento $GL_{RC}6$ aunque éste está localizado en uno de los grupos más

pequeños (GL6) por lo que no se puede relacionar con ninguno de los posibles cromosomas de lenteja.

Tabla 4.10: Relación de los grupos de ligamiento de los mapas descritos en el presente trabajo con los grupos descrito por Sharpe *et al.* (2013) y su posible equivalencia con los cromosomas de *Medicago truncatula*.

Mapa basado en retrotransposones	Mapa con retrotransposones y otros marcadores	Mapa Fedoruk *et al*. (2013)	Cromosoma equivalente en *M. truncatula*
GL_R1	$GL_{RC}1$	GL1	1 o 5
GL_R2	$GL_{RC}2$	GL6	7
GL_R3	$GL_{RC}3$	GL2	2 o 6
GL_R4	$GL_{RC}3$	GL2	2 o 6
GL_R5	$GL_{RC}1$	GL1	1 o 5
GL_R6	$GL_{RC}6$	-	-
GL_R7	$GL_{RC}7$	-	-
GL_R8	$GL_{RC}8$	-	-
GL_R9	$GL_{RC}9$	-	-
GL_R10	$GL_{RC}10$	-	-
GL_R11	$GL_{RC}11$	-	-
GL_R12	$GL_{RC}12$	-	-
	$GL_{RC}13$		
	$GL_{RC}14$		
	$GL_{RC}15$		

El último mapa de lenteja publicado hasta la fecha es el de Andeden *et al*. (2015). Este mapa, basado en el uso de marcadores SSR analizados en una población F_2, incluye 47 marcadores localizados en siete grupos de ligamiento. Presenta una longitud de 303,9 cM y una distancia media entre marcadores de 7,06 cM.

Hasta el momento, los marcadores basados en retrotransposones se han utilizado, sobre todo, para realizar estudios de evolución y filogenia (Baloch *et al*., 2015; Branco *et al*., 2007; Ellis *et al*., 1998; Kalendar *et al*., 1999 y 2010; Nasri *et al*., 2013; Tam *et al*., 2005; Villano *et al*., 2014; Waugh *et al*., 1997) aunque también se han descrito mapas genéticos basados en este tipo de marcadores como es el caso de la avena (Yu y Wise, 2000), los géneros *Citrus* y *Poncirus* (Bernet y Asíns, 2003), el trigo (Queen *et al*., 2004), la lechuga (Syed *et al*., 2006) o la batata (Monden *et al*., 2015).

Un marcador ideal debería ser aquel que permita obtener un gran número de polimorfismos que estén distribuidos a lo largo de todo el genoma además de ser baratos y que se puedan analizar mediante técnicas sencillas y fácilmente reproducibles. Las técnicas REMAP e iPBS utilizadas en este trabajo cumplen todas estas características ya que su nivel de polimorfismo es más alto que el obtenido con los otros tipos de marcadores utilizados hasta el momento y las técnicas necesarias para su análisis son baratas y sencillas, los SSAP presentan algunos inconvenientes como el encarecimiento de los reactivos y el material de análisis necesario, sin embargo permite obtener un gran número de polimorfismos por cada reacción lo que disminuye el tiempo de trabajo considerablemente.

CONCLUSIONES

5. Conclusiones

1. Se ha conseguido aislar, secuenciar e identificar regiones conservadas de retrotransposones de los tipos *Ty1-copia* y *Ty3-gypsy* a partir de cebadores degenerados obtenidos del gen de la retrotranscriptasa y de otros cebadores diseñados a partir de secuencias de *Lens culinaris* que presentaban homología con retrotransposones. Se han podido definir siete tipos de retrotransposones *Ty1-copia* y tres tipos de *Ty3-gypsy*.

2. Se han secuenciado LTRs completas y fragmentos de éstas utilizando cebadores diseñados a partir de la RNasaH y de las regiones PBS. Las LTRs descritas pertenecen a diferentes familias de retrotransposones incluyendo *Ty1-copia* (*SIRE1-13* y *Angela*), *Ty3-gypsy* (*Ogre*, *Peabody*) y TRIM (*Cassandra*).

3. Se ha secuenciado un retrotransposón completo del tipo TRIM (*Cassandra*) en el genoma de *Lens culinaris*, comprobándose que carece de la maquinaria necesaria para su replicación.

4. Se han desarrollado 343 marcadores moleculares basados en retrotransposones, más concretamente en dos regiones: en la región LTR, utilizando las técnicas SSAP (Sequence-Specific Amplification Polymorphism) y REMAP (REtrotransposon-Microsatellite Amplified Polymorphisms), y en la región PBS mediante la técnica iPBS (inter-Primer Binding Site).

5. Se ha descrito un mapa en una población RIL F_7 del cruzamiento entre la forma cultivada *Lens culinaris* ssp. *culinaris* cv. Lupa y la silvestre *Lens culinaris* ssp. *orientalis* BG16880. Dicho mapa consta de 256 marcadores distribuidos en 12 grupos de ligamiento.

6. Se ha obtenido un segundo mapa genético en el que se han añadido otros 119 marcadores previamente analizados en la misma población RIL. Este mapa consta de 375 y 13 grupos de ligamiento.

7. Se han podido relacionar tres de los grupos de ligamiento obtenidos con tres de los grupos descritos por Fedoruk *et al.* (2013) que se corresponden con tres cromosomas de lenteja y que, a su vez, muestran sintenia con los cromosomas correspondientes de *Medicago truncatula*.

BIBLIOGRAFÍA

6. Bibliografía

Abe, H., Ohbayashi, F., Sugasaki, T., Kanehara, M., Terada, T., Shimada, T., Kawai, S., Mita, K., Kanamori, Y., Yamamoto, M.T., Oshiki, T. (2001). Two novel *Pao*-like retrotransposons (Kamikaze and *Yamato*) from the silkworm species *Bombyx mori* and *B. mandarina*: Common structural features of *Pao*-like elements. Mol. Genet. Genomics 265: 375-385.

Agarwal, M., Shrivastava, N., Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. Plant Cell. Rep. 27: 617-631.

Alefeld, F. (1866). Landwirtschaftliche Flora. Reprint 1966, ed. Otto Koeltz, Koenigstein-Taunus.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Andeden, E.E, Bachol, F.S., Çakir, E., Toklu, F., Özkan, H. (2015). Development, characterization and mapping of microsatellite markers for lentil (*Lens culinaris* Medik.). Plant Breed. 134: 589-598.

Antonius-Klemola, K., Kalendar, R., Schulman, A.H. (2006). TRIM retrotransposons occur in Apple and are polymorphic between varieties but not sports. Theor. Appl. Genet. 112: 999-1008.

Aranzana, M.J., Monforte, A.J., Viruel, M.A., Arús, P., Graziano, E., García Mas, J., Ballester, J., Oliver, M. (2000). Marcadores de AD. En: Nuez, F., Carrillo, J.M. Los marcadores genéticos en la mejora vegetal. Universidad Politécnica de Valencia.

Arumuganathan, K., Earle, E. (1991). Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208-218.

Avramova, Z., San Miguel, P., Georgieva, E., Bennetzen, J.L. (1995). Matrix attachment regions and transcribed sequences within a long chromosomal continium containing maize Adh1. Plant Cell 7: 1667-1680.

Baloch, F.S., Derya, M., Andeden, E.E., Alsaleh, A., Cömertpay, G., Kilian, B., Özkan, H. (2015). Inter-primer binding site retrotransposon and inter-simple sequence repeat diversity among wild *Lens* species. Biochem. Syst. Ecol. 58: 162-168.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226: 1209-1211.

Barulina, H. (1930). Lentils of the USSR and other countries. Bull. Appl. Bot. Genet. Pl. Breed. 40: 265-304.

Bernet, G.P., Asíns, M.J. (2003). Identification and genomic distribution of *gypsy* like retrotransposons in *Citrus* and *Poncirus*. Theor. Appl. Genet. 108: 121-130.

Bi, Y-A., Laten, H.M. (1996). Sequence analysis of a cDNA containing the *gag* and *prot* regions of the soybean retrovirus-like element, *SIRE-1*. Plant Mol. Biol. 30: 1315-1319.

Bingham, P.M., Zachar, Z. (1989). Retrotransposons and FB elements from *Drosophila melanogaster*. En Mobile DNA. Ed. DE Berg, M.M. Howe, pp. 485-502.

Boeke, J.D., Corces, V.G. (1989). Transcription and reverse-transcription of retrotransposons. Annu. Rev. Microbiol. 43: 403-434.

Branco, C.J.S., Vieira, E.A., Malone G., Kopp, M.M., Malone, E., Bernardes, A., Mistura, C.C., Carvalho, F.I.F., Oliveira, C.A. (2007). IRAP and REMAP assessments of genetic similarity in rice. J. Appl. Genet. 48: 107-113.

Brandes, A., Heslop-Harrison, J.S., Kamm, A., Kubis, S., Doudrick, R.L., Schmidt, T. (1997). Comparative analysis of the chromosomal and genomic organization of *Ty1-copia* like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant. Mol. Biol. 33: 11-21.

Brik, A.F., Kalendar, R.N., Statula, O.P., Sivolap, Iu.M. (2006). IRAP and REMAP analyses of barley varieties of *Odessa breeding*. Tsitol. Genet. 40: 24-33.

Broeck van den, D., Maes, T., Saber, M., Zethof, J., De Keukeleire, P., D'hauw, M., van Montagu, M., Gerats, T. (1998). Transposon Display identifies individual transposable elements in high copy number lines. Plant J. 13: 121-129.

Bureau, T.E, Wessler, S.R. (1992). Tourist: A large family of small inverted repeat elements frequently associated with maize gene. Plant Cell 4: 1283-1294.

Buxó, R. (1997). Arqueología de las plantas. Crítica, Barcelona, España. 368 pp.

Chadha, S., Gopalakrishna, T. (2005). Retrotranspson-microsatellite amplified polymorphism (REMAP) markers for genetic diversity assessment of the rice blast pathogen (*Magnaporthe grisea*). Genome 48: 943-945.

Castro, I., D'Onofrio, C., Martín., J.P., Ortiz, J.M., De Lorenzis, G., Ferreira, V., Pinto-Carnide, O. (2012). Effectiveness of AFLPs and retrotransposon-based markers for the identification of Portuguese grapevine cultivars and clones. Mol. Biotechnol. 52: 26-39.

Ceccarelli, M., Sarri, V., Polizzi, E., Andreozzi, G., Cionini, P.G. (2010). *Ogre* retrotrasposons in *Lathyrus* species. Accesión Gene Bank: FR717828

Chavanne, F., Zhang, D.X., Liaud, M.F., Cerff, R. (1998). Structure and evolution of *Cyclops:* A novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. Plant Mol. Biol. 37: 363-375.

Clarke, G.C.S., Kupicha, F.K. (1976). The relatioships of the genus *Cicer* L. (Leguminoseae): the evidence from pollen morphology. Bot. J. Lin. Soc. 72: 35-44.

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the *Arabidopsis thaliana* 5S rRNA genes. Plant J. 35: 251-261.

Cubero, J.I. (1981). Origin, domestication and evolution. En: Webb, C. and Hawtin, G.C. (eds) Lentils. Commonwealth Agricultural Bureau, Slough, UK, pp. 15-38.

Cubero, J.I. (2003). Introducción a la mejora genética vegetal. 2ª Edición. Mundi-Prensa. Madrid.

Cubero, J.I., Pérez de la Vega, M., Fratini, R. (2009). Origin, Phylogeny, domestication and spread. En: Erskine, W., Muehlbauer, F.J., Sarker, A. y Sharma, B. (eds) The Lentil: Botany, Production and Uses. CABI Press, Wallingford, UK, pp 13-33.

Devos, K.M., Brown, J.K.M., Bennetzen, J.L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. 12: 1075-1079.

Ding, D., Lipshitz, H.D. (1994). Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. Gent. Res. 64: 167-181.

Dixit, A., Ma, K., Yu, J., Cho, E., Park, Y. (2006). Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: Sequence characterization and phylogenetic analysis. Plant Cell. Rep. 25: 100-111.

Doolittle, R.F., Feng, D.F., Johnson, M.S., McClure, M.A. (1989). Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. 64: 1-30.

Durán, Y., Fratini, R., García, P., Pérez de la Vega, M. (2004). An intersubspecific genetic map of *Lens*. Theor. Appl. Genet. 108: 1265-1273.

Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V., Ambrose, M.J. (1998) Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol. Gen. Genet. 260: 9-19.

Ellwood, S.R., Phan, H.T., Jordan, M., Hane, J., Torres, A.M., Avila, C.M., Cruz-Izquierdo, S., Oliver, R.P. (2008). Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genomics 9: 380-391.

Erskine, W., Adham, Y., Holly, L. (1989). Geographic distribution of variation in quantitative traits in a world lentil collection. Euphytica 43: 97-103.

Erskine, W., Ellis, R.H., Summerfield, R.J., Roberts, E.H., Hussain, A. (1990). Characterization of responses to temperature and photoperiod for time to flowering in a world lentil collection. Theor. Appl. Genet. 80: 193-199.

Eujayl, I., Baum, M., Erskine, W., Pehu, E., Muehlbauer, F.J. (1997). The use of RAPD markers for lentil genetic mapping and the evaluation of distorted F_2 segregation. Euphytica 96: 405-412.

Eujayl, I., Baum, M., Powell, W., Erskine, W., Pehu, E. (1998). A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. Theor. Appl. Genet. 97: 83-89.

Fablet, M., Rebollo, R., Biémont, C., Vieira, C. (2007). The evolution of retrotransposon regulatory regions and its consequences on the *Drosophila melanogaster* and *Homo sapiens* host genomes. Gene 390: 84-91.

FAOSTAT. 2015. http://faostat.fao.org/.

Fawcett, D.H., Lister, C.K., Kellett, E., Finnegan, D. J. (1986). Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian *LINEs*. Cell 47: 1007-1015.

Fedoroff, N. (2000). Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. U.S.A. 97: 7002-7007.

Fedoruk, M.J., Vandenberg, A., Bett, K.B. (2013). Quantitative Trait Loci analysis of seed quality characteristics in Lentil (*Lens culinaris* ssp. *culinaris* Medik.) using SNP markers. The Plant Genome 6: 1-10.

Ferguson, M.E., Maxted, N., Van Slageren, M., Robertson, L.D. (2000). A reassessment of the taxonomy of *Lens* Miller (Leguminosae, Papilionoideae, Vicieae). Bot. J. Lin. Soc. 133: 41-59.

Ferguson, M.E., Newbury, H.J., Mated, N., Ford-Lloyd, B.V., Robertson, L.D. (1998). Population genetic structure in *Lens* taxa revealed by isozyme and RAPD analysis. Genet. Res. Crop Evol. 45: 549-559.

Flavell, R.B., Bennet, M.D., Smith, J.B., Smith, D.B. (1974). Genome size and proportion of repeated nucleotide sequence DNA in plants. Biochem. Genet. 12: 257-269.

Flavell, A.J., Bolshakov, V.N., Booth, A., Jing, R., Russell, J., Ellis, T.H., Isaac, P. (2003). A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. Nucleic Acids Res. 31: e115.

Flavell, A.J., Knox, M.R., Pearce, S.R., Ellis, T.H. (1998). Retrotransposon based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J. 16: 643-650.

Flavell, A.J., Smith, J.B., Kumar, A. (1992). Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol. Gen. Genet. 231: 233-242.

Flavell, A.J., Pearce, S.R., Heslop-Harrison, L., Kumar, A. (1997). The evolution of *Ty1-copia* group retrotransposons in eukaryote genomes. Genetica 100: 185-195.

Forbes, E.M., Nieduszynska, S.R., Brunton, F.K., Gibson, J., Glover L.A., Stansfield, I. (2007). Control of *gag-pol* gene expression in the *Candida albicans* retrotransposon *Tca2*. Mol. Biol. 8: 94-107.

Fratini, R., Ruiz, M.L. (2006). Interspecific hybridization in the genus *Lens* applying in vitro embryo rescue. Euphytica 150: 271-280.

Fratini, R., Ruiz, M.L., Pérez de la Vega, M. (2004). Intra-specific and inter-sub-specific crossing in lentil (*Lens culinaris* Medik.). Can. J. Plant Sci. 84: 981-986.

Friesen, N., Brandes, A., Heslop-Harrison, J.S. (2001). Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. Mol. Biol. Evol. 18: 1176-1188.

Galasso, I. (2003). Distribution of highly repeated DNA sequences in species of the genus *Lens* Miller. Genome 46: 1118-1124.

Galeano, C., Fernández, A., Gómez, M., Blair, M. (2009). Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (*Phaseolus vulgaris* L.). BMC Genomics 10: 629.

Galindo, L.M., Gaitán-Solís, E., Baccam, P., Tohme, J. (2004). Isolation and characterization of RNase LTR sequences of *Ty1-copia* retrotransposons in common bean (*Phaseolus vulgaris* L.). Genome 47: 84-95.

Gao, X., Havecker, E.R., Baranov, P.V., Atkins, J.F, Voytas, D.F. (2003). Translational recoding signals between retrotransposons in diverse LTR *pol* and gag. RNA. 9: 1422-1430.

Givry de, S., Bouchez, M., Chabrier, P., Milan, D., Schiex, T. (2005). CARTHAGENE: Multipopulation integrated genetic and radiated hybrid mapping. Bioinformatics. 21: 1703-1704.

Godwin, I.D., Aitken, E.A., Smith, L.W. (1997). Application of inter simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis. 18: 1524-1528.

Goedken, E. R., Marqusee, S. (2001). Co-crystal of *Escherichia coli* RNaseHI with Mn^{2+} ions reveals two divalent metals bound in the active site. J. Biol. Chem. 276: 7266-7271.

Gonçalves, I., Duret, L., Mouchiroud, D. (2000). Nature and structure of human genes that generate retropseudogenes. Genome Res. 10: 672-678.

Goodenow, M.M., Bloom, G., Rose, S.L., Pomeroy, S.M., O'Brien, P.O., Perez, E.E., Sleasman, J.W., Dunn, B.M. (2002). Naturally occurring amino acid polymorphisms in human immunodeficiency virus type 1 (HIV-1) Gag p7 (NC) and the C-cleavage site impact Gag-Pol processing by HIV-1 protease. Virology 292:137-149.

Grandbastien, M.A. (1998). Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3: 181-187.

Grandbastien, M.A., Lucas, H., Morel, J-B., Mhiri, C., Vernhettes, S., Casacuberta, J.M. (1997). The expression of the tobacco *Tnt1* retrotransposon is linked to the plant defense responses. Genetica 100: 241-252.

Grandbastien, M.A., Spielmann, A., Caboche, C. (1989). *Tnt1*, a mobile retroviral transposable element of tobacco isolated by plant cell genetics. Nature 337: 376-380.

Gribbon, B.M., Pearce, S.R., Kalendar, R., Schulman, A.H., Paulin, L., Jack, P., Kumar, A., Flavell, A.J. (1999). Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. Mol. Gen. Genet. 261: 883-891.

Gunn, C.R. (1969). Genera, types and lectotypes in the tribe Vicieae (Fabaceae). Taxon 18: 725-733.

Gunn, C.R., Kluve, J. (1976.) Androecium and pistil characters for Tribe Vicieae (Fabaceae). Taxon 25: 563-575.

Haldane, J.B.S. (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. J .Genet. 8: 299-309.

Hamon, P., Duroy, P.O., Dubreuil-Tranchant, C., Mafra D'Almeida Costa, P., Duret, C., Razafinarivo, N.J., Couturon, E., Hamon, S., de Kochko, A., Poncet, V., Guyot, R. (2011). Two novel *Ty1-copia* retrotransposons isolated from coffee trees can effectively reveal evolutionary relationships in the *Coffea* genus (Rubiaceae). Mol. Genet. Genomics 285: 447-460.

Hamwieh, A., Udupa, S.M., Choumane, W., Sarker, A., Dreyer, F., Jung, C., Baum M. (2005). A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. Theor. Appl. Genet. 110: 669-677.

Harlan, J. (1992). Crops and man. 2ª edición. Am. Soc. Agron., Madison, Wisconsin, USA.

Harlan, J.R., De Wet, J.M.J. (1971). Towards a rational classification of cultivated plants. Taxon 20: 509-517.

Hansen, C., Heslop-Harrison, J.S. (2004). Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. Adv. Bot. Res. 41: 165-193.

Havecker, E.R., Gao, X., Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biol. 5: 225.

Havey, M.J., Muehlbauer, F.J. (1989). Linkages between restriction fragment length, isozyme, and morphological markers in lentil. Theor. Appl. Genet. 77: 395-401.

Hirochika, H., Fukuchi, A., Kikuchi, F. (1992). Retrotransposon families in rice. Mol. Gen. Genet. 233: 209-216

Hirochika, F., Hirochika, R. (1993). *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn. J. Genet. 68: 35-46.

Hirochika, H., Okamoto, H., Kakutani, T. (2000). Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. Plant Cell 12: 357-368.

Hirochika, H., Otsuki, H., Yoshikawa, M., Otsuki, Y., Sugimoto, K., Takeda, S. (1996A). Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. Plant Cell 8: 725-734.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., Kanda, M. (1996B). Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA. 93: 7783-7788.

Huang, C.R.L., Burns, K.H., Boeke, J.D. (2012). Active transposition in genomes. Annu. Rev. Genetic. 46: 651-675.

Huang, H., Chopra, R., Verdine, G.L., Harrison, S.C. (1998). Structure of a catalytic complex of Hiv-1 reverse transcriptase: implications for nucleoside analog drug resistance. Science 282: 1669-1675.

Hunter, R.L., Markert, C.L. (1957). Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125: 1294-1295.

Huo, H., Joann A. Conner, J.A., Ozias-Akins, P. (2009). Genetic mapping of the apospory-specific genomic region in *Pennisetum squamulatum* using retrotransposon-based molecular markers. Theor. Appl. Genet. 119:199-212.

Inoue, H., Nojima, H., Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene 96: 23-28.

Jaskolski, M., Alexandratos, J.N., Bujacz, G., Wlodawer, A. (2009). Piecing together the structure of retroviral integrase, an important target in AIDS therapy. FEBS J. 276: 2926-2946.

Jeffreys. A. (1979). DNA sequence variants in the Gγ-, Aγ-, δ and β-globin genes of man. Cell 18: 1-10.

Jeffreys, A., Wilson, V., Thein, S.L. (1985). Hypervariable "minisatellite" regions in human DNA. Nature 314: 67-73.

Jing, R., Bolshakov, V., Flavell, A.J. (2007). The tagged microarray marker (TAM) method for high-throughput detection of single nucleotide and indel polymorphisms. Nat. Protoc. 2: 168-177.

Jing, R., Knox, M.R., Lee, J.M., Vershinin, A.V., Ambrose, M.J., Ellis, T.H.N., Flavell, A.J. (2005). Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species. Genetics 171: 741-752.

Jing, R., Vershinin, A., Grzebyta, J., Shaw, P., Smýkal, P., Marshall, D., Ambrose, M.J., Ellis, T.H.N., Flavell, A.J. (2010). The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evol. Biol. 10: 44.

Johnson, P.C., Webster, L.M., Adam, A., Buckland, R., Dawson, D.A., Keller, L.F. (2006). Abundant variation in microsatellites of the parasitic nematode *Trichostrongylus tenuis* and linkage to a tandem repeat. Mol. Biochem. Parasitol. 148: 210-218.

Kahraman, A., Kusmenoglu, I., Aydin, N., Aydogan, A., Erskine, W., Muehlbauer, F. (2004). QTL mapping of winter hardiness genes in lentil. Crop. Sci. 44:13-22.

Kalendar, R., Antonius, K., Smýkal, P., Schulman, A.H. (2010). iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor. Appl. Genet. 121: 1419-1430.

Kalendar, R., Flavell, A.J., Ellis, T.H.N., Sjakste, T., Moisy, C., Schulman, A.H. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106: 520-530.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A.H. (1999). IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. Theor. Appl. Genet. 98: 704-711.

Kalendar, R., Schulman, A.H. (2006). IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat. Protoc. 1: 2478-2484.

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., Schulman A.H. (2008). *Cassandra* retrotransposons carry independently transcribed 5S RNA. Proc. Natl. Acad. Sci. 105: 5833-5838.

Kalendar, R., Vicient, C.M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A., Schulman, A.H. (2004). Large retrotransposon derivatives: abundant, conserved, but nonautonomous retroelements of barley and related genomes. Genetics 166: 1437-1450.

Kazazian, H.H.Jr., (2004). Mobile elements: drivers of genome evolution. Science 303:1626-1632.

Kim, H., Terakami, S., Nishitani, C., Kurita, K., Kanamori, H., Katayose, Y., Sawamura, Y., Saito, T., Yamamoto, T. (2012). Development of cultivar-specific DNA markers based on retrotransposon based insertional polymorphism in Japanese pear. Breed Sci. 62: 53-62.

Koike, M., Kawaura, K., Ogihara, Y., Torada, A. (2006). Isolation and characterization of SSR sequences from the genome and TAC clones of common wheat using the PCR technique. Genome 49: 432-444.

Kolano, B., Bednara, E., Weiss-Schneeweiss, H. (2013). Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of *Chenopodium quinoa* (Amaranthaceae). Plant Cell Rep. 32: 1575-1588.

Konieczny, A., Ausubel, F.M. (1993). Procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4: 403-410.

Kosambi, D.D. (1944). The estimation of map distances from recombination values. Ann. Eugenic. 12: 172-175.

Kumar, A., Bennetzen, J.L. (1999). Plant retrotransposons. Annu. Rev. Genet. 33: 479-532.

Ladizinsky, G. (1979). The origin of lentil and its wild gene pool. Euphytica 28: 179-187.

Ladizinsky, G. (1993). Wild lentils. Crit. Rev. Plant Sci. 12: 169-184.

Ladizinsky, G. (1997). A new species of *Lens* from south-east Turkey. Bot. J. Lin. Soc. 123: 257-260.

Ladizinsky, G. (1999). Identification of the lentil wild genetic stock. Genet. Resour. Crop Evol. 46: 115-118.

Ladizinsky, G., Braun, D., Goshen, D., Muehlbauer, F.J. (1984). The biological species of the genus *Lens*. Bot. Gaz. 145: 253-261.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Laten, H.M., Majumdar, A., Gaucher, E.A. (1998). SIRE-1, a *copia/Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc. Natl. Acad. Sci. USA 95: 6897-6902.

Laten, H.M., Havecker, E.R., Farmer, L.M., Voytas, D.F. (2003). *SIRE1*, an endogenous retrovirus family from *Glycine max*, is highly homogeneous and evolutionarily young. Mol. Biol. Evol. 20: 1222-1230.

Law, J.A., Jacobsen. S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11: 204-220.

Leeton, P.R.J., Smith, D.R. (1993). An abundant *LINE*-like element amplified in the genome of *Lilium speciosum*. Mol. Gen. Genet. 237: 97-107.

Lev-Yadun, S., Gopher, A., Abbo, S. (2000). The cradle of agriculture. Science 288: 1602-1603.

Lewin, B. (1997). Genes VI. Oxford University Press. Oxford. pp: 597-619.

Li. G., Quiros, C.F. (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theor. Appl. Genet. 103: 455-461.

Llorente, A. (2001). Análisis de genes de leguminosas en el género *Lens*. Universidad de León. Tesis doctoral.

Lu, Q., Teare, J.M., Granok, H., Swede, M.J., Xu, J., Elgin, S.C.R. (2003). The capacity to form H-DNA cannot substitute for GAGA factor binding to a (CT)n (GA)n regulatory site. Nucl. Acid Res. 31: 2483-2494.

Macas, J., Neumann, P. (2007). *Ogre* elements - A distinct group of plant *Ty3/gypsy*-like retrotransposons. Gene 390: 108-116.

Macas, J., Neumann, P., Navrátilová, A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8: 427-443.

Malik, H.S. (2005). Ribonuclease H evolution in retrotransposable elements. Cytogenet. Genome Res. 110: 392-401.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chistas, F., Derbyshire, M. K., De Weese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N.R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchel, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., Bryant, S. H. (2010). CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39: 225-229.

Martín Sanz, A., Gilsanz Gonzalez, S., Syed, N.H., Suso, M.J., Caminero Saldaña, C., Flavell, A.J. (2007). Genetic diversity analysis in *Vicia* species using retrotransposon-based SSAP markers. Mol. Genet. Genomics 278: 433-441.

Mayer, M.S., Bagga, S.K. (2002). The phylogeny of *Lens* (Leguminosae): new insight from ITS sequence analysis. Plant. Syst. Evol. 232: 145-154.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36: 344-355.

Meglecz, E., Anderson, S.J., Bourguet, D., Butcher, R., Caldas, A., Cassel-Lundhagen, A., d'Acier, A.C., Dawson, D.A., Faure, N., Fauvelot, C., Franck, P., Harper, G., Keyghobadi, N., Kluetsch, C., Muthulaksshmi, M., Nagaraju, J., Patt, A., Péténiean, F., Silvain, J.F., Wilcock, H.R. (2007). Microsatellite flanking region similarities among different loci within insect species. Insect Mol. Biol. 16: 175-185

Michel, F., Lang, B. F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature 316: 641-643.

Miller, J.H. (1972). Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Monden, Y., Hara, T., Okada, Y., Jahana, O., Kobayashi, A., Tabuchi, H., Onaga, S., Tahara, M. (2015). Construction of a linkage map based on retrotransposon insertion polymorphisms in sweetpotato via high-throughput sequencing. Breeding Sci. 65: 145-153.

Moore, G., Lucas, H., Batty, N., Flavell, R. (1991). A family of retrotransposons and associated genomic variation in wheat. Genomics 10: 461-468.

Muehlbauer, F.J., Weeden, N.F., Hoffman, D.L. (1989). Inheritance and linkage relationships of morphological and isozyme loci in lentil (*Lens* Miller). J. Hered. 80: 298-303.

Mullis, K.B., Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155: 335-350.

Muthukumar, B., Bennetzen, J.L. (2004). Isolation and characterization of genomic and transcribed retrotransposon sequences from sorghum. Mol. Gen. Genomics 271: 308-316.

Nair, G.R, Dash, C., Le Grice, S.F.J., DeStefano, J.J. (2012). Viral reverse transcriptases show selective high affinity binding to DNA-DNA primer-templates that resemble the polypurine tract. PLOS One 7: e41712.

Nasri, S., Mandoulakani, B.A., Darvishzadeh, R., Bernousi, I. (2013). Retrotransposon insertional polymorphism in Iranian bread wheat cultivars and breeding lines revealed by IRAP and REMAP markers. Biochem. Genet. 51:927-943.

Neumann, P., Koblízková, A., Navrátilová, A., Macas, J. (2006). Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 93: 315-320.

Neumann, P., Pozaárková, D., Macas, J. (2003). Highly abundant pea LTR retrotransposon *Ogre* is constitutively transcribed and partially spliced. Plant Mol. Biol. 53: 399-410.

Neuvéglise, C., Feldmann, H., Bon, E., Gaillardin, C., Casaregola, S. (2002). Genomic evolution of the long terminal repeat retrotransposons in Hemiascomycetous yeasts. Genome Res. 12: 930-43.

Neyra, C. D. (2014). Estudio de genes codificantes de proteínas de reserva, mapeo genético, análisis de QTLs y estudio de genes de resistencia a *Ascochyta lentis* en *Lens*. Universidad de León. Tesis doctoral.

Novikov, A., Smyshlyaev, G., Novikova, O. (2012). Evolutionary history of LTR retrotransposon chromodomains in plants. Int. J. Plant Genomics 2012: 874743.

Nowotny, M., Gaidamakov, S. A., Crouch, R.J., Yang, W. (2005). Crystal structures of RNaseH bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121: 1005-1016.

Nowotny, M., Gaidamakov, S.A., Ghirlando, R., Cerritelli, S.M., Crouch, R.J. Yang, W. (2007). Structure of human RNase H1 complexed with an RNA/DNA hybrid: Insight into HIV reverse transcription. Mol. Cell 28: 264-276.

Oda, Y., Yoshida, M., Kanaya, S. (1993). Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli*. J. Biol. Chem. 268: 88-92.

Otshima, K., Okada, N. (2005). SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. Cytogenet. Genome Res. 110: 475-490.

Ohtsubo, H., Kumekawa, N., Ohtsubo, E. (1999). RIRE2, a novel *gypsy*-type retrotransposon from rice. Genes Genet. Syst. 74: 83-91.

Okada, N. (1991). SINEs. Curr. Opin. Genet. Dev. 1: 498-504.

Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K., Yeung, A.T. (1998). Mutation detection using a novel plant endonuclease. Nucleic Acids Res. 26: 4597-4602.

Pearce, S.R. (2007). SIRE-1, a putative plant retrovirus is closely related to a legume *Ty1-copia* retrotransposon family. Cell. Mol. Biol. Lett. 12: 120-126.

Pearce, S.R., Harrison, G., Heslop-Harrison, J.S., Flavell, A.J., Kumar, A. (1997). Characterization and genomic organization of *Ty1-copia* group retrotransposons in rye (*Secale cereale*). Genome. 40: 617-625.

Pearce, S.R., Knox, M., Ellis, T.H.N., Flavell, A.J., Kumar, A. (2000). Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol. Gen. Genet. 263: 898-907.

Pearce, S.R., Kumar, A., Flavell, A.J. (1996). Activation of the *Ty1-copia* group retrotransposons of potato (*Solanum tuberosum*) during protoplast isolation. Plant Cell Rep. 15: 949-953.

Pearce, S., Stuart-Rogers, C., Knox, M.R., Kumar, A., Ellis, T.H.N., Flavel, A.J. (1999). Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. Plant J. 19: 711-717.

Pelsy, F. (2007). Untranslated leader region polymorphism of *Tvv1*, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*. Theor. Appl. Genet. 116: 15-27.

Peredo, E.L., Arroyo-García, R, Revilla, M.A. (2009). Epigenetic changes detected in micropropagated hop plants. J. Plant Physiol. 166: 1101-1111.

Pereira, H.S., Barão, A., Delgado, M., Morais-Cecílio, L., Viegas, W. (2005). Genomic analysis of Grapevine Retrotransposon 1 (*Gret 1*) in *Vitis vinifera*. Theor. Appl. Genet. 111: 871-878.

Pérez de la Vega, M., Fratini, R., Muehlbauer, F.J. (2011). Lentil (*Lens culinaris* Medik.). En: Pérez de la Vega, M., Torres, A.M., Cubero, J.I. y Kole, C. (eds). Genetics, genomics and breeding of cool season grain legumes. 1ª edición. Enfield: Science Publisher, pp 98-150.

Petit, J., Bourgeois, E., Stenger, W., Bès, M., Droc, G., Meynard, D., Courtois, B., Ghesquière, A., Sabot, F., Panaud, O., Guiderdoni, E. (2009). Diversity of the *Ty-1 copia* retrotransposon *Tos17* in rice (*Oryza sativa* L.) and the AA genome of the *Oryza* genus. Mol. Genet. Genomics 282: 633-652.

Petterson, D.S., Sipsas, S., Mackintosh, J.B. (1997). The chemical composition and nutritive value of Australian pulses. 2nd ed. Grain Research & Development Corporation, Camberra, Australia, p 65.

Peterson-Burch, B., Nettleton, D., Voytas, D.F. (2004). Genomic neighborhoods for *Arabidopsis* retrotransposons: a role for targeted integration in the distribution of the Metaviridae. Genome Biol. 5: R78.

Peterson-Burch, B.D., Voytas, D.F. (2002). Genes of the pseudoviridae (*Ty1-copia* retrotransposons). Mol. Biol. Evol. 19: 1832-1845.

Peterson-Burch, B.D., Wright, D.A., Laten, H.M., Voytas, D.F. (2000). Retroviruses in plants? Trends Genet. 16: 151-152.

Phan, H.T.T., Ellwood, S.R., Hane, J.K., Ford, R., Materne, M., Oliver, R.P. (2007). Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris*. Theor. Appl. Genet. 114: 549-558.

Porceddu, A., Albertini, E., Barcaccia, G., Marconi, G., Bertoli, F.B., Veronesi, F. (2002). Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol. Genet. Genomics 267: 107-114.

Pouteau, S., Huttner, E., Grandbastien, M.A., Caboche, M. (1991). Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. EMBO J. 10: 1911-1918.

Preston, B.D. (1996). Error-prone retrotransposition: rime of the ancient mutators. Proc. Natl. Acad. Sci. USA 93: 7427-7431.

Promega Corporation (2007). Technical manual. pGem®-T and pGem®-T easy vector uses. Madison. USA. Revised 5/07.

Puente de la, R. (2012). Mapa genético y análisis de secuencias nucleotídicas de interés en *Lens*. Universidad de León. Tesis doctoral.

Qi, X., Vargas, E., Larsen, L., Knapp, W., Hatfield, G.W., Lathrop, R., Sandmeyer, S. (2013). Directed DNA shuffling of retrovirus and retrotransposon integrase protein domains. PLOS One. 8: e63957.

Queen, R.A., Gribbon, B.M., James, C., Jack, P., Flavell, A.J. (2004). Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol. Genet. Genomics 271: 91-97.

Rajput, M.K., Upadhyaya, K.C. CARE1, a *Ty3-gypsy* like LTR-retrotransposon in the food legume chickpea (*Cicer arietinum* L.). (2009). Genetica 136: 429–437.

Rashkova, S., Karam, S.E., Pardue, M.L. (2002). Element-specific localization of *Drosophila* retrotransposon Gag proteins occurs in both nucleus and cytoplasm. Proc. Natl. Acad. Sci. 99: 8457-8460.

Rico, L. (2005). Nuevos retrotransposones de naranjo. Desarrollo de marcadores basados en retrotransposones para la evaluación de la biodiversidad de los cítricos. Universidad de Barcelona. Tesis doctoral.

Rubeena, Ford, R., Taylor, P.W. (2003). Construction of an intraspecific linkage map of lentil (*Lens culinaris* ssp. *culinaris*). Theor. Appl. Genet. 107: 910-916.

Rubeena, Taylor, P.W.J., Ades, P.K., Ford, R. (2006). QTL mapping of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight (*Ascochyta lentis*). Plant Breed. 125:506-512.

Sabot, F., Schulman, A.H. (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97: 381-388.

Saha, G.C. Sarker, A., Chen, W., Vandemark, G.J., Muehlbauer, F.J. (2010). Inheritance and linkage map positions of genes conferring resistance to *Stemphylium blight* in lentil. Crop. Sci. 50: 1831-1839.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., Yuki, S. (1984). Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature 312: 659-661.

Saitou, N., Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol. Biol. Evol. 4: 406-425.

SanMiguel, P., Bennetzen, J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. 82: 37-44

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat. Genet. 20: 43-45.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765-768.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schmidt, T., Kubis, S., Heslop-Harrison, J.S. (1995). Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.) *LINEs* and *Tyl-copia*-like elements as major components of the genome. Chromosome Res. 3: 335-345.

Schulman, A.H. (2007). Molecular markers to assess genetic diversity. Euphytica 158: 313-321.

Schulman, A.H., Flavell, A.J., Ellis, T.H.N. (2004). The application of LTR retrotransposons as genetic markers in plants. En: Miller, W., Capy, P. (eds) Mobile genetic elements: protocols and genomic applications. Humana Press Inc., Totowa.

Sharpe, A.G., Ramsay, L., Sanderson, L.A., Fedoruck, M.J., Clarke, W.E., Li, R., Kagale, S., Vijayan, P., Vandenberg, A., Bett, K.E. (2013). Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genomics. 14: 192-204.

Sinjushin, A.A., Akopian, J.A. (2011). On seedling structure in *Pisum* L., *Lathyrus* L. and *Vavilovia* Fed. (Fabeae: Fabaceae). Wulfenia 18: 81-93.

Sobrino, B., Briona, M., Carracedoa, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci. Int. 154: 181-194.

Sonnante, G., Galasso, I., Pignone, D. (2003). ITS sequence analysis and phylogenetic inference in the genus *Lens* Mill. Ann. Bot. 91: 49-54.

Smýkal P. (2006). Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. J. Appl. Genet. 47: 221-30.

Smýkal, P., Horacek, J., Dostalova, R., Hybl, M. (2008A). Variety discrimination in pea (*Pisum sativum* L.) by molecular, biochemical and morphological markers. J. Appl. Genet. 49: 155-66.

Smýkal, P., Hýbl, M., Corander, J., Jarkovský, J., Flavell, A.J., Griga, M. (2008B). Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. Theor. Appl. Genet. 117: 413-424.

Smýkal, P., Kalendar, R., Ford, R., Macas, J., Griga, M. (2009). Evolutionary conserved lineage of *Angela*-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity 103: 157-167.

Su, P.Y., Brown, T.A. (1997). *Ty3/gypsy*-like retrotransposon sequences in tomato. Plasmid 38: 148-157.

Suoniemi, A., Anammthawat-Johnsson, K., Arna, T., Schulman, A.H. (1996A). Retrotransposon *BARE-1* is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol. Biol. 30: 1321-1329.

Suoniemi, A., Narvanto, A., Schulman, A.H. (1996B). The *BARE-1* retrotransposon is transcribed in barley from an LTR promotor active in transient assays. Plant Mol. Biol. 31: 295-306.

Suoniemi, A., Tanskanen, J. Schulman, A.H. (1998). Gypsy-like retrotransposons are widespread in the plant kingdom. Plant J. 13: 699-705.

Syed, N.H., Flavell, A.J. (2006). Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. Nat. Protoc. 1: 2746-2752.

Syed, N.H, Sørensen, A.P., Antonise, R., van de Wiel, C., van der Linden C.G., van´t Westende, W., Hooftman, D.A.P., den Nijs, H.C.M., Flavell, A.J. (2006). A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. Theor. Appl. Genet. 112: 517-527.

Syed, N.H., Sureshsundar, S., Wilkinson, M.J., Bhau, B.S., Cavalcanti, J.J.V., Flavell, A.J. (2005). *Ty1-copia* retrotransposon-based SSAP marker development in Cashew (*Anacardium occidentale* L.) Theor. Appl. Genet. 110: 1195-1202.

Tadmor, Y., Zamir, D., Ladizinsky, G. (1987). Genetic mapping of an ancient translocation in the genus *Lens*. Theor. Appl. Genet. 73: 883-892.

Tahir, M., Muehlbauer, F.J. (1994). Gene mapping in lentil with recombinant inbred lines. J. Hered. 85: 306-310.

Tahir, M., Simon, C.J., Muehlbauer, F.J. (1993). Gene map of lentil: a review. *Lens* Newslet. 20: 3-10.

Takeda, S., Sugimoto, K., Otsuki, H., Hirochika, H. (1998). Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol. Biol. 36: 365-376.

Tam, S.M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S.R., Grandbastien, M. (2005). Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor. Appl. Genet. 110: 819-831.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol. Biol. Evol. 9: 678-687.

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.

Tan, Y.D., Fu, Y.X. (2006). A novel method for estimating linkage maps. Genetics 173: 2283-2290.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Brown, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M.S., Wing, R.A., Wu, W., Young, N.D. (1992). High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141-1160.

Tanyolac, B., Ozatay, S., Kahraman, A., Muehlbauer, F. (2010). Linkage mapping of lentil (*Lens culinaris*) genome using recombinant inbred lines revealed by AFLP, ISSR, RAPD and some morphologic markers. J. Agric. Biotech. Sust. Develop. 2: 1-6.

Tautz, D., Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl. Acids Res. 12: 4127-4138.

Temin, H.M., Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226: 1211-1213.

Tero, N., Neumeier, H., Gudavalli, R., Schlötterer, C. (2006). *Silene tatarica* microsatellites are frequently located in repetitive DNA. J. Evol. Biol. 19: 1612-1619.

This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., Crespan, M., Dangl, G.S., Eisenheld, C., Ferreira-Monteiro, F., Grando, S., Ibáñez, J., Lacombe, T., Laucou, V., Magalhães, R., Meredith, C.P., Milani, N., Peterlunger, E., Regner, F., Zulini, L., Maul, E. (2004). Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor. Appl. Genet. 109: 1448-1458.

Thompsom, J.D., Higgins, D.G., Gibson, T.J. (1984). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Tisdale, M., Schulze, T., Larder, B.A., Moelling, K. (1991). Mutations within the RNaseH domain of human immunodeficiency virus type 1 reverse transcriptase abolish virus infectivity. J. Gen. Virol. 72: 59-66.

Todorovska, E. (2007). Retrotransposons and their role in plant genome evolution. Biotechnol. & Biotechnol. Eq. 21: 294-305.

Toh, H., Hayashida, H., Miyata, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature 305: 827-829.

Tullu, A., Taran, B., Warkentin, T., Vandenberg, A. (2008). Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. Crop. Sci. 48: 2254-2264.

Turcich, M.P., Bokhari-Riza, A., Hamilton, D.A., He, C., Messier, W., Stewart, C.B., Mascarenhas, J.P. (1996) *PREM-2*, a *copia*-type retroelement in maize is expressed preferentially in early microspores. Sex. Plant Reprod. 9: 65-74.

Urbano, G., Porres, J.M., Frías J., Vidal-Valverde, C. (2007). Nutricional value. En: Yadav, S.S., McNeil, D., Stevenson, P.C. (eds) Lentil. An ancient crop for modern times. Springer, Dordrecht, The Netherlands. pp 47–93.

Vaillancourt, R.E., Slinkard, A.E. (1993). Linkage of morphological and isozyme loci in lentil, *Lens culinaris*. Can. J. Plant Sci. 73: 917-926.

Vallejos, C.E., Sakiyama, N.S., Christine, D.C. (1992). A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131: 733-740.

Van der Linden, C.G., Wouters, D.C., Mihalka, V., Kochieva, E.Z., Smulders, M.J., Vosman, B. (2004). Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet. 109: 384-393.

Van Ooijen, J.W., Voorrips, R.E. (2001). JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Van Oss, H., Aron, Y., Ladizinsky, G. (1997). Chloroplast DNA variation and evolution in the genus *Lens* Mill. Theor. Appl. Genet. 94: 452-457.

Vanin, E.F. (1985). Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. 19: 253-272.

Varagona, M.J., Purugganan, M., Wessler, S.R. (1992). Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. Plant Cell 4: 811-820.

Vershinin, A.V., Ellis, T.H.N. (1999). Heterogeneity of the internal structure of *PDR1*, a family of *Ty1/copia* - like retrotransposons in pea. Mol. Gen. Genet. 262: 703-713.

Vicient, C.M., Kalendar, R., Schulman, A.H. (2001). Envelope-Class Retrovirus-Like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res. 11: 2041-2049.

Villano, C., Carputo, D., Frusciante, L., Santoro X., Aversano, R. (2014). Use of SSR and retrotransposon-based markers to interpret the population structure of native grapevines from southern Italy. J. Appl. Genet. 48: 107-113.

Vitte, C., Ishii, T., Lamy, F., Brar, D., Panaud, O. (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol. Genet. Genomics 272: 504-511.

Volman, G.P., Shilo, J., Weinhouse, H., Ohana, P., Benziman, M. (1998). *Peabody*, a new *Ty3-type* retrotransposon in *Pisum sativum*. Accesión Gene Bank: AF083074.

Voorrips, R.E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93: 77-78.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic. Acids Res. 23: 4407-4414.

Voytas, D. F., Cummings, M. P., Konieczny, A., Ausubel, F. M., Rodermel S. (1992). *Copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA 89: 7124-7128.

Walbot, V., Petrov, D.A. (2001). Gene galaxies in the maize genome. Proc. Natl. Acad. Sci. USA 98: 8163-8164.

Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T., Powell, W. (1997). Genetic distribution of *BARE -1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. 253: 687-694.

Wawrzynski, A., Ashfield, T., Chen, N.W., Mammadov, J., Nguyen, A., Podicheti, R., Cannon, S.B., Thareau, V., Ameline-Torregrosa, C., Cannon, E., Chacko, B., Couloux, A., Dalwani, A., Denny, R., Deshpande, S., Egan, A.N., Glover, N., Howell, S., Ilut, D., Lai, H., Del Campo, S.M., Metcalf, M., O'Bleness, M., Pfeil, B.E., Ratnaparkhe, M.B., Samain, S., Sanders, I., Ségurens, B., Sévignac, M., Sherman-Broyles, S., Tucker, D.M., Yi, J., Doyle, J.J., Geffroy, V., Roe, B.A., Maroof, M.A., Young, N.D., Innes, R.W. (2008). Replication of nonautonomous retroelements in soybean appears to be both recent and common. Plant Physiol. 148: 1760-1771.

Weeden, N.F., Muehlbauer, F.J., Ladizinsky, G. (1992). Extensive conservation of linkage relationships between pea and lentil genetic maps. J. Hered. 83: 123-129.

Welsh, J., McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res. 18: 7213-7218.

Wessler, S.R. (1996). Turned on by stress. Plant retrotransposons. Curr. Biol. 6: 959-961.

White, S.E., Habera, L.F., Wessler, S.R. (1994). Retrotransposons in the flanking regions of normal plant genes: A role for *copia*-like elements in the evolution of gene structure and expression. Proc. Natl. Acad. Sci. 91: 11792- 11796.

Wilder, J., Holloche, H. (2001). Mobile elements and the genesis of microsatellites in dipterans. Mol. Biol. Evol. 18: 384-392.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. (1991). DNA polymorphisms amplifies by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18: 6531-6535.

Witte, C.P., Le, Q.H., Bureau, T., Kumar, A. (2001). Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc. Natl. Acad. Sci. USA 98: 13778-13783.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. Clin. Chem. 49: 853-860.

Woodrow, P., Ciarmiello, L.F., Fantaccione, S., Annunziata, M.G., Pontecorvo, G., Carillo, P. (2012). *Ty1-copia* group retrotransposons and the evolution of retroelements in several angiosperm plants: evidence of horizontal transmission. Bioinformation 8: 267-271.

Wright, D.A., Voytas, D.F. (1998). Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana Ty3/gypsy* retrotransposons that encode envelope-like proteins. Genetics 149: 703-715.

Wright, D.A., Voytas, D.F. (2002). *Athila4* of *Arabidopsis* and Calypso of soybean define a lineage of endogenous plant retroviruses. Genome Res. 12: 122-131.

Xiong, Y., Eickbush, H.E. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9: 3353-3362.

Yang, G., Nagel, D.H., Feschotte, C., Hancock, C.N., Wessler, S.R. (2009). Tuned for transposition: molecular determinants underlying the hyperactivity of a stowaway MITE. Science 325: 1391-1394.

Yañez, M., Verdugo, I., Rodríguez, M., Prat, S., Ruiz-Lara, S. (1998). Highly heterogeneous families of *Ty1-copia* retrotransposons in the *Lycopersicon chilense* genome. Gene 222: 223-228.

Yilmaz Temel, H., Göl, D., Akkale, H.B., Kahriman, A., Tanyolaç, M.B. (2015). Single nucleotide polymorphism discovery through Illumina-based transcriptome sequencing and mapping in lentil. Turk. J. Agric. For. 39: 470-488.

Yoder, J.A, Walsh, C.P., Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13: 335-340.

Yu, G.X., Wise, R.P. (2000). An anchored AFLP- and retrotransposon-based map of diploid Avena. Genome 43: 736-49.

Zamir, D., Ladizinsky, G. (1984). Genetics of allozyme variants and linkage groups in lentil. Euphytica 33: 329-336.

Zimniak-Przybylska, Z., Przybylska, J., Krajewski, P. (2001). Electrophoretic seed globulin patterns and species relationships in the genus *Lens* Miller. J. Appl. Genet. 42: 435-447.

Zohary, D. (1999). Monophyletic vs. polyphyletic origin of crops on which agriculture founded in the Near East. Genet. Resour. Crop Evol. 46: 133-142.