

# UNIVERSIDAD DE LEÓN

### FACULTAD DE CIENCIAS BIOLÓGICAS Y AMBIENTALES

Departamento de Biología Molecular

Área de Genética

## **TESIS DOCTORAL**

# ANÁLISIS DE SECUENCIAS NBS-LRR EN EL GÉNERO Lens

MARÍA NATIVIDAD TARTILÁN MENÉNDEZ

León, 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 LEGRESIST (GEN2006-27798-C6-1-E/VEG), AGL2009-07853 y AGL2013-44714-R.

A mi familia

La realización de esta tesis ha sido posible con la ayuda de mucha gente, por ello quiero agradecerles cada granito de arena que han ido aportando.

A mis directores, el Dr. Marcelino Pérez de la Vega y el Dr. Pedro García García, por brindarme esta oportunidad y por saber dirigirla a buen término.

Al Dr. Richard Fratini por todo lo que he aprendido con él y a la Dra. Ana Llorente y a Juan Guerrero por hacernos la vida más fácil a todos en el laboratorio.

A la Dra. Mª Luisa Ruiz porque su buen humor y su risa pueden hacer mejorar un mal día. Al Dr. Luis E. Sáenz de Miera por su inestimable ayuda con los programas informáticos. Y al resto de profesores y doctores del área de Genética, Dra. Francisca Vaquero, Dr. Fco Javier Vences, Dra. Mª Isabel Peláez, Dr. Carlos Polanco y Dra. Ana Isabel González por la ayuda prestada en las encrucijadas que iban apareciendo en el camino.

A todos los doctorandos con los que he coincidido, porque junto a los profesores del área han sido una gran familia laboral para mí. Especialmente a Raúl, por tener tanta paciencia conmigo y ayudarme siempre con las cuestiones de laboratorio e informáticas en nuestro pequeño mundo aparte y a Rita, por aportar claridad en algunos momentos de bloqueo y con la que tantas cosas he compartido a lo largo de estos años.

A mi madre y a mis hermanos por su amor y compresión, por soportarme como soy y sobre todo por sus esfuerzos y sacrificios para que pudiera llegar hasta aquí, porque ellos son mi mayor apoyo. A Rubén, mi cuñado, que al entrar a formar parte de la familia se convirtió en un pilar fundamental de ese apoyo.

A mi padre, porque su recuerdo me sirve de inspiración para avanzar en la vida.

A mis familiares y amigos que pese a no llegar a comprender qué les hacia a las "pobres lentejas" siempre preguntaban qué tal iba con la investigación y a las amigas que conocí durante la carrera, que entendiendo mejor lo que hacía aportaron ideas y reflexiones, en especial a Ana y su familia por acogerme como una más en su casa tantas veces.

A Modesta, por su cariño y porque me abrió las puertas de su casa cuando hizo falta.

Finalmente a David, primero como un gran amigo y posteriormente como pareja, por cada palabra de aliento y cada gesto de ánimo, que me ayudaron a seguir adelante, incluso en los peores momentos.

Por si me olvido de nombrar a alguien quiero agradecer su ayuda a todas y cada una de las personas que de alguna forma u otra han contribuido a que este proyecto se haya llevado a cabo.

## ÍNDICE

1. INTRODUCCIÓN	1	
<b>1</b> .1 La lenteja	3	
1.1.1 Características	3	
Características botánicas	3	
Características genéticas	4	
Características bioquímicas nutritivas	5	
1.1.2 Taxonomía y clasificación	6	
1.1.3 Origen, cultivo y usos	8	
Origen	8	
Cultivo	9	
Usos	14	
1.2 Sistemas de resistencia a patógenos y plagas de plantas	14	
1.2.1 Genes <i>R</i>	15	
Secuencias NBS-LRR	17	
• TIR-NBS-LRR o TNL	20	
• no TIR-NBS-LRR, CC-NBS-LRR o CNL	20	
1.3 Sintenia entre especies vegetales	24	
2. OBJETIVOS	27	
3. MATERIAL Y MÉTODOS	31	
3.1 Material vegetal	33	
3.2 Extracción de ADN genómico	33	
3.2.1 Valoración y cuantificación del ADN	33	
3.3 Amplificación de ADN. Reacción en cadena de la polimerasa	34	
3.3.1 Cebadores	34	
3.3.2 Reacciones de amplificación		
3.3.3 Valoración y cuantificación del ADN amplificado	38	

3.4 Purificación de ADN de geles de agarosa	38		
3.5 Clonación de los productos de PCR en pGEM <sup>®</sup> -T	38		
3.6 Transformación en células competentes	39		
3.6.1 Extracción de ADN plasmídico	41		
3.7 Secuenciación	41		
3.8 Nomenclatura de las secuencias	42		
3.9 Conservación de clones de interés	43		
3.10 Análisis de secuencias			
4. RESULTADOS	47		
4.1 Resultados de la secuenciación	49		
4.2 Pseudogenes	52		
4.3 Secuencias utilizadas en los análisis filogenéticos	55		
4.4 Análisis de los aminoácidos y de los motivos conservados	67		
4.4.1 Contenido de aminoácidos	67		
4.4.2 Motivos conservados	69		
4.5 Comparación con datos previos y con otras especies	76		
5. DISCUSIÓN	105		
5.1 Pseudogenes y secuencias descartadas	109		
5.2 Análisis filogenéticos	111		
5.2.1 Cambios sinónimos y no sinónimos	112		
5.3 Motivos conservados	113		
5.3.1 Contenido en nucleótidos y aminoácidos	116		
5.4 Comparación con otras especies	117		
6. CONCLUSIONES	125		
7. BIBLIOGRAFÍA	129		
8. ANEXOS	CD		

#### **ABREVIATURAS**

ADN: ácido desoxirribonucleico

Amp: ampicilina

CViT: Chromosome Visualization Tool. Herramienta de visualización de cromosomas

dNTP: desoxinucleótido trifosfato

EDTA: ácido etilendiaminotetraacético

ETI: effector triggered immunity. Inmunidad activada por efectores.

HR: hypersensitive response. Respuesta hipersensible

**IPTG:** isopropil-β-D-tiogalactósido

**LB:** Luria-Bertani (medio de cultivo)

LccA: Lens culinaris subespecie culinaris variedad ILL5588

LccB: Lens culinaris subespecie culinaris variedad Lupa

Lco: Lens culinaris subespecie orientalis

Ler: Lens ervoides

Lla: Lens lamottei

Lni: Lens nigricans

Lod: Lens odemensis

Lto: Lens tomentosus

LRR: leucine rich repeat. Repeticiones ricas en leucina

MAMP: microbe-associated molecular pattern. Patrones moleculares asociados a microbios

Mpb: mega pares de bases

NBS: nucleotide binding site. Sitios de unión a nucleótidos

NOR: nucleolar organizing region. Region organizadora nucleolar

ORF: open reading frame. Marco abierto de lectura

PAMP: pathogen associated molecular pattern. Patrones moleculares asociados a patógenos

Pb: pares de bases

PCoA: principal components analysis. Análisis de componentes principales

PCR: polymerase chain reaction. Reacción en cadena de la polimerasa

PRR: pattern recognition receptor. Receptor de reconocimiento de patrones

PTI: PAPMs triggered immunity. Inmunidad activada por PAPMs

**p/v:** peso/volumen

RGA: resistance gene analogs. Análogos a genes de resistencia

**SDS:** dodecilsulfato sódico

Syn: sinónimo

TAE: tampón de tris-acetato-EDTA

U: unidades

X-Gal: 5-bromo-4-cloro-3-indolil-β-D-galactósido

## Relación de figuras

Figura 1.1. Lámina ilustrativa de <i>Lens</i>	3
Figura 1.2. Mapa mundial que representa el nivel de producción por países en el año 2013	9
Figura 1.3. Niveles de producción de los 5 principales productores de lenteja en el año 2013	10
Figura 1.4. Países y cantidades en toneladas a) exportados b) importados en 2011	10
Figura 1.5. Gráficos de producción y superficie sembrada de leguminosas grano en España distribuida por comunidades autónomas para la campaña 2013/2014	11
Figura 1.6. Datos de a) producción y b) superficie sembrada de las últimas cinco campañas en España para yeros, vezas, garbanzos, lentejas y judías secas de la campaña 2013/2014	12
Figura 1.7. Datos de a) exportación, b) importación entre los años 1991 y 2011 en España	13
Figura 1.8. Representación esquemática de una proteína NBS-LRR típica	17
Figura 1.9. Modelo en tres dimensiones de la estructura del motivo NB-ARC de Gpa2 de patata	17
Figura 1.10. Representación esquemática de una proteína NBS-LRR típica donde se señalan los motivos conservados con líneas discontinuas	19
Figura 1.11. a) Secuencia del dominio CC de la proteína MLA10 que muestra la localización de las tres hélices. b) Esquema del entrelazado de dos dominios CC formando un dímero	21
Figura 1.12. Modelo de activación de la proteína NBS-LRR	23
Figura 3.1. Localización de los cebadores con respecto a los dominios conservados de la región NB-ARC	35
Figura 3.2. Vector de clonación pGEM <sup>®</sup> -T Easy Vector	39
Figura 3.3. Desglose de la nomenclatura utilizada para las secuencias obtenidas en este estudio	43
Figura 4.1. Geles de agarosa donde se muestran los patrones de bandas obtenidos con cada una de las combinaciones de cebadores	49
<b>Figura 4.2.</b> Gráfica de distribución que muestra el aumento de secuencias diferentes con respecto al número de amplificaciones realizadas en <i>L. culinaris</i> subsp. <i>culinaris</i> Lupa con las combinaciones C6 y C10	52
Figura 4.3. Secuencias de nucleótidos en las que se puede observar una deleción central de 153 nucleótidos	53
Figura 4.4. Secuencias de aminoácidos que muestran una sucesión de cambios en el marco de lectura	55

Figura 4.5. Secuencia con duplicación en tándem	55
Figura 4.6. Árboles filogenéticos obtenidos mediante el algoritmo de Kimura-2 parámetros con el programa MEGA 5.2, a) el árbol general con las 130 secuencias muestra una marcada división entre los grupos TNL y CNL y la denominación de los distintos grupos identificados, b) árboles individuales de cada grupo	57
Figura 4.7. Secuencias de aminoácidos en las que se muestra la localización de los motivos conservados característicos de las secuencias NBS-LRR. Se han mantenido los fragmentos de secuencia correspondientes a los cebadores ya que en algunos de los casos estos coinciden con un motivo conservado	70
Figura 4.8. Secuencias consenso de aminoácidos para las secuencias del grupo CNL entre los dominios p-loop y GLPL (no incluidos). a) Representación en frecuencias b) representación en bits	71
Figura 4.9. Secuencias consenso de aminoácidos para las secuencias del grupo TNL entre los dominios p-loop y GLPL (no incluidos). a) Representación en frecuencias b) representación en bits	72
Figura 4.10. Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en <i>Lens</i> . a) grupo CNL, b) grupo TNL.	77
<b>Figura 4.11.</b> Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en <i>Lens</i> junto con las secuencias de <i>Medicago truncatula</i> . a) grupo CNL, b) grupo TNL	79
Figura 4.12. Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en <i>Lens</i> junto con las secuencias de <i>Glycine max</i> . a) grupo CNL, b) grupo TNL	81
<b>Figura 4.13.</b> Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en <i>Lens</i> junto con las secuencias de <i>Cicer arietinum</i> . a) grupo CNL, b) grupo TNL	83
Figura 4.14. Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en <i>Lens</i> junto con las secuencias de <i>Phaseolus vulgaris</i> . a) grupo CNL, b) grupo TNL	85
Figura 4.15. Gráficos 3D que muestran la agrupación de secuencias obtenida por análisis de componentes principales	87
Figura 4.16. Árboles en los que se muestran los grupos de secuencias NBS-LRR y la separación por subgrupos en los grupos donde se hayan formado en a) CNL y b) TNL	89
Figura 4.17. Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el grupo NT-1 de lenteja	91
Figura 4.18. Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el grupo NT-2 de lenteja	91

<b>Figura 4.19.</b> Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el grupo NT-3 de lenteja	92
<b>Figura 4.20.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el grupo NT-4 de lenteja	92
<b>Figura 4.21.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el grupo NT-5 de lenteja	94
<b>Figura 4.22.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el grupo NT-6 de lenteja	95
<b>Figura 4.23.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el subgrupo T-1.1 de lenteja	96
<b>Figura 4.24.</b> Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el subgrupo T-1.2 de lenteja	97
<b>Figura 4.25.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el subgrupo T-1.3 de lenteja	97
<b>Figura 4.26.</b> Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el subgrupo T-2.1 de lenteja	99
<b>Figura 4.27.</b> Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el subgrupo T-2.2 de lenteja	99
<b>Figura 4.28.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el subgrupo T-3.1 de lenteja	100
<b>Figura 4.29.</b> Representación de la localización de las secuencias de <i>M. truncatula</i> que muestran similitud con el subgrupo T-3.2 de lenteja	101
<b>Figura 4.30.</b> Representación de la localización de las secuencias de a) <i>M. truncatula</i> y b) <i>G. max</i> que muestran similitud con el grupo T-4 de lenteja	101
<b>Figura 4.31.</b> Representación de la localización de las secuencias homólogas de <i>M. truncatula</i> para a) las secuencias seleccionadas de los grupos CNL y b) las secuencias seleccionadas de los grupos y subgrupos TNL	103
<b>Figura 4.32.</b> Representación de la localización de las secuencias homólogas de <i>G. max</i> para a) las secuencias seleccionadas de los grupos CNL y b) las secuencias seleccionadas de los grupos y subgrupos TNL	104
Figura 5.1. Fragmento de secuencias de aminoácidos del dominio NBS en el que se muestran sombreados dos puntos de conservación característicos	116
<b>Figura 5.2.</b> Distribución de genes <i>NBS-LRR</i> codificados en los ocho cromosomas de <i>M. truncatula</i> según Ameline-Torregrosa et al. (2008) adaptado a la base de datos Mt v3.5.1	117
<b>Figura 5.3.</b> Distribución de genes <i>NBS-LRR</i> de <i>M. truncatula</i> que mostraron similitud con las secuencias de lenteja	118

Figura 5.4. Distribución de genes NBS-LRR en Glycine max	119
Figura 5.5. Distribución de genes NBS-LRR de Glycine max que mostraron similitud con las secuencias de lenteja	120
Figura 5.6. Gráfica de correlación entre el número de secuencias de la especie <i>Lens</i> culinaris subsp culinaris de cada grupo o subgrupo y el número de secuencias	
homólogas de <i>M. truncatula</i>	122

#### Relación de tablas

<b>Tabla 1.1</b> . Fórmula de los cariotipos (n = 7) para las cinco especies de Lens que sedescribieron en el trabajo de Balyan et al. (2002)	5
<b>Tabla 1.2.</b> Producción de los 20 principales países productores de lenteja en el año2013	11
Tabla 1.3.      Secuencia de los motivos conservados en las proteínas de resistencia	18
Tabla 3.1.      Secuencias de los cebadores degenerados utilizados	35
Tabla 3.2. Combinaciones de cebadores degenerados	36
Tabla 3.3.      Secuencias de los cebadores no degenerados utilizados	36
<b>Tabla 4.1</b> . Número de muestras totales analizadas mediante secuenciación desglosadaspor especie y por combinación de cebadores	50
Tabla 4.2.Número de secuencias que muestran alta similitud con NBS-LRR,desglosadas por especie y por combinación de cebadores	51
Tabla 4.3. Número de secuencias utilizadas en los estudios filogenéticos desglosadaspor especie y por combinación de cebadores	56
Tabla 4.4. Tamaño de las secuencias en número de nucleótidos una vez eliminados los cebadores de los extremos	63
Tabla 4.5. Valores medios de diversidad evolutiva dentro de los distintos grados de agrupación de secuencias	64
Tabla 4.6. Rango de frecuencias nucleotídicas para el total de las secuencias, por subfamilias y por grupos	64
T <b>abla 4.7.</b> Valores promedio de los cambios sinónimos (dS), no sinónimos (dN), su diferencia (dS-dN) y el cociente entre ambas ( $\omega$ = dS/dN) de todas las secuencias, por subfamilias y por grupos.	66
Tabla 4.8. Contenido en porcentaje de aminoácidos en las secuencias	68
Tabla 4.9.Comparación de los motivos conservados en las secuencias comunespropuestos para los genes TIR (TNL) y No-TIR (CNL) (líneas superiores) y losencontrados en los correspondientes grupos de secuencias de lenteja	73

Tabla 4.10. Comparación de los motivos conservados en las secuencias comunespropuestos para los genes CNL (líneas superiores) y los encontrados en los distintosgrupos de secuencias de lenteja	74
Tabla 4.11. Comparación de los motivos conservados en las secuencias comunespropuestos para los genes TNL (líneas superiores) y los encontrados en los distintosgrupos de secuencias de lenteja	75
T <b>abla 4.12.</b> Porcentajes de variación en la matriz de distancias de las secuencias que representa cada eje de los gráficos obtenidos con el programa STATISTICA y el total de la representación acumulada	88
<b>Tabla 5.1.</b> Tipo de secuencias obtenidas con las combinaciones de cebadores degenerados basados en los motivos conservados p-loop y GLPL en lenteja (este estudio), guisante (Djebbi <i>et al.,</i> 2015), haba y garbanzo (Palomino <i>et al.,</i> 2006) y patata (Leister <i>et al.,</i> 1996)	108
patata (Leister et al., 1996)	108

# Introducción

#### 1. INTRODUCCIÓN

#### 1.1. La lenteja

La lenteja (*Lens culinaris* Medik.) forma parte de las denominadas leguminosas grano, ya que su utilidad primordial es el consumo humano de la semilla. Es una de las primeras especies de plantas domesticada en el Creciente Fértil y ha sido cultivada desde hace unos 10.000 años en los entornos agrícolas más difíciles (Pérez de la Vega *et al.,* 2011). En la alimentación humana, la lenteja es considerada, además de un aporte calórico, una fuente de hierro, proteínas, fibra y antioxidantes, aunque también contiene factores antinutritivos (Bhatty, 1988).

#### 1.1.1 Características

#### Características botánicas

Desde el punto de vista botánico la lenteja es una planta herbácea anual, adaptada a crecer en climas secos, considerada un cultivo de secano. Se la clasifica como planta de día largo ya que necesita más de 12 horas de luz para florecer, pero se han descrito cultivares de día neutro (Summerfield y Roberts, 1988).

Es planta una erecta 0 semierecta y su altura puede variar entre los 15 y 75 cm, dependiendo del genotipo y de las condiciones ambientales; aunque la mayoría de las variedades tiene una altura de entre 25 y 30 cm. Generalmente es una planta poco pubescente (a excepción de algunos cultivares de Palestina, Siria, Turquía e India). Tanto los tallos como las ramas de la planta son delgados y el radicular sistema está constituido por una raíz pivotante con numerosas raíces laterales fibrosas.

Tiene hojas pinnadas o imparipinnadas con hasta 14 foliolos con una longitud de uno a tres cm. Cada hoja tiene dos pequeñas estípulas y puede acabar o no en zarcillo. La flor es completa y típicamente



Figura 1.1. Lámina ilustrativa de *Lens*. Obtenida de Deutschlands Flora in Abbidulgen de Johann Georg Sturm (1796)

papilionácea, de color blanco, morado pálido o morado azulado y con un tamaño entre los cuatro y los ocho mm de longitud. Se originan pedúnculos florales de una a tres flores, raramente cuatro, en los nodos superiores de la planta.

El fruto o vaina es romboidal, comprimido lateralmente, con una longitud de 6 a 20 mm y una anchura de 3,5 a 11 mm. Contiene una o dos semillas, raramente tres (Muehlbauer *et al.*, 1995). Generalmente las vainas de la parte inferior de la planta son más grandes y es más frecuente que sean éstas las que contienen dos semillas mientras que las de la parte superior suelen tener un tamaño más pequeño. La semilla tiene forma lenticular y su tamaño varía mucho de unas variedades a otras. Esta diferencia de tamaño llevó a Barulina (1930) a clasificar las lentejas cultivadas en dos tipos, macrosperma, que englobaba las variedades con semillas más grandes (entre seis y nueve mm de diámetro) y microsperma, con semillas más pequeñas (de dos a seis mm de diámetro).

La germinación es hipogea. La temperatura óptima para la germinación de la semilla está entre los 18 y los 21 °C, necesitando una temperatura mínima de 15 °C para que comience la germinación (Alonso Ponga y Cristóbal, 1996).

Las lentejas, como todas las leguminosas, son fijadoras de nitrógeno, debido a la relación simbiótica que mantienen con bacterias del género *Rhizobium*. Las raíces de la capa superficial del suelo tienen gran cantidad de nódulos que no suelen superar los cinco mm de tamaño y que albergan a *Rhizobium leguminosarum* (Alonso Ponga y Cristóbal, 1996).

#### **Características genéticas**

La lenteja es una leguminosa predominantemente autógama con una frecuencia de fecundación cruzada generalmente baja (Skibinski *et al.,* 1984; Wilson y Law, 1972). Horneburg (2006) estima la tasa de cruzamiento en el rango de 0,01 al 5,66%.

Todas las especies del género *Lens* Miller son diploides, con un número cromosómico 2n = 14 y comparten un cariotipo parecido que consiste en tres pares de cromosomas metacéntricos o submetacéntricos, tres pares de cromosomas acrocéntricos y un par de cromosomas metacéntricos con región organizadora nucleolar (NOR) próxima al centrómero (Ladizinsky, 1993).

Balyan *et al.* (2002) describieron el cariotipo de la especie *Lens culinaris:* dos pares de cromosomas metacéntricos, uno de ellos con una región NOR; dos pares de cromosomas submetacéntricos y tres pares de cromosomas acrocéntricos. Los cariotipos descritos en ambos casos muestran la misma estructura, pero mientras Ladizinsky (1993) lo extendió a todo el género *Lens,* Balyan *et al.* (2002) establecieron diferencias entre *L. culinaris* y otras especies en el número de cromosomas metacéntricos, submetacéntricos y acrocéntricos (Tabla 1.1).

El contenido C de lenteja determinado por citometría de flujo es de 4,41 pg que equivale a 4.036 Mpb (Arumuganathan y Earle, 1991). Esto coincide con el resultado de 4,6 pg obtenido por Bennet y Smith (1976) usando microdensitometría Feulgen. El tamaño genómico de la lenteja es muy similar al de especies cercanamente relacionadas, como el guisante (3.947-4.397 Mpb).

Especie	Cariotipo <sup>1</sup>
L. nigricans	1 M(A) + 3 SM(A) + 1 Acro(B)+ 2 Acro(A)
L. ervoides	3 SM(A) + 1 SM(B)+ 3 Acro(A)
L. odemensis	2 M(A) + 1 M(B) + 1 SM(A)+ 3 Acro(A)
L. orientalis	1 M(B) + 4 SM(A)+ 2 Acro(A)
L. culinaris	1 M(A) + 1 M(B) + 2 SM(A)+ 3 Acro(A)

**Tabla 1.1**. Fórmula de los cariotipos (n = 7) para las cinco especies de *Lens* que se describieron en el trabajo de Balyan *et al.* (2002).

<sup>1</sup> M: metacéntrico, SM: submetacéntrico, Acro: acrocéntrico, A: sin constricción secundaria y B: con constricción secundaria (modificado de Balyan *et al.*, 2002).

#### Características bioquímicas nutritivas

La lenteja es llamada la "carne del pobre" ya que constituye un gran aporte de energía y proteínas, además de fibra y antioxidantes. Las cantidades difieren entre variedades. Los rangos de los componentes por 100 g de materia seca de lenteja cruda son (Urbano *et al.*, 2007):

- Energía: 1.483 2.010 KJ.
- Proteínas: 20,6 31,4 g. las proteínas son principalmente globulinas, pero también hay albúminas.
- Grasas: 0,7 4,3 g. El contenido en grasas es bajo, como en la mayoría de las leguminosas.
  Gran parte son triglicéridos.
- Carbohidratos: 43,4 69,9 g. La mayoría es almidón, aunque también hay otros azúcares, como celulosa, hemicelulosa y lignina.
- Fibra: 5,0 26,9 g.
- Ceniza: 2,2 4,2 g.

La lenteja también contiene vitaminas, como riboflavina y oligoelementos: el más destacado es el hierro, ya que las lentejas son consideradas una fuente de este nutriente dentro de la dieta humana.

La variación en la cantidad de algunos componentes presentes en la semilla, sobre todo la fibra, se debe en parte a la manera en que se procesa y consume la lenteja, habiendo dos maneras principales: como semilla completa o como semilla descascarillada y partida (Pérez de la Vega *et al.*, 2011). Aunque las cubiertas de la semilla son como máximo el 20% del peso total de la semilla, el descascarillado altera la composición nutricional aumentando proporcionalmente la cantidad de grasa, proteínas y tiamina y disminuyendo otros componentes como la fibra, tanto soluble como insoluble, el hierro y el calcio (Cuadrado *et al.*, 2002; Ghavidel y Prakash, 2007; Urbano *et al.*, 2007). Esta composición también se ve alterada por el proceso de cocinado donde aumenta proporcionalmente la cantidad de carbohidratos, proteínas y fibra insoluble (Almeida-Costa *et al.*, 2007).

Aunque la concentración de proteína es relativamente alta, éstas tienen una baja cantidad de aminoácidos azufrados y triptófano. Sin embargo, presentan un alto contenido en lisina, que en los

cereales es un aminoácido limitante, por eso se considera que estos dos alimentos se complementan nutricionalmente (Alonso Ponga y Cristóbal, 1996; Castell y Cliplef, 1990).

Además del gran valor nutricional, la lenteja también tiene factores que actúan como tóxicos o provocan problemas digestivos o metabólicos al ser ingeridos, aunque es una de las leguminosas que presenta menor cantidad de estos elementos. Dos componentes importantes son el fitato y los taninos que forman complejos con las proteínas y provocan una disminución en la disponibilidad de éstas.

Otros factores de este tipo son los inhibidores de proteasas, fundamentalmente inhibidores de tripsina y quimotripsina. Estos componentes provocan muchos problemas en el uso de las leguminosas en alimentación animal.

Otras sustancias potencialmente nocivas presentes en la lenteja son las saponinas y las lectinas o fitohemaglutininas; estas últimas provocan una mala absorción intestinal. La mayoría de los factores antinutritivos pueden ser eliminados poniendo las lentejas a remojo un tiempo antes de ser cocinadas o por calentamiento en el proceso de cocinado.

#### 1.1.2 Taxonomía y clasificación

En 1787 un botánico y médico alemán llamado Medikus identificó la lenteja con el nombre científico *Lens culinaris* y este nombre es el utilizado en la actualidad (Cubero *et al.,* 2009), aunque el nombre de *Lens esculenta* dado por Moench puede aparecer todavía en algunas publicaciones.

El nombre del género *Lens* no tiene una autoría clara, en el Congreso de Nomenclatura Botánica de 1966 se le atribuyó a Miller. Esto hace que el nombre del género quede como *Lens* Miller con prioridad sobre nombres más antiguos (Cubero *et al.*, 2009; Gunn 1969). El género *Lens* pertenece a la Tribu *Vicieae* DC (syn. *Fabeae* Rchb) junto a los géneros *Lathyrus* L., *Pisum* L. y *Vicia* L. Otros géneros como *Cicer* L. y *Vavilovia* anteriormente formaban parte de la tribu (Gunn y Kluve, 1976), pero análisis filogenéticos posteriores hicieron que *Cicer* fuera excluido y pusieron en duda la inclusión de *Vavilovia* (Steele y Wojciechowski, 2003).

Esta tribu forma parte de la familia *Fabaceae*, que en la Sistemática clásica recibe el nombre de *Papilionaceae*.

Los estudios realizados por Jansen *et al*. (2008) y por Wojciechowski (2006) situaron a la tribu *Vicieae* filogenéticamente cerca de géneros como *Trifolium*, *Medicago* y el excluido *Cicer*.

Tradicionalmente la taxonomía de *Lens* se basaba en la morfología, pero algunas similitudes o diferencias morfológicas no tienen por qué deberse a relaciones filogenéticas, sino que pueden ser debidas a la convergencia evolutiva. Los estudios dependen mucho de las accesiones escogidas y de las características estudiadas (Ahmad *et al.*, 1997), pero el uso de datos moleculares ayuda a alcanzar una filogenia más precisa del género (Durán y Pérez de la Vega, 2004; Mayer y Bagga, 2002; Sonnante *et al.*, 2003).

El primer estudio detallado de la lenteja cultivada lo realizó en 1930 Barulina. Como ya se ha mencionado, en ese trabajo ella agrupó las variedades de lenteja en dos subespecies tomando como criterio principal el tamaño de la semilla: *macrosperma* (semillas grandes, de seis a nueve mm) y

*microsperma* (semillas pequeñas y medias, de dos a seis mm), además de otros factores como el color de los cotiledones y la abundancia de pigmentación en las flores y partes vegetativas de la planta. También tuvo en cuenta la distribución geográfica de grupos de caracteres, definiendo grupos regionales.

Los principales caracteres usados en la clasificación de las subespecies son vainas, semillas y diferencias en las flores, tamaño de los foliolos y altura de las plantas. Dentro de las subespecies los grupos se definen con caracteres como dehiscencia, longitud de los dientes del cáliz y número de flores por pedúnculo (Cubero *et al.*, 2009).

Un estudio de Erskine *et al.* (1989) sobre la variación morfológica del germoplasma identificó tres grandes grupos regionales. Un grupo formado por las muestras de Egipto, Jordania, Líbano y Siria. Un segundo grupo más al norte compuesto por Grecia, Irán, Turquía y los territorios de la antigua URSS. Y por último, un tercer grupo formado por accesiones de India y Etiopia que presentan caracteres morfológicos cuantitativos asombrosamente similares.

La clasificación de *Lens culinaris* y las relaciones filogenéticas entre las especies del género *Lens* han suscitado muchas discusiones a lo largo de los años.

La taxonomía clásica, basada en análisis morfológicos y/o marcadores bioquímicos, consideraba cuatro especies en el género *Lens: L. culinaris, L. orientalis, L. nigricans* y *L. ervoides* (Ladizinsky, 1979). Una clasificación posterior propone únicamente dos especies: *L. culinaris* y *L. nigricans*. En esta clasificación la especie *L. culinaris* tendría como subespecies *culinaris, orientalis* y *odemensis* y la especie *L. nigricans* a *nigricans* y a *ervoides* (Ladizinsky *et al.*, 1984).

En 1997 se añadieron dos nuevas especies al género *Lens: L. tomentosus* (ex *L. orientalis* Ladizinsky, 1997) y *L. lamottei* (ex *L. nigricans,* Van Oss *et al.,* 1997).

Usando marcadores moleculares y morfológicos Ferguson *et al.* (2000) consideraron que *odemensis* y *tomentosus* eran subespecies de *L*. *culinaris*, proponiendo la clasificación que actualmente utiliza el NCBI (National Center for Biotechnology Information): *L. culinaris*, con cuatro subespecies (*culinaris, orientalis, tomentosus* y *odemensis*), *L. nigricans*, *L. ervoides* y *L. lamottei*.

Wong *et al.* (2015) en un estudio basado en SNPs, realizado utilizando métodos de genotipado por secuenciación (Genotyping-by-sequencing o GBS), concluyeron que el género *Lens* consiste en siete taxones diferenciados en cuatro "gene pools": *L. culinaris/ L. orientalis/ L. tomentosus* como "pool" primario; *L. lamottei/L. odemensis* forman el pool secundario; *L. ervoides* el terciario y *L. nigricans* el cuaternario.

Por último, la clasificación más aceptada en la actualidad sugiere seis especies dentro del género: *L. odemensis* (Godr) Ladiz., *L. tomentosus* Ladiz., *L. nigricans* (Bieb) Godr., *L. ervoides* (Bring.) Grande, *L. lamottei* Czfr. y *L. culinaris* Medik., que tiene dos subespecies, *culinaris* (la lenteja cultivada o cultigen) y *orientalis* (Boiss) Ponert (el ancestro silvestre). Las dificultades en la hibridación apoyan la idea de las seis especies diferenciadas de *Lens* (Pérez de la Vega *et al.*, 2011).

Al ser el ancestro silvestre de la especie cultivada a *L. orientalis* se le da el nivel de subespecíe atendiendo a los criterios de nomenclatura para plantas cultivadas de Harlan y De Wet (1971).

#### 1.1.3 Origen, cultivo y usos

#### Origen

La lenteja lleva siendo cultivada desde hace unos 10.000 años en los entornos agrícolas más difíciles además de ser una de las primeras plantas domesticadas en el Creciente Fértil. El origen de la forma cultivada es la especie silvestre L. culinaris subsp. orientalis (syn. L. orientalis). Teorías basadas en los datos arqueológicos, la distribución de las especies silvestres y la aparición de restos de las especies silvestres y la cultivada en la misma región, hacen de la zona de Oriente Próximo la mejor candidata como origen de domesticación. Barulina (1930) propuso la región entre Afganistán, India y Turquistán como posible centro de origen de domesticación. Basó esta propuesta en la gran cantidad de variedades endémicas localizadas en esa zona. Pero esa gran cantidad de variedades se puede explicar por una selección artificial llevada a cabo por diferentes poblaciones humanas, que conlleva una drástica fijación genética. Una localización más concreta del origen podría encontrarse entre el sur de Turquía y el norte de Siria (Ladizinsky, 1999; Lev-Yadun et al., 2000; Zohary, 1999) donde L. culinaris subsp. culinaris pudo originarse debido a la domesticación inconsciente de alguna población de la especie silvestre L. culinaris subsp. orientalis. Alo et al. (2011) sugirieron un posible centro de domesticación de la lenteja localizado en el sur de Turquía que se extendía entre Denizli (al suroeste) y Antakya (sur-central). El Creciente Fértil también ha sido propuesta como área de domesticación de otras leguminosas, como los guisantes y de algunos cereales como la cebada y el trigo (Lev-Yadun et al., 2000). Históricamente, en muchas culturas las leguminosas se consumían junto con los cereales ya que se complementan nutricionalmente.

El estudio de los rasgos de domesticación y la diversidad de la especie, además del polimorfismo cromosómico revelaron un efecto fundador que lleva a la conclusión de que la lenteja fue domesticada en una única ocasión o muy pocas veces (Zohary, 1999, tomado de Pérez de la Vega *et al.*, 2011).

El cultivo de la lenteja tuvo tres ramas de expansión primarias: Europa por el oeste, el centro y sur de Asia por el este y el norte de África y la Península Arábica por el sur. La llegada del cultivo de lenteja al Delta del Nilo fue temprana, debido a la proximidad geográfica y cultural con el Creciente Fértil. Los antiguos egipcios tenían extensos cultivos de lenteja, se han encontrado semillas en tumbas pertenecientes a la XII dinastía (Cubero *et al.*, 2009).

La llegada a Etiopia se produjo probablemente por el Nilo. A la India llega hace unos 4.000 años probablemente llevada por la invasión Indo-Europea a través de Afganistán (De Candolle, 1882, tomado de Pérez de la Vega *et al.*, 2011). El cultivo de lenteja llegó a América en el siglo XVI a través de España.

Los restos más antiguos encontrados en la Península Ibérica datan del Neolítico temprano. Junto a lentejas se encontraron restos de otros cultivos que tradicionalmente estaban asociados, cereales como los trigos *Triticum monococcum*, *T. dicocum*, *T. aestivum/durum* o la cebada (*Hordeum vulgare*) u otras leguminosas como habas (*Vicia faba*), guisantes (*Pisum sativum*) y almorta (*Lathyrus cicera/sativus*) (Alonso, 2008; Buxó, 1991; Buxó, 1997).

Según Cubero *et al.* (2009) si el cultivo de la lenteja ha sido mantenido por los agricultores a través de los años, es porque crece en suelos pobres y climas duros y en muchos casos era la única fuente de proteínas disponible.

#### Cultivo

Según los datos obtenidos a través de FAOSTAT (http://faostat.fao.org), en 2013 más de 50 países de todo el mundo cultivaron lenteja (Figura 1.2). Se obtuvo una producción global de más de 4,9 millones de toneladas en una superficie total de cultivo de 4.326.781 hectáreas. Estos datos equivalen al 6,80% de la producción y al 5,34% de la superficie cultivada de legumbre seca a nivel mundial para ese año.



**Figura 1.2**. Mapa mundial que representa el nivel de producción por países en el año 2013. Tomado de la página web de FAOSTAT.

Las variedades de lenteja que se cultivan en cada región dependen en gran medida del destino último de la cosecha. En los países donde se consumen tradicionalmente en la dieta, la mayoría de la cosecha es destinada a demanda local y esto será lo que determine principalmente el tipo de lenteja que se cultiva, mientras que en los países donde lo predominante es la exportación, el comercio internacional determinará las variedades de lenteja sembradas, ya que se han de tener en cuenta las preferencias del mercado al que vayan destinadas. Por ejemplo, en Europa se consumen semillas enteras y con el cotiledón verde o amarillo, mientras que en el subcontinente Indio prefieren un cotiledón rojo, descascarillada y la semilla se consume partida (Pérez de la Vega *et al.*, 2011).

Hasta 2008 el mayor productor de lenteja a nivel mundial era la India (excepto en 2005, que fue Canadá). Esto coincide con el hecho de que también es el país con mayor número de hectáreas cultivadas (casi 1,9 millones de hectáreas en 2013). Pero a partir de 2008 Canadá se convierte en el mayor productor, con una producción de 1,88 millones de toneladas en 2013 (Figura 1.3), debido al progresivo aumento de la cantidad de hectáreas cultivadas y a la amplia superioridad de rendimiento de la cosecha canadiense con respecto a la india. Este incremento del área de cultivo se mantiene en Canadá hasta 2010, llegando a superar el 1,3 millones de hectáreas, cifra que desciende los tres años siguientes.



**Figura 1.3.** Niveles de producción de los 5 principales productores de lenteja en el año 2013. Tomado de la página web de FAOSTAT.

En 2011, Canadá es también el mayor exportador de lenteja, seguido por Australia, Turquía, Estados Unidos y los Emiratos Árabes (Figura 1.4a). Mientras que el primer país importador es Turquía, seguido de la Unión Europea (Figura 1.4b).



**Figura 1.4.** Países y cantidades en toneladas a) exportados b) importados en 2011. Tomado de la página web de FAOSTAT.

En las lista de los 20 países principales en producción de lenteja en el año 2013 España se encuentra en la posición número 13 (Tabla 1.2), siendo el primer productor de la Unión Europea, con un total de producción de 40.700 toneladas en 31.400 hectáreas cultivadas.

Posición	Región	Producción (T) <sup>1</sup>	
1	Canadá	1.880.500,0	
2	India	1.134.000,0	
3	Turquía	417.000,0	
4	Australia	324.100,0	*
5	Estados Unidos de América	227.658,0	
6	Nepal	226.931,0	
7	China, Continental	150.000,0	*
8	Etiopía	129.833,0	
9	República Árabe Siria	129.370,0	
10	Bangladesh	93.000,0	
11	Irán (República Islámica de)	73.000,0	F
12	Marruecos	42.948,0	
13	España	40.700,0	
14	Francia	22.724,6	
15	Federación Rusa	16.625,0	
16	Pakistán	9.883,0	
17	Yemen	9.415,0	*
18	Argelia	6.318,0	
19	Kenia	4.100,0	F
20	Perú	4.037,5	

**Tabla 1.2.** Producción de los 20 principales países productores de lenteja en el año 2013. Datos tomados de la página web de FAOSTAT.

<sup>1</sup> F: estimación FAO

\*: cifras no oficiales

Los datos del Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA) indican que la comunidad autónoma española con mayor producción y área de cultivo de leguminosas grano es Castilla la Mancha, seguida de Castilla y León (Figura 1.5). En el caso de la lenteja, la provincia de Cuenca monopoliza el cultivo con casi un 80% de la producción.



(\*\*) Avances

**Figura 1.5.** Gráficos de producción y superficie sembrada de leguminosas grano en España distribuida por comunidades autónomas para la campaña 2013/2014. Datos obtenidos de Avances de superficies y producciones agrícolas de la SGT del MAGRAMA.

Según unos estudios de la Secretaría General Técnica (SGT) del MAGRAMA, en el año 2013 la superficie cosechada de leguminosas grano en España era un 6% mayor que la media de las cinco últimas campañas (Figura 1.6a) y la producción (Figura 1.6b) se incrementó un 28% con respecto a dicha media y hasta un 30% con respecto a la campaña anterior, afectada por la sequía.



a)





(\*) Estimaciones

(\*\*) Avances

**Figura 1.6.** Datos de a) producción y b) superficie sembrada de las últimas cinco campañas en España para yeros, vezas, garbanzos, lentejas y judías secas de la campaña 2013/2014. Datos obtenidos de Avances de superficies y producciones agrícolas de la SGT del MAGRAMA.

Utilizando datos de la FAO específicos de lenteja se aprecia que entre 2000 y 2013, tanto la extensión cosechada, como el rendimiento y la producción han sido muy irregulares en España, destacando el año 2005, en el cual el área cultivada (36.151 ha) es similar a otros años, como 2010 (37.280 ha) pero la producción no alcanza las 7.000 toneladas, disminuyendo un 75% con respecto al año anterior. Esto se debió a los efectos de la sequía que sufrió el país y que redujo a la mitad la producción total española de leguminosas ese año (anuario de la UPA, 2006). Los datos de producción para el 2013 presentados por el MAGRAMA y la FAO son 40.700 toneladas en un área de 31.400 ha. Los resultados de la encuesta sobre superficies y rendimiento de cultivos (ESYRCE) revelan que la superficie de cultivo destinada a lenteja en España disminuye un 1,2% (hasta las 29.926 ha) en 2014.

España, como muchos países productores de lenteja, participa en el comercio de exportaciónimportación de este producto (Figura 1.7). Tanto España como toda la UE son importadoras netas de leguminosas grano. Las importaciones de lentejas proceden fundamentalmente de Canadá y EEUU (datos del MAGRAMA).



**Figura 1.7**. Datos de a) exportación, b) importación entre los años 1991 y 2011 en España. Tomado de la página web de FAOSTAT.

Al utilizar datos recogidos de FAOSTAT y del MAGRAMA se aprecia una discordancia en dichos datos en algunos años, llegando a contradecirse datos y gráficas obtenidos de la web del MAGRAMA para un mismo año. Esto dificulta la interpretación y la fiabilidad de los análisis.

a)

#### Usos

La utilidad primordial de la lenteja es el consumo humano de la semilla.

Un uso indirecto de las leguminosas es su utilización como fertilizante de las tierras de cultivo, debido al incremento en la disponibilidad de nitrógeno en el suelo. Por ese motivo siempre han sido parte indispensable en el sistema de rotación de cultivos en la agricultura tradicional. Las leguminosas grano, tienen una utilidad secundaria como forraje (Erskine *et al.*, 1990), generalmente en pequeñas granjas. También se pueden usar como abono verde o para ayudar al control de la erosión del suelo en algunos países. La paja o no se utiliza o el uso es mínimo.

Se han propuesto algunos usos alternativos: por ejemplo, la harina de lenteja se puede utilizar como fuente de nitrógeno junto con células de levadura, para la producción de ácido L-(+) láctico por fermentación anaerobia de *Lactobacillus amylophilus* GV6 (Altaf *et al.*, 2007), y la cascara se puede utilizar para la producción de  $\alpha$ - amilasa alcalina extracelular por la fermentación en estado sólido de cepas de *Bacillus* spp (Baysal *et al.*, 2008), o como absorbente de metales pesados como el Cu (II) de soluciones acuosas (Aydin *et al.*, 2008).

Varios estudios relacionan algunos componentes de la lenteja con una reducción en la incidencia de varios cánceres, por ejemplo un tipo de inhibidor de proteasas es capaz de inhibir el crecimiento de células de un adenocarcinoma de colon (Caccialupi *et al.*, 2010; Roy *et al.*, 2010).

#### 1.2. Sistemas de resistencia a patógenos y plagas de plantas

Existe una gran variedad de factores que pueden afectar al crecimiento y desarrollo de los cultivos vegetales, produciendo grandes pérdidas económicas en el sector de la agricultura. Dichos factores pueden ser abióticos, como la temperatura y otras condiciones climáticas o del suelo, pero también pueden ser bióticos como los patógenos y las plagas. Entre los patógenos que pueden atacar a las plantas se incluyen virus, bacterias, hongos y nematodos. En el caso de la lenteja probablemente la enfermedad más importante a nivel mundial sea la podredumbre o marchitamiento de la raíz (Muehlbauer et al., 1995; Rubiales y Fernandez-Aparicio, 2011) que está provocada por los hongos Fusarium oxysporium f. sp. lentis, Rhizoctonia solani y Sclerotinia sclerotiorum. Otra grave enfermedad que afecta al cultivo de lenteja es la provocada por el hongo Ascochyta lentis (syn. A. fabae f. sp. lentis, teleomorfo: Didimella lentis) que ataca tallos, hojas y vainas de la planta causando puntos necróticos y que puede infectar a las semillas. La base de datos Plant Virus Online (http://pvo.bio-mirror.cn/famly078.htm#Lens%20culinaris ) muestra una lista de 27 virus a los que la especie Lens culinaris es susceptible. En esta lista se encuentran virus de géneros como potivirus, luteovirus y cucumovirus. Una de las peores plagas por las que se ven amenazadas la lenteja y otras leguminosas en el área Mediterránea es el jopo, una mala hierba parásita de la especie Orobanche crenata (Muehlbauer et al., 1995). Por último, otros agentes muy dañinos son los insectos, teniendo gran importancia los áfidos (Aphis craccivora, Acyrthosiphon pisum), el gorgojo de la hoja (Sitona spp.), el gorgojo del grano (Bruchus lentis) y el gusano cortador (Agrostis ipsilon) (Stevenson et al., 2007).

Las plantas han desarrollado un evolucionado sistema inmune innato estratificado para resistir o contrarrestar el ataque de patógenos (Jones y Dangl, 2006). Un primer nivel en esta defensa innata consiste en un conjunto de proteínas de la membrana plasmática con un dominio extracelular, los

receptores de reconocimiento de patrones PRRs (Pattern Recognition Receptors), que perciben unos patrones moleculares asociados a patógenos (PAMPs: Pathogen Associated Molecular Patterns, también llamados MAMPs: Microbe-Associated Molecular Patterns) iniciando la inmunidad activada por PAMPs o PTI (PAMPs Triggered Immunity). Reconocen moléculas conservadas de los patógenos, tales como flagelina o quitina, y proporcionan resistencia de amplio espectro (Boller y Felix, 2009; Jacobs *et al.*, 2013).

Pero los patógenos han evolucionado para evitar el reconocimiento de los receptores o suprimir la PTI; una de las vías de esta supresión es mediante proteínas de genes de avirulencia llamadas efectores o factores de virulencia (avr). En respuesta a estas proteínas Avr las plantas han desarrollado un segundo nivel de defensa, que consiste en la producción de proteínas de resistencia codificadas por los genes de resistencia a enfermedades o genes *R*, que reconocen los efectores del patógeno iniciando la inmunidad activada por efectores o ETI (Effector Triggered Immunity) (Buscaill y Rivas, 2014; Cesari *et al.*, 2014; Chisholm *et al.*, 2006; Cui *et al.*, 2015; Dodds y Rathjen, 2010; Jones y Dangl, 2006; Maekawa *et al.*, 2011b; Monaghan y Zipfel, 2012; Qi e Innes, 2013; Sekhwal *et al.*, 2015; Takken y Goverse, 2012; Tameling y Takken, 2008), que amplifica las respuestas más débiles de resistencia basal PTI, y que a menudo lleva a la muerte celular programada localizada en repuesta a la infección produciendo una respuesta hipersensible (HR).

El reconocimiento de los efectores por las proteínas de resistencia puede ser directo, por interacción física, o indirectamente, mediante una proteína intermediaria llamada proteína protectora o "guardee".

#### 1.2.1 Genes R

Se han clonado genes *R* en numerosas especies de plantas, donde confieren resistencia a un amplio rango de patógenos como bacterias, virus, hongos, oomicetos y nematodos (Baker *et al.*, 1997; Bent, 1996; Dangl y Jones, 2001; De Young y Innes, 2006; Ellis *et al.*, 2000; Hammond-Kosack y Jones, 1997; McHale *et al.*, 2006; Meyers *et al.*, 2003). A pesar del amplio rango de patógenos, los genes *R* parecen codificar un grupo limitado de proteínas con dominios muy conservados.

La familia de genes *R* es una de las familias génicas más amplia en el genoma de plantas, con al menos cinco clases diferentes de genes conocidas hasta el momento (Marone *et al.,* 2013; Van Ooijen *et al.,* 2007). Que el número de secuencias sea muy variable de unas especies a otras, sin una clara correlación con la filogenia, sugiere mecanismos específicos en cada especie para la expansión y/o contracción de estos genes (Jacob *et al.,* 2013). Por ejemplo en *Medicago truncatula* se han encontrado más de 300 secuencias, aunque se estima que hay entre 400 y 500 en todo el genoma (Ameline-Torregrosa *et al.,* 2008). Otro estudio analiza 617 secuencias NBS-LRR en *Medicago,* de las cuales 490 reciben el nombre de genes regulares al tener el tamaño que consideran correcto en comparación con otras secuencias NBS-LRR, y 127 son llamados genes no regulares ya que son secuencias más cortas (2/3 de lo normal) de lo que esperaban (Song y Nan, 2014). *Medicago* tiene aproximadamente tres veces más secuencias NBS-LRR que las identificadas en *Arabidopsis thaliana,* donde hay 149 genes codificantes y 58 genes relacionados que carecen de las repeticiones ricas en leucina (LRR) (Meyers *et al.,* 1999; Meyers *et al.,* 2003) o 159 genes según Guo *et al.* (2011). Shang *et al.* (2009) sugieren que *Oryza sativa* puede tener 653 secuencias NBS-LRR, y Kang *et al.* (2012) encuentran 319 para *Glycine max.* 

Esta familia génica tiene una distribución desigual en el genoma de las plantas, ya que muchos genes *R* forman grupos locales en los cromosomas. Esto proporciona un reservorio de variación genética donde pueden evolucionar nuevas especificidades. La estructura en grupos de estos genes y la evolución de nuevas especificidades parecen producirse por duplicaciones, sobrecruzamiento desigual, recombinación ectópica, diversificación en la selección y conversión génica (Marone *et al.*, 2013; Michelmore y Meyers, 1998; Sun *et al.*, 2001; Young, 2000; Zhou *et al.*, 2004). Los genes que derivan de duplicaciones génicas son muy similares entre sí (Ameline-Torregrosa *et al.*, 2008; Jia *et al.*, 2015).

*M. truncatula* muestra dos grandes grupos de genes *R* en su genoma, uno en el cromosoma 3 y otro en el cromosoma 6, y un tercer grupo más reducido en el cromosoma 4 (Ameline-Torregrosa *et al.*, 2008). También hay grupos localizados en *Arabidopsis*, en los cromosomas At-1 y At-5 (Meyers *et al.*, 1999; Meyers *et al.*, 2003), en arroz, en el cromosoma 11 (Monosi *et al.*, 2004) y en *Glycine max*, en los cromosomas 3, 6, 13, 15, 16 y 18 (Kang *et al.*, 2012; Nepal y Beson, 2015). El tamaño de los grupos de genes *R* podría estar correlacionado con la densidad de elementos transponibles en el cromosoma y estos elementos pueden estar involucrados en la evolución de los grupos de genes, ya que generan inestabilidad genómica y probabilidades de recombinación (Ameline-Torregrosa *et al.*, 2008; Li *et al.*, 2010b).

Aunque muchos genes R se encuentran formando grupos, otros existen como secuencias aisladas. Estas secuencias aisladas pueden tener homólogos cercanos en alguna otra parte del genoma o pueden haber evolucionado independientemente. Aunque estas secuencias aisladas sean raras pueden ser fuentes de nuevos grupos de genes R (Ameline-Torregrosa *et al.*, 2008).

También se han identificado pseudogenes *R* en diferentes especies de plantas. Los pseudogenes son secuencias muy similares a genes conocidos pero que no producen una proteína funcional. Se originan por mutaciones, inserciones, deleciones y/o sustituciones, que interrumpen el marco de lectura por la aparición de codones de fin prematuros que incapacitan su función (Chandrasekaran y Betrán, 2008; Podlaha y Zhang, 2010; Tutar, 2012). También pueden ser secuencias parciales (Marone *et al.*, 2013). La abundancia de pseudogenes en un determinado genoma normalmente depende de las tasas de duplicación y pérdida de genes (Chandrasekaran y Betrán, 2008).

La mayor parte de los genes *R* clonados pertenecen a la familia NBS-LRR. Esta familia tiene unos motivos muy conservados en su secuencia, lo que ha permitido el diseño de cebadores degenerados basados en los motivos conservados de genes *R* conocidos para aislar secuencias análogas a genes de resistencia o *RGAs* (Resistance Gene Analogs) en otras especies vegetales mediante técnicas que utilizan la reacción en cadena de la polimerasa (PCR) (Bouktila *et al.*, 2014; Deng *et al.*, 2000; Donald *et al.*, 2002; Graham *et al.*, 2000; Meyers *et al.*, 2003; Palomino *et al.*, 2006; Timmerman-Vaughan *et al.*, 2000; Wan *et al.*, 2010; Xu *et al.*, 2005; Zhang *et al.*, 2011; Zhu *et al.*, 2002), ya sea con ADN total y utilizando geles de agarosa (Djebbi *et al.*, 2015; Kanazin *et al.*, 1996; Leister *et al.*, 1996; Reddy *et al.*, 2015; Wan *et al.*, 2012; Yaish *et al.*, 2004; Yu *et al.*, 1996) o geles de poliacrilamida (Chen *et al.*, 1998) para la separación de los fragmentos, o utilizando ADNc en la técnica llamada "NBS profiling" que combina la digestión del ADN con enzimas de restricción con una posterior amplificación por PCR (Brugmans *et al.*, 2008; Sanz *et al.*, 2013; van der Linden *et al.*, 2004).

Los *RGAs* pueden considerarse una fuente potencial de genes de resistencia (Hunger *et al.,* 2003) y como secuencias parciales de genes *R* (Noir *et al.,* 2001) basándonos en:

- 1\_ la alta identidad de secuencia con genes *R* conocidos y con *RGAs* de otras especies.
- 2\_la presencia de los motivos conservados característicos de los genes *R* NBS-LRR.
- 3\_ tener un marco de lectura abierto (ORF: Open Reading Frame) ininterrumpido.

#### Secuencias NBS-LRR

La clase mayoritaria de proteínas R es la de las proteínas NBS-LRR. Estas proteínas tienen un subdominio central de unión a nucleótidos o NBS (Nucleotide-Binding Site) y un dominio C-terminal rico en repeticiones de leucina o LRR (Leucine-Rich Repeat). El subdominio NBS, también denominado en ocasiones NB, forma parte de un dominio mayor llamado NB-ARC (Nucleotide Binding Adaptor) (Figura 1.8).



**Figura 1.8**. Representación esquemática de una proteína NBS-LRR típica. Los dominios y subdominios se representan como cajas coloreadas (modificada de Lukasik y Takken, 2009).

Basándose en modelos 3D, el dominio NB-ARC (Figura 1.9) parece contener tres subdominios que forman una cavidad cerrada de unión a nucleótidos cuando la proteína está en reposo (Slootweg *et al.*, 2013; Takken *et al.*, 2006):

- NBS: es una lámina beta de cinco hebras paralelas rodeadas de siete alfa hélices.
- ARC1: es un conjunto de cuatro hélices.
- ARC2: plegamiento en hélices dispuestas como alas.



**Figura 1.9**. Modelo en tres dimensiones de la estructura del motivo NB-ARC de Gpa2 de patata. Tomado de Slootweg *et al.* (2013).

Las proteínas NB-ARC forman una subclase de la superfamilia STAND (Signal Transduction ATPases with Numerous Domains), una clase de interruptores moleculares involucrados en multitud de procesos, tales como inmunidad, apoptosis (Apaf 1 y CED-4) y regulación transcripcional (Danot *et al.*, 2009). Debido a su estructura modular las proteínas STAND pueden funcionar simultáneamente como sensor, interruptor y factor de respuesta (Takken y Goverse, 2012).

Meyers *et al.* (1999) describieron una serie de motivos conservados que caracterizan las secuencias de estas proteínas como TIR (siglas del dominio Toll interleukin receptor) y No-TIR. El análisis se basaba en las secuencias de 14 genes de resistencia conocidos y más de 400 homólogos de 26 géneros de especies de mono y dicotiledóneas. Entre los genes TIR se incluían los genes de resistencia *N* (de *Nicotiana tabacum* y resistencia a TMV), *M* y *L6* (*Linum usitatissimum* a *Melampsora lini*), *RPP1* (de *Arabidopsis*) y *RPP5* y entre los No-TIR *RPS2* (de *Arabidopsis*), *RPM1*, *I2*, *Mi* (de tomate), *Dm3*, *Pi-B*, *Xa-1* (de arroz), *RPP8*, *RPS5* y *Prf*. Este conjunto de motivos es recogido por Lukasik y Takken (2009) en una revisión sobre proteínas de resistencia, aunque introducen algunas modificaciones en la nomenclatura designando a los genes TIR y No-TIR como TNL (siglas de TIR-NBS-LRR) y CNL (Coiled-coil-NBS-LRR) respectivamente.

La tabla 1.3 resume las secuencias consenso de los distintos motivos propuesto en estas publicaciones:

<b>Motivos</b> <sup>1</sup>	Grupos de genes		
	TNL	CNL	TNL/CNL
TIR-1	VFPSFRGEDVRDKTFLSH		
TIR-2	YASSSWCLDEL		
EDVID		EDVID	
TIR-3	VIPIFYKDPSDVRKQTGEFG		
hhGRExE			xxGRExE
Pre-P-Loop	VRMVGIWG		
P-Loop/ Kinase-1a/ WalkerA			GVGKTT
RNBS-A	FLENIRExSKKHGLEHLQKKLLSKLL	FDLxAWVCVSQxF	
Kinase-2/ Walter-B			LLVDDVW
RNBS-B/ Sensor1			GSRIIITTRD
RNBS-C			YEVxxLSEDEAWELFCKxAF
GLPL			CGGLPLA
RNBS-D	FLHIACFF	CFLYCALFPED	
MHD			MHDV

Tabla 1.3. Secuencia de los motivos conservados en las proteínas de resistencia.

1 Los motivos están ordenados, de arriba a abajo, desde la región N terminal a la C terminal. Se recogen las distintas denominaciones que han recibido algunos de estos motivos. Ver Figura 1.10.
Dentro de este grupo de motivos conservados los presentes en el dominio NB-ARC son (Lukasik y Takken, 2009; Meyers *et al.*, 1999; Pan *et al.*, 2000; Takken *et al.*, 2006; Van der Biezen y Jones, 1998a; Yaish *et al.*, 2004; Yue *et al.*, 2012) (Figura 1.10):

- P-loop (phosphate-binding loop) también llamado Walker A (Walker *et al.*, 1982) o Kinase 1a (Traut, 1994).
- Walker B (Walker et al., 1982) nombrado como Kinase 2 por Traut (1994).
- Kinase 3a (Traut, 1994).
- RNBS (Resistance Nucleotide Binding Site), seriados de A a D
- GLPL (nombre dado por los aminoácidos que lo componen más habitualmente)
- MHD (este motivo conservado también fue nombrado en función de los residuos presentes).

Los motivos "P-loop y Kinase 2" son importantes para la unión de ATP y GTP (Traut, 1994). El "P-loop" interacciona directamente con el fosfato del NTP ligado (Saraste *et al.*, 1990) y el motivo "Kinase 2" tiene un residuo de aspartato crítico para la coordinación del ión metálico Mg <sup>2+</sup> requerido para la reacción de fosfo-transferencia (Traut, 1994). El motivo "Kinase 3a" contiene una arginina que en otras proteínas interacciona con la base púrica del ATP (Traut, 1994). El motivo RNBS-B parece tener la misma función que el "Kinase 3a", pero son llamados de diferente manera porque hay una similitud posicional pero no de secuencia entre los motivos RNBS-B y "Kinase 3a" (Meyers *et al.*, 1999). El GLPL es un motivo hidrofóbico (Hammond-Kosack y Jones, 1997). Un estudio de mutaciones realizado por van Ooijen *et al.* (2008b) indica que la histidina del motivo MHD es un componente clave en la función del subdominio ARC2 de percepción del patógeno para la activación de la proteína.



**Figura 1.10.** Representación esquemática de una proteína NBS-LRR típica donde se señalan los motivos conservados con líneas discontinuas (modificada de Lukasik y Takken, 2009).

El subdominio NBS muestra algunos motivos conservados que se encuentran típicamente en proteínas de unión a ATP y GTP (Saraste *et al.*, 1990; Traut, 1994; Walker *et al.*, 1982) y parece ser el responsable de la unión e hidrólisis de ATP y GTP (Tameling *et al.*, 2002). Por otro lado, el subdominio ARC parece jugar un papel en el reclutamiento del dominio LRR (Rairdan y Moffett, 2006). Modelos de homología del dominio NB-ARC basados en Apaf 1 ligado a ADP muestran una estructura globular compacta (Takken y Goverse, 2012).

El dominio LRR de las proteínas R tiene una gran variabilidad de secuencia en cuanto al número de repeticiones, esto sugiere diferentes clases estructurales entre las proteínas NBS-LRR de plantas (Takken y Goverse, 2012). Parece que el dominio LRR tiene dos funciones, preservar la proteína en ausencia de patógeno y conferir especificidad en el reconocimiento de patógeno. Esta última función parece debida a la variabilidad de secuencia (Takken y Goverse, 2012).

Estudios estructurales de la proteína Lr10 de trigo (Sela *et al.*, 2012) muestran una forma en herradura altamente compacta. El dominio LRR muestra dos zonas claramente diferenciadas, una parte N-terminal que contiene un grupo de residuos cargados positivamente y una parte C-terminal que contiene un alto número de aminoácidos aromáticos que pueden participar en interacciones hidrofóbicas. Parece que la zona N-terminal del LRR podría intervenir en la transducción intramolecular de señales (Takken y Goverse, 2012), mientras que hay evidencias, basadas en mutaciones individuales de intercambio de dominio que modifican la especificidad, de que la zona C-terminal parece estar involucrada en el reconocimiento de patógenos (Bendahmane *et al.*, 2002; Famham y Baulcombe, 2006; Sela *et al.*, 2012).

Basándose en las secuencias N-terminales los NBS-LRR de plantas pueden clasificarse en dos subfamilias:

• **TIR-NBS-LRR** o **TNL**: tienen un dominio TIR que presenta homología con los dominios de señalización intracelular de los receptores Toll de *Drosophila* e interleuquina 1 humano (Hammond-Kosack y Jones, 1997; Meyers *et al.*, 1999; Pan *et al.*, 2000).

El pliegue del dominio TIR consiste en una lámina beta con cinco hebras paralelas rodeada por cinco hélices alfa.

El dominio TIR de la proteína L6 de lino muestra una unidad asimétrica de dos monómeros obteniéndose una forma globular (Bernoux *et al.*, 2011; Ve *et al*, 2011) que también parece tener la proteína AtTIR de *Arabidopsis* (Chan *et al.*, 2010). L6 y AtTIR contienen una hélice alfa D extra, característica única de los dominios TIR de plantas (Bernoux *et al.*, 2011).

• **No-TIR-NBS-LRR**: también llamados **CC-NBS-LRR** o **CNL** porque muchos contienen en la región N-terminal un dominio "coiled-coil" (CC) (Pan *et al.*, 2000), pero también pueden poseer un dominio "leucine zipper" (LZ) (Vidhyasekaran, 2007).

Estudios de estructura de la proteína Mla10 de cebada muestran la formación de un homodímero entrelazándose dos dominios CC (Figura 1.11b), cada uno con una estructura hélice-vuelta-hélice (Figura 1.11a), que forma una vara creando una superficie de interacción amplia (Maekawa *et al.*, 2011a). El estudio de la proteína RPM1 de *Arabidopsis* muestra que también forma un dímero y el análisis *in silico* de la proteína Lr10 de trigo espelta parece apoyar la estructura de hélice-vuelta-hélice (Sela *et al.*, 2012).



**Figura 1.11.** a) Secuencia del dominio CC de la proteína MLA10 que muestra la localización de las tres hélices. b) Esquema del entrelazado de dos dominios CC formando un dímero (tomado de Maekawa et *al.*, 2011a).

Los genes CNL están ampliamente distribuidos tanto en monocotiledóneas como en dicotiledóneas, mientras que los TNL parecen estar restringidos a especies dicotiledóneas (Hammond-Kosack y Jones, 1997; Meyers *et al.*, 1999). Se han identificado algunos genes con dominio TIR en arroz (*Oryza sativa*) pero no tienen dominio LRR y difieren mucho de los genes TNL típicos (Bai *et al.*, 2002). En 2009 Tarr y Alexander realizaron un estudio sobre TNL en monocotiledóneas estableciendo la presencia de TNL en angiospermas basales, sugiriendo su presencia en las plantas primigenias, habiéndose reducido significativamente en las monocotiledóneas actuales.

Algunos estudios sugieren que las clases CNL y TNL podrían estar más subdivididas basándose en motivos específicos, posiciones intrónicas y distribución genómica. Ameline-Torregrosa *et al.* (2008) en su estudio sobre *Medicago* concluyen que la subfamilia TNL es más diversa en condiciones de ordenación de dominios que la CNL, ya que aparecen 10 tipos de secuencias TIR (TNL, TN, N, NT, NL, TTNL, NTNL, TNLT, TNTNL, TNLTNL) pero sólo aparecen dos tipos de secuencias CC (CNL y CN). Esto puede deberse a la estructura exón-intrón de los TNL, que usualmente están codificados por múltiples exones, a diferencia de los CNL que generalmente sólo son codificados por uno (Bai *et al.*, 2002; Meyers *et al.*, 1999; Meyers *et al.*, 2003). Las combinaciones inusuales más habitualmente encontradas en las secuencias NBS tanto en *Medicago* como en *Arabidopsis* son secuencias carentes del dominio LRR y en el caso de TNLs además hay variaciones en el número de dominios. Estudios realizados con RPP2a (un TNTNL) sugieren que no es una quimera inactiva sino que puede funcionar como gen de resistencia (Meyers *et al.*, 2003; Sinapidou *et al.*, 2004).

21

Los TNLs y los CNLs además de tener diferentes dominios en la región N-terminal también tienen diferencias de secuencias en tres motivos del dominio NB-ARC: el motivo RNBS-C tiene baja similitud entre las dos subfamilias y en el motivo RNBS-A, que es parte del subdominio NBS y en el motivo RNBS-D, que pertenece al subdominio ARC2, tienen una secuencia completamente distinta en TNL y en CNL (Lukasik y Takken, 2009; McHale *et al.*, 2006; Meyers *et al.*, 2003). Además, Meyers *et al.* (1999) concluyen que el residuo final en el motivo "kinase 2" también puede usarse para predecir la subfamilia a la que pertenece la secuencia (al menos en el 95% de los casos), ya que aparece un residuo de triptófano en la clase CNL que en los TNLs corresponde a un residuo de ácido aspártico. Otra diferencia entre las dos subfamilias de NBS-LRRs está en el requerimiento de componentes en las vías de señalización corriente abajo (downstream) en la respuesta de defensa, ya que parece que las secuencias TNL (y un tipo de CNLs, HRT) requieren el complejo EDS1 (Enhanced Disease Susceptibility 1)/ PAD4 (Phytoalexin Deficient 4)/ SAG101 (Senescence Associated Gene 101) y las CNL requieren la proteína integrina asociada a membrana plasmática NDR1 (Non-race-specific Disease Resistance) (Aarts *et al.*, 1998; Bhattacharjee *et al.*, 2011; Glazebrook, 1999; Heidrich *et al.*, 2011; Hu *et al.*, 2005; Peart *et al.*, 2002; Rietz *et al.*, 2011; Zhu *et al.*, 2011).

Ameline-Torregrosa *et al.* (2008) identifican en la región corriente arriba (upstream) de los NBS-LRRs cuatro elementos reguladores implicados en respuesta a patógenos o estreses de plantas: cajas WBOX, asociadas con los factores de transcripción WRKY (Dong *et al.*, 2003); cajas CBF y DRE (Sakuma *et al.*, 2006), y el motivo GCC asociado a los factores de transcripción ERF (Ohme-Takagi *et al.*, 2000). Los elementos WBOX son los más numerosos, pudiendo aparecer varios. Los otros tres elementos promotores aparecen generalmente una sola vez por cada región corriente arriba (Ameline-Torregrosa *et al.*, 2008).

Hay estudios que sugieren que el dominio NB-ARC se comporta como interruptor molecular en las proteínas R, ya que define el estado de activación de la proteína dependiendo del nucleótido ligado, ADP en el estado inactivo y ATP en el estado activo (Rairdan y Moffett, 2006; Takken *et al.*, 2006; Tameling *et al.*, 2006). El subdominio ARC1 interviene en la interacción intramolecular con el LRR y el subdominio ARC2 transduce la percepción del patógeno por el LRR en la activación de las proteínas R (Rairdan y Moffett, 2006; Tameling *et al.*, 2006).

Takken y Goverse (2012) proponen una estructura general para todas las proteínas NBS-LRR basada en datos bioquímicos. En esta estructura el dominio NB-ARC interactúa con la parte N-terminal del dominio LRR, manteniendo a la proteína en una conformación cerrada en ausencia del efector producido por el patógeno (Ade *et al.*, 2007; Deslandes y Rivas, 2012; Moffett *et al.*, 2002; Van Ooijen *et al.*, 2008a). En este modelo el dominio CC interactúa simultáneamente con el NB-ARC y con el LRR creando una estructura compacta y los datos indican que en el caso del dominio TIR globular se produce una estructura similar. Parece ser que en ambos modelos el dominio N-terminal (CC/TIR) está próximo a la zona C-terminal del dominio LRR cuando la proteína está en reposo para mantener estable la proteína en un estado de auto-inhibición (Takken y Goverse, 2012).

Mantener el plegamiento correcto de las proteínas NBS-LRR requiere algunas moléculas como pueden ser HSP90 (Heat Shock Protein 90), co-chaperonas Rar1, PP5 (Protein Phosphatase 5) y SGT1 (Suppressor of the *G2* allele of *Skp 1*) y pequeñas HSPs que evitan el despliegue de la proteína y la agregación (Bieri, 2004; Holt, 2005; Muskett *et al.*, 2002; Van Ooijen *et al.*, 2010). Estas moléculas actúan como reguladores positivos del estado estable de las NBS-LRRs, pero SGT1 puede actuar

como efector negativo en la acumulación de la proteína cuando se une al complejo SFC (Skin-1-collin-F-box) (Catlett y Kaplan, 2006; Kadota *et al.*, 2008), que marca proteínas para su degradación en el proteosoma. Otros dos genes que actúan como efectores negativos de la actividad de NBS-LRR son *SRFR1* (Suppressor of rds 4-RDL1) y *CRT1* (Compromised Recognition of TCV) (Kim *et al.*, 2010; Kwon *et al.*, 2009; Li *et al.*, 2010c). La proteína CRT1 parece tener un papel en los primeros pasos del plegamiento de NBS-LRR y la función de SRFR1 no parece muy clara (Kang *et al.*, 2008; Kang *et al.*, 2010).

La liberación de la auto-inhibición de la proteína NBS-LRR desencadena un cambio conformacional en el dominio NB-ARC permitiendo el intercambio de ADP (del estado inactivo) por ATP (del estado activo) y así mostrar una estructura más abierta donde el dominio N-terminal y el dominio NBS quedan más expuestos para activar la señalización de defensa (Figura 1.12) (Heidrich *et al.,* 2012; Takken y Goverse, 2012).



**Figura 1.12.** Modelo de activación de la proteína NBS-LRR. En ausencia del patógeno la proteína NBS-LRR está en estado de reposo (ADP), en el cual el dominio LRR estabiliza la conformación cerrada. La plataforma de reconocimiento de la proteína AVR (triangulo negro) la proporciona la parte C-terminal del LRR junto al dominio CC/TIR, que podría estar vinculado a una proteína accesoria (nombrado como guardián o señuelo, en inglés "guardee or decoy"-G/D, hexágono amarillo). La percepción de AVR (directamente o vía G/D) cambia la superficie de interacción entre la parte N-terminal del LRR y del subdominio ARC2, liberando así la autoinhibición conferida por el LRR. Este cambio provoca un segundo cambio conformacional, alterando la interacción del dominio NB-ARC con el dominio CC/TIR y el LRR (estado inducido). En el estado activado el subdominio NBS está disponible para interactuar con su compañero de señalización corriente abajo. La hidrólisis del ATP retorna la proteína a su estado relajado (modificada de Lukasik y Takken, 2009).

Como ya se ha comentado antes, algunas proteínas R reconocen directamente el efector patógeno en una interacción gen a gen (Flor, 1955; Flor 1971). Las proteínas M y L6 de lino constituyen un ejemplo de este modelo, ya que reconocen a los factores de la roya del lino Avr1567 y AvrM (Catanzariti *et al.*, 2010; Dodds *et al.*, 2006). Pero otras proteínas de esta familia necesitan proteínas accesorias para el reconocimiento de los efectores patógenos, en un sistema de pares de genes. Por ejemplo para la resistencia contra *Pseudomonas syringae*, en tomate el gen *Prf* requiere la presencia de *Pto* para activar la resistencia, y en *Arabidopsis PRM1/PRS2* necesita el gen *RIN4* (Van der Biezen y Jones, 1998b; Van der Hoom y Kamoun, 2008). Estos estudios llevaron a la propuesta de un modelo llamado cebo-interruptor ("bait and switch" en inglés), en el que proteínas "guardee" o accesorias actúan como efectores de reconocimiento (Collier y Moffett, 2009; McDowell y Woffender, 2003). Las interacciones con proteínas accesorias pueden proporcionar gran flexibilidad en la percepción de patógenos y la subsiguiente señalización. Esto podría ser especialmente beneficioso si los efectores patógenos evolucionaran para perder asociación con la proteína accesoria "original".

Esta asociación por pares de genes NBS-LRR puede incluir parejas con dominios homotípicos y heterotípicos y pueden estar genéticamente vinculadas o desvinculadas (Cesari *et al.*, 2014; Eitas y Dangl, 2010).

Análisis basados en la mutación y en la estructura de los dominios CC y TIR de los receptores MLA10 de cebada y L6 de lino, respectivamente, sugieren que la homodimerización del N-terminal es necesaria para la activación de la defensa que finaliza en muerte celular (Bernoux *et al.*, 2011; Maekawa *et al.*, 2011a).

#### **1.3. Sintenia entre especies vegetales**

En el "*Dictionary of Gene Technology*" Kahl (2004) define la sintenia como la conservación del orden, secuencia y orientación de genes en el genoma de especies estrechamente relacionadas. Se han detectado regiones sinténicas en muchas familias de plantas. Mientras las secuencias génicas y su orden en el mapa están altamente conservadas, el tamaño, secuencia y composición de ADN de las secuencias intergénicas no lo están. Por lo tanto, la mayoría de la diversidad genética entre especies reside en las regiones intergénicas (Kahl, 2004).

Dentro de las leguminosas se ha demostrado la existencia de sintenia entre *Medicago* y otras especies. *M. truncatula, Medicago sativa, Pisum sativum, Lotus japonicus, Vigna radiata, Phaseolus vulgaris* y *Glycine. max* (Choi *et al.,* 2004), *Medicago sativa* con *Pisum sativum* (Kalo *et al.,* 2004), garbanzo, lupino, alfalfa, lenteja y haba con *M. truncatula* (Phan *et al.,* 2005), guisante y *M. truncatula* (Aubert *et al.,* 2006) y *M. truncatula* con lenteja (Phan *et al.,* 2007) *M. truncatula* y *Phaseolus vulgaris* además de otras especies vegetales (Hougaard *et al.,* 2008) y *Cicer,* soja, lenteja, *Trifolium* spp., *Cajanus cajan, Phaseolus mungo* con *Medicago truncatula* (Choudhary *et al.,* 2009).

En estudios con *Medicago* también se ha observado sintenia interna en grupos génicos en distintos cromosomas, como en el caso de algunos genes *R* NBS-LRR. Parece deberse a duplicaciones a gran escala o a transposiciones ocurridas durante de la evolución del genoma (Ameline-Torregrosa *et al.*, 2008; Chen *et al.*, 2015). Estas zonas sinténicas internas en *Medicago* podrían explicarse por un posible episodio temprano de poliploidía en las leguminosas (Cannon *et al.*, 2006).

Un estudio de mapeo genético en el género *Lens* hecho por Muehlbauer *et al.* (1989) muestra por primera vez evidencias de sintenia compartida entre *Lens* y *Pisum* ya que algunos de los grupos de ligamiento obtenidos en ese trabajo aparecían conservados en *Pisum*. Weeden *et al.* (1992) describen una relación genética entre guisante y lenteja estudiando marcadores moleculares. La creación de mapas genéticos con gran cantidad de marcadores moleculares facilita la comparación entre especies. Por ejemplo, un estudio de Elwood *et al.* (2008) compara *L. culinaris* y *Vicia faba* y muestra zonas de sintenia de ambas especies con *M. truncatula*.

Simon y Muehlbauer (1997) establecieron relaciones de sintenia por primera vez entre lenteja, garbanzo y guisante empleando isoenzimas. En 2006 Moffet realiza un estudio identificando regiones de conservación sinténica entre guisantes, judías y lenteja.

En 2011 se desarrollaron varios estudios de comparación de lenteja y *M. truncatula* utilizando ESTs, llegando a utilizar tecnología de segunda generación (Alo *et al.*, 2011; Kaur *et al.*, 2011).

Trabajos más recientes, que evidencian la alta conservación de sintenia entre *Lens* y *M. truncatula*, evaluaron la divergencia entre *L. culinaris* y *L. ervoides* y de ambas con el genoma secuenciado de *M. truncatula* a partir de SNP (Sharpe *et al.*, 2013) o con TOGs ("tentative orthologous genes") (Gujaria-Verma *et al.*, 2014). Este último también incluye una comparación con el genoma de garbanzo.

Han sido varios los trabajos que utilizan genes *R* o secuencias NBS-LRR para estudiar la sintenia entre plantas. Por ejemplo, Leister *et al.* (1996) analizaron genes de resistencia a patógenos de patata localizados en regiones sinténicas de cromosomas de tomate que no habían sido descritos hasta ese momento. En 2008 David *et al.* utilizaron secuencias de genes *R* para comparar *Phaseolus vulgaris* con las especies modelo *M. trucatula* y *L. japonicus*. También se han utilizados RGAs para estudiar la sintenia y microcolinearidad entre cacahuete, uva, soja, *Medicago* y *Arabidopsis* (Ratnaparkhe *et al.*, 2011). Los RGHs u homólogos a genes de resistencia a enfermedades también se han usado para comparar garbanzo con otras especies de leguminosas (Varshney *et al.*, 2013). Rispail *et al.* (2010) hicieron un estudio de sintenia y colinearidad de leguminosas grano con *M. trucatula* y *L. japonicus* utilizando secuencias NBS-LRR entre otros marcadores.

# **Objetivos**

#### **2. OBJETIVOS**

El objetivo general de esta Tesis ha sido el profundizar en el conocimiento de las secuencias de análogos a genes de resistencia presentes en las especies del género *Lens*. Concretamente en la familia de genes descrita de forma más general como genes *NBS-LRR*. Para ello se ha procedido a:

- Amplificar mediante PCR y aislar secuencias NBS-LRR de especies de lenteja tanto cultivada como silvestre, usando cebadores degenerados basados en motivos conservados de secuencias conocidas y cebadores no degenerados diseñados a partir de secuencias de otras especies vegetales.
- Caracterizar secuencias incluyéndolas en las diferentes clases de NBS-LRR y reconstruir su filogenia.
- Analizar las relaciones filogenéticas con las secuencias de otras especies, particularmente de leguminosas.
- Caracterizar las posibles agrupaciones génicas y zonas sinténicas de *Lens* con otras especies de leguminosas.

# Materiales y métodos

## **3. MATERIALES Y MÉTODOS**

### 3.1 Material vegetal

El ADN utilizado para este estudio se obtuvo en algunos casos de hoja y en otros de semilla seca de dos variedades de lenteja cultivada *Lens culinaris* Medik. y de las seis especies o subespecies silvestres del género *Lens: L. culinaris* subsp. *orientalis* BG16880 (Boiss) Ponert, *L. ervoides* BG16877 (Bring.) Grande, *L. lamottei* 244 Czfr., *L. nigricans* BG16873 (Bieb) Godr., *L. tomentosus* 133 Ladiz. y *L. odemensis* ILWL235 (Godr.) Ladiz.

Las dos variedades de *Lens culinaris* subsp. *culinaris* usadas fueron ILL5588 y el cultivar Lupa (obtenido por el Dr. Alonso Ponga a partir de la variedad pardina). *Lens culinaris* ILL5588 fue escogida por ser una variedad resistente a *Ascochyta,* mientras que *Lens culinaris* cv. Lupa se eligió por haber sido utilizada en el Área de Genética de la Universidad de León en varios cruzamientos con la forma silvestre *L. c. orientalis*.

### 3.2 Extracción de ADN genómico

Para la extracción de ADN genómico se han utilizado los artículos del "Dneasy<sup>®</sup> Plant Mini Kit" de Qiagen siguiendo el protocolo recomendado en el manual de uso (Dneasy Plant Mini Handbook, 1999). Para la extracción se utilizó el material vegetal pulverizado, por lo que antes de comenzar el proceso se preparaba la muestra según su procedencia. Si las muestras eran de semillas, éstas se molían poniéndolas entre dos placas de bronce y golpeándolas con un martillo. En cambio si las muestras eran de hojas, éstas se homogeneizaban introduciéndolas en nitrógeno líquido y triturándolas con una varilla de metal. Una vez pulverizado el material, se siguió el protocolo de extracción. En el último paso del protocolo se hizo una pequeña modificación cambiando el tampón incluido en el "kit" por agua "milliQ".

Para la posterior valoración de la calidad y cuantificación del ADN se dejaron pasar al menos tres horas a 4°C después de la extracción.

### 3.2.1 Valoración y cuantificación del ADN

Uno de los métodos de valoración de la calidad del ADN genómico extraído fue la electroforesis en gel de agarosa (D1 Low EEO de Pronadisa). Tanto para la electroforesis como para la preparación de los geles se utilizó el tampón TAE (tris, acetato, EDTA). La composición por litro en una concentración 50X es la siguiente:

Tris base	242 g
Ácido acético glacial	57,1 ml
EDTA 0,5 M pH 8,0	100 ml
Agua destilada hasta u	n litro

El tampón TAE se utilizó en una concentración final 1X.

Se cargaron 5  $\mu$ l de cada muestra en un gel de agarosa al 1% (p/v) preparado en tampón TAE 1X. Como tampón de carga se utilizó en todos los geles de agarosa el denominado "Blue juice" (0,025% de azul de bromofenol y 50% de glicerol). Se añadieron 5  $\mu$ l de tampón de carga por muestra. Junto con las muestras, pero en un carril diferente, se cargó la escalera comercial de fragmentos "1 Kb Plus DNA ladder" (Invitrogen) como referencia para estimar el tamaño de las bandas. La electroforesis se realizó a 80-90 voltios usando como referencia visual el frente del tampón de carga, la electroforesis se detuvo cuando había recorrido tres cuartas partes del gel.

Posteriormente los geles se tiñeron en agua con bromuro de etidio (0,5  $\mu$ g/ml) en un agitador oscilante a velocidad lenta durante 15 minutos. A continuación, se fotografiaron los geles en un transiluminador con sistema de fotografía Gel Doc <sup>®</sup> (BioRad<sup>®</sup>) de luz ultravioleta usando el programa Quantity One.

Un método complementario para cuantificar la cantidad y pureza de las muestras fue la utilización de un espectrofotómetro ND-1000 Nanodrop. Para ello se tuvieron en cuenta la concentración de ADN, medida en  $\mu g/\mu l$ , y la relación entre la absorbancia a 260 nm y a 280 nm.

### 3.3 Amplificación de ADN. Reacción en cadena de la polimerasa

La reacción en cadena de la polimerasa o PCR es una técnica empleada para amplificar fragmentos de ADN situados entre dos regiones de secuencia conocida (Mullis *et al.*, 1986; Mullis y Faloona, 1987). Como ADN molde para la reacción se utilizó el ADN genómico total y como cebadores secuencias diseñadas a partir de dominios conservados de la región NBS-LRR de análogos a genes de resistencia (*RGAs*). Las PCRs se llevaron a cabo en un termociclador Gene Amp PCR System 9700 de Applied Byosystems.

## 3.3.1. Cebadores

Para obtener las secuencias de tipo NBS-LRR en este trabajo se han utilizado cebadores tanto degenerados, diseñados a partir de secuencias conservadas de diversas especies de plantas, como no degenerados o específicos, diseñados para lenteja o para otras especies de leguminosas.

Los cebadores degenerados fueron diseñados a partir de dominios conservados de la región NB-ARC de *RGAs*. Los cebadores "forward" o directos s1, s2 (Leister *et al.*, 1996) y B1 (Rikvin *et al.*, 1999) se diseñaron a partir del dominio p-loop. Los cebadores "reverse" o inversos corresponden a dos dominios distintos, los cebadores as1, as3 (Leister *et al.*, 1996) y as4 (Kanazin *et al.*, 1996) fueron diseñados a partir del dominio GLPL y el cebador L2 (Yaish *et al.*, 2004) a partir del dominio MHD (Figura 3.1) de los genes *N*, *L6* y *RPS2*.





Las secuencias de los siete cebadores degenerados utilizados son las siguientes (Tabla 3.1):

Tabla 3.1. Secuencias de los cebadores degenerados utilizados.	

Cebador	Secuencia de oligonucleótidos (5-→ 3 <sup>°</sup> )
s1	GGTGGGGTTGGGAAGACAACG
s2	GGIGGIGTIGGIAAIACIAC
B1	GGIGGIRTIGGIAARACIAC
as1	CAACGCTAGTGGCAATCC
as3	IAGIGCIAGIGGIAGICC
as4	ARIGCTARIGGIARICC
L2	CCKNSAGYMNRTCRTGCAT

I: inosina; R: A/G; Y: C/T; K: G/T; N: A/C/G/T; M: A/C; S: G/T

Estos cebadores fueron emparejados en combinaciones como muestra la tabla 3.2:

C1 (I)   s1/as1     C2 (II)   s2/as3     C3 (III)   s2/as4     C4 (IV)   s1/as3     C5 (V)   s1/as4     C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	<b>Combinación</b> <sup>1</sup>	Cebadores
C2 (II)   s2/as3     C3 (III)   s2/as4     C4 (IV)   s1/as3     C5 (V)   s1/as4     C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	<b>C1</b> (I)	s1/as1
C3 (III)   s2/as4     C4 (IV)   s1/as3     C5 (V)   s1/as4     C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	C2 (II)	s2/as3
C4 (IV)   s1/as3     C5 (V)   s1/as4     C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	<b>C3</b> (III)	s2/as4
C5 (V)   s1/as4     C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	<b>C4</b> (IV)	s1/as3
C6 (VI)   s2/as1     C7   s1/L2     C8   s2/L2     C9   B1/L2	<b>C5</b> (∨)	s1/as4
C7 s1/L2 C8 s2/L2 C9 B1/L2	<b>C6</b> (VI)	s2/as1
C8 s2/L2 C9 B1/L2	C7	s1/L2
C9 B1/L2	C8	s2/L2
	С9	B1/L2

Tabla 3.2. Combinaciones de cebadores degenerados.

<sup>1</sup> Los números romanos entre paréntesis corresponden al orden dado por Palomino *et al.* (2006) para las combinaciones.

El tamaño de los productos de PCR que se esperaba en coincidencia con secuencias NBS en las combinaciones de C1 a C6 era de 510-550 pb, mientras que en las combinaciones C7, C8 y C9 era de aproximadamente 850 pb, de acuerdo con las secuencias de genes de resistencia *RPS2* de *Arabidopsis, N* de tabaco y *L6* de lino.

Las parejas de cebadores no degenerados utilizadas fueron obtenidas de secuencias consenso de distintas clases de *RGAs* de *Cicer* y *Vicia* publicadas por Palomino *et al.* (2006). El tamaño esperado de los productos de PCR obtenidos en estas especies se distribuía entre 430 y 510 pb. Las secuencias de los cebadores específicos son (Tabla 3.3):

Pareja de cebadores <sup>1</sup>	Cebador directo (5′→ 3′)	Cebador inverso (5′→ 3′)
<b>C10</b> (RGA01)	ACCCTTGCACAACTTGTTTAC	GCAATCCTCCACACTTTCTTG
<b>C11</b> (RGA03)	GTACAACCATGACACTATAA	TTTCTAACAATCTCTTTGCC
<b>C12</b> (RGA04)	GATAGTAGAATTGCTAACCATTT	CAATCCTTCACATTTCACCAC
<b>C13</b> (RGA05)	GCGATTTTGAATGTAGGAG	GGCAATCCCCCAGAATAC
<b>C14</b> (RGA10)	CAATTTCCTGTGTACTGCC	CCATTGGCATACCTTAGTA

Tabla 3.3. Secuencias de los cebadores no degenerados utilizados.

<sup>1</sup>Los nombres entre paréntesis son los utilizados en el trabajo de Palomino *et al*. (2009)

#### 3.3.2 Reacciones de amplificación

Las reacciones de amplificación se prepararon en tubos de 0,2 ml en un volumen final de 25  $\mu$ l con la siguiente composición:

Cebadores (5 pmol/ µl)	2 µl + 2 µl
Tampón 10X	2,5 µl
Mezcla de dNTP	2,5 µl
DNA (100 ng/ μl)	2,5 µl
Agua milliQ	13,25 µl
Taq polimerasa (5U/ μl)	0,25 μl

Utilizando la enzima "Taq DNA polymerase" (5-Prime).

La composición del tampón utilizado en la reacción fue: Tris-Cl 100 mM, pH 9,0; KCl 500 mM; Triton X-100 1% p/v; MgCl<sub>2</sub> 2,5 mM y la concentración de la mezcla de los cuatro dNTPs fue de 2,0 mM cada uno.

Las reacciones se llevaron a cabo siguiendo los programas de amplificación aplicados en otros estudios y teniendo en cuenta la temperatura óptima de anclaje de las distintas parejas de cebadores y la longitud de los productos esperados.

La temperatura óptima de anclaje es:

- 1. 44°C para las combinaciones C1 y C3.
- 2. 45°C para las combinaciones C7, C8 y C9.
- 3. 46°C para las combinaciones C2 y C6.
- 4. 55°C para las combinaciones C4, C5 y no degenerados (C10, C11, C12, C13 Y C14)

Los programas de PCR utilizados son:

Para las combinaciones C1, C2, C3 y C6 (tomado de Palomino et al., 2006)

Desnaturalización inicial: 94°C, 3 min 35 ciclos de amplificación 94°C, 1 min 46°C, 30 s 72°C, 30 s Finalización: 72°C, 7 min

Para las combinaciones C4 y C5 se utilizó un programa "touch down" (modificado de Palomino *et al.*, 2006)

Desnaturalización inicial: 94°C, 3 min 6 x 3 ciclos de amplificación Desnaturalización: 94°C, 1 min Anclaje: 30 s a 54°C, 53°C... hasta 49°C Extensión: 72°C, 30 s 21 ciclos de amplificación 94°C, 1 min 48°C, 30 s 72°C, 30 s Finalización: 72°C, 7 min

Para las parejas de cebadores no degenerados (C10, C11, C12, C13 Y C14) (tomado de Palomino *et al.*, 2009)

Desnaturalización inicial: 94°C, 5 min 35 ciclos de amplificación 94°C, 1 min 55°C, 1 min

#### 72°C, 2 min Finalización: 72°C, 10 min

Para las combinaciones C7, C8 y C9 (tomado de Yaish et al., 2004)

Desnaturalización inicial: 95°C, 1 min 38 ciclos de amplificación 94°C, 40 s 45°C, 1 min 72°C, 2 min Finalización: 72°C, 7 min

### 3.3.3 Valoración y cuantificación del ADN amplificado

Para valorar el resultado de las PCR se cargaron 5  $\mu$ l de cada muestra en un gel de agarosa al 2% (p/v) preparado en tampón TAE 1X al que se añadió o bromuro de etidio (5  $\mu$ l de una concentración de 10 mg/ml por cada 100 ml de agua) o Red Safe (0,1  $\mu$ l de la solución comercial por cada ml de TAE). Las electroforesis se realizaron a 80-100 voltios, usando como referencia visual el frente del tampón de carga, y se paraban cuando habían recorrido tres cuartas partes del gel. Posteriormente se fotografiaban según se ha descrito en el apartado anterior.

#### 3.4 Purificación de ADN de geles de agarosa

En algunos casos el producto obtenido de la amplificación de ADN sólo mostraba una banda en el gel, pero generalmente se obtenían varias bandas. En estos casos, con el fin de aislar las distintas bandas, las muestras de PCR se cargaban en un gel de agarosa de bajo punto de fusión (Boehring Mannheim) al 2% p/v, en las mismas condiciones que para visualizarlas después de la amplificación. Una vez que las muestras habían recorrido tres cuartas partes del gel, éste se ponía sobre un transiluminador de luz ultravioleta y se cortaba con un bisturí la sección del gel que contenía la banda de interés. Cada fragmento de gel cortado se introducía en un tubo tipo eppendorf.

La purificación de ADN desde los geles de agarosa se realizó siguiendo el protocolo recomendado para el producto "QIAquick<sup>®</sup> Kit extraction" (Qiagen).

Para verificar el éxito del aislamiento de cada banda se cargaban 5  $\mu$ l del producto resultante en un gel de agarosa al 2% en TAE 1X.

En caso de que la banda purificada apareciera débil en el gel se procedía a amplificarla de nuevo utilizando las mismas cantidades y el mismo programa de PCR que para la muestra inicial.

### 3.5 Clonación de los productos de PCR en pGEM®-T

Las muestras de ADN utilizadas para la ligación en este paso fueron el producto de PCR directamente cuando el gel mostraba una sola banda, o el producto resultante de la purificación de ADN cuando en éste aparecían varias bandas.

En el "kit pGEM<sup>®</sup>-T Easy Vector System" (Figura 3.2), de Promega Corporation, el vector viene previamente cortado con la enzima *Eco*RV y con una timidina 3' terminal añadida en ambos

extremos. Esta 3'-T mejora la eficiencia de la ligación del producto de PCR en el plásmido porque previene la recircularización del vector y aporta una base compatible con las A protuberantes en 3' que genera la polimerasa en el producto de PCR (Promega, 2007).



Figura 3.2. Vector de clonación pGEM®-T Easy Vector.

Siguiendo el manual técnico de Promega para "pGEM®-T y pGEM®-T Easy Vector Systems" se procedió a la ligación de las muestras de ADN y el vector en un volumen total de 10 µl, utilizando las siguientes cantidades:

Tampón de ligación rápida 2X	5 µl
Vector pGEM <sup>®</sup> -T Easy (50 ng)	1 µl
Producto de PCR	3 µl
ADN ligasa T4 (3U Weiss/ μl)	1 µl

Las reacciones de ligación se realizaron en tubos de 1,5 ml. El tiempo de reacción dependió de la temperatura a la que se llevara a cabo, siendo de 2 a 4 horas a temperatura ambiente o toda la noche a 4°C.

#### 3.6 Transformación en células competentes

Se obtuvieron células competentes siguiendo el protocolo basado en una modificación del sistema utilizado por Cohen *et al.* (1972). Dicho protocolo se describe a continuación:

- 1. Inocular una colonia de la bacteria *E. coli* DH5 $\alpha$  en 1,5 ml de medio LB e incubar a 37°C durante toda la noche.
- 2. Inocular 200  $\mu$ l de células en 7 ml de medio LB, esperar una hora hasta obtener una densidad óptica de 0,2 a 600 nm.
- 3. Centrifugar durante 8 minutos a 3100 xg. Retirar el sobrenadante con cuidado.
- 4. Resuspender en 3,4 ml de CaCl<sub>2</sub> 50 mM a 4°C.

- 5. Dejar 30 minutos en hielo. Centrifugar 8 minutos a 3100 xg.
- 6. Resuspender en 0,7 ml de CaCl<sub>2</sub> 50 mM a 4°C.
- 7. Mantener a 4°C de 12 a 48 horas.

Las células competentes obtenidas con este protocolo se deben utilizar dentro de las 48 horas posteriores a su elaboración pero también pueden utilizarse células competentes mantenidas a -80°C.

La transformación de células competentes de la cepa DH5 $\alpha$  se realizó según el siguiente protocolo:

- 1. Si se utilizan células competentes congeladas, deben dejarse descongelar en un baño de hielo.
- 2. Añadir 100  $\mu$ l de células competentes al tubo que contiene los 10  $\mu$ l de la mezcla de ligación especificada previamente.
- 3. Dejar 20-30 minutos en hielo.
- 4. Dar un choque térmico a 42°C de 45-50 segundos.
- 5. Poner en hielo 2 minutos.
- 6. Añadir 1 ml de medio LB.
- 7. Mezclar suavemente. Mantener una hora a 37°C con agitación suave.
- 8. Añadir 100 μl de X-Gal 5% y 100 μl de IPTG 100 mM al tubo de la transformación.
- 9. Extender la mezcla de transformación en varias placas con medio LB con ampicilina.
- 10. Poner a incubar las placas cerradas y boca abajo a 37°C.

Las placas no deben estar más de 20 horas a 37°C, ya que pasado ese tiempo la ampicilina se degrada por el calor y podrían crecer colonias que no contengan el plásmido.

El medio LB o medio de Luria-Bertani (Miller, 1972) se utilizó tanto en el cultivo como en la conservación de las cepas bacterianas transformadas. Su composición para un volumen final de un litro es:

Bacto triptona	10 g
Extracto de levadura	5 g
NaCl	10 g

Se ajustó el pH a 7,5 y se esterilizó en autoclave. Para hacer medio LB sólido se añadieron 15 g de agar antes de la esterilización.

En medios suplementados con ampicilina, ésta se añadía en una concentración de 50 mg/l cuando el medio esterilizado se había enfriado a 45°C.

El plásmido tiene el gen de resistencia a ampicilina para seleccionar las bacterias con el plásmido. Además, también presenta el gen que codifica la proteína  $\beta$ -Galactosidasa, que es funcional cuando el plásmido se recirculariza sin el inserto y no funcional cuando tiene el inserto. Por esta razón en este proceso se añade X-Gal (5-bromo-4-cloro-3-indolil- $\beta$ -D-galactósido) e IPTG (isopropil- $\beta$ -Dtiogalactósido), ya que el IPTG es inductor de la  $\beta$ -Galactosidasa y el X-Gal actúa como sustrato. Esto hace que las colonias que portan el inserto sean blancas y más grandes que las que no lo contienen que presentan un color azul.

### 3.6.1 Extracción de ADN plasmídico

Para comprobar la existencia del inserto en las colonias seleccionadas se llevó a cabo la extracción ultrarrápida de las colonias positivas:

- 1. Resembrar en placa de LB + ampicilina las colonias blancas. Dejar toda la noche a 37°C.
- 2. Transferir media colonia a un tubo "eppendorf".
- 3. Añadir 50 µl de EDTA 0,01 M, pH 8,0. Resuspender.
- 4. Añadir 50 µl de tampón de lisis 2X recién hecho. Mezclar por agitación vórtice.
- 5. Incubar a 70°C 5 min.
- 6. Dejar enfriar a temperatura ambiente 5 min.
- 7. Añadir 1,5  $\mu l$  de KCl 4M y 2  $\mu l$  de bromofenol 0,4 %
- 8. Agitar 5 segundos en vórtice.
- 9. Dejar en hielo 5 min.
- 10. Centrifugar 4 min a 4°C (a 21.000 xg).
- 11. Dejar 5 min (o más si es en nevera).
- 12. Cargar el gel con el sobrenadante.

Como control negativo del inserto se utilizaron algunas colonias azules.

El tampón de lisis se preparaba poco antes de ser usado. Las cantidades necesarias para 20 muestras son:

0,2 g Sacarosa 40 μl NaOH 5M 50 μl SDS 10% 900 μl agua MilliQ

En esta ocasión los geles utilizados presentaban una concentración de agarosa del 1% en TAE 1X. La electroforesis se realizó a 80-100 voltios.

Los geles se tiñeron con agua con bromuro de etidio (0,5  $\mu$ l/ml) en un agitador oscilante a velocidad lenta durante 15 minutos, o en TAE 1X con Red Safe también en agitador durante una hora (0,1  $\mu$ l por cada ml de TAE del gel).

#### 3.7 Secuenciación

Después de comprobar que las colonias contenían inserto se procedió a la secuenciación del mismo.

La muestra para la reacción de secuenciación se obtenía mediante este protocolo:

- 1. Inocular una parte de la colonia en un tubo con 20 ml de LB líquido al que se le han añadido 20  $\mu$ l de ampicilina (50 mg/l).
- 2. Incubar toda la noche a 37°C y agitación a 300 rpm.
- 3. Centrifugar 5 minutos a 3400 xg.
- 4. Eliminar el sobrenadante.

Las reacciones de secuenciación se llevaron a cabo en el Servicio de Análisis de Ácidos Nucleicos del Laboratorio de Técnicas Instrumentales (LTI) de la Universidad de León. Para la reacción de

secuenciación se utilizó el "DYEnamic ET Dye Terminator kit" (MegaBACE, Amersham). La amplificación se llevó a cabo en un termociclador MJ Research PTC-200, y los fragmentos de ADN se analizaron en un secuenciador MegaBACE 500 (Amersham Biosciences).

Para la visualización de los electroferogramas de secuenciación se utilizó la versión 1.45 del programa Chromas (McCarthy, 1998).

#### 3.8 Nomenclatura de las secuencias

Se ha dado un nombre identificativo a cada una de las secuencias obtenidas en este trabajo a partir de los siguientes criterios. Primero, tres letras que identifican la especie y, en caso de ser necesario, la subespecie de lenteja a la que corresponde la muestra, seguidas, en el caso de *L. culinaris* subespecie *culinaris*, por una cuarta letra mayúscula (A o B) que diferencia la variedad. El código de letras utilizado es el siguiente:

- LccA : L. culinaris subespecie culinaris variedad ILL5588
- LccB : L. culinaris subespecie culinaris variedad Lupa
- Lco : L. culinaris subespecie orientalis
- Ler : L. ervoides
- Lni : L. nigricans
- Lla : L. lamottei
- Lto: L. tomentosus
- Lod : L. odemensis

Después de identificar la especie, en el nombre de la secuencia aparece un código de tres números, que hace referencia a la muestra de transformación de la que se ha obtenido la secuencia, correspondiendo el número antes del punto a la tanda del experimento de transformación y los dos números situados después del punto a la colonia de *E. coli* que contenía el inserto. A continuación se identifica la combinación de cebadores mediante el nombre dado en las tablas 3.2 y 3.3, que consiste en una C mayúscula, seguido por el número correspondiente. Por último, en el caso de la secuencia LccB2.06C10, que contiene dos subdominios NBS seguidos (uno de ellos incompleto), se añade un guion bajo seguido de un 1 o un 2, esto indica el fragmento de secuencia al ser dividida quedando como LccB2.06C10\_1 y LccB2.06C10\_2.

En la figura 3.3 se muestra un ejemplo de la nomenclatura creada para identificar las secuencias de este estudio.



Figura 3.3. Desglose de la nomenclatura utilizada para las secuencias obtenidas en este estudio.

### 3.9 Conservación de clones de interés

Una vez secuenciadas las muestras, se procedió a la conservación de las colonias de *E. coli* que contenían fragmentos de ADN similares a secuencias NBS-LRR conocidas. Para ello se utilizó glicerol como crioprotector, siguiendo el protocolo descrito a continuación:

- 1. En cabina de flujo laminar, se toma parte de la colonia seleccionada con un asa de siembra y se resuspende en un tubo con 3 ml de LB líquido suplementado con ampicilina.
- 2. Se incuba a 37°C y con una agitación de 300 rpm toda la noche
- 3. Al día siguiente, en cabina de flujo laminar se añaden 800 μl del resultado de la incubación a un tubo "eppendorf "que contiene 200 μl de glicerol.
- 4. Se agita vigorosamente para que el cultivo bacteriano se mezcle bien con el glicerol.
- 5. Se almacena el tubo a la temperatura correspondiente debidamente etiquetado.

Para cada colonia se prepararon tres tubos "eppendorf," que fueron almacenados a diferentes temperaturas. Uno de ellos a -20°C y los otros dos a -80°C.

#### 3.10 Análisis de secuencias

El análisis de las secuencias obtenidas fue realizado con el programa MEGA 5.2 (Tamura *et al.*, 2011). Dicho análisis consistió en el alineamiento y agrupación de secuencias por similitud, utilizando para ello los programas de alineamientos múltiples ClustalW (Thompson *et al.*, 1994) y MUSCLE (Edgar, 2004).

Las secuencias obtenidas en este estudio fueron comparadas con las secuencias generales presentes en la base de datos Nucleotide Database del NCBI (mediante la función BLAST) para determinar la similitud con NBS-LRR. Posteriormente, tanto las 130 secuencias diferentes obtenidas en este estudio como las 32 secuencias de *Lens* obtenidas por el Dr. Yaish en un estudio anterior realizado en el área de Genética de la Universidad de León en 2004, se compararon con secuencias específicas de *Medicago truncatula, Glycine max, Phaseolus vulgaris* y *Cicer arietinum.* En el caso de *Medicago, Glycine* y *Phaseolus* se obtuvieron datos de genes anotados a través de la base de datos Phytozome 5.0 mientras que los datos de comparación con *Cicer* se obtuvieron del Nucleotide Database (NCBI, National Center for Biotechnology information).

Debido a que los nombres descriptivos de las secuencias obtenidas desde la base de datos del NCBI eran demasiado largos y dificultaban el manejo de las muestras se han acortado, utilizando como nombre el número identificador gi correspondiente seguido entre paréntesis de la abreviatura "ca" para la especie *Cicer arietinum* o de las tres letras correspondientes a la especie de *Lens* en el caso de las secuencias obtenidas previamente por el Dr. Yaish. El código de tres letras para las especies de *Lens* es el mismo utilizado para nombrar las secuencias obtenidas en este trabajo.

Algunos de los genes anotados de *Medicago truncatula* y dos de las secuencias de *Cicer arietinum* tienen dos o tres dominios NBS en su secuencia, por lo que, una vez alineadas con las secuencias obtenidas en este trabajo, esos dominios han sido numerados con un número del 1 al 3 entre paréntesis. Para la elaboración de los árboles filogenéticos se han utilizado únicamente los fragmentos más similares a las secuencias de lenteja estudiadas.

Se consideraron las secuencias con un exponente de "e-value" de -60 o menor (ver archivo "Excel" A2 que se incluye en el CD anexo) y se realizaron árboles con una selección de secuencias de *Lens* (tanto de este trabajo como del anterior del Dr. Yaish) y cada una de las especies comparadas. Para la realización de estos árboles, para cada secuencia de *Lens* se tuvieron en cuenta solamente los tres primeros resultados obtenidos en el BLAST de cada secuencia analizada, debido a que el gran volumen de datos dificultaba los análisis. Para simplificar los esquemas, al hacer los árboles también se dividieron las secuencias comparadas en las dos grandes subfamilias de NBS-LRR, TNL y CNL.

Además de la creación de árboles filogenéticos por el método estadístico Neighbor-Joining, usando el modelo de Kimura-2 parámetros (Kimura, 1980) y para una mayor fiabilidad de las posiciones obtenidas, se aplicó un "bootstrap" de 1000 repeticiones.

Se crearon secuencias consenso de los grupos de secuencias mediante el programa WebLogo (Crooks *et al.,* 2004), versión 2.8.2 (08, Sept., 2005) en sus dos alternativas, frecuencias y "bits", lo que facilitó el análisis del grado de conservación de los motivos señalados para la región NBS-LRR.

La secuencia consenso obtenida a partir del programa consiste en filas de letras, una columna por cada posición en la secuencia, donde la altura de cada letra indica la frecuencia relativa de cada aminoácido o nucleótido en esa posición mientras que la altura total de la columna representa la conservación, medida en bits, para esa posición en particular. Las letras de cada columna están ordenadas de mayor a menor frecuencia.

El programa asigna colores a las letras en función de las propiedades químicas (Crooks *et al.*, 2004), en este caso las secuencias consenso se obtuvieron en secuencia de aminoácidos y los colores se basan en grupos: verde para los aminoácidos polares (G, S, T, Y, C) excepto para Q y N que se representan en rosa, en rojo aparecen los aminoácidos ácidos (D, E), azul para los básicos (K,R, H) y por último los aminoácidos hidrofóbicos (A, V, L, I, P, W, F, M) en negro.

Mediante una matriz de distancias sin estandarización de datos, obtenida con el programa MEGA 5.2 y un análisis de componentes principales (PCoA de sus siglas en inglés) se obtuvieron los valores de variación e independencia entre los grupos de las secuencias obtenidas en este estudio cuando eran añadidas las secuencias, de las especies *M. truncatula* y *C. arietinum*, que mostraban alta similitud

con ellas. Utilizando los datos obtenidos para los tres primeros ejes se realizaron gráficos en 3D a partir del programa STATISTICA (StatSoft, Inc, 1995).

Para determinar la posición cromosómica de las secuencias NBS-LRR homólogas a las de lenteja, tanto en *Medicago truncatula* como en *Glycine max*, se utilizó la aplicación implementada en la web LIS (Legume Information System, http://cvit.comparative-legumes.org/), utilizando como base de datos Medicago truncatula v 3.5.1 y Glyma1.01 (=Glyma.Wm82.a1), con un blast del tipo blastx (nucleótido  $\rightarrow$  proteína), un valor de "e" de 1e<sup>-20</sup> y un porcentaje de identidad mínima del 60%. Para la visualización de los datos se ha utilizado el programa CViT (Chromosome Visualization Tool). En el análisis se han utilizado las secuencias de lenteja seleccionadas para los árboles comparativos con otras especies.

# Resultados

#### **4. RESULTADOS**

#### 4.1 Resultados de la secuenciación

En este trabajo se han utilizado 14 combinaciones diferentes de cebadores para la amplificación de secuencias NBS-LRR a partir de ADN total de lenteja. El producto de la amplificación de 13 de las 14 combinaciones utilizadas ha mostrado diversos patrones de bandas en los geles de agarosa, bien de una única banda o de varias bandas (Figura 4.1). La única combinación que no proporcionó un resultado claro fue la C8 (combinación de cebadores s2/L2, ver sección 3.3.1 en Materiales y Métodos), dado que no se obtuvieron amplicones para su posterior secuenciación.



**Figura 4.1.** Geles de agarosa donde se muestran los patrones de bandas obtenidos con cada una de las combinaciones de cebadores. En todos los casos se pueden observar los ocho carriles correspondientes a las especies y cultivares de lenteja, excepto en el caso de C7, C8 y C9 en la que sólo aparece un carril que corresponde a la muestra *Lens culinaris* subsp. *culinaris* cultivar Lupa. El orden de las muestras es: *L. culinaris* subsp. *culinaris* LL5588, *L. culinaris* subsp. *orientalis*, *L. ervoides*, *L. nigricans*, *L. culinaris* subsp. *culinaris* Lupa, *L. lamottei*, *L. tomentosus*, *L. odemensis*.

Todas las combinaciones formadas por los cebadores degenerados incluían como cebador directo una secuencia derivada del motivo conservado p-loop. La diferencia principal radicaba en el cebador inverso, puesto que en las secuencias obtenidas con las combinaciones C1 a C6 está basado en el dominio conservado GLPL, mientras que en C7, C8 y C9 el dominio del que deriva el cebador es MHD. En el caso de las combinaciones de C1 a C6, la banda de interés esperada era la que tenía un tamaño de 510 a 550 pb mientras que para las combinaciones C7, C8 y C9 tenían un tamaño de aproximadamente 850 pb. El tamaño esperado en las combinaciones de C10 a C14 estaba entre los 430 y los 510 pb.

Con el ADN amplificado por las 13 combinaciones que dieron resultados positivos, se procedió a la clonación y transformación de células competentes de *E. coli* para la secuenciación de los fragmentos que presentaban el tamaño esperado. En el caso de las combinaciones donde la banda de interés fue única, o muy intensa con respecto a las demás, el ADN utilizado en la clonación fue directamente el producto de PCR, pero en caso de patrones con varias bandas se procedió a aislar la banda del tamaño deseado desde el gel y purificarla antes de la clonación. La única excepción se hizo con la combinación C9, que muestra dos bandas de aproximadamente 850 nucleótidos donde se utilizó el producto de PCR como si fuera una única banda.

A partir de los clones de *E. coli* se obtuvieron 268 secuencias, de las cuales 129 correspondían a la especie cultivada *L. culinaris* subsp. *culinaris*, 23 de ellas de la accesión ILL5588 y las otras 106 del cultivar Lupa. Las restantes 139 secuencias se repartieron entre las especies silvestres: 28 de *L. culinaris* subsp. *orientalis*, 19 de *L. ervoides*, 21 de *L. nigricans*, 25 de *L. lamottei*, 28 de *L. tomentosus* y 18 de *L. odemensis* (Tabla 4.1).

Tabla 4	4.1.	Número	de	muestras	totales	analizadas	mediante	secuenciación	desglosadas	por	especie	y por
combir	nació	on de ceba	ado	res <sup>1</sup>								

$\searrow$	C1	C2	C3	C4	C5	<b>C6</b>	C7	<b>C9</b>	C10	C11	C12	C13	C14	total
LccA <sup>2</sup>	1	2	2	np³	4	1	np	np	6	2	3	-	2	23
LccB	2	3	3	np	np	30	1	17	46	1	1	1	1	106
Lco	1	1	2	2	4	1	np	np	5	2	3	6	1	28
Ler	3	-	1	2	np	1	np	np	2	1	-	8	1	19
Lni	3	1	1	np	4	1	np	np	2	3	3	3	-	21
Lla	2	2	3	-	np	1	np	np	4	2	9	1	1	25
Lto	2	-	4	np	3	1	np	np	7	2	2	4	3	28
Lod	2	3	2	np	np	1	np	np	3	2	1	-	4	18
total	16	12	18	4	15	37	1	17	75	15	22	23	13	268

<sup>1</sup>Excepto C8 de la que no se obtuvieron datos. Las combinaciones de C1 a C9 corresponden a parejas de cebadores degenerados y las combinaciones de C10 a C14 a cebadores no degenerados.

<sup>2</sup> LccA: *L. culinaris* subsp. *culinaris* ILL5588; LccB: *L. culinaris* subsp. *culinaris* Lupa; Lco: *L. culinaris* subsp. *orientalis*; Ler: *L. ervoides*; Lni: *L. nigricans*; Lla: *L. lamottei*; Lto: *L. tomentosus*; Lod: *L. odemensis*.

<sup>3</sup> np: no probado. -: no se obtuvieron amplificaciones.

Debido a que el cultivar Lupa de *L. culinaris* subsp. *culinaris* se ha utilizado ampliamente en otros trabajos del Área de Genética de la Universidad de León fue la muestra principal de este estudio. En

el caso de las combinaciones C7, C8 y C9 fue la única muestra analizada y en otros casos, como las combinaciones C6 y C10, fue amplificada y secuenciada en mayor número que las demás muestras. Se escogieron las combinaciones de cebadores C6, C9 y C10 para obtener un mayor número de muestras de *L. culinaris* subsp. *culinaris* cultivar Lupa por la variedad de secuencias que aportaban.

De las 268 secuencias obtenidas, 212 mostraron alta similitud con secuencias NBS-LRR, tanto de lenteja como de otras especies vegetales, al compararlas mediante el programa BLAST del NCBI (Tabla 4.2 y anexo A1). Las otras 56 fueron secuencias que presentaban similitudes con diferentes elementos genéticos. La mayoría mostraron similitud con secuencias no clasificadas todavía o con retrotransposones de tipo Ogre, secuencias de retrotransposones codificantes de proteínas gag de *Medicago truncatula* y *Glycine max*, y transposones de tipo Mariner de *Pisum sativum*. En el caso de la combinación de cebadores C4, gran parte de la secuencia de aminoácidos coincide con proteínas de unión a ácidos nucleicos y a una serina-carboxipepsidasa de *Medicago*, pero en algunos casos la coincidencia consistió en pequeños fragmentos de nuestra secuencia con secuencias de *Medicago* sin caracterizar. De estas 56 secuencias, 51 tenían ambos cebadores mientras que en cinco aparecía sólo uno de ellos. Este tipo de secuencias se ha obtenido con varias de las combinaciones de cebadores estudiadas pero en el caso de las combinaciones C4, C5 y C7 (degenerados) ninguna de las secuencias obtenidas presentó similitud con las del tipo NBS-LRR.

$\sim$	C1 <sup>2</sup>	C2	С3	<b>C6</b>	<b>C9</b>	C10	C11	C12	C13	C14	total
LccA <sup>1</sup>	1	2	2	1	np	6	1	3	-	2	18
LccB	2	1	3	30	11	46	-	1	-	1	95
Lco	1	1	1	1	np	5	1	2	4	1	17
Ler	3	-	1	1	np	2	-	-	6	1	14
Lni	3	1	1	1	np	2	2	2	3	-	15
Lla	1	2	3	1	np	4	2	5	1	-	19
Lto	1	-	3	1	np	7	2	1	3	3	21
Lod	-	1	2	-	np	3	2	1	-	4	13
total	12	8	16	36	11	75	10	15	17	12	212

**Tabla 4.2**. Número de secuencias que muestran alta similitud con NBS-LRR, desglosadas por especie y por combinación de cebadores.

<sup>1</sup>Véase leyenda de la tabla 4.1.

<sup>2</sup> Además de C8, de la que no se obtuvieron datos, también se omiten las columnas correspondientes a C4, C5 y
C7 ya que ninguna de las secuencias obtenidas mostró similitud con secuencias NBS-LRR.

Veintiuna de las secuencias similares a NBS-LRR fueron secuencias parciales, ya que estaban incompletas por alguno de sus extremos, faltándoles alguno de los dos cebadores, siendo mucho más frecuente la ausencia del cebador directo.

Al analizar las secuencias se encontraron varios casos en los que dos muestras o más eran idénticas (llegando a encontrar hasta seis veces la misma secuencia). Estas secuencias idénticas podían aparecer en distintas especies o en muestras de la lenteja cultivada y su subespecie silvestre, pero lo más frecuente es que fueran amplicones de una misma especie. A veces las secuencias idénticas se han obtenido utilizando combinaciones distintas de cebadores. Por tanto, debido a que las posiciones de los cebadores eran diferentes (diferentes motivos conservados de secuencias NBS),

algunas secuencias tenían una longitud menor que otras, pero eran idénticas a lo largo de la secuencia común. Una de estas secuencias más cortas (Lod1.02C14) coincidió en toda su longitud con dos muestras de mayor tamaño que no eran idénticas entre sí, puesto que tenían alguna diferencia de nucleótidos en la zona que se encontraba antes del cebador directo de la muestra más corta.

Para la creación de los árboles sólo se tuvo en cuenta una secuencia de cada grupo de muestras idénticas y en las ocasiones en las que había diferencia de tamaño se escogió siempre la de mayor longitud.

La figura 4.2 muestra la distribución del número de amplificaciones que se realizaron del ADN de la especie *L. culinaris* subsp. *culinaris* Lupa con las combinaciones C6 y C10 y el número de secuencias diferentes similares a NBS-LRR obtenidas



**Figura 4.2.** Gráfica de distribución que muestra el aumento de secuencias diferentes con respecto al número de amplificaciones realizadas en *L. culinaris* subsp. *culinaris* Lupa con las combinaciones C6 y C10.

#### 4.2 Pseudogenes

Dado que el fragmento amplificado en este estudio corresponde a un único exón se esperaba un marco de lectura continuo pero dentro de las secuencias con similitud a NBS-LRR se encontraron varias muestras que no tenían un marco de lectura correcto, bien sea por la presencia de codones prematuros de fin, por la falta o la inclusión de algún nucleótido que alteraba el marco de lectura en la zona secuenciada, la deleción de alguna zona dentro de la secuencia, o varias de estas alteraciones en la misma muestra. A estas secuencias las consideramos genéricamente como pseudogenes.

Un grupo de 14 secuencias mostraba una amplia deleción central (153 nucleótidos) en relación al alineamiento con las secuencias resultantes obtenidas de la base de datos del NCBI. Observando las secuencias que aparecieron de la base de datos y las secuencias completas obtenidas, pudo observarse que se repetía un grupo de cinco nucleótidos (AGTTG) en ambos extremos de la deleción. Dos de estas secuencias mantenían un marco de lectura correcto mientras que las 12 restantes habían perdido un nucleótido en la posición 104. Estas 14 secuencias mostraron otra diferencia con respecto a las secuencias de su mismo grupo que no presentan la deleción: en la posición 212 (211 en las que ya habían perdido un nucleótido) de la secuencia de nucleótidos había una T en las

secuencias con deleción, mientras que en las normales aparecía una C, dando lugar a leucina (L) en la secuencia de aminoácidos de las muestras con deleción, mientras que en las muestras con la secuencia completa el aminoácido que aparecía era prolina (P) (Figura 4.3).

Lto1.05C10	<b><u>ACCCTTGCACAACTTGTTTAC</u>AATGAAGAAAAGTTGAGCAACATTTTGATTTCAAAGTCTGGGTTTGTGTAT</b>
LccB6. 07C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAAATAGTTCAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
LccA4. 02C10	<b><u>ACCCTTGCACAACTTGTTTAC</u></b> AATGATGAAAATAGTTCAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
LccB3.09C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAGCATTTTGATTTCAAAGCTTGGGTTTGTGTGT
LccA3.04C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAACATTTTGATTTCAAAGCACGGGTTTGTGTGT
LccA3. 07C10	<b><u>ACCCTTGCACAACTTGTTTAC</u></b> AATGATGAAATAGTTCAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lco3.01C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATGGTTCAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lco3.03C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lco3.07C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAACATTCTGATTTCAAAGCATGGGTTTGTGTGT
LccB3. 02C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
LccB6. 18C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
LccB6. 20C10	<b>ACCCTTGCACAACTTGTTTAC</b> TATGATGAAAATAGTTCAAACAACATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lto3.02C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lto3.03C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTTCAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lto3.08C10	<b>ACCCTTGCACAACTTGTTTAC</b> AATGATGAAATAGTACAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lto3.09C10	<b><u>ACCCTTGCACAACTTGTTTAC</u></b> AATGATGAAATAGTTCAACAGCATTTTGATTTCAAAGCATGGGTTTGTGTGT
Lto1.05C10	CAGAGGACTTTGATGTTGTGAGAGTAACAAAGTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTTC
LccB6. 07C10	CTGAGGACTTTGATGTTGTGAGAGTAACTAAGTCTCTCCTTGAATCTGTTATTAGAAATACAACATCTGCGTC
LccA4. 02C10	CTGAGGACTTTGATGTTGTGAGAGTAACTAAGTCTCTCCTTGGATCTGTTGTTAGAAATACAACATCGGCTGC
LccB3.09C10	CTGAGGACTTCAATGTTGTGGGAGTAACCA <b>A</b> GTCTCTCCTTGAATCTGTCGTTAGAAATACAACATCTGCTGC
LccA3.04C10	CTCAGGACTTCGATGTTGCGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
LccA3.07C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
Lco3.01C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
Lco3.03C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
Lco3.07C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
LccB3. 02C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTAAATCTGTTGTTAGAAATACAACATCGGCTGC
LccB6. 18C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACGACATCGGCTGC
LccB6. 20C10	CTCAGGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
Lto3.02C10	CTCAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCCGTTGTTAGAAATACAACATCGGCTGC
Lto3.03C10	CTCAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
Lto3.08C10	CTCAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
1 +03 00010	
L103.07010	CTCAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC
2103.09010	
Lto1. 05C10	TTCAAAGGTATGGGGAAAGCGATAATCTTGATATCTTCGAGTGGGAGTGAAAAAAAA
Lto1. 05C10 LccB6. 07C10	TTCAAAGGTATGGGGAAAGCGATAATCTTGATATCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCG <b>AGTTG</b> AATTAAAAAAAAACACGAGGATGAAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAAGAAACACGAGGAGAAAAAGA TTCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAAGAAACACCAAGGGAGAAAAAGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccB3. 09C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCG <b>AGTTG</b> AATCAGTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAAAAAAACACGAGGGATGAAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAAGAAACACCCAAGGGAGAAAAAGA TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCG <b>A</b> TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCCA <b>A</b>
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccB3. 09C10 LccA3. 04C10	TTCAAAGGTATGGGAAAGCGATAATCTTGGATATCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAAAAAACACACGAGGATGAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACTCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGGATATCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGACAATCTTGGATATTCTTCGAGTTGAATTAAAAAAACACACGAGGATGAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACTCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCG <b>AGTTG</b> AATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAATAAAAACACGGAGGATGAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCG <b>AGTTG</b> AATTAAAGAAACACCCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCG <b>A</b>
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lco3. 01C10 Lco3. 03C10	TTCAAAGGTATGGGAAAGCGATAATCTTGGATATCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGACAATCTTGGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccB3. 09C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10 Lcc3. 01C10 Lco3. 03C10 Lco3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccB3. 09C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lco3. 03C10 Lco3. 07C10 Lco3. 07C10 LccB3. 02C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGTATTCTTCGAGTTGAATTAAAAAAACACGAGGATGAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACCTCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA3. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lco3. 03C10 Lco3. 07C10 LccB3. 02C10 LccB3. 02C10 LccB6. 18C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGTATTCTTCGAGTTGAATTAAAAAAACACGAGGATGAAAAGA TTCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACCTCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCGATAAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 01C10 Lcc3. 03C10 LccB3. 02C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGTATTCTTCGAGTTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 01C10 Lcc3. 03C10 Lcc3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 02C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAACACACGAGGATGAAAAGA TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACTCAAGGGAGAAAAGA TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10	TTCAAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCTCCATGTTGTTAGAAATACAACATCGGCTGC     TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAAAAAACACACGAGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACTCAAGGGAGAAAAGA     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 Lcc3. 01C10 Lcc3. 03C10 Lcc3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 02C10 Lto3. 03C10 Lto3. 03C10	TTCAAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC     TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 03C10 Lcc3. 03C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10	TTCAAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 03C10 Lcc3. 03C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10	TTCAAAGACTTCGATGTTGTGAGAGTAACAA-GTCTCCCTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTAGAAATACAACATCGGCTGC TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 03C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCGAGTTGAATCTGTTGTAGAAATACAACATCGGCTGC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB3. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccA4. 02C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCGAGTTGAATCTGTTGTAGAAATACAACATCGGCTGC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 01C10 Lcc3. 01C10 LccB3. 02C10 LccB3. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccB3. 09C10	TTCAAAGACTTCGATGTTGTGGGGAGGTAACAA-GTCTCCTTCGTTGTTGTTGTTGTGTGAGAATTACAACATCGGCTGC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAACACCGAGGATGAAAAGA     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAC     TTCCAAAGGTATGGGAAAGCGATAATCTTGATGTAT
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTGGAAATAAAAACACGGAGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 07C10 LccB6. 18C10 LccB6. 20C10 Lto3. 02C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccB6. 07C10 LccB3. 09C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGTTGTGAGAACACCGGGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAACACCGAGGATGAAAAGA     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 LccB6. 07C10 LccB6. 07C10 LccB3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGATTTCGATGTTGTGAGAGTAACAA-GTCTCTCCTTGATCTGTTGTTGTTAGAAATACAACATCGGCTGC     TTCAAAGGTATGGGAAAGCCGATAATCTTGATATTCTTCGGAGTTGAATTAAAGAAACACCCAAGGGAGAAAAGA     TTCAAAGATATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 02C10 Lto3. 03C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccB6. 07C10 LccB3. 09C10 LccA3. 04C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAATAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 02C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccA3. 09C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGATTGGTAGTAGAAATAAAAACACGAGGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGATTGAATTAAAGAAACACCGAGGGATGAAAAGA     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccA3. 09C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTGAAATAAAAACACCGAGGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACCGAGGGATGAAAAGA     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 02C10 LccB6. 18C10 LccB6. 20C10 Lto3. 02C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccB6. 07C10 LccA3. 09C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10	TTCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATCTGTTGAAAAACACCGAGGGATGAAAAGA     TCCAAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGAGTTGAATTAAAAAAACACCGAGGGATGAAAAGA     TTCAAAGGTATGGGAAAGCCAATCTTGATATTCTTCGA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 04C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 Lto3. 02C10 Lto3. 02C10 Lto3. 02C10 Lto3. 03C10 Lto3. 09C10 Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 04C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10	TTCAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGATGTGTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccB3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 01C10 Lcc3. 01C10 Lcc3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 18C10 Lto3. 02C10 Lto3. 09C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10	TTCAAGGTATGGGAAAGCGATAATCTTGATATTCTTCGATGTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB3. 02C10 LccB3. 02C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 Lto1. 05C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB3. 02C10 LccB6. 18C10 LccB6. 02C10 LccB6. 02C10 Lto3. 03C10 Lto3. 03C10	TTCAAAGATTCGGAAAGCGATAATCTTGATATTCTTCGATTGAATTAAAAAAAA
Lto1. 05C10 LccB6. 07C10 LccA4. 02C10 LccA3. 09C10 LccA3. 07C10 LccA3. 07C10 Lcc3. 07C10 Lcc3. 07C10 Lcc3. 07C10 LccB6. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 20C10 Lto3. 03C10 Lto3. 03C10 Lto3. 09C10 Lto3. 09C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccA3. 07C10 LccB6. 18C10 LccB6. 07C10 LccB6. 18C10 LccB6. 18C10 LccB6. 18C10 LccB6. 07C10 LccB6. 18C10 LccB6. 18C10	TTCAAAGATTCGATGTTGTGAGAGTAACAA-GTCTCCTTCCTTCGAGTTGAAATCAAACACCGAGGATGAAAAGA     TTCAAAGGTATGGGAAAGCGACAATCTTGATATTCTTCGAGTTGAATTAAAGAAACACTCAAGGGGAAAAGA     TTCAAAGGTATGGGAAAGCTATAATCTTGATATTCTTCGA

1 +01 05010	ΑΤCCAAAACCTCCAACTCACTCATTATAACAACACCCCAACAA
LCCA4. 02010	
LCCB3.09C10	GIIGCAGAGAIGACACACAIIICI
LCCA3. 04C10	GIIGCAGAGAIGACACACACAIIICI
LCCA3.0/C10	GIIGCAGAGAIGACACACACAIIICI
Lco3. 01C10	<b>GTTG</b> CAGAGATGACACACATTT <b>C</b>
Lco3. 03C10	<b>GTTG</b> CAGAGATGACACACATTT <b>C</b> T
Lco3. 07C10	<b>GTTG</b> CAGAGATGACGCACACATTT <b>CT</b>
LccB3. 02C10	<b>GTTG</b> CAGAGATGACACACATTT <b>CT</b>
LccB6. 18C10	<b>GTTG</b> CAGAGATGACACACACATTT <b>CT</b>
LccB6. 20C10	<b>GTTG</b> CAGAGATGACACACACATTT <b>CT</b>
Lto3.02C10	<b>GTTG</b> CAGAGATGACACACACATTT <b>CT</b>
Lto3.03C10	<b>GTTG</b> CAGAGATGACACACACATCT <b>C</b> T
Lto3.08C10	<b>GTTG</b> CAGAGATGACACATACATTT <b>CT</b>
Lto3.09C10	<b>GTTG</b> CAGAGATGACACACACATTT <b>CT</b>
Lto1.05C10	<b>A</b> ATTTATAAACTAGATCCTATGTCACATGAAGATTGTTGGTCTATACTCTCAAAACATGCATTTGGAAGTGAT
LccB6.07C10	ATTCATAAATTAGACCCTTTGTCACAAGAAGAATGTTGGTCATTACTATCAAAGCATGCAT
LccA4. 02C10	ATTTATAAGTTAGACCCTTTGTCACAAGAAGAATGTTGGTCTTTACTATCAAAGCATGCAT
LccB3. 09C10	AATTTATAAATTAGACCCTTTGTCACAAGAAGAAGAATGTTGGTCTTTACTATTAAAGCATGCAT
LccA3. 04C10	AGTTTATAAATTAGACCCTTTGTCACAAGAAGAAGAATGTTGGTCTTTACTATCAAAGCATGCAT
LCCA3_07C10	AATTTATAAATTAGACCCTTTGTCACAAGAAGAAGAATGTTGGTCTTTACTATCAAAGCATGCAT
L co3, 01C10	AATTTATAAATTAGACCCTTTGTCACAAAAAGAATGTTGGTCTTTACTATCAAAGCATGCAT
	A ATTATA A ATTAGA CCTTTGTCA CA A GA A GA
L co3 07010	AATTATAAATTAGACCCTTIGTCACAAGAAGAAGATGTIGGTCTTACTATCAAAGCATGCATTGGGAAGIGAT
$L_{CC}B_{3}$ 02010	ΔΑΤΤΑΤΑΔΑΤΤΑΘΑΓΟΥΤΤΕΓΟΛΟΑΓΑΘΟΑΓΑΤΕΤΕΘΕΓΤΤΑΓΤΑΓΛΑΘΟΑΘΟΑΓΟΟΤΗ
LCCB6 18C10	ΑΛΤΤΑΤΑΛΑΤΤΑΘΑΘΟΤΤΙΤΟΓΟΛΟΛΟΛΟΛΙΟΤΙΤΟΓΙΑΓΑΤΑΓΙΑΙΑΙΟΑΠΟΟΤΙΟΤΙΑΓΙΑ
LCCB0. 10010	ΑΛΤΤΑΤΑΛΑΤΤΑΘΑΘΕΟΤΤΙGTΟΛΟΛΟΘΑΤΟΤΙGTΟΤΤΑΟΤΑΓΑΘΑΠΟΘΙΟΤΟΤΙΑΟΤΑΓΟΛΟΛΟΘΟΙΟ
Lto2 02010	Αντταταγάττας ανογοτητοτοια και ανα ανα αντοττοριτοτη τη αναιτο αναφοριατο αντογραφικά του
$1 \pm 03.02010$	ΑΛΤΤΑΤΑΛΑΤΤΑΘΑΘΕΟΤΤΙGΤΟΛΟΑΘΑΠΟΠΤΟΓΤΟΓΤΑΓΙΑΓΙΑΘΑΑΙΘΕΑΤΟΓΙΟ
$1 \pm 02$ 09010	A TTATAAATTACACCTTTCTCACAACAACAATGTCTCCTTTACTACAACCATCCAT
$L_{103}$ , 00010	A TTATAAATTACACCTTTCTCACAACAACAATCTCCTCTTACTAC
L103.09010	ATTTATAATTAGACCCTTTGTCACAAGAAGAAGAATGTTGGTCTTTACTATCAAAGCATGCAT
1 + 01 05010	
	GAAAA TCACGCLAG TAAAAAGCGCAGCCCTTCAAGAAGATTCGCTAGAAAAGATTCGCAGAAAGTGTGGGAGGATTCC
LCCB6. 07010	GATTTCACGATAGTAAAAACACAGCCCTTCAAGAGATTCACAAGAATTCACAAGAAGTGTGGGGGGGATTCA
LCCA4. 02010	GATTITCACGATAGTAAAAAACACAGCCCTTGAAGAGATTGCAAGAAAGA
LCCB3. 09010	GATTITCACGATAGTAAAAAAAAAACACAGCCTTGAAGAGAATTGGCAGGAAGATTACAAGAAGTGTGGCAGGATTGC
LCCA3. 04010	GATTITICACGATAGTAAAAAACACAGCCCTTGAAGAAGATTGGCAGGAAGATTG
LCCA3. 07010	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGCCAGGAAGATTGCCAAGAAGTGTGGAGGATTGC
Lco3. 01C10	GATTTCCCGATAGTAAAAACACAGCCCTTGAAGAGATTGCCAGGAAGATTGCCAAGAAAGTGTGGAGGATTGC
Lco3. 03C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGCCAGGAAGATTGCCAAGAAAGTGTGGAGGATTGC
Lco3.07C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGC
LccB3. 02C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGC
LccB6. 18C10	GATITTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTG
LccB6. 20C10	GATTTTCACGATAGTAAAAACACAGCCCTTGTAGAGATTGGCAGGAAGATTG <u>CAAGAAAGTGTGGAGGATTGC</u>
Lto3.02C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGC
Lto3.03C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGC
Lto3.08C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGCAAGAAAGTGTGGAGGATTGC
Lto3.09C10	GATTTTCACGATAGTAAAAACACAGCCCTTGAAGAGATTGGCAGGAAGATTGC

Figura 4.3. Secuencias de nucleótidos en las que se puede observar una deleción central de 153 nucleótidos. Las secuencias Lto1.05C10 y LccB6.07C10 son secuencias completas tomadas como referencia. En negrita y subrayado se indican las secuencias correspondientes a los cebadores, y en negrita se indica un nucleótido (A) delecionado adicionalmente en la mayoría de las secuencias con la deleción central, la secuencia repetida que flanquea los extremos de la secuencia de la deleción de mayor tamaño y por último el triplete de nucleótidos correspondiente a la sustitución prolina→leucina.

Dentro del conjunto de muestras que denominamos pseudogenes se encuentran algunas que, aunque su secuencia de nucleótidos con respecto a las correctas no se ve gravemente alterada, han sido descartadas por la presencia de un codón de fin prematuro debido a un cambio de nucleótido.

Al analizar el alineamiento de las secuencias de aminoácidos deducidas *in silico* se encontró un grupo de tres secuencias de *Lens culinaris* subsp. *culinaris* cultivar Lupa en el que se observó una posible sucesión de cambios en el marco de lectura, ya que las tres mostraban la misma alteración de la secuencia en la misma posición (deleción de un nucleótido), adicionalmente dos de ellas mostraron una segunda mutación (la aparición de un codón de fin por un cambio de nucleótido) en una posición
corriente abajo y en una de éstas dos una tercera diferencia (nuevamente un codón de fin por cambio de nucleótido) más próxima al extremo carboxilo (Figura 4.4).

LccB3. 04C10	TLAQLVYNDEI VQQHFDFKAWVCVSEDFDVVRVTKSLLESVVRNTTSADSKVWESDNL <b>?</b> I LRVELKKHSREI I F
LccB5. 03C10	TLAQLVYNDEI VQLHFDFKAWVCVSEDFNVVRVTKSLLESVVRNTTSADSKVWESDNL <b>?</b> I LRVELKKHSREI I F
LccB5. 04C10	TLAQLVYNDEI VQQHFDFKAWVCVSEDFDVVRVTKSLLESVVRNTTSADSKVWESDNL <b>?</b> I LRVELKKHSREI I F
LccB3. 04C10	LFVLDDLWNDNYNDWDELVSPLI EGKPGSSMI I TTRQQKVAEMTETFPI CKLDPMSHEDCWSI LSNHALGSDEF
LccB5. 03C10	LFVLDDLWNDNYNDWDELVSPLI EGKPGSSMI I TTRHQKVAEMTDTFPI HKLDP*SQEECWSLLSKHALGSDDF
LccB5. 04C10	LFVLDDLWNDNYNDWDELVSPLI EGKPGSSMI I TTRHQKVAEMTDTFPI HKLDP*SQEECWSLLS*HALGSDDF
LccB3. 04C10	HASKSVALEEI GRKI ARKCGGL
LccB5. 03C10	HDSKNTSLEEI GRKI ARKCGGL
LccB5. 04C10	HDSKNTSLEEI GRKI ARKCGGL

**Figura 4.4.** Secuencias de aminoácidos que muestran una sucesión de cambios en el marco de lectura. En negrita se resalta la interrogación, que muestra la falta de un nucleótido en ese codón, y los asteriscos que señalan codones de fin debidos a un cambio de nucleótido.

Otras muestras presentaban una leve variación en el número de nucleótidos (de uno a tres), que podrían ser disminución o aumento, o ambas cosas, en distintos puntos de la secuencia, que llevaban a la pérdida del marco de lectura correcto.

El resultado obtenido al secuenciar el amplicón LccB2.06C10 (Figura 4.5) fue una secuencia compuesta por dos partes, una primera parte (LccB2.06C10\_1) a la que le faltaba el segmento final (que incluye el cebador inverso) y que además no tenía un marco de lectura correcto debido a la presencia de un nucleótido adicional en la posición 151, y una segunda zona, esta vez completa y con un marco de lectura correcto. Para el estudio filogenético sólo se ha tenido en cuenta esta segunda parte de la secuencia bajo el nombre de LccB2.06C10\_2.

**Figura 4.5.** Secuencia con duplicación en tándem. Se muestran en negrita y subrayados los cebadores directos de las dos partes de la secuencia y en negrita el cebador inverso de la segunda parte (la primera parte de la secuencia es una secuencia incompleta en la que no aparece cebador inverso). También se remarca en negrita, cursiva y sombreado el nucleótido extra (en la posición 151) que desplaza el marco de lectura en esa primera secuencia.

#### 4.3 Secuencias utilizadas en los análisis filogenéticos

Una vez descartados los pseudogenes, las secuencias parciales y las secuencias idénticas, quedaron 130 secuencias diferentes (Tabla 4.3 y anexo A1) con las que se realizaron los análisis filogenéticos. Las secuencias LccB6.08C10 y LccB6.10C10 presentaron una pequeña deleción con respecto al grupo de secuencias con las que se agrupan, pero este cambio no afectó al marco de lectura por lo que se decidió mantenerlas en los estudios filogenéticos.

Estas 130 secuencias fueron alineadas en primer lugar usando el programa ClustalW (Thompson et al., 1994) y posteriormente el alineamiento fue mejorado con el programa MUSCLE (Edgar, 2004).

Los dendrogramas (Figura 4.6) se obtuvieron a partir de la matriz de distancias utilizando el modelo de Kimura-2 parámetros (Kimura, 1980) y un posterior agrupamiento mediante el método Neighbor-

Joining. Para comprobar la consistencia de los agrupamientos se realizó un remuestreo "bootstrap" consistente en 1000 pseudorréplicas. También se aplicaron métodos de máxima verosimilitud utilizando distintos modelos evolutivos, obteniéndose en todos los casos topologías muy similares a la obtenida mediante Kimura-2 parámetros y Neighbor-Joining, por lo que solamente se muestran los resultados obtenidos con esta metodología.

 Tabla 4.3.
 Número de secuencias utilizadas en los estudios filogenéticos desglosadas por especie y por combinación de cebadores

$\searrow$	C1	C2	С3	<b>C6</b>	<b>C9</b>	C10	C11	C12	C13	C14	total
LccA <sup>1</sup>	1	1	1	1	np	2	1	3	-	2	12
LccB	2	1	3	21	10	23	-	1	-	1	62
Lco	1	1	-	1	np	2	-	1	-	-	6
Ler	2	-	1	1	np	2	-	-	2	1	9
Lni	1	1	-	1	np	1	1	2	2	-	9
Lla	1	2	1	1	np	3	2	3	1	-	14
Lto	1	-	-	1	np	3	1	1	1	2	10
Lod	-	1	2	-	np	2	2	-	-	1	8
total	9	7	8	27	10	38	7	11	6	7	130

<sup>1</sup> Véase leyenda de las tablas 4.1 y 4.2.

Analizando la figura 4.6a se aprecia una clara división en el árbol que, debido a los análisis posteriores, asociamos a la separación por subfamilias característica de las secuencias NBS-LRR, es decir, que como señala la figura 4.6a la parte superior del árbol correspondería a la subfamilia no TIR o CNL y la parte inferior a la subfamilia TIR o TNL. De las 130 secuencias utilizadas, 42 fueron del tipo TNL y 88 pertenecían a la subfamilia CNL.

Analizándolo por combinaciones de cebadores, las secuencias obtenidas a partir de las combinaciones C2, C3, C13 y C14 (degenerados y no degenerados) son únicamente del tipo TNL y las combinaciones C10, C11 y C12 (no degenerados) han dado secuencias solamente CNL, mientras que con las combinaciones C1, C6 y C9 (degenerados) se amplificaron secuencias de ambos tipos.

Además de la separación evidente de las dos subfamilias, el árbol muestra la formación de grupos dentro de cada subfamilia de secuencias, apareciendo cuatro grupos dentro de la subdivisión TNL o TIR, que llamaremos T-1, T-2, T-3 y T-4 y seis grupos en la subdivisión CNL o no TIR, nombrados como NT-1, NT-2, NT-3, NT-4, NT-5 y NT-6 (Figura 4.6b).

En el apartado b de la figura 4.6 se puede ver como los grupos NT-2, NT-3, NT-4, NT-6 y T-4 muestran una configuración compacta, en NT-1, NT-5 y T-3 se observan dos ramificaciones claras y los grupos T-1 y T-2 son muy ramificados.





#### b)

NT-1



58

NT-2



NT-3



NT-4



NT-5



NT-6





0.02

T-4



**Figura 4.6**. Árboles filogenéticos obtenidos mediante el algoritmo de Kimura-2 parámetros con el programa MEGA 5.2, a) el árbol general con las 130 secuencias muestra una marcada división entre los grupos TNL y CNL y la denominación de los distintos grupos identificados, b) árboles individuales de cada grupo. En cada caso se indican los valores de "bootstrap" y la escala de distancia.

En cuanto al número de secuencias englobadas en cada grupo hay una gran diversidad. Hay grupos muy numerosos, como por ejemplo NT-1 (38 secuencias) y NT-2 (26 secuencias) mientras que los grupos menos numerosos son NT-3, NT-5 y T-2 con 2, 4 y 5 secuencias respectivamente.

Con algunas combinaciones de cebadores las secuencias obtenidas fueron altamente similares entre sí, llegando a englobarse todas en un mismo grupo dentro del árbol. En el caso de las combinaciones C11 y C12 (no degenerados), sus secuencias forman en exclusividad los grupos NT-4 y NT-6 respectivamente. En otros casos estas combinaciones comparten grupo con otras; todas las secuencias de la combinación C3 forman el grupo T-4 con la mayoría de secuencias de la combinación C2; todas las secuencias de la combinación C14 se agrupan con dos secuencias de C2 para formar el grupo T-3; y todas las secuencias de la combinación C13 forman parte del muy diverso grupo T-1.

Combinaciones como C2, C6 y C10 tienen representación en dos grupos cada una. En el caso de C2 ambos grupos están en la subfamilia TNL (T-3 y T-4), lo opuesto ocurre con la combinación C10 en la que los dos grupos en los que se divide pertenecen al tipo CNL (NT-1 y NT-2), mientras que las secuencias de la combinación C6 se encuentran divididas entre las dos subfamilias, estando en el grupo NT-1 de CNL y en el T-1 de TNL.

Las dos combinaciones que muestran más diversidad a la hora de agruparse son C1, presente en tres grupos (NT-2, NT-5 y T-1) y C9, en cuatro grupos (NT-3, NT-5, T-1 y T-2). Ambas combinaciones (de cebadores degenerados) muestran secuencias de las dos subfamilias, TNL y CNL.

Esta uniformidad o diversidad dentro de las combinaciones también se ve reflejada en el tamaño de las secuencias (número de nucleótidos) que se ha obtenido con ellas, como se puede ver en la tabla 4.4.

Tabla 4.4. Tamaño de las secuencias en número de nucleótidos una vez eliminados los cebadores de los extremos.

	C1	C2	C3	C6	C9	C10	C11	C12	C13	C14
					778					
TNL	471-474 <sup>1</sup>	475	475	475	784	-	-	-	411-414	402
					790					
	471				847	421				
CNL	480	-	-	499-502	868-871	448	404	424	-	-
	400				000 071	466-469				

<sup>1</sup>Los datos consistentes en dos números separados por un guion indican secuencias que difieren entre sí en una longitud de tres nucleótidos contiguos (un aminoácido).

Como se ha explicado previamente, las combinaciones desde C1 a C9 fueron amplificadas usando cebadores degenerados mientras que las combinaciones de la C10 a la C14 se obtuvieron a partir de parejas de cebadores no degenerados.

Las combinaciones de cebadores no degenerados (C10 a C14) dieron siempre secuencias de una única subfamilia y un mismo grupo, excepto C10, con la que se obtuvieron secuencias que se agrupan en dos grupos diferentes (NT-1 y NT-2)

Dentro del conjunto de cebadores degenerados, algunos presentaban un grado de degeneración más elevado que otros. Por ejemplo, la combinación con más posiciones variables en su secuencia fue la C9. Esto podría explicar por qué resultó ser la combinación que produjo fragmentos con mayor diversidad, tanto de tamaño como de distribución de sus secuencias dentro del árbol.

Dentro de las seis combinaciones cuya secuencia se enmarca entre el p-loop y el GLPL pueden apreciarse unos patrones de amplificación. Cuando el cebador inverso fue as1 las secuencias que se obtuvieron pertenecían a las dos subfamilias (CNL o TNL) indistintamente de cual fuera el cebador directo, mientras que los cebadores inversos as3 y as4 dieron resultados muy diferentes dependiendo del cebador directo que los acompañaba. Si este cebador era s1 las secuencias obtenidas no mostraban similitud alguna con NBS-LRR, pero si era s2 las secuencias amplificadas presentaban similitud con el tipo TNL.

A diferencia de lo obtenido con combinaciones de cebadores no degenerados, ninguna combinación de cebadores degenerados ha producido secuencias que pertenecieran únicamente a la subfamilia CNL.

Se calculó la diversidad evolutiva o el número de sustituciones de bases por sitio utilizando el modelo Kimura-2 parámetros a través del MEGA 5.2. Para ello se analizó cada posición en el codón (1º+ 2º+3º+ no codificante). Todas las posiciones ambiguas fueron retiradas para cada par de secuencias.

Debido a la manifiesta diferencia de secuencia entre las dos subfamilias cuando se analiza la media de la diversidad evolutiva de la población total (Tabla 4.5), ésta es muy alta en el conjunto global de secuencias (d: 0,920), pero se ve drásticamente reducida al calcularla por subfamilias, siendo más alta en el caso de las TNL (d: 0,603) que en las CNL (d: 0,487). Analizándola por grupos, se advierte un

gran contraste en las correspondientes a la subfamilia TNL, ya que dentro de esta subdivisión se encuentra el grupo más diverso (T2, d: 0,435) pero también el menos diverso (T4, d: 0,013). En las CNL la diversidad de secuencia dentro de los grupos es mucho menor ya que el rango está entre el d: 0,014 del grupo NT-6 y el d: 0,178 del grupo NT-5.

Tabla 4.5. Valores medios de diversidad evolutiva dentro de los distintos grados de agrupación de secuencias.

Todas	CNL	TNL	NT-1	NT-2	NT-3	NT-4	NT-5	NT-6	T-1	T-2	T-3	T-4
0,920	0,487	0,603	0,050	0,034	0,083	0,033	0,178	0,014	0,361	0,435	0,121	0,013

En la tabla 4.6 se muestran los rangos de frecuencias nucleotídicas para el conjunto de todas las secuencias, por subfamilias y para cada uno de los grupos.

Para obtener unos rangos de frecuencia nucleotídicas comparables sólo se ha tenido en cuenta la secuencia común en las 130 secuencias, es decir, la porción de secuencia que coincide con la secuencia más corta.

Las secuencias NBS, por lo general, son ricas en adenina (A) y timina (T) y pobres en citosina (C). La cantidad de guanina (G) es más variable (Tabla 4.6).

	Α	т	С	G
Todas	25,8 - 38,1	25,2 - 34,5	11,5 – 19,0	17,3 – 29,8
CNL	25,8 – 36,0	25,2 - 34,5	12,3 - 18,5	17,9 – 29,8
TNL	31,3 - 38,1	28,0 - 32,7	11,5 – 19,0	17,3 – 24,5
NT-1	31,9 – 33,3	28,5 – 30,3	13,7 – 16,0	22,0 - 24,2
NT-2	30,5 - 33,3	28,5 – 30,7	15,8 – 17,0	20,9 – 22,4
NT-3	36,0 - 36,0	28,8 – 29,5	16,7 – 16,9	17,9 - 18,3
NT-4	25,8 – 27,9	32,4 - 34,5	16,4 - 18,5	21,1 - 22,2
NT-5	31,5 – 33,9	25,2 – 27,2	13,0 - 14,4	26,4 - 28,4
NT-6	26,2 – 26,9	31,0 - 31,7	12,3 - 13,6	28,3 - 29,8
T-1	31,3 – 38,1	28,2 – 32,7	11,5 – 19,0	20,1 - 24,5
T-2	33,4 – 35,1	28,0 - 29,6	13,3 – 15,0	22,0 - 24,0
Т-3	32,8 – 33,8	29,9 – 32,1	13,2 – 17,5	19,2 – 21,9
T-4	36,0 – 36,7	30,6 - 31,3	14,7 – 15,2	17,3 – 18,0

Tabla 4.6. Rango de frecuencias nucleotídicas para el total de las secuencias, por subfamilias y por grupos.

Al analizar el total de las secuencias se observa que los rangos son amplios y la única tendencia clara es la baja frecuencia de citosina. Por subfamilias, en los grupos TNL se puede apreciar que las frecuencias de adenina y timina son más altas que las de citosina y guanina, mientras que en los CNL esto no es tan claro. Además la diferencia entre los valores de los rangos es menor para TNL que para CNL en todos los nucleótidos excepto C. Esto podría deberse a las diferencias entre grupos dentro de una misma subfamilia.

Al estudiar los grupos de secuencias, C siempre es el nucleótido menos frecuente, seguido por G en la mayoría de los casos, excepto en NT-5 y NT-6. El nucleótido más frecuente para el mayor número de grupos es A, excepto para los grupos NT-4 y NT-6 que es T.

Atendiendo a la diferencia entre los valores del rango, el NT-3 es el grupo más homogéneo, ya que es un grupo compuesto únicamente de dos secuencias muy parecidas entre sí. Dentro de los grupos con más de dos secuencias, el que tiene unos rangos más cortos es T-4 seguido de NT-6, esto refuerza la teoría de que son grupos muy homogéneos. En el lado opuesto está T-1, que presenta las diferencias más notables en las frecuencias nucleotídicas de sus secuencias.

Hay rangos de frecuencias nucleotídicas para algunos grupos que son más amplios debido a que podrían dividir a estos en subgrupos. NT-5, T-1, T-2 y T-3 tienen valores muy distintos para alguno o todos los nucleótidos, dependiendo de a qué secuencia correspondan. La tabla con las frecuencias nucleotídicas de cada secuencia se encuentra en el CD adjunto (anexo A3).

Para analizar la divergencia evolutiva de los codones entre grupos de secuencias se han obtenido los valores promedio de los cambios sinónimos por sitio sinónimo (dS), no sinónimos por sitio no sinónimo (dN), la diferencia entre ambos (dS-dN) y el cociente entre dS y dN ( $\omega$ = dS/dN) entre parejas de secuencias (Tabla 4.7). Los análisis se realizaron por el modelo Nei-Gojobori (Nei y Gojobori, 1986) a través del programa MEGA 5.2.

Una sustitución sinónima es un cambio de nucleótido que no supone cambio de aminoácido, lo que produce una mutación neutra, mientras que el cambio de nucleótido en las sustituciones no sinónimas conlleva un cambio de aminoácido, pudiendo afectar a la función de la proteína, lo que hace que las sustituciones no sinónimas estén sometidas a selección. La comparación de la frecuencia con la que se fijan los cambios sinónimos frente a los no sinónimos permite estimar la importancia de la selección natural sobre un gen determinado.

El valor obtenido con el cociente dS/dN ( $\omega$ ) determina el tipo de selección entre parejas de secuencias. Un valor de  $\omega$  superior a 1 indica una permanencia más estable de los cambios sinónimos debido a una selección depuradora, un resultado menor de 1 sugiere una selección positiva y un  $\omega$  igual a 1 apunta a que no existe una selección neta.

Como se ve en la tabla 4.7 los valores promedio de sustituciones sinónimas fueron sustancialmente más altos que los de las sustituciones no sinónimas, además, en todos los casos el cociente  $\omega$  fue mayor que 1, por lo tanto, se aprecia una selección depuradora entre las secuencias de este estudio.

Como era de esperar, donde más cambios, tanto sinónimos como no sinónimos, se produjeron fue en el conjunto total de secuencias NBS-LRR. Como fueron datos altos en ambos tipos de sustituciones el coeficiente  $\omega$  resultante fue bajo ( $\omega$ = 1,838). Analizando las subfamilias se vio que el valor de  $\omega$  obtenido fue muy parecido ( $\omega$ = 2,289 para CNL y  $\omega$ = 2,378 para TNL); sin embargo, el número de sustituciones, tanto dS como dN, fue superior para TNL, aunque el análisis por grupos reveló que fue el grupo NT-5, de la subfamilia CNL, el que mostró mayor selección depuradora entre sus secuencias y que el grupo con un menor valor de  $\omega$  pertenece a la subfamilia TNL. Este grupo es T-3.

**Tabla 4.7.** Valores promedio de los cambios sinónimos (dS), no sinónimos (dN), su diferencia (dS-dN) y el cociente entre ambas ( $\omega$ = dS/dN) de todas las secuencias, por subfamilias y por grupos.

	dS	dN	dS-dN	ω =dS/dN
Todas	1,325	0,721	0,641	1,838
CNL	0,904	0,395	0,570	2,289
TNL	1,201	0,505	0,704	2,378
NT-1	0,107	0,035	0,072	3.057
NT-2	0,064	0,025	0,039	2,560
NT-3 <sup>1</sup>	0,071	0,032	0,040	2,219
NT-4	0,062	0,024	0,038	2,583
NT-5	0,522	0,097	0,425	5,381
NT-6	0,031	0,009	0,022	3,444
T-1	0,669	0,286	0,383	2,339
Т-2	1,188	0,319	0,868	3,724
Т-3	0,207	0,097	0,111	2,134
Т-4	0,028	0,008	0,020	3,500

<sup>1</sup> No son valores promedio al presentar este grupo solo dos secuencias.

Dentro de los CNL, NT-6 es el que tuvo los niveles promedio de sustituciones más bajos de la subfamilia (dS= 0,031 y dN= 0,009), seguido de NT-4 y NT-2, con valores muy parecidos, dS= 0,062 y 0,064 y dN= 0,024 y 0,025 respectivamente. NT-6 pese a tener un menor número de sustituciones tiene el segundo valor de  $\omega$  más alto de la subfamilia ( $\omega$ = 3,444).

El grupo NT-3 solo tiene dos secuencias, por lo que sus datos no son valores promedio si no valores absolutos. Este pequeño grupo es el que tiene una selección depuradora menor entre su par de secuencias de todos los grupos que componen la subfamilia CNL.

La mayor selección depuradora de todos los grupos se observó en el grupo NT-5 ( $\omega$ = 5,381). Este grupo está compuesto por cuatro secuencias y al revisar la matriz obtenida con las distancias por parejas se vio que entre las dos secuencias obtenidas con la combinación de cebadores C1 el número de cambios era muy pequeño (dS= 0,010 y dN= 0,008), lo mismo pasaba entre las dos secuencias amplificadas con la combinación C9 (dS= 0,011 y dN= 0,006). Fue al hacer la comparación entre las

secuencias de distintas combinaciones de cebadores donde aumentó el promedio de cambios, sobre todo sinónimos, dando valores muy significativos en el test de neutralidad (archivo A4 de los anexos).

Algo parecido se observó en NT-1, donde las secuencias LccA4.04C10 y LccB6.07C10, que forman una ramificación en el árbol de la figura 4.6b, hicieron aumentar los valores promedio de dS y dN de dicho grupo. Excluyendo estas dos secuencias los valores se reducen a dS= 0,079 y dN= 0,028, más parecidos a los valores obtenidos para la mayoría de grupos de la subfamilia CNL.

Los resultados obtenidos en el análisis por grupos de la subfamilia TNL son mucho más dispares, aunque el rango de  $\omega$  es menor (entre  $\omega$ = 2,134 de T-3 y  $\omega$ = 3,724 de T-2) que en CNL (entre  $\omega$ = 2,219 de NT-3 y  $\omega$ = 5,381 de NT-5).

La subfamilia TNL tiene a los dos grupos con más sustituciones tanto sinónimas como no sinónimas, entre sus pares de secuencias, T-2 (dS= 1,188 y dN= 0, 319) y T-1 (dS= 0,669 y dN= 0, 286) pero también tiene el grupo con menor promedio de cambios de los diez grupos obtenidos en este estudio, T-4, con unos valores de dS= 0,028 y dN= 0,008.

Observando los valores para cada pareja de secuencias en las matrices de distancia para cambios sinónimos y no sinónimos del grupo T-1 se pueden deducir tres subgrupos de secuencias, donde la secuencia LccB2.08C9 es la que muestra más sustituciones con respecto a las demás.

Eso mismo sugieren los datos en el caso de T-2, donde la secuencia LccB2.10C9 es la que más cambios muestra en la comparación.

T-3 mostró el valor de  $\omega$  más bajo de todos los grupos de NBS-LRR, lo que sugiere, como ya se ha mencionado antes, que es el grupo con una selección depuradora menor entre sus secuencias. Este grupo, al igual que NT-1, manifestó unos niveles de cambio más elevados a causa de los valores obtenidos en la comparación de dos de sus secuencias con el resto. En este caso, las dos secuencias pertenecientes a la amplificación con la combinación de cebadores degenerados C2 que se integraron en este grupo, Lod1.01C2 y LccB1.15C2. Al excluir estas dos secuencias del análisis las secuencias que quedaron pertenecían todas a la combinación de cebadores no degenerados C14 y los valores se redujeron a dS= 0,035 y dN= 0,012.

Las matrices de distancia con los valores obtenidos de sustituciones sinónimas y no sinónimas para cada pareja de secuencias están disponibles en el CD adjunto (anexo A4).

# 4.4 Análisis de los aminoácidos y de los motivos conservados

# 4.4.1 Contenido en aminoácidos

Se realizó un análisis del contenido de aminoácidos de las secuencias, tanto con todas las secuencias como de una forma más individualizada, por subfamilias y por grupos (Tabla 4.8). El aminoácido más abundante en casi todos los grupos es la leucina, excepto en NT-1 que es la valina. Sin embargo, cuando se analizó el segundo aminoácido más abundante hubo una gran diversidad de resultados, ya que fue serina en NT-1 y T-1, ácido aspártico en NT-2 y NT-4, lisina en el caso de NT-3 y T-2, ácido glutámico para NT-5 y NT-6, isoleucina en T-3 y asparagina en T-4. El aminoácido isoleucina, además de ser el segundo más abundante en T-3, también puede dar una idea de la subfamilia a la que

Tabla 4.8. Contenido en porcentaje de aminoácidos en las secuencias.

	Ala	Cys	Asp	Glu	Phe	Gly	His	lle	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
todas	4,857	2,204	7,834	6,704	4,256	4,285	2,743	6,059	7,629	11,384	1,217	5,139	2,481	3,435	5,420	7,639	4,285	7,815	2,309	2,304
CNL	5,024	1,962	8,442	7,357	4,474	3,725	2,548	4,995	7,778	10,804	1,085	4,674	2,847	2,612	5,502	8,135	4,753	8,542	3,004	1,734
TNL	4,521	2,692	6,608	5,384	3,815	5,413	3,138	8,206	7,328	12,554	1,483	6,075	1,742	5,096	5,255	6,637	3,340	6,349	0,907	3,455
NT-1	4,256	1,836	8,479	7,602	3,882	3,151	2,615	4,825	7,846	9,535	1,494	5,344	2,469	2,469	5,019	9,698	4,548	10,055	3,086	1,787
NT-2	5,037	2,224	9,793	6,494	6,085	3,656	2,659	4,807	7,696	10,841	0,051	4,577	2,634	1,406	6,801	7,287	6,392	7,057	3,017	1,483
NT-3	4,490	1,900	5,700	6,218	4,145	5,354	2,763	7,254	8,290	12,608	0,345	4,663	3,454	4,145	4,318	7,772	4,145	6,908	2,763	2,763
NT-4	3,945	2,985	10,981	4,158	3,731	3,198	2,026	5,117	6,930	11,194	2,026	5,011	2,345	5,117	2,452	8,209	6,183	9,701	2,985	1,706
NT-5	5,695	2,733	6,150	9,112	4,100	5,239	1,594	5,581	7,973	13,554	2,961	3,303	2,847	3,189	6,264	5,125	3,645	6,834	2,961	1,139
NT-6	8,511	0,774	5,674	9,929	3,546	5,029	2,772	4,900	7,930	13,282	0,709	2,837	4,964	3,804	5,996	5,867	1,418	7,157	2,772	2,128
T-1	3,712	2,688	6,186	6,101	4,309	5,119	2,218	7,125	7,850	10,964	2,645	5,333	1,749	4,607	5,717	8,020	4,309	6,527	1,621	3,200
T-2	3,689	2,229	6,457	7,071	4,458	5,842	2,613	7,071	8,993	14,835	0,922	3,613	1,614	3,920	4,304	8,301	3,305	7,148	1,230	2,383
T-3	5,774	1,524	7,618	4,892	3,288	4,491	3,609	10,425	6,255	13,152	1,283	6,014	1,443	6,255	5,533	5,934	1,764	6,496	0,722	3,528
T-4	5,209	3,700	6,572	3,797	3,165	6,037	4,236	8,812	6,329	12,561	0,633	8,520	1,996	5,696	5,161	4,430	3,213	5,550	0,000	4,382

pertenece una secuencia, ya que en las CNL está mayoritariamente entre el 4 y el 7% y en las TNL supera el 7%.

Por otro lado, al analizar los aminoácidos menos abundantes se pueden observar patrones diferentes para CNL y TNL. Por ejemplo, la metionina es un aminoácido poco frecuente en ambas subfamilias pero en las TNL es menos frecuente el triptófano en tres de los cuatro grupos, llegando a ser inexistente en T-4. Otro aminoácido con presencia característicamente baja en TNL es la prolina. Mientras que en los CNL, además de la metionina, los niveles más bajos de aminoácidos corresponden a tirosina, cisteína, histidina y en el caso de NT-2 glutamina.

#### 4.4.2 Motivos conservados

Una vez alineadas las secuencias, se ha procedido a buscar los motivos conservados en los distintos grupos a partir de las secuencias de aminoácidos deducidas (Figura 4.7), según el agrupamiento realizado en el análisis filogenético, tanto en el conjunto de secuencias tipo TNL y CNL, como en los subgrupos definidos en el análisis de similitud. Para deducir las secuencias consenso se usó el programa WebLogo versión 2.8.2 (Crooks et al., 2004), en sus dos alternativas, frecuencias y "bits", en la que se representa gráficamente la frecuencia de cada aminoácido o la información aportada por cada uno de ellos en cada posición. Los resultados obtenidos pueden observarse en las figuras 4.8 y 4.9 y A5 de los anexos. En ellas se han agrupado las secuencias según los grupos definidos en los dendrogramas. Sólo se han considerado los grupos que incluían al menos siete secuencias para obtener los posibles motivos conservados.

En los motivos incluidos en todas las secuencia de lenteja analizadas puede observarse que su grado de conservación es variable y en algunos casos inexistente (Tabla 4.9). Mientras que los motivos "Walker B" y "RNBS-B/sensor" están relativamente bien conservados en todos los grupos, el motivo "RNBS-C" tiene un parecido bajo en las secuencias de tipo CNL y es prácticamente irreconocible en las TNL. El motivo "RNBS-A", descrito como muy distinto entre genes CNL y TNL, se reconoce en las secuencias de tipo CNL en lenteja, mientras que es de nuevo prácticamente irreconocible en las secuencias de tipo TNL (Tablas 4.10 y 4.11).

		P-loop	RNBS-A		Walker B	RNBS-B	RNBS-C
CNL	LccB1.02C1 LccB2.01C6 LccB2.18C9 LccB5.11C10 LccA1.13C11 LccB2.01C12	GGVGKTTLAQLAYNDEKVHKHFDLIAW GGVGKTTLAQLVYNNEKVEQHFDFKVW GGMGKTTLAKLLYNDPDLKGKFEVRGL TLAQLVYNDEKVRKHFDLVAW YNHDTIKQIFDVQAW DSRIANHFECKAW	VC-VSEDFDVFRVTKTLLESVTKRF /C-VSEDFDVVRLTKSLLESVVRNT VH-IPKDFDHVTVTKTLLESVTKTF /C-VSEDFDVFRVTKTLLESVTKTF /C-VSDFDDVLKVTKFIAEEVGSSC /F-VSEEYRDKRKDVLQGILRGVDPLA	PWETNNFDLLRVELKKN TSASSKVWESDNLDILRVELNKN TRDTNFDKLQVQLKKS PCETNNFDLLRVELKKN TNTNNLNILHRDLKEK QGVEIDKLPEQELVNKLHSV	LRDKRFFIVLDDLWNDKYSD TRTKRFLFVLDDLWNDNYND SDKKFLLVLDDLWYGNYYG LRDRFFIVLDDLWNDKYSD LGKRFLIVLDDVWTEDDDS LAEKRYLIVLDDVWGMEV	WDELVSPFTSGKTGSRVIITTRHRKVAD-AA WDELVSPLVGGKPGSSVIITTRQQRVAE-MA WNSLSDIFNVGKIGSKIIITTRDVRVALPKQ WDELVSPLISGKTGSRVIITTRHEKVAD-AA WNSLLKPLQYGTMGSKILVTTRIEKVAS-MV WDGLKYAFPKRKLGSKILLTTRNWEVALHAD	HTF-PIYELDPLSDEDSWSLLSKHAFGSG CTF-PIHKLDPMSHEECWSILSKHALGCD CSL-YVHHLTSLETEDSWSLLARHAFVGR TF-PIYKLDPLSDEDSWSLLSKHAFGSG QTV-QPYSLHHLSDQDCWSVFANNACLSP AQS-HPHQLRPLNQEESFALLRSKAFPGA
TNL	Lni1.01C1 LccB1.15C2 LccA1.01C3 LccB1.04C6 LccB2.10C9 Ler1.06C13 LccA1.03C14	GGVGKTTIAKATYNHMSGSFEAKSFI GGVGKTTLATVLYDRISQQFVACCF: GGVGKTTLAGVLYGRISHQFNACCL: GGVGKTTIAKATYNHMSGSFEGKSFI GGMGKTTLARAVYNSISRKFDSSSFI DFECRSFI QFPVYCL:	VDVKTVWDSGKGLLSLQKQLLSDVCKT- DDVSKNYRLHDGPLGVQKQILDQTLGQU IDDVSKVFK-YDGPIGVQKQILHQTLGEE EDVNKVWEGGKGLVSLQKQLLSDVYKTT VDVRENS-MKHGLVHLQETLLLHLLDE- ANIREVWEQVGGQVCLQEQLMYDIFKE- DDLGKTFR-DNGPIGAQKLILHQTLIEE	SIHIDSVESGKRVLQHR )NHQIHNHYNATNLIRCR EHNQIYNLHDGTNLIQNR SIHIYSVESGKRVLQHR NIKLDDVSQGIPILERR TKIQNIELGKSILKRR EQLQACNIYNACNLIQSR	LGYKRIFLVLDNVNSLDQI LCHQKALMILDNVDHIEQ LCRRRALLIFDNVDHNEQI GYKRIFLVLDNVNSLDQI LCNKKVLLILDDVDNLQQI LSQKRVLLVLDVNKLDQI HGVKALIILDNVDQVEQI	LNALCGSLEWFGPGSRIIITTSDKHILRCLQ LEXIAVHREWLGAGSRIIITSDCHILKQYG LEKLAVNPKSLGSGSRIIIVSRDAHILKEYG LNALCGSLEWFGPGSRIIITTSNKHILCCLQ LRSLVGRTEWFGLGSRIIITTRDKHLLTTHG LNALCGSCRWFAPGSKIIITTRDEDILRGNR LEKLAVNREWLGAGGRIIIVSRDEHILKEYG	VDHVYKMKYMDNNESLEFFSWHAFNIP VDAVYQVPLLDSTNSLQLLCRKAFKLD VNTLYKVQPLNHTNSLQLFCRKAFKCD VDHYYKMKYMDNNESLEFFSWNAFNIP VEKEKLYEVKELNDRESLELFSWRAFRKS VDKVYVMKEMDESESLELFSWHAFKQN VDVYKVPLMNRIDSFQLFSRKAFKLD
		GLPL		RN	BS-D		MHD
CNL	LccB1.02C1 LccB2.01C6 LccB2.18C9 LccB5.11C10 LccA1.13C11 LccB2.01C12	NFCETQCQNLEAIGRKIARKCGGLPLA EYHASKSVALEEIGRQIARKCGGLPLA NYQQHPNLEIIGREIAKKCGGLPLA NFCETQCQNLEAIGKKIARKCGGL DENMDLQKIGKEIVR SAIPSEFENLAREIVVKCEGL	IALGGILCTKLSQDHWRDVLNSSIWKQ1	NDEVQPALLLSYRYLPSSLKGCFAY	CSIFPKNSILRKEMVVQLWI	AEGLIPQPKNESSWEKVAEECFNELVWRSLI	RRRSIHDESEQFIGDNEVFFEMHDRLP
TNL	Lni1.01C1 LccA1.15C2 LccA1.01C3 LccB1.04C6 LccB2.10C9 Ler1.06C13 LccA1.03C14	SPMESYAELCGDVVEYCGGLPLA HILNSYEELVKYILMYVNGLPLA NIMNDAYKELTYGILDIACGLPLA SPKESYAELCRDVVEYCGGLPLA MPDPCYVEIARSVVQYAKGHPLA SPSKDFSVISRNIIMYSGGL HIMSSYNKLAFDILRYAN	- - NVIGSDLFGK-TVEEWKSALNKYE-TIF	PSKEILNVLKVSYENLDDNEKEIFLD	IACFFKG	YPKADVENTLDASRFYSKYGIGVLV	DKSLVTIGESNS-VKMHDVLE

**Figura 4.7.** Secuencias de aminoácidos en las que se muestra la localización de los motivos conservados característicos de las secuencias NBS-LRR. Se han mantenido los fragmentos de secuencia correspondientes a los cebadores ya que en algunos de los casos estos coinciden con un motivo conservado.



**Figura 4.8.** Secuencias consenso de aminoácidos para las secuencias del grupo CNL entre los dominios p-loop y GLPL (no incluidos). a) Representación en frecuencias b) representación en bits.



**Figura 4.9.** Secuencias consenso de aminoácidos para las secuencias del grupo TNL entre los dominios p-loop y GLPL (no incluidos). a) Representación en frecuencias b) representación en bits.

**Tabla 4.9**. Comparación de los motivos conservados en las secuencias comunes propuestos<sup>1</sup> para los genes TIR (TNL) y No-TIR (CNL) (líneas superiores) y los encontrados en los correspondientes grupos de secuencias de lenteja.

Motivo	CNL	TNL
RNBS-A	FDLxAWVCVSQxF <sup>2</sup> FDI xaWVCVSEDF	FLENI RExSKKHGLEHLQKKLLSKLL Fl eni rev <mark>fexxxxGpi g</mark> l Qk <mark>q</mark> l Lsq <mark>t</mark> L
Walker B	LLVLDDVW Lfvlddi W	LLVLDDVW LI vLDdV <mark>x</mark>
RNBS-B/sensor	GSRI I I TTRD GSx <mark>VI I TTRx</mark>	GSRIIITTRD GSRIIIttRD
RNBS-C	YEVxxLS-EDEAWELFCFKxAF y <mark>kL</mark> DOLS <mark>x</mark> ED <mark>c-WSI LS</mark> -KHAf	YEVxxLSEDEAWELFCFKxAF YxvxxI xxxSL-eLFcxx-AF

<sup>1</sup> Según Meyers et al. (1999) y Lukasik y Takken (2009).

<sup>2</sup> Las letras en negro indican coincidencia en la posición y las rojas diferencias; las minúsculas indican que aunque el correspondiente aminoácido es el más frecuente en esa posición, no es mayoritario.

**Tabla 4.10.** Comparación de los motivos conservados en las secuencias comunes propuestos<sup>1</sup> para los genes CNL (líneas superiores) y los encontrados en los distintos grupos de secuencias de lenteja.

Motivo	NT-1	NT-2	NT-4	NT-6
RNBS-A	FDLxAWVCVSQxF	FDLxAWVCVSQxF	FDLxAWVCVSQxF	FDLxAWVCVSQxF
	FDFK <mark>V</mark> WVCVSEDF	FDLVAW <mark>A</mark> CVS <mark>E</mark> DF	FDVQAWVCVSDDF	<mark>¿ECKAWVFVSEEY</mark>
Walker B	LLVLDDVW	LLVLDDVW	LLVLDDVW	LLVLDDVW
	LFVLDDLW	FIVLDDLW	LI VLDDVW	LIVLDDIW
RNBS-B/sensor	GSRI I I TTRD	GSRI I I TTRD	GSRI I I TTRD	GSRI I I TTRD
	GS <mark>SVI I TTRQ</mark>	GSR <mark>V</mark> I I TTR <del>H</del>	G <mark>SKI LV</mark> TTRI	GS <mark>KI LL</mark> TTRN
RNBS-C	YEVxxLS-EDEAWELFCFKxAF	YEVxxLSEDEA-WELFCFKxAF	YEVxxLSEDEA-WELFCFKxAF	YEVxxLSEDEAWELFCFKxAF
	yKLDPMSHEDC-WSILSKHALGsDE	YELDPLS-DE <mark>DSWSLLS-</mark> KHAF	YSLHHLS-DQDCWSVFANNACL	HQLRPLNQEESFALLRSK-AF

<sup>1</sup> Véase leyenda en la tabla 4.9.

**Tabla 4.11.** Comparación de los motivos conservados en las secuencias comunes propuestos<sup>1</sup> para los genes TNL (líneas superiores) y los encontrados en los distintos grupos de secuencias de lenteja.

Motivo	T-1	T-3	T-4
RNBS-A	FLENI REx-SKKHGLEHLQKKLLSKLL	FLENI REXSKKHGLEHLQKKLLSKLL	FLENI REXSKKHGLEHLQKKLLSKLL
	FLenI reVWEQXXGI VSLQkQLLs <mark>Di v</mark>	fi DDLSKTFR-I DNGPI GAQKLI LHQ	LI DDVSKVFKYDGPI GVQKQI LHQ
Walker B	LLVLDDVW	LLVLDDVW	LLVLDDVW
	FLVLDDVN	LIILDNVD	LLI FDNVD
RNBS-B/sensor	GSRI I I TTRD	GSRI I I TTRD	GSRI I I TTRD
	GSRI I I TTrd	GSRI I I <mark>VS</mark> RD	GSRI I I <mark>VS</mark> RD
RNBS-C	YEVxxLSEDEAWELFCFKxAF	YEVxxLSE-DEAWELFCFKxAF	YEVxxLSEDEAWELFCFKxAF
	Y <mark>kM</mark> ke <mark>MD</mark> exE <mark>SL</mark> ELF <mark>SWh</mark> -AF	YKVPL <mark>MNRI</mark> D <mark>SF-Q</mark> LF <mark>SR</mark> K-AF	YKVQLNdTNSL-QLFCRK-AF

<sup>1</sup> Véase leyenda en la tabla 4.9.

# 4.5 Comparación con datos previos y con otras especies

Cuando se analiza la similitud de las secuencias obtenidas en este estudio con las secuencias de lenteja resultantes del trabajo realizado por Yaish *et al.* (2004) en este mismo laboratorio, vemos que estas últimas se incluían exclusivamente en la subfamilia TNL, distribuyéndose entre los cuatro grupos que se han detectado ahora en estas secuencias (Figura 4.10).

También se han realizado comparaciones con los datos de otras especies relacionadas, elegidas por su proximidad taxonómica al género *Lens* y por la extensa información disponible al haber sido secuenciado su genoma. Estas especies son *Medicago truncatula* (Figura 4.11), *Glycine max* (Figura 4.12), *Cicer arietinum* (Figura 4.13) y *Phaseolus vulgaris* (Figura 4.14).

Para *Medicago, Glycine* y *Phaseolus* se ha utilizado la base de datos Phytozome 5.0 obteniéndose como resultado datos de genes anotados de estas especies. Como dicha base no contiene datos para *Cicer,* para esta especie se han utilizado los datos del NCBI y las secuencias obtenidas pueden no corresponder con genes anotados.

En ocasiones, para *Cicer arietinum* se obtenían varias secuencias diferentes aunque idénticas en el segmento comparado con lenteja. En este caso sólo se ha incluido una (generalmente la que presentaba un número de locus (Loc) dentro del nombre descriptivo que aparece en la base de datos). Esta coincidencia podría deberse a que se trate de la misma secuencia introducida en la base de datos con distinto registro procedente de diferentes estudios (en este caso las coincidencias comprenden grandes extensiones de los transcritos), o simplemente al hecho de que coincida la secuencia comparada (la coincidencia se limita a la zona NBS) aunque sean de genes distintos.

Los datos escogidos para la comparación tenían un valor "–e "inferior a -60, pero para los árboles filogenéticos se han incluido sólo hasta las tres primeras posiciones entre las secuencias similares a cada secuencia de lenteja (tabla de "Excel" A2 incluida en el CD de anexos). También se han seleccionado algunas secuencias de *Lens* (tanto las obtenidas en este trabajo como en el trabajo previo realizado en el laboratorio del Área de Genética de la Universidad de León) de cada grupo, ya que el gran volumen de secuencias en algunos grupos hacía difícil la visualización de los árboles. Dichos árboles (Figuras 4.10, 4.11 4.12, 4.13 y 4.14) se obtuvieron utilizando el modelo de Kimura-2 parámetros (Kimura, 1980) y un posterior agrupamiento mediante el método Neighbor-Joining. Para comprobar la consistencia de los agrupamientos se realizó un remuestreo "bootstrap" consistente en 1000 repeticiones. En todas las figuras se incluyeron dos árboles, uno correspondiente a CNL y otro a TNL, para darle uniformidad a la visualización de los datos, aunque en el caso de la figura 4.10a no se incluyen secuencias nuevas debido a que Yaish *et al.* (2004) no obtuvieron ninguna del tipo CNL.

a)



Fig 4.10a

b)



**Figura 4.10.** Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en *Lens*. a) grupo CNL, b) grupo TNL.

a)



Fig. 4.11a



**Figura 4.11.** Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en *Lens* junto con las secuencias de *Medicago truncatula*. a) grupo CNL, b) grupo TNL.

a) NT-3 NT-2 1580 NT-4 160 Loi 1.05C11 100 Loi 1.05C11 Constant 1,02001 1,000 1 681.01C10 LCCA4.04C1010 100 NT-1 49 **98** 100\_1200B1.03C1 <sup>CB2.01C6</sup> Glyma01g04200.2 100 99 80 37 Glyma02g03520.1 Glyma09g02420.1 Glyma02903010.2 5 ST. GIMMADI GOALANT, wayina ngo 8640.1 LccB2.23C9 8 **NT-5** 1.3005 100 9\_1234GN/ma/1907680.2 Lla1.06C127207 nie orczą COB2.01C12 0.1 NT-6

Fig. 4.12a



**Figura 4.12.** Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en *Lens* junto con las secuencias de *Glycine max*. a) grupo CNL, b) grupo TNL.

a)



Fig. 4.13a

b)



**Figura 4.13.** Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en *Lens* junto con las secuencias de *Cicer arietinum*. a) grupo CNL, b) grupo TNL.

a)



Fig. 4.14a

b)



**Figura 4.14.** Árbol comparativo entre una selección de secuencias obtenidas en este estudio y en el trabajo previo realizado por el Dr. Yaish en *Lens* junto con las secuencias de *Phaseolus vulgaris*. a) grupo CNL, b) grupo TNL.

Al observar los datos para la comparación de las secuencias de *Lens* con otras especies se aprecia que para los grupos NT-1 y NT-2 de CNL se obtienen en algún caso los mismos resultados al comparar las secuencias de ambos con *Medicago* y *Cicer*, mientras que en los grupos T-3 y T-4 de TNL la coincidencia de resultados se observa tanto en *Medicago* y *Cicer* como en *Glycine*. Mediante un análisis de componentes principales (PcoA) y la creación de gráficos en 3D con el programa STATISTICA se observa que las secuencias de estos grupos aparecen entremezcladas al añadir los

datos tanto de *Medicago* como de *Cicer*, formando un macrogrupo NT-1\_NT-2 en las CNL y otro T-3\_T-4 en las TNL (Figura 4.15). También se puede observar el efecto contrario en el grupo T-2, en el que al introducir secuencias de otras especies en el análisis, se ve una clara separación en dos nuevos grupos que también queda patente en los gráficos 3D.



**Figura 4.15.** Gráficos 3D que muestran la agrupación de secuencias obtenida por análisis de componentes principales.

En los ejes de los gráficos de la figura 4.15 se pueden observar los porcentajes de la variación explicada de cada eje. En la tabla 4.12 aparece el porcentaje total alcanzado al utilizar tres ejes. El primer eje es el más representativo al explicar entre un 40 y un 47,5% de la variación obtenida en la matriz de distancias de las secuencias. Se obtiene un porcentaje mayor de la variación total en la

comparación de las secuencias CNL con *Medicago* y *Cicer* que al realizar el mismo análisis de las secuencias TNL.

	Eje 1	Eje 2	Eje 3	% total
CNL + Medicago	40,53	20,13	17,90	78,55
TNL + Medicago	41,57	12,83	9,81	64,20
CNL + Cicer	47,51	20,29	14,52	82,32
TNL + Cicer	44,79	13,65	10,41	68,86

**Tabla 4.12.** Porcentajes de variación en la matriz de distancias de las secuencias que representa cada eje de los gráficos obtenidos con el programa STATISTICA y el total de la representación acumulada.

Para analizar la posible localización de las secuencias de lenteja en superagrupaciones o como secuencias aisladas dentro de los cromosomas se realizó un estudio de las secuencias de lenteja obtenidas en este trabajo con secuencias homólogas de *Medicago truncatula* y *Glycine max* con los datos de la web LIS (Legume Information System), utilizando como bases de datos Medicago truncatula v 3.5.1 y Glyma1.01 (=Glyma.Wm82.a1), con un blast del tipo blastx (nucleótido->proteína), un valor de "e" de 1e<sup>-20</sup> y un porcentaje de identidad mínima del 60%. La visualización de los datos se realizó mediante CViT (Chromosome Visualization Tool). En el análisis se han utilizado las secuencias de lenteja seleccionadas para los árboles comparativos con otras especies (Figuras 4.10- 4.14).

Al introducir cada una de las secuencias de lenteja aparecía un patrón de secuencias homólogas, representadas por puntos sobre un esquema de los cromosomas de *Medicago* y *Glycine*, que indicaban la localización de los genes anotados de estas especies que mostraban similitud con la secuencia de lenteja estudiada, dentro de los parámetros seleccionados.

Al comparar las secuencias de la subfamilia CNL con *Medicago* y *Glycine*, los patrones de distribución en el genoma obtenidos para cada secuencia de un mismo grupo son muy parecidos, llegando a ser idénticos en algunos casos. Pero no ocurre lo mismo para la subfamilia TNL, donde algunas secuencias de un mismo grupo mostraron una homología con secuencias de *Medicago* y/o *Glycine* muy diferente de la mostrada por el resto del grupo, hasta el punto de poder considerar distintos subgrupos en tres de los cuatro grupos de la subfamilia TNL, que se pueden observar también en los árboles filogenéticos (Figura 4.16). Mientras que al comparar las secuencias de *Lens* con *M. truncatula* se obtuvo un patrón de similitudes para todas las secuencias, sin embargo, en el caso de *Glycine max* sólo se halló similitud en algunos grupos o subgrupos.

Como representación de los patrones elaborados se escogió el patrón de una única secuencia de cada grupo o subgrupo, en el caso de que los hubiera (Figuras 4.17 a 4.30). Todos los patrones obtenidos por este método para cada secuencia se incluyen en el anexo A6 del CD adjunto.



NT-6

89

b)



**Figura 4.16.** Árboles en los que se muestran los grupos de secuencias NBS-LRR y la separación por subgrupos en los grupos donde se hayan formado en a) CNL y b) TNL.

Analizando por grupos se observa que todas las secuencias de NT-1 mostraron un patrón de similitud muy uniforme. Todas revelaron una gran cantidad de puntos de similitud con *Medicago* en el cromosoma 3 y un único punto de similitud en el cromosoma 4 (figura 4.17), la única diferencia apreciable fue que algunas secuencias presentaron dos puntos de similitud en el cromosoma 8 de *M. truncatula* y otras solo uno.


**Figura 4.17.** Representación de la localización de las secuencias de *M. truncatula* que muestran similitud con el grupo NT-1 de lenteja. El patrón de puntos corresponde a la secuencia LccB2.01C6 de este estudio.

El patrón de puntos obtenido con el grupo NT-2 (Figura 4.18) mostró un gran parecido al de NT-1, ya que la gran mayoría de las similitudes con *Medicago* se encontraron en la misma porción del cromosoma 3 que las halladas para NT-1 (Figura 4.31a).También compartían el punto de similitud en el cromosoma 4, además de que algunas de las secuencias de NT-2 presentaron dos similitudes en el cromosoma 8 y otras solo una, al igual que pasaba en NT-1, con la diferencia de que en NT-2 la que apareció en todas las secuencias es la que se encuentra en la posición superior y la que variaba es la de la posición inferior, justo al contrario de lo que ocurría en el grupo NT-1. Otra diferencia con NT-1 es que muchas de las secuencias de NT-2 exhibieron una similitud con *Medicago* en el cromosoma 7.



**Figura 4.18.** Representación de la localización de las secuencias de *M. truncatula* que muestran similitud con el grupo NT-2 de lenteja. El patrón de puntos corresponde a la secuencia LccB5.21C10 de este estudio.

NT-3 es el grupo más pequeño de los formados en este estudio, ya que lo componen solamente dos secuencias. Estas dos secuencias exhibieron las mismas similitudes al compararlas con los genes anotados de *M. truncatula* (Figura 4.19). Localizadas todas en el cromosoma 3, en una posición más distal en el brazo corto del cromosoma que las similitudes con el cromosoma 3 que se apreciaban en los grupos NT-1 y NT-2 (Figura 4.31a).



**Figura 4.19.** Representación de la localización de las secuencias de *M. truncatula* que muestran similitud con el grupo NT-3 de lenteja. El patrón de puntos corresponde a la secuencia LccB1.24C9 de este estudio.

Los grupos NT-1, NT-2 y NT-3 no mostraron similitud con *Glycine max* (Figura 4.20b) bajo los criterios escogidos para la comparación, por lo que NT-4 es el primero de los grupos de la subfamilia CNL que ofreció similitud con dicha especie además de con *M. truncatula* (Figura 4.20a) en este análisis.





**Figura 4.20.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el grupo NT-4 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccA1.13C11 de este estudio.

Con respecto a *M. truncatula*, presentó similitud con un grupo de secuencias localizadas en el cromosoma 2, además de mostrar similitud con una secuencia situada en el cromosoma 5 y otra en el 7 (Figura 4.20a).

Analizando los resultados obtenidos de la comparación con *Glycine* vimos que las similitudes de las secuencias de *Lens* del grupo NT-4 quedaban localizadas en el cromosoma 3 de *G. max*, variando bastante el número de secuencias similares de *Glycine* según a qué secuencia de *Lens* corresponda (Figura 4.20b).

NT-5 también exhibió similitud específica con *G. max* además de con *M. truncatula*. Este grupo mostró una similitud en el cromosoma 3 y un conjunto de similitudes en el centro del cromosoma 5 de *M. truncatula* (Figura 4.21a).

En cuanto a *G. max* reveló un patrón de similitud muy disperso por los cromosomas, al contrario que NT-4(Figura 4.32a), donde quedaba concentrado en un único cromosoma, con secuencias similares a NBS-LRR de *Lens* en los cromosomas 1, 2, 9, 12, 15 y 19, aunque en ninguno de ellos en gran número (figura 4.21b)



**Figura 4.21.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el grupo NT-5 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccB1.03C1 de este estudio.

El último grupo dentro de la subfamilia CNL, NT-6, también arrojó resultados en la comparación tanto con *M. truncatula* como *G. max*.

Todas las secuencias de *Lens* pertenecientes a este grupo presentaron una única secuencia de *Medicago truncatula* similar en el cromosoma 5 (Figura 4.22a) y dos puntos de similitud en *Glycine max* localizados en los cromosomas 1 y 11 (Figura 4.22b). Estos resultados coinciden con los obtenidos en los árboles de las figuras 4.11a y 4.12a.



**Figura 4.22.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el grupo NT-6 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccB2.01C12 de este estudio.

Analizando la subfamilia TNL por grupos vimos que T-1 era un grupo muy disperso al compararlo con *Medicago* y *Glycine*, hasta el punto de aparecer patrones de distribución de secuencias homólogas tan diferentes como para considerar dividir este grupo en tres subgrupos (figura 4.16), que hemos llamado T-1.1, formado por las secuencias de la combinación C13 además de la secuencia Lla1.01C6; T-1.2, que englobó las secuencias LccB3.14C6, Lni1.01C1, Lto1.02C1 y Lto1.03C6; y T-1.3, con una única secuencia, LccB2.08C9.

En la comparación con *M. truncatula* el subgrupo T-1.1 mostró un nutrido conjunto de similitudes en la zona distal del brazo corto del cromosoma 4 y alguna similitud más a lo largo de dicho cromosoma, además todas menos Lni1.05C13 exhibieron alguna similitud en el cromosoma 7. La secuencia Lla1.01C6 tiene también una similitud en el cromosoma 1 (Figura 4.23a).



**Figura 4.23.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el subgrupo T-1.1 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia Lla1.01C6 de este estudio.

En el caso de *Glycine* aparecieron similitudes en el cromosoma 3, un grupo en la parte superior y alguna en la parte central. También se localizaron similitudes en el cromosoma 1 aunque en menor número (Figura 4.23b).

El subgrupo T-1.2 no ofreció patrón de similitud con *Glycine* con los criterios marcados para esta comparación y los resultados obtenidos con *Medicago* tuvieron similitudes con lo obtenido para T-1.1 (Figura 4.31b), ya que aparecieron puntos de similitud en los cromosomas 4 y 7, pero en distinto número, ya que en el cromosoma 4 aparecía una única similitud en lugar de un grupo (Figura 4.24).



**Figura 4.24.** Representación de la localización de las secuencias de *M. truncatula* que muestran similitud con el subgrupo T-1.2 de lenteja. El patrón de puntos corresponde a la secuencia Lto1.03C6 de este estudio.

La secuencia LccB2.08C9 forma el subgrupo T-1.3 (Figura 4.25 a y b) y es la que tiene un patrón de similitudes más diferente del resto de T-1 (Figura 4.31b).





**Figura 4.25.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el subgrupo T-1.3 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccB2.08C9 de este estudio.

Al comparar con *Medicago truncatula* tenía similitudes localizadas en los cromosomas 3, 5, 7 y 8, en este último se encontraron la mayoría de las secuencias homólogas (Figura 4.25a).

El subgrupo T-1.3 también mostró similitudes con *Glycine max*, donde encontramos grupos de secuencias similares a los NBS de lenteja en los cromosomas 3 y 16, además de una similitud en la parte inferior del cromosoma 12 (Figura 4.25b).

En el grupo T-2 también se apreció la separación en subgrupos al observar los patrones de puntos, en este caso se formaron dos subgrupos, el primero formado por cuatro de las cinco secuencias del grupo, lo nombramos como T-2.1 y el segundo, T-2.2, formado por la secuencia restante, LccB2.10C9 (Figura 4.16). En este grupo ninguno de los dos subgrupos presentó similitudes con *Glycine max*.

Con *Medicago* mostraron similitudes muy localizadas, en el caso de T-2.1 tiene una única similitud en el cromosoma 3 y otra en la parte distal del brazo largo del cromosoma 7 (Figura 4.26), mientras que la secuencia que forma el subgrupo T-2.2 tiene una sola similitud en el cromosoma 6 (Figura 4.27).









El grupo T-3 mostró un patrón de similitud muy homogéneo al compararlo con *M. truncatula* en todas las secuencias, pero no pasó lo mismo en la comparación con *G. max* donde se puede hacer una separación en dos subgrupos (Figura 4.16) a partir de la presencia o ausencia de secuencias homólogas en esta especie. Los subgrupos fueron llamados T-3.1, formado por las secuencias obtenidas a través de la combinación de cebadores C14, que sí presentaron similitudes con secuencias de *Glycine*, y T-3.2, donde se localizaron las secuencias LccB1.15C2 y Lod1.01C2 que no exhibieron similitudes significativas con *Glycine*.

En comparación con *Medicago* todas las secuencias del grupo T-3 presentan dos agrupaciones de secuencias similares (con más puntos de homología en el caso de T-3.2), uno en el cromosoma 4 y otro en la zona distal del brazo largo del cromosoma 6. Además también tienen alguna similitud en los cromosomas 2 y 3 (Figuras 4.28a y 4.29)



**Figura 4.28.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el subgrupo T-3.1 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccA1.04C14 de este estudio.

Las secuencias que fueron asignadas como subgrupo T-3.1 mostraron todas una similitud en el cromosoma 12 de *Glycine max* y además la secuencia LccA1.04C14 mostró una similitud en la parte final del cromosoma 6 (Figura 4.28b).



**Figura 4.29.** Representación de la localización de las secuencias de *M. truncatula* que muestran similitud con el subgrupo T-3.2 de lenteja. El patrón de puntos corresponde a la secuencia LccB1.15C2 de este estudio.

Por último, el grupo T-4 obtuvo unos patrones de comparación muy similares en todas sus secuencias (Figuras 4. 30 a y b). Mostró algunos puntos en común con el patrón del grupo T-3 para ambas especies (Figuras 4.31b y 4.32b).





**Figura 4.30.** Representación de la localización de las secuencias de a) *M. truncatula* y b) *G. max* que muestran similitud con el grupo T-4 de lenteja. Los patrones de puntos de ambas especies corresponden a la secuencia LccA1.01C2 de este estudio.

Al analizar la comparación del grupo T-4 con *Medicago truncatula* (Figura 4.30a) se originaron patrones con un grupo de similitudes en el cromosoma 4, en el mismo lugar que el del grupo T-3 pero una sola similitud en el cromosoma 6, a diferencia de T-3, donde había un grupo. También apareció la similitud en el cromosoma 3 observada en el grupo T-3 (Figura 4.31b).

Por otra parte, al comparar este grupo con *Glycine max* (Figura 4.30b) se observó el mismo punto de similitud en el cromosoma 12 localizado en el subgrupo T-3.1 (Figura 4.32b).

Las figuras 4.31 y 4.32 muestran los resultados conjuntos de las secuencias escogidas como representación de cada grupo y/o subgrupo por subfamilias para facilitar la visualización de las agrupaciones de secuencias homólogas.



**Figura 4.31.** Representación de la localización de las secuencias homólogas de *M. truncatula* para a) las secuencias seleccionadas de los grupos CNL y b) las secuencias seleccionadas de los grupos TNL.



**Figura 4.32.** Representación de la localización de las secuencias homólogas de *G. max* para a) las secuencias seleccionadas de los grupos CNL y b) las secuencias seleccionadas de los grupos y subgrupos TNL.

Estos patrones de homología coinciden con los resultados obtenidos en los árboles de las figuras 4.11 a y b en el caso de *M. truncatula* y 4.12 a y b para *G. max.* También apoya la división por subgrupos de alguno de los grupos que se observó al analizar los valores promedio de cambios sinónimos y no sinónimos entre pares de secuencias. Igualmente refuerza la similitud de resultados para los grupos NT-1/NT-2 y T-3/T-4 y la separación en dos subgrupos en T-2 que se observó en el análisis de componentes principales que se hizo con las secuencias de este estudio con secuencias de *Medicago* (figura 4.15).

## Discusión

## 5. DISCUSIÓN

Entre los genes relacionados con la resistencia a enfermedades en plantas se han descrito diferentes familias génicas entre las que se encuentra la denominada *NBS-LRR* por la presencia de ciertos dominios (Nucleotide Binding Site – Leucine Rich Repeat). Estas secuencias se clasifican en dos grandes grupos o subfamilias que codifican para proteínas denominadas respectivamente CNL y TNL y que se caracterizan por el tipo de secuencia presente en el dominio N-terminal además de algunas diferencias características en los motivos conservados que forman el dominio NBS.

La familia génica de los genes *NBS-LRR* es la más numerosa de las familias de genes de resistencia a enfermedades y es una de las familias génicas más amplias en el genoma de plantas.

El número de genes *NBS-LRR* presente en el genoma varía dependiendo de la especie, por ejemplo, para *Medicago truncatula* se estiman entre 400 y 500 genes (Ameline-Torregrosa *et al.*, 2008; Song y Nan, 2014) mientras que para *Arabidopsis thaliana* se han aislado un total de 207 (Meyers *et al.*, 1999). Jia *et al.* (2015) encontraron también esta gran variación en el número de secuencias NBS-LRR comparando entre cinco especies de cucurbitáceas y cinco especies de rosáceas, ya que las cucurbitaceas tenían entre 45 y 80 genes *NBS* codificantes y las rosáceas entre 346 y 1303 genes *NBS* codificantes. Sekhwal *et al.* (2015) en su análisis de RGAs, en el que recogieron datos de varias especies de plantas, tanto monocotiledóneas como dicotiledóneas, observaron valores muy diversos, tales como, 20 secuencias NBS en *P. patens*, 84 en *L. japonicus*, 179 en *Z. mays*, 201 en *A. thaliana*, 352 en *V. vinifera*, 661 en *M. truncatula*, 960 en *M. domestica* y 971 en *T. aestivum*. Esta variabilidad en el número de secuencias NBS se observa en las dos subfamilias, CNL y TNL (Jia *et al.*, 2015; Lozano *et al.*, 2015; Sekhwal *et al.*, 2015; Sun *et al.*, 2014; Zhong *et al.*, 2015)

En este trabajo se han obtenido 130 secuencias NBS-LRR diferentes mediante amplificación por PCR utilizando cebadores tanto degenerados como no degenerados. La utilización de cebadores degenerados para obtener secuencias NBS-LRR e incluso diferenciar entre secuencias TNL y CNL ha sido extensa. Centrándonos en las leguminosas, este método se ha utilizado por ejemplo en alfalfa (Cordero y Skinner, 2002; Zhu *et al.*, 2002), alubia (López *et al.*, 2003), garbanzo (Huettel *et al.*, 2002; Palomino *et al.*, 2006), lenteja (Yaish *et al.*, 2004), haba (Palomino *et al.*, 2006) y guisante (Djebbi *et al.*, 2015).

Yaish *et al.* (2004) sugirieron una asociación entre la combinación de cebadores y el tipo de secuencias amplificadas. Así, en leguminosas, los cebadores derivados de los motivos conservados P-loop y GLPL amplifican secuencias TNL con preferencia sobre las secuencias CNL (Bertioli *et al.*, 2003; Palomino *et al.*, 2006; Yu *et al.*, 1996). De este estudio se han obtenido resultados que apoyan esta hipótesis, ya que con todas las combinaciones formadas a partir de estos cebadores que dieron resultados válidos (C1, C2, C3 y C6) se obtuvieron secuencias de la subfamilia TNL, mientras que solamente las combinaciones C1 y C6 amplificaron secuencias del tipo CNL. Debido al uso de cebadores degenerados puede suceder que una única pareja de cebadores amplifique más de un tipo de secuencias (Palomino *et al.*, 2006). Esto explicaría por qué las combinaciones C1, C6 y C9 amplificaron tanta variedad de secuencias, ya que en los tres casos amplificaron secuencias para ambas subfamilias (CNL y TNL) y distribuidas en varios de los diez grupos en los que se separaron finalmente las secuencias.

El tamaño de banda esperado para las secuencias obtenidas a partir de las combinaciones de los cebadores derivados de los motivos conservados p-loop y GLPL fue de entre 500 y 550 pb porque es la distancia nucleotídica entre estos dos motivos en los genes *N* de tabaco, *L6* de lino y *RPS2* de *Arabidopsis* (Leister *et al.*, 1996). Analizando los geles de agarosa que muestran los patrones de bandas obtenidos como resultado de la amplificación por PCR (Figura 4.1), se pudo observar que en las combinaciones de C1 a C6 había una banda claramente visible aunque su intensidad variaba según la combinación, con un tamaño próximo a 500 pb. Estos resultados coinciden con lo obtenido por Leister *et al.* (1996) en patata, Yaish *et al.* (2004) en lenteja, Palomino *et al.* (2006) en haba y garbanzo y Djebbi *et al.* (2015) en guisante.

Una vez secuenciadas las muestras que contenían el fragmento de ADN con el tamaño de banda deseado (hay que tener en cuenta que los tamaños indicados incluyen las bases de los cebadores) se pudo comprobar que las secuencias obtenidas a partir de las combinaciones C4 y C5 eran algo menores de lo esperado (481-482) y que además no mostraban similitud con NBS-LRR, esto llevó al descarte de las secuencias obtenidas con ambas combinaciones. Al igual que para lenteja, en garbanzo y haba tampoco se obtuvo banda de 500 pb con estas combinaciones (Palomino *et al.,* 2006), al contrario de lo que ocurre en guisante, donde C5 amplifica una banda de 500 pb formada por secuencias con similitud a la subfamilia TNL (Tabla 5.1).

Para lenteja las combinaciones C1, C2, C3 y C6 sí que amplificaron secuencias con tamaños dentro de los esperados y con similitud a NBS-LRR. Estos resultados parecen variar según la especie de estudio, por ejemplo, C1 en lenteja ha sido una de las combinaciones que ha obtenido resultados para las dos subfamilias de la familia génica de *NBS-LRR*, también obtuvo resultados válidos en patata (Leister *et al.,* 1996) y en guisante (Djebbi *et al.,* 2015) aunque en este último únicamente de la subfamilia CNL, mientras que en haba y garbanzo ni mostraba similitud con secuencias NBS-LRR ni amplificaba un tamaño de banda adecuado (Tabla 5.1).

Combinación	lenteja	guisante	haba	garbanzo	patata
<b>C1</b> (s1/as1)	TNL/CNL	CNL	-	-	TNL/CNL
<b>C2</b> (s2/as3)	TNL	np <sup>1</sup>	TNL	TNL	TNL
<b>C3</b> (s2/as4)	TNL	TNL	-	-	np
<b>C4</b> (s1/as3)	-	np	-	-	np
<b>C5</b> (s1/as4)	-	TNL	-	-	np
<b>C6</b> (s2/as1)	TNL/CNL	CNL	TNL/CNL	TNL/CNL	np

**Tabla 5.1.** Tipo de secuencias obtenidas con las combinaciones de cebadores degenerados basados en los motivos conservados p-loop y GLPL en lenteja (este estudio), guisante (Djebbi *et al.*, 2015), haba y garbanzo (Palomino *et al.*, 2006) y patata (Leister *et al.*, 1996).

<sup>1</sup> np: no probado. -: no se obtuvieron secuencias con similitud a NBS-LRR

Algo parecido ocurre con los resultados de la combinación C3. En este estudio dicha combinación genera secuencias similares a la subfamilia TNL de los genes *NBS-LRR*. Esto se asemeja a lo conseguido por Djebbi *et al.* (2015) para guisante (Tabla 5.1), donde además es la combinación de cebadores más prolífica con siete de los diez tipos de RGAs hallados, pero contradice nuevamente lo obtenido para haba y garbanzo por Palomino *et al.* (2006), que excluyen esta combinación de cebadores de su estudio debido a los nulos resultados producidos con ella.

En lo que coinciden, tanto el trabajo de Palomino *et al.* (2006) como los resultados de esta Tesis, es en el tipo de secuencias obtenido por las combinaciones C2 y C6, puesto que en ambos trabajos todas las secuencias derivadas de la combinación C2 pertenecían al tipo TNL mientras que la combinación C6 tuvo representación en las dos subfamilias, TNL y CNL. En el análisis de guisante realizado por Djebbi *et al.* (2015) sin embargo con la combinación C6 sólo encuentran secuencias de la subfamilia CNL (Tabla 5.1).

Para las combinaciones C7, C8 y C9 esperábamos que las secuencias tuvieran un tamaño de entre 817 y 832 pb como en el trabajo de Yaish *et al.* (2004), pero la combinación C8 no llegó a mostrar ninguna banda nítida en los geles de agarosa, y la combinación C7 no amplificó secuencias similares a NBS-LRR, a pesar de mostrar la banda correspondiente. La única combinación de este grupo que dio resultados analizables fue C9 pero sólo estaban dentro del tamaño esperado las secuencias que finalmente se englobaron en la subfamilia TNL ya que las secuencias CNL obtenidas en este estudio con esta combinación tuvieron un tamaño mayor (entre 886 y 910 pb). Muy probablemente esto se debe a que el tamaño de referencia dado por Yaish *et al.* (2004) correspondía únicamente a secuencias TNL, ya que no obtuvo secuencias de la otra subfamilia.

Por último, los tamaños de referencia para las combinaciones de C10 a C14 se tomaron del estudio realizado por Palomino *et al.* (2009). En él la combinación C10 (llamada originalmente RGA01) daría lugar a un producto de 511 pb, C11 (RGA03) 456 pb, C12 (RGA04) 462 pb, C13 (RGA05) 486 pb y C14 (RGA10) 440 pb. En nuestro caso C10 dio un resultado más diverso con secuencias de un tamaño algo menor al indicado, desde los 463 a los 508 pb, pero todas con similitud a NBS-LRR y los dominios conservados característicos.

El resto de combinaciones también amplificó fragmentos que presentaron diferencias de tamaño con respecto a los resultados de Palomino *et al.* (2009) excepto la combinación C14 que amplificó secuencias de 440 pb tanto en este trabajo como en el de Palomino *et al.* (2009). Los resultados obtenidos a partir de este estudio fueron: con C11, 444 pb, con C12, 468 pb y con C13, 448-451 pb (algunas secuencias tienen un triplete de diferencia). En lo que sí coincidieron ambos trabajos fue, nuevamente, en el tipo de secuencias obtenido con cada combinación, ya que las combinaciones C10, C11 y C12 amplificaron siempre secuencias del tipo CNL y C13 y C14 de la subfamilia TNL.

#### 5.1 Pseudogenes y secuencias descartadas

Los pseudogenes fueron originalmente definidos como secuencias que se asemejan a genes conocidos pero no pueden producir una proteína funcional. El estudio de estas secuencias ha puesto de manifiesto no sólo la frecuencia con la que degeneran los genes, sino también, que muchas secuencias que se creían pseudogenes son realmente genes que codifican ARN funcionales (Chandrasekaran y Betrán, 2008).

Los pseudogenes se originan debido a la acumulación de mutaciones incapacitantes en las secuencias como, por ejemplo, inserciones nucleotídicas, deleciones y/o sustituciones que o bien desplazan el marco de lectura o lo interrumpen debido a la formación de uno o varios codones de fin prematuros (Chandrasekaran y Betrán, 2008; Podlaha y Zhang, 2010; Tutar, 2012).

La abundancia de pseudogenes en un determinado genoma normalmente depende de las tasas de duplicación y pérdida de genes (Chandrasekaran y Betrán, 2008). Algunos estudios han revelado que los genomas de plantas contienen muchos pseudogenes, por ejemplo *Arabidopsis thaliana* tiene aproximadamente 2700 pseudogenes bien definidos (Benovoy y Drouin, 2006; Zou *et al.*, 2009) mientras que *Oryza sativa* tiene alrededor de 5600 (Zou *et al.*, 2009).

En este estudio se ha utilizado ADN total de lenteja como molde para la PCR, por lo tanto, no se puede tener la certeza de que las secuencias NBS-LRR encontradas pertenezcan a genes de resistencia funcionales, de hecho, han aparecido secuencias cuyo marco de lectura no es el esperado y que han sido consideradas como pseudogenes. De las amplificaciones realizadas con las 13 combinaciones de cebadores que dieron bandas que se pudieron clonar, se obtuvieron 56 secuencias que no mostraron similitud con secuencias NBS-LRR, sino con secuencias similares a retrotransposones, a otras proteínas, o directamente no mostraban una similitud significativa con ninguna de las secuencias depositadas en las bases de datos.

De un total de 212 secuencias con alta similitud a NBS-LRR (anexo A1), 82 de ellas fueron descartadas para los análisis por diversos motivos, la mayoría por ser consideradas pseudogenes.

Catorce de estas secuencias descartadas mostraron una gran deleción central que afectaba a motivos conservados, de las cuales 12 poseían además otra mutación incapacitante, la pérdida de un nucleótido que desplazaba el marco de lectura. Shao *et al.* (2014) ya observaron secuencias cortas con una gran deleción en cuatro especies de leguminosas, *M. truncatula, G. max, C. cajan y P. vulgaris,* aunque en diversas posiciones, alguna de ellas afectando a la misma zona del dominio NBS. Luo *et al.* (2012) sugiere que estas grandes deleciones pueden ser producidas por eventos de transposición y salto de exones. Esto parece explicar lo que ha pasado en las 14 secuencias de este estudio ya que el fragmento delecionado está flanqueado por una secuencia de cinco nucleótidos (AGTTG) (Figura 4.3) que ha sido definida como TSD o "target site duplication" (Zhang, 2003).

Otras 21 secuencias fueron desestimadas para el estudio al ser secuencias parciales, a las que le faltaba al menos uno de los cebadores. Este tipo de secuencias también fue observado por Marone et al. (2003)

También se han eliminado secuencias del estudio cuando presentaban un pequeño aumento o disminución en el número de nucleótidos, generalmente entre uno y tres nucleótidos no consecutivos que llevó a la pérdida del marco de lectura correcto.

Otras secuencias descartadas al ser consideradas pseudogenes, pese a mostrar una secuencia de aminoácidos muy parecida a las secuencias válidas son las que presentaron uno o unos pocos codones de fin prematuros.

Palomino *et al.* (2006) también descartaron de sus análisis 14 secuencias que mostraron similitud a retrotransposones, 21 a otras proteínas y 41 secuencias para las que no encontraron secuencias con la similitud suficiente en la búsqueda en GenBank. Además también descartaron cuatro secuencias

que sí mostraron similitud con genes *NBS-LRR*, dos de ellas por contener uno o más codones de fin prematuros y otras dos por tener un tamaño de secuencia inadecuado. Por tanto es habitual que, además de secuencias NBS-LRR estos cebadores amplifiquen secuencias de retrotransposones y secuencias desconocidas.

Se han caracterizado 49 pseudogenes para genes NBS-LRR de *M. truncatula* (Ameline-Torregrosa *et al.,* 2008) aunque estiman que hay más, 10 en *A. thaliana* (Meyers *et al.,* 2003), 184 en *Oryza sativa* (Shang *et al.,* 2009) y 62 en *Lotus japonicus* (Li *et al.,* 2010a). Estos pseudogenes coinciden en sus características con las secuencias descartadas en esta Tesis.

Por último, varias secuencias fueron excluidas del estudio aunque no se las puede considerar pseudogenes propiamente dichos. Estas secuencias son idénticas o redundantes y únicamente se mantuvo una de cada pareja o grupo de secuencias iguales, descartando las demás. Este tipo de resultados también se obtuvieron en guisante donde Djebbi *et al.* (2015) descartaron siete de las secuencias obtenidas en su estudio por ser redundantes.

## 5.2 Análisis filogenéticos

En este estudio se ha pretendido demostrar la existencia de diferentes tipos de secuencias NBS-LRR en lenteja y analizar las relaciones filogenéticas entre ellas, para lo cual se han obtenido dendrogramas usando el modelo de Kimura-2 parámetros (Kimura, 1980). Al hacer el árbol filogenético se han obtenido 10 grupos de secuencias claramente diferenciables, con un número variable en cada grupo. Seis grupos quedaron englobados en la subfamilia CNL mientras que los otros cuatro pertenecían a la TNL. Estos cuatro grupos de secuencias TNL han coincidido con los obtenidos anteriormente por Yaish *et al.* (2004) en lenteja. En la figura 4.10b se muestra la integración de los grupos obtenidos por el Dr. Yaish con los obtenidos en este estudio.

Palomino *et al.* (2006) también obtuvieron 10 grupos al crear un árbol en su estudio de las especies *Vicia faba* y *Cicer arietinum*, pero en este caso fueron cuatro los grupos de CNL y seis los de TNL. Djebbi *et al.* (2015) también diferenciaron sus resultados en diez tipos de secuencias que nombraron de RGA1 a RGA10 aunque, al igual que Palomino *et al.* (2006) y al contrario que este estudio, había más TNLs que CNLs, siete frente a tres. En otros estudios se han obtenido números de grupos similares para otras especies, ocho para la judía, de los cuales seis muestran características TNL y dos CNL (Rivkin *et al.*, 1999), y para soja nueve similares a TNL (Kanazin *et al.*, 1996) u once, de los cuales nueve se clasificarían en TNL y dos en CNL (Yu *et al.*, 1996).

Las 130 secuencias similares a NBS-LRR analizadas en la presente Tesis se han dividido en 42 secuencias de tipo TNL y 88 secuencias de la subfamilia CNL, por lo que se podría sugerir que en *Lens* hay un mayor número de secuencias CNL que TNL. Esto coincide con los resultados obtenidos por Ameline-Torregrosa *et al.* (2008) y Song y Nan (2014) para *Medicago truncatula*, por Jupe *et al.* (2012) y Lozano *et al.* (2012) para patata y Zhong *et al.* (2015) y Jia *et al.* (2015) en rosáceas y cucurbitáceas y Lozano *et al.* (2015) para yuca. Por el contrario, Marone *et al.* (2013) afirman que los genes de tipo TNL están presentes en el genoma de dicotiledóneas con mayor frecuencia que los genes de tipo CNL ya que los genomas de especies como *A. thaliana, A. lyrata* y soja contienen de dos a seis veces más genes TNL que CNL (Guo *et al.,* 2011; Kang *et al.,* 2012).

Los análisis filogenéticos de las secuencias NBS tanto del género *Lens* como de otros géneros de plantas muestran que algunas secuencias de distintas especies, tanto del mismo género como de géneros diferentes, son más similares entre ellas, pudiendo llegar a ser idénticas, que con otras secuencias de la misma especie. Pan *et al.* (2000) han propuesto para explicar este hecho una hipótesis de diversificación génica en la evolución de los genes de resistencia a enfermedades de tipo NBS-LRR que sostiene que muchas de las duplicaciones génicas son anteriores a la divergencia de los taxones superiores de plantas. En esta Tesis se ha observado la coincidencia en secuencia entre especies diferentes, por ejemplo, Lto1.01C13, Lto1.02C3 y Lto1.04C3, que pertenecen a la especie *L. tomentosus,* son idénticas a la secuencia Lco1.05C2 de *L. culinaris* subsp *orientalis*. Esto sugiere una duplicación génica antes de la separación de las dos especies.

La mayoría de las secuencias de esta Tesis descartadas por ser idénticas a otras secuencias utilizadas en los análisis pertenecen a los grupos NT-1, NT-2 y T-4, que son los grupos más numerosos y/o homogéneos. Esto puede deberse a que la evolución de las secuencias NBS-LRR por duplicación ha producido un gran número de secuencias similares, agrupadas y con altas tasas de recombinación local (Baumgarten *et al.*, 2003; Guo *et al.*, 2011; Mondragón-Palomino y Gaut, 2005)

## 5.2.1 Cambios sinónimos y no sinónimos

Las sustituciones sinónimas son cambios de nucleótidos que no conllevan a cambios de aminoácido durante la traducción y por lo tanto serian probablemente mutaciones neutras, sin embargo las sustituciones no sinónimas sí acarrean un cambio de aminoácido, lo que puede afectar a la función de la proteína, y por lo tanto están sometidas a selección. La comparación entre la frecuencia de fijación de las sustituciones sinónimas por sitio sinónimo (dS) y no sinónimas por sitio no sinónimo (dN) permite estimar la importancia de la selección natural sobre genes concretos.

El valor de dS puede ser usado como reloj molecular para medir el tiempo transcurrido desde la duplicación génica, ya que el número de cambios sinónimos presentes puede considerarse proporcional al tiempo transcurrido desde la divergencia de las poblaciones (Peterson y Masel, 2009)

El cociente entre dS y dN ( $\omega$ ) se utiliza para caracterizar el tipo de selección y su magnitud entre parejas de secuencias. Un valor de  $\omega$  mayor de uno indica una selección depuradora o prevalencia de los cambios sinónimos, en caso de que  $\omega$  sea menor que uno refleja una selección positiva, lo que lleva a una fijación de los cambios no sinónimos y si  $\omega$  es igual a uno no hay selección neta de ningún tipo.

Analizando los datos mostrados en la tabla 4.7 vemos que los valores promedio de sustituciones sinónimas fueron más altos que los de las sustituciones no sinónimas. También se observa que en todos los casos el cociente  $\omega$  fue mayor que 1, por lo tanto, se puede sugerir que una selección depuradora predomina en este estudio, tanto a nivel global como de subfamilia. Analizando por grupos, también se observa selección depuradora en todos ellos, aunque los valores del test de neutralidad para cada pareja de secuencias (Anexo A4) demuestran que para los grupos T-1, T-2, T-3, NT-1 y NT-5 la diferencia de cambios sinónimos y no sinónimos entre las agrupaciones de sus secuencias es significativa, y para el resto de grupos muestran muy pocos o ningún valor significativo. Unos resultados similares fueron observados en el trabajo de Yaish *et al.* (2004), donde casi todos los grupos y subgrupos muestran selección depuradora, excepto el subgrupo NB que muestra selección positiva y que se agrupa con las secuencias del grupo T-4 de este estudio. Palomino *et al.* (2006)

obtienen unos resultados parecidos a Yaish *et al.* (2004), es decir una selección depuradora para casi todos los grupos excepto uno, el RGA 06, que muestra una selección positiva o diversificadora.

Donde más cambios, tanto sinónimos como no sinónimos, se produjeron fue en el conjunto total de secuencias NBS-LRR. El coeficiente  $\omega$  resultante fue bajo ( $\omega$ = 1,838), lo que se interpreta como que la selección depuradora ejercida sobre el conjunto de las secuencias analizadas no es muy intensa ya que es un valor próximo a 1. Analizando las dos subfamilias (Tabla 4.7) se vio que los valores de  $\omega$  obtenidos fueron algo mayores y muy parecido entre ellas ( $\omega$ = 2,289 para CNL y  $\omega$ = 2,378 para TNL), sin embargo, el número de sustituciones, tanto dS como dN, fue superior para TNL, aunque el análisis por grupos reveló que fue el grupo NT-5 ( $\omega$ = 5,381), de la subfamilia CNL, el que mostró mayor selección depuradora entre sus secuencias y T-3, perteneciente a la subfamilia TNL es el grupo con un menor valor de  $\omega$  ( $\omega$ = 2,134).

Zhong *et al.* (2015) en su estudio de NBS-LRR en rosáceas también llegaron a la conclusión de que ha actuado una selección depuradora sobre dichas secuencias para cinco especies de esta familia. Al igual que los datos obtenidos por subfamilias en este estudio, los TNL de los genomas de manzana, pera y albaricoque japonés mostraron distribuciones más amplias de sustituciones sinónimas que los no TNL (engloba a CNL y XNL). Estos autores concluyeron que esta observación indica que TNL tiene un proceso de duplicación anterior que los no TNL en estas tres especies.

Varios estudios sugieren que los genes presentan diferentes patrones evolutivos en función de si son de tipo TNL o CNL. Por ejemplo, Chen et al. (2010) concluyen que los genes TNL evolucionan más rápidamente que los CNL en A. thaliana y A. lyrata, y Zhang et al. (2011) llegan a la misma conclusión para soja. Por otra parte Zhong et al., (2015) estiman que el hecho de que la tasa de sustituciones sinónimas sea mayor en las secuencias TNL que en las CNL demuestra que la subfamilia TNL se duplicó antes y que tiene una tasa de evolución más rápida comparada con la CNL. Yue et al. (2012) sugieren que los genes TNLs se originaron antes que los CNL, porque las TNL se observan en briófitas, un linaje evolutivamente más antiguo que licófitas, donde se observaron los primeros genes CNL. Sin embargo, McHale et al. (2006), Yang et al. (2008) y Song y Nan (2014) afirman que la subfamilia CNL se originó antes de la divergencia de las monocotiledóneas y las dicotiledóneas, mientras que la subfamilia TNL ha aparecido más recientemente en eudicotiledóneas, después de la divergencia de las monocotiledóneas. Song y Nan (2014) atribuyen la contradicción de conclusiones a la diferencia de objetivos de la investigación, ya que McHale et al. (2006) y Yang et al. (2008) utilizaron plantas con flores como muestra y Yue et al. (2012) hizo una comparación entre los cinco reinos usando plantas no vasculares para el reino vegetal. Zhong et al. (2015), exponen la posibilidad de que las subfamilias TNL y CNL tengan diferentes tasas de evolución en respuesta a la coevolución con los patógenos. Los resultados obtenidos con las secuencias de este estudio no muestran una tendencia clara sobre el origen de cada subfamilia ya que la subfamilia TNL tiene más sustituciones, tanto sinónimas como no sinónimas totales, pero ambas subfamilias muestran grupos con un alto promedio de cambios y otros grupos con los promedios muy bajos (Figura 4.7).

#### 5.3 Motivos conservados

La mayor parte de los genes *R* clonados en plantas pertenecen a la familia NBS-LRR. Esta familia tiene unos motivos muy conservados en su secuencia, lo que ha permitido el diseño de cebadores degenerados basados en dichos motivos para aislar secuencias análogas a genes de resistencia o

*RGAs* (Resistance Gene Analogs) en un gran número de especies vegetales (Bouktila *et al.*, 2014; Djebbi *et al.*, 2015; Graham *et al.*, 2000; Kanazin *et al.*, 1996; Leister *et al.*, 1996; Meyers *et al.*, 2003; Palomino *et al.*, 2006; Timmerman- Vaughan *et al.*, 2000; Wan *et al.*, 2012; Yaish *et al.*, 2004; Yu *et al.*, 1996; Zhang *et al.*, 2011; Zhu *et al.*, 2002)

Meyers *et al.* (1999) y Lukasik y Takken (2009) describieron una serie de motivos conservados que caracterizan las secuencias de estos como TIR o TNL (siglas del dominio Toll interleukin receptor) y No-TIR o CNL.

La forma más clara de distinguir una secuencia perteneciente a la subfamilia TNL de una secuencia de la subfamilia CNL es mediante los dominios específicos localizados en la región N-terminal, pero, dado que las secuencias obtenidas en este estudio no contienen dicha región, para su clasificación se utilizaron los motivos diferenciadores que se han descrito dentro de la región o dominio NB-ARC. La conservación de los motivos en el dominio NB-ARC puede jugar un papel importante en la función de los genes de resistencia, porque alguno de los alelos con pérdida de función de los genes *NBS-LRR* son debidos a mutaciones puntuales en dichos motivos conservados dentro del dominio (Pan *et al.*, 2000; Palomino *et al.*, 2006). Los estudios de Meyers *et al.* (1999, 2003), McHale *et al.* (2006) y Lukasik y Takken (2009) se han basado en las diferencias dentro de los motivos conservados de esta región para clasificar las secuencias en la subfamilia correspondiente, bien sea TNL o CNL. Según estos autores, los motivos conservados más relevantes para esta separación son el RNBS-A y el RNBS-D que muestran una secuencia completamente distinta para TNL y CNL, tanto en composición de aminoácidos como en el tamaño del motivo conservado, y RNBS-C que tiene muy baja similitud de secuencia entre las dos subfamilias. La figura 4.6 muestra ejemplos de secuencias obtenidas en este estudio en los que se pueden apreciar estas diferencias de los motivos conservados.

Los resultados obtenidos para este estudio parecen entrar en contradicción con los datos presentados por Meyers et al. (1999) en el grado de conservación de los motivos específicos, ya que al compararlos (Tablas 4.9-4.11) parece que los motivos estarían más conservados entre genes de especies vegetales evolutivamente más distantes que entre especies del mismo género, ya que las secuencias consenso obtenidas a partir de las secuencias de Lens muestran muchas variaciones nucleotídicas (marcadas en rojo en las tablas 4.9-4.11) con respecto a los motivos conservados consensuados, sobre todo para los motivos RNBS-A de TNL y RNBS-C de las dos subfamilias. Una explicación plausible para esta contradicción sería el procedimiento usado por Meyers et al. (1999) en su estudio para buscar las secuencias con las que dedujeron los motivos conservados. Utilizaron los motivos presentes en algunos genes concretos ya descritos como genes de resistencia para buscar secuencias de aminoácidos similares en la base de datos, sin tener en cuenta que podrían existir muchos más genes NBS-LRR con secuencias distintas en los motivos conservados. Por tanto, muy posiblemente al hacer la búsqueda, las secuencias encontradas poseían unos motivos iguales o muy similares a los de la secuencia de referencia, mientras que las secuencias NBS-LRR con los motivos conservados menos similares a los escogidos no aparecían en la búsqueda, lo que habría introducido un sesgo en el grado de conservación de los motivos.

Además de usar las diferencias entre motivos conservados enteros, Meyers *et al.* (1999) utilizaron el residuo final del motivo "Kinase 2" o "Walker B" como sistema de predicción para la pertenencia a una subfamilia u otra, con una fiabilidad del 95% de los casos. Según estos autores, la presencia del aminoácido triptófano (W) en dicha posición determina que la secuencia pertenece a la subfamilia

CNL, mientras que las proteínas TNL tendrían un residuo de aspártico (D). En la figura 4.8 relativa a secuencias CNL de lenteja se puede comprobar cómo en la posición 86 de la secuencia consenso aparece únicamente el residuo de triptófano (W) que se esperaba, ya que todas las secuencias agrupadas como CNL en esta Tesis muestran dicho aminoácido en esta posición excepto Lla1.06C12, que muestra un residuo de leucina (L).

En el caso de la subfamilia TNL esta predicción no se cumple de una manera tan rotunda. La figura 4.9, relativa a las secuencias TNL de lenteja, muestra que, pese a que el aspártico (D) aparece en la posición 86 como aminoácido mayoritario, no es el único ya que también hay secuencias que en dicha posición tienen residuos de asparagina (N) o serina (S).

Analizando la subfamilia TNL por grupos, se puede observar que los grupos de secuencias aminoacídicas de lenteja que sí cumplen con la predicción del estudio de Meyers *et al.* (1999) son las correspondientes a T-3 y T-4, ya que todas las secuencias agrupadas contienen aspártico en la posición descrita. Pero de las 15 secuencias de lenteja que componen el grupo T-1 sólo una tiene ácido aspártico en la posición correspondiente (Lto1.02C1), otra (LccB2.08C9) tiene treonina (T) mientras que 13 tienen asparagina (N). En el grupo T-2 sólo la secuencia LccB2.10C9 contiene aspártico (D) mientras que las otras cuatro secuencias del grupo muestran un residuo de serina (S).

Observando las secuencias de las especies de *Medicago, Cicer, Phaseolus* y *Glycine* recogidas para la comparación entre especies, se ve un patrón similar al obtenido en *Lens* en este estudio, una gran consistencia en el caso de la subfamilia CNL y una mayor variabilidad en el caso de la subfamilia TNL. En estas especies también se mantiene en gran medida la distribución por grupos observada en las secuencias TNL de lenteja.

La variabilidad del residuo final del motivo conservado "Kinase 2" en la subfamilia TNL ya fue descrita por Yaish *et al.* (2004) para lenteja, observando los mismos cambios del ácido aspártico (D) por los aminoácidos asparagina (N), serina (S) y treonina (T), por Palomino *et al.* (2006) en haba y garbanzo, en este caso únicamente se observaron cambios por asparagina (N) o serina (S). También Djebbi *et al.* (2015) en su estudio sobre guisante observaron un cambio de residuo donde debería estar el aspártico (D), pero esta vez por un residuo diferente, glicina (G).

Además de la conservación de los motivos característicos del dominio NB-ARC, algunos aminoácidos aislados están conservados entre muchas de las secuencias contribuyendo a la estabilidad de la secuencia consenso (Leister *et al.,* 1996). Como mostraron Pan *et al.* (2000) y Palomino *et al.* (2006) en sus respectivos trabajos, donde analizaron la conservación invariable de una fenilalanina (F) en un fragmento de secuencia que denominaban como NBS-II (coincide en el espacio con el dominio RNBS-A en las secuencias CNL) tanto en la subfamilia TNL como en la CNL y, cinco residuos corriente abajo del residuo F, se mantienen altamente conservados una fenilalanina (F) o un triptófano (W) en el grupo TNL y en el grupo CNL respectivamente. La figura 5.1 muestra que en las secuencias obtenidas en este trabajo ese grado de conservación se mantiene en las posiciones mencionadas.

Aunque GGMGKTT es la secuencia de aminoácidos más común para el motivo P-loop en lenteja (Yaish *et al.,* 2004), en este estudio solo las secuencias de la combinación C9 presentaron mayoritariamente este patrón mientras que las secuencias de las combinaciones C1 a C6 tuvieron como secuencia consenso GGVGKTT (Figura 5.1). Esto puede deberse a que, al ser ésta la zona utilizada para la unión del cebador en muchas amplificaciones, la secuencia resultante viene marcada

por la propia secuencia de dicho cebador, que solo en el caso del cebador B1 está degenerada en la posición correspondiente a ese aminoácido, por lo tanto, C9 es la única combinación de cebadores con la que se puede obtener amplicones con un P-loop distinto a GGVGKTT.

		P-loop	RNBS-A
CNL	LccB1.02C1 LccB2.01C6 LccB2.18C9 LccB5.11C10 LccA1.13C11 LccB2.01C12	GGVGKTTLAQLAYNDEKVHKI GGVGKTTLAQLVYNNEKVEQI GGMGKTTLAKLLYNDPDLKGI TLAQLVYNDEKVRKI YNHDTIKQ DSRIAN	HFDLIAWAC-VSEDFDVFRVTKTLLESVTKRP- HFDFKVWVC-VSEDFDVVRLTKSLLESVVRNTT KFEVRGLAH-IPKDFDHVTVTKTILESVTLGT- HFDLVAWAC-VSEDFDVFRVTKTLLESVTKTP- IFDVQAWVC-VSDDFDVLKVTKFIAEEVGSSC- HFECKAWVF-VSEEYRDKRKDVLQGILRGVDPLA-
TNL	Lni1.01C1 LccB1.15C2 LccA1.01C3 LccB1.04C6 LccB2.10C9 Ler1.06C13 LccA1.03C14	GGVGKTTIAKATYNHMSG GGVGKTTLATVLYDRISQ GGVGKTTLAGVLYGRISH GGVGKTTIAKATYNHMSG GGMGKTTLARAVYNSISR	SFEAKSFLVDVKTVWDSGKGLLSLQKQLLSDVCKT QFVACCFIDDVSKNYRLHDGPLGVQKQILDQTLGQD- QFNACCLIDDVSKVFK-YDGPIGVQKQILHQTLGEE- SFEGKSFLEDVNKVWEGGKGLVSLQKQLLSDVYKTT- KFDSSSFLVDVRENS-MKHGLVHLQETLLLHLLDE DFECRSFLANIREVWEQVGGQVCLQEQLMYDIFKE QFPVYCLIDDLGKTFR-DNGPIGAQKLILHQTLIEE-

**Figura 5.1.** Fragmento de secuencias de aminoácidos del dominio NBS en el que se muestran sombreados dos puntos de conservación característicos. (Fragmento modificado de la figura 4.7 de los resultados). También aparece encuadrado el fragmento de secuencia correspondiente al motivo conservado P-loop.

## 5.3.1 Contenido en nucleótidos y aminoácidos

Las secuencias NBS analizadas en este estudio, por lo general, son ricas en adenina (A) y timina (T) y pobres en citosina (C). La cantidad de guanina (G) es más variable (Tabla 4.6). Al analizar todas las secuencias se observa que los rangos de variación para los cuatro nucleótidos son amplios y la única tendencia clara es la baja frecuencia de citosina. Por subfamilias, en los grupos TNL se puede apreciar que las frecuencias de adenina y timina son más altas que las de citosina y guanina, mientras que en los CNL esto no es tan claro. Además la diferencia entre los valores de los rangos es menor para TNL que para CNL en todos los nucleótidos excepto C. Esto podría deberse a las diferencias entre grupos dentro de una misma subfamilia. Neyra (2015) obtiene resultados equiparables en otros tipos de secuencias de *Lens*, que también mostraron un alto contenido de adenina y timina y baja cantidad de citosina.

Las lentejas como el resto de leguminosas grano son deficitarias en aminoácidos azufrados. Por lo tanto, al analizar los aminoácidos menos abundantes se puede observar que la metionina es un aminoácido poco frecuente en ambas subfamilias pero en las secuencias TNL es menos frecuente el triptófano en tres de los cuatro grupos, llegando a ser inexistente en T-4. Otro aminoácido con presencia característicamente baja en la subfamilia TNL es la prolina, esto podría deberse a la baja frecuencia del nucleótido citosina, codificado por codones con dos o tres citosinas. En la subfamilia CNL, además de la metionina, los niveles más bajos de aminoácidos corresponden a tirosina, cisteína, histidina y en el caso de NT-2 glutamina (Tabla 4.8).

Por el contrario, el aminoácido más abundante en casi todos los grupos es la leucina excepto en NT-1 que es la valina. En el caso del segundo aminoácido más abundante hubo una gran diversidad de resultados: serina para NT-1 y T-1, aspártico para NT-2 y NT-4, lisina en NT-3 y T-2, glutámico para NT-5 y NT-6, isoleucina en T-3 y asparagina en T-4. En las secuencias de *Lens* estudiadas por Neyra (2015) la leucina también es uno de los aminoácidos más abundantes, junto con glutámico, arginina, alanina y glicina.

Iqbal *et al.* (2005) hacen un análisis de composición de aminoácidos en proteínas obtenidas a partir de semilla en leguminosas, donde se puede comprobar que los aminoácidos menos abundantes en lenteja, guisante, garbanzo y alubia son triptófano, metionina y cisteína y los más abundantes son glutámico, aspártico y leucina. Este hecho pone de manifiesto que las secuencias NBS-LRR de lenteja siguen la pauta general de composición aminoacídica de los genomas de leguminosas.

#### 5.4 Comparación con otras especies

Los genes *R* del tipo NBS-LRR se distribuyen de una manera desigual en el genoma de las plantas y muchos forman grupos multigénicos locales. Esto puede favorecer la variación génica (Michelmore y Meyers, 1998; Sun *et al.*, 2001; Young, 2000). Un claro ejemplo de agrupación localizada en el genoma lo constituye *Medicago truncatula* que presenta un gran grupo de NBS-LRR, mayoritariamente de tipo CNL, en el cromosoma 3 y otro, esta vez exclusivamente TNL, en el cromosoma 6, además de un tercer grupo, menos numeroso en el cromosoma 4 (Figura 5.2) (Ameline-Torregrosa *et al.*, 2008; Song y Nan, 2014).



**Figura 5.2.** Distribución de genes *NBS-LRR* codificados en los ocho cromosomas de *M. truncatula* según Ameline-Torregrosa et al. (2008) adaptado a la base de datos Mt v3.5.1. En color morado se representan los genes de la subfamilia TNL y en color amarillo los genes correspondientes a la subfamilia CNL.

Se ha demostrado la existencia de sintenia entre *Medicago* y otras especies de leguminosas (Alo *et al.*, 2011; Aubert *et al.*, 2006; Choi *et al.*, 2004; Choudhary *et al.*, 2009; Gujaria-Verma *et al.*, 2014; Hougaard *et al.*, 2008; Kalo *et al.*, 2004; Kaur *et al.*, 2011; Phan *et al.*, 2005; Phan *et al.*, 2007; Sharpe *et al.*, 2013). Lenteja y *Medicago* son especies estrechamente relacionadas (Gujaria-Verma *et al.*, 2014; *et al.*, 2013).

2014) y se estima que divergieron hace entre 24 (Lavin *et al.,* 2005) y 38 (Sharpe *et al.,* 2013) millones de años, por lo que en este estudio se postula la idea de que los grupos de secuencias NBS-LRR obtenidos de *Lens* que tengan una mayor homología con los grupos de genes de *Medicago* puedan estar localizados en el genoma formando grupos de genes sinténicos en posiciones equivalentes.

En el archivo de excel A2 (adjunto en el CD que acompaña a esta Tesis) se recogen los resultados de la comparación de las secuencias de *Lens* obtenidas en este estudio con genes anotados de *M. truncatula* a través de Phytozome 5.0. Se puede comprobar que secuencias que forman parte de la superagrupación de genes *NBS-LRR* localizado en la parte distal del cromosoma 3 de *Medicago* (Figura 5.2) muestran homología con los grupos de la subfamilia CNL (Figura 5.3), siendo muy numerosas las coincidencias de los dos grandes grupos NT-1 y NT-2 en la misma zona (Figuras 4.17, 4.18 y 4.31a) y de NT-3 en una zona algo superior (Figuras 4.19 y 4.31a). Para el resto de grupos de la subfamilia CNL la homología con el cromosoma 3 de *Medicago* es mucho menor, por lo que en el análisis comparativo realizado con genes anotados en la web Legume Information System (LIS), de los tres grupos restantes, sólo NT-5 mostró similitud en la zona correspondiente del cromosoma 3 de *Medicago* y solamente con una secuencia (Figuras 4.21a y 4.31a). Ni NT-4 ni NT-6 mostraron similitud con el cromosoma 3 bajo los criterios escogidos en el análisis.



**Figura 5.3.** Distribución de genes *NBS-LRR* de *M. truncatula* que mostraron similitud con las secuencias de lenteja. En color morado se representan los genes de la subfamilia TNL y en color amarillo los genes correspondientes a la subfamilia CNL.

Aunque esta superagrupación de *Medicago truncatula* la forman mayoritariamente genes de la subfamilia CNL, también contiene algunos genes de la subfamilia TNL (Ameline-Torregrosa *et al.*, 2008) (Figura 5.2), lo que coincide con el hecho de que el subgrupo T-1.3 de este estudio (Figuras 4.25 y 4.31b), perteneciente a la subfamilia TNL (Figura 5.3), también mostrara homología con la zona del cromosoma 3 de *Medicago* que alberga dicha superagrupación.

La otra superagrupación mencionada por Ameline-Torregrosa *et al.* (2008), localizada en la parte distal del cromosoma 6 de *M. truncatula* (Figura 5.2), estaría formada por genes de la subfamilia TNL. En este estudio se han encontrado similitudes con dicha superagrupación de los grupos T-3 y T-4, en mayor o menor medida y del subgrupo T-2.2 (Figura 4.31b). Como era lógico esperar y coincidiendo con lo descrito por Ameline-Torregrosa *et al.* (2008), las similitudes encontradas con la superagrupación del cromosoma 6 se ha limitado a varios grupos, todos ellos pertenecientes a la subfamilia TNL (Figura 5.3).

Dos de los subgrupos de T-1, T-1.1 y T-1.2 muestran similitud con el cromosoma 4 de *Medicago*, al igual que los grupos T-3 y T-4 (Figura 5.3), pero sólo los subgrupos de T-1 coinciden en el lugar donde Ameline-Torregrosa *et al*. (2008) mencionaron otra gran agrupación de genes *NBS-LRR* (Figura 5.2), ya que las secuencias similares encontradas para T-3 y T-4 se encontraban por debajo de la zona indicada (Figura 4.31b).



También se han localizado superagrupaciones de genes *NBS-LRR* en *Glycine max,* en los cromosomas 3, 6, 13, 15, 16 y 18 (Kang *et al.,* 2012; Nepal y Beson, 2015) (Figura 5.4).

**Figura 5.4.** Distribución de genes *NBS-LRR* en *Glycine max*. En color morado se representan los genes de la subfamilia TNL y en color amarillo los genes correspondientes a la subfamilia CNL.

Al analizar los resultados de la localización de homologías de las secuencias NBS-RR de *Glycine max* con las secuencias de lenteja se vio que hay un menor número de secuencias homólogas (Figura 5.5) que en *Medicago* (Figura 5.3). Esto puede deberse a que el ancestro común más próximo entre lenteja y soja es anterior al ancestro entre lenteja y *Medicago*.

De los grupos que sí mostraron homología con *G. max,* NT-4 (Figuras 4.20 y 4.32a) parece ser el único de la subfamilia CNL (Figura 5.5) que coincide con una superagrupación de secuencias NBS-LRR descritas en *Glycine* (Kang *et al.,* 2012), concretamente con la localizada en el cromosoma 3 (Figura 5.4).

También hay homología de la subfamilia TNL de lenteja con la superagrupación del cromosoma 3 de *Glycine* (Figura 5.5), en concreto del subgrupo T-1.1 (Figuras 4.23b y 4.32b). Por otro lado, el subgrupo T-1.3 (Figuras 4.25b y 4.32b) muestra similitud con otra gran superagrupación de secuencias TNL de *G. max* localizada en el cromosoma 16.



**Figura 5.5.** Distribución de genes *NBS-LRR* de *Glycine max* que mostraron similitud con las secuencias de lenteja. En color morado se representan los genes de la subfamilia TNL y en color amarillo los genes correspondientes a la subfamilia CNL.

Por otro lado, además de estas superagrupaciones de secuencias en los genomas también pueden encontrarse similitudes con secuencias aisladas. Por ejemplo, el grupo NT-6 de la subfamilia CNL está formado por 11 secuencias de distintas especies de lenteja: cuatro de *L. culinaris* subsp *culinaris*, una de *L. culinaris* subsp *orientalis*, dos de *L. nigricans*, tres de *L. lamottei* y una de *L. tomentosus*. Las once muestran una alta similitud entre ellas y todas se han obtenido con la combinación de cebadores C12. Al comparar las secuencias de este grupo con otras especies se observó que presentaba similitud con un único gen concreto en el cromosoma 5 en el caso de *Medicago* (Figuras 4.11a y 4.22a) y en el cromosoma 2 de *Phaseolus* (Figura 4.14a), una secuencia sin localización cromosómica específica en *Cicer* (Figura 4.13a) y con dos genes de *Glycine*, uno en el cromosoma 1 y

otro en el cromosoma 11 (Figura 4.12a y 4.22b). Esta exclusividad en la homología con genes específicos también se observó en las secuencias incluidas en el diverso grupo T-2 de la subfamilia TNL, donde se puede ver la división entre subgrupos comentada anteriormente. La secuencia LccB2.10C9, que da lugar al subgrupo T-2.2, muestra similitud con una única secuencia en cada una de las otras especies comparadas, excepto *Phaseolus* donde coincide con más de una. Esta secuencia está localizada en la zona distal del cromosoma 6 de *Medicago* (Figuras 4.11b y 4.27). Según lo que se observa en la Figura 4.12b, la secuencia LccB2.10C9 tiene cierta homología con una secuencia de *Glycine* localizada en el cromosoma 6, aunque el grado de homología es menor al umbral usado en el análisis comparativo con la web LIS, por lo que no se localizó en este último análisis. Por otro lado, las cuatro secuencias que forman el subgrupo T-2.1 muestran similitud con una sola secuencia de *Cicer* (Figura 4.13b) pero en el caso de *Medicago* tienen similitud con dos secuencias, una localizada en el cromosoma 7 (Figura 4.26), si bien al analizar el árbol de la figura 4.11b se puede determinar que dos secuencias de T-2.1, LccB1.34C9 y LccB2.27C9, muestran una mayor similitud con el gen localizado en el cromosoma 3, mientras que las dos secuencias restantes de dicho subgrupo parecen asemejarse más al presente en el cromosoma 7.

Para hacer una estimación de la cobertura sobre el tipo y número de secuencias NBS- LRR presentes en el genoma de *Lens* que se ha conseguido en el presente trabajo, se han realizado dos aproximaciones. En primer lugar se ha analizado el número de secuencias diferentes que se obtenían progresivamente en los procesos de clonación, esperando que a medida que se fuera llegando al número real de secuencias existentes en lenteja, dicho número se fuera reduciendo progresivamente. En la figura 4.2 no se detecta esta tendencia, lo que sugiere que todavía quedarían un número elevado de secuencias diferentes por describir, al menos para los grupos y subgrupos amplificados con las combinaciones C6 (NT-1, T-1.1 y T-1.2) y C10 (NT-1 y NT-2).

En segundo lugar se ha hecho un estudio comparativo entre las secuencias obtenidas en la especie cultivada *L. culinaris* subsp *culinaris* con las secuencias descritas en *Medicago* obtenidas de la base de datos Medicago truncatula v 3.5. En relación a los tipos de secuencias, únicamente las pertenecientes al subgrupo T-1.1 no se han encontrado en la especie cultivada de lenteja, si bien si se han detectado en las especies silvestres. En cuanto al número de secuencias dentro de cada grupo y el número de secuencias en *M. truncatula* (Figura 5.6) se puede comprobar que la mayoría de ellos muestra un número mayor de secuencias en *Medicago*, esto podría ser un indicativo de que no se han encontrado todas las secuencias NBS-LRR de *Lens*.

También encontramos el caso contrario, que varias secuencias de *Lens* muy parecidas entre sí coincidan con una sola secuencia de *M. truncatula*, es el caso del grupo NT-6 comentado anteriormente, en el cual cuatro secuencias de *L. culinaris* presentan homología con una única secuencia de *M. truncatula*, o el caso del subgrupo T-2.1, donde cuatro secuencias de *Lens* coinciden con dos de *Medicago*. Esto implica que no se puedan deducir los ortólogos directos con genes de otras especies en ningún caso, excepto la secuencia LccB2.10C9, ya que hay varias secuencias de *Lens* muy parecidas entre sí que presentan similitud con una sola secuencia o un grupo de secuencias homólogas de las otras especies de la comparación y no se podría concretar fielmente a cual correspondería. Este resultado es esperable en genes que frecuentemente presentan procesos de duplicación en tándem, como son los genes *NBS-LRR* (Zhu *et al.*, 2002). Adicionalmente, la no identificación de genes ortólogos puede deberse a un alto grado de divergencia entre los genes



ortólogos en las diferentes especies, la presencia de genes específicos en lenteja o a la descripción incompleta de los genes en los genomas de las especies modelo actuales (Sharpe *et al.*, 2013).

Nº de secuencias por grupo enLens culinaris subsp. culinaris

**Figura 5.6.** Gráfica de correlación entre el número de secuencias de la especie *Lens culinaris* subsp *culinaris* de cada grupo o subgrupo y el número de secuencias homólogas de *M. truncatula*.

La gráfica 5.6 también muestra que los grupos más numerosos de *Lens*, NT-1 y NT-2, presentan homología con una de las agrupaciones más numerosas de secuencias NBS-LRR de *Medicago*, esto puede deberse a las frecuentes duplicaciones durante la evolución de las secuencias NBS-LRR de las leguminosas (Shao *et al.*, 2014).

Varios análisis de comparativos realizados en esta Tesis muestran unos resultados distintos para cada uno de los diez grupos dependiendo de la especie con la que se busca homología, por ejemplo, *M. truncatula* muestra similitud con los diez grupos en los que se dividen las secuencias de lenteja de este estudio tanto en los árboles y en el análisis de componentes principales como en la búsqueda de genes anotados realizada en la web LIS. *Cicer*, también muestra similitudes en todos los grupos, aunque NT-1 y NT-2 queden agrupadas en la comparación debido a que muestran similitud con las mismas secuencias de *Cicer*, al igual que T-3 y T-4 (Figura 4.13). Pero *Glycine* y *Phaseolus* sólo muestran similitud con algunos de los grupos de las secuencias de lenteja. En el caso de la subfamilia CNL ambas especies muestran similitud solamente con dos grupos, NT-5 y NT-6 (Figuras 4.12a y 4.14a) mientras que para la subfamilia TNL las secuencias de *Glycine* se agrupan con todos los grupos de dicha subfamilia en el árbol de la figura 4.12b y *Phaseolus* muestra secuencias similares con los grupos T-1 y T-2 (Figura 4.14b). Esto coincide con los resultados de Palomino *et al.*, (2006) que

consiguen obtener similitudes de las secuencias amplificadas con las combinaciones C12 (forman el grupo NT-6), C13 (forman parte del grupo T-1) y C14 (forman parte del grupo T-3, concretamente T-3.1) con *Glycine max*. Algo parecido les ocurre a Song y Nan (2014) en su trabajo de comparación entre secuencias NBS-LRR de *M. truncatula* con *G. max* donde observan que algunos grupos sólo tienen secuencias de *Medicago* y otros tienen secuencias de las dos especies, indicando el mismo origen de los genes *NBS-LRR* de *G. max* y *M. truncatula* que forman parte de estos últimos grupos. Esto puede extrapolarse a los datos comparativos para estas dos especies obtenidos en este estudio. Por lo tanto, después de analizar los resultados de esta Tesis y compararlos con los obtenidos en otros estudios queda probada la existencia de homología entre las secuencias NBS-LRR que forman parte del genoma de *Lens* y las secuencias NBS-LRR de las especies *M. truncatula*, *G. max*, *C. arietinum* y *P. vulgaris*, aunque la similitud entre secuencias potencialmente homólogas es variable, y teniendo en cuenta que los datos sobre los genes *NBS-LRR*, son parciales en *Lens*, y probablemente incompletos en otras especies, la identificación de genes ortólogos es aún muy imprecisa.

# Conclusiones
## 6. CONCLUSIONES

Cumpliendo el objetivo principal de esta Tesis de ampliar el conocimiento de las secuencias NBS-LRR presentes en las especies del género *Lens* se puede concluir que:

1. La mayoría de las combinaciones de cebadores degenerados y no degenerados diseñados a partir de motivos conservados de secuencias NBS-LRR de otras especies amplificaron secuencias similares a NBS-LRR en las especies del género *Lens.* Se obtuvieron 213 secuencias de las cuales 130 pudieron utilizarse en los análisis filogenéticos.

2. El 90% del total de las secuencias similares a NBS-LRR presentaron la secuencia completa para la región amplificada., de las cuales aproximadamente un 13% estaban duplicadas y un 16% se consideraron pseudogenes porque mostraban alguna mutación sin sentido y/o de cambio en el marco de lectura.

3. Se han amplificado y caracterizado secuencias de las dos subfamilias en las que se dividen los genes *NBS-LRR*, observándose el doble de secuencias de la subfamilia CNL (88) que de la subfamilia TNL (42).

4. Ambas subfamilias mostraron una división en grupos de secuencias similares entre sí, distinguiéndose seis en CNL y cuatro en TNL. Tres de los grupos de la subfamilia TNL fueron a su vez divididos en subgrupos.

5. Se caracterizaron los motivos conservados descritos en estas secuencias. Generalmente el grado de conservación de los motivos fue mayor en las secuencias de la subfamilia CNL que en las TNL, particularmente en el motivo "RNBS-A" y el nucleótido final del motivo "Kinase 2", ambos puntos clave para la distinción entre subfamilias.

6. Los análisis filogenéticos indican que en todos los grupos ha actuado una selección depuradora.

7. Se ha podido establecer una relación de homología entre las secuencias NBS-LRR halladas en el género *Lens* con secuencias NBS-LRR conocidas de las especies *M. truncatula, G. max, C. arietinum y P. vulgaris.* 

8. En lenteja se presentan agrupaciones génicas con una gran cantidad de secuencias muy similares, comparables a las superagrupaciones que muestra *M. truncatula* en varios de sus cromosomas.

## Bibliografía

## 7. BIBLIOGRAFÍA

- Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE (1998) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two *R* gene-mediated signaling pathways in *Arabidopsis*. Proc Natl Acad Sci USA 95: 10306-10311.
- Ade J, De Young BJ, Golstein C, Innes RW (2007) Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. Proc Natl Acad Sci USA 104: 2531-2536.
- Ahmad M, McNeill D, Seedcole J (1997) Phylogenetic relationships in *Lens* species and their interspecific hybrids as measured by morphological characters. Euphytica 94: 101–111.
- Almeida-Costa GE, Queiroz-Monici KS, Machado Reis SMP, Costa de Oliveira A (2007) Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. Food Chem 94: 327–330.
- Alo F, Furman BJ, Akhunov E., Dvorak J, Gepts P (2011) Leveraging genomic resources of model species for the assessment of diversity and phylogeny in wild and domesticated lentil. J Hered 102: 315-329.
- Alonso N (2008) Crops and agriculture during the Iron Age and Late Antiquity in Cerdanyola del Vallès (Catalonia, Spain). Veg Hist Archaeobot 17: 75–84.
- Alonso Ponga J, Cristóbal D (1996) La lenteja. En: F Franco Jubete, A Ramos Monreal (coords) El cultivo de las leguminosas grano en Castilla y León. Consejería de Agricultura y Ganadería. Junta de Castilla y León, pp: 399-442.
- Altaf M, Venkateshwar M, Srijana M, Reddy G (2007) An economic approach for L–(+) lactic acid fermentation by *Lactobacillus amylophilus* GV6 using inexpensive carbon and nitrógeno sources. J Appl Microbiol 103: 372–380.
- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of Nucleotide-Binding Site-Leucine-Rich Repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146: 5-21.
- Anuario de la UPA (2006) Leguminosas grano. Informe socioeconómico de la agricultura española. Agricultura española 2005/sectores. www.upa.es/anuario\_2006/pag\_184-207\_sectores.pdf.
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–218.
- Aubert G, Morin J, Jacquin F, Loridon K, Quillet M, Petit A, Rameau C, Lejeune-Henaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112: 1024– 1041.
- Aydin H, Buluta Y, Yerlikaya C (2008) Removal of copper (II) from aqueous solution by adsorption onto low-cost adsorbents. J Environ Manag 87: 37–45.

- Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site leucine- rich repeat genes in cereals. Genome Res 12: 1871–1884.
- Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP (1997) Signaling in plant-microbe interactions. Science 276: 726–733.
- Balyan HS, Houben A, Ahne R (2002) Karyotype analysis and physical mapping of 18S–5.8S–25S and a 5S ribosomal RNA loci in species of genus *Lens* Miller (*Fabaceae*). Caryologia 55: 121–128.
- Barulina H (1930) Lentils of the USSR and other countries. Bull Appl Bot Genet Plant Breed 40: 265–304 USSR.
- Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics 165: 309–19.
- Baysal Z, Uyar F, Dogru M, Alkan H (2008) Production of extracellular alkaline alpha-amylase by solid state fermentation with a newly isolated *Bacillus* sp. Prep Biochem Biotechnol 38: 191–197.
- Bendahmane A, Farnham G, Moffett P, Baulcombe DC (2002) Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the *Rx* locus of potato. Plant J 32: 195-204.
- Bennet MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Phil Trans Roy Soc Lond B 274: 227–274.
- Benovoy D, Drouin G (2006) Processed pseudogenes, processed genes, and spontaneous mutations in the *Arabidopsis* genome. J Mol Evol 62: 511–522.
- Bent AF (1996) Plant disease resistance genes, function meets structure. Plant Cell 8: 1757–1771.
- Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN (2011) Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. Cell Host Microbe 9: 200-211.
- Bertioli DJ, Leal-Bertioli SCM, Lion MB, Santos VL, Pappas G Jr, Cannon SB, Guimaraes PM (2003) A large scale analysis of resistance gene homologues in *Arachis*. Mol Genet Genom 270: 34-45.
- Bhattacharjee S, Halane MK, Kim SH, Gassmann W (2011) Pathogen effectors target *Arabidopsis* EDS1 and alter its interactions with immune regulators. Science 334: 1405–1408.
- Bhatty RS (1988) Composition and quality of lentil (*Lens culinaris* Medikus): Rev Can Food Sci Tech J 21: 144-160.
- Bieri S (2004) RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. Plant Cell 16: 3480-3495.
- Boller T, Felix G (2009) A renaissance of elicitors: perception of microbe- associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol 60: 379–406.

- Bouktila D, Habachi-Houimli Y, Khalfallah Y, Mezghani-Khemakhem M, Makni M, Makni H (2014) Characterization of novel wheat NBS domain containing sequences and their utilization, *in silico*, for genome scale *R* gene mining. Mol Genet Genom 289: 599-613.
- Brugmans B, Wouters D, van Os H, Hutten R, van der Linden CG, Visser RGF, van Eck HJ, van der Vossen EAG (2008) Genetic mapping and transcription analyses of resistance gene loci in potato using NBS profiling. Theor Appl Genet 117: 1379-1388.
- Buscaill P, Rivas S (2014) Transcriptional control of plant defence responses. Curr Opin Plant Biol 20: 35–46.
- Buxó R (1991) Algunos aspectos sobre la presencia de leguminosas en el Mediterraneo peninsular: nuevos datos de investigación de restos paleocarpológicos. En A. Vila (eds) Arqueología 101-114. Madrid: CSIC.
- Buxó R (1997) Arqueología de las Plantas. Crítica, Barcelona, Spain.
- Caccialupi P, Ceci LR, Siciliano RA, Pignone D, Clemente A, Sonnante G (2010) Bowman–Birk inhibitors in lentil: heterologous expression, functional characterization and anti-proliferative properties in human colon cancer cells. Food Chem 120: 1058–1066.
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KFX, Rogers J, Quétier F, Oldroyd GE, Debellé F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, van de Peer Y, Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA 103: 14959–14964.
- Castell AG, Cliplef RL (1990) Methionine supplementation of barley diets containing lentils (*Lens culinaris*) or soybean meal: live performance and carcass responses by gilts fed by libitum. Can J Anim Sci 70: 329-332.
- Catanzariti AM, Dodds PN, Ve T, Kobe B, Ellis JG, Staskawicz BJ (2010) The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. Mol Plant Microbe Interact 23: 49-57.
- Catlett MG, Kaplan KB (2006) Sgt1p is a unique co-chaperone that acts as a client adaptor to link Hsp90 to Skp1p. J Biol Chem 281: 33739-33748.
- Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds P (2014) A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. Front Plant Sci 5: 606: 1-10.
- Chan SL, Mukasa T, Santelli E, Low LY, Pascual J (2010) The crystal structure of a TIR domain from *Arabidopsis thaliana* reveals a conserved helical region unique to plants. Protein Sci 19: 155-161.
- Chandrasekaran C, Betrán E (2008) Origins of new genes and pseudogenes. Nat Educ 1: 181-199.
- Chen QH, Han ZX, Jiang HY, Tian DC, Yang SH Strong (2010) Positive Selection Drives Rapid Diversification of *R*-Genes in *Arabidopsis* Relatives. J Mol Evol 70: 137–48.

- Chen JY, Huang JQ, Li NY, Ma XF, Wang JL, Liu C, Liu YF, Liang Y, Bao YM, Dai XF (2015) Genomewide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium* wilt. BMC Plant Biol 15: 148
- Chen XM, Line RF, Leung H (1998) Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. Theor Appl Genet 97: 345-355.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124: 803–814.
- Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101: 15289–15294.
- Choudhary S, Sethy NK, Shokeen B, Bhatia S (2009) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. Theor Appl Genet 118: 591-608.
- Cohen SN, Chang ACY, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic Transformation of *Escherichia coli* by R-factor DNA. Proc Natl Acad Sci USA 69: 2110-2114.
- Collier SM, Moffett P (2009) NB-LRRs work a 'bait and switch' on pathogens. Trends Plant Sci 14: 521-529.
- Cordero JC, Skinner DZ (2002) Isolation from alfalfa of resistance gene analogues containing nucleotide binding sites. Theor Appl Genet 104: 1283–1289.
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Res 14: 1188-1190.
- Cuadrado C, Grant G, Rubio LA, Muzquiz M, Bardocz S, Pusztai A (2002) Nutritional utilization by the rat of diet based on lentil (*Lens culinaris*) seed meal or its fractions. J Agri Food Chem 50: 4371–4376.
- Cubero JL, Pérez de la Vega M, Fratini R (2009) Origin, phylogeny, domestication and spread. En: W Erskine, FJ Muehlbauer, A Sarker, B Sharma (eds) The Lentil: Botany, Production and Uses. CABI Press, Wallingford, UK, pp 13–33.
- Cui H, Tsuda K, Parker JE (2015) Effector-triggered immunity: From pathogen perception to robust defense. Annu Rev Plant Biol 66: 487–511.
- Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 411: 826-833.
- Danot O, Marquenet E, Vidal-Ingigliardi D, Richet E (2009) Wheel of life, wheel of death: a mechanistic insight into signaling by STAND proteins. Structure 17: 172-182.
- David P, Sévignac M, Thareau V, Catillon Y, Kami J, Gepts P, Langin T, Geffroy V (2008) BAC end sequences corresponding to the B4 resistance gene cluster in common bean: a resource for markers and synteny analyses. Mol Genet Genom 280: 521–533.

- De Candolle AP (1882, reimpreso en 1967) Origins of cultivated species. Hafner, London, UK, p 468.
- De Young BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7: 1243-1249.
- Deng Z, Huang S, Ling P, Chen C, Yu C, Weber CA, Moore GA, Gmiter FG Jr (2000) Cloning and characterization of NBS-LRR class resistance gene candidate sequences in citrus. Theor Appl Genet 101: 814–822.
- Deslandes L, Rivas S (2012) Catch me if you can: Bacterial effectors and plant targets. Trends Plant Sci 17: 644–655.
- Djebbi S, Bouktila D, Makni H, Makni M, Mezghani-Khemakhem M (2015) Identification and characterization of novel NBS-LRR resistance genes analogues from the pea. Genet Mol Res 14: 6419-6428.
- Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CI, Ayliffe MA, Kobe B, Ellis JG (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Natl Acad Sci USA 103: 8888-8893.
- Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet 11: 539-548.
- Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR, Dry IB (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. Theor Appl Genet 104: 610–618.
- Dong J, Chen C, Chen Z (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol 51: 21–37.
- Durán Y, Pérez de la Vega M (2004) Assessment of genetic variation and species relationships in a collection of Lens using RAPD and ISSR. Spanish J Agri Res 4: 538–544.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res 32: 1792-1797.
- Eitas TK, Dangl JL (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr Opin Plant Biol 13: 472-477.
- Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3: 278–284.
- Ellwood SR, Phan HTT, Jordan M, Hane J, Torres AM, Avila CM, Cruz-Izquierdo S, Oliver RP (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genomics 9: 380.
- 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, Rihawe S, Capper BS (1990) Variation in lentil straw quality. Anim Feed Sci Tech 28: 61-69.
- FAOSTAT: http://faostat.fao.org. Datos revisados el día 24-04-2015.
- Farnham G, Baulcombe DC (2006) Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. Proc Natl Acad Sci USA 103: 18828-18833.
- Ferguson ME, Maxted N, Van Slageren M, Robertson LD (2000) A re-assessment of the taxonomy of *Lens* Miller (*Leguminosae, Papilionoideae, Vicieae*). Bot J Linn Soc 133: 41–59.
- Flor HH (1955) Host-Parasite interaction in flax rust-its genetics and other implications. Phytopath 45: 680-685.
- Flor HH (1971) Current status of the gene-for-gene concept. Annu Rev Phytopath 9: 275-296.
- Ghavidel R, Prakash J (2007) The impact of germination and dehulling on nutrients, antinutrients, *in vitro* iron and calcium bioavailability and *in vitro* starch and protein digestibility of some legume seeds. Food Sci Technol 40: 1292–1299.
- Glazebrook J (1999) Genes controlling expression of defense responses in *Arabidopsis*. Curr Opin Plant Biol 2: 280–286.
- Graham MA, Fredrick Marek L, Lohnes D, Cregan P, Shoemaker RC (2000). Expression and genome organization of resistance gene analogs in soybean. Genome 43: 86-93.
- Gujaria-Verma N, Vail SL, Carrasquilla-Garcia N, Penmetsa R, Cook DR, Farmer AD, Vandenberg A, Bett KE (2014) Genetic mapping of legume orthologs reveals high conservation of synteny between lentil species and the sequenced genomes of *Medicago* and chickpea. Front Plant Sci 5: 676.
- Gunn CR (1969) Genera, types and lectotypes in the tribe Vicieae (Fabaceae). Taxon 18: 725–733.
- Gunn CR, Kluve J (1976) Androecium and pistil characters for Tribe *Vicieae (Fabaceae)*. Taxon 25: 563–575.
- Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich-repeat-encoding genes in *Arabidopsis*. Plant Physiol 157: 757–769.
- Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48: 575–607.
- Harlan JR, De Wet JMJ (1971) Towards a rational classification of cultivated plants. Taxon 20: 509–517.
- Heidrich K, Blanvillain-Baufumé S, Parker JE (2012) Molecular and spatial constraints on NB-LRR receptor signaling. Curr Opin Plant Biol 15: 385-391.

- Heidrich K, Wirthmueller L, Tasset C, Pouzet C, Deslandes L, Parker JE (2011) *Arabidopsis* EDS1 connects pathogen effector recognition to cell compartment-specific immune responses. Science 334: 1401–1404.
- Holt BF (2005) Antagonistic control of disease resistance protein stability in the plant immune system. Science 309: 929-932.
- Horneburg B (2006) Outcrossing in lentil (*Lens culinaris*) depends on cultivar, location and year, and varies within cultivars. Plant Breed 125: 638–640.
- Hougaard BK, Madsen LH, Sandal N, Moretzsohn MC, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, Bertioli DJ, Stougaard J (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris, Lotus japonicus, Medicago truncatula* and *Arachis*. Genetics 179: 2299–2312.
- Hu G, de Hart AK, Li Y, Ustach C, Handley V, Navarre R, Hwang CF, Aegerter BJ, Williamson VM, Baker B (2005) EDS1 in tomato is required for resistance mediated by TIR-class *R* genes and the receptor-like R gene *Ve*. Plant J 42: 376–391.
- Huettel B, Santra D, Muehlbauer FJ, Kahl G (2002) Resistance gene analogues of chickpea (*Cicer arietinum*): isolation, genetic mapping and association with *Fusarium* resistance gene cluster. Theor Appl Genet 105: 479-490.
- Hunger S, Di Gaspero G, Möhring S, Bellin D, Schäfer-Pregl R, Borchardt DC, Charles-Durel E, Werber M, Weisshaar B, Salamini F, Schneider K (2003) Isolation and linkage analysis of expressed disease resistance gene analogues of sugar beet (*Beta vulgaris* L). Genome 46: 70-82.
- Iqbal A, Khalil IA, Ateeq N, Khan MS (2006) Nutritional quality of important food legumes. Food Chem 97: 331–335
- Jacob F, Vemaldi S, Maekawa T (2013) Evolution and conservation of plant NLR functions. Front Immunol 4: 297.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, Daniell H (2008) Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (*Leguminosae*). Mol Phylogenet Evol 48: 1204–1217.
- Jia YX, Yuan Y, Zhang Y, Yang S, Zhang X (2015) Extreme expansion of NBS-enconding genes in *Rosaceae*. BMC Genet 16: 48.
- Jones JDG, Dangl JL (2006) The plant immune system. Nature 444: 323-329.
- Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JDG, Hein I (2012) Identification and localisation of the *NB-LRR* gene family within the potato genome. BMC Genomics 13: 75.
- Kadota Y, Amigues B, Ducassou L, Madaoui H, Ochsenbein F, Guerois R, Shirasu K (2008) Structural and functional analysis of SGT1–HSP90 core complex required for innate immunity in plants. EMBO Rep 9: 1209-1215.

- Kahl G (2004) The Dictionary of Gene Technology: Genomics, Transcriptomics, Proteomics. Volume 2. Third Edition. Wiley-VCH.
- Kalo P, Seres A, Taylor S, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis T, Kiss G (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genom 272: 235–246.
- Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad. Sci USA 93: 11746-11750.
- Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, Van K, Lee SH (2012) Genome-wide mapping of *NBS-LRR* genes and their association with disease resistance in soybean. BMC Plant Biol 12: 139.
- Kang HG, Kuhl JC, Kachroo P, Klessig DF (2008) CRT1, an *Arabidopsis* ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. Cell Host Microbe 3: 48-57.
- Kang HG, Oh CS, Sato M, Katagiri F, Glazebrook J, Takahashi H, Kachroo P, Martin GB, Klessig DF (2010) Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in *Arabidopsis* and tobacco. Plant Cell 22: 918-936.
- Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics 12: 265.
- Kim SH, Gao F, Bhattacharjee S, Adiasor JA, Nam JC, Gassmann W (2010) The *Arabidopsis* resistance-like gene *SNC1* is activated by mutations in *SRFR1* and contributes to resistance to the bacterial effector AvrRps4. PLoS Pathog 6: e1001172.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Molec Evol 16: 111-120.
- Kwon SI, Kim SH, Bhattacharjee S, Noh J-J, Gassmann W (2009) SRFR1, a suppressor of effectortriggered immunity, encodes a conserved tetratricopeptide repeat protein with similarity to transcriptional repressors. Plant J 57: 109-119.
- 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 Linn Soc 123: 257-260.
- Ladizinsky G (1999) Identification of the lentil's wild genetic stock. Genet Res Crop Evol. 46: 115–118.
- Ladizinsky G, Braun D, Goshen D, Muehlbauer FJ (1984) The biological species of the genus *Lens*. Bot Gaz 145: 253–261.
- Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. Syst Biol 54, 575–594.

- Legume Information System (LIS): http://cvit.comparative-legumes.org/. Datos revisados el día 10-11-2015.
- Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet 14: 421-429.
- Lev-Yadun S, Gopher A, Abbo S (2000) The cradle of agriculture. Science 288: 1602–1603.
- Li X, Cheng Y, Ma W, Zhao Y, Jiang, H, Zhang M (2010a) Identification and characterization of NBS-encoding disease resistance genes in *Lotus japonicus*. Plant Syst Evol 289, 101–110.
- Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen J-Q, Tian D, Yang S (2010b) Unique evolutionary pattern of numbers of gramineous *NBS–LRR* genes. Mol Genet Genom 283: 427–38.
- Li Y, Li S, Bi D, Cheng YT, Li X, Zhang Y (2010c) SRFR1 negatively regulates plant NB-LRR resistance protein accumulation to prevent autoimmunity. PLoS Pathog 6: e1001111.
- Lopez CE, Acosta LF, Jara C, Pedraza F, Gaitán-Solis E, Gallego G, Beebe S, Tohme J (2003) Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathology 93: 88-95.
- Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015) Identification and distribution of the NBS-LRR gene family in the cassava genome. BMC Genomics 16: 360.
- Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G (2012) Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group Phureja. PLoS One 7: e34775.
- Lukasik E, Takken FLW (2009) STANDing strong, resistance proteins instigators of plant defence. Curr Opin Plant Biol 12: 427-436.
- Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H (2012) Dynamic nucleotidebinding-site and leucine-rich-repeat-encoding genes in the grass family. Plant Physiol 159: 197– 210.
- Maekawa T, Cheng W, Spiridon LN, Töller A, Lukasik E, Saijo Y, Liu P, Shen QH, Micluta MA, Somssich IE, Takken FLW, Petrescu AJ, Chai J, Schulze-Lefert P (2011a) Coiled-coil domaindependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. Cell Host Microbe 9: 187-199.
- Maekawa T, Kufer TA, Schulze-Lefert P (2011b) NLR functions in plant and animal immune systems: so far and yet so close. Nat Immunol 12: 817–826.
- MAGRAMA: Ministerio de Agricultura, Alimentación y Medio Ambiente. www.magrama.gob.es . Datos revisados el día 25-04-2015.

- Marone D, Russo M, Laidò G, De Leonardis A, Mastrangelo A (2013) Plant nucleotide binding site–leucine-rich repeat (*NBS-LRR*) genes: active guardians in host defense responses. Int J Mol Sci 14: 7302–26.
- Mayer MS, Bagga SK (2002) The phylogeny of *Lens* (*Leguminosae*): new insight from ITS sequence analysis. Plant Syst Evol 232: 145–154.
- McCarthy C (1998) Chromas 1.45. School of Health Science, Griffith University, Southport.
- McDowell JM, Woffenden BJ (2003) Plant disease resistance genes: Recent insights and potential applications. Trends Biotechnol 21: 178–183.
- McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7: 212.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20: 317-332.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15: 809-834.
- Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8: 1113-1130.
- Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor. New York, U.S.A.
- Moffet MD (2006) Identifying regions of conserved synteny between pea (*Pisum spp.*), lentil (*Lens spp.*) and bean (*Phaseolus spp.*). Tesis doctoral. Montana State University. Bozeman, Montana.
- Moffett P, Farnham G, Peart J, Baulcombe DC (2002) Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. Embo J 21: 4511-4519.
- Monaghan J, Zipfel C (2012) Plant pattern recognition receptor complexes at the plasma membrane. Curr Opin Plant Biol 15: 349–357.
- Mondragon-Palomino M, Gaut BS (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Mol Biol Evol 22: 2444–2456.
- Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109: 1434-1447.
- Muehlbauer FJ, Kaiser WJ, Clement SL, Summerfield RJ (1995) Production and breeding of lentil. Adv Agron 54: 283–332.
- Muehlbauer FJ, Weeden NF, Hoffman DL (1989) Inheritance and linkage relationships of morphological and isozyme loci in lentil (*Lens* Miller). J Hered 80: 298–303.

- Mullis KB, Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Method Enzymol 155: 335-350.
- Mullis KB, Faloona F, Scharf S, Saiki R, Horn GT, Erlich HA (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Sym 51: 263-273.
- Muskett PR, Kahn K, Austin MJ, Moisan LJ, Sadanandom A, Shirasu K, Jones JDG, Parker JE (2002) *Arabidopsis* RAR1 exerts rate-limiting control of *R* gene-mediated defenses against multiple pathogens. Plant Cell 14: 979-992.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418-426.
- Nepal MP, Benson BV (2015) CNL disease resistance genes in soybean and their evolutionary divergence. Evol Bioinform 11: 49-63.
- Neyra CD (2015) Estudio de genes codificantes de proteínas de reserva, mapeo genético, análisis de QTLs y estudio de genes de resistencia a *Ascochyta lentis* en *Lens*. Tesis doctoral. Universidad de León. León.
- Noir S, Combes MC, Anthony F, Lashermes P (2001) Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea L*). Mol Genet Genom 265: 654-662.
- Ohme-Takagi M, Suzuki K, Shinshi H (2000) Regulation of ethylene-induced transcription of defense genes. Plant Cell Physiol 41: 1187–1192
- Palomino C, Fernández-Romero MD, Rubio J, Torres A, Moreno MT, Millán T (2009) Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance. Theor Appl Genet 118: 671-682.
- Palomino C, Satovic Z, Cubero JL, Torres AM (2006) Identification and characterization of NBS-LRR class resistance gene analogs in faba bean (*Vicia faba* L.) and chickpea (*Cicer arietinum* L.). Genome 49: 1227-1237.
- Pan Q, Jonathan W, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50: 203-213.
- Peart JR, Cook G, Feys BJ, Parker JE, Baulcombe DC (2002) An EDS1 orthologue is required for Nmediated resistance against tobacco mosaic virus. Plant J 29: 569–579
- Pérez de la Vega M, Fratini RM, Muehlbauer FJ (2011) Lentil. En: M Pérez de la Vega, AM Torres, JL Cubero, C Kole (eds) Genetics, genomics and breeding of cool season grain legumes. Boca Raton, FL: CRC Press; Jersey, British Isles; Enfield, NH: Science Publishers, pp: 98-150.
- Peterson GI, Masel J (2009) Quantitative Prediction of Molecular Clock and K-a/K-s at Short Timescales. Mol Biol Evol. 26: 2595–603.
- Phan HT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris*. Theor Appl Genet 114: 549-558.

- Phan HT, Ellwood SR, Moolhuijen P, Williams A, Hane JK, Ford R, Bellgard M, Oliver RP (2005) Genomic synteny in legumes: Application to crop breeding. En: 2<sup>nd</sup> Australian Model Legume Workshop, 5-8 April 2005, Rottnest Island, Western Australia.
- Plant Virus Online: http://pvo.bio-mirror.cn/famly078.htm#Lens%20culinaris . Datos revisados el día 01-10-2014.
- Podlaha O, Zhang J (2010) Pseudogenes and their evolution. En: Encyclopedia of life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
- Promega Corporation (2007) Technical Manual. pGEM®-T and pGEM®-T Easy Vector Uses. Madison. USA. Revised 5/07.
- Qi D, Innes RW (2013) Recent advances in plant NLR structure, function, localization, and signaling. Front Immunol 4: 348.
- Rairdan GJ, Moffett P (2006) Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. Plant Cell 18: 2082–2093.
- Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, Kim C, Tang H, Paterson AH (2011) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. New Phytol 192: 164-178.
- Reddy AC, Venkat S, Singh TH, Aswath C, Reddy KM, Reddy DCL (2015) Isolation characterization and evolution of NBS-LRR encoding disease-resistance gene analogs in eggplant against bacterial wilt. Eur Plant Pathol 143: 417-426.
- Rietz S, Stamm A, Malonek S, Wagner S, Becker D, Medina- Escobar N, Vlot AC, Feys BJ, Niefind K, Parker JE (2011) Different roles of Enhanced Disease Susceptibility1 (EDS1) bound to and dissociated from Phytoalexin Deficient4 (PAD4) in *Arabidopsis* immunity. New Phytol 191: 107-119.
- Rispail N, Kaló P, Kiss GB, Ellis N, Gallardo K, Thompson RD, Prats E, Larrainzar E, Ladrera R, González EM, Arrese-Igor C, Ferguson BJ, Gresshoff PM, Rubiales D (2010) Model legumes contribute to faba bean breeding. Field Crops Res 115: 253–269
- Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42: 41-47.
- Roy F, Boye JI; Simpson BK (2010) Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. Food Res Int 43: 432–442.
- Rubiales D, Fernandez-Aparicio M (2011) Broomrape management in lentils. Grain Legumes 57: 37-38.
- Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi- Shinozaki K (2006) Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proc Natl Acad Sci USA 103: 18822–18827.

- Sanz MJ, Loarce Y, Fominaya A, Vossen JH, Ferrer E (2013) Identification of RFLP and NBS/PK profiling markers for disease resistance *loci* in genetic maps of oats. Theor Appl Genet 126 1: 203-218.
- Saraste M, Sibbald PR, Wittinghofer A (1990) The p-loop-a common motif in ATP- and GTPbinding proteins. Tr Biochem Sci 15: 430-434.
- Sekhwal MK, Li P, Lam I, Wang X, Cloutier S, You FM (2015) Disease resistance gene analogs (RGAs) in plants. Int J Mol Sci 16: 19248-19290.
- Sela H, Spiridon LN, Petrescu AJ, Akerman M, Mandel-Gutfreund Y, Nevo E, Loutre C, Keller B, Schulman AH, Fatima T (2012) Ancient diversity of splicing motifs and protein surfaces in the wild emmer wheat (*Triticum dicoccoides*) LR10 coiled coil (CC) and leucine-rich repeat (LRR) domains. Mol Plant Pathol 13: 276-287.
- Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z, Xu J, Cheng Z, Wan J, Zhu L (2009) Identification of a new rice blast resistance gene, *Pid3*, by genome wide comparison of paired nucleotide-binding site leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. Genetics 182: 1303–1311.
- Shao ZQ, Zhang YM, Hang YY, Xue JY, Zhou GC, Wu P, Wu XY, Wu XZ, Wang Q, Wang B, Chen JQ (2014) Long-Term Evolution of Nucleotide-Binding Site-Leucine-Rich Repeat Genes: Understanding Gained from and beyond the Legume Family. Plant Physiol 166: 217–34.
- Sharpe AG, Ramsay L, Sanderson LA, Fedoruk MJ, Clarke WE, Li R, Kagale S, Vijayan P, Vandenberg A, Bett KE (2013) Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genomics 14: 192.
- Simon CJ, Muehlbauer FL (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88: 115-119.
- Sinapidou E, Williams K, Nott L, Bahkt S, Tör M, Crute I, Bittner-Eddy P, Beynon J (2004) Two TIR: NB: LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. Plant J 38: 898–909.
- Skibinski DOF, Rasool D, Erskine W (1984) Aspartate aminotransferase allozyme variation in a germplasm collection of the domesticated lentil (*Lens culinaris*). Theor Appl Genet 68: 441-448.
- Slootweg EJ, Spiridon LN, Roosien J, Butterbach P, Pomp R, Westerhof L, Wilbers R, Bakker E, Bakker J, Petrescu AJ, Smant G, Goverse A (2013) Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. Plant Physiol 162: 1510–1528.
- Song H, Nan Z (2014) Genome-wide analysis of nucleotide-binding site disease resistance genes in *Medicago truncatula*. Chin Sci Bull 59: 1129-1138.
- Sonnante G, Galasso I, Pignone D (2003) ITS sequence analysis and phylogenetic inference in the genus *Lens* Mill. Ann Bot 91: 49–54.

- StatSoft, Inc (1995) STATISTICA for Windows [Computer program manual]. Tulsa, OK: Statsoft, Inc, 2300 East 14th street, Tulsa; OK, 74104-4442.
- Steele KP, Wojciechowski MF (2003) Phylogenetic analyses of tribes *Trifolieae* and *Vicieae*, based on sequences of the plastid gene, *matk* (*Papilionoidae: Leguminosae*). En: BB. Klitgaard, A. Bruneau (eds) Advances in Legume Systematics, part 10, Higher Level Systematic. Royal Botanic Gardens, Kew, pp 355-370.
- Stevenson PC, Dhillon MK, Sharma HC, El-Bouhssini M (2007) Insect pest of lentil and their management. En: SS Yadav, D McNeil, PC Stevenson (eds) Lentil. An Ancient Crop for Modern Times. Springer, Dordrecht, The Netherlands, pp 331–348.
- Summerfield RJ, Roberts EH (1988) Photothermal regulation of flowering in pea, lentil and chickpea. En: World Crops: Cool Season food legume. Curr Plant Sci Biot 5: 911-922
- Sun Q, Collins NC, Ayliffe M, Smith SM, Drake J, Pryor T, Hulbert SH (2001) Recombination between paralogues at the *Rp1* rust resistance locus in maize. Genetics 158: 423-438.
- Sun X, Pang H, Li M Chen J, Hang Y (2014) Tracing the origin and evolution of plant TIR-encoding genes. Gene 546: 408-416.
- Takken FLW, Albrecht M, Tameling WI (2006) Resistance proteins: molecular switches of plant defense. Curr Opin Plant Biol 9: 383-390.
- Takken FLW, Goverse A (2012) How to build a pethogen detector: structural basis of NB-LRR function. Curr Opin Plant Biol 15: 375-384.
- Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL, Haring MA, Cornelissen BJ (2002) The tomato *R* gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell 14: 2929–2939.
- Tameling WI, Takken FLW (2008) Resistance proteins: scouts of the plant innate immune system. Eur J Plant Pathol 121: 243-255.
- Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJ, Takken FLW (2006) Mutations in the NB-ARC domain of 1-2 that impair ATP hydrolysis cause autoactivation. Plant Physiol 140: 1233-1245.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.
- Tarr DEK, Alexander HM (2009) TIR- NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2: 197.
- Thompson JD, Higgins DG, Gibson TJ (1994) 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.

- Timmerman-Vaughan GM, Frew TJ, Weeden NF (2000) Characterization and linkage mapping of *R*-gene analogous DNA sequences in pea (*Pisum sativum L*.). Theor Appl Genet 101: 241-247.
- Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 222: 9–19.
- Tutar Y (2012) Pseudogenes. Comp Funct Genom 2012: 6-9.
- Urbano G, Porres, JM, Frías J, Vidal-Valverde C (2007) Nutricional value. En: SS Yadav, D McNeil, PC Stevenson (eds) Lentil. An Ancient Crop for Modern Times. Springer, Dordrecht, The Netherlands, pp 47–93.
- Van der Biezen EA, Jones JD (1998a) The NB-ARC domain: a novel signalling motif shared by plant resistance genes products and regulators of cell death in animals. Curr Biol 8: 226-227.
- Van der Biezen EA, Jones JD (1998b) Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci 23: 454-456.
- Van der Hoorn RA, Kamoun S (2008) From Guard to Decoy: a new model for perception of plant pathogen effectors. Plant Cell 20: 2009-2017.
- Van der Linden CG, Wouters D, Mihalka V, Kochieva E, Smulders M, Vosman B (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet 109: 384-393.
- Van Ooijen G, Lukasik E, Van Den Burg HA, Vossen JH, Cornelissen BJ, Takken FLW (2010) The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2. Plant J 63: 563-572.
- Van Ooijen G, Mayr G, Albrecht M, Cornelissen BJ, Takken FLW (2008a) Transcomplementation, but not physical association of the CC-NB-ARC and LRR domains of tomato R protein Mi-1.2 is altered by mutations in the ARC2 subdomain. Mol Plant 1: 401-410.
- Van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJ, Takken FLW (2008b) Structurefunction analysis of the NB-ARC domain of plant disease resistance proteins. J Exp Bot 59: 1383-1397.
- Van Ooijen G, van den Burg HA, Cornelissen B.J, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol 45: 43–72.
- Van Oss H, Aron Y, Ladizinsky G (1997) Chloroplast DNA variation and evolution in the genus *Lens* Mill. Theor Appl Genet 94: 452–457.
- Varshney RK, Song, C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31: 240–246.

- Ve T, Williams SJ, Stamp A, Valkov E, Dodds PN, Anderson PA, Kobe B (2011) Crystallization and X-ray diffraction analysis of the C-terminal domain of the flax rust effector protein AvrM. Acta Crystallogr Sect F Struct Biol Cryst Commun 67: 1603-1607.
- Vidhyasekaran P (2007) Fungal Pathogenesis in Plants and Crops: Molecular Biology and Host Defense Mechanisms. Second Edition. CRC Press, Taylor & Francis Group.
- Walker JE, Saraste M, Runswich MJ, Gay NJ (1982) Distantly related sequences in the  $\alpha$  and  $\beta$ subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1: 945-951.
- Wan H, Yuan W, Ye Q, Wang R, Ruan M, Li Z, Zhou G, Yao Z, Zhao J, Liu S, Yang Y (2012) Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: Characterization, genetic variation, functional divergence and expression patterns. BMC Genomics 13: 502.
- Wan H, Zhao Z, Malik AA, Qian C, Chen J (2010) Identification and characterization of potential NBS-encoding resistance genes and induction kinetics of a putative candidate gene associated with downy mildew resistance in Cucumis. BMC Plant Biol 10: 186.
- Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83: 123–129.
- Wilson VE, Law AG (1972) Natural crossing in *Lens esculenta* Moench. J Am Soc Hortic Sci 97: 142-143.
- Wojciechowski MF (2006) IRLC (Inverted Repeat-lacking clade). Version 11 July 2006: http://tolweb.org/IRLC\_(Inverted\_Repeat-lacking\_clade)/60358/2006.07.11 in The Tree of Life Web Project, http://tolweb.org/.
- Wong MM, Gujaria-Verma N, Ramsay L, Yuan HY, Caron C, Diapari M, Vandenberg A, Bett KE (2015) Classification and characterization of species within the genus *Lens* using Genotyping-by-Sequencing (GBS). PLoS One 10 (3): e0122025.
- Xu Q, Wen X, Deng X (2005) Isolation of TIR and nonTIR NBS–LRR resistance gene analogues and identification of molecular markers linked to a powdery mildew resistance locus in chestnut rose (*Rosa roxburghii* Tratt). Theor Appl Genet 111: 819–830.
- Yaish MWF, Sáez de Miera LE, Pérez de la Vega M (2004) Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (NBS) type from *Lens* species. Genome 47: 650-659.
- Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genom 280: 187–198.
- Young ND (2000) The genetic architecture of resistance. Curr Opin Plant Biol 3: 285-290.
- Yu YG, Buss GR, Sahai-Maroof MA (1996) Isolation of a superfamily of candidate diseaseresistance gene in soybean based on a conserved nucleotide- binding site. Proc Natl Acad Sci USA 93: 11751-11756.

- Yue JX, Meyers BC, Chen JQ, Tian DC, Yang SH (2012) Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol 193: 1049–1063.
- Zhang X (2003) Characterization of the *PIF/Pong* superfamily of DNA transposons and their relationship with *tourist*-like miniature inverted-repeat transposable elements (MITES). Tesis doctoral. The University of Georgia, Athens, Georgia.
- Zhang N, Wang S, Wang H, Liu D (2011) Isolation and characterization of NBS-LRR class resistance homologous gene from wheat. Agr Sci China 10: 1151-1158.
- Zhong Y, Yin H, Sargent DJ, Malnoy M, Cheng ZM (2015) Species-specific duplications driving the recent expansion of NBS-LRR genes in five *Rosaceae* species. BMC Genomics 16: 77.
- Zhou T, Wang Y, Chen J-Q, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genom 271: 402–15.
- Zhu H, Cannon SB, Young ND, Cook DR (2002) Phylogeny and genomic organization of the TIR and non-TIR-NBS-LRR resistance gene family in *Medicago truncatula*. Mol Plant Microbe Int 15: 529-539.
- Zhu SF, Jeong RD, Venugopal SC, Lapchyk L, Navarre D, Kachroo A, Kachroo P (2011) SAG101 forms a ternary complex with EDS1 and PAD4 and is required for resistance signaling against Turnip Crinkle Virus. PLoS Pathog 7 (11): e1002318.
- Zohary D (1999) Monophyletic vs. polyphyletic origin of crops on which agriculture founded in the Near East. Genet Resour Crop Evol 46: 133–142.
- Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR Shiu (2009) Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. Plant Physiol 151: 3–15.